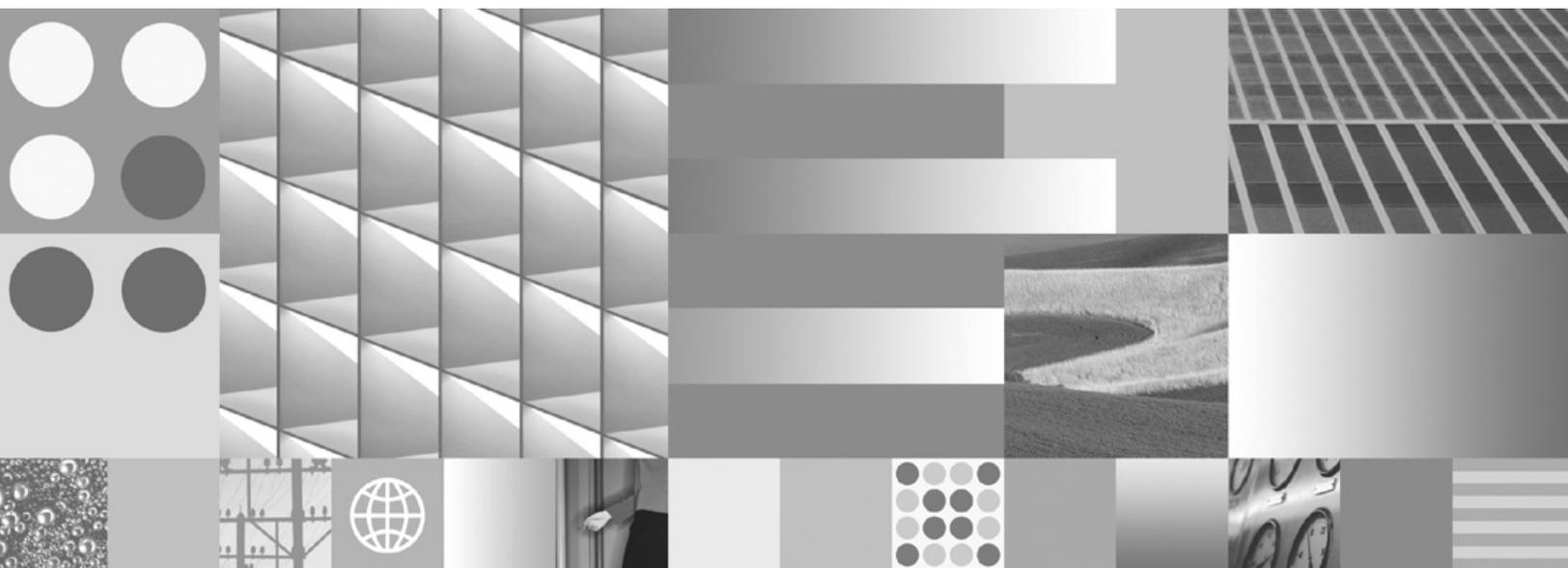


Управление поисковой системой предприятия



Управление поисковой системой предприятия

Замечание

Перед тем как использовать данный документ и продукт, описанный в нем, прочтите информацию под заголовком “Замечания и товарные знаки” на стр. 423.

Замечания по выпуску

Это издание применимо к версии 8, выпуску 5, модификации 0 системы IBM OmniFind Enterprise Edition (номер продукта 5724-C74) и ко всем последующим выпускам и модификациям, пока в новых изданиях не будет указано иное.

Отсылая информацию IBM, вы тем самым даете IBM неисключительное право использовать или распространять эту информацию любым способом, как фирма сочтет нужным, без каких-либо обязательств перед вами.

© Copyright International Business Machines Corporation 2004, 2008. Все права защищены.

Содержание

ibm.com и соответствующие ресурсы vii

Как послать свои комментарии vii

Как связаться с IBM viii

Что такое поисковая система предприятия? 1

Типы источников данных, поддерживаемые поисковой системой предприятия 2

Обзор компонентов поисковой системы предприятия 3

Искатели поисковой системы предприятия 4

Анализаторы поисковой системы предприятия 5

Индексы поисковой системы предприятия 7

Серверы поиска для поисковой системы предприятия 8

Консоль администратора поисковой системы предприятия 9

Слежение за поисковой системой предприятия 10

Файлы журналов поисковой системы предприятия 10

Настройка поисковой системы предприятия 11

Пример программы поиска для поисковой системы предприятия 12

Поток данных поисковой системы предприятия 12

Управление поисковой системой предприятия 15

Регистрация на консоли администратора 19

Изменение пароля администратора поисковой системы предприятия для конфигурации с одним сервером 20

Изменение пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами 21

Номера портов TCP для поисковой системы предприятия 23

Изменение номера порта для поисковой системы предприятия 24

Изменение имен хостов или IP-адресов поисковой системы предприятия 26

Конфигурирование поддержки двух IP-адресов 27

Включение поддержки протокола IPv6 28

Собрания поисковой системы предприятия 33

Создание собрания при помощи мастера по собраниям 33

Создание собрания при помощи окна Собрания 35

Изменение собрания 36

Удаление собрания 38

Определение ID собрания 38

Управление искателями 41

Создание искателя 43

Изменение свойств искателя 44

Изменение пространства для искателя 44

Удаление искателя 45

Расписания искателя 45

Искатели Content Edition 46

Доступ к репозиториям Content Edition в непосредственном режиме 47

Доступ к репозиториям WebSphere II Content Edition в режиме сервера 48

Искатели DB2 51

Конфигурирование сервера искателей в UNIX для искателей DB2 53

Конфигурирование сервера искателей в Windows для искателей DB2 54

Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2 55

Конфигурирование WebSphere MQ для искателей DB2 57

Просмотр баз данных DB2 на сервере источника данных Classic 59

Искатели DB2 Content Manager 60

Конфигурирование сервера искателей в UNIX для искателей DB2 Content Manager 61

Конфигурирование сервера искателей в Windows для искателей DB2 Content Manager 63

Искатели Domino Document Manager 64

Искатели Exchange Server 66

Искатели баз данных JDBC 67

Отображения взаимосвязей для баз данных JDBC 69

Просмотр искателем нескольких структурированных таблиц баз данных JDBC 71

Искатели NNTP 75

Искатели Notes 76

Советы по просмотру баз данных Lotus Domino 78

Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino 79

Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino 81

Конфигурирование серверов, использующих протокол DIIOP 83

Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino 84

Искатели QuickPlace 86

Искатели с начальным списком 89

Искатели файловой системы UNIX 91

Искатели Web 91

Конфигурация пользовательского агента 92

Как искатель Web использует протокол Robots Exclusion 93

Поддержка JavaScript 95

Правила ограничения пространства для искателя Web 96

Проверка соединений URL для искателя Web 100

Задание интервала повторного просмотра для искателя Web 101

Опции посещения URL искателем Web 101

Как искатель Web обрабатывает произвольные страницы ошибок	102
Поддержка просмотра защищенных сайтов.	103
Сайты с доступом через прокси-серверы	106
Управление опознавательными файлами (cookie)	107
Конфигурирование глобального пространства для искателя Web	108
Директивы no-follow и no-index	110
Переопределение директив no-follow и no-index на Web-страницах	111
Конфигурирование дат, которые будут использоваться искатель Web для обрабатываемых им документов	111
Искатели Web Content Management	113
Искатели WebSphere Portal	114
Копирование URL для просмотра искателем из WebSphere Portal	116
Искатели файловой системы Windows	117
Конфигурирование поддержки для программ приема данных	119
Пользовательские подключаемые модули искателей	120
Поддержка просмотра архивных файлов искателем	122
Форматы URI для индекса поисковой системы предприятия	123

Управление анализаторами 133

Работа с категориями	134
Категории на основе правил	134
Деревья категорий	136
Выбор типа категоризации	137
Конфигурирование категорий	138
Работа с полями поиска XML	139
Поля поиска XML	139
Отображение элементов XML на поля поиска	140
Работа с полями поиска HTML	143
Поля поиска HTML	143
Отображение элементов метаданных HTML на поля поиска	144
Пользовательская обработка текста	145
Добавление механизмов анализа текста в систему	146
Связывание механизма текстового анализа с собранием	147
Отображение элементов XML на общую структуру анализа	148
Отображение общей структуры анализа на индекс	150
Отображение общей структуры анализа на реляционную базу данных	151
Конфигурирование потоков для службы анализатора	152
Включение расширенного анализа для составных терминов	152
Включение поддержки собственного поиска XML	153

Определение форматов документов 155

Поддерживаемые типы документов по умолчанию	156
Типы документов, связанные с анализаторами собраний и анализаторами Stellent	157
Связывание типов документов с анализатором собраний	157

Правила по умолчанию для служб анализаторов собраний	159
Анализ документов неизвестных типов	160
Изменение правил замены для некоторых тегов HTML	161
Правила замены HTML по умолчанию	162
Связывание типов документов с анализатором Stellent.	163
Правила анализа по умолчанию для анализаторов Stellent	165

Поддержка языков и кодовых страниц. 169

Автоматическое определение языка	170
Автоматическое определение кодовой страницы	171
Лингвистический анализ документов на китайском, японском и корейском языках	172
Сегментация энграмм.	172
Удаление пробелов из текста	172

Управление индексами. 175

Планирование операций построения индекса	176
Изменение расписания индексирования	177
Включение и выключение расписаний индекса	177
Конфигурирование одновременных операций построения индекса	178
Построение индексов только при обнаружении изменений	179
Остановка построения индексов	181
Опции, влияющие на представление индекса для поиска	181
Индексированные опции для поиска документов	182
Обнаружение дубликатов документов	186
Символы подстановки в запросах	187
Области действия	191
Конфигурирование областей действия	191
Сворачивание URI	192
Объединение URI в результатах поиска	193
Удаление URI из индекса.	194

Управление сервером поиска 197

Кэши поиска	198
Конфигурирование кэша поиска	198
Пользовательские словари синонимов	198
Добавление словарей синонимов в систему	200
Связывание словаря синонимов с собранием	201
Пользовательские словари стоп-слов	201
Добавление словарей стоп-слов в систему	202
Связывание словаря стоп-слов с собранием.	203
Повторное внедрение пользовательских словарей	203
Динамическое составление сводок	205
Настройка сводок документов на консоли администратора	205
Настройка сводок документов при помощи редактирования свойств	206
Работа с прямыми ссылками	207
Прямые ссылки	207
Конфигурирование прямых ссылок	207

Ранжирование документов 209

Назначение оценок на основе текста	209
Статическое ранжирование	210
Восстановление значений по умолчанию для статического ранжирования документов.	211
Пользовательские словари значимых слов	213
Добавление словарей значимых слов в систему	214
Связывание словаря значимых слов с собранием	215
Ранжирование документов на основе шаблонов URI	215
Влияние на оценки документов, соответствующих шаблону URI	216
Ранжирование документов на основе классов значимости	217
Отображение полей на классы значимости	219
Конфигурирование коэффициентов значимости для классов значимости	220
Значения классов значимости по умолчанию	220

Прикладные программы поиска для поисковой системы предприятия . . . 223

Задание прикладных программ поиска для собраний	224
Функции примера программы поиска.	224
Свойства программы поиска	226
Редактирование свойств примера программы поиска	243
Настройка программы поиска	244
Клонирование примера программы поиска	246
Анализ частых результатов	247
Обращение к программе поиска	251
Конфигурирование серверов поиска для приема только защищенных требований (по протоколу SSL).	252
Конфигурирование серверов поиска для приема требований через прокси-сервер	253

Поддержка внешних источников 255

Добавление внешних источников в систему.	255
Задание прикладных программ поиска для внешних источников	258

Защита поисковой системы предприятия 259

Защита при установке.	260
Сравнение аутентификации и управления доступом	261
Административные роли.	262
Конфигурирование пользователей-администраторов	263
Защита уровня собрания	264
Анализ повторных документов и защита собрания	264
Идентификаторы программ поиска	264
Защита на уровне документов	265
Предварительная фильтрация и постфильтрация результатов поиска	266
Проверка на основе хранимых маркеров защиты	267
Проверка текущих параметров аутентификации при обработке запроса	268
Анализ привязок текста	276
Включение защиты для поисковой системы предприятия	277

Конфигурирование глобальной защиты и реестра пользователей LDAP в WebSphere Application Server	278
Включение защиты для поисковой системы предприятия с одним сервером.	280
Включение защиты для поисковой системы предприятия с несколькими серверами	281
Требования к конфигурации искателя для поддержки защиты	282
Проверка доступа к защищенным документам Exchange Server	285
Обеспечение защиты уровня документов для документов Lotus Domino	286
Обеспечение защиты уровня документов для документов файловых систем Windows	289
Отключение защиты для поисковой системы предприятия	292
Отключение защиты для прикладной программы предприятия в WebSphere Application Server.	292
Отключение защиты на уровне документов.	294
Отключение защиты для объединенных результатов поиска	295

Запуск и остановка поисковой системы предприятия 297

Запуск поисковой системы предприятия	297
Остановка поисковой системы предприятия	299
Управление набором запускаемых или останавливаемых компонентов.	301
Управление серверами поиска в автономном режиме	302

Слежение за активностью поисковой системы предприятия . . . 305

Оценка числа документов в собрании.	305
Слежение за собранием	306
Просмотр подробностей об URI	306
Отслеживание работы искателей	308
Просмотр подробностей об активности искателя Web	309
Подробности потока искателя Web	310
Активные сайты искателя Web	310
Скорость просмотра искателя Web	311
Создание отчетов искателя Web	312
Код состояния HTTP, возвращаемые искателем Web	313
Отслеживание работы анализатора	317
Отслеживание активности индекса для собрания	318
Слежение за очередью индексации поисковой системы предприятия	319
Отслеживание работы серверов поиска	320
Изменение способа подсчета статистики.	321
Слежение за работой системы приема данных	322
Отслеживание документов	323
Конфигурирование файлов журнала для отслеживания документов	323
Просмотр отчетов об отброшенных документах	324
Просмотр файлов журналов отброшенных документах	325

Файлы журналов и оповещения . . . 327

Оповещения	327
Конфигурирование оповещений уровня собрания	328
Конфигурирование оповещений уровня системы	329
Конфигурирование файлов журнала	330
Конфигурирование информации о SMTP-сервере	332
Получение сообщений журнала по электронной	
почте	332
Изменение размера журнала запросов	334
Просмотр файлов журнала	335

Резервное копирование и восстановление поисковой системы предприятия 337

Резервное копирование поисковой системы	
предприятия	338
Восстановление поисковой системы предприятия	339
Экспорт и импорт конфигураций собраний	340

Интеграция с Lotus Notes Версии 8 345

Создание сайта обновления подключаемых модулей	
для поисковой системы предприятия	345
Установка подключаемого модуля поисковой	
системы предприятия в клиенте Lotus Notes версии 8 .	345

Интеграция с WebSphere Portal . . . 347

Сценарии конфигурирования для интеграции	
поисковой системы предприятия с WebSphere Portal .	348
Конфигурирование поисковой системы предприятия	
в WebSphere Portal версии 5.1	349
Конфигурирование панели поиска WebSphere	
Portal Версии 5.1 для использования поисковой	
системы предприятия	352
Удаление поисковой системы предприятия из	
WebSphere Portal Версии 5.1	353
Конфигурирование поисковой системы предприятия	
в WebSphere Portal версии 6	354
Конфигурирование Центра поиска WebSphere	
Portal Версии 6 для поисковой системы	
предприятия	357
Конфигурирование панели поиска WebSphere	
Portal Версии 6 для использования поисковой	
системы предприятия	358

Конфигурирование портлета поисковой системы	
предприятия для Lotus Quickr	360
Удаление поисковой системы предприятия из	
WebSphere Portal Версии 6	361
Интеграция поисковой системы предприятия с	
кластеризованной системой WebSphere Portal	362
Конфигурирование поисковой системы	
предприятия в кластерной системе WebSphere	
Portal	362
Удаление поисковой системы предприятия из	
кластерной системы WebSphere Portal.	366

Перенастройка из WebSphere Portal в поисковую систему предприятия . 369

Перенос собрания из WebSphere Portal	369
Перенос параметров собраний	371
Файл журнала мастера по перенастройке	372

Команды, коды возврата и ID сеансов поисковой системы предприятия 373

Регистрозависимость в поисковой системе предприятия 403

Документация поисковой системы предприятия 405

Возможности доступности 407

Глоссарий терминов поисковой системы предприятия 409

Замечания и товарные знаки 423

Замечания	423
Товарные знаки.	425

Индекс 427

ibm.com и соответствующие ресурсы

Поддержку продукта и документацию можно получить на веб-сайте [ibm.com](http://www.ibm.com).

Поддержка и содействие

Поддержка данного продукта доступна в Web.

IBM OmniFind Enterprise Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-enterprise/support.html>

IBM OmniFind Discovery Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-discovery/support.html>

IBM OmniFind Yahoo! Edition

<http://www.ibm.com/software/data/enterprise-search/omnifind-yahoo/support.html>

Информационный центр

Документацию по продукту можно просматривать с помощью веб-браузера в информационном центре, работающем на основе технологии Eclipse.

Информационный центр можно найти по адресу <http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/>.

Публикации в формате PDF

PDF-файлы можно просматривать на компьютере, пользуясь версией Adobe Acrobat Reader, соответствующей вашей операционной системе. Если Acrobat Reader у вас не установлен, то его можно скачать с веб-сайта Adobe по адресу <http://www.adobe.com>.

Публикации в формате PDF находятся на следующих веб-сайтах:

Продукт	Адрес веб-сайта
OmniFind Enterprise Edition, версия 8.5	http://www.ibm.com/support/docview.wss?rs=63&uid=swg27010938
OmniFind Discovery Edition, версия 8.4	http://www.ibm.com/support/docview.wss?rs=3035&uid=swg27008552
OmniFind Yahoo! Edition, версия 8.4	http://www.ibm.com/support/docview.wss?rs=3193&uid=swg27008932

Как послать свои комментарии

Ваши отзывы важны, потому что они помогают нам повышать точность и качество информации.

Высылайте свои комментарии с помощью электронного бланка читательских отзывов по адресу https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Как связаться с IBM

Чтобы обратиться в службу поддержки IBM в США или Канаде, позвоните по телефону 1-800-IBM-SERV (1-800-426-7378).

О существующих возможностях обслуживания можно узнать, позвонив по одному из следующих номеров:

- В США: 1-888-426-4343
- В Канаде: 1-800-465-9600

Подробнее о том, как обратиться в IBM, вы узнаете на веб-сайте Contact IBM <http://www.ibm.com/contact/us/>.

Что такое поисковая система предприятия?

Поисковая система предприятия предоставляет возможность с помощью одного запроса выполнять поиск по любому числу структурированных или неструктурированных источников данных. Малое время отклика на запрос и сгруппированный и ранжированный набор результатов, для получения которых применяются мощные средства анализа текста, позволяют не только находить нужные документы, но и извлекать из них нужное содержимое.

Компоненты поисковой системы предприятия, устанавливаемые с IBM OmniFind Enterprise Edition, собирают информацию со всего предприятия. Введя запрос в браузере, можно выполнять поиск одновременно в локальных и удаленных базах данных, вспомогательных системах, системах управления содержимым, файловых системах, внутренних и внешних сайтах.

Поисковая система предприятия спроектирована для полной интеграции с существующими системами; она выполняет всю работу по сбору данных из разнородных источников и построению индексов для быстрого извлечения данных. Используя лингвистический анализ и другие типы анализа, поисковая система предприятия возвращает результаты поиска с высоким уровнем соответствия запросу. Вам не придется изучать разные интерфейсы для поиска по различным типам репозиториям.

Можно добавить поддержку поиска в источниках данных, которые вы не хотите включать в индекс поисковой системы предприятия. При помощи функции объединенного поиска, включенной в поисковую систему предприятия, можно выполнять поиск в этих внешних источниках одновременно с поиском в индексированных источниках данных.

Качество поиска

OmniFind Enterprise Edition поддерживает архитектуру управления неструктурированной информацией IBM (UIMA - Unstructured Information Management Architecture), которая позволяет пользователям находить нужную им информацию. UIMA - это открытая структура, определяющая общий стандартный интерфейс для анализа текста. Всесторонний семантический анализ позволяет поисковой системе предприятия выявлять в неструктурированном тексте понятия, латентные значения, связи, факты и другие релевантные данные, которые часто бывают скрыты. Извлеченную при анализе информацию можно использовать для повышения качества результатов поиска или для повышения качества других прикладных программ, например, бизнес-интеллекта и исследования данных.

Защита

Защита - существенный элемент поисковой системы предприятия. Системой могут управлять только авторизованные для этого пользователи. Используя механизмы защиты IBM WebSphere Application Server, можно настроить роли администраторов и предоставлять разным пользователям доступ к разным функциям управления.

Кроме того, можно задать опции установки маркеров защиты данных при сборе данных. Если ваши программы поиска поддерживают защиту, эти маркеры (хранящиеся в индексе вместе с документами) можно использовать для обеспечения

управления доступом, чтобы только пользователи с соответствующими параметрами регистрации могли запрашивать данные и просматривать результаты поиска.

Для некоторых типов источников данных можно конфигурировать опции проверки учетных данных регистрации пользователя с текущим управлением доступом, выполняемой при обработке запросов. Этот дополнительный уровень защиты обеспечивает проверку привилегий пользователей в реальном времени с собственным источником данных. Эта опция обеспечивает защиту от таких экземпляров, где учетные данные пользователя изменились после индексации документа и его маркеров защиты.

Обучающая программа

На странице <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> доступна электронная обучающая программа. В ней описаны действия по установке и конфигурированию, показано, как выполнять поиск в источниках данных различных типов, и описано использование интерфейсов прикладного программирования продукта для расширения поисковой системы предприятия. Эта обучающая программа написана для более старой версии OmniFind Enterprise Edition, но большинство понятий и процедур все еще применимы.

Понятия, связанные с данным

“Обзор компонентов поисковой системы предприятия” на стр. 3

“Поток данных поисковой системы предприятия” на стр. 12

Защита поисковой системы предприятия



Интеграция пользовательского анализа текста



Основные понятия текстового анализа

Ссылки, связанные с данной

“Типы источников данных, поддерживаемые поисковой системой предприятия”

Типы источников данных, поддерживаемые поисковой системой предприятия

В системе предусмотрена поддержка поиска по различным типам источников данных.

После установки IBM OmniFind Enterprise Edition можно собирать данные из источников данных следующих типов:

- Типы элементов IBM DB2 Content Manager (документы, ресурсы и элементы)
- Базы данных IBM DB2
- Базы данных IBM Domino Document Manager (прежнее название - Domino.Doc)
- Базы данных IBM Lotus Notes
- Базы данных IBM Lotus QuickPlace
- Библиотеки контента IBM Lotus Quickr
- Репозитории IBM WebSphere Information Integrator Content Edition, в том числе Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), Microsoft SharePoint, OpenText Livelink Enterprise Server и WebSphere Portal Document Manager (PDM)
- Таблицы псевдонимов IBM WebSphere Information Integrator для систем баз данных многих типов, в том числе IBM DB2 for z/OS, IBM Informix, Microsoft SQL Server, Oracle и Sybase
- Сайты IBM WebSphere Portal

- Сайты IBM Workplace Web Content Management
- Общедоступные папки Microsoft Exchange Server
- Базы данных Microsoft SQL Server
- Файловые системы Microsoft Windows
- Группы новостей NNTP
- базы данных Oracle
- Файловые системы UNIX
- Сайты Интернета или внутренней сети

Кроме того, можно добавить поддержку поиска во внешних источниках следующих типов без добавления документов из этих источников в индекс поисковой системы предприятия:

- Базы данных, поддерживающие протокол Java database connectivity (JDBC) (только системы баз данных DB2 и Oracle). Для каждой таблицы, в которой вы разрешаете поиск, создается отдельный внешний источник данных.
- Серверы LDAP (Lightweight Directory Access Protocol). Для каждого сервера LDAP создается один внешний источник.

Последнюю информацию о поддерживаемых типах источников данных смотрите на странице требований к системе на сайте OmniFind Enterprise Edition Support.

Понятия, связанные с данным

Поддержка внешних источников

“Поток данных поисковой системы предприятия” на стр. 12

Обзор компонентов поисковой системы предприятия

Компоненты поисковой системы предприятия собирают данные по предприятию, анализируют и категоризируют информацию и создают индекс, который пользователи используют при поиске.

Собрание поисковой системы предприятия - это набор источников, в которых пользователи могут выполнять поиск при помощи одного запроса. При создании собрания вы задаете источники, которые должны в него входить, и конфигурируете опции, определяющие возможности пользователей при поиске индексированных данных.

Можно создать несколько собраний, каждое из которых может содержать данные из разнообразных источников данных. Например, можно создать собрание, включающее документы из баз данных IBM DB2 Universal Database, IBM Lotus Notes и IBM DB2 Content Manager. Когда пользователи выполняют поиск в этом собрании, результаты поиска могут включать документы из каждого из этих источников данных.

Поддержка объединенного поиска позволяет пользователям выполнять поиск в нескольких собраниях при помощи одного запроса. Результаты поиска могут включать документы из всех собраний и внешних источников в вашей поисковой системе предприятия.

Создание собрания и управление им включает в себя следующие действия:

Сбор данных

Компоненты *искателя* собирают документы из источников данных либо

непрерывно, либо по заданному вами расписанию. Частый просмотр информации искателями гарантирует, что у пользователей всегда будет доступ к последним данным.

Анализ данных

Компоненты *анализатора* извлекают текст из документов и выполняют лингвистический и другие типы анализа для каждого документа, обработанного искателем. Подробный анализ содержимого повышает качество результатов поиска.

Индексация данных

Компоненты *индексации* работают на основе регулярных расписаний, добавляя в индекс информацию о новых и измененных документах. Компоненты индексации выполняют также глобальный анализ документов в собрании, улучшая качество результатов поиска.

Поиск данных

Компоненты *поиска* выполняют поиск по индексу и взаимодействуют с программами поиска; они обрабатывают требования поиска и возвращают результаты поиска.

Другие компоненты OmniFind Enterprise Edition позволяют задать предпочтения защиты, отслеживать активность системы и устранять возникающие ошибки. Кроме того, в продукт включен пример рабочей программы поиска, которую можно использовать как образец для создания своих собственных программ поиска.

Понятия, связанные с данным

Управление искателями

Управление анализаторами

Управление индексами

Управление сервером поиска

“Поток данных поисковой системы предприятия” на стр. 12

Искатели поисковой системы предприятия

Искатели поисковой системы предприятия собирают из источников данных документы, чтобы для них можно было выполнить анализ, индексацию и поиск.

Компонент искателя, поставляемый с OmniFind Enterprise Edition, выполняет следующие функции:

- При конфигурировании искателя процессы *обнаружения* находят информацию об источниках, доступных для просмотра искателем, например, имена всех представлений и папок в базе данных Lotus Notes или имена всех файловых систем на сервере UNIX.
- После выбора источников для просмотра искателем и запуска искателя компоненты искателя собирают из этих источников данные, чтобы их можно было проанализировать и проиндексировать.

В одном собрании может быть несколько искателей, при этом каждый искатель используется для сбора данных из источников данных определенного типа. Например, можно создать три искателя, чтобы объединить в одном собрании данные из файловых систем, баз данных Notes и реляционных баз данных. Можно также создать несколько искателей одного типа и сконфигурировать для них отличающиеся расписания просмотра в соответствии с тем, насколько часто изменяются данные, просматриваемые каждым искателем.

После запуска искателя Web он работает в непрерывном режиме. Вы задаете адреса URL для просмотра, и искатель периодически возвращается к ним, проверяя новые и измененные данные. Искатели других типов можно запускать и останавливать вручную или задать для них расписания просмотра. При задании расписания для искателя нужно задать время его начального запуска и частоту, с которой он должен посещать источники данных для просмотра новых и измененных документов.

Свойства искателя - это набор правил, регулирующих поведение отдельного искателя во время просмотра им данных. Например, можно задать правила, управляющие использованием искателем системных ресурсов. Набор источников, отобранных для просмотра, составляет *пространство для искателя*. Свойства искателя можно отредактировать в любое время после его создания, изменив правила сбора данных искателем. Можно также отредактировать пространство для искателя, изменив расписание просмотра, добавив новые источники или удалив источники, поиск в которых больше не требуется.

Понятия, связанные с данным

Управление искателями

Задачи, связанные с данной

Отслеживание работы искателей

Анализаторы поисковой системы предприятия

Синтаксический анализатор поисковой системы предприятия анализирует документы, собранные искателем, и подготавливает их для индексирования.

Компонент синтаксического анализа, поставляемый с OmniFind Enterprise Edition, анализирует содержимое документов и их метаданные. Этот компонент сохраняет результаты анализа на складе данных для компонента индексирования.

Синтаксический анализатор выполняет следующие задачи:

- Извлекает текст из документов любого формата. Например, синтаксический анализатор извлекает текст из тегов в документах XML и HTML. Используя технологию Oracle (ранее Stellent) Outside In Content Access for IBM OmniFind Enterprise Edition, анализатор извлекает также текст из двоичных форматов, таких как Microsoft Word и Adobe Acrobat Portable Document Format (PDF).
- Определяет кодовую страницу документа. Анализатор использует эту информацию, чтобы преобразовать весь текст в Unicode, прежде чем выполнять лингвистический анализ.
- Определяет исходный язык каждого документа.
- Извлекает текст и добавляет маркеры, чтобы облегчить извлечение данных. На этой фазе синтаксический анализатор выполняет следующие задачи:
 - Нормализация символов, например, перевод символов в верхний регистр и удаление диакритических знаков, таких как умляуты в немецком.
 - Анализ структуры абзацев, предложений, слов и пустого пространства. Используя лингвистический анализ, анализатор разбивает составные слова на части и назначает маркеры, позволяющие выполнять просмотр словаря и синонимов.
- Применяет правила синтаксического анализа, заданные для собрания. При конфигурировании синтаксического анализатора можно настроить следующие действия анализа:

Правила отображения полей для документов XML и HTML

Эта возможность позволяет пользователям выполнять поиск по структурированному и неструктурированному содержимому в документах XML и HTML. Если отображать элементы XML или элементы метаданных

HTML в поля поиска в индексе поисковой системы предприятия, пользователи смогут указывать в запросах имена полей и выполнять поиск по определенным частям документов XML и HTML. Запросы с указанием конкретных полей могут давать более точные результаты поиска, чем произвольные запросы текста, выполняющие поиск во всем содержимом документа.

Категории

Эта возможность позволяет пользователям выполнять поиск по определенной категории документов. Кроме того, пользователи могут выбрать в результатах поиска категории и просмотреть документы по категориям, к которым они относятся.

При создании собрания можно выбрать, хотите ли вы использовать категоризацию. Для категорий *на основе правил* документы будут разбиваться по категориям на основе определенных вами правил. Сконфигурировать категории на основе правил можно для создаваемых собраний поисковой системы предприятия и для собраний, переносимых из IBM WebSphere Portal.

Пользовательский анализ текста

Разработчики могут создавать свои программы для сложного лингвистического анализа данных, в которых выполняется поиск. Эти программы можно подключать к поисковой системе предприятия и использовать для аннотирования содержимого ваших собраний. Индексация аннотаций делает возможным семантический поиск в собраниях.

Например, пользователи могут выполнять поиск терминов запроса, встречающихся недалеко друг от друга или встречающихся в одном предложении, или поиск различных взаимосвязей между терминами запроса. Например, вам могут понадобиться документы, в которых упоминается тот Смит из IBM, который занимается продажами, а не тот Смит из IBM, который инженер.

Поддержка сегментации энграмм

Чтобы облегчить извлечение документов на китайском, японском и корейском, можно разрешить такой метод лексического анализа, как сегментация энграмм. В этом способе анализа пробелы не считаются разделителями слов. После создания собрания изменить метод сегментации нельзя.

Поддержка поиска по документам XML с использованием собственного поиска XML

Собственный поиск XML может давать более точные результаты благодаря учету разметки XML. Например, в запросе можно задать, что некоторое слово должно встретиться в определенном элементе XML.

Классы значимости для оценки относительной важности полей

Отображая поля на классы значимости, можно влиять на ранжирование документов в результатах поиска. Например, можно повысить оценку полей названий, чтобы документы, у которых термин запроса входит в название, получали более высокий ранг в результатах поиска.

Понятия, связанные с данным

Работа с категориями

поля поиска XML

поля поиска HTML



Интеграция пользовательского анализа текста

 Включение анализа текста в поисковую систему предприятия

Задачи, связанные с данной

Отслеживание работы анализатора

Индексы поисковой системы предприятия

Компоненты индексации поисковой системы предприятия работают по регулярным расписаниям, добавляя в индекс информацию о новых и измененных документах.

Чтобы у пользователей всегда был доступ к последней информации в источниках, где они выполняют поиск, применяются два варианта построения индекса:

Полное построение индекса

При полном построении индекса перестраивается весь индекс, что гарантирует оптимальность его структуры. Процессы индексирования считывают все данные, собранные искателями и проанализированные анализаторами.

Дельта-построение индекса

При дельта-построении индекса в индекс добавляется информация, собранная искателем с момента последней операции полного построения индекса.

При конфигурировании опций индексации для собрания можно задать расписания для полного построения и дельта-построения индекса. Периодичность, с которой следует выполнять построение индекса, определяется ресурсами системы и содержимым индексируемых источников - статическим или динамическим.

Чтобы гарантировать доступность новой информации, нужно запланировать достаточно частое дельта-построение индекса. Следует также запланировать периодическое полное построение индекса, чтобы объединять новую информацию, анализировать новое содержание и оптимизировать производительность индекса.

Процессы индексации можно запускать также вне расписаний. Например, если вы изменили некоторые правила анализа и хотите, чтобы эти изменения стали доступны вашим программам поиска, можно запустить полное построение индекса после повторного просмотра и анализа данных, не дожидаясь запуска построения индекса по расписанию.

Чтобы управлять использованием ресурсов, можно задать число собраний, которые могут одновременно использовать процессы индексации и отправлять требования построения индексов. Одновременное построение индексов позволяет добиться того, что операция полного построения очень большого индекса не будет блокировать операции дельта-построения индексов для других собраний. Построение индекса может быть ресурсоемким процессом, поэтому чтобы настроить расписания операций полного построения и дельта-построения индекса для большой системы, нужно выполнить мониторинг системной нагрузки.

При построении индекса процессы индексации выполняют глобальный анализ документов. Во время этой фазы применяются алгоритмы для идентификации повторяющихся документов, анализа структуры ссылок документов и выполнения специальной обработки текста привязок в документах Web (текста, описывающего страницу назначения в гиперссылке).

Можно задать опции для следующих операций индексации:

- Чтобы разрешить пользователям задавать символы подстановки, можно или встроить в индекс поддержку для расширения терминов запроса, или указать, что расширение терминов запроса надо выполнять при обработке запроса. Выбор опции зависит от того, что именно вы хотите минимизировать - использование ресурсов или время ответа на запрос.
- Можно сконфигурировать области действия. *Область действия* позволяет ограничить данные, которые пользователи могут просмотреть в собрании. Например, можно создать одну область действия, включив в нее все идентификаторы URI для документов из вашего технического отдела, а в другую область действия - URI документов из отдела кадров. Если программа поиска поддерживает области действия, пользователи смогут выполнять поиск документов и получать их только из этих поднаборов документов в собрании.
- Можно задать опции для объединения в результатах поиска документов с одинаковым префиксом URI. Можно также задать имя группы, чтобы в результатах поиска можно было объединять документы с разными префиксами URI.
- После построения индекса можно удалить URI, для которых пользователи не должны выполнять поиск.

Понятия, связанные с данным

Управление индексами

Символы подстановки в запросах

Области действия

Сворачивание URI

Ранжирование документов на основе шаблонов URI

Задачи, связанные с данной

Планирование операций построения индекса

Конфигурирование одновременных операций построения индекса

Удаление URI из индекса

Отслеживание активности индекса для собрания

Слежение за очередью индексации поисковой системы предприятия

Серверы поиска для поисковой системы предприятия

Серверы поиска для поисковой системы предприятия работают с прикладными программами поиска, они обрабатывают запросы, выполняют поиск по индексу и возвращают результаты поиска.

Серверы поиска для поисковой системы предприятия устанавливаются при установке OmniFind Enterprise Edition. При конфигурировании серверов поиска для собрания можно задать опции поиска по собранию:

- Можно сконфигурировать кэш поиска для хранения часто запрашиваемых результатов поиска. Кэширование поиска может повысить производительность поиска и получения документов.
- Можно задать язык по умолчанию для поиска документов в этом собрании.
- Если ваши разработчики программы создадут пользовательские словари, можно связать эти словари с собраниями:
 - При запросах к собранию, использующему *словарь синонимов*, документы, содержащие синонимы терминов запроса, включаются в результаты поиска.
 - При запросах к собранию, использующему *словарь стоп-слов*, стоп-слова удаляются из запроса перед его обработкой.

- При запросах к собранию, использующему *словарь значимых слов*, важность документов, содержащих эти слова, увеличивается или уменьшается (в зависимости от показателя значимости, заданного для этих слов в словаре).
- Если вы заранее определили, что определенным запросам соответствуют определенные документы, можно настроить прямые ссылки. *Прямая ссылка* связывает определенный URI с определенными ключевыми словами и словосочетаниями. Если в запросе содержится какое-либо из ключевых слов или словосочетаний, заданных в определении прямой ссылки, в результатах поиска автоматически возвращается соответствующий URI.

Для конфигурации с несколькими серверами защита от сбоя доступна также на уровне собрания, а не только на уровне сервера. Если по какой-либо причине собрание на одном из серверов поиска станет недоступно, запросы для этого собрания будут автоматически направлены на другой сервер поиска.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Кэши поиска

Пользовательские словари синонимов

Пользовательские словари стоп-слов

Пользовательские словари значимых слов

Прямые ссылки

Задачи, связанные с данной

Отслеживание работы серверов поиска

Консоль администратора поисковой системы предприятия

Консоль администратора поисковой системы предприятия запускается в браузере, что позволяет пользователем-администраторам обращаться к ней из любого места в любое время. Механизмы защиты гарантируют, что сделать это смогут только пользователи с правами доступа к функциям управления.

Консоль администратора поисковой системы предприятия устанавливается на серверах поиска при установке OmniFind Enterprise Edition.

Консоль администратора содержит мастера, помогающие выполнять некоторые основные задачи управления. Например, мастер по собраниям помогает создать собрание и позволяет сохранять работу в черновом варианте. Мастера искателей (отдельные для каждого из типов источников данных) помогают выбрать источники, которые вы хотите сделать доступными для поиска.

В других задачах управления можно выбрать отдельные элементы для управления. Например, при редактировании собрания можно выбрать страницу Индекс, чтобы изменить расписание индексации, или страницу Анализ, чтобы изменить правило синтаксического анализа документов XML.

Понятия, связанные с данным

Управление поисковой системой предприятия

Административные роли

Задачи, связанные с данной

Регистрация на консоли администратора

Слежение за поисковой системой предприятия

С помощью консоли администратора поисковой системы предприятия можно выполнять слежение за системой и при необходимости настраивать операции.

После установки OmniFind Enterprise Edition и создания как минимум одного собрания можно просмотреть подробную статистику по каждому из основных процессов (таких как просмотр искателем, анализ, индексация и поиск). В эту информацию входит среднее время отклика и информация о ходе выполнения, например, число документов, просмотренных или проиндексированных в ходе отдельного сеанса просмотра искателем или построения индекса.

Большинство процессов вы можете запускать и останавливать. Например, можно приостановить процесс, изменить его конфигурацию или исправить ошибку и перезапустить обработку, когда вы будете готовы ее продолжить.

Можно также сконфигурировать оповещения, позволяющие получать по электронной почте сообщения при возникновении определенных событий для отслеживаемых процессов. Например, можно получать оповещение, когда время ответа поиска превысило заданный порог.

Если документ был отброшен из поисковой системы предприятия, можно отследить этот документ и выяснить, когда, где и почему он был отброшен. Возможно, этот документ не смог проанализировать синтаксический анализатор или же его удалил из индекса администратор.

Понятия, связанные с данным

Отслеживание активности поисковой системы предприятия

Запуск и остановка поисковой системы предприятия

Файлы журналов поисковой системы предприятия

Файлы журналов поисковой системы предприятия создаются для отдельных собраний и для сеансов системного уровня.

При конфигурировании опций записи в журналы для собрания поисковой системы предприятия или для системы нужно задать типы сообщений, записываемых в журнал, например, сообщения об ошибках и предупреждения. Нужно также задать, как часто система должна переходить от более старых журналов к новым, освобождая место для свежих сообщений. Можно выбрать опции для получения сообщения электронной почты о конкретных сообщениях (в том числе оповещений) или всех сообщениях об ошибках, когда они происходят.

При просмотре файлов журналов можно выбрать файл журнала, который нужно просмотреть. Имя файла содержит информацию о времени создания файла и о компоненте, сгенерировавшем сообщения. Можно также задать фильтры просмотра. Например, можно просмотреть только сообщения об ошибках или только сообщения от конкретного сеанса поисковой системы предприятия.

Понятия, связанные с данным

Файлы журналов и оповещения

Оповещения



Сообщения для поисковой системы предприятия

Задачи, связанные с данной

Конфигурирование файлов журнала

Конфигурирование информации о SMTP-сервере

Получение сообщений журнала по электронной почте
Просмотр файлов журнала

Настройка поисковой системы предприятия

Интерфейсы прикладного программирования для поисковой системы предприятия позволяют создавать пользовательские программы поиска, пользовательские программы для обновления содержимого собраний, пользовательские программы для анализа текста и пользовательские словари синонимов, стоп-слов и значимых слов.

После установки OmniFind Enterprise Edition можно использовать следующие семейства API, позволяющие расширить собрания поисковой системы предприятия:

API поиска и индексации (SI-API)

Этот API используется для построения пользовательских программ поиска и настроенного интерфейса администратора.

Подключаемые модули искателей

API подключаемых модулей используются для добавления метаданных в документы при их просмотре искателем или для привязки маркеров защиты, обеспечивающих выполнение бизнес-правил и правил защиты, принятых в вашей организации.

Можно повысить возможность получения информации, интегрировав пользовательские программы лингвистического анализа в собрания поисковой системы предприятия. Добавив в систему пользовательские механизмы анализа текста, можно связать эти механизмы с собраниями. Пользователи, запрашивающие собрание, могут использовать преимущества информации о взаимосвязях между словами, которую пользовательские программы встраивают в индекс. Например, пользователи могут выполнять поиск понятий и взаимосвязей между терминами, а не просто самих терминов.

Можно также повысить возможность получения информации, интегрировав пользовательские словари, отражающие, например, акронимы, сокращения и специфические словарные термины для вашей области деятельности. Добавив в систему словари, можно связать эти словари с собраниями. Пользователи, запрашивающие собрание, получают следующие преимущества:

- Если запрос содержит слова, определенные как синонимы, документы, содержащие синонимы терминов запроса, включаются в результаты поиска.
- Если запрос содержит стоп-слова, они удаляются из запроса, чтобы исключить из результатов поиска нерелевантные документы.
- Если запрос содержит значимые слова, документы, содержащие эти слова, получают более высокий или более низкий ранг в результатах поиска в зависимости от показателей значимости, заданных в словаре для этих слов.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Пользовательские словари синонимов

Пользовательские словари стоп-слов

Пользовательские словари значимых слов



Обзор API поиска и индексирования



Подключаемые модули искателей

Пример программы поиска для поисковой системы предприятия

Можно использовать этот пример программы поиска как образец для создания пользовательских программ поиска.

Пример программы поиска устанавливается при установке OmniFind Enterprise Edition. Этот пример программы поиска демонстрирует большинство функций поиска и получения, доступных для поисковой системы предприятия. Эта программа - рабочий образец, позволяющий выполнять поиск во всех активных собраниях и внешних источниках в поисковой системе предприятия. Пример программы поиска можно использовать для тестирования новых собраний и внешних источников перед тем, как делать их доступными для пользователей.

Пример программы поиска демонстрирует поддержку объединенного поиска, позволяя одновременно выполнять поиск в нескольких собраниях и внешних источниках.

Для некоторых типов искателей можно использовать компонент управления идентификацией для поисковой системы предприятия, чтобы проверять текущие параметры регистрации при обращении пользователя к программе поиска. Если домен для поиска защищен с помощью защиты с единой регистрацией (SSO), для проверки пользователя в сеансе поиска можно использовать механизмы SSO. Если единая регистрация не используется, можно с помощью компонента управления идентификацией зашифровать параметры регистрации пользователя и сохранить их в профиле, а затем использовать эти параметры регистрации для исключения из результатов поиска документов, доступ к которым запрещен.

Для настройки примера программы поиска можно использовать Настройщик программ поиска, графический пользовательский интерфейс которого позволяет сразу видеть влияние вносимых изменений. Для настройки программы поиска можно также отредактировать файл конфигурации для этой программы.

Для создания собственной программы поиска можно использовать API поиска и индексации поисковой системы предприятия.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Функции примера программы поиска

 Обзор API поиска и индексирования

Задачи, связанные с данной

Обращение к программе поиска

Редактирование свойств примера программы поиска

Настройка программы поиска

Поток данных поисковой системы предприятия

Компоненты поисковой системы предприятия, которые устанавливаются вместе с OmniFind Enterprise Edition, тесно взаимодействуют между собой, чтобы обеспечить поток данных через систему.

Искатели собирают документы из источников данных по всему предприятию. Синтаксический анализатор извлекает из собранных искателем документов полезную информацию и генерирует маркеры, которые позволяют, например, распределять

документы по категориям и помогают определить степень соответствия документов терминам в требовании поиска. В индексе хранятся данные для эффективного получения документов.

Используя браузер и прикладные программы поиска, пользователи выполняют поиск по проиндексированным собраниям и внешним источникам. Программа поиска может вывести список результатов, по которым пользователь может щелкнуть в браузере; прикладная программа может также быть более сложной и возвращать динамически генерируемое содержание на основе информации из разных источников.

Например, программа поиска каталога может настраивать способ вывода продуктов, удовлетворяющих требованию поиска. С помощью одного запроса можно выполнять поиск по документам в разных источниках данных, например, по документам в репозиториях IBM DB2 Content Manager и Lotus Notes.

Администраторы задают, какие данные надо собирать и как их надо просматривать анализатором, анализировать, индексировать и искать. Наблюдая за работой системы, администраторы выполняют также настройку для оптимизации потоков данных.

На следующей диаграмме показан поток информации в поисковой системе предприятия.

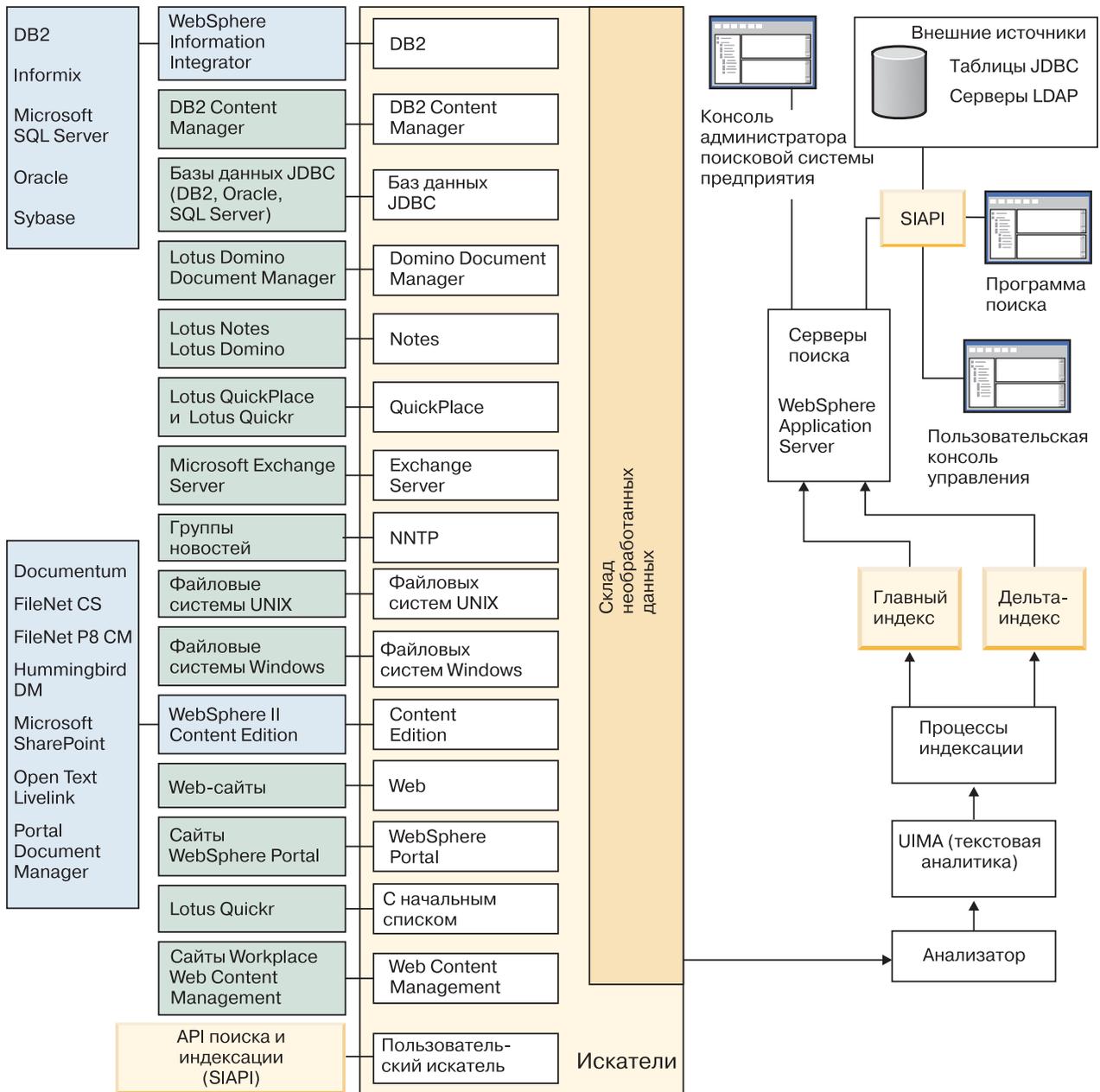


Рисунок 1. Как перемещаются данные по поисковой системе предприятия

Понятия, связанные с данным

“Что такое поисковая система предприятия?” на стр. 1

“Обзор компонентов поисковой системы предприятия” на стр. 3

Ссылки, связанные с данной

“Типы источников данных, поддерживаемые поисковой системой предприятия” на стр. 2

Управление поисковой системой предприятия

При помощи консоли управления поисковой системой предприятия можно создавать собрания и внешние источники и управлять ими, запускать и останавливать компоненты, следить за работой системы и файлами журналов, конфигурировать пользователей - администраторов и связывать программы поиска с собраниями.

Совет: На странице <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> доступна электронная обучающая программа. В ней описаны действия по установке и конфигурированию, показано, как выполнять поиск в источниках данных различных типов, и описано использование интерфейсов прикладного программирования продукта для расширения поисковой системы предприятия. Эта обучающая программа написана для более старой версии OmniFind Enterprise Edition, но большинство понятий и процедур все еще применимы.

Подробные примеры способов конфигурирования искателей и включения защиты в небольших, средних и крупномасштабных организациях смотрите в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Окно Собрания

При создании первого собрания и управлении системой следуйте приведенным ниже инструкциям.

Регистрация

Зарегистрируйтесь на консоли управления поисковой системой предприятия. Окно Собрания - это точка входа для создания собраний и управления ими.

Совет: Чтобы получить справку по консоли администратора, нажмите кнопку **Справка** на панели инструментов или **Справка для этой страницы**. При наличии подробных пояснений к сообщениям и описаний действий по исправлению щелкните по значку  **Дополнительная информация** в конце сообщения, чтобы увидеть эти подробности.

Создание собрания

Новое, пустое собрание можно создать одним из следующих способов:

- Чтобы создать собрание при помощи мастера по собраниям, выберите **Мастер по собраниям** и следуйте подсказкам мастера.
- Чтобы создать собрание при помощи окна Собрания, выберите **Создать собрание**, заполните поля на странице Создать собрание и нажмите кнопку **ОК**.

Конфигурирование собрания

Нужно отредактировать новое, пустое собрание, чтобы добавить в него содержимое и задать опции для просмотра данных искателем и доступа к этим данным.

Нажмите для вашего нового собрания кнопку  **Изменить**, затем выберите страницу, чтобы задать опции для этого собрания.

Внимание: Чтобы перейти на предыдущую страницу или обновить информацию на консоли администратора, используйте кнопки **Предыдущая** и **Обновить** на консоли администратора поисковой системы предприятия. Нажатие кнопок **Назад** или **Обновить** в браузере может привести к несогласованным результатам и потере данных.

- На странице Общие можно задать опции, применяемые ко всему собранию:
 - Редактируя общие опции, можно изменить имя или описание собрания, а также изменить ожидаемый размер собрания.
 - Можно просмотреть информацию о собрании, которую нельзя изменить, такую как ID собрания или метод статистического ранжирования для ранжирования документов в результатах поиска.
 - Если защита для собрания была включена во время его создания, можно включать и выключать управление защитой на уровне документов.
- На странице Искатель сконфигурируйте как минимум один искатель. Одно собрание может содержать данные из разнообразных источников данных. Для каждого используемого типа источников данных необходимо сконфигурировать как минимум один искатель. При создании искателя сконфигурировать его помогает мастер, зависящий от типа данных этого искателя.
- На странице Анализ можно сконфигурировать опции синтаксического анализа данных, просматриваемых искателем, для более эффективного поиска данных:
 - Можно указать, будет ли выполняться анализ документов в формате XML для поиска собственными средствами поиска XML.
 - Можно связывать документы с категориями, что дает пользователям возможности поиска в поднаборе собрания и просмотра документов в результатах поиска по категориям, к которым принадлежат эти документы.
 - Можно отображать элементы XML и элементы метаданных HTML на поля поиска в индексе, что дает пользователям возможность указывать имена этих полей в запросах и находить необходимые части документов.
 - Если вы добавили в поисковую систему предприятия пользовательские механизмы анализа, можно выбрать один из них для использования с собранием, а затем задать опции обработки текста для улучшения возможностей получения информации и поддержки семантического поиска.
 - Поля можно связать с классами значимости, чтобы изменить ранжирование полей, удовлетворяющих условиям запроса, в результатах поиска.
- На странице Индекс сконфигурируйте расписания для построения индексов. Запланируйте достаточно частое построение индексов, чтобы у пользователей всегда был доступ к последней информации. Можно также внести следующие необязательные изменения:
 - Позволить пользователям использовать символы подстановки в условиях запросов.
 - Сконфигурировать области действия, чтобы пользователи выполняли поиск в ограниченной части собрания, а не во всех документах, содержащихся в индексе.

- Объединять результаты поиска, чтобы документы из одного и того же источника были объединены в результатах поиска.
- Присваивать коэффициенты значимости, чтобы изменить порядок ранжирования в результатах поиска документов, соответствующих шаблону URI.
- Удалять URI из индекса. Например, можно запретить пользователям просматривать определенные документы после создания собрания.
- На странице Поиск можно задать опции для поиска документов в собрании:
 - Можно отвести пространство под кэш для результатов поиска и изменять язык собрания по умолчанию.
 - Если вы добавили в поисковую систему предприятия пользовательские словари для синонимов, стоп-слова и значимые слова, можно выбрать словари для пользователей, выполняющих поиск в собрании.
 - Можно указать длину вывода на экран для сводок документов в результатах поиска.
 - Если нужно, чтобы отдельные идентификаторы URI выводились автоматически в результатах поиска, когда запрос содержит конкретные ключевые слова или фразы, можно сконфигурировать прямые ссылки.
- На странице Журнал можно:
 - Задать типы записываемых в журнал сообщений и частоту перезапуска файлов журналов.
 - Задать опции для получения оповещений о действиях собрания. Например, можно настроить оповещение, информирующее вас о том, что среднее время отклика поиска превысило определенный предел.
 - Задать опции для получения сообщений по электронной почте, когда в журнал записываются конкретные сообщения или сообщения конкретных типов.
 - Задать опции для информации регистрации, позволяющие определять время, место и причину отбрасывания документа из поисковой системы предприятия.

Запуск компонентов

После задания источников данных, просматриваемых искателем, и опций для сбора и поиска данных можно запустить процессы для построения собрания. Порядок запуска компонентов критичен. Чтобы данные можно было проанализировать и проиндексировать, сначала их должны просмотреть искатели; чтобы серверы поиска смогли обработать требования поиска, должно быть выполнено полное построение индекса.

Окно Внешние источники

Если нужно выполнить поиск в источниках данных без их просмотра искателем или индексирования, нажмите кнопку **Внешние источники** на панели инструментов, чтобы задать опции, делающие источники данных доступными для поиска. Необходимо задать информацию, разрешающую поисковой системе предприятия доступ к базам данных JDBC (Java Database Connectivity) и серверам LDAP (Lightweight Directory Access Protocol). Связав внешние источники с прикладными программами, поиска, вы позволите пользователям выполнять поиск в этих источниках одновременно с поиском в собраниях с обработанным искателем, проанализированными и проиндексированными данными.

Окно Система

Если вы - администратор поисковой системы предприятия, на панели инструментов можно выбрать **Система**, чтобы выполнить следующие действия. Администраторы, операторы и пользователи мониторов собраний могут обращаться к этому окну, только если им разрешил администратор поисковой системы предприятия.

- Добавьте в систему пользовательские механизмы анализа текста.
- Добавьте в систему пользовательские словари для синонимов, стоп-слова и значимые слова.
- Задайте число собраний для одновременного построения индексов и укажите, возможно ли для одного собрания одновременное полное построение индекса и дельта-построения индексов.
- Сконфигурируйте оповещения для событий на уровне системы.
- Задайте опции для записи сообщений, генерируемых сеансами на уровне системы.
- Укажите сведения о вашем почтовом сервере, чтобы получать по электронной почте сообщения о работе поисковой системы предприятия.

Окно Защита

Если вы - администратор поисковой системы предприятия, можно выбрать **Защита**, чтобы задать опции защиты. У администраторов, операторов и наблюдателей собраний нет доступа к этому окну.

Если включена защита на сервере WebSphere Application Server, в окне Защита можно сконфигурировать административные роли. Конфигурирование административных ролей позволяет разрешить управление системой большему числу пользователей, но с ограничением доступа каждого пользователя к конкретным функциям и собраниям.

Кроме того, в окне Защита можно сконфигурировать опции управления идентификацией. Например, можно задать опции для хранения параметров аутентификации пользователей в профилях, используемых для проверки правильности текущих параметров аутентификации пользователя при обработке запроса. Если для защиты источника, в котором выполняется поиск, используется защита с единой регистрацией (SSO), можно также задать опции использования методов аутентификации SSO для проверки правильности текущих параметров аутентификации пользователя при обработке запроса.

Пока вы не создали ваши собственные программы поиска, можно использовать пример программы поиска, выполняющей поиск во всех собраниях и внешних источниках. Создав пользовательскую программу поиска, при помощи окна Защита свяжите созданную программу с собраниями и внешними источниками, в которых она может выполнять поиск.

Окно Слежение

В любое время можно щелкнуть по значку  **Слежение**, чтобы начать слежение за компонентами системы или собраниями. Если позволяет ваша административная роль, при мониторинге можно также запускать и останавливать процессы компонентов.

Понятия, связанные с данным

“Управление искателями” на стр. 41

“Слежение за активностью поисковой системы предприятия” на стр. 305

Задачи, связанные с данной

“Запуск поисковой системы предприятия” на стр. 297

“Управление серверами поиска в автономном режиме” на стр. 302

“Остановка поисковой системы предприятия” на стр. 299

“Создание собрания при помощи мастера по собраниям” на стр. 33

“Создание собрания при помощи окна Собрания” на стр. 35

Регистрация на консоли администратора

Для управления поисковой системой предприятия надо ввести URL в браузере и зарегистрироваться на консоли администратора.

Прежде чем приступать к работе...

Регистрироваться нужно под ID пользователя, авторизованным для доступа к консоли администратора поисковой системы предприятия:

- Если в WebSphere Application Server не включена глобальная защита, только администратор поисковой системы предприятия, заданный при установке OmniFind Enterprise Edition, может использовать консоль администратора.
- Если вы включили глобальную защиту в WebSphere Application Server, можно использовать консоль администратора поисковой системы предприятия для конфигурирования административных ролей. Конфигурируемые ID пользователей должны существовать в реестре пользователей WebSphere Application Server. При конфигурировании ролей администраторов вы позволяете регистрироваться на консоли администратора большему числу пользователей, но зато можете управлять тем, к каким функциям и собраниям есть доступ у каждого пользователя-администратора.

Процедура

Чтобы зарегистрироваться на консоли администратора поисковой системы предприятия:

1. В браузере введите URL консоли администратора. Например:

`http://SearchServer.com/ESAdmin/`

SearchServer.com - имя хоста поискового сервера для поисковой системы предприятия.

В зависимости от конфигурации вашего Web-сервера может потребоваться задание номера порта. Например:

`http://SearchServer.com:9080/ESAdmin/`

2. На странице приветствия введите свои ID пользователя и пароль, после чего нажмите кнопку **Регистрация**.

Откроется окно Собрания - точка входа для управления системой и собраниями. Если используются роли администраторов, возможные варианты действий и видимые собрания зависят от вашей роли администратора.

Если сеанс не активен в течение некоторого времени, система автоматически отменяет регистрацию. Чтобы продолжить управлять системой, зарегистрируйтесь заново.

Закончив операции управления собраниями, можете выбрать **Выход**, чтобы выйти с консоли. Затем вы можете снова зарегистрироваться с другим ID и паролем или закрыть браузер, чтобы выйти с консоли администратора.

Понятия, связанные с данным

“Административные роли” на стр. 262

Изменение пароля администратора поисковой системы предприятия для конфигурации с одним сервером

Пароль администратора поисковой системы предприятия хранится в зашифрованном виде. Чтобы изменить этот пароль, используйте сценарий **eschangepw**.

Ограничения

Пароли могут включать следующие специальные символы:

! @ # \$ % ^ & * () - _ = + , . / < > ?

В системах AIX, Linux и Solaris при задании пароля со специальными символами надо весь пароль заключить в одинарные кавычки.

Например: 'mypwd@\$%'

В Windows при задании пароля со специальными символами надо весь пароль заключить в двойные кавычки.

Например: "my?+!pwd"

Об этой задаче

Пароль для исходного ID администратора поисковой системы предприятия задается при установке OmniFind Enterprise Edition.

Чтобы изменить пароль, надо запустить сценарий **eschangepw**, который распространит изменение по всей поисковой системе предприятия. Программа установки создает две переменные среды, которые можно использовать со сценарием **eschangepw**:

КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES

Каталог установки поисковой системы предприятия.

КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES

Каталог данных поисковой системы предприятия. Пароль для ID администратора поисковой системы предприятия хранится в файле `es.cfg` в этом каталоге.

Процедура

Для изменения пароля администратора поисковой системы предприятия в конфигурации с одним сервером:

1. Зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему:
`esadmin system stopall`

Важное замечание: После остановки системы пользователи не смогут отправлять требования поиска.

2. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы (в AIX, Linux и Solaris) или утилитой изменения пароля (в Windows).
3. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 2:

AIX, Linux или Solaris

`eschangepw.sh` *новое_значение*

Windows

`eschangepw` *новое_значение*

4. Перезапустите поисковую систему предприятия:

`esadmin system startall`

Ссылки, связанные с данной

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373

Изменение пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами

Пароль администратора поисковой системы предприятия хранится в зашифрованном виде. Чтобы изменить пароль, запустите сценарий **eschangepw**, который изменит его на всех компьютерах в поисковой системе предприятия.

Ограничения

Пароли могут включать следующие специальные символы:

! @ # \$ % ^ & * () - _ = + , . / < > ?

В системах AIX, Linux и Solaris при задании пароля со специальными символами надо весь пароль заключить в одинарные кавычки.

Например: `'mypwd@$%'`

В Windows при задании пароля со специальными символами надо весь пароль заключить в двойные кавычки.

Например: `"my?+!pwd"`

Об этой задаче

Исходный пароль для ID администратора поисковой системы предприятия, заданный при установке OmniFind Enterprise Edition, должен быть одинаковым на всех серверах поисковой системы предприятия.

Чтобы изменить пароль и распространить это изменение по всей поисковой системе предприятия, надо запустить сценарий **eschangepw** на каждом компьютере поисковой системы предприятия. Ниже приведена процедура, описывающая порядок смены паролей на всех серверах. Придерживаться этого порядка не обязательно, необходимо, однако, выполнить все обязательные шаги для каждого типа сервера.

Программа установки создает две переменные среды, которые можно использовать со сценарием **eschangepw**:

КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES

Каталог установки поисковой системы предприятия.

КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES

Каталог данных поисковой системы предприятия. Пароль для ID администратора поисковой системы предприятия хранится в файле `es.cfg` в этом каталоге.

Процедура

Чтобы изменить пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами:

1. На сервере индексов поисковой системы предприятия зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему.
`esadmin system stopall`

Важное замечание: После остановки системы пользователи не смогут отправлять требования поиска.

- a. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы (в AIX, Linux и Solaris) или утилитой изменения пароля (в Windows).
- b. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 1a:

AIX, Linux или Solaris

```
eschangerw.sh новое_значение
```

Windows

```
eschangerw новое_значение
```

2. Выполните следующие шаги на других компьютерах в поисковой системе предприятия:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия.
 - b. Остановите слой общей связи (common communication layer, CCL) для поисковой системы предприятия:

AIX, Linux или Solaris

```
stopccl.sh
```

Окно командной строки Windows

```
stopccl
```

Администрирование - Службы Windows

- 1) Запустите службы Windows: **Пуск → Программы → Инструменты управления → Службы.**
 - 2) Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить.**
- c. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы (в AIX, Linux и Solaris) или утилитой изменения пароля (в Windows). Этот пароль должен соответствовать паролю, который вы задали на шаге 1a.
 - d. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 1a:

AIX, Linux или Solaris

```
eschangerw.sh новое_значение
```

Windows

```
eschangerw новое_значение
```

- e. Перезапустите CCL:

AIX, Linux или Solaris

```
startccl.sh -bg
```

Окно командной строки Windows

```
startccl
```

Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- 1) Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
 - 2) Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Свойства**.
 - 3) Щелкните по вкладке **Вход в систему**.
 - 4) Измените пароль, задав новое значение пароля, и затем нажмите кнопку **ОК**.
 - 5) Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
3. На сервере индексов поисковой системы предприятия зарегистрируйтесь как администратор поисковой системы предприятия и перезапустите поисковую систему.

```
esadmin system startall
```

Ссылки, связанные с данной



Задание одинакового зашифрованного пароля администратора на всех серверах

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373

Номера портов TCP для поисковой системы предприятия

Посмотрите, какие порты используются по умолчанию в поисковой системе предприятия, чтобы избежать конфликтов портов при конфигурировании ресурсов или назначении номеров портов для других прикладных программ.

Если вы конфигурируете брандмауэр, надо явным образом разрешить доступ к портам с конкретными номерами. Необходимо также убедиться, что все серверы поиска предприятия находятся по одну сторону брандмауэра.

Таблица 1. Конфигурации номеров портов для поисковой системы предприятия

Имя порта	Номера портов	Где задаются
Common communication layer (CCL)	6002	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/nodeinfo/es.cfg и КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/nodes.ini на всех серверах поиска предприятия
HTTP на серверах поиска	80	HTTP_SERVER_ROOT/conf/http.conf на серверах поиска
Консоль администратора WebSphere Application Server Версии 6	9060	На серверах поиска
Консоль администратора WebSphere Application Server Версии 5.1	9090	На серверах поиска
Искатель DB2	6000, 6001, 6002, 60003, 50000	На сервере искателя
Информационный центр	8888	На серверах поиска

Таблица 1. Конфигурации номеров портов для поисковой системы предприятия (продолжение)

Имя порта	Номера портов	Где задаются
Анонимные или динамические порты для CCL, передача файлов (ESFTP) и копирование индекса	от 49152 до 65535	На всех серверах поисковой системы предприятия
Apache Derby Network Server	1527	На сервере искателя
Пользовательская связь	8890	На сервере искателя
Соединения удаленного клиента с сервером DB2 (используется только с выпусками до OmniFind Enterprise Edition Версии 8.4)	50000	На сервере искателя
База данных WebSphere Information Integrator Content Edition FastObjects	6001 (как альтернатива - 6002)	На консоли администратора WebSphere Information Integrator Content Edition
Прокси-коннектор удаленного вызова метода (RMI) WebSphere Information Integrator Content Edition	1251 (порт RMI)	На консоли администратора WebSphere Information Integrator Content Edition

Изменение номера порта для поисковой системы предприятия

Если номер порта, используемый поисковой системой предприятия для связи, конфликтует с номером порта, используемым другим продуктом, нужно изменить номер порта поисковой системы предприятия.

Об этой задаче

А Номер порта для поисковой системы предприятия задается при установке OmniFind Enterprise Edition. В многосерверной конфигурации один и тот же номер порта задается на всех серверах.

Если данный номер порта нельзя использовать, например, если он назначен для другого продукта на том же сервере, в результате конфликта в файл CCLServer_дата.log, здесь дата - дата создания файла журнала) будет записано следующее сообщение об ошибке:

```
FFQ00273W Внутреннее предупреждение - Сообщение об исключительной ситуации: {0}
at java.net.PlainSocketImpl.socketBind(Native Method)
at java.net.PlainSocketImpl.bind(PlainSocketImpl.java:357)
at java.net.ServerSocket.bind(ServerSocket.java:341)
at java.net.ServerSocket.<init>(ServerSocket.java:208)
at java.net.ServerSocket.<init>(ServerSocket.java:120)
```

Процедура

Чтобы изменить номер порта, используемого поисковой системой предприятия:

1. Перейдите к компьютеру, на котором нужно изменить номер порта, зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему предприятия:

```
esadmin system stopall
```

Важное замечание: После остановки системы пользователи не смогут отправлять требования поиска.

2. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/nodeinfo/es.cfg`, найдите следующее свойство, задайте новое значение для номера порта, а затем сохраните и закройте файл:

`CCLPort=новый_номер_порта`

3. Перезапустите слой общей связи (common communication layer, CCL) для поисковой системы предприятия:

AIX, Linux или Solaris

`startccl.sh`

Окно командной строки Windows

`startccl`

Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
 - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
4. Перейдите на сервер индексов, зарегистрируйтесь как администратор поисковой системы предприятия и остановите CCL:

AIX, Linux или Solaris

`stopccl.sh`

Окно командной строки Windows

`stopccl`

Администрирование - Службы Windows

- a. Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
 - b. Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить**.
5. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/nodeinfo/es.cfg` :

- a. Найдите следующее свойство (где *имя_компьютера* - имя компьютера, где был изменен номер порта на шаге 2. *N* в свойстве *nodeN* - номер, идентифицирующий сервер поисковой системы предприятия:

`nodeN.destination=имя_компьютера`

- b. Найдите следующее подсвойство, введите номер порта, заданный вами для сервера на шаге 2, а затем сохраните и закройте файл:

`nodeN.port=новый_номер_порта`

6. Перезапустите поисковую систему предприятия:

`esadmin system startall`

Ссылки, связанные с данной

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373

“Номера портов TCP для поисковой системы предприятия” на стр. 23

Изменение имен хостов или IP-адресов поисковой системы предприятия

Вы можете изменить имена хостов или IP-адреса, указанные в конфигурации серверов поисковой системы предприятия.

Например, попробуйте изменить IP-адрес, если при использовании на каждом сервере поиска поисковой системы предприятия нескольких плат сетевого интерфейса будет обнаружено, что сервер индексов сконфигурирован для использования медленной сети. Для возможности использования сервером индексов более быстрой сети можно изменить значения параметров в файлах конфигурации.

Совет: Если вы не хотите редактировать файлы конфигурации, можно переустановить OmniFind Enterprise Edition и задать новые имена хостов или IP-адреса при запуске программы установки.

Порядок действий:

Чтобы изменить имена хостов или IP-адреса:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. При конфигурации с несколькими серверами зарегистрируйтесь на любом поисковом сервере предприятия.
2. Остановите сеансы поисковой системы предприятия:
`esadmin system stopall`
3. Остановите CCL (common communication layer, общий уровень связи). При конфигурации с несколькими серверами остановите CCL на каждом сервере поисковой системы предприятия одним из следующих способов:

AIX, Linux или Solaris

```
stopccl.sh
```

Окно командной строки Windows

```
stopccl
```

Администрирование - Службы Windows

- a. Запустите службы Windows: **Пуск** → **Программы** → **Администрирование** → **Службы**.
 - b. Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить**.
4. Выполните на каждом сервере поисковой системы предприятия следующие действия:
 - a. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/configurations/ccl.properties` и задайте имя хоста или IP-адрес, которые вы хотите использовать для этого сервера, в параметре `es_server_hostName`.
 - b. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/nodeinfo/es.cfg` и задайте имя хоста или IP-адрес, которые вы хотите использовать для этого сервера, в параметре `LocalHostName`.
 - c. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/nodes.ini` и измените все вхождения параметра `nodeN.destination` на имя хоста или IP-адрес, которые вы хотите использовать для этого сервера.
 - d. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/config/nodes.ini` и измените все вхождения параметра `nodeN.destination` на имя хоста или IP-адрес, которые вы хотите использовать для этого сервера.

5. Перезапустите CCL. При конфигурации с несколькими серверами перезапустите CCL на каждом сервере поисковой системы предприятия одним из следующих способов:

AIX, Linux или Solaris

```
startccl.sh -bg
```

Окно командной строки Windows

```
startccl
```

Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск** → **Программы** → **Администрирование** → **Службы**.
 - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
6. Перезапустите сеансы поисковой системы предприятия:

```
esadmin system startall
```

Конфигурирование поддержки двух IP-адресов

Чтобы сконфигурировать для серверов, где устанавливается поисковая система предприятия, поддержку двух IP-адресов, поисковые серверы предприятия нужно вручную сконфигурировать для запуска в этой среде.

Администратор рабочего стола может, например, установив адаптер обратной связи Microsoft, создать виртуальную сеть, поддерживающую требования к сети для определенных продуктов, таких как Microsoft SQL Server.

Порядок действий:

Чтобы сконфигурировать для поисковой системы предприятия возможность запуска на серверах, поддерживающих два IP-адреса:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. При конфигурации с несколькими серверами зарегистрируйтесь на любом поисковом сервере предприятия.
2. Остановите сеансы поисковой системы предприятия:

```
esadmin system stopall
```
3. Остановите CCL (common communication layer, общий уровень связи). При конфигурации с несколькими серверами остановите CCL на каждом сервере поисковой системы предприятия одним из следующих способов:

AIX, Linux или Solaris

```
stopccl.sh
```

Окно командной строки Windows

```
stopccl
```

Администрирование - Службы Windows

- a. Запустите службы Windows: **Пуск** → **Программы** → **Администрирование** → **Службы**.
- b. Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить**.

4. На каждом поисковом сервере предприятия, поддерживающем два IP-адреса, отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/nodeinfo/es.cfg` и добавьте параметр **LocalIPAddress**. Назначьте для этого значения IP-адрес, который может быть разрешен DNS.
5. Перезапустите CCL. При конфигурации с несколькими серверами перезапустите CCL на каждом сервере поисковой системы предприятия одним из следующих способов:

AIX, Linux или Solaris

```
startccl.sh -bg
```

Окно командной строки Windows

```
startccl
```

Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск → Программы → Администрирование → Службы**.
 - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
6. Перезапустите сеансы поисковой системы предприятия:

```
esadmin system startall
```

Включение поддержки протокола IPv6

Чтобы включить поддержку адресов и URL, использующих протокол IP версии 6 (IPv6), надо задать переменную среды, которая сообщит поисковой системе предприятия, что надо использовать только адреса гнезд версии IPv6.

Прежде чем приступать к работе...

1. Убедитесь, что значения свойства **LocalHostName** и свойства **LocalIPAddress** (если оно задано) в `%ES_CFG%` - либо имя хоста, либо допустимый IPv6-адрес локального компьютера. Можно изменить эти значения вручную, если требуется, а затем сохранить `%ES_CFG%`.
2. Убедитесь, что значение свойства **DerbyServerHostName** в `%ES_CFG%` - имя хоста. Это значение не может быть IPv4-адресом или IPv6-адресом. Если вы изменили это свойство, сохраните файл `%ES_CFG%`.
3. Убедитесь, что значения свойств **destination** и **serverhost** в файле `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/nodes.ini` на сервере индексов - либо допустимый IPv6-адрес, либо допустимое имя хоста. Эти значения не могут быть IPv4-адресами. Если вы изменили этот файл, сохраните его.

Ограничения

Поддержка протокола IPv6 доступна только для поисковых систем предприятия, установленных на серверах Windows 2003. После включения поддержки IPv6-адресов поисковая система предприятия не будет больше использовать IPv4-адреса для связи гнезд.

Поддерживаемые браузеры

Браузеры Internet Explorer и Mozilla Firefox обрабатывают IPv6-адреса по-разному.

Mozilla Firefox

Для запуска консоли администратора поисковой системы предприятия или программы поиска можно задать в URL IPv6-адрес или имя хоста. Например:

```
http://[2001::db8]/ESAdmin  
http://SearchServer.com/ESSearchApplication/
```

Internet Explorer

Для запуска консоли администратора поисковой системы предприятия или программы поиска нельзя задавать в URL IPv6-адрес. Надо использовать приведенный ниже формат и обеспечить отображение имени хоста на IPv6-адрес в DNS или в файле `c:\windows\system32\etc\hosts`:

```
http://SearchServer.com/ESAdmin/  
http://SearchServer.com/ESSearchApplication/
```

Если задать в URL IPv6-адрес, выводится сообщение **Ошибка: недопустимый синтаксис**. Дополнительную информацию об этом ограничении смотрите в документе <http://support.microsoft.com/kb/325414>.

Поддержка источников данных IPv4

Источники данных, находящиеся на сервере IPv4, поддерживаются искомателями поисковой системы предприятия. Если при конфигурировании искомателя Web вы задает начальные URL в формате IPv6, не забудьте заключить такие URL в квадратные скобки. Например:

```
http://[2001:db8:0:1:0:0:0:1]  
http://[2001:db8:0:1::1]
```

Процедура

Чтобы включить поддержку протокола IPv6:

1. На сервере индексов поисковой системы предприятия зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему.
`esadmin system stopall`

Важное замечание: После остановки системы пользователи не смогут отправлять требования поиска.

2. Остановите слой общей связи (common communication layer, CCL) на всех серверах поисковой системы предприятия:
 - a. Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
 - b. Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить**.
3. Определите системную переменную среды **ES_IPV6=1**. Для установки поисковой системы предприятия на нескольких серверах выполните этот шаг на серверах поиска.
4. Обновите пользовательские свойства виртуальной Java-машины (JVM) в WebSphere Application Server. Для установки поисковой системы предприятия на нескольких серверах выполните эти шаги на серверах поиска.
 - a. Остановите сервер прикладных программ ESSearchServer в WebSphere Application Server, если он работает.

- b. Откройте консоль администратора WebSphere Application Server и перейдите на панель свойств виртуальной Java-машины. Выберите **Серверы → Серверы прикладных программ → ESSearchServer → Управление процессом и Java → Определение процесса → Виртуальная Java-машина → Пользовательские свойства**.
 - c. Сконфигурируйте следующие свойства:


```
java.net.preferIPv4Stack=false
java.net.preferIPv6Addresses=true
```

Если пользовательского свойства еще нет, создайте новое свойство и введите имя свойства в поле **Имя**, а допустимое значение - в поле **Значение**.
 - d. Если вы используете WebSphere Application Server Версии 6.1, добавьте свойство **LocalIPAddress** в файл ES_NODE_ROOT/nodeinfo/es.cfg; его значением будет IPv6-адрес поискового сервера.
 - e. Перезапустите программу ESSearchServer.
5. Если вы используете WebSphere Application Server Версии 6.1, выполните следующие шаги. Для установки поисковой системы предприятия на нескольких серверах выполните эти шаги на серверах поиска.
 - a. Отредактируйте файл httpd.conf для IBM HTTP Server.
 - b. Удалите символ # в начале следующей строки, чтобы раскомментировать инструкцию для сервера IBM HTTP Server ожидать соединений IPv6 на порте 80:


```
# Listen [::]:80
```
 - c. Перезапустите IBM HTTP Server.
 6. Перезапустите CCL на всех серверах поисковой системы предприятия:
 - a. Запустите службы Windows: **Пуск → Программы → Инструменты управления → Службы**.
 - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
 7. Перезапустите поисковую систему предприятия:


```
esadmin system startall
```
 8. Посмотрите файл КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/node/logs/cc10.log, чтобы убедиться, что сервер CCL запущен без ошибок. Если протокол IPv6 разрешен правильно, будут записаны примерно следующие сообщения:


```
INFO: CCL server ready for business. Now, waiting for external requests.
CCL host name is fe80::250:56ff:feb4:27d1
CCL host dot.ip address is fe80:0:0:0:250:56ff:feb4:27d1
CCL server port is 6002
Total JVM Runtime memory is 33022Kb
Current free memory is 21346Kb
java.net.preferIPv6Addresses=true
java.net.preferIPv4Stack=false
INFO: Session "TraceDaemonSession" was attached PID=4992
```
 9. Выполните следующие действия, чтобы проверить правильность работы программ поиска и управления:
 - a. Запустите серверы поиска для любого собрания, затем откройте пример программы поиска и введите запрос. Этот шаг проверяет, что процедура принятия запроса действует, даже если в индексе поиска нет документов.
 - b. Откройте консоль администратора и убедитесь, что вам удастся без ошибок зарегистрироваться.

Отключение поддержки протокола IPv6

1. Посмотрите описание в разделе **Прежде чем начать** и замените каждое значение, представляющее собой IPv6-адрес, на имя хоста или на IPv4-адрес.
2. Выполните те же действия, что и для включения поддержки протокола IPv6, за следующими исключениями:
 - На шаге 3 на стр. 29 удалите переменную среды **ES_IPV6**.
 - На шаге 4 на стр. 29 удалите пользовательские свойства JVM properties, заданные для программы ESSearchServer в WebSphere Application Server.
3. Если свойство **LocalIPAddress** в файле ES_NODE_ROOT/nodeinfo/es.cfg содержит IPv6-адрес, удалите его.
4. Выполните действия шага 9 на стр. 30, чтобы проверить программы поиска и управления.

Собрания поисковой системы предприятия

Собрание поисковой системы предприятия содержит полный набор источников, в которых пользователи могут выполнять поиск при помощи одного запроса. Используя объединение, пользователи при помощи одного запроса могут выполнять поиск в нескольких собраниях.

При создании собрания вы задаете опции, применяемые ко всему собранию. Новое собрание будет пустым, пока вы не добавите в него содержимое.

Добавить собрания в поисковую систему предприятия можно двумя способами:

- Если вы еще не освоили консоль администратора поисковой системы предприятия или только изучаете, как компоненты собрания работают совместно, создайте собрание при помощи мастера по собраниям. Мастер по собраниям помогает выполнять задачи и позволяет сохранять вашу работу в качестве черного варианта создаваемого собрания.
- Если вы знакомы с консолью администратора поисковой системы предприятия, возможно, вы предпочтете создать собрание, выбирая для управления конкретные страницы в окне Собрания.

После создания собрания, используя элементы управления в окне Собрания, можно изменить собрание, поисковую систему предприятия и опции защиты, а также задать для них параметры слежения.

Объединение собраний

Если в программу поиска встроена поддержка объединения, пользователи могут выполнять поиск одновременно в нескольких собраниях. Объединение позволяет преодолеть предельно допустимый размер для собрания, равный 20 миллионам документов на собрание. Например, пользователь может выполнить поиск в двух собраниях, содержащих по 20 миллионов документов каждое.

Качество поиска зависит от областей действия, генерируемых отдельными собраниями, которые затем сливаются, образуя конечный набор результатов. Результаты будут такими же, как и при передаче двух отдельных запросов поиска с последующим объединением и ранжированием их результатов.

Задачи, связанные с данной

“Слежение за собранием” на стр. 306

Создание собрания при помощи мастера по собраниям

Если вы не знакомы с поисковой системой предприятия, мастер может помочь вам при создании собрания. Мастер дает подробную информацию о каждом этапе процесса и позволяет сохранять параметры по ходу работы.

Прежде чем приступить к работе...

Создать собрание может только пользователь с ролью администратора поисковой системы предприятия.

Добавить содержимое в собрание или задать опции анализа, индексирования и поиска содержимого собрания может администратор поисковой системы предприятия или администратор этого собрания.

Ограничения

Мастер по собраниям можно использовать для создания искателей таких типов:

- Content Edition
- DB2
- DB2 Content Manager
- Exchange Server
- Примечания
- Файловая система UNIX
- Web
- Файловая система Windows

Об этой задаче

При создании собрания можно сохранить его в черновом состоянии. Если собрание находится в черновом состоянии, любой администратор, имеющий право управлять этим собранием, может вносить в него изменения. Например, вы можете попросить администратора собрания, имеющего опыт работы с источниками Lotus Notes, сконфигурировать искатель Примечания. Затем администратор собрания, имеющий опыт работы с системами UNIX, может изменить черновое собрание, чтобы сконфигурировать искатель Файловая система UNIX.

Процедура

Чтобы использовать мастер по собраниям для создания собрания:

1. Выберите **Собрания**, чтобы открыть окно Собрания.
2. Выберите **Мастер по собраниям**.
3. Следуйте указаниям мастера, чтобы создать пустое собрание и добавить в него содержимое.
Нужно сконфигурировать общую информацию о собрании и создать хотя бы один искатель. Можно принять значения по умолчанию для оставшихся опций конфигурации или задать опции для вашего нового собрания.
4. Чтобы сохранить собрание до того, как вы закончите его создание, выберите **Сохранить как черновик**.
Ваше собрание находится в списке остальных черновых собраний в окне Собрания. Если для этого собрания включена защита, рядом с именем собрания будет показан значок  **Защита собрания включена**.
5. Чтобы вернуться к собранию, которое вы все еще создаете, выберите **Вернуться в мастер** в окне Собрания.
6. Нажмите кнопку **Закончить**, чтобы создать собрание.
Ваше собрание будет внесено в список остальных собраний в окне Собрания.

После создания собрания нужно запустить процессы искателя, анализатора, индексирования и поиска для этого собрания. Пока вы не будете готовы связать собрание с программами поиска, которые будут искать информацию в нем, можно использовать для поиска в вашем новом собрании пример программы поиска (под именем Default).

Создание собрания при помощи окна Собрания

Для создания пустого собрания служит окно Собрания. После этого можно отредактировать это собрание, указав опции для добавления данных в собрание и создания системы поиска в собрании.

Прежде чем приступить к работе...

Создать собрание может только пользователь с ролью администратора поисковой системы предприятия.

Добавить содержимое в собрание или задать опции анализа, индексирования и поиска содержимого собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Чтобы получить информацию о значениях, которые можно задать для нового собрания, при создании собрания нажмите кнопку **Справка**.

Процедура

Чтобы создать собрание из окна Собрания:

1. В окне Собрания выберите **Создать собрание**.
2. На странице Создать собрание введите или выберите информацию для следующих полей:
 - **Имя собрания.** Введите имя, соответствующее содержанию или назначению собрания.
 - **Защита собрания.** Укажите, хотите ли вы включить защиту для этого собрания. После создания собрания этот параметр нельзя будет изменить. Если защита собрания включена, далее можно указать опции для обеспечения управления доступом на уровне отдельных документов.
 - **Важность документов (статическая модель ранжирования).** Укажите стратегию присвоения критерия статического ранжирования, который будет использоваться для упорядочения документов в результатах поиска. После создания собрания изменить это значение нельзя.
 - **Тип категоризации.** Укажите, хотите ли вы искать документы по категориям, к которым они принадлежат.
 - **Используемый язык.** Укажите язык по умолчанию для поиска документов в собрании.
3. Примите для следующих полей значения по умолчанию или задайте опции, которые надо использовать для этого собрания:
 - **Описание.** По умолчанию описание не создается.
 - **Оценка числа документов.** По умолчанию для одного собрания принимается оценка 1000000 документов. Система использует это значение для оценки необходимых для собрания ресурсов памяти и дискового пространства, но не для ограничения размера самого собрания.
 - **Положение данных собрания.** По умолчанию файлы, имеющие отношение к собранию, находятся на сервере индексов. После создания собрания изменить это значение нельзя.
 - **ID собрания.** По умолчанию ID собрания основывается на имени собрания. После создания собрания изменить это значение нельзя. Если вы указали

пользовательский ID собрания, поисковые программы будут вызывать это собрание по этому идентификатору вместо созданного системой и, возможно, зашифрованного идентификатора.

- **Сегментирование энграмм.** По умолчанию используется метод сегментирования пробелами на основе Unicode. Выберите эту опцию, чтобы использовать сегментирование энграмм, только в том случае, когда ваше собрание включает в себя документы на китайском, японском или корейском языках и вы хотите, чтобы анализатор использовал для ограничения слов сегментирование энграмм вместо сегментирования пробелами. После создания собрания изменить это значение нельзя.

Дополнительные сведения о том, как сконфигурировать поддержку полного анализа и разбиения энграмм в собраниях поисковой системы предприятия, смотрите в документе <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011088>.

4. Нажмите кнопку **ОК**.

В окне Собрания среди других собраний в поисковой системе предприятия появится новое собрание. Если для этого собрания включена защита, рядом с именем собрания будет показан значок  **Защита собрания включена**.

Новое собрание будет пустым, пока вы не добавите в него содержимое. Чтобы добавить содержимое в новое собрание, выберите это собрание в окне Собрания, отредактируйте его, создайте хотя бы один искатель и задайте опции для анализа, индексации и поиска данных.

Затем надо запустить процессы для работы искателя, анализа, индексации и поиска в собрании. До того, как вы сделаете новое собрание доступным для поиска при помощи пользовательских поисковых программ, можно использовать пример поисковой программы.

Изменение собрания

При редактировании собраний вы задаете информацию о документах, которые надо включить в собрание.

Прежде чем приступить к работе...

Редактировать собрание может пользователь с ролью администратора поисковой системы предприятия или администратор этого собрания.

Об этой задаче

При редактировании собрания надо задать опции просмотра искателем источников данных, анализа документов, построения индекса, поиска в индексированном содержимом и записи в журнал сообщений об ошибках. При создании собрания его надо отредактировать, чтобы добавить в него содержимое. Позднее можно редактировать собрание для изменения содержимого или для изменения способа просмотра информации искателем, ее анализа, индексирования, поиска или записи в журнал.

Процедура

Чтобы отредактировать собрание:

1. Выберите на панели инструментов **Собрания**, чтобы открыть окно Собрания.

2. Найдите в списке собраний собрание, которое нужно отредактировать, и выберите  Удалить .
3. Измените параметры на любых из следующих страниц:

Просмотр

Задайте источники данных, которые должен просматривать искатель, и опции просмотра искателем содержимого. В каждом собрании должен быть как минимум один искатель; в одном собрании могут содержаться данные из нескольких типов источников данных. Для каждого типа источника данных, который нужно включить в собрание, надо сконфигурировать хотя бы по одному искателю.

Анализ

Задайте опции для просмотра документов искателем, их синтаксического и лексического анализа. Можно сконфигурировать категории, позволяющие пользователям выполнять поиск в подмножествах собрания, а также сконфигурировать правила, позволяющие пользователям выполнять поиск по определенным частям документов XML и HTML. Если в поисковую систему предприятия добавлены пользовательские механизмы тестового анализа, можно выбрать механизм для анализа и составления аннотаций содержимого в этом собрании. Поля можно также связать с классами значимости, чтобы изменить ранжирование документов в результатах поиска.

Индекс

Задайте расписания полного построения индекса и обновления индекса с учетом нового и измененного содержимого. Можно также сконфигурировать опции для использования символов подстановки в запросах, для ограничения просмотра собрания диапазоном URI, для объединения результатов поиска с одного сайта и для удаления URI из индекса.

Поиск Задайте опции для процессов поиска, например, конфигурирования кэша поиска и выбора языка поиска. Кроме того, можно сконфигурировать прямые ссылки - эта функция позволяет возвращать заранее определенные URI, когда пользователь включает в запрос определенные слова и словосочетания. Если в поисковую систему предприятия были добавлены пользовательские словари, можно выбрать словари, которые вы хотите использовать для поиска в этом собрании.

Журнал

Задайте типы сообщений, которые должны записываться в журнал, и опции для создания файлов журнала и перехода к новым файлам журнал. Можно сконфигурировать оповещения, чтобы получать уведомления об определенных событиях, задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов. Можно также указать опции для записи в журнал информации о документах, отброшенных из поисковой системы предприятия.

Общие Задайте общую информацию о собрании и просмотрите опции, которые нельзя изменить. Если при создании собрания для него включена защита, можно сконфигурировать опции управления защитой на уровне документов.

Удаление собрания

Удаление собрания полностью удаляет всю информацию о собрании из поисковой системы предприятия.

Прежде чем приступить к работе...

Удалить собрание может только пользователь с ролью администратора поисковой системы предприятия.

Перед удалением собрания нужно остановить все процессы, связанные с этим собранием.

Об этой задаче

Операция удаления собрания может занять определенное время. Когда вы подтвердите, что хотите удалить это собрание, система удаляет все данные в системе, связанные с этим собранием.

Совет: Сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажмите кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Процесс удаления завершен, когда имя этого собрания более не выводится в списке собраний.

Процедура

Чтобы удалить собрание:

1. Выберите **Собрания**, чтобы открыть окно Собрания.
2. В списке собраний найдите собрание, которое хотите удалить, и выберите  **Удалить**.

Определение ID собрания

Для многих задач управления требуется знать ID собрания.

Прежде чем приступить к работе...

Смотрите ID собрания может только пользователь с ролью администратора поисковой системы предприятия.

Об этой задаче

При создании собрания вы либо задаете значение ID, либо позволяете системе назначить его. Чтобы определить ID собрания после его создания, можно использовать консоль администратора или посмотреть файл конфигурации.

Процедура

1. Чтобы узнать ID собрания при помощи консоли администратора:
 - a. Выберите **Собрания**, чтобы открыть окно Собрания.
 - b. На странице Общие выберите **Посмотреть параметры собрания**.

На странице Параметры собрания выводится ID собрания, полный путь к месту хранения данных собрания и модель статического ранжирования, используемая для документов в этом собрании.

2. Чтобы узнать ID собрания при помощи файла конфигурации:
 - a. Откройте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/collections.ini. Для удобства просмотра отсортируйте этот файл. В следующем примере вывода coll - ID собрания:

```
% sort $ES_NODE_ROOT/master_config/collections.ini | more
collection1.configfile=coll_config.ini
collection1.datadir=/home/eseach/node/data/coll
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=coll
collection1.sectiontype=collection
collection1.type=1
...
```

Управление искателями

Вы конфигурируете искатели для различных типов данных, которые нужно включить в собрание. Одно собрание может содержать любое число искателей.

Совет: На странице <http://www.ibm.com/developerworks/edu/dm-dw-dm-0503buehler-i.html> доступна электронная обучающая программа. В ней описаны действия по установке и конфигурированию, показано, как выполнять поиск в источниках данных различных типов, и описано использование интерфейсов прикладного программирования продукта для расширения поисковой системы предприятия. Эта обучающая программа написана для более старой версии OmniFind Enterprise Edition, но большинство понятий и процедур все еще применимы.

Подробные примеры способов конфигурирования искателей и включения защиты в небольших, средних и крупномасштабных организациях смотрите в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурирование искателей

При помощи консоли администратора поисковой системы предприятия можно создавать, редактировать и удалять искатели. Обычно конфигурированием искателя занимается специалист по типу данных для искателя. Например, чтобы настроить искатель для работы с источниками Lotus Notes, администратор собрания должен быть администратором Notes или работать в тесном сотрудничестве с кем-то, кто знает особенности баз данных, которые будет просматривать искатель.

При создании искателя мастер для конкретного типа данных, с которыми будет работать искатель, поможет задать свойства, управляющие способом использования искателем системных ресурсов. Этот мастер поможет также выбрать источники, в которых будет выполняться поиск.

В любое время можно внести изменения в существующие искатели. При необходимости можно изменить свойства искателя или частей пространства для искателя. Внести эти изменения помогут также мастера по искателям.

Заполнение нового искателя основными значениями

При создании искателя можно использовать системные значения по умолчанию или скопировать значения, заданных для другого искателей того же типа. Если взять за основу для нового искателя существующий искатель, можно быстро создать несколько искателей с одинаковыми свойствами, а затем сконфигурировать их, например, для просмотра различных источников или для работы по различным расписаниям.

Создав копии искателя, можно разделить рабочую нагрузку по нескольким искателям, использующим одни и те же правила работы с данными. Например, можно создать копию искателя Примечания, чтобы использовать те же свойства и правила просмотра полей искателем на другом сервере Lotus Notes. Они могут отличаться только наборами баз данных, просматриваемых каждым искателем, и установками защиты на уровне документов.

Комбинирование типов искателей в собрании

Искатели поисковой системы предприятия предназначены для сбора информации из источников данных различных типов. При конфигурировании искателей для собрания нужно решить, как скомбинировать эти различные типы источников данных, чтобы пользователи могли легко выполнить поиск данных вашего предприятия. Например, чтобы пользователи могли выполнять поиск данных в файловых системах Microsoft Windows и общедоступных папках Microsoft Exchange Server при помощи одного запроса, создайте собрание, включающее искатели Файловая система Windows и искатели Exchange Server.

Если в одном собрании будут работать искатели нескольких типов, убедитесь, что все они могут использовать один метод статического ранжирования. (Метод статического ранжирования задается при создании собрания.) Например, если вы объединяете источники Web (для которых для ранжирования используются ссылки на документы) и источники (для которых для ранжирования обычно используются даты документов), это может негативно повлиять на качество результатов поиска.

Конфигурирование защиты на уровне документов

Если при создании собрания вы включили для него защиту, можно сконфигурировать опции защиты на уровне документов. Каждый искатель может связать с просматриваемыми им документами маркеры защиты. Если при конфигурировании искателя задать использование защиты на уровне документов, искатель свяжет заданные вами маркеры защиты с каждым документом, и они будут добавлены с документами в индекс.

Если в пользовательских программах поиска включить защиту, эти программы смогут использовать маркеры защиты, связываемые искателями с документами, для аутентификации пользователей. Эта возможность позволяет ограничить доступ к некоторым документам в собрании и разрешить всем пользователям выполнять поиск в других документах. Например, в одном собрании можно разрешить доступ ко всем документам в общедоступных папках Microsoft Exchange Server всем пользователям, но доступ к документам в базе данных Lotus Notes разрешить только пользователям с конкретными ID пользователей.

Можно применять пользовательские бизнес-правила для определения значения маркеров защиты путем шифрования правил в классе Java. При конфигурировании свойств искателя вы задаете имя подключаемого модуля, который будет использовать искатель при просмотре документов. Маркеры защиты, добавляемые этим подключаемым модулем, сохраняются в индексе и могут использоваться для управления доступом к документам.

При конфигурировании определенных типов искателей можно задать дополнительные элементы управления защитой. Например, можно указать, что вы хотите выполнять проверку пользователей во время обработки запросов. Если включить эту опцию, учетные данные пользователя сравниваются с текущими списками управления доступом, поддерживаемыми источниками данных, в которых будет выполняться поиск. Такая проверка текущих учетных данных может выполняться вместо или помимо проверки на основе маркеров защиты в индексе поисковой системы предприятия.

Понятия, связанные с данным

“Защита на уровне документов” на стр. 265

Задачи, связанные с данной

“Отслеживание работы искателей” на стр. 308

Создание искателя

При создании искателя вы задаете его тип. Мастер помогает задать информацию о данных, которую вы хотите включить в собрание.

Прежде чем приступить к работе...

Создать искатель может пользователь с ролью администратора поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

Для собрания нужно создать хотя бы один искатель. Тип создаваемого искателя зависит от типа данных, которые вы хотите включить в это собрание. Мастер для создаваемого типа искателя поможет задать опции искателя. Например, мастер поможет задать опции использования ресурсов системы для искателя. Мастер поможет также выбрать источники данных, которые вы хотите включить в собрание.

Процедура

Чтобы создать искатель:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Создать искатель**.
2. Выберите тип искателя и основные значения для искателя:
 - a. Выберите тип искателя, поддерживающий тип данных, в которых вы собираетесь выполнять поиск - сайты, базы данных Lotus Notes или файловые системы UNIX.
Когда вы выберете тип искателя, будут показаны варианты создания искателя.
 - b. Выберите основные значения искателя:

Использовать для нового искателя системные значения по умолчанию

Заполняет исходные значения для этого искателя установочными значениями по умолчанию.

Если вы выбрали эту опцию, нажмите кнопку **Далее**, чтобы начать конфигурирование нового искателя.

Скопировать значения из существующего искателя для нового искателя

Заполняет начальные значения для этого искателя значениями, сконфигурированными для другого искателя этого типа.

Если вы выбрали эту опцию, будет показан список искателей этого типа. Выберите искатель, который вы хотите использовать как основу для нового искателя, и нажмите кнопку **Далее**, чтобы начать конфигурирование нового искателя.

Откроется мастер для создаваемого типа искателя. Чтобы создать искатель, следуйте подсказкам мастера. Чтобы узнать больше об опциях, которые вы можете задать для этого типа искателя, выберите **Помощь** на любой странице мастера.

Ваш новый искатель находится в списке других искателей, принадлежащих этому собранию, на странице Искатель. Если вам нужно внести изменения в искатель, выберите Опции, чтобы изменить свойства искателя и пространство для искателя.

Изменение свойств искателя

Можно изменять информацию о искателе и о том, как он обрабатывает данные. Например, можно изменить использование искателем ресурсов системы.

Прежде чем приступать к работе...

Редактировать свойства искателя может только администратор поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

При редактировании свойств искателя выберите **Помощь**, чтобы узнать больше об изменениях, которые вы можете внести. Доступные для изменения свойства зависят от типа искателя.

Процедура

Чтобы изменить свойства искателя:

1. Откройте собрание для редактирования, на странице Искатель найдите искатель, который вы хотите отредактировать, и выберите  **Свойства искателя**.
2. Измените свойства искателя и нажмите кнопку **ОК**.
3. Чтобы изменения вступили в силу, остановите и перезапустите искатель. (Если вы изменили только описание искателя, перезапуск не требуется.)

Изменение пространства для искателя

Можно изменять информацию о источниках данных, которые просматривает искатель. Например, можно добавлять источники данных, удалять источники данных, изменять расписание просмотра и правила просмотра документов из определенного источника данных.

Прежде чем приступать к работе...

Редактировать пространство искателя может только администратор поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

Чтобы узнать об изменениях, которые вы можете внести для типа искателя, который вы редактируете, выберите **Помощь** во время изменения пространства для искателя.

Процедура

Чтобы редактировать пространство для искателя:

1. Откройте собрание для редактирования, на странице Искатель найдите искатель, который вы хотите отредактировать, и выберите  **Пространство для искателя**.
2. Измените пространство для искателя, выбрав опции, которые вы хотите изменить.
Опции, которые можно выбрать, зависят от типа искателя. Для некоторых опций, таких, как добавление источников данных в собрание, открывается мастер для этого типа искателя, который поможет вам изменить пространство для искателя.
3. Чтобы изменения вступили в силу, остановите и перезапустите искатель.

Удаление искателя

При удалении искателя вся информация о нем из поисковой системы предприятия удаляется. Информация, найденная ранее этим искателем, останется в индексе до следующего полного построения индекса.

Прежде чем приступить к работе...

Удалить искатель может пользователь с ролью администратора поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

Операция удаления искателя может занять определенное время. Когда вы подтвердите, что хотите удалить этот искатель, система удаляет все данные в системе, связанные с этим искателем.

Совет: Для этого может потребоваться время, и сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажимайте кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Когда имя искателя исчезнет из списка искателей, процесс удаления завершился.

Процедура

Чтобы удалить искатель:

1. Откройте собрание для редактирования и выберите страницу Искатель.
2. Найдите искатель, который вы хотите удалить, и выберите  **Удалить**.

Расписания искателя

Искатели, создаваемые для источников Web, работают постоянно. После запуска искателя Web, как правило, его можно не останавливать, пока не потребуются изменить конфигурацию искателя. При конфигурировании искателей всех остальных типов нужно составить расписание просмотра ими данных.

Для некоторых типов источников данных одно расписание управляет временем обращения искателя ко всем источникам данных в пространстве искателя. Для других источников данных можно задать разные расписания обращения к конкретным источникам данных. Например, можно задать отдельные расписания просмотра для каждой просматриваемой искателем базы данных Lotus Notes.

При конфигурировании расписания задается тип выполняемой работы искателя. Можно запланировать полную обработку искателем всех документов в пространстве работы искателя, запланировать работу искателя, включающую все обновления его рабочего пространства (новые документы, измененные документы и удаленные документы) или запланировать работу искателя, включающую в себя только новые или измененные документы. Больше всего времени требуется для полной обработки искателем. Работа искателя с удалением удаленных документов требует больше времени чем работа, при которой удаленные документы игнорируются.

При редактировании пространства работы искателя можно задать второе расписание работы искателя. Например, можно сконфигурировать одно расписание для

обработки всех документов в пространстве работы искателя в субботу вечером, а другое расписание - для более частой обработки новых и измененных документов.

Создание нескольких расписаний искателя позволяет точнее управлять временем обращения искателя к источникам назначения. Например, для просмотра искателем баз данных в разных часовых поясах можно задать расписание работы искателя на время наиболее вероятного окончания рабочего дня пользователями в этих поясах.

Искатели Content Edition

Чтобы включить репозитории IBM WebSphere Information Integrator Content Edition в собрание поисковой системы предприятия, нужно сконфигурировать искатель Content Edition.

Искатель Content Edition можно использовать для просмотра репозиторий следующих типов:

- Documentum
- FileNet P8 Content Manager
- FileNet Panagon Content Services
- Hummingbird Document Management (DM)
- Microsoft SharePoint
- OpenText Livelink
- Portal Document Manager (PDM)

При конфигурировании искателя нужно задать опции для просмотра искателем всех репозиторий в пространстве искателя. Нужно также выбрать типы классов, которые должен просматривать искатель в каждом репозитории.

Совет:

Подробные примеры конфигурирования соединителей и защищенного искателя Content Edition смотрите в сценарии для средней организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурация сервера искателя

Подготовка репозиторий для обработки искателем зависит от того, планируется ли использовать для соединения с просматриваемыми данными непосредственный или серверный режим. При выборе непосредственного режима нужно сконфигурировать соединитель в WebSphere Information Integrator Content Edition. При использовании серверного режима на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю Content Edition взаимодействовать с серверами WebSphere Information Integrator Content Edition.

Если планируется использовать серверный режим, перед созданием искателя Content Edition выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition” на стр. 49.
- “Конфигурирование сервера искателей в Windows для WebSphere II Content Edition” на стр. 50.

Подробные инструкции о том, как сконфигурировать поисковую систему предприятия для поиска в репозиториях WebSphere Information Integrator Content

Edition, смотрите в статье *IBM developerWorks Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* по адресу <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-06061ee/>.

Конфигурирование - обзор

Чтобы создать или изменить искатель Content Edition, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех репозиториях в пространстве искателя.
- Указать, использует ли искатель при доступе к репозиториям прямой или же серверный режим. Для серверного режима можно также задать информацию, обеспечивающую искателю доступ к серверу прикладных программ Web.
- Выбрать репозитории, которые будет просматривать искатель.
- Задать ID пользователей и пароли для возможности обращения искателя к содержимому в выбранных репозиториях.
- Составить расписания просмотра искателем репозиториях.
- Выбрать в каждом репозитории классы элементов, просматриваемые искателем.
- Задать опции, обеспечивающие доступность поиска для свойств классов элементов. Например, можно исключить из пространства искателя определенные типы документов или указать, что нужен просмотр искателем конкретной версии репозитория.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволят программам поиска применять управление доступом на основе списков управления доступом или маркеров защиты.

Для классов элементов Documentum, FileNet Panagon Content Services, Hummingbird DM, Portal Document Manager и SharePoint можно также выбрать опцию для проверки правильности параметров регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает учетные данные пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных. Для других типов репозиториях этот тип проверки правильности текущих параметров аутентификации недоступен.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Доступ к репозиториям Content Edition в непосредственном режиме

Для искателя Content Edition можно сконфигурировать доступ к репозиториям WebSphere Information Integrator Content Edition в непосредственном режиме.

Об этой задаче

В непосредственном режиме искатель использует соединитель WebSphere Information Integrator Content Edition, устанавливаемый на сервере искателя при установке OmniFind Enterprise Edition. Для непосредственного соединения с просматриваемыми репозиториями искатель использует API интеграции содержимого. При работе сервера интеграции содержимого в непосредственном режиме доступны не все его функциональные возможности. Информацию о работе сервера интеграции содержимого в непосредственном режиме и различиях в его функциональных возможностях при работе в непосредственном и серверном режимах смотрите в документации WebSphere Information Integrator Content Edition.

В эту процедуру включены действия, необходимые для настройки доступа в непосредственном режиме. Подробные инструкции смотрите в статье IBM developerWorks *Search WebSphere Portal Document Manager using WebSphere Information Integrator OmniFind Edition* (URL <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0606lee/>).

Процедура

Чтобы сконфигурировать в системе возможность обращения искателя к репозиториям в непосредственном режиме:

1. Убедитесь, что переменные среды VBR_HOME и JAVA_HOME в файле *каталог_установки_ice/bin/config.sh* (в UNIX) или в файле *каталог_установки_ice\bin\config.bat* (в Microsoft Windows) задают правильный каталог.
2. Чтобы сконфигурировать запуск консоли администратора WebSphere Information Integrator Content Edition в непосредственном режиме, добавьте системное свойство Java **-Dvbr.as.operationMode=direct** в файл *каталог_установки_ice/bin/Admin.sh* (в UNIX) или в файл *каталог_установки_ice\bin\Admin.bat* (в Windows). Например:

Файл Admin.sh

```
java -classpath \  
"$VBR_CLASSPATH" \  
-Dvbr.home="$VBR_HOME" \  
-Dvbr.as.operationMode=direct \  
-Dlog4j.category.com.venetica.vbr.tools.admin=WARN \  
com.venetica.vbr.tools.admin.AdminFrame $1 $2 $3 $4
```

Файл Admin.bat

```
java -classpath "%VBR_CLASSPATH%" ^  
-Dvbr.home="%VBR_HOME%" ^  
-Dvbr.as.operationMode=direct ^  
-Dlog4j.category.com.venetica.vbr.tools.admin=WARN ^  
com.venetica.vbr.tools.admin.AdminFrame %*
```

3. Запустите консоль администратора WebSphere Information Integrator Content Edition в непосредственном режиме и сконфигурируйте соединитель для сервера искателя OmniFind Enterprise Edition.
4. При конфигурировании искателя Content Edition с помощью консоли управления поисковой системой предприятия выберите опцию непосредственного режима.

Доступ к репозиториям WebSphere II Content Edition в режиме сервера

Для искателя Content Edition можно сконфигурировать доступ к репозиториям в режиме сервера.

В режиме сервера соединитель WebSphere Information Integrator Content Edition, используемый искателем для доступа к данным, устанавливается как прикладная

программа предприятия на сервер WebSphere Application Server, и искатель обращается к репозиториям через сервер. Этот подход позволяет воспользоваться преимуществами среды сервера прикладных программ J2EE.

Прежде чем конфигурировать для искателя доступ к репозиториям WebSphere Information Integrator Content Edition в режиме сервера, необходимо запустить сценарий на сервере искателя. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю Content Edition обращаться к репозиториям на сервере.

Завершите задачу, соответствующую вашей среде:

- “Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition”.
- “Конфигурирование сервера искателей в Windows для WebSphere II Content Edition” на стр. 50.

Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition

Если вы устанавливаете OmniFind Enterprise Edition на компьютере, где запускается IBM AIX, Linux или операционная среда Solaris, и конфигурируете для искателя Content Edition использование режима сервера при обращении к репозиториям, нужно запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю Content Edition обращаться к репозиториям WebSphere Information Integrator Content Edition.

Об этой задаче

Искатель Content Edition использует библиотеки Java WebSphere Information Integrator Content Edition как клиент Java. В режиме сервера для этих библиотек Java требуются библиотеки Java WebSphere Application Server, связанные с EJB. Чтобы искатель Content Edition мог работать с этими библиотеками Java, после установки WebSphere Application Server нужно запустить на сервере искателя сценарий установки, который OmniFind Enterprise Edition записывает на сервер искателя после установки WebSphere Application Server.

WebSphere Information Integrator Content Edition устанавливается на сервере искателя при установке OmniFind Enterprise Edition. Для возможности использования искателя Content Edition в режиме сервера на сервере искателя нужно скопировать файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition.

Процедура

Чтобы сконфигурировать сервер искателя для возможности просмотра репозитория WebSphere Information Integrator Content Edition:

1. Если устанавливается OmniFind Enterprise Edition в многосерверной конфигурации, установите и свяжите библиотеки Java WebSphere Application Server.
2. На сервере искателя запустите сценарий конфигурирования для искателя Content Edition:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия.
 - b. Запустите следующий сценарий (устанавливаемый в каталоге `$ES_INSTALL_ROOT/bin`) и ответьте на вопросы программы:
`escrvbr.sh`
3. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall
esadmin system startall
```

4. Скопируйте файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition на сервер искателя.

Скопируйте из:

Файл `vbr_access_services.jar` находится в следующем положении по умолчанию:

корневой_каталог_уст_was/installedApps/имя_серв/имя_программы

корневой_каталог_уст_was - каталог установки WebSphere Application Server, *имя_серв* - имя, заданное для этого сервера, *имя_программы* - имя, заданное для прикладной программы WebSphere Information Integrator Content Edition в WebSphere Application Server.

Скопируйте в:

Каталог назначения на сервере искателя - *корневой_каталог_установки_ице/lib*, где *корневой_каталог_установки_ице* - каталог установки WebSphere Information Integrator Content Edition на сервере искателя.

Конфигурирование сервера искателей в Windows для WebSphere II Content Edition

Если вы устанавливаете OmniFind Enterprise Edition на компьютере Microsoft Windows и конфигурируете для искателя Content Edition использование режима сервера при обращении к репозиториям, нужно запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю Content Edition обращаться к репозиториям WebSphere Information Integrator Content Edition.

Об этой задаче

Искатель Content Edition использует библиотеки Java WebSphere Information Integrator Content Edition как клиент Java. В режиме сервера для этих библиотек Java требуются библиотеки Java WebSphere Application Server, связанные с EJB. Чтобы искатель Content Edition мог работать с этими библиотеками Java, после установки WebSphere Application Server нужно запустить на сервере искателя сценарий установки, который OmniFind Enterprise Edition записывает на сервер искателя после установки WebSphere Application Server.

WebSphere Information Integrator Content Edition устанавливается на сервере искателя при установке OmniFind Enterprise Edition. Для возможности использования искателя Content Edition в режиме сервера на сервере искателя нужно скопировать файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition.

Процедура

Чтобы сконфигурировать сервер искателя для возможности просмотра репозитория WebSphere Information Integrator Content Edition:

1. Если устанавливается OmniFind Enterprise Edition в многосерверной конфигурации, установите и свяжите библиотеки Java WebSphere Application Server.
2. На сервере искателя запустите сценарий конфигурирования для искателя Content Edition:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке OmniFind Enterprise Edition).

- b. Запустите следующий сценарий (устанавливаемый в каталоге %ES_INSTALL_ROOT%\bin) и отвечайте на вопросы программы:
escrvbr.vbs
3. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall  
esadmin system startall
```

4. Скопируйте файл vbr_access_services.jar с сервера WebSphere Information Integrator Content Edition на сервер искателя.

Скопируйте из:

Файл vbr_access_services.jar находится в следующем положении по умолчанию:

корневой_к-лог_уст_was\installedApps\имя_серв\имя_пр-мы

корневой_каталог_уст_was - каталог установки WebSphere Application Server, *имя_серв* - имя, заданное для этого сервера, *имя_программы* - имя, заданное для прикладной программы WebSphere Information Integrator Content Edition в WebSphere Application Server.

Скопируйте в:

Каталог назначения на сервере искателя - *корневой_каталог_установки_ice\lib*, где *корневой_каталог_установки_ice* - каталог установки WebSphere Information Integrator Content Edition на сервере искателя.

Искатели DB2

С помощью искателя DB2 можно добавить базы данных IBM DB2 в собрание.

Если для объединения и создания таблиц псевдонимов для систем баз данных следующих типов используется IBM WebSphere Information Integrator, для просмотра таблиц по их псевдонимам можно использовать искатель DB2:

- CA-Datcom
- IBM DB2 для z/OS
- DB2 для iSeries
- IBM Informix
- IMS
- Oracle
- Microsoft SQL Server
- Software AG Adabas
- Sybase
- VSAM

Для каждого просматриваемого сервера баз данных нужно сконфигурировать отдельный искатель. При конфигурировании искателя нужно задать опции просмотра искателем всех баз данных на одном сервере. Кроме того, в каждой базе данных надо выбрать конкретные таблицы, которые будет просматривать искатель.

В качестве таблиц, обрабатываемых искателем, можно выбрать таблицы баз данных, таблицы псевдонимов или производные таблицы. Искатель DB2 не поддерживает объединенные таблицы.

Совет:

Подробные примеры конфигурирования защищенного искателя DB2 смотрите в сценарии для крупномасштабной организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурация сервера искателя

Для возможности просмотра таблиц баз данных искателем на сервере искателей надо установить клиент администратора DB2. Затем на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю DB2 взаимодействовать с серверами баз данных.

Перед тем, как сконфигурировать искатель DB2 с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для искателей DB2” на стр. 53.
- “Конфигурирование сервера искателей в Windows для искателей DB2” на стр. 54.

Публикация событий

Если вы используете WebSphere Information Integrator Event Publisher Edition и связали базы данных, которые хотите обработать искателем, с картами очередей публикации, искатель DB2 может использовать эти карты для просмотра обновлений таблиц этих баз данных.

Карта очередей публикации идентифицирует очередь WebSphere MQ, получающую сообщения XML при публикации обновлений таблицы базы данных. Искатель принимает из этой очереди информацию о публикации событий и обновляет свое пространство работы при обновлении таблиц (при первой обработке таблицы искателем обрабатываются все документы).

Публикация событий позволяет сделать новые и измененные документы доступными для поиска быстрее, чем при работе искателя по расписанию.

Если некоторые или все таблицы сконфигурированы для использования публикации событий, при конфигурировании искателя вы можете задать информацию, которая обеспечит искателю доступ к WebSphere MQ и картам очередей публикации.

Необходимо также проверить, что WebSphere MQ и WebSphere Information Integrator Event Publisher Edition сконфигурированы на сервере, на котором будет работать искатель, и что модуль клиента WebSphere MQ сконфигурирован на сервере искателя. Чтобы использовать публикацию событий с искателем DB2, выполните следующие задачи:

- “Конфигурирование WebSphere MQ для искателей DB2” на стр. 57.
- “Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2” на стр. 55.

Конфигурирование - обзор

Чтобы создать или изменить искатель DB2, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех баз данных на конкретном сервере баз данных.
- Задать информацию о типах баз данных, которые будет просматривать искатель. Если планируется просматривать удаленные базы данных, не внесенные в каталог локального сервера баз данных, для возможности использования искателя DB2 для просмотра этих баз данных необходимо запустить сервер администратора DB2. При конфигурировании искателя нужно также задать имя хоста и порт удаленного сервера баз данных.
- Задать базы данных, которые будет просматривать искатель.
- Задать ID пользователей и пароли для возможности обращения искателя к базам данных, где используется управление доступом.
- Составить расписания просмотра искателем баз данных.
- Выбрать в каждой базе данных таблицы, просматриваемые искателем.
Внимание: Чтобы оптимизировать производительность процессов обнаружения (и предотвратить истечение срока ожидания для процесса конфигурирования искателя), выбирайте просмотр искателем всех таблиц, только если в базе данных немного таблиц или в таблицах немного столбцов. Если сейчас выбрать определенные таблицы, впоследствии можно будет изменить пространство для искателя и добавить в собрание дополнительные таблицы.
- Выберите таблицы, которые будут обрабатываться искателем при публикации их обновлений в очереди публикации событий, и укажите информацию для доступа искателя к очереди публикации событий.
- Задать опции, обеспечивающие доступность поиска для столбцов в определенных таблицах. Например, можно включить использование конкретных столбцов в параметрических запросах или задать, какие столбцы будут возвращаться в результатах поиска.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование сервера искателей в UNIX для искателей DB2

При установке OmniFind Enterprise Edition на компьютере с IBM AIX, Linux или операционной средой Solaris необходимо запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю DB2 связываться с серверами баз данных. Если вы используете публикацию событий, сценарий позволяет также искателю обращаться к менеджерам очередей и очередям WebSphere MQ.

Об этой задаче

Чтобы искатель DB2 мог просматривать таблицы баз данных, запустите сценарий установки `escrdb2.sh`, который OmniFind Enterprise Edition помещает на сервер искателей.

Прежде, чем запустить этот сценарий, убедитесь, что на сервере искателей установлен клиент администратора DB2.

Если вы используете публикацию событий, необходимо установить на сервер искателей модули WebSphere MQ 5.3 для Java Messaging, чтобы искатель DB2 имел доступ к диспетчерам очередей и очередям WebSphere MQ. Запустите сценарий установки `escrdb2.sh` после того, как установите модули WebSphere MQ.

Процедура

Чтобы сконфигурировать сервер искателей для поддержки просмотра искателями DB2:

1. Необязательно: Если вы планируете использовать публикацию событий, установите модули WebSphere MQ 5.3 для Java Messaging на сервере искателей:
 - a. Зарегистрируйтесь как пользователь `root` и введите команду:
`export LD_ASSUME_KERNEL=2.4.19`
 - b. Вставьте компакт-диск WebSphere MQ в дисковод компакт-дисков.
 - c. Перейдите в каталог, где находятся модули MQ для Java Messaging.
 - d. Введите следующую команду, чтобы установить модули:
`rpm -i MQSeriesJava-5.3.0-1.i386.rpm`
2. На сервере искателя запустите сценарий конфигурирования для искателя DB2:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия (этот ID пользователя задается при установке OmniFind Enterprise Edition).
 - b. Запустите следующий сценарий, который установлен в каталоге `$ES_INSTALL_ROOT/bin`, и ответьте на вопросы программы:
`escrdb2.sh`
3. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall  
esadmin system startall
```

Конфигурирование сервера искателей в Windows для искателей DB2

При установке OmniFind Enterprise Edition на компьютере Microsoft Windows нужно запустить сценарий, чтобы сконфигурировать сервер искателя. Этот сценарий позволяет искателю DB2 связываться с серверами баз данных. Если вы используете публикацию событий, сценарий позволяет также искателю обращаться к менеджерам очередей и очередям WebSphere MQ.

Об этой задаче

Чтобы искатель DB2 мог просматривать таблицы баз данных, запустите сценарий установки `escrdb2.vbs`, который OmniFind Enterprise Edition помещает на сервер искателей.

Прежде, чем запустить этот сценарий, убедитесь, что на сервере искателей установлен клиент администратора DB2.

Если вы используете публикацию событий, необходимо установить на сервер искателей модули WebSphere MQ 5.3 для Java Messaging, чтобы искатель DB2 имел доступ к диспетчерам очередей и очередям WebSphere MQ. Запустите сценарий установки `escrdb2.vbs` после того, как установите модули WebSphere MQ.

Процедура

Чтобы сконфигурировать сервер искателей для поддержки просмотра искателями DB2:

1. Необязательно: Если вы планируете использовать публикацию событий, установите модули WebSphere MQ 5.3 для Java Messaging на сервере искателей:
 - a. Вставьте компакт-диск WebSphere MQ в дисковод компакт-дисков.
 - b. Запустите утилиту установки WebSphere MQ.
 - c. В окне Выберите возможности продукта выберите опцию установки **Java Messaging**.
2. На сервере искателя запустите сценарий конфигурирования для искателя DB2:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке OmniFind Enterprise Edition).
 - b. Запустите следующий сценарий (устанавливаемый в каталоге %ES_INSTALL_ROOT%\bin) и отвечайте на вопросы программы:
escrdb2.vbs
3. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall  
esadmin system startall
```

Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2

Перед тем, как сконфигурировать искатель DB2 для использования публикации событий, убедитесь, что на сервере, который будет обрабатываться искателем, сконфигурирована IBM WebSphere Information Integrator Event Publisher Edition.

Об этой задаче

При конфигурировании WebSphere Information Integrator Event Publisher Edition для использования с искателем DB2 руководствуйтесь следующими указаниями:

- В таблицах-источниках для публикации должны быть выбраны как измененные, так и неизмененные столбцы.
- Удаленные строки в таблицах-источниках должны быть выбраны для публикации.
- Очередь публикации событий не может совместно использоваться несколькими базами данных.
- У одной базы данных может быть несколько карт очередей и несколько очередей.
- У таблицы должна быть одна публикация XML, связанная с одной картой очередей публикации. (У таблицы не может быть нескольких публикаций XML, связанных с одной картой очередей публикации. У таблицы может быть несколько публикаций XML, если каждая из публикаций XML связана со своей картой очередей публикации.)

Процедура

Выполните следующие шаги, чтобы сконфигурировать сервер баз данных так, чтобы искатель DB2 мог получать обновления таблиц, опубликованные в очереди публикации событий. (В качестве помощи при этих шагах используйте документацию к WebSphere Information Integrator Publisher Edition.)

1. Установите WebSphere Information Integrator Event Publisher Edition на сервере баз данных, который будет обрабатываться искателем.
2. Запустите панель запуска центра репликации:

AIX, Linux или Solaris

db2rc

Окно командной строки Windows

Выберите **Пуск** → **Центр репликации IBM DB2**

3. Создайте управляющие таблицы Q Capture:
 - a. Выберите окно панели запуска **Публикация событий**, выберите **Создать управляющие таблицы Q Capture** и нажмите кнопку **Далее**.
 - b. В поле **Сервер Q Capture** выберите из списка доступных серверов баз данных сервер, который вы хотите использовать в качестве сервера Q Capture, и нажмите кнопку **ОК**.
 - c. Укажите ID пользователя и пароль для доступа к выбранному серверу Q Capture. Измените схему Q Capture или примите имя схемы по умолчанию, затем нажмите кнопку **Далее**.
 - d. Укажите имена менеджера очередей и очереди управления и перезапустите очередь, указанную вами при конфигурировании WebSphere MQ на этом сервере баз данных, а затем нажмите кнопку **Далее**.
 - e. Нажмите кнопку **Готово**. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - f. Выберите опцию обработки **Запустить сейчас** и нажмите кнопку **ОК**. Когда появится сообщение о завершении сценариев SQL, нажмите кнопку **Заккрыть**.
4. Создайте публикацию XML:
 - a. На панели запуска Центра репликации выберите окно **Публикация событий**, выберите **Создать публикацию XML** и нажмите кнопку **Далее**.
 - b. На странице **Пуск** нажмите кнопку **Далее**.
 - c. На странице **Сервер** и карта очередей подтвердите правильность сервера Q Capture и схемы Q Capture, выберите опцию рядом с полем **Карта очередей публикации** и нажмите кнопку **Создать**, чтобы создать карту очередей публикации.
 - d. На странице **Общие** введите имя карты очередей.
 - e. На странице **Свойства** укажите имя очереди отправления (например, имя очереди данных, указанное вами при конфигурировании WebSphere MQ на этом сервере), для типа содержимого сообщения выберите либо **Операция со строками** или **Транзакция**, выключите переключатели для отправки сообщений о работоспособности и добавления заголовков сообщений JMS и нажмите кнопку **ОК**.
 - f. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - g. Выберите опцию обработки **Запустить сейчас** и нажмите кнопку **ОК**. Когда появится сообщение о завершении сценариев SQL, нажмите кнопку **Заккрыть**.
 - h. На странице **Выберите карту очередей публикации** выберите созданную вами очередь публикации и нажмите кнопку **ОК**.
 - i. На странице **Сервер** и карта очередей подтвердите правильность имени очереди публикации и нажмите кнопку **Далее**.
 - j. На странице **Исходная таблица** нажмите кнопку **Добавить**, выберите **Получить все**, выделите таблицу, для которой вы хотите включить публикацию событий, и нажмите кнопку **ОК**, а затем **Далее**.
 - k. На странице **Столбцы** и строки выберите столбец (или все столбцы), который должен обрабатывать искатель DB2, затем выберите столбцы ключа. На странице, на которой вы выбрали строки для обработки искателем (или

задали все строки), выберите опцию публикации удалений из исходной таблицы. После завершения конфигурирования этих опций нажмите кнопку **Далее**.

- l. На странице Содержимое сообщения выберите данных столбцов опцию включения как измененных, так и не измененных столбцов и опцию Только для новых значений данных. Убедитесь, что включен переключатель для автоматического запуска публикаций XML, и нажмите кнопку **Далее**.
 - m. На странице Обзор и заполнение публикаций XML нажмите кнопку **Далее**.
 - n. На странице Сводка нажмите кнопку **Готово**. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - o. Выберите опцию обработки **Запустить сейчас** и нажмите кнопку **ОК**. Когда появится сообщение о завершении сценариев SQL, нажмите кнопку **Заккрыть**.
5. Запустите сервер Q Capture:
- a. Закройте панель запуска Центра репликации и запустите Центр репликации.
 - b. В дереве объектов щелкните по **Q-репликация** → **Определения** → **Серверы Q Capture**.
 - c. Щелкните правой кнопкой мыши по значку сконфигурированного вами сервера Q Capture и выберите **Подключить базу данных для Q-репликации**.
 - d. Когда появится сообщение с предупреждением, нажмите кнопку **ОК**.
 - e. Когда появится страница с сообщениями DB2, нажмите кнопку **Заккрыть**.
 - f. В дереве объектов щелкните правой кнопкой мыши по значку для сервера Q Capture и выберите **Запустить программу Q Capture**.
 - g. Выберите опцию обработки **Запустить сейчас**, укажите имя системы, ID пользователя и пароль пользователя DB2, путь к каталогу, в котором хранятся журналы, и имя экземпляра DB2, а затем нажмите кнопку **ОК**.
 - h. Когда появится сообщение об отправке требования, нажмите кнопку **Заккрыть**.
 - i. В дереве объектов щелкните правой кнопкой мыши по значку для сервера Q Capture и выберите **Проверить состояние**.

Будет показано состояние сервера Q Capture. При возникновении ошибок в сообщении о состоянии будет указано, что сервер предположительно находится в нерабочем состоянии. Для просмотра журналов и определения причин ошибок введите в командной строке команду:

```
asnqcap Capture_Server=имя_сервера_Capture LOGSTDOUT=y
```

Конфигурирование WebSphere MQ для искателей DB2

Перед тем, как конфигурировать искатель DB2 для использования публикации событий, убедитесь, что на сервере, прием с которого будет осуществлять сервер, сконфигурирована IBM WebSphere MQ.

Прежде чем приступить к работе...

Проверьте, что на сервере назначения баз данных установлены DB2, WebSphere Information Integrator Event Publisher Edition и WebSphere MQ.

Ограничения

Если сервер назначения баз данных установлен на компьютере с Linux, все пользователи DB2, пользователи WebSphere MQ и пользователи OmniFind Enterprise Edition должны задать следующую переменную среды:

```
export LD_ASSUME_KERNEL=2.4.19
```

Эта переменная среды позволяет экспортировать реализацию потоков LinuxThread из любой среды, в которой была выполнена установка, при запуске как управляющих команд WebSphere MQ, так и прикладных программ WebSphere MQ. Для работы WebSphere MQ необходимо, чтобы эта переменная среды была экспортирована.

Искатели DB2, которые используют публикацию событий, соединяются с очередями WebSphere MQ через клиентское соединение. Чтобы разрешить клиентские соединения, зарегистрируйтесь в WebSphere MQ как администратор и введите команду, которая установит для CCSID менеджера очередей значение 819:

```
runmqsc имя_менеджера_очередей  
ALTER QMGR CCSID(819)  
END
```

Об этой задаче

Искатель DB2 поддерживает режим соединения клиентов с сервером WebSphere MQ. Искатель принимает сообщения XML, опубликованные в очереди публикации событий. Искатель не может принимать сообщения XML, транспорт которых осуществляется посредством нескольких очередей.

После того, как вы сконфигурировали WebSphere MQ, искатель DB2 использует имя менеджера очередей, имя очереди, имя хоста сервера, номер порта сервера и имя канала сервера для получения сообщений XML из очереди публикации. Искатель анализирует эти сообщения и обновляет свое пространство при помощи информации об обновленных таблицах.

Процедура

Выполните следующие шаги, чтобы сконфигурировать сервер баз данных так, чтобы искатель DB2 мог принимать сообщения от очереди публикации событий. (В качестве помощи при этих шагах используйте документацию к WebSphere MQ.)

1. Зарегистрируйтесь в роли администратора WebSphere MQ и введите следующие команды для создания менеджера очередей и очередей.
 - a. Введите в командной строке команду:
crtmqm QM1
 - b. После появления сообщения Установка завершена введите команду:
strmqm QM1
 - c. После появления сообщения 'QM1' запущен введите команду:
runmqsc QM1
 - d. После появления сообщения Запускается MQSC для менеджера очередей QM1 введите следующую команду, чтобы создать очередь управления:
DEFINE QLOCAL('ASN.QM1.ADMINQ')
 - e. После появления сообщения Очередь WebSphere MQ создана введите следующую команду, чтобы создать очередь перезапуска:
DEFINE QLOCAL('ASN.QM1.RESTARTQ')
 - f. После того, как сообщение Очередь WebSphere MQ создана появится снова, введите следующую команду, чтобы создать очередь данных:
DEFINE QLOCAL('ASN.QM1.DATAQ')
 - g. После того, как сообщение Очередь WebSphere MQ создана появится снова, введите следующую команду для выхода:
end

2. Введите следующую команду, чтобы запустить MQ Listener на сервере баз данных (MQ Listener должен быть запущен, когда вы создаете искатель DB2, использующий публикацию событий). В следующем примере 1414 - номер порта сервера, а канал по умолчанию - SYSTEM.DEF.SVRCONN:

```
runmq1sr -m QM1 -t TCP -p 1414 &
```

3. Введите следующие команды для авторизации доступа пользователя DB2 к менеджеру очередей и очередям через интерфейс Message Queuing Interface (MQI) для публикации событий (в этом примере ID пользователя - db2inst1):

```
setmqaut -m QM1 -t qmgr -p db2inst1 +allmqi  
setmqaut -m QM1 -t queue -n ASN.QM1.DATAQ -p db2inst1 +allmqi  
setmqaut -m QM1 -t queue -n ASN.QM1.ADMINQ -p db2inst1 +allmqi  
setmqaut -m QM1 -t queue -n ASN.QM1.RESTARTQ -p db2inst1 +allmqi
```

4. Введите следующие команды для ID пользователя, который использовался для создания и запуска искателя DB2 с публикацией событий. Эти команды авторизуют данный ID пользователя для доступа к менеджеру очередей и очередям для публикации событий через интерфейс Message Queuing Interface (MQI). В этом примере используется ID пользователя esuser:

```
setmqaut -m ASN.QM1.QM2 -t qmgr -p esuser +allmqi  
setmqaut -m ASN.QM1.QM2 -t queue -n ASN.QM1.DATAQ -p esuser +allmqi
```

Просмотр баз данных DB2 на сервере источника данных Classic

Искатель DB2 может просматривать базу данных DB2 на сервере источника данных Classic с помощью WebSphere Information Integrator Classic Federation.

Об этой задаче

Чтобы можно было просмотреть базу данных DB2 на сервере источника данных Classic, база данных должна быть интегрирована с базой данных DB2 на неклассическом сервере источника данных с помощью оболочки ODBC WebSphere Information Integrator Classic Federation.

Процедура

Чтобы интегрировать базу данных:

1. Установите WebSphere Information Integrator Classic Federation на сервер источника данных Classic.
2. Установите клиентский модуль WebSphere Information Integrator Classic Federation на иной сервер источника данных (не Classic), где находится база данных, которая будет служить базой данных объединения для сервера источника данных Classic.
3. Сконфигурируйте драйвер ODBC WebSphere Information Integrator Classic Federation, чтобы установить соединение с сервером источника данных Classic.
4. Установите соединение с базой данных для интеграции и создайте оболочку ODBC, чтобы выполнить интеграцию с базой данных на сервере источника данных Classic.
5. Сконфигурируйте искатель DB2 для просмотра базы данных объединения на ином сервере источника данных (не Classic). После этого искатель сможет просматривать базу данных на сервере источника данных Classic с помощью WebSphere Information Integrator Classic Federation.

Искатели DB2 Content Manager

Чтобы включить типы элементов IBM DB2 Content Manager в собрание поисковой системы предприятия, нужно сконфигурировать искатель DB2 Content Manager.

Конфигурация сервера искателя

Для возможности просмотра сервера DB2 Content Manager искателем на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю DB2 Content Manager взаимодействовать с серверами DB2 Content Manager.

Перед тем, как сконфигурировать искатель DB2 Content Manager с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для искателей DB2 Content Manager” на стр. 61.
- “Конфигурирование сервера искателей в Windows для искателей DB2 Content Manager” на стр. 63.

Совет:

Подробные примеры конфигурирования защищенного искателя DB2 Content Manager смотрите в сценарии для крупномасштабной организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурирование - обзор

Искатель DB2 Content Manager можно использовать для просмотра любого числа серверов DB2 Content Manager. При конфигурировании искателя нужно задать опции для просмотра искателем всех серверов DB2 Content Manager в пространстве искателя. Нужно также выбрать конкретные типы элементов, которые должен просматривать искатель на каждом сервере.

Чтобы создать или изменить искатель DB2 Content Manager, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют просмотром искателем элементов всех типов на всех серверах DB2 Content Manager в пространстве искателя.
- Выбрать серверы DB2 Content Manager, которые будет просматривать искатель.
- Задать ID пользователей и пароли для доступа искателя к содержимому на серверах DB2 Content Manager.
- Составить расписания просмотра искателем серверов.
- Выбрать типы элементов для просмотра искателем на каждом сервере DB2 Content Manager.

- Задать опции, обеспечивающие доступность поиска для атрибутов в определенных типах элементов. Например, можно исключить из пространства для искателя конкретные типы документов или задать, какие атрибуты будут возвращаться в результатах поиска.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование сервера искателей в UNIX для искателей DB2 Content Manager

При установке OmniFind Enterprise Edition на компьютере с IBM AIX, Linux или операционной средой Solaris необходимо запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю DB2 Content Manager связываться с серверами IBM DB2 Content Manager.

Об этой задаче

Искатель DB2 Content Manager использует соединитель Java для DB2 Content Manager Версии 8, чтобы обращаться к серверам DB2 Content Manager. Этот соединитель устанавливается при установке на сервере искателей одного из следующих продуктов:

- IBM DB2 Information Integrator for Content, Версия 8.3 for AIX, Solaris или Linux
- IBM DB2 Information Integrator for Content, Версия 8.2 для AIX или Solaris
- IBM DB2 Content Manager Toolkit, Версия 8.2 для Linux

Чтобы искатель DB2 Content Manager мог работать с DB2 Content Manager, после установки соединителя запустите сценарий настройки, установленный OmniFind Enterprise Edition на сервере искателя.

Процедура

Чтобы сконфигурировать сервер искателя для просмотра серверов DB2 Content Manager:

1. Установите на сервере искателя соединитель Java для DB2 Content Manager Версии 8:
 - a. Зарегистрируйтесь на сервере искателя в качестве пользователя root:


```
su - root
```
 - b. Запустите файл db2profile. Например:


```
./home/db2inst/sql1lib/db2profile
```
 - c. Экспортируйте переменную среды JAVAHOME. Например:


```
export JAVAHOME=/usr/IBMJava2-141
```
 - d. Добавьте каталог Java в переменную среды PATH:


```
export PATH=$PATH:$JAVAHOME/bin
```

- e. Вставьте установочный компакт-диск DB2 Information Integrator for Content и запустите мастер по установке.
 - f. В окне Выбор компонентов выполните следующие действия. (Если вы работаете с DB2 Information Integrator for Content Версии 8.3, вы увидите окно Выбор компонентов с опцией установки Пользовательская.)
 - 1) Выберите **Локальные соединители** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - 2) Выберите **Комплекты и примеры соединителей** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - g. Укажите имя базы данных, имя пользователя и пароль для библиотеки DB2 Content Manager и примите значения по умолчанию для остальных окон.
2. Зарегистрируйтесь на сервере искателя под ID пользователя, входящим в группу администраторов DB2.
 3. Добавьте в каталог базу данных удаленного библиотечного сервера DB2 Content Manager и убедитесь, что сервер искателя может соединиться с сервером DB2 Content Manager:

```
db2 catalog tcpip node имя_узла remote имя_хоста server порт
db2 catalog database имя_базы_данных as алиас at node имя_узла
```

Где:

node_name

Это краткое имя хоста для сервера DB2 Content Manager (например, ibmes).

hostname

Это полное имя хоста для сервера DB2 Content Manager (например, ibmes.ibm.com).

port

Это номер порта сервера DB2 Content Manager.

имя_базы_данных

Это имя базы данных DB2 Content Manager (например, ICMNLSDB).

алиас

Это алиас базы данных DB2 Content Manager (например, CMSVR).

4. Необязательно: Зарегистрируйтесь как пользователь root и проверьте соединение с базой данных:

```
. каталог_установки_Information_Integrator_for_Content/bin/cmbenv81.sh
cd каталог_установки_Information_Integrator_for_Content/samples/java/icm
javac *.java
java SConnectDisconnectICM имя_базы_данных_ICM ID_администратора_CM
пароль_администратора_CM
```

5. На сервере искателя запустите сценарий конфигурирования для искателя DB2 Content Manager:
 - a. Перейдите в каталог КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/bin:


```
cd $ES_INSTALL_ROOT/bin
```
 - b. Запустите следующий сценарий и отвечайте на вопросы:


```
escrsm.sh
```
6. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall
esadmin system startall
```

Конфигурирование сервера искателей в Windows для искателей DB2 Content Manager

При установке OmniFind Enterprise Edition на компьютере Microsoft Windows нужно запустить сценарий, чтобы сконфигурировать сервер искателя. Этот сценарий позволяет искателю DB2 Content Manager связываться с серверами IBM DB2 Content Manager.

Об этой задаче

Искатель DB2 Content Manager использует соединитель Java для DB2 Content Manager Версии 8, чтобы обращаться к серверам DB2 Content Manager. Чтобы установить этот соединитель, на сервере искателя надо установить IBM DB2 Information Integrator for Content Версии 8.2 или Версии 8.3. Чтобы искатель DB2 Content Manager мог работать с DB2 Content Manager, после установки соединителя запустите сценарий настройки, установленный OmniFind Enterprise Edition на сервере искателя.

Процедура

Чтобы сконфигурировать сервер искателя для просмотра серверов DB2 Content Manager:

1. Установите на сервере искателя соединитель Java для DB2 Content Manager Версии 8:
 - a. Вставьте установочный компакт-диск DB2 Information Integrator for Content. Будет автоматически запущена программа установки.
Откроется мастер по установке DB2 Content Manager Enterprise Information Portal.
 - b. В окне Выбор компонентов выполните следующие действия. (Если вы работаете с DB2 Information Integrator for Content Версии 8.3, вы увидите окно Выбор компонентов с опцией установки Пользовательская.)
 - 1) Выберите **Локальные соединители** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - 2) Выберите **Комплекты и примеры соединителей** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - c. Укажите имя базы данных, имя пользователя и пароль для библиотеки DB2 Content Manager и примите значения по умолчанию для остальных окон.
2. Добавьте в каталог базу данных удаленного библиотечного сервера DB2 Content Manager и убедитесь, что сервер искателя может соединиться с сервером DB2 Content Manager. На сервере искателя введите в командной строке следующие команды:

```
db2 catalog tcpip node имя_узла remote имя_хоста server порт
db2 catalog database имя_базы_данных as алиас at node имя_узла
```

Где:

имя_узла

Это краткое имя хоста для сервера DB2 Content Manager (например, ibmes).

имя_хоста

Это полное имя хоста для сервера DB2 Content Manager (например, ibmes.ibm.com).

порт

Это номер порта сервера DB2 Content Manager.

имя_базы_данных

Это имя базы данных DB2 Content Manager (например, ICMNLSDB).

алиас

Это алиас базы данных DB2 Content Manager (например, CMSVR).

3. Необязательно: Проверьте соединение базы данных, введя в командной строке команды:

```
smbenv81.bat
cd каталог_установки_Information_Integrator_for_Content/samples/java/icm
javac *.java
java SConnectDisconnectICM имя_базы_данных_ICM ID_администратора_CM
пароль_администратора_CM
```

4. На сервере искателя запустите сценарий конфигурирования для искателя DB2 Content Manager:

- a. Перейдите в каталог КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES\bin:

```
cd %ES_INSTALL_ROOT%\bin
```

- b. Запустите следующий сценарий и отвечайте на вопросы:

```
esrcm.vbs
```

5. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall
esadmin system startall
```

Искатели Domino Document Manager

Чтобы включить библиотеки и шкафы Domino Document Manager в собрание поисковой системы предприятия, нужно сконфигурировать искатель Domino Document Manager.

Конфигурация сервера искателя

Если на сервере Domino Document Manager, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий конфигурирования. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю Domino Document Manager взаимодействовать с серверами, использующими NRPC.

Если на сервере Domino Document Manager, который планируется просматривать искателем, используется протокол ДИОР (Domino Internet Inter-ORB Protocol), запускать сценарий установки на сервере искателя не требуется. Однако нужно сконфигурировать сервер Domino Document Manager, чтобы к нему мог обращаться искатель Domino Document Manager.

Важное замечание: Если сервер Domino Document Manager использует протокол ДИОР, а для шифрования данных, передаваемых между искателем и сервером данных, для искателя конфигурируется использование HTTPS или ДИОР через SSL, нужно скопировать файл TrustedCerts.class (например, c:\certs или /data/certs) с сервера Domino Document Manager на сервер искателей. В конфигурации с двумя или четырьмя серверами файл TrustedCerts.class нужно также скопировать на серверы,

где установлен компонент поиска. Кроме того, нужно убедиться, что положение этого файла на сервере искателей совпадает с положением на серверах поиска. Путь каталога для файла `TrustedCerts.class` задается при конфигурировании искателя.

Если OmniFind Enterprise Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель Domino Document Manager с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 83.
- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84.

Защита на уровне документов

Если включена защита собрания и на сервере, который планируется просматривать искателем, используется протокол NRPC (Notes Remote Procedure Call - удаленный вызов процедур Notes), на сервере искателя нужно сконфигурировать Lotus Domino Trusted Server. Trusted Server используется для обеспечения соблюдения управления доступом на уровне документов. Перед тем, как сделать собрания доступными пользователям для поиска, выполните следующие задачи:

- Сконфигурируйте доверенные серверы Lotus Domino для проверки параметров аутентификации пользователей
- Включите глобальную защиту в WebSphere Application Server и конфигурирование прикладных программ поиска для использования защиты. Этот шаг гарантирует проверку параметров аутентификации при попытках пользователей использовать программу поиска. Серверы поиска используют параметры аутентификации для проверки полномочий каждого из пользователей на доступ к документам Lotus Domino.

Конфигурирование - обзор

Искатель Domino Document Manager можно использовать для просмотра любого числа библиотек Domino Document Manager. При создании искателя вы выбираете библиотеку для просмотра искателем с одного сервера Domino Document Manager. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера Domino Document Manager, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все шкафы в библиотеках, выбранных для просмотра или только отдельные шкафы.

Чтобы создать или изменить искатель Domino Document Manager, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Идентифицировать сервер и протокол связи Domino Document Manager.
- Если сервер сконфигурирован для использования протокола DIIOP, можно задать, как искатель будет соединяться с объектами Domino. Например, можно задать опции для использования HTTPS или SSL (Secure Sockets Layer) для шифрования соединений.
- Выбрать библиотеки, которые будет просматривать искатель.
- Составить расписания просмотра искателем библиотек.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все шкафы в библиотеке или только документы, находящиеся в выбранных вами шкафах.
- Задать опции, обеспечивающие доступность поиска для полей в различных библиотеках и шкафах. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Советы по просмотру баз данных Lotus Domino” на стр. 78

“Обеспечение защиты уровня документов для документов Lotus Domino” на стр. 286

Задачи, связанные с данной

“Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79

“Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81

“Конфигурирование серверов, использующих протокол DIIOP” на стр. 83

“Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84

“Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей” на стр. 286

Искатели Exchange Server

Чтобы включить общедоступные папки Microsoft Exchange Server в собрание поисковой системы предприятия, нужно сконфигурировать искатель Exchange Server.

Искатель Exchange Server можно использовать для просмотра любого числа папок и подпапок на серверах общедоступных папок Exchange Server. При создании искателя

нужно выбрать содержимое, просматриваемое искателем на сервере общедоступных папок. Позже можно отредактировать пространство для искателя, добавив в него содержимое с другого сервера общедоступных папок.

Чтобы создать или изменить искатель Exchange Server, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подпапок на всех серверах в пространстве для искателя.
- Задать информацию о сервере общедоступных папок Exchange Server, просматриваемого искателем.

Для возможности обращения искателя к содержимому на сервере нужно задать ID пользователя и пароль. Если сервер использует протокол SSL, можно задать опции для доступа искателя к файлу склада ключей на сервере искателя.

- Составить расписание просмотра искателем сервера общедоступных папок.
- Выбрать папки и подпапки для просмотра.
- Задать опции, обеспечивающие доступность поиска для документов в подпапках. Например, можно исключить из пространства для искателя определенные типы документов.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Задачи, связанные с данной

“Проверка доступа к защищенным документам Exchange Server” на стр. 285

Искатели баз данных JDBC

Искатель База данных JDBC позволяет включить в собрание поисковой системы предприятия базы данных, к которым можно обращаться при помощи протокола Java Database Connectivity (JDBC).

Для каждого типа просматриваемых систем баз данных нужно сконфигурировать отдельный искатель. При создании искателя надо задать опции для просмотра одной базы данных. Впоследствии можно добавить в пространство для искателя базы данных того же типа.

Каждая строка в таблице базы данных считается документом, а значения столбцов базы данных анализируются и индексируются как поля, доступные для поиска. Искатель можно сконфигурировать для просмотра нескольких структурированных таблиц, связав с ним подключаемый модуль при конфигурировании свойств искателя. Этот подключаемый модуль позволяет объединить строки из нескольких таблиц в реляционной базе данных с одинаковыми ключевыми полями и

обрабатывать их как один документ. Когда пользователь выполняет в базе данных поиск, при выводе документа в результатах поиска данные из объединенных таблиц выводятся как дополнительные поля.

Поддерживаемые системы и драйверы баз данных

Чтобы искатель использовал протокол JDBC для обработки таблиц базы данных, соответствующий драйвер JDBC должен существовать на сервере искателя. Искатель База данных JDBC поддерживает следующие системы баз данных и драйверы JDBC типа 4:

Система баз данных	Имя драйвера JDBC типа 4	Пути классов стандартного драйвера JDBC
IBM DB2 Universal Database Version 8.2 and IBM DB2 Enterprise Server Edition Версии 9.1 для Linux, UNIX и Windows	com.ibm.db2.jcc.DB2Driver	<i>корневой_каталог_установки_db2/java/db2jcc.jar</i> <i>корневой_каталог_установки_db2/java/db2jcc_license_cu.jar</i>
Oracle 9i and 10g	Oracle.jdbc.driver.OracleDriver	<i>oracle_home/jdbc/lib/ojdbc14.jar</i>
Microsoft SQL Server 2000	com.microsoft.jdbc.sqlserver.SQLServerDriver	<i>домашний_каталог_jdbc_mssql/lib/mssqlserver.jar</i> <i>домашний_каталог_jdbc_mssql/lib/msbase.jar</i> <i>домашний_каталог_jdbc_mssql/lib/msutil.jar</i>
Microsoft SQL Server 2005	com.microsoft.sqlserver.jdbc.SQLServerDriver (драйвер JDBC для SQL Server 2005 не поддерживается в системах AIX.)	<i>к-лог_уст-ки/sqljdbc_1.0/локаль/ sqljdbc.jar</i> где <i>локаль</i> - ваша локаль, например, <i>каталог_установки/sqljdbc_1.0/enu/sqljdbc.jar</i>

Сравнение искателей База данных JDBC и искателей DB2

Если в текущий момент используется искатель DB2, его можно продолжать использовать. Данные, хранимые для искателя DB2, нельзя перенастроить для искателя База данных JDBC.

Вместо искателя База данных JDBC используйте искатель DB2:

- Для просмотра баз данных DB2 с драйвером JDBC типа 2.
- Для просмотра баз данных Oracle и SQL Server, объединенных с базой данных DB2. При помощи искателя DB2 можно обращаться к базам данных всех типов по псевдонимам.
- Для просмотра баз данных DB2 for z/OS, DB2 for iSeries, Informix, Sybase, VSAM, IMS, CA-Datcom и Software AG Adabas. Базы данных этих типов следует объединить с базами данных DB2 и обращаться к ним при помощи искателя DB2 по псевдонимам.
- Если при публикации обновлений базы данных для обновления индекса поисковой системы предприятия предполагается использовать публикацию событий.

Конфигурирование - обзор

Чтобы создать или изменить искатель База данных JDBC, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать

искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех баз данных в пространстве для искателя.
- Задать тип базы данных для обработки искателем.
- Выбрать базу данных для просмотра и при необходимости задать ID пользователя и пароль, позволяющие искателю обращаться к базе данных.
- Составить расписания просмотра искателем баз данных.
- Выбрать таблицы, которые будет просматривать искатель.

Внимание: Чтобы оптимизировать производительность процессов обнаружения и предотвратить истечение срока ожидания для процесса конфигурирования искателя, выбирайте просмотр искателем всех таблиц, только если в базе данных немного таблиц или в таблицах немного столбцов. Если сейчас выбрать определенные таблицы, впоследствии можно будет изменить пространство для искателя и добавить в собрание дополнительные таблицы.

- Задать опции, обеспечивающие доступность поиска для столбцов в определенных таблицах. Например, можно включить использование конкретных столбцов в параметрических запросах или задать, какие столбцы будут возвращаться в результатах поиска.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Отображения взаимосвязей для баз данных JDBC

При создании правил для подключаемого модуля, просматривающего несколько структурированных таблиц баз данных JDBC задается информация о корневой таблице и способах объединения родительских и дочерних таблиц.

Поставляемый для поисковой системы предприятия подключаемый модуль позволяет искателю База данных JDBC объединить несколько структурированных таблиц. Для создания подключаемого модуля задаются правила в файле `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/crawler_rdb_plugin.xml`. После конфигурирования искателя для использования подключаемого модуля строки из таблиц с одинаковыми ключевыми полями объединяются и обрабатываются как один документ. Когда пользователь выполняет в базе данных поиск, при выводе документа в результатах поиска данные из объединенных таблиц выводятся как дополнительные поля.

Объединение таблиц при помощи ключевых столбцов

На следующей диаграмме показано, как построено отображение взаимосвязей для нескольких таблиц. Искатель База данных JDBC просматривает в базе данных корневую таблицу. Некоторые столбцы в таблице представляют собой ключевые поля, которые можно использовать для объединения таблицы с другими таблицами. После этого столбцы в объединенных таблицах можно использовать в качестве

ключей для объединения дополнительных таблиц. Строки в нескольких таблицах рассматриваются в пространстве искателя как один документ. Значения столбцов считаются метаданными документа. Корневая таблица в этом отношении является *родительской*, а таблицы, объединенные на первом уровне - *дочерними*. Дочерние таблицы на первом уровне могут также быть родительскими для таблиц, объединенных на втором уровне.

В этом примере столбцы Key 1 и Key 2 в корневой (родительской) таблице - это ключевые поля, позволяющие объединить таблицу с дочерними таблицами, у которых тоже есть столбцы Key 1 и Key 2. У одной таблицы, объединенной на этом первом уровне, есть ключевые поля Key 3 и Key 4, позволяющие объединить таблицу с дополнительными таблицами.

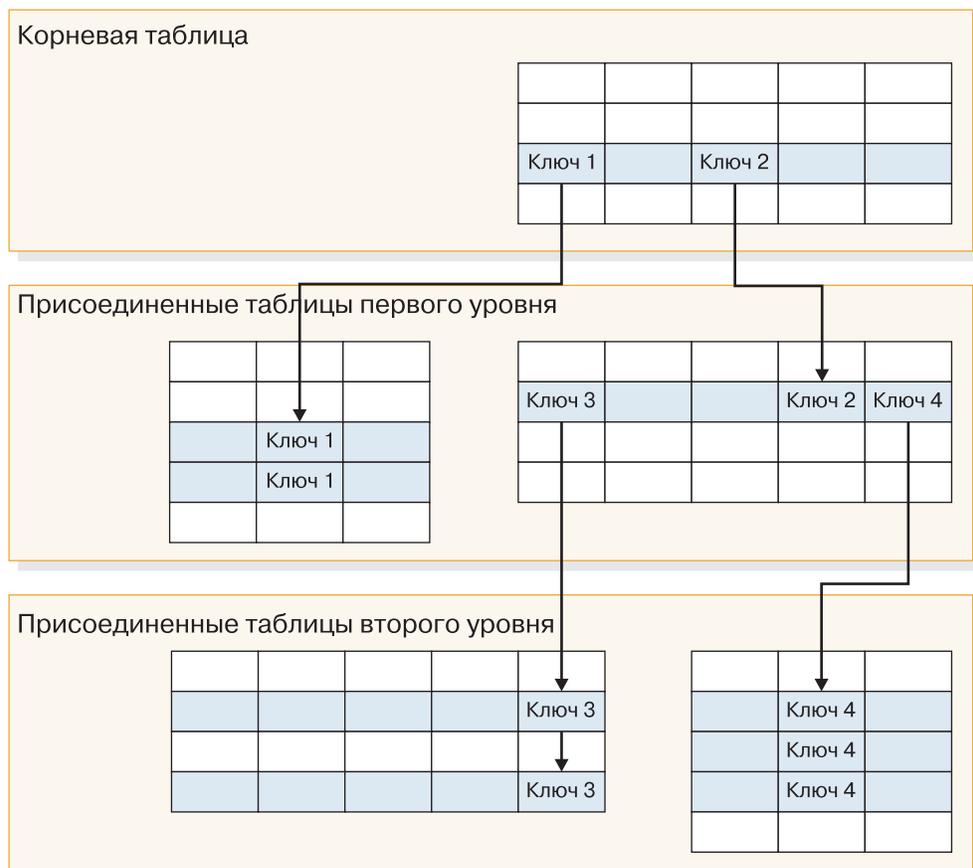


Рисунок 2. Таблицы баз данных JDBC, объединенные при помощи ключевых полей

Просмотр результатов поиска

На приведенной ниже диаграмме показан способ представления в результатах поиска данных из нескольких структурированных таблиц в виде одного документа. Без подключаемого модуля пользователь, выполнивший поиск в таблице EMPLOYEE, мог просмотреть строку из корневой таблицы, выведенной в результатах поиска, и увидеть только значения для столбцов таблицы EMPLOYEE (ID, Name и Office).

С подключаемым модулем искатель, используя столбец Office в качестве ключа, может объединить таблицу EMPLOYEE с таблицей OFFICE. Столбец Country в таблице OFFICE служит ключом для объединения этой таблицы с таблицей COUNTRY. После объединения этих таблиц пользователи, выполняющие поиск в таблице EMPLOYEE, смогут просмотреть в результатах поиска значения из столбцов

таблиц OFFICE и COUNTRY как дополнительные поля.

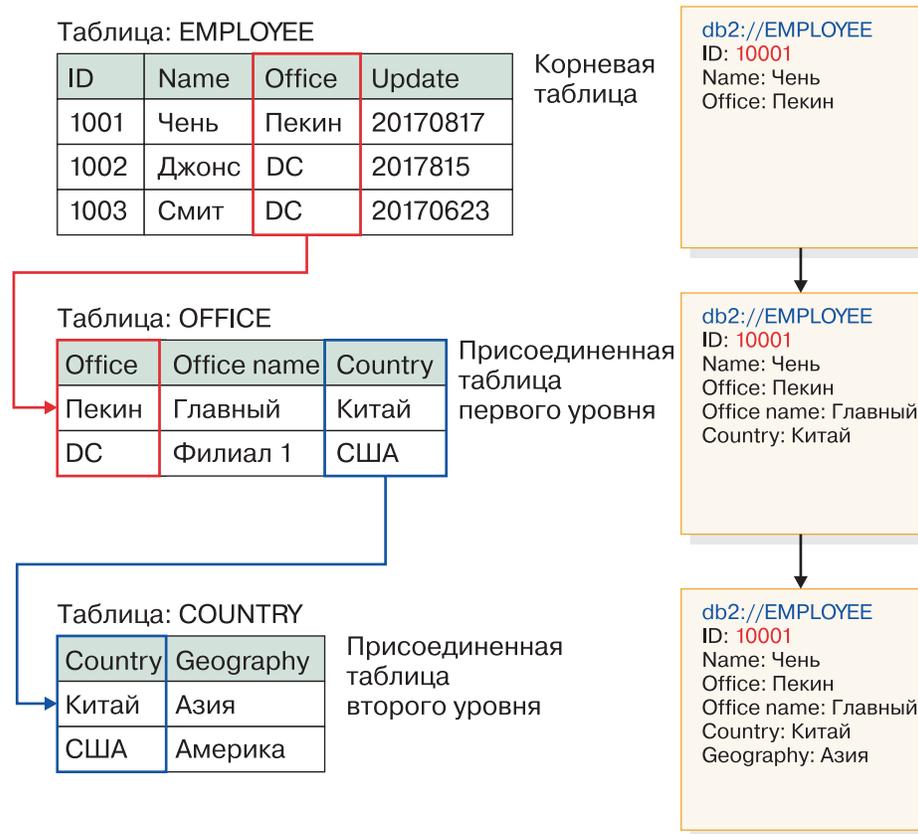


Рисунок 3. В результатах поиска выводятся значения из объединенных таблиц JDBC

Просмотр искателем нескольких структурированных таблиц баз данных JDBC

Искатель База данных JDBC можно сконфигурировать для объединения нескольких структурированных таблиц с одинаковыми ключевыми полями.

Прежде чем приступить к работе...

Для выполнения этой задачи у вас должны быть полномочия администратора поисковой системы предприятия.

Об этой задаче

При конфигурировании свойств искателя База данных JDBC можно задать подключаемый модуль для просмотра искателем нескольких структурированных таблиц, связанных друг с другом посредством ключевых полей. Без этого подключаемого модуля строки в таблице базы данных рассматриваются как отдельные документы, а значения столбцов базы данных доступны для поиска как отдельные поля. С подключаемым модулем строки из нескольких таблиц в реляционной базе данных с одинаковыми ключевыми полями объединяются и рассматриваются как один документ. Искатель добавляет данные, которые он получает из подсоединенных таблиц, к метаданным исходной строки таблицы базы данных. Когда пользователь выполняет поиск по такой базе данных, дополнительные данные появляются как дополнительные поля при выводе документов в результатах поиска.

Ограничения

Типы данных, недоступные для просмотра

В объединенных вами таблицах искатель не может просматривать поля со следующими двоичными типами данных:

BLOB
CHARACTER FOR BIT DATA
VARCHAR FOR BIT DATA
LONG VARCHAR FOR BIT DATA

Ограничения на область действия пространства искателя

Объединяемые таблицы должны находиться в одной и той же реляционной базе данных. Таблицы разных баз данных объединять нельзя.

Если для таблицы в базе данных сконфигурировано объединение с другими таблицами, эта конфигурация универсальна для всех искателей в собрании, для которых включено использование подключаемого модуля. Однако можно создать несколько собраний и сконфигурировать для отдельных искателей просмотр разных корневых таблиц и объединение различных таблиц.

Ограничения на использование других подключаемых модулей

Если для искателя конфигурируется использование подключаемого модуля для просмотра нескольких структурированных таблиц, с этим искателем нельзя связать другой подключаемый модуль. Например, нельзя задать пользовательский подключаемый модуль для применения бизнес-правил и правил защиты. С искателем нельзя связать несколько подключаемых модулей.

Ограничения на число таблиц, строк, полей и ключей

Максимальное число объединенных таблиц для одной базы данных равно пяти, а суммарное число строк в этих таблицах должно быть меньше одного миллиона. Максимальное число полей, которые можно прочитать из таблицы, равно 10. Для объединения таблиц используется пара ключей. Это означает, что большее число ключей для объединения таблиц использовать нельзя.

Гарантированный просмотр изменений в объединенных таблицах

Если строки в корневой таблице между просмотрами не изменяются и искатель не сконфигурирован для выполнения полного просмотра, неизменные строки искателем игнорируются. Если строки в таблице, объединенной с корневой таблицей, изменяются, а сама корневая таблица остается без изменений, нужно выполнить одно из следующих действий, гарантирующих, что изменения будут обнаружены и просмотрены искателем:

- В базе данных назначения корневая таблица должна содержать поле отметки времени. Сконфигурируйте базу данных назначения, чтобы она содержала поле отметки времени, изменяющееся при изменении строки в корневой таблице или строк в любой из объединенных с ней дочерних таблиц. При конфигурировании искателя База данных JDBC обязательно задайте это поле отметки времени в качестве поля, которое искатель должен использовать для определения момента времени изменений в таблицах.
- При конфигурировании расписания искателя укажите, что искатель должен выполнять полный просмотр. Эта опция гарантирует, что все таблицы будут просматриваться искателем каждый раз, независимо от внесения в них изменений.

Процедура

Чтобы сконфигурировать для искателя База данных JDBC просмотр нескольких структурированных таблиц:

1. Зарегистрируйтесь как администратор поисковой системы предприятия на сервере искателей и скопируйте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/crawler_rdb_plugin.xml для создания файла КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/crawler_rdb_plugin.xml.
2. Отредактируйте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/crawler_rdb_plugin.xml в текстовом редакторе, поддерживающем кодировку UTF-8.
 - a. Отредактируйте элемент `<Server DBURL="jdbc:db2://db_server_url:50000/SAMPLE">` и замените `jdbc:db2://db_server_url:50000/SAMPLE` на URL базы данных JDBC, которую будет просматривать искатель. Обязательно задайте этот же URL для просматриваемой базы данных при конфигурировании искателя.
 - b. Если искатель будет просматривать не базу данных DB2, отредактируйте элемент `<JDBCdriver>com.ibm.db2.jcc.DB2Driver</JDBCdriver>` и замените `com.ibm.db2.jcc.DB2Driver` на нужный драйвер JDBC. Обязательно задайте этот же драйвер для просматриваемой базы данных при конфигурировании искателя.
 - c. Отредактируйте элемент `<User>username</User>` и замените `username` на ID пользователя, у которого есть право доступа к просматриваемой базе данных.
 - d. Отредактируйте элемент `<Password Encryption="True">encrypted_password</Password>` и замените `encrypted_password` на зашифрованный пароль для заданного ID пользователя. Зашифрованный пароль можно скопировать из файла КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/col_имя_собрания.JDBC_имя_искателя/jdbccrawler.xml и вставить его сюда. Если шифрование пароля не требуется, замените `Encryption="True"` на `Encryption="False"` и замените `encrypted_password` на простой текстовый пароль.
 - e. Если оставить элемент `<Delimiters Use="True">` как есть, несколько терминов в столбце будут разделяться символом (,), определенным в элементе `<Delimiter>`. Наборы терминов для одной таблицы будут разделяться символом (;), определенным в элементе `<SecondDelimiter>`. Если вместо `<Delimiters Use="True">` задать `<Delimiters Use="False">`, символы разделителей использоваться не будут, а в качестве метаданных документов будут добавлены несколько полей метаданных с одним именем поля.
 - f. При использовании элемента `<Delimiters Use="True">` отредактируйте элементы `<Delimiter>`, `</Delimiter>` и `<SecondDelimiter>`, `</SecondDelimiter>`, чтобы задать символы, используемые в качестве разделителей значений.
 - g. Отредактируйте элемент `<RelationMap Root="DB2INST1.TABLE_0">` и замените `DB2INST1.TABLE_0` на имя корневой таблицы, которая будет просматриваться искателем.
 - h. Отредактируйте элемент `<Relation Parent="DB2INST1.TABLE_0" ParentAlias="T0" ParentKey="ID" Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID"/>`.
 - Замените `Parent="DB2INST1.TABLE_0"` на имя таблицы, которая в данном отношении является родительской.
 - Замените `ParentKey="T0"` на алиас родительской таблицы. Этот алиас должен быть уникален и не должен повторяться в файле `crawler_rdb_plugin.xml`.

- Замените ParentKey="ID" на имя столбца, используемого в данном отношении в качестве ключевого поля.
- Замените Child="DB2INST1.TABLE_1" ChildAlias="T1" ChildKey="ID" на информацию о просматриваемой дочерней таблице.

Эта структура определяет, как должны быть объединены таблицы. Например, следующее отображение взаимосвязей задает, что следует выполнять просмотр искателем корневой таблицы под именем DB2INST1.TABLE_A. Таблицы DB2INST1.TABLE_B и DB2INST1.TABLE_C объединены при помощи условия DB2INST1.TABLE_A.ID=DB2INST1.TABLE_B.ID AND DB2INST1.TABLE_B.ID=DB2INST1.TABLE_C.ID.

```
<RelationMap Root="DB2INST1.TABLE_A">
<Relation Parent="DB2INST1.TABLE_A" ParentAlias="TA" ParentKey="ID"
  Child="DB2INST1.TABLE_B" ChildAlias="TB" ChildKey="ID"/>
<Relation Parent="DB2INST1.TABLE_B" ParentAlias="TB" ParentKey="ID"
  Child="DB2INST1.TABLE_C" ChildAlias="TC" ChildKey="ID"/>
```

- Повторите шаг 2h на стр. 73, чтобы создать элементы <Relation> для всех отношений, объединяющих таблицы из корневой таблицы.
 - Отредактируйте элемент <Target TableAlias="T1"> и замените TableAlias="T1" на значение ChildAlias, определенное вами на шаге 2h на стр. 73.
 - Отредактируйте элемент <Field Name="ID" FieldName="ID_1" Enabling="True" Searchable="True" FieldSearchable="True" IsContent="True"/>.
 - Замените Name="ID" на имя столбца в просматриваемых документах.
 - Замените FieldName="ID_1" на имя поля метаданных в просматриваемых документах. Это значение используется как имя для вывода столбца в консоли администратора поисковой системы предприятия и результатах поиска.
 - Замените Enabling="True" на "False", если этот столбец не следует включать в метаданные документов.
 - Замените Searchable="True" на "False", чтобы запретить пользователям выполнять поиск в этом столбце при помощи свободного текстового запроса.
 - Замените FieldSearchable="True" на "False", чтобы запретить пользователям выполнять в этом столбце поиск по имени столбца.
 - Замените IsContent="True" на "False", чтобы указать, что в столбце отсутствует доступное для поиска содержимое. Если задать Searchable="True" и IsContent="True", значение столбца будет использоваться для определения повторяющихся документов и войдет в состав динамической сводки документов в результатах поиска.
 - Повторите шаг 2k, чтобы создать элементы <Field> для всех просматриваемых искателем столбцов.
 - Повторите шаги 2j и 2k, чтобы создать элементы <Target> и <Field> для всех дочерних таблиц, на которые есть ссылки в отображении взаимосвязей (<RelationMap>).
 - Повторите шаги 2g на стр. 73 - 2m, чтобы создать несколько отображений взаимосвязей для нескольких корневых таблиц.
 - Повторите шаги 2a на стр. 73 - 2n, чтобы сконфигурировать отображение взаимосвязей для другой базы данных.
- Сконфигурируйте искатель для использования подключаемого модуля:
 - Откройте консоль администратора поисковой системы предприятия, отредактируйте собрание и выберите страницу Искатель.

- b. Создайте искатель База данных JDBC или найдите искатель, который нужно изменить, и щелкните по значку  **Свойства искателя**.
 - c. В поле **Имя класса подключаемых модулей** введите имя подключаемого модуля для просмотра нескольких структурированных таблиц.
com.ibm.es.plugin.rdb.RDBPlugin
 - d. В поле **Путь класса подключаемых модулей** введите полные пути для подключаемого модуля и используемых им драйверов JDBC. Например, путем драйвера JDBC базы данных DB2 в системе Windows может быть:
C:\Program Files\IBM\es\lib\plugin_rdb.jar;C:\Program Files\IBM\SQLLIB\java\db2jcc.jar;C:\Program Files\IBM\SQLLIB\java\db2jcc_license_cu.jar;
 - e. Нажмите кнопку **Далее**, чтобы продолжить создание искателя, или кнопку **ОК**, чтобы сохранить внесенные изменения.
4. Чтобы внедрить файл crawler_rdb_plugin.xml в конфигурацию системы, перезапустите поисковую систему предприятия:

```
esadmin system stop  
esadmin system start
```

Искатели NNTP

Чтобы включить группы новостей NNTP в собрание поисковой системы предприятия, нужно сконфигурировать искатель NNTP.

Искатель NNTP можно использовать для просмотра любого числа серверов NNTP. При конфигурировании искателя нужно выбрать группы новостей для просмотра с одного сервера NNTP. Позднее, при редактировании пространства для искателя, можно будет добавить другие серверы NNTP, которые должен просматривать искатель.

При идентификации групп новостей для просмотра можно выбрать группы, которые будут включены в пространство для искателя или исключены из него. Тем самым можно разрешить искателю просматривать большую часть групп новостей на сервере и запретить ему просматривать небольшое число групп новостей, которые вы не хотите включать в поиск.

Например, можно задать правила, включающие все группы новостей на некотором сервере NNTP, затем исключить из них группы новостей, имена которых содержат строку private.

Чтобы создать или изменить искатель NNTP, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех групп новостей в пространстве искателя.
- Идентифицировать сервер NNTP для просмотра искателем. Если сервер защищен паролем, для искателя надо указать ID пользователя и пароль, которые будут использоваться при обращении к группам новостей на сервере.
- Составить расписание просмотра искателем сервера.

- Задать шаблоны для включения групп новостей и шаблоны для исключения определенных групп новостей из пространства для искателя.
- Задать, должен ли искатель автоматически определять язык и кодовую страницу просматриваемых статей или использовать конкретный язык и кодовую страницу.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Искатели Notes

Чтобы включить базы данных IBM Lotus Notes в собрание поисковой системы предприятия, нужно сконфигурировать искатель Примечания.

Совет:

Подробные примеры конфигурирования защищенного искателя Примечания смотрите в сценарии для крупномасштабной организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурация сервера искателя

Если на сервере Lotus Notes, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю Примечания взаимодействовать с серверами, использующими NRPC.

Если на сервере Lotus Notes, который планируется просматривать искателем, используется протокол DIIOP (Domino Internet Inter-ORB Protocol), запускать сценарий установки на сервере искателя не требуется. Однако нужно сконфигурировать сервер Lotus Notes, чтобы к нему мог обращаться искатель Примечания.

Важное замечание: Если сервер Lotus Notes использует протокол DIIOP, а для шифрования данных, передаваемых между искателем и сервером данных, для искателя задано использование HTTPS или DIIOP через SSL, нужно скопировать файл `TrustedCerts.class` (например, `c:\certs` или `/data/certs`) с сервера Lotus Notes Manager на сервер искателей. В конфигурации с двумя или четырьмя серверами файл `TrustedCerts.class` нужно также скопировать на серверы, где установлен компонент поиска. Кроме того, нужно убедиться, что положение этого файла на сервере искателей совпадает с положением на серверах поиска. Путь каталога для файла `TrustedCerts.class` задается при конфигурировании искателя.

Если OmniFind Enterprise Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель Примечания при помощи консоли управления поисковой системой предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 83.
- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84.

Защита на уровне документов

Если включена защита собрания и на сервере Lotus Notes, который планируется просматривать искателем, используется протокол NRPC, на сервере искателя нужно сконфигурировать Lotus Domino Trusted Server. Trusted Server используется для обеспечения соблюдения управления доступом на уровне документов. Перед тем, как сделать собрания доступными пользователям для поиска, выполните следующие задачи:

- Сконфигурируйте доверенные серверы Lotus Domino для проверки параметров аутентификации пользователей
- Включите глобальную защиту в WebSphere Application Server и конфигурирование прикладных программ поиска для использования защиты. Этот шаг гарантирует проверку параметров аутентификации при попытках пользователей использовать программу поиска. Серверы поиска используют параметры аутентификации для проверки полномочий каждого из пользователей на доступ к документам Lotus Domino.

Конфигурирование - обзор

Искатель Примечания может просматривать любое число стандартных баз данных Lotus Notes (файлы .nsf). При создании искателя вы выбираете базы данных или каталоги для просмотра искателем с одного сервера Lotus Notes. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера Lotus Notes, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все базы данных или каталоги на сервере или только конкретные базы данных, представления и папки.

Чтобы создать или изменить искатель Примечания, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Указать имя хоста, порт и протокол связи сервера Lotus Notes.
- Если сервер сконфигурирован для использования протокола DIIOP, можно задать, как искатель будет соединяться с объектами Domino. Например, можно задать опции для использования HTTPS или SSL (Secure Sockets Layer) для шифрования соединений.
- Выбрать базы данных или каталоги, которые будет просматривать искатель. При просмотре искателем каталогов можно задать шаблоны включения или исключения баз данных, которые помогут разделить задачу просмотра больших каталогов между несколькими искателями.

- Составить расписания просмотра искателем баз данных или каталогов.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все документы в каталоге, все документы в базе данных или документы в выбранных представлениях и папках базы данных.
- Задать опции, обеспечивающие доступность поиска для полей в различных базах данных, представлениях и папках. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Советы по просмотру баз данных Lotus Domino”

“Обеспечение защиты уровня документов для документов Lotus Domino” на стр. 286

Задачи, связанные с данной

“Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79

“Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81

“Конфигурирование серверов, использующих протокол DIIOP” на стр. 83

“Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84

“Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей” на стр. 286

Советы по просмотру баз данных Lotus Domino

Перед тем, как конфигурировать искатель Примечания, просмотрите указания по обработке искателем баз данных Lotus Domino.

- Базы данных Notes, основанные на стандартных шаблонах (как база данных примера) - самый удачный тип баз данных для работы искателя.
- Искатель Примечания использует следующие правила отображения полей:
 - Имена главных полей из стандартных шаблонов Domino зарегистрированы изначально.
 - Значения из полей Notes, указанные в таблице правил отображения, используются в результатах поиска в качестве сводок документов.
 - Значения из полей Notes, не указанные в таблице правил отображения, не используются в сводках документов.
 - Значения из полей Notes, отображенные на поле заголовка, используются в результатах поиска в качестве заголовка документа.

- Поля в следующей таблице отображаются на имена полей поиска по умолчанию:

Таблица 2. Правила отображения полей по умолчанию

Имя поля базы данных Notes	Имя поля поиска
Title	Title
EventTitle	Title
Subject	Title
Body	Body
Mission	Body
From	Creator
Author	Creator
Keywords	Categories
Categories	Categories
TeamRoomName	Organization
TeamName	Organization
Department	Organization

- Искатель Примечания может обрабатывать все типы полей за исключением полей, вычисляемых для вывода.
- Статичный текст и изображения, помещенные в документ Notes, искателем не обрабатываются.
- При конфигурировании искателя включите переключатель **Все** под переключателем **Обрабатывать искателем**, чтобы обрабатывать искателем все поля и максимально увеличить набор данных полей для обработки (для ограничения числа обрабатываемых искателем полей можно использовать поле **Обрабатывать искателем все поля, кроме**).

Чтобы сократить обработку искателем ненужных полей, выключите переключатель **Обрабатывать искателем** для всех полей за исключением тех, которые отображены на поля поиска.

Понятия, связанные с данным

“Искатели Domino Document Manager” на стр. 64

“Искатели Notes” на стр. 76

“Искатели QuickPlace” на стр. 86

Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino

При установке OmniFind Enterprise Edition на компьютере с операционной средой IBM AIX, Linux или Solaris, если вы планируете просматривать серверы, которые используют протокол Notes remote procedure call (NRPC), надо запустить сценарий для конфигурирования сервера искателя. Этот сценарий разрешает искателям Примечания, QuickPlace и Domino Document Manager обмениваться информацией с серверами базы данных.

Ограничения

Сервер Domino нельзя запустить одновременно и на одном компьютере с искателем Примечания, QuickPlace или Domino Document Manager, сконфигурированным для

использования протокола NRPC. При попытке запустить один из этих искателей во время работы Domino Server появится сообщение об ошибке, и искатель будет остановлен.

Об этой задаче

Искатели, работающие по протоколу NRPC, используют библиотеки Domino как клиент. Эти библиотеки устанавливаются при установке Lotus Domino Server на сервере искателя. Чтобы искатели могли работать с библиотеками Domino, запустите сценарий настройки, который OmniFind Enterprise Edition помещает на сервер искателя после установки библиотек Domino.

Процедура

Чтобы сконфигурировать сервер искателя для просмотра серверов Lotus Notes, Lotus QuickPlace и Domino Document Manager:

1. Создайте на сервере искателя пользователя `server` и группу `notes`:
 - a. Зарегистрируйтесь в качестве пользователя `root`:
`su - root`
 - b. Добавьте пользователя:
`useradd server`
 - c. Добавьте пароль для этого пользователя:
`passwd server`
Вас попросят изменить этот пароль.
2. Установите Lotus Domino Server на сервере искателя:
 - a. Вставьте компакт-диск Domino Server и смонтируйте его. (Если у вас нет этого компакт-диска, можно загрузить его образ.)
 - b. Перейдите в папку для вашей операционной системы.

AIX: `cd /mnt/cdrom/aix`
Linux: `cd /mnt/cdrom/linux`
Solaris: `cd /mnt/cdrom/solaris`
 - c. Запустите программу установки:
`./install`
 - d. Ответьте на запросы и примите значения по умолчанию или задайте другие параметры установки (например, пути для каталога установки и каталога данных).

При необходимости смотрите информацию об установке Domino Server в документации по Domino.
 - e. Убедитесь, что у ID администратора поисковой системы предприятия есть разрешение на доступ к каталогу `home/server`. Этот ID администратора задается при установке OmniFind Enterprise Edition.
3. На сервере искателя запустите сценарий конфигурирования, переданный OmniFind Enterprise Edition:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия (этот ID пользователя задается при установке OmniFind Enterprise Edition).
 - b. Запустите следующий сценарий, который установлен в каталоге `$ES_INSTALL_ROOT/bin`:
`escrnote.sh`
 - c. Отвечайте на вопросы программы:

- Ответьте Y (Да) на следующий вопрос, если Domino Server установлен в каталоге по умолчанию, и N (Нет), если это не так:
Найден путь каталога Lotus Notes /opt/lotus/notes/latest/linux.
Это правильный путь каталога Lotus Notes?

Для AIX путь по умолчанию - /opt/lotus/notes/latest/ibmpow.
Для Linux путь по умолчанию - /opt/lotus/notes/latest/linux.
Для Solaris путь по умолчанию - /opt/lotus/notes/latest/sunspa.

- Если Domino Server установлен на сервере искателя не в каталоге по умолчанию, задайте путь установки Domino в ответ на следующий вопрос:
Введите путь каталога Lotus Notes

Например, на компьютере Linux это может быть /opt/lotus/notes/latest/linux.

- Ответьте на следующий вопрос Д (Да), если каталог данных Domino Server установлен в каталоге по умолчанию, и Н (Нет), если это не так:
Найден путь каталога данных Lotus Notes /local/notesdata.
Это правильный путь каталога данных Lotus Notes?

Путь по умолчанию - /local/notesdata.

- Если каталог данных Domino Server не внедрен в положение по умолчанию на сервере искателя, задайте путь данных Domino в ответ на следующий вопрос:
Введите путь каталога данных Lotus Notes.

4. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall
esadmin system startall
```

Понятия, связанные с данным

“Искатели Domino Document Manager” на стр. 64

“Искатели Notes” на стр. 76

“Искатели QuickPlace” на стр. 86

Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino

Если вы устанавливаете OmniFind Enterprise Edition на компьютере с Microsoft Windows и планируете просмотр серверов, использующих протокол Notes Remote remote procedure (NRPC), надо запустить сценарий для конфигурирования сервера искателя. Этот сценарий разрешает искателям Примечания, QuickPlace и Domino Document Manager обмениваться информацией с серверами базы данных.

Ограничения

Сервер Lotus Domino и клиент Lotus Notes не могут работать одновременно и на одном компьютере с искателем Примечания, QuickPlace или Domino Document Manager, сконфигурированным для использования протокола NRPC. При попытке запустить один из этих искателей во время работы Domino Server появится сообщение об ошибке, и искатель будет остановлен.

Об этой задаче

Искатели, работающие по протоколу NRPC, используют библиотеки клиента Lotus Domino. Эти библиотеки устанавливаются при установке Lotus Domino Server на сервере искателя. Чтобы искатели могли работать с библиотеками Domino, запустите сценарий настройки, который OmniFind Enterprise Edition помещает на сервер искателя после установки библиотек Domino.

Процедура

Чтобы сконфигурировать сервер искателя для просмотра серверов Lotus Notes, Lotus QuickPlace и Domino Document Manager:

1. Зарегистрируйтесь на сервере искателя с ID пользователя, входящего в группу администраторов. У этого ID пользователя должны быть полномочия на установку Lotus Notes.
2. Установите Lotus Notes:
 - a. Вставьте компакт-диск Domino Server в дисковод компакт-дисков. (Если у вас нет этого компакт-диска, можно загрузить его образ.)
 - b. Запустите программу установки: `setup.exe`
 - c. Ответьте на запросы и примите значения по умолчанию или задайте другие параметры установки (например, пути для каталога установки и каталога данных).

Если вам нужна помощь, обратитесь к документации по Lotus Domino.
 - d. Убедитесь, что у ID администратора поисковой системы предприятия есть разрешение на доступ к каталогу данных Domino. Этот ID администратора задается при установке OmniFind Enterprise Edition.
3. На сервере искателя запустите сценарий конфигурирования, переданный OmniFind Enterprise Edition:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке OmniFind Enterprise Edition).
 - b. Запустите следующий сценарий (он установлен в каталоге `%ES_INSTALL_ROOT%\bin`):
`escrnote.vbs`
 - c. Отвечайте на вопросы программы:
 - Для следующего вопроса ответьте Y, если Lotus Notes установлен в каталог по умолчанию, или N, если Lotus Notes установлен в другой каталог:
Найден каталог Lotus Notes `c:\lotus\notes`.
Это правильный путь каталога Lotus Notes?

Типичный путь установки на компьютере Windows - `c:\lotus\notes` или `c:\lotus\domino`.
 - Если Lotus Notes установлен на сервере искателя не в каталоге по умолчанию, в ответ на следующий запрос задайте путь установки Lotus Notes:
Введите путь каталога Lotus Notes
 - Для следующего вопроса ответьте D, если каталог данных Lotus Notes размещен по умолчанию, или H, если это не так:
Найден каталог данных Lotus Notes `c:\lotus\notes\data`.
Это правильный путь каталога данных Lotus Notes?

Типичный путь каталога данных на компьютере Windows - `c:\lotus\notes\data` или `c:\lotus\domino\data`.

- Если каталог данных Lotus Notes размещен на сервере искателя не в положении по умолчанию, в ответ на следующий запрос задайте путь каталога данных:
Введите путь каталога данных Lotus Notes.
4. На сервере искателей остановите и перезапустите поисковую систему предприятия:

```
esadmin system stopall
esadmin system startall
```

Понятия, связанные с данным

“Искатели Domino Document Manager” на стр. 64

“Искатели Notes” на стр. 76

“Искатели QuickPlace” на стр. 86

Конфигурирование серверов, использующих протокол DIIOP

Чтобы искатель просматривал серверы, использующие протокол Domino Internet Inter-ORB Protocol (DIIOP), надо сконфигурировать сервер так, чтобы искатели Примечания, QuickPlace и Domino Document Manager могли использовать этот протокол..

Прежде чем приступить к работе...

На сервере, который вы хотите просматривать, должны быть запущены задачи DIIOP и HTTP.

Процедура

Чтобы сконфигурировать серверы, использующие протокол DIIOP:

1. Сконфигурируйте документ server:
 - a. Откройте документ server на сервере Lotus Notes, Lotus QuickPlace или Domino Document Manager, который вы хотите просматривать. Этот документ хранится в каталоге Domino.
 - b. На странице Configuration (Конфигурация) раскройте раздел server.
 - c. На странице Security (Защита) в области **Programmability Restrictions** (Ограничения программируемости) задайте соответствующие ограничения защиты для вашей среды в следующих полях:
 - **Run restricted Lotus Script/Java agents** (Запустить агенты Lotus Script/Java с ограничениями)
 - **Run restricted Java/Javascript/COM**(Запустить Java/Javascript/COM с ограничениями)
 - **Run unrestricted Java/Javascript/COM** (Запустить Java/Javascript/COM)

Например, можно разрешить неограниченный доступ агентов Lotus Script/Java, задав звездочку (*), или же задать имена пользователей, зарегистрированных в Domino Directory, на которых распространяются ограничения Java/Javascript/COM.

Важное замечание: Имена пользователей, заданные вами в этих полях, должны быть доступны для искателя, который вы сконфигурировали для просмотра этого сервера с протоколом DIIOP.

- d. Откройте страницу Internet Protocol (Протокол интернета), затем откройте страницу HTTP и задайте для опции **Allow HTTP clients to browse database** (Разрешить клиентам HTTP просматривать базу данных) значение **Yes** (Да).
2. Сконфигурируйте документ user:
 - a. Откройте документ user на сервере Lotus Notes, Lotus QuickPlace или Domino Document Manager, который вы хотите просматривать. Этот документ хранится в каталоге Domino.
 - b. На странице Basics (Основная информация) в поле **Internet password** (Пароль интернета) задайте пароль.
 Если для конфигурирования опций просмотра для этого сервера используется консоль администратора поисковой системы предприятия, задайте этот ID пользователя и пароль на странице, где задается просматриваемый сервер. Искатель использует эти параметры аутентификации для доступа к серверу.
 3. Перезапустите задачу DIIOP на сервере.
Понятия, связанные с данным
 “Искатели Domino Document Manager” на стр. 64
 “Искатели Notes” на стр. 76
 “Искатели QuickPlace” на стр. 86

Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino

Перед использованием искателей Примечания, QuickPlace или Domino Document Manager в системе IBM AIX необходимо установить модуль IOCP (I/O completion port, порт завершения ввода-вывода) и сконфигурировать его для использования искателем.

Об этой задаче

Без модуля IOCP при попытке создать искатель возникнет ошибка процессов поиска. Появится сообщение об ошибке:

FFQM0105E Получена ошибка от сервера -
 Сообщение: FFQG0024E Перехвачена непредвиденная исключительная ситуация: поиск

В файл \$ES_NODE_ROOT/logs/system_ггггммдд.лог записывается приведенное ниже сообщение, включающее ошибку ENOEXEC. (Часть текста сообщения разбита на несколько строк для лучшей читаемости.)

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]
ies10.yamato.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86
FFQ00277E Перехвачена исключительная ситуация, подробности:
'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a'
and a stack trace of 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a
at java.lang.ClassLoader$NativeLibrary.load(Native Method)
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)
at java.lang.Runtime.loadLibrary0(Runtime.java:824)
at java.lang.System.loadLibrary(System.java:908)
at lotus.domino.NotesThread.load(NotesThread.java:306)
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)
at com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover
(Unknown Source)
```

```

at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)
at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover
(ComponentDiscoveryW.java:72)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
at sun.reflect.DelegatingMethodAccessorImpl.invoke
(DelegatingMethodAccessorImpl.java:60)
at java.lang.reflect.Method.invoke(Method.java:391)
at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)

```

Процедура

Чтобы установить модуль ИОСР на сервер искателя и проверить правильность установки,

необходимо:

1. Установить модуль ИОСР (bos.iocr.rte) с компакт-диска продукта AIX на сервер искателя.

После установки модуля ИОСР, но до создания искателя Примечания, QuickPlace или Domino Document Manager примените программное исправление для модуля. Инструкции смотрите здесь:

<http://www.ibm.com/support/docview.wss?uid=swg21086556>

2. Чтобы проверить правильность установки модуля ИОСР на сервере искателя, введите команду:

```
$ ls1pp -l bos.iocr.rte
```

Вывод команды ls1pp должен быть аналогичен следующему:

Набор файлов	Уровень	Состояние	Описание

Path: /usr/lib/objrepos			
bos.iocr.rte	5.2.0.10	ВЫПОЛНЕНО	API ИОСР
Путь: /etc/objrepos			
bos.iocr.rte	5.2.0.10	ВЫПОЛНЕНО	API ИОСР

3. Введите следующую команду и убедитесь, что порт ИОСР находится в состоянии **Доступен**:

```
$ lsdev -Cc iocr
```

Вывод команды lsdev должен соответствовать следующему:

```
iocr0 Доступен ИОСР
```

4. Если состояние порта ИОСР - **Определен**, измените его на **Доступен**:
 - a. Зарегистрируйтесь на сервере искателя как root и введите команду:

```
# smit iocr
```
 - b. Выберите **Изменить / Показать характеристики ИОСР** и измените **СОСТОЯНИЕ, КОНФИГУРИРУЕМОЕ ПРИ ПЕРЕЗАПУСКЕ СИСТЕМЫ** с **Определен** на **Доступен**.
 - c. Перезагрузите сервер искателя.
 - d. Введите команду lsdev еще раз и убедитесь, что состояние порта ИОСР изменилось на **Доступен**.

Понятия, связанные с данным

“Искатели Domino Document Manager” на стр. 64

“Искатели Notes” на стр. 76

“Искатели QuickPlace” на стр. 86

Искатели QuickPlace

Чтобы включить площадки и комнаты Lotus QuickPlace в собрание поисковой системы предприятия, нужно сконфигурировать искатель QuickPlace.

Для просмотра площадок, управляемых службами Quickr, для Lotus Domino, можно также использовать искатель QuickPlace. Если используются службы Lotus Quickr для WebSphere Portal, при помощи искателя Список заполнения в собрание поисковой системы предприятия можно добавить документы, хранящиеся в библиотеках Lotus Quickr.

Совет:

Подробные примеры конфигурирования защищенного искателя QuickPlace смотрите в сценарии для небольшой организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурация сервера искателя

Если на сервере, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий конфигурирования. Этот сценарий, поставляемый с OmniFind Enterprise Edition, позволяет искателю QuickPlace взаимодействовать с серверами, использующими NRPC.

Если на сервере, который планируется просматривать искателем, используется протокол DIIOP (Domino Internet Inter-ORB Protocol), запускать сценарий конфигурирования на сервере искателя не требуется. Однако нужно сконфигурировать сервер назначения, чтобы к нему мог обращаться искатель QuickPlace.

Если сервер, на котором вы планируете запускать искатель, использует сервер Lightweight Directory Access Protocol (LDAP), сервер назначения необходимо сконфигурировать для использования протокола DIIOP (искатель QuickPlace не может использовать протокол NRPC для обработки данных LDAP). Необходимо также сконфигурировать базу данных Directory Assistance, а сервер назначения сконфигурировать для использования сервера LDAP в качестве вторичного сервера Domino.

Важное замечание: Если сервер назначения использует протокол DIIOP, а для шифрования данных, передаваемых между искателем и сервером данных, для искателя конфигурируется использование HTTPS или DIIOP через SSL, нужно скопировать файл TrustedCerts.class (например, c:\certs или /data/certs) с сервера назначения на сервер искателя. В конфигурации с двумя или четырьмя серверами файл TrustedCerts.class нужно также скопировать на серверы, где установлен компонент поиска. Кроме того, нужно убедиться, что положение этого файла на сервере искателей совпадает с положением на серверах поиска. Путь каталога для файла TrustedCerts.class задается при конфигурировании искателя.

При конфигурировании искателя и задании ID пользователя, который будет использовать искатель, нужно задать ID с полномочиями, достаточными для обращения ко всем площадкам QuickPlace в домене Domino. Для этого включите ID пользователя в особую группу под названием QuickPlaceAdministratorsSUGroup.

Если OmniFind Enterprise Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель QuickPlace с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 83.
- “Конфигурирование сервера QuickPlace для использования защиты Local User” на стр. 288.
- “Конфигурирование Directory Assistance на сервере QuickPlace” на стр. 289.
- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84.

Защита на уровне документов

Если включена защита собрания и на сервере, который планируется просматривать искателем, используется протокол NRPC (Notes Remote Procedure Call - удаленный вызов процедур Notes), на сервере искателя нужно сконфигурировать Lotus Domino Trusted Server. Trusted Server используется для обеспечения соблюдения управления доступом на уровне документов. Перед тем, как сделать собрания доступными пользователям для поиска, выполните следующие задачи:

- Сконфигурируйте доверенные серверы Lotus Domino для проверки параметров аутентификации пользователей
- Включите глобальную защиту в WebSphere Application Server и конфигурирование прикладных программ поиска для использования защиты. Этот шаг гарантирует проверку параметров аутентификации при попытках пользователей использовать программу поиска. Серверы поиска используют параметры аутентификации для проверки полномочий каждого из пользователей на доступ к документам Lotus Domino.

Просмотр вложений

В Lotus QuickPlace можно импортировать и публиковать документы Microsoft Office (опции выбираются при включении импорта страниц Imported Page, Microsoft Word Page, Microsoft Excel Page, Microsoft PowerPoint Page и Multiple Imported Pages). Искатель QuickPlace может обрабатывать эти типы импортированных документов в качестве вложений только при следующих условиях:

- На сервере используется протокол DIIOP.
- При конфигурировании опций просмотра для искателя был разрешен просмотр вложений.
- Искатель сконфигурирован для просмотра поля "\$FILE" или всех полей.

Конфигурирование - обзор

Искатель QuickPlace можно использовать для просмотра любого числа площадок QuickPlace. При создании искателя вы выбираете площадки для просмотра искателем с одного сервера QuickPlace. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера QuickPlace, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все обрабатываемые площадки, или только комнаты, которые вы зададите.

Ограничение: При задании ID пользователя, который будет использовать искатель, обязательно задайте ID с полномочиями, достаточными для обращения ко всем площадкам QuickPlace в домене Domino. Это можно сделать, сконфигурировав сервер QuickPlace и назначив ID пользователя для зарезервированной группы QuickPlaceAdministratorsSUGroup. Имейте в виду, что имя группы не содержит пробелов и оно регистрозависимо.

Чтобы создать или изменить искатель QuickPlace, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Идентифицировать сервер и протокол связи QuickPlace.
- Если сервер сконфигурирован для использования протокола DPOP, можно задать, как искатель будет соединяться с объектами Domino. Например, можно задать опции для использования HTTPS или SSL (Secure Sockets Layer) для шифрования соединений.
- Указать информацию о каталоге пользователя, связанном с сервером (эта информация нужна искателю для соблюдения управления доступом при выполнении пользователями поиска в собрании).
- Выбрать площадки, которые будет просматривать искатель.
- Составить расписания просмотра искателем площадок.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все комнаты на площадке или только документы, находящиеся в выбранных вами комнатах.
- Задать опции, обеспечивающие доступность поиска для полей в различных площадках и комнатах. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Советы по просмотру баз данных Lotus Domino” на стр. 78

“Обеспечение защиты уровня документов для документов Lotus Domino” на стр. 286

Задачи, связанные с данной

“Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79

“Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81

“Конфигурирование серверов, использующих протокол DIIOP” на стр. 83

“Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84

“Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей” на стр. 286

“Конфигурирование сервера QuickPlace для использования защиты Local User” на стр. 288

“Конфигурирование Directory Assistance на сервере QuickPlace” на стр. 289

Искатели с начальным списком

Если вы применяете службы IBM Lotus Quickr для WebSphere, можно использовать искатель Список заполнения crawler для добавления документов из библиотек Lotus Quickr в собрания поисковой системы предприятия.

Библиотека Lotus Quickr - это контейнер для файлов документов. Искатель Список заполнения не поддерживает просмотр содержимого Web, такого как wiki-материалы и блоги.

Если вы применяете службы Lotus Quickr для Lotus Domino, можно использовать искатель QuickPlace для добавления документов Lotus Quickr в собрание.

Конфигурирование сервера WebSphere Portal

Если на сервере WebSphere Portal Версии 6 установить Lotus Quickr, искатель Список заполнения позволит просматривать документы библиотек Lotus Quickr. Опции для просмотра искателем этих документов можно сконфигурировать отдельно от опций, задаваемых вами для сайтов портала, просматриваемых искателем WebSphere Portal.

Перед созданием искателя Список заполнения нужно выполнить действия по конфигурированию поисковой системы предприятия в WebSphere Portal. Для конфигурирования поисковой среды предприятия, на серверах поиска нужно запустить сценарий (wp6_install.sh - в AIX, Linux или Solaris и wp6_install.bat - в Windows), поставляемый с OmniFind Enterprise Edition.

Строка пользовательского агента указывает, какой браузер или робот обращается к серверу. При просмотре сервера Lotus Quickr искатель Список заполнения использует строку пользовательского агента OmniFind SeedlistCrawler/1.0.

Конфигурирование - обзор

Искатель Список заполнения позволяет просматривать любое число документов Lotus Quickr. Сервер для обработки задается при конфигурировании искателя. После этого искатель просматривает все документы в библиотеках Lotus Quickr на этом сервере.

Просматриваемые документы должны быть доступны по одному и тому же ID администратора и паролю Lotus Quickr. Для обработки искателем сайтов, на которых используются другие параметры аутентификации, нужно сконфигурировать отдельный искатель Список заполнения.

Чтобы создать или изменить искатель Список заполнения, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать

искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем документов Lotus Quickr.
- Задать сервер, с которыми должен работать искатель, и информацию, позволяющую искателю соединиться с этим сервером.

При создании или редактировании искателя можно проверить способность искателя соединиться с просматриваемыми документами. Еще до запуска искателя сообщения покажут, сможет ли искатель обращаться к документам, с которыми он должен работать.

- Укажите опции защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют управлять доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

Важное замечание: При поиске в защищенных документах Lotus Quickr для передачи требований поиска должен использоваться портлет поиска для поисковой системы предприятия из WebSphere Portal. У поиска, запущенного из примера поисковой программы, ESSearchApplication, нет необходимых параметров аутентификации, и он не может проверить полномочия пользователя для доступа к документам.

- Задать информацию, позволяющую искателю связываться с прокси-сервером, если для обслуживания страниц используется прокси-сервер.
- Если для защиты сервера WebSphere Portal и документов Lotus Quickr используется другой продукт (например, IBM Tivoli Access Manager WebSEAL или CA SiteMinder SSO Agent for PeopleSoft), задать полномочия единой регистрации, которые позволят искателю обращаться к документам на сервере.
- Задать информацию о файле склада ключей, чтобы искатель мог использовать для связи с сервером протокол Secure Sockets Layer (SSL).
- Укажите язык и кодовую страницу документов, просматриваемых искателем.
- Задать опции для обработки искателем и поиска метаданных в документах Lotus Quickr.
- Задать расписания для просмотра искателем документов Lotus Quickr.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Интеграция с WebSphere Portal” на стр. 347

Задачи, связанные с данной

“Конфигурирование поисковой системы предприятия в WebSphere Portal версии 6” на стр. 354

“Конфигурирование портлета поисковой системы предприятия для Lotus Quickr” на стр. 360

Искатели файловой системы UNIX

Чтобы включить документы, хранимые в файловых системах AIX, Linux или Solaris, в собрание поисковой системы предприятия, нужно сконфигурировать искатель Файловая система UNIX.

Искатель Файловая система UNIX можно использовать для просмотра любого числа файловых систем. При конфигурировании искателя нужно выбрать локальные и удаленные каталоги и подкаталоги, которые будет просматривать искатель.

При установке сервера искателя на компьютере Windows этот сервер нельзя использовать для просмотра источников файловых систем AIX, Linux или Solaris (искатель Файловая система UNIX не выводится в списке доступных типов искателей).

Искатель Файловая система UNIX обрабатывает документы в соответствии с разрешениями на чтение, указанными для администратора поисковой системы предприятия.

Чтобы создать или изменить искатель Файловая система UNIX, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подкаталогов в пространстве искателя.
- Составить расписания просмотра искателем файловых систем.
- Выбрать подкаталоги и уровни подкаталогов, с которыми должен работать искатель.
- Задать опции, обеспечивающие доступность поиска для документов в подкаталогах. Например, можно исключить из пространства для искателя определенные типы документов.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Искатели Web

Чтобы включить страницы с сайтов в собрание поисковой системы предприятия, нужно сконфигурировать искатель Web.

При помощи искателя Web можно просматривать любое число серверов HTTP. Искатель обращается к сайту и читает с него данные. Затем по ссылкам в документах он может просмотреть дополнительные документы. Искатель Web может просматривать и извлекать ссылки с отдельных страниц или из *наборов фреймов* (страниц, созданных с использованием фреймов HTML).

Просматриваемые искателем данные могут быть одного из многих обычных форматов; эти данные искатель может получать из разнообразных источников во внутренней сети или Интернете. К обычным форматам относятся HTML, PDF, Microsoft Word, Lotus WordPro, Extensible Markup Language (XML) и так далее.

Совет:

Подробные примеры конфигурирования защищенного искателя Web смотрите в сценарии для средней организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Чтобы создать или изменить искатель Web, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Сконфигурировать искатель может только администратор поисковой системы предприятия или администратор собрания, которому принадлежит этот искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех страниц Web в пространстве для искателя.
- Задать правила для разрешения и запрета обращения к определенным сайтам. Задав правила для искателя, можно протестировать их и проверить, что искатель может обращаться к сайтам, которые вы хотите включить в пространство искателя.
- Задайте опции, чтобы включить определенные типы файлов и исключить файлы с определенными расширениями файла.
- Задать правила обработки искателем Web произвольных страниц ошибок.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.
- Задать опции для просмотра искателем сайтов, защищенных паролем (просматриваемые искателем Web-серверы должны использовать для предложения ввода паролей базовую аутентификацию HTTP или формы HTML).
- Задать опции для просмотра искателем сайтов, обслуживаемых прокси-сервером.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурация пользовательского агента

Для просмотра сайта, на котором используется протокол Robots Exclusion, нужно, чтобы в файле robots.txt на этом сайте был разрешен доступ к этому сайту для имени пользовательского агента, заданного в конфигурации искателя Web.

При запуске поисковой системы предприятия искатель Web загружает имя пользовательского агента, заданное в его конфигурации. Прежде чем загружать страницу с сайта, который искатель еще не посещал (или не посещал некоторое время), искатель сначала пытается загрузить файл с именем robots.txt. Этот файл находится в корневом каталоге сайта.

Если файл robots.txt не существует, это значит, что данный сайт открыт для неограниченного просмотра. Если этот файл существует, он указывает, какие области

сайта (каталоги) недоступны для искателей. В файле `robots.txt` задаются разрешения для искателей (искатели идентифицируются по их именам пользовательских агентов).

Протокол Robots Exclusion не является обязательным, но искатели Web поисковой системы предприятия стараются следовать ему:

- Если в файле `robots.txt` есть запись для имени пользовательского агента, заданного в конфигурации искателя Web, этот искатель Web подчиняется ограничениям, наложенным на данный пользовательский агент.
- Если этого имени пользовательского агента нет в файле `robots.txt`, но в последней записи в файле задано `User-agent: *` (что означает "любой пользовательский агент") и задано ограничение `Disallow: /` (что означает "не разрешать никакой доступ, начиная с корневого каталога сайта"), искатель Web не сможет выполнять просмотр этого сайта.
- Если этого имени пользовательского агента нет в файле `robots.txt`, но в последней записи в файле задано `User-agent: *` и задано ограничение `Allow: /`, искателю Web разрешено выполнять просмотр этого сайта.

Администраторы сайтов часто задают такую последнюю запись, запрещающую доступ всем искателям, кроме тех, которым доступ разрешен явным образом. Если вы конфигурируете новый искатель Web и знаете, что на некоторых сайтах, для которых вы хотите выполнять просмотр, используется протокол Robots Exclusion, попросите администраторов этих сайтов добавить записи для вашего искателя в соответствующие файлы `robots.txt`.

В свойствах искателя Web и во всех файлах `robots.txt` на интересующих вас сайтах должно быть задано одно и то же имя пользовательского агента.

Если ни один из сайтов, для которых нужно выполнять просмотр, не использует протокол Robots Exclusion, значение, заданное для имени пользовательского агента, обычно несущественно. Однако некоторые серверы прикладных программ, JSP и сервлеты генерируют разные ответы для разных имен пользовательского агента. Например, это могут быть различные ответы для обработки несовместимостей браузеров. В таких случаях имя пользовательского агента, заданное для искателя Web, может быть существенным (независимо от протокола Robots Exclusion). Если нужно выполнять просмотр сайтов таких типов, обратитесь к администраторам этих сайтов, чтобы убедиться, что искателю Web разрешен доступ.

Как искатель Web использует протокол Robots Exclusion

Искатель Web учитывает требования протокола Robots Exclusion и не просматривает сайты, если правила в файле `robots.txt` на сервере не разрешают работу искателей.

Загрузка считается успешной, когда искатель может получить файл `robots.txt` с Web-сервера или получить подтверждение того, что файла `robots.txt` не существует. Загрузка считается неудачной, когда искатель не может получить правила или получить подтверждение, что файл `robots.txt` существует.

Успешная загрузка не означает, что у искателя есть разрешение на работу, так как правила в файле `robots.txt` могут запрещать работу искателей. Неудачная загрузка временно запрещает работу искателя, так как искатель не может определить, что задано в правилах.

Шаги, которые искатель выполняет при попытке загрузки файла `robots.txt`:

1. Когда искатель обнаруживает новый сайт, он пытается получить IP-адрес сервера. Если эта попытка не удалась, искатель не сможет выполнить просмотр.
2. Когда доступен как минимум один IP-адрес, искатель пытается загрузить файл `robots.txt` при помощи команды протокола **HTTP (или HTTPS) GET**.
3. Если при этом истекает срок ожидания, или соединение гнезда разрывается, или происходит другая ошибка низкого уровня (такая, как ошибка сертификата SSL), искатель записывает эту ошибку в журнал и повторяет попытку для всех IP-адресов, известных для сервера назначения.
4. Если соединение не установлено после того, как искатель использовал все адреса, он ждет две секунды, а затем пытается установить соединение по этим адресам еще раз.
5. Если соединение установлено и выполнен обмен заголовков HTTP, изучается возвращенное состояние. Если код состояния 500 или выше, искатель считает соединение плохим и переходит к другим IP-адресам. При любом другом состоянии искатель прекращает попытки соединений по другим IP-адресам и действует в соответствии с кодом состояния.

После того, как искатель получит код состояния HTTP ниже 500 или испробует все IP-адреса дважды, он действует следующим образом:

1. Если не получено состояние HTTP меньше 500, сайт считается неопределенным в данное время.
2. Если получено состояние HTTP 400, 404 или 410, сайт определяется для просмотра искателем без правил.
3. Если получено состояние HTTP от 200 до 299, дальнейшие действия определяются следующими условиями:
 - Если содержимое усечено, сайт считается неопределенным в данное время.
 - Если анализ содержимого выполнен без ошибок, сайт определяется для просмотра искателем с найденными правилами.
 - Если анализ содержимого выполнен с ошибками, сайт определяется для просмотра искателем без правил.
4. Если получено любое другое состояние HTTP, сайт считается неопределенным в данное время.

Когда искатель пытается загрузить файл `robots.txt` для сайта, он обновляет для этого сайта отметку времени, то есть дату файла `robots`. Если сайт признан неопределенным из-за недоступности информации в файле `robots.txt`, увеличивается значение счетчика ошибок `robots`.

По истечении заданного времени искатель снова пытается получить информацию из файла `robots.txt` для сайта. Когда число ошибок достигает максимального заданного значения, искатель прекращает попытки получения файла `robots.txt` и считает сайт неопределенным для просмотра.

После того, как сайт определен для просмотра искателем (проверка правил в файле `robots.txt` выполнена успешно), счетчик попыток устанавливается на ноль. Искатель использует результаты загрузки, пока не истечет период времени для проверки правил. После этого сайт должен быть определен снова.

Совет:

- Если сервер возвращает содержимое, но оно содержит синтаксические ошибки, либо сервер использует протокол `robots`, отличный от версии 1994, либо в файле содержится что-либо отличное от правил `robots` (например, страница ошибки программного обеспечения), искатель действует так, как будто у сервера

отсутствуют применимые правила, и просматривает сайт. Обычно такое действие является правильным, так как администраторы собраний не управляют содержимым сайта или поведением сервера по умолчанию. Если администратор Web-сервера не хочет, чтобы сайт просматривался искателем, но при этом не хочет устанавливать соответствующий файл правил, администратор собрания может заблокировать сайт для искателя Web, указав домен, IP-адрес или префикс HTTP сайта в правилах искателя.

- Если сервер возвращает код состояния 302 или другие коды перенаправления, искатель интерпретирует этот код как означающий, что на сайте есть файл `robots.txt`, который должен быть использован, но этот файл находится в неправильном положении (не в корне файла). Администратор Web-сервера должен переместить файл в правильное положение, чтобы искатель Web мог соблюдать правила в этом файле.
- Если возникли проблемы с сертификатами (например, сертификат просрочен, сертификат не является доверенным или сертификат самоподписанный, а искатель не сконфигурирован для принятия самоподписанных сертификатов), искатель интерпретирует это как ошибку связи с сайтом и считает сайт неопределенным. В общем случае те же самые проблемы могут препятствовать просмотру искателем других страниц сайта. Чтобы сделать сайт доступным для просмотра искателем, администратор собрания должен разрешить самоподписанные сертификаты, добавить сертификата сайта в файл склада доверенных ключей или попросить администратора Web-сервера получить обновленный сертификат.
- Искатель Web должен быть сконфигурирован для использования базовой аутентификации HTTP (включая базовую аутентификацию HTTP прокси-сервера). При правильном конфигурировании аутентификации требуется также для загрузки файлов `robots.txt`. Код состояния 403, 407 или другие ответы, связанные с аутентификацией, указывают на ошибки авторизации, и искатель считает сайт неопределенным. (Поддерживается только базовая аутентификация HTTP.)
- Если длина файла `robots.txt` для сайта превышает максимальную длину для страницы `robots`, администратор собрания может увеличить сконфигурированное максимальное значение (значение по умолчанию - один миллион байтов - должно быть достаточным).

Чтобы помочь исправлению ошибок, при слежении за искателем Web можно затребовать отчет с сайта. Выберите опции для просмотра содержимого файла `robots.txt` (чтобы увидеть, запрещают ли правила доступ искателю Web на сайт), просмотрев дату и время последней попытки искателя загрузить файл `robots.txt` (искатель не будет делать следующей попытки, пока не истечет заданный между попытками период времени), а также сколько последовательных попыток искателя загрузить файл `robots.txt` закончилось неудачно. Во время слежения за искателем Web выберите **Справка**, чтобы узнать больше об этих опциях отчетов сайтов, и о том, как интерпретировать полученные результаты.

Дополнительную информацию о протоколе Robots Exclusion смотрите по адресу <http://www.robotstxt.org/wc/exclusion.html>.

Поддержка JavaScript

Искатель Web для поисковой системы предприятия может находить некоторые ссылки (адреса URL), входящие в разделы JavaScript документов Web.

Искатель Web может находить как относительные, так и абсолютные ссылки. Если документ HTML содержит элемент `BASE`, искатель использует этот элемент для

получения из относительных ссылок абсолютных. В противном случае искатель использует для этого URL самого документа.

Поддержка JavaScript ограничена извлечением ссылок. Искатель не выполняет синтаксический анализ JavaScript, не строит объектную модель документа (Document Object Model - DOM), не интерпретирует и не выполняет операторы JavaScript. Искатель ищет в содержимом документа (в том числе в разделах JavaScript, но не только в них) строки, которые похожи на адреса URL в операторах JavaScript. Это означает две вещи:

- Будут найдены некоторые URL, которые были бы проигнорированы более строгим синтаксическим анализатором HTML. Искатель отвергнет все строки, не являющиеся синтаксически правильными адресами URL, но некоторые из допустимых URL, возвращенных шагом просмотра, могут не представлять интереса для поиска.
- Содержимое документа, генерируемое в JavaScript (например, когда пользователь просматривает страницу и браузер выполняет JavaScript), будет недоступно искателю Web и, следовательно, не будет индексироваться.

Поскольку искатель Web не анализирует JavaScript в файлах HTML, URL в JavaScript им не обрабатываются. Чтобы включить поддержку просмотра искателем Web этих URL в JavaScript, можно выполнить одно из следующих действий:

- Отредактировать на консоли администратора поиска искатель Web, а на странице Пространство для искателя Web добавить эти URL в список URL, который искатель будет использовать в качестве отправной точки для добавления нужных URL в собрание (**Начальные URL**). Чтобы изменения вступили в силу, следует перезапустить искатель Web (запускать полный просмотр не требуется).
- С помощью тега привязки () задать нужные URL в файле HTML в качестве гипертекстовых ссылок.

Правила ограничения пространства для искателя Web

Чтобы пользователи могли обращаться только к сайтам, на которых вы хотите разрешить им поиск, нужно задать правила, ограничивающие пространство просмотра данных искателем.

Когда искатель Web просматривает Web-страницы, он находит ссылки на другие страницы и помещает эти ссылки в очередь для следующего просмотра. Просмотр и поиск могут повторяться столько раз, сколько позволяют ресурсы времени и памяти. При конфигурировании искателя Web нужно задать URL, с которых искатель должен начинать просмотр. С этих URL (они называются *начальными URL*) искатель Web может дойти до любого документа в Web, перейдя к нему по прямым ссылкам или через промежуточные страницы.

Чтобы ограничить пространство для искателя, сконфигурируйте искатель Web, чтобы он детально просматривал определенные URL и игнорировал ссылки на области, которые вас не интересуют. Поскольку по умолчанию искатель принимает любые обнаруживаемые им URL, необходимо задать правила, задающие URL, которые вы хотите включить в собрание, и исключать остальные страницы.

Просматриваемую и не просматриваемую искателем Web область можно задать несколькими способами. Вы можете задать:

- Список начальных URL, с которых искатель должен начинать просмотр
- Три типа правил просмотра искателем: правила доменов, IP-адресов и префиксов URL
- Список типов MIME для включения документов

- Список расширений файлов для исключения документов
- Максимальное число каталогов в пути URL

Для правил просмотра используется следующий формат:
действие тип назначение

где действие - запрет или разрешение (forbid или allow), тип - домен (domain), IP-адрес (address) или префикс URL (HTTP или HTTPS), а назначение зависит от значения типа. Для ограничения путей можно использовать в качестве символа подстановки звездочку (*), чтобы задать назначения, соответствующие шаблону.

Правила доменов

Назначение правила доменов - это имя DNS. Например, можно задать, что искатель должен просмотреть весь домен `www.ibm.com`:

```
allow domain www.ibm.com
```

В качестве символа подстановки можно задать звездочку, означающую, что правило применимо ко всем именам хостов, где есть последовательность символов из остальной части шаблона. Например, можно задать, что искатель не должен просматривать домены, имена которых начинаются с `server` и заканчиваются `ibm.com`

```
forbid domain server*.ibm.com
```

И для явно заданного имени домена, и для шаблона имен доменов при определении соответствия учитывается регистр символов. Например, `*.user.ibm.com` соответствует `joe.user.ibm.com` и `mary.smith.user.ibm.com`, но не соответствует `joe.user.IBM.com`.

Правило доменов, в котором не задан номер порта, применяется ко всем портам в этом домене. Следующий пример разрешает все порты в домене `sales`:

```
allow domain sales.ibm.com
```

Если в правиле домена указан номер порта, правило применяется только к этому порту. Следующий пример разрешает только порт 443 в домене `sales`:

```
allow domain sales.ibm.com:443
```

Правила префиксов

Правило префиксов управляет просмотром искателем URL, начинающихся с заданной строки. В качестве назначения задается один URL, содержащий обычно одну или несколько звездочек для шаблона. Например, часто задают звездочку в качестве последнего символа в строке префикса.

Правило префиксов позволяет просмотреть искателем весь сайт или его часть. Можно задать путь каталога или шаблон, а затем разрешить или запретить просмотр всего из этой точки в дереве каталогов. Например, следующие правила, работая совместно, разрешают искателю просматривать все в каталоге `public` на сайте `sales.ibm.com`, но запрещают ему обращаться ко всем остальным страницам на этом сайте:

```
allow prefix http://sales.ibm.com/public/*
forbid prefix http://sales.ibm.com/*
```

При задании правила префиксов можно задать сразу несколько звездочек, причем в любой позиции строки префикса, кроме последней. Например, следующее правило запрещает искателю просматривать любые документы в каталогах верхнего уровня

на сайте sales.ibm.com, если имя каталога заканчивается на fs. (У вас, например, могут быть смонтированы файловые системы, не содержащие информации, полезной для индекса поиска информации.)

```
forbid http://sales.ibm.com/*fs/*
```

Правила адресов

При помощи правила адресов можно управлять просмотром целых хостов или сетей, задавая в качестве назначения задать IP-адрес и маску сети. Например:

```
IPv4 allow address 9.0.0.0 255.0.0.0
```

IPv6 Если вы используете поисковую систему предприятия на сервере Windows 2003 и разрешили применение протокола IP версии 6 (IPv6), надо заключать адрес в квадратные скобки.

```
allow address [2001:db8:0:1:0:0:0:1]
```

Маска сети позволяет задать соответствие шаблону. Чтобы правило адресов применялось к IP-адресу кандидата, IP-адрес в правиле и IP-адрес кандидата должны быть идентичны, за исключением тех мест, где в маске сети стоят нули. Правило адресов задает шаблон, а маска сети определяет значащие биты в шаблоне адреса. Ноль в маске сети действует как символ подстановки и означает, что совпадает любое значение, задаваемое в этой же битовой позиции в адресе.

В предыдущем примере применяется разрешающее правило к любому IP-адресу с 9 в первом октете и любым значением в последних трех октетах.

В качестве заключительного адреса в список правил полезно включить следующее правило. Это правило соответствует всем адресам, поскольку маска сети делает все биты незначащими (это правило запрещает все адреса, не разрешенные предшествующим правилом в списке правил).

```
IPv4 forbid address 0.0.0.0 0.0.0.0
```

IPv6

```
forbid address :: ::
```

Ограничения для прокси-серверов: Если вы планируете работу искателя с сайтами, обслуживаемыми прокси-серверами, не задавайте правила IP-адресов. Обычно прокси-сервер используется, когда у пользовательского агента (браузера или искателя) нет прямого доступа к сетям, в которых находятся нужные Web-серверы. Например, прокси-сервер HTTP может передавать запросы HTTP от искателя на Web-сервер, а потом ответы - обратно искателю.

Когда Web-искатель использует прокси-сервер, IP-адрес этого прокси-сервера - единственный IP-адрес другого хоста, известный искателю. Если правила IP-адресов разрешают искателю доступ только к IP-адресам подсети, почти все адреса URL будут классифицированы с кодом состояния 760 (указывающим, что это запрещенный адрес в пространстве Web).

Порядок правил просмотра искателем

Искатель в процессе обнаружения и просмотра URL применяет правила просмотра в различное время. Порядок правил важен, но только в пределах каждого типа правил. Важен порядок использования одного правила адресов относительно другого правила адресов, но порядок правил адресов относительно правил префиксов не важен.

В пределах набора правил одного типа искатель проверяет домен, адрес или URL кандидата для каждого правила, с первого заданного правила до последнего, пока не найдет применимое правило. Используется действие, заданное для первого применимого правила.

Зависимость от порядка определяет обычную структуру большей части правил просмотра:

- Набор правил доменов обычно начинается с запрещающих правил, исключающих из пространства искателя отдельные домены. Например, администратор собрания может указать конкретные домены, не содержащие полезной информации.
- После списка запрещающих правил обычно следуют серии разрешающих правил (с символами подстановки), позволяющих искателю посещать любые домены, имена которых заканчиваются одним из имен доменов верхнего уровня, определяющих внутреннюю сеть предприятия (например, *.ibm.com или *.lotus.com).

Завершите набор правил доменов приведенным ниже правилом по умолчанию, исключающим домены, которые не были разрешены предшествующими правилами:

```
forbid domain *
```

Это последнее правило играет важную роль - оно не дает включить в пространство искателя всю сеть Интернет.

- Набор правил адресов обычно начинается с небольшого числа разрешающих правил, позволяющих искателю просматривать сети верхнего уровня (класса А, В или С) во внутренней сети предприятия.

Примеры задания окончательного правила в списке правил адресов, предотвращающего просмотр искателем сайтов вне корпоративной сети, смотрите в предыдущем разделе.

- Набор правил префиксов обычно самый большой, поскольку содержит произвольные подробные спецификации разрешенных и запрещенных регионов, которые задаются как деревья и поддеревья. Рекомендуется сначала разрешить или запретить наиболее тесно локализованные регионы, а затем задать противоположное правило в более общем шаблоне, чтобы разрешить или запретить все остальное.

Как правило, раздел префиксов не заканчивается обычным правилом. Задание предлагаемых заключительных правил для доменов и адресов гарантирует, что искатель не будет выполнять просмотр вне сети предприятия; это проще, чем проверка префиксов URL.

Искатель сможет применять правила префиксов более эффективно, если сгруппировать эти правила по действиям (запрещающим или разрешающим). Например, вместо коротких, альтернативных друг другу последовательностей разрешающих и запрещающих правил задайте длинную последовательность правил, предусматривающих одно действие, а затем задайте длинную последовательность правил, предусматривающих другое действие. Чтобы получить нужную конфигурацию пространства искателя, разрешающие и запрещающие правила можно чередовать. Но группирование вместе разрешающих правил и группирование запрещающих правил может улучшить производительность искателя.

Расширения файлов, типы MIME и максимальная глубина просмотра искателем

Следующие опции обеспечивают дополнительные способы задания содержимого пространства для искателя. Можно исключать определенные типы документов на основе их расширений и включать определенные типы документов на основе их типов

MIME. При задании типов MIME для просмотра искателем учтите, что тип MIME в документах Web часто задается неправильно.

Максимальная глубина просмотра искателем - это число дробных черт в URL от корня сайта. Эта опция позволяет предотвратить просмотр искателем рекурсивных структур файловых систем на неопределенную глубину. Глубина просмотра не соответствует уровням, по которым искатель переходит по ссылкам от одного документа к другому.

Начальные URL

Начальные URL - это URL, с которых искатель начинает просмотр; они включаются в просмотр при каждом запуске искателя. Если начальные URL уже обнаружены, они не будут просматриваться или повторно просматриваться искателем быстрее, чем другие сайты, разрешенные вами в правилах просмотра.

Начальный URL важен только при первом запуске искателя Web, когда пространство искателя пусто. Кроме того, начальный URL важен, если вы добавили в список начальных URL пространства искателя не обнаруженный ранее URL.

В качестве начальных URL нужно задавать полные адреса URL, а не просто имена доменов. Необходимо задать протокол и номер порта, если это не порт 80.

Допустимы следующие начальные URL:

```
http://w3.ibm.com/  
http://sales.ibm.com:9080/
```

Следующий начальный URL недопустим:

```
www.ibm.com
```

Начальные URL нужно включить в правила просмотра. Например, искатель не сможет начать просмотр с заданного начального URL, если правила просмотра не разрешают его просматривать.

Поддержка адресов IPv6: Если вы используете поисковую систему предприятия на сервере Windows 2003 и разрешили применение протокола IP версии 6 (IPv6), надо заключать начальные URL в квадратные скобки. Например:

```
http://[2001:db8:0:1:0:0:0:1]  
http://[2001:db8:0:1::1]
```

Задачи, связанные с данной

“Включение поддержки протокола IPv6” на стр. 28

Проверка соединений URL для искателя Web

Задав URL для искателя Web, можно протестировать конфигурацию правил искателя.

Можно либо нажать кнопку **Проверка**, когда вы задаете домены, префиксы HTTP или IP-адреса для работы искателя, либо выбрать страницу Проверить адреса URL, чтобы проверить, может ли искатель соединиться с начальными URL в дополнение к заданными вами URL.

Результаты тестирования показывают, доступны ли искателю URL при использовании имени пользовательского агента, заданного в свойствах искателя. Кроме того, результаты тестирования показывают, допускают ли просмотр URL

правила исключения (например, просмотр документа может быть запрещен из-за того, что расширение файла соответствует расширению, исключенному из пространства искателя).

Когда сайт просмотрен искателем хотя бы один раз, можно протестировать URL для получения дополнительной информации. Например, отчет тестирования может включать последний код состояния HTTP (который показывает успешность просмотра URL искателем), время последнего просмотра URL, время следующего просмотра согласно расписанию и информацию о том, использует ли пользовательский агент текущий файл `robots.txt` Web-сервера.

Задание интервала повторного просмотра для искателя Web

Для управления частотой обращения искателя Web к адресам URL нужно задать опции в свойствах искателя Web.

Большинство искателей других типов в поисковой системы предприятия запускаются согласно расписаниям, задаваемым администратором. В отличие от них, искатель Web после запуска обычно работает непрерывно. Для управления частотой повторных посещений URL, уже просмотренных искателем, нужно задать минимальный и максимальный интервалы повторного просмотра.

Если при создании искателя Web или редактировании его свойств используется консоль администратора поисковой системы предприятия, можно выбрать опцию, позволяющую сконфигурировать дополнительные свойства. На странице *Дополнительные свойства искателя Web* можно задать опции для минимального и максимального интервалов повторного просмотра. Искатель Web будет использовать заданные вами значения для вычисления интервала повторного просмотра данных.

При первом просмотре страницы искатель использует дату и время просмотра, а также среднее значение от заданных вами минимального и максимального интервалов повторного просмотра, чтобы установить дату ее повторного просмотра. До этой даты страница просматриваться не будет. Время повторного просмотра страницы после этой даты зависит от нагрузки искателя и соотношения между новыми и старыми URL в пространстве для искателя.

При каждом повторном просмотре данной страницы искатель проверяет, не изменилось ли ее содержимое. Если содержимое изменилось, следующий интервал повторного просмотра будет сокращен, но он никогда не станет короче заданного минимального интервала повторного просмотра. Если содержимое не изменилось, следующий интервал повторного просмотра будет увеличен, но он никогда не станет длиннее заданного максимального интервала повторного просмотра.

Опции посещения URL искателем Web

Вы можете указать искателю Web посетить определенные URL, как только он сможет это сделать.

Если необходимо выполнить обновление информации с конкретных сайтов пространства для искателя, можно открыть страницу *Слежение этого искателя*, выбрать опцию **Адреса URL для посещения или повторного посещения**, затем задать адреса или шаблоны URL для страниц, которые необходимо просмотреть или просмотреть повторно.

Например, если ваш отдел связи добавляет Web-страницу во внутреннюю сеть или пересматривает страницу для внесения важного изменения, вы можете задать URL новой или измененной страницы. Если искатель запущен, он помещает заданный URL в очередь для просмотра в следующий раз при проверке страниц, ожидающих посещения (обычно каждые десять минут). Если искатель не запущен, он ставит указанный URL в очередь для просмотра в следующий раз, когда будет запущен.

Убедитесь, что правила просмотра позволяют искателю посещать заданные вами URL. Искатель может посетить заданные вами URL раньше, чем при нормальной работе. Однако чтобы такое посещение произошло, должно существовать правило просмотра, разрешающее просмотр данного URL.

Новая найденная информация станет доступна для поиска, когда произойдет следующее полное построение индекса.

Как искатель Web обрабатывает произвольные страницы ошибок

Можно сконфигурировать искатель Web для обработки настроенных страниц, созданных администраторами сайтов в случаях, когда они не хотят возвращать стандартный код ошибки в ответ на требования получения определенных страниц.

Если сервер HTTP не может вернуть страницу, запрошенную клиентом, он в обычном режиме возвращает ответ, содержащий заголовок и код состояния. Код состояния указывает, что возникла ошибка (например, ошибка 404 - файл не найден). Администраторы некоторых сайтов создают специальные страницы, в которых ошибка объясняется более подробно, и конфигурируют сервер HTTP, чтобы он возвращал именно эти страницы. Такие настроенные страницы называют *произвольными страницами ошибок*.

Произвольные страницы ошибок могут искажать результаты работы искателя Web. Например, вместо заголовка, в котором указана ошибка, искатель получает произвольную страницу ошибки и код состояния 200, означающий успешную загрузку правильной страницы HTML. Но загруженная произвольная страница ошибки не связана с запрошенным URL, и вместо разных страниц возвращается почти одинаковое содержимое страницы ошибки. Такие посторонние по содержанию и почти идентичные страницы искажают индекс и результаты поиска.

Для обработки таких ситуаций можно задать при конфигурировании искателя Web опции для обработки произвольных страниц ошибок. Искателю Web нужна следующая информация о каждом сайте, с которого возвращаются произвольные страницы ошибок:

- Шаблон URL для сайта, для которого используются произвольные страницы ошибок. Этот шаблон URL состоит из протокола (HTTP или HTTPS), имени хоста, номера порта (если он нестандартный) и имени пути. Можно использовать звездочку (*) в качестве символа подстановки, заменяющего один или несколько символов до следующего обычного символа в шаблоне. Задаваемый шаблон регистрозависим.
- Шаблон заголовка для текста, соответствующего тегу <TITLE> документа HTML. При задании этого шаблона можно использовать звездочку (*) в качестве символа подстановки. Задаваемый шаблон регистрозависим.
- Шаблон содержимого для текста, соответствующего содержимому документа HTML. Это содержимое включает не только содержимое тега <BODY> (если тег <BODY> присутствует). Содержимым считается все, что следует в файле после

заголовка HTTP. При задании этого шаблона можно использовать звездочку (*) в качестве символа подстановки. Задаваемый шаблон регистрозависим.

- Целочисленное значение, представляющее код состояния для документов, соответствующих заданным шаблонам URL, заголовка и содержимого.

Пример

Следующая конфигурация указывает искателю Web, что он должен сравнить все правильные страницы HTML (код состояния 200), возвращенные с сайта `http://www.mysite.com/hr/*`, с заданными шаблонами заголовка и содержимого. Если тег `<TITLE>` страницы начинается с "Sorry, the page", а содержимое документа - любой текст (*), искатель обрабатывает такую страницу, как если бы он получил код состояния 404 (страница не найдена).

Таблица 3. Пример обработки страниц ошибок

Шаблон URL	Шаблон заголовка	Шаблон содержимого	Код состояния HTTP
<code>http://www.mysite.com/hr/*</code>	<code>Sorry, the page*</code>	<code>*</code>	404

Для одного сайта можно создать несколько записей для обработки разных кодов состояния. Для каждого кода состояния с данного сайта требуется отдельная запись в конфигурации искателя Web.

Использование символов подстановки

Шаблоны URL, заголовка и содержимого - это выражения с символами подстановки. Символ подстановки (звездочка) соответствует любым символам, идущим до следующего обычного символа. Например:

`*404` означает *любые символы*404

404: `*` означает 404: *любые символы*

`http://*.mysite.com/*` означает `http://любой хост.mysite.com/любой файл`

`*` означает *любые символы*

Влияние на производительность

Сконфигурировав опции для обработки страниц ошибок, вы увеличиваете время, затрачиваемое искателем на обработку, поскольку искателю придется проверить все успешно полученные страницы. Дополнительное время обработки требуется на проверку соответствия шаблонам, чтобы выяснить, что нужно вернуть - страницу или заменяющий код состояния.

Поддержка просмотра защищенных сайтов

Чтобы искатель Web мог обращаться к защищенному содержимому (например, документам, для доступа к которым требуется пароль), можно задать параметры аутентификации на консоли администратора поисковой системы предприятия.

Если для ограничения доступа к сайтам Web-сервер использует базовую аутентификацию HTTP или аутентификацию на основе форм HTML, можно задать в конфигурации искателя Web параметры аутентификации, чтобы этот искатель Web мог работать со страницами сайтов, защищенным паролями. Можно также задать опции для конфигурирования опознавательных файлов (cookie) вручную.

Сайты, защищенные базовой аутентификацией HTTP

Если для ограничения доступа к сайтам Web-сервер использует базовую аутентификацию HTTP, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.

Чтобы узнать, разрешено ли пользователю (или прикладной программе) обращаться к страницам на некотором сайте, многие Web-серверы используют для идентификации пользователя схему аутентификации клиента, которая называется базовой аутентификацией HTTP. Обычно это интерактивный процесс:

- Когда пользовательский агент HTTP (например, браузер) запрашивает страницу, защищенную базовой аутентификацией HTTP, Web-сервер возвращает код состояния 401, указывающий, что у запрашивающего нет полномочий доступа к запрошенной странице.
- Кроме того, Web-сервер предлагает передать ему параметры аутентификации, по которым сервер сможет определить, разрешен ли данному пользователю доступ к защищенному содержимому.
- Браузер выводит пользователю диалоговое окно, в котором предлагает ввести имя пользователя, пароль и другую информацию, необходимую для составления параметров аутентификации пользователя.
- Браузер кодирует параметры аутентификации и затем включает их в повторное требование защищенной страницы.
- Если эти параметры аутентификации успешно прошли проверку, Web-сервер возвращает код возврата 200 и содержимое запрошенной страницы.
- В последующие требования страниц с этого же Web-сервера обычно включаются те же параметры аутентификации; таким образом авторизованный пользователь может обращаться к дополнительному защищенному содержимому без дополнительного ввода параметров аутентификации.

Когда пользователь идентифицирован, Web-сервер и пользовательский агент HTTP обычно обмениваются опознавательными сервер файлами (*cookies*), позволяющими сохранять информацию о регистрации пользователя между требованиями HTTP.

Поскольку искатель Web работает не в интерактивном режиме, параметры аутентификации для просмотра защищенных паролями страниц нужно задать до начала операции просмотра. При создании искателя Web или редактировании пространства искателя задайте информацию о каждом защищенном сайте, который нужно просматривать.

Чтобы задать такую информацию, нужно работать в тесном взаимодействии с администраторами сайтов или Web-серверов, для защиты которых используется базовая аутентификация HTTP. Администраторы должны сообщить вам требования защиты для сайтов, на которых будет выполняться просмотр, включая всю информацию, используемую для аутентификации искателей Web и определения того, разрешен ли искателю доступ к защищенным страницам.

Если при создании собрания для него включена защита, можно при конфигурировании искателя задать маркеры защиты (такие как ID пользователей, ID групп или роли пользователей) для управления доступом к документам. Искатель Web связывает эти маркеры защиты с каждым документом, просматриваемым им в дереве файловой системы для заданного корневого URL. Эти маркеры используются в дополнение ко всем маркерам защиты уровня документа, сконфигурированным для всего пространства для искателя Web.

Имеет значение порядок, в котором заданы URL. Добавив информацию о защищенном паролем сайте, нужно задать его в соответствии с требуемым порядком обработки искателем. Сначала задайте более конкретные URL, а более общие поместите в конце списка. Когда искатель Web оценивает URL-кандидат, он использует данные аутентификации, которые заданы для первого URL в списке, соответствующего URL-кандидату.

Сайты, защищенные аутентификацией на основе форм

Если для ограничения доступа к сайтам Web-сервер использует формы HTML, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.

Чтобы узнать, разрешено ли пользователю (или прикладной программе) обращаться к страницам на некотором сайте, многие Web-серверы используют для идентификации пользователя формы HTML. Обычно это интерактивный процесс:

- Когда пользовательский агент HTTP (например, браузер) запрашивает страницу, защищенную аутентификацией на основе форм, Web-сервер проверяет, содержит ли требование опознавательные файлы (cookie), позволяющие идентифицировать пользователя.
- Если опознавательных файлов нет, Web-сервер предлагает пользователю ввести данные защиты в форму. Когда пользователь передаст форму, Web-сервер возвращает необходимые опознавательные файлы и разрешает выполнить требование получения защищенной паролем страницы.
- Для последующих требований, в которых есть необходимые опознавательные файлы, также будет разрешено выполнение. Авторизованный пользователь может обращаться к дополнительному защищенному содержимому, не заполняя форму и не вводя свои параметры аутентификации для каждого требования.

Поскольку искатель Web работает не в интерактивном режиме, параметры аутентификации для просмотра защищенных паролями страниц нужно задать до начала операции просмотра. При создании искателя Web или редактировании пространства искателя задайте информацию о каждом защищенном сайте, который нужно просматривать.

Задаваемые поля соответствуют полям, которые пользователь заполняет в интерактивном режиме по запросу браузера; кроме того, нужно задать все скрытые поля и поля статистики, необходимые для успешной регистрации.

Чтобы задать такую информацию, нужно работать в тесном взаимодействии с администраторами сайтов или Web-серверов, для защиты которых используется аутентификация на основе форм. Администраторы должны сообщить вам требования защиты для сайтов, на которых будет выполняться просмотр, включая всю информацию, используемую для аутентификации искателей Web и определения того, разрешен ли искателю доступ к защищенным страницам.

Имеет значение порядок, в котором заданы шаблоны URL. Добавив информацию о защищенном паролем сайте, нужно задать его в соответствии с требуемым порядком обработки искателем. В начало списка поместите наиболее конкретные шаблоны URL, а в конец списка - наиболее общие. Когда искатель Web оценивает URL-кандидат, он использует данные формы, которые заданы для первого шаблона URL в списке, соответствующего URL-кандидату.

Использование подключаемого модуля для просмотра защищенных узлов WebSphere Portal

Если на WebSphere Application Server включена глобальная защита и вы хотите просматривать искателем Web защищенные сайты WebSphere Portal, надо создать подключаемый модуль искателя для обработки требований аутентификации на основе форм. Обсуждение аутентификации на основе форм и пример программы, которую можно приспособить в качестве вашего собственного подключаемого модуля искателя Web, смотрите в документе <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0707nishitani>.

Подключаемый модуль необходим, если вы используете искатель Web для просмотра через WebSphere Portal любых сайтов, в том числе сайтов Workplace Web Content Management и сайтов Lotus Quickr.

Сайты с доступом через прокси-серверы

Если искателю Web не разрешен прямой доступ к сети, можно сконфигурировать искатель, чтобы он использовал для доступа к просматриваемому содержимому прокси-сервер HTTP.

Если у компьютера, где должен работать искатель Web, нет доступа к сети TCP/IP (или доступ разрешен только привилегированным процессам), можно сконфигурировать искатель Web, чтобы он использовал прокси-сервер HTTP. Прокси-сервер HTTP - это процесс, ожидающий требований HTTP на заданном порте заданного хоста. Прокси-сервер пересылает требования серверу Web и затем пересылает ответы сервера Web передавшему требованию клиенту (искателю Web). Прокси-сервер может работать на том же компьютере, что и искатель Web, или на другом компьютере.

Если искатель не использует прокси, требование для получения страницы с данным URL отправляется напрямую хосту. Если искатель работает через прокси, такое требование отправляется прокси-серверу.

При создании искателя Web или редактировании пространства для искателя задайте информацию о прокси-серверах, которые этот искатель Web использует просмотре страниц в домене прокси-сервера. Перед добавлением прокси-сервера в пространство для искателя получите имена доменов, обслуживаемых прокси-сервером, имя хоста или IP-адрес прокси-сервера и используемый прокси-сервером номер порта.

Если для прокси-сервера требуется аутентификация, получите также имя пользователя и пароль, которые искатель Web сможет использовать для доступа к страницам, обслуживаемым прокси-сервером. Искатель Web поддерживает только основную аутентификацию прокси HTTP, как описано в RFC2616 (<http://rfc.net/rfc2616.html>). Другие типы авторизации, в том числе Windows NT LAN Manager (NTLM), не поддерживаются.

Добавив прокси-сервер, надо выбрать его и разместить в ряду других прокси-серверов в соответствии с требуемым порядком их обработки искателем. Сначала задайте наиболее конкретные имена доменов, а более общие имена поместите в конец списка. Когда искатель Web проверяет некоторый URL, он использует данные прокси-сервера, заданные для первого домена в списке, соответствующего этому URL. (Адреса URL, которым не соответствует ни одно из правил прокси, считаются напрямую доступными для искателя.)

Управление опознавательными файлами (cookie)

Обычно управление опознавательными файлами (cookie) выполняется автоматически и не требует от администратора поисковой системы предприятия никаких действий. При необходимости можно вручную задать файлы cookie для сеанса просмотра искателем Web.

Опознавательные файлы (cookie) - это внешние маркеры, возвращаемые Web-сервером пользователю как часть заголовка ответа HTTP. Они значимы только для сгенерировавшего их Web-сервера и используются для поддержания состояния между требованиями HTTP. Например, при аутентификации клиента Web-сервер может вернуть опознавательный файл, позволяющий серверу определить, что аутентифицируемый пользователь уже зарегистрирован. Наличие опознавательных файлов позволяет пользователю запускать дополнительные требования для страниц на этом Web-сервере без повторной регистрации.

Искатель Web сохраняет опознавательные файлы, получаемые с Web-серверов, и использует их, пока существует экземпляр искателя. Искатель Web хранит опознавательные файлы в файле `cookies.ini`; искатель перезаписывает этот файл в конце каждого своего сеанса. При остановке искатель Web сохраняет все опознавательные файлы с неистекшим сроком действия и перезагружает их при запуске следующего сеанса.

При задании опознавательных файлов вручную сохраните их в отдельном файле, а затем, когда потребуется, объедините с опознавательными файлами в файле `cookies.ini`. Искатель не отбрасывает опознавательные файлы с неистекшим сроком действия, но на случай ошибки, при которой запись всех опознавательных файлов окажется невозможной, лучше застраховаться от потери файлов, заданных вручную. Эти файлы следует объединить с опознавательными файлами, поддерживаемыми искателем автоматически, перед запуском сеанса искателя.

Формат опознавательных файлов (cookie)

Опознавательные файлы, которые вы планируете объединить с файлом `cookies.ini` поисковой системы предприятия, должны иметь определенный формат.

- Каждая запись должна состоять из одной строки. Допускаются пустые строки и комментарии, но они не будут сохранены в файле `cookies.ini`.
- Каждая запись должна иметь следующий формат:

`CookieN(длина_cookie,длина_URL)текст_cookie,URL_проверки`

Cookie

Обязательное ключевое слово, указывающее начало записи.

Ключевое слово Cookie не может содержать пробелов, и на конце должна быть присоединена одна цифра - 0, 1 или 2. Эта цифра указывает тип cookie: `version-0` (Netscape), `version-1` (RFC2109) или `version-2` (RFC2965). Списки портов не поддерживаются в файлах cookie RFC2965.

длина_cookie

Число символов в соответствующем тексте.

длина_URL

Число символов в соответствующем URL проверки.

текст_cookie

Содержимое, которое будет отправлено на исходный сервер Web. Эта строка (представляющая собой правую часть директивы Set-Cookie в заголовке ответа HTTP) задает пару имя - значение и любое другое содержимое (например, путь,

параметры защиты и так далее), отправляемое в качестве cookie. После строки должен следовать разделитель - запятая (,).

URL_проверки

URL, по которому был найден этот опознавательный файл. Этот URL используется для определения, куда отправить запись cookie (например, имя домена и имя пути). URL проверки должен удовлетворять ограничениям защиты и конфиденциальности исходного сервера, установленным для опознавательных файлов.

Для удобства чтения следующий пример разбит на две строки; ваши записи надо задавать в одну строку:

```
Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSINFLSNL;path=/,  
https://www.ibm.com:443/help/solutions/
```

Конфигурирование опознавательных файлов (cookie) для искателя Web

Можно вручную задать опознавательные файлы (cookie) для сеанса искателя Web и объединить их с файлами, задаваемыми искателем Web.

Прежде чем приступить к работе...

Чтобы вручную сконфигурировать опознавательные файлы (cookie) для искателя Web, нужно быть администратором поисковой системы предприятия.

Процедура

Чтобы вручную сконфигурировать опознавательные файлы для искателя Web:

1. С консоли администратора поисковой системы предприятия запустите монитор нужного собрания и остановите искатель Web.
2. Зарегистрируйтесь на сервере искателя как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
3. Перейдите в каталог данных для нужного искателя, (*crawler_session_ID* здесь - ID, назначенный сеансу искателя поисковой системой предприятия). Например: `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/col_56092.WEB_88534`
4. Откройте файл `cookies.ini`, введите свои записи cookie в дополнение к существующим, затем сохраните файл и закройте его. Убедитесь, что ваши опознавательные файлы не переопределяют существующие.
5. С консоли администратора поисковой системы предприятия снова запустите остановленный искатель Web.

Конфигурирование глобального пространства для искателя Web

Для искателей Web можно сконфигурировать глобальное пространство, позволяющее лучше управлять удалением URL из индекса.

Для каждого искателя Web задается пространство, определяющее адреса URL, которые искатель должен (или не должен) просматривать. Если обнаруженный URL входит в пространство для искателя, он сохраняется (в базе данных) для последующего просмотра; если URL не входит в пространство для искателя, он отбрасывается. Если искатель запущен с пустой базой данных, во время его работы определение пространства для искателя и база данных остаются согласованными.

В некоторых случаях искатель может быть остановлен, а его пространство для искателя - сокращено (например, могут быть добавлены новые правила, запрещающие просмотр определенных страниц). При перезапуске искателя его определение пространства для искателя и база данных становятся несогласованными. База данных содержит адреса URL (для некоторых просмотр выполнен, для других - нет), которые не входят в новое, более узкое пространство для искателя.

Если у собрания только один искатель Web, этот искатель может восстановить согласованность, изменив коды состояния HTTP для этих URL на 760 (что означает, что они должны быть исключены) и затребовав удаление этих вновь исключенных страниц из индекса.

Если вы делите пространство для искателя между двумя или несколькими искателями Web (например, чтобы просмотр определенных страниц выполнялся чаще, чем остальных), каждый искатель Web будет работать с собственными таблицами базы данных (исходно пустыми) и будет просматривать свою часть пространства для искателей Web. Затем пространство для исходного искателя сокращается - из него удаляются части пространства, просмотр которых будут выполнять другие искатели. Проблемы возникают, когда исходный искатель пытается восстановить согласованность, удалив из индекса страницы, которые более не входят в его пространство для искателя. Поскольку эти страницы просматриваются теперь другими искателями, они должны оставаться в индексе.

Можно сконфигурировать пространство более высокого уровня - глобальное пространство для искателя, задающее адреса URL, которые исходный искатель не должен просматривать, но удалять их из индекса нельзя. Адреса URL, которые не входят ни в одно из пространств для искателей, по-прежнему будут отмечаться процессом поиска как исключенные и будут удаляться из индекса при повторном просмотре.

Глобальное пространство для искателя определяется в файле конфигурации `global.rules`, который должен находиться в каталоге конфигурации искателя (присутствие файла `global.rules` включает функцию глобального пространства для искателя). Если этот файл существует, он читается во время инициализации искателя. Если этот файл не существует, искатель работает с одноуровневым пространством искателя и удаляет из индекса документы, которые нужно удалить для согласованности определения пространства для этого искателя и базы данных.

Если глобальное пространство искателя определено, оно не влияет на то, как искатель вносит адреса URL в базу данных, но искатель будет затребовать удаление URL из индекса только в том случае, когда этот URL не входит ни в одно из пространств для искателей Web.

Синтаксис файла `global.rules` такой же, как и у файла `crawl.rules`, но он содержит только правила для имен доменов. Это ограничение позволяет разделять пространство для искателей между искателями только на основе DNS-имен хостов, но не на основе IP-адресов или шаблонов префиксов HTTP. Глобальное пространство для искателя не влияет на адреса URL, которые исключены из локального пространства для искателя правилами префиксов URL или IP-адресов (заданными в файле `crawl.rules`); такие URL будут по-прежнему исключаться.

Глобальное пространство для искателя используется только для того, чтобы предотвратить удаление из индекса адресов URL, которые исключены из пространства для данного искателя его локальным правилом доменов. Следующие правила применяются в указанной последовательности:

1. Если URL из базы данных искателя исключен в соответствии с локальным правилом префиксов или адресов, этому URL назначается код состояния 760 и он удаляется из индекса. Для этого URL не будет выполняться повторный просмотр.
2. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов и глобальное пространство для искателя не определено, этому URL назначается код состояния 760 и он удаляется из индекса. Для этого URL не будет выполняться повторный просмотр.
3. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов, но для него есть явное разрешающее правило в глобальном пространстве для искателя, этому URL назначается код состояния 761. Искатель не будет выполнять для этого URL повторный просмотр, но он не удаляется из индекса (подразумевается, что этот URL входит в локальное пространство для другого искателя).
4. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов, но для него нет явного разрешающего правила в глобальном пространстве для искателя, этому URL назначается код состояния 760 и он удаляется из индекса.

Поскольку искатель использует определение глобального пространства искателя только для того, чтобы определить, нужно ли удалять адреса URL, уже исключенные из локального пространства для этого искателя, то если в глобальном пространстве искателя нет правила для данного URL, для него по умолчанию запрещен просмотр.

Файл `global.rules` должен существовать в каталоге `master_config` каждого из искателей, работающих в данном глобальном пространстве. Нужно внимательно отредактировать все копии файла `global.rules` и файлы локальных пространств `crawl.rules`, чтобы обеспечить их взаимную согласованность.

Директивы `no-follow` и `no-index`

Можно улучшить качество поиска, указав для искателя Web директивы, управляющие следованием ссылкам на страницах и индексированием страниц.

На некоторых страницах Web есть директивы `no-follow` или `no-index`, указывающие роботам (например, искателю Web), что не надо переходить по ссылкам с этих страниц, не надо включать содержимое этих страниц в индекс или не надо делать ни того, ни другого.

Управление этими параметрами может улучшить качество поиска. Например, на некоторых страницах каталогов могут быть тысячи ссылок и никакой другой полезной информации; искатель должен просматривать эти страницы и переходить с них по ссылкам, но индексирование самих страниц каталога не имеет смысла.

В некоторых случаях нужно, чтобы искатель не переходил вниз по иерархической структуре страниц, но на требуемых конечных страницах есть ссылки и нет директив `no-follow`. Поскольку некоторые из этих страниц генерируются автоматически, у них нет владельцев, которые могли бы вставить нужные директивы.

Чтобы задать правила просмотра таких страниц, надо создать или отредактировать файл конфигурации `followindex.rules`. При задании правил в этом файле руководствуйтесь следующим:

- В конфигурируемых правилах указываются префиксы URL (нельзя задавать сайты Web по IP-адресу или имени хоста DNS).
- Префиксы URL могут содержать звездочки (*) в качестве символов подстановки, что позволяет разрешить или запретить несколько файлов с подобными URL.

- Порядок имеет значение (искатель применяет первое правило, соответствующее проверяемому URL).
- Правила, в явном виде разрешающие и запрещающие переход по ссылкам или индексирование, переопределяют правила, указанные в обрабатываемом документе.

Переопределение директив no-follow и no-index на Web-страницах

В файле конфигурации можно задать правила, управляющие переходами искателя Web по ссылкам на страницы и индексацией страниц, содержащих директивы no-follow или no-index.

Прежде чем приступить к работе...

Чтобы задать директивы no-follow и no-index для искателя Web, нужно быть администратором поисковой системы предприятия. Заданные вами директивы переопределяют директивы на страницах, обрабатываемых искателем.

Процедура

Чтобы переопределить директивы no-follow и no-index:

1. С консоли администратора поисковой системы предприятия запустите монитор нужного собрания и остановите искатель Web.
2. Зарегистрируйтесь на сервере искателя как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
3. Перейдите в каталог конфигурации для нужного искателя, (*crawler_session_ID* здесь - ID, назначенный сеансу искателя поисковой системой предприятия).
Например:
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/col_56092.WEB_88534
4. Создайте или отредактируйте файл с именем followindex.rules.
5. Введите правила для искателя в приведенном ниже формате, где *префикс_URL* - начальные символы сайтов, для которых вы хотите разрешить (allow) или запретить (forbid) переход (follow) или индексацию (index):

```
forbid follow префикс_URL
allow follow префикс_URL
forbid index префикс_URL
allow index префикс_URL
```
6. Сохраните и закройте файл.
7. С консоли администратора поисковой системы предприятия снова запустите остановленный искатель Web.

Конфигурирование дат, которые будет использовать искатель Web для обрабатываемых им документов

В файле конфигурации можно задать опцию, управляющую датой, которую искатель Web будет использовать в качестве даты обрабатываемого им документа.

Прежде чем приступить к работе...

По умолчанию искатель Web задает в качестве даты обрабатываемого им документа значение поля Дата. При желании конфигурацию искателя Web можно изменить, задав в качестве этой даты значение даты и времени в поле Время последнего изменения, возвращаемое Web-сервером.

Данные в поле Время последнего изменения могут оказаться доступными не для всех документов. Если для искателя сконфигурировано использование этого значения, а Web-сервер не возвращает данные Время последнего изменения для документа, искатель задает в качестве значения поля Дата обрабатываемого документа дату и время просмотра этого документа искателем.

Таблица 4. Как для искателя Web конфигурируется поле Дата для просмотренных документов

Конфигурация искателя	Данные, возвращаемые Web-сервером	Поле метаданных Дата для просмотренных искателем документов
Конфигурация по умолчанию	Web-сервер возвращает для просмотренного документа значение Время последнего изменения	Дата и время, полученные искателем
	Web-сервер не возвращает для просмотренного документа значение Время последнего изменения	Дата и время, полученные искателем
Сконфигурировано использование даты в поле Время последнего изменения	Web-сервер возвращает для просмотренного документа значение Время последнего изменения	Значение Время последнего изменения, возвращаемое Web-сервером
	Web-сервер не возвращает для просмотренного документа значение Время последнего изменения	Дата и время, полученные искателем

Чтобы задать, какую дату должен использовать искатель Web для обрабатываемых документов, нужно быть администратором поисковой системы предприятия.

Процедура

Чтобы создать конфигурацию искателя Web, в которой в качестве даты обрабатываемых им документов использовались бы дата и время в поле Время последнего изменения, возвращаемые Web-серверами:

1. Зарегистрируйтесь на сервере искателя как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Отредактируйте файл `crawl.properties` для искателя Web, который вы хотите сконфигурировать, (`ID_сеанса_искателя` - это ID, назначенный сеансу искателя поисковой системой предприятия).

`КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_сеанса_искателя/crawl.properties`

Например:

```
/home/esadmin/master_config/col_00112.WEB_23344/crawl.properties
C:\Program Files\IBM\esadmin\master_config\col_55667.WEB_78899\crawl.properties
```

3. Добавьте в файл следующую строку, а затем сохраните этот файл:
`which_date=LastModified`
4. С консоли администратора поисковой системы предприятия перезапустите искатель Web.

При возврате Web-сервером данных в поле Время последнего изменения для документов, просматриваемых этим искателем Web, искатель будет использовать эти возвращенные дату и время в качестве Даты обрабатываемых им документов.

Искатели Web Content Management

Чтобы включить документы IBM Workplace Web Content Management в собрание поисковой системы предприятия, нужно сконфигурировать искатель Web Content Management.

Конфигурирование сервера WebSphere Portal

Если установить Workplace Web Content Management на сервере WebSphere Portal Версии 6, искатель Web Content Management можно использовать для просмотра сайтов Web Content Management. Опции для просмотра искателем этих сайтов можно сконфигурировать отдельно от опций, задаваемых вами для сайтов портала, просматриваемых искателем WebSphere Portal.

Перед созданием искателя Web Content Management нужно выполнить действия по конфигурированию поисковой системы предприятия в WebSphere Portal. Для конфигурирования поисковой среды предприятия, на серверах поиска нужно запустить сценарий (wpb_install.sh - в AIX, Linux или Solaris и wpb_install.bat - в Windows), поставляемый с OmniFind Enterprise Edition.

Формат URL

Адреса URL для просмотра надо задавать в таком формате:

протокол_`http://имя_хоста_портала:номер_порта/префикс_портала`
/путь_нач_сервлета_поиска_WCM/searchseed?site=имя_сайта_WCM&lib=имя_библ_WCM

Ниже показан URL для сайта с путем установки по умолчанию Workplace Web Content Management на WebSphere Portal:

`http://portal.server.ibm.com:80/wps/wcmsearchseed/
searchseed?site=SiteTest01&lib=Web+Content`

Если имя сайта содержит пробелы, их надо заменить знаками плюс (+). Например, вместо Web Content следует указать Web+Content.

Конфигурирование - обзор

Искатель Web Content Management можно использовать для просмотра любого числа сайтов Web Content Management. При конфигурировании искателя нужно задать адреса URL для просматриваемых сайтов. После этого искатель будет загружать страницы, доступные по ссылкам на заданных сайтах.

Просматриваемые сайты должны быть доступны по одному и тому же ID администратора и паролю WebSphere Portal. Для обработки искателем сайтов, на которых используются другие параметры аутентификации, нужно сконфигурировать отдельный искатель Web Content Management.

Чтобы создать или изменить искатель Web Content Management, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве для искателя.
- Задать адреса URL для сайтов, с которыми должен работать искатель, и информацию, позволяющую искателю соединяться с этими сайтами.

При создании или редактировании искателя можно протестировать соединение искателя с просматриваемыми URL. Еще до запуска искателя сообщения покажут, сможет ли искатель обращаться к документам, с которыми он должен работать.

- Укажите опции защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют управлять доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

Важное замечание: Для поиска на защищенных страницах Web Content Management требования поиска нужно передать при помощи портлета поиска для поисковой системы предприятия из WebSphere Portal. У требований поисков, переданных из программы поиска примера ESSearchApplication, не будет правильных параметров аутентификации, и для них не удастся проверить полномочия пользователя для доступа к документам.

- Задать информацию, позволяющую искателю соединяться с прокси-сервером, если на сайтах Web Content Management для документов используется прокси-сервер.
- Если для защиты сервера WebSphere Portal и сайтов используется другой продукт (например, IBM Tivoli Access Manager WebSEAL или CA SiteMinder SSO Agent for PeopleSoft), задать полномочия единой регистрации, которые позволят искателю обращаться к документам на сервере.
- Задать информацию о файле склада ключей, чтобы искатель мог соединяться с сайтами Web Content Management при помощи протокола SSL (Secure Sockets Layer).
- Укажите язык и кодовую страницу документов, просматриваемых искателем.
- Задать опции просмотра искателем и поиска метаданных в Web Content Management.
- Задать расписания просмотра искателем сайтов Web Content Management.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Интеграция с WebSphere Portal” на стр. 347

Задачи, связанные с данной

“Конфигурирование поисковой системы предприятия в WebSphere Portal версии 6” на стр. 354

Искатели WebSphere Portal

Чтобы включить страницы с сайта IBM WebSphere Portal в собрание поисковой системы предприятия, нужно сконфигурировать искатель WebSphere Portal.

Конфигурирование сервера WebSphere Portal

Перед созданием искателя WebSphere Portal необходимо запустить сценарий для конфигурирования поисковой среды предприятия в WebSphere Portal. Для разных версий WebSphere Portal требуются разные сценарии. Эти сценарии устанавливаются на поисковых серверах при установке OmniFind Enterprise Edition.

- Для WebSphere Portal версии 5.1.0 или новее в операционных системах AIX, Linux и Solaris запускается сценарий `wp5_install.sh`, а в операционной системе Windows - сценарий `wp5_install.bat`.
- Для WebSphere Portal версии 6 или новее в операционных системах AIX, Linux и Solaris запускается сценарий `wp6_install.sh`, а в операционной системе Windows - сценарий `wp6_install.bat`.

Совет:

Подробные примеры конфигурирования защищенного искателя WebSphere Portal смотрите в сценарии для средней организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Конфигурирование - обзор

Можно использовать искатель WebSphere Portal для просмотра одного сайта WebSphere Portal. При конфигурировании искателя нужно задать URL нужного сайта портала. Затем искатель загружает портлеты и страницы со ссылками из URL заданного портала. Для обработки другого сайта портала создайте другой искатель.

Чтобы создать или изменить искатель WebSphere Portal, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех страниц на сайте.
- Задайте URL сайта портала, с которым должен работать искатель, и информацию, которая позволит искателю соединиться с этим сайтом. Поскольку такие URL могут быть большой длины и содержать символы, не входящие в кодовый набор ASCII, лучше скопировать URL с сервера WebSphere Portal и вставить на консоль администратора поисковой системы предприятия.

При создании или редактировании искателя можно проверить способность искателя связываться с URL, в которых он будет работать. Еще до запуска искателя сообщения покажут, сможет ли искатель обращаться к документам, с которыми он должен работать.

- Укажите опции защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют управлять доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

Важное замечание: При поиске на защищенных страницах WebSphere Portal необходимо запускать поиск через портлет поиска поисковой системы предприятия из WebSphere Portal. У поиска, запущенного из примера поисковой программы, ESSearchApplication, нет необходимых параметров аутентификации, и он не может проверить полномочия пользователя для доступа к документам.

- Укажите информацию, позволяющие искателю связываться с прокси-сервером, если сайт WebSphere Portal использует прокси-сервер для обслуживания страниц.
- Если для защиты сервера WebSphere Portal и сайтов используется другой продукт (например, IBM Tivoli Access Manager WebSEAL или CA SiteMinder SSO Agent for PeopleSoft), задать полномочия единой регистрации, которые позволят искателю обращаться к документам на сервере.
- Укажите информацию о файле склада ключей, чтобы искатель мог использовать для связи с сайтом WebSphere Portal протокол SSL.
- Укажите язык и кодовую страницу документов, просматриваемых искателем.
- Укажите опции для обработки искателем и поиска метаданных в документах WebSphere Portal.
- Укажите расписания для просмотра искателем документов WebSphere Portal.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Интеграция с WebSphere Portal” на стр. 347

Задачи, связанные с данной

“Копирование URL для просмотра искателем из WebSphere Portal”

“Конфигурирование поисковой системы предприятия в WebSphere Portal версии 6” на стр. 354

“Конфигурирование поисковой системы предприятия в WebSphere Portal версии 5.1” на стр. 349

Копирование URL для просмотра искателем из WebSphere Portal

Чтобы снизить вероятность ввода неправильного URL, при конфигурировании искателя WebSphere Portal или Web Content Management, скопируйте и вставьте в соответствующее поле URL сайта, на котором вы хотите запустить искатель.

Об этой задаче

При создании искателя WebSphere Portal или Web Content Management вы задаете URL сайта на сервере WebSphere Portal, с которым должен работать искатель. Поскольку эти URL - большой длины и обычно содержат символы, не входящие в кодовый набор ASCII, лучше, как описано ниже, скопировать URL с сервера WebSphere Portal и вставить на консоли администратора поисковой системы предприятия.

Процедура

Чтобы задать URL для просмотра искателем:

1. Когда вы готовы указать один или несколько URL для работы искателя на консоли администратора поисковой системы предприятия, убедитесь, что сервер WebSphere Portal запущен, и зарегистрируйтесь на WebSphere Portal как администратор.

2. Если вы используете WebSphere Portal версии 5.1, выполните следующие действия на сервере WebSphere Portal:
 - a. В верхнем правом углу выберите **Управление**.
 - b. Выберите **Параметры портала** в области навигации слева, затем выберите **Управление поисковой системой**.
 - c. На странице Управление собраниями поисковой системы в области Собрания поиска выберите **PortalCollection**. Можно выбрать и другое собрание, если оно есть.
 - d. В области Источники содержимого в собрании выберите **Добавить источник содержимого**.
 - e. В поле **Тип источника искателя** выберите **Сайт портала**. URL сайта появится в поле **Собирать документы по ссылкам из этого URL**.
 - f. Скопируйте URL в буфер обмена. Например, выделите URL и, удерживая клавишу Ctrl, нажмите клавишу Insert.
3. Если вы используете WebSphere Portal версии 6, выполните следующие действия на сервере WebSphere Portal:
 - a. В нижнем левом углу выберите **Управление**.
 - b. Выберите **Управление поиском** в области навигации слева.
 - c. На странице Управление поиском щелкните по ссылке **Собрания поиска**.
 - d. На странице Управление поиском в опциях службы поиска выберите **Служба поиска портала по умолчанию**.
 - e. В таблице Собрания для поиска выберите собрание **Содержимое портала**.
 - f. В таблице Источники содержимого справа щелкните по значку Редактировать (значок с изображением карандаша) рядом со строкой **Источник содержимого портала**.
 - g. В поле **Тип источника содержимого** выберите **Сайт портала**. URL сайта появится в поле **Собирать документы по ссылкам из этого URL**.
 - h. Скопируйте URL в буфер обмена. Например, выделите URL и, удерживая клавишу Ctrl, нажмите клавишу Insert.
4. Вернитесь на консоль администратора поисковой системы предприятия и вставьте скопированный URL в поле URL сайта.

Понятия, связанные с данным

“Искатели WebSphere Portal” на стр. 114

Искатели файловой системы Windows

Чтобы включить документы, хранимые в файловых системах Microsoft Windows, в собрание поисковой системы предприятия, нужно сконфигурировать искатель для файловой системы Файловая система Windows.

Искатель Файловая система Windows можно использовать для просмотра любого числа файловых систем Windows. При конфигурировании искателя нужно выбрать локальные и удаленные каталоги и подкаталоги, которые будет просматривать искатель.

При установке сервера искателя на системе AIX, Linux или Solaris этот сервер нельзя использовать для просмотра источников файловых систем Windows (искатель Файловая система Windows не выводится в списке доступных типов искателей).

Совет:

Подробные примеры конфигурирования защищенного искателя Файловая система Windows смотрите в сценарии для небольшой организации в руководстве IBM Redbook на странице IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Работа искателей в совместно используемых сетевых каталогах

Искатель Файловая система Windows обрабатывает документы в соответствии с разрешениями на чтение, указанными для администратора поисковой системы предприятия. Администратор - это учетная запись служб OmniFind Enterprise Edition.

Вы можете указать ID пользователя и пароль для каталогов, которые будут обрабатываться искателем. Однако эти ID пользователя и пароль используются только для связи с совместно используемыми сетевыми каталогами. Искатель обрабатывает файлы в соответствии с разрешениями на чтение, установленными для этого пользователя для совместно используемых сетевых каталогов, но не для локальных дисков.

Соединения с сетевыми каталогами не разрываются до перезапуска службы OmniFind Enterprise Edition. После установки соединения возможен доступ к каталогу с неправильными ID пользователя и паролем. Однако такое соединение разрешается только для сеансов обнаружения Файловая система Windows и работы искателя, управляемых поисковой системой предприятия. Чтобы избежать возможного нарушения защиты, проверьте правильность настройки авторизации для учетной записи администратора поисковой системы предприятия (под которой запущена служба OmniFind Enterprise Edition).

Чтобы в будущем избежать проблем с соединением с сетевым каталогом, указывайте одинаковые ID пользователя и пароль для одного и того же сетевого каталога. Если вы укажете неправильные ID пользователя и пароль и перезапустите службу OmniFind Enterprise Edition, работа искателя Файловая система Windows может быть нарушена, так как он будет пытаться связаться с каталогом с неправильными параметрами аутентификации. Работа искателя будет успешной, если сетевое соединение будет установлено другим искателем Файловая система Windows, использующим правильные ID пользователя и пароль.

Конфигурирование - обзор

Чтобы создать или изменить искатель Файловая система Windows, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подкаталогов в пространстве искателя.
- Составить расписания просмотра искателем файловых систем.
- Выбрать подкаталоги для просмотра искателем.

Можно задать, сколько уровней подкаталогов должен просматривать искатель. Для просмотра удаленных файловых систем нужно также задать ID пользователя и пароль, позволяющие искателю обращаться к данным.

- Задать опции, обеспечивающие доступность поиска для документов в подкаталогах. Например, можно исключить из пространства для искателя конкретные типы документов или задать ID пользователя и пароль, чтобы искатель мог обращаться к файлам в определенном каталоге.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности параметров аутентификации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры аутентификации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

Чтобы использовалась защита уровня документов, надо убедиться, что на сервере искателя правильно сконфигурирована информация об учетной записи пользователя и домена.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Понятия, связанные с данным

“Обеспечение защиты уровня документов для документов файловых систем Windows” на стр. 289

“Защищенный поиск для доверенных доменов Windows” на стр. 291

Конфигурирование поддержки для программ приема данных

Можно расширить поисковую систему предприятия, используя API приема данных для создания внешнего искателя. Пользовательские программы приема данных смогут добавлять данные в собрание или будут указывать искателю Web посещать определенные URL или повторно их просматривать.

Прежде чем приступить к работе...

Чтобы конфигурировать программы приема данных, надо быть администратором поисковой системы предприятия.

Важное замечание: Система приема данных в будущих выпусках не будет поддерживаться. Используйте при разработке клиентских программ для поисковой системы предприятия вместо API системы приема данных API поиска и индексирования (SI-API). Информация ниже приводится для тех пользователей, кто ранее создавал программы системы приема данных.

Об этой задаче

Клиентская программа приема данных позволяет выполнять поиск в таких типах источников данных, с которыми не могут работать искатели поисковой системы предприятия по умолчанию. Для использования программы приема данных надо сконфигурировать параметры аутентификации, обеспечивающие этой программе доступ к собраниям и их изменение.

Когда клиентская программа приема данных связывается с системой приема данных, она должна передать ID клиентской программы и пароль, а также ID изменяемого собрания. Эта информация должна совпадать с тем, что сконфигурировано для этой программы с консоли администратора.

Система приема данных запускается автоматически при запуске поисковой системы предприятия. Если вы изменили номер порта после того, как сконфигурировали программу с консоли администратора, необходимо перезапустить систему приема данных.

Процедура

Чтобы сконфигурировать программы приема данных:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Система приема данных нажмите кнопку **Конфигурировать программы приема данных**.
4. На странице Программы приема данных задайте число потоков, которые система приема данных может создавать для обработки требований от клиентских программ, и номер порта, на котором она принимает требования. Задайте также максимальное число документов на одно собрание, которые можно хранить во временном хранилище до начала их обработки синтаксическим анализатором.
5. Выберите **Добавить программу системы приема данных**, чтобы добавить информацию о клиентской программе.
6. На странице Добавить программу приема данных укажите информацию об аутентификации, позволяющую клиентским программам приема данных получать доступ к собраниям поисковой системы предприятия. ID клиентов приема данных должны быть уникальными в поисковой системе предприятия.
7. Выберите собрания, которые сможет обновлять программа приема данных:
 - Чтобы разрешить программе обновлять все собрания, выберите **Все собрания**.
 - Чтобы разрешить программе обновлять только указанные вами собрания, выберите **Отдельные собрания**.
Если выбрана эта опция, появляется список имен собраний. Включите переключатели **Выбрать** для всех собраний, которые сможет обновлять эта программа.
8. Нажмите кнопку **ОК**.
9. Если вы изменили номер порта системы приема данных или число документов, которые могут храниться во временном хранилище, перезапустите систему приема данных:
 - a. Выберите  **Слежение**, чтобы перейти в окно слежения за системой.
 - b. На странице Система приема данных нажмите кнопку  **Перезапустить**.

Задачи, связанные с данной

“Слежение за работой системы приема данных” на стр. 322

Пользовательские подключаемые модули искателей

При конфигурировании свойств для искателей можно задать класс Java, обеспечивающий управление доступом на уровне отдельных документов. Изменить индекс путем добавления, изменения или удаления метаданных и содержимого документа можно также при помощи класса Java. Написав подключаемый модуль, вы можете также расширить возможность искателя просматривать архивированные файлы.

Подключаемый модуль содержит класс Java, который вызывается для каждого документа, просматриваемого искателем. Классу Java передается идентификатор документа (URI) из индекса поисковой системы предприятия, маркеры защиты, метаданные и содержимое документа. Этот класс может возвращать новые или измененные маркеры защиты, метаданные и содержимое или же удалять маркеры защиты, метаданные и содержимое.

После первого просмотра искателем всех документов в пространстве искателя подключаемый модуль вызывается только для новых или измененных документов. Чтобы изменить маркеры защиты, метаданные или содержимое документов, включенных в индекс поисковой системы предприятия, но не измененных в исходном источнике данных, запустите полный просмотр искателем всех документов в пространстве искателя, а затем постройте заново главный индекс.

Подключаемые модули для обеспечения защиты

Защита на уровне документов реализуется при помощи связывания одного или нескольких элементов защиты (строка с разделителями-запятыми) с каждым документом, просматриваемым искателем. В качестве маркеров защиты обычно используются идентификаторы групп.

По умолчанию каждый документ связан с маркером общедоступности, делающим документ доступным каждому. Этот маркер можно заменить значением, заданным администратором, или значением, извлеченным из поля в просмотренном искателем документе.

Подключаемый модуль позволяет применить ваши собственные бизнес-правила для определения значения маркеров защиты для документов, просмотренных искателем. Маркеры защиты, связанные с документами, хранятся в индексе. При помощи маркеров защиты можно фильтровать соответствующие им документы и добиться возвращения в результатах поиска только тех документов, для просмотра которых у пользователя есть разрешение.

Подключаемые модули для добавления, изменения и удаления метаданных

Метаданные документов, такие как дата последнего изменения документа, создаются для всех документов, просмотренных искателем. Подключаемый модуль искателя позволяет применить ваши собственные бизнес-правила для определения значения метаданных, которые должны индексироваться для каждого документа.

Метаданные создаются как пара имя - значение. Пользователи могут выполнять поиск в метаданных при помощи свободного текстового запроса или запроса, в котором задается имя поля метаданных.

Подключаемые модули для добавления, изменения и удаления содержимого документа

Содержимое документа включает части документа с доступным для поиска содержимым, а также содержимым, которое может стать частью динамической сводки документа в результатах поиска. Подключаемый модуль искателя позволяет применить ваши собственные бизнес-правила для определения содержимого, которое должно индексироваться для каждого документа.

Подключаемые модули искателя Web

Интерфейсы прикладного программирования для искателя Web позволяют управлять тем, как искатель просматривает документы и как документы готовятся для анализа. Например, можно добавить поля в заголовок требования HTTP, который будет использоваться, когда искатель затребует документ. После того, как документ просмотрен искателем, но перед его синтаксическим и лексическим анализом можно изменить содержимое, маркеры защиты и метаданные. Можно также остановить отправку документа анализатору.

Обсуждение аутентификации на основе форм и пример программы, которую можно приспособить в качестве вашего собственного подключаемого модуля искателя Web, смотрите в документе <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0707nishitani>.

Подключаемые модули для архивированных файлов

Написав подключаемый модуль, вы можете расширить возможности искателя и дать ему возможность просмотра архивированных файлов в форматах, отличных от ZIP и TAR. Например, можно написать подключаемый модуль для просмотра документов в формате RAR.

Неизолированный режим

При конфигурировании искателя (кроме искателя Web) можно выбрать опцию его работы в неизолированном режиме. В этом режиме процесс подключаемого модуля выполняется внутри процесса искателя, что может повысить производительность.

Важное замечание: Если подключаемый модуль сталкивается с ошибкой в таком режиме, процесс искателя может быть прерван.

Понятия, связанные с данным



Подключаемые модули искателей

Поддержка просмотра архивных файлов искателем

Искатели поисковой системы предприятия могут извлекать файлы из архивных файлов (таких как ZIP или TAR), поэтому в архиве можно индексировать отдельные файлы и выполнять в них поиск.

Поддерживаемые форматы архивных файлов

Поддерживаются следующие форматы архивных файлов:

Таблица 5. Форматы архивных файлов, поддерживаемые искателями поисковой системы предприятия

Расширение файла	Тип MIME	Тип данных	Примечания
.zip, .ZIP	application/zip	zip	<ul style="list-style-type: none">Зависит от возможностей пакета <code>java.util.zip</code>Поддерживает метод сжатия "deflated" (метод 8):<ul style="list-style-type: none">Зашифрованные файлы не поддерживаютсяzip64 не поддерживается

Таблица 5. Форматы архивных файлов, поддерживаемые искомателями поисковой системы предприятия (продолжение)

Расширение файла	Тип MIME	Тип данных	Примечания
.tar	application/tar	tar	Поддерживаемые форматы tar: <ul style="list-style-type: none"> • GNU tar 1.13 • POSIX 1003.1-1998 (ustar) • POSIX 1003.1-2001 (pax)
.tar, .gz, .tgz	неприменимо	tgz	Зависит от возможностей пакета java.util.zip

Ограничения и рекомендации

Для файлов, извлекаемых из архивного файла, автоматическое определение кодовой страницы недоступно. При извлечении файлов искоматель использует параметр кодовой страницы, сконфигурированный для просмотра простых текстовых документов и документов неизвестных типов. Кодовая страница, которую искоматель должен использовать для простых текстовых документов и документов, кодовую страницу которых не удастся определить автоматически, задается при конфигурировании параметров языка и кодовой страницы для искомателя при помощи консоли управления поисковой системой предприятия.

Чтобы определить, когда требуется повторный просмотр файлов в архивном файле, искоматель использует дату изменения в данных заголовка записи архива для каждого файла. При слежении за искомателем выводится статистика для просмотренных документов, в том числе статистика для вставленных, измененных и удаленных документов, включая информацию о файлах, извлеченных из архивных файлов.

Чтобы включить просмотр искомателями файлов архивов других форматов (например, файлов LZH), нужно написать подключаемый модуль искомателя, после чего сконфигурировать искоматель для использования этого подключаемого модуля.

Перенастройка

Для возможности обработки архивных файлов искомателями, существовавшими до установки OmniFind Enterprise Edition Версии 8.4, требуется отредактировать пространство для искомателей. Убедитесь, что в параметрах для исключаемых типов MIME и исключаемых расширений файлов нет ссылок на файлы .zip, .tar, .tgz и .gz.

Важное замечание: Если изменить эти параметры для искомателя Content Edition или DB2 Content Manager, для применения внесенных изменений нужно выполнить повторный просмотр всех документов.

Форматы URI для индекса поисковой системы предприятия

Универсальный идентификатор ресурса (URI) каждого документа в индексе поисковой системы предприятия определяет тип искомателя, добавившего документ в собрание.

При конфигурировании категорий, областей действия и прямых ссылок для собрания можно задавать URI или шаблоны URI. Кроме того, URI задается при удалении документов из индекса или при просмотре подробной информации о состоянии определенного URI.

Чтобы определить URI или шаблоны URI для документа, выполните поиск по собранию. Для получения интересующего вас документа щелкните по его URI в результатах поиска. Можно скопировать URI найденного документа, чтобы использовать его на консоли администратора поисковой системы предприятия. Например, можно задать шаблон URI для автоматической связи документов, соответствующих этому шаблону URI, с определенной категорией поисковой системы предприятия.

Архивные файлы

Вот формат URI для документов, извлеченных из архивного файла (например, .zip или .tar) и затем обработанных искателем:

Исходный_URI(?|&)ArchiveEntry=*Имя_записи*(&ArchiveEntry=*Имя_записи*)

Параметры

Исходный_URI

Положение архивного файла в источнике данных.

Имя_записи

Имя в кодировке URL архивной записи в архивном файле.

Примеры

file:///d:/Archive1.zip

file:///d:/Archive1.zip?ArchiveEntry=Folder1/PowerPoint.ppt

file:///d:/Archive1.zip?ArchiveEntry=Folder2/Text.txt

Искатели Content Edition

Формат URI для документов, обработанных искателем Content Edition в режиме доступа к серверу:

vbr://*Имя_сервера/Системный_ID_репозитория/Постоянный_ID_репозитория*
/ID_элемента/ID_версии
/Тип_элемента/?[Page=Номер_страницы&] Свойства_JNDI

Формат URI для документов, обработанных искателем Content Edition в режиме прямого доступа:

vbr:///Системный_ID_репозитория/Постоянный_ID_репозитория
/ID_элемента/ID_версии
/Тип_элемента/[?Page=Номер_страницы]

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера WebSphere Information Integrator Content Edition.

Системный_ID_репозитория

Системный ID для репозитория.

Постоянный_ID_репозитория

Постоянный ID для репозитория.

ID_элемента

ID для элемента.

ID_версии

ID версии. Если ID версии пуст, используется последняя версия документа.

Тип_элемента

Тип элемента (CONTENT или FOLDER)).

Номер_страницы

Номер страницы.

Свойства_JNDI

Свойства JNDI для клиента прикладных программ J2EE. Свойства бывают двух видов:

java.naming.factory.initial

Имя класса сервера прикладных программ, используемого для создания хэндла EJB.

java.naming.provider.url

URL службы имен для сервера прикладных программ, используемого для получения хэндла EJB.

Примеры

Documentum:

```
vbr://vbrsrv.ibm.com/Documentum/c06b/094e827780000302//CONTENT/?  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

FileNet PanagonCS:

```
vbr://vbrsrv.ibm.com/PanagonCS/4a4c/003671066//CONTENT/?Page=1&  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

Искатели DB2

Формат URI для документов, обработанных искателем DB2:

```
db2://Имя_базы_данных/Имя_таблицы  
/Имя_столбца_идентиф_уникальности1/Значение_идентиф_уникальности1  
[/Имя_столбца_идентиф_уникальности2/Значение_идентиф_уникальности2/...  
/Имя_столбца_идентиф_уникальностиN/Значение_идентиф_уникальностиN]
```

Параметры:

Ко всем полям применяется кодировка URL.

Имя_базы_данных

Внутреннее имя базы данных или ее алиас.

Имя_таблицы

Имя таблицы назначения, включая имя схемы.

Имя_столбца_идентиф_уникальности1

Имя первого столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальности1

Значение первого столбца идентификатора уникальности.

Имя_столбца_идентиф_уникальностиN

Имя *n*-ного столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальностиN

Значение *n*-ного столбца идентификатора уникальности.

Примеры

Локальная, каталогизированная база данных:

```
db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20
```

Удаленная, некаталогизированная база данных:

```
db2://myserver.mycompany.com:50001/REMOTEDB/SCHEMA2.TABLE2/NAME/DAVID
```

Искатели DB2 Content Manager

Формат URI для документов, обработанных искателем DB2 Content Manager:

cm://Имя_сервера/Имя_типа_элементов/PID

Параметры

К параметру *PID* применяется кодировка URL.

Имя_сервера

Имя библиотечного сервера IBM DB2 Content Manager.

Имя_типа_элементов

Имя типа элементов назначения.

PID

Постоянный идентификатор DB2 Content Manager.

Пример

cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icm1sdb12+ITEMTYPE159+26+A1001001A
03F27B94411D1831718+A03F27B+94411D183171+14+1018

Искатели Domino Document Manager

Формат URI для документов, обработанных искателем Domino Document Manager:

dominodoc://Имя_сервера:Номер_порта/ID_реплики_базы_данных/Путь_и_имя_базы_данных
/Универсальный_ID_представления/Универсальный_ID_документа
/?AttNo=Номер_вложения&AttName=Имя_файла_вложения

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Domino Document Manager.

Номер_порта

Необязательно: Номер порта для сервера Domino Document Manager.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла для базы данных NSF документов на сервере назначения Domino Document Manager.

Универсальный_ID_представления

Универсальный ID представления, используемый для просмотра искателем документов Domino Document Manager.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просмотренном искателем.

Номер_вложения

Необязательно: Порядковый номер, начиная с нуля, для каждого вложения.

Имя_файла_вложения

Необязательно: Первоначальное имя файла вложения.

Примеры

Документ Domino Document Manager:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

Вложение документа:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc

Искатели Exchange Server

Формат URI для документов, обработанных искателем Exchange Server:
exchange://Путь_OWA[?useSSL=true]

Параметры

Путь_OWA

Путь Outlook Web Access (OWA) без указания протокола.

useSSL=true

Добавляется, если протокол первоначального пути OWA - HTTPS.

Примеры

Тело документа:

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML

Вложение документа:

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/
AttachedFile.doc

Включена возможность SSL:

exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML
?useSSL=true

Искатели База данных JDBC

Формат URI для документов, обработанных искателем База данных JDBC:

jdbc://URL_базы_данных/Имя_таблицы
/Имя_столбца_идентиф_уникальности1/Значение_идентиф_уникальности1
/[Имя_столбца_идентиф_уникальности2/Значение_идентиф_уникальности2
/.../Имя_столбца_идентиф_уникальностиN/Значение_идентиф_уникальностиN]

Параметры

Ко всем полям применяется кодировка URL.

URL_базы_данных

URL для базы данных.

Имя_таблицы

Имя таблицы назначения, включая имя схемы.

Имя_столбца_идентиф_уникальности1

Имя первого столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальности1

Значение первого столбца идентификатора уникальности.

Имя_столбца_идентиф_уникальностиN

Имя n-ного столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальностиN

Значение n-ного столбца идентификатора уникальности.

Примеры:

База данных DB2:

jdbc:db2://host01.svl.ibm.com:50000/SAMPLE/DB2INST1.ORG/DEPTNUMB/51

База данных Oracle:

jdbc:oracle:thin:@/host01.svl.ibm.com:1521:ora/SCOTT.EMP/EMPNO/7934

База данных MS SQL Server 2000:

jdbc:microsoft:sqlserver://host01.svl.ibm.com:1433;
DatabaseName=Northwind/dbo.Region/RegionID/100

База данных MS SQL Server 2005:

jdbc:sqlserver://host01.svl.ibm.com:1433;
DatabaseName=Northwind/dbo.Region/RegionID/100

Искатели Примечания

Формат URI для документов, обработанных искателем Примечания:

domino://Имя_сервера[:Номер_порта]/ID_реплики_базы_данных/Путь_и_имя_базы_данных
/[Универсальный_ID_представления]/Универсальный_ID_документа
[?AttNo=Номер_вложения&AttName=Имя_файла_вложения]

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Lotus Notes.

Номер_порта

Номер порта сервера Lotus Notes. Задавать номер порта необязательно.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла базы данных NSF на сервере назначения Lotus Notes.

Универсальный_ID_представления

Универсальный ID представления, определенный в удаленной базе данных. Этот ID задается только при выборе документа из представления или папки. Если не указано представление или папка для просмотра искателем (например, если задан просмотр искателем всех документов в базе данных), универсальный ID представления не задается.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просматриваемом искателем.

Номер_вложения

Порядковый номер, начиная от нуля, для каждого вложения. Указывать номер вложения необязательно.

Имя_файла_вложения

Первоначальное имя файла вложения. Указывать имя файла вложения необязательно.

Примеры

Документ, выбранный для просмотра по представлению или папке:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

Документ, не выбранный для просмотра по представлению или папке:

domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//
0205F44FA3F45A9049256DB20042D226

Вложение документа:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

Искатели QuickPlace

Формат URI для документов, обработанных искателем QuickPlace:

```
quickplace://Имя_сервера:Номер_порта/ID_реплики_базы_данных/Путь_и_имя_базы_данных  
/Универсальный_ID_представления/Универсальный_ID_документа  
/?AttNo=Номер_вложения&AttName=Имя_файла_вложения
```

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Lotus QuickPlace.

Номер_порта

Необязательно: Номер порта для сервера QuickPlace.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла для базы данных NSF документов на сервере назначения QuickPlace.

Универсальный_ID_представления

Универсальный ID представления, используемый для просмотра искателем документов QuickPlace.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просмотренном искателем.

Номер_вложения

Необязательно: Порядковый номер, начиная с нуля, для каждого вложения.

Имя_файла_вложения

Необязательно: Первоначальное имя файла вложения.

Примеры

Документ:

```
quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498
```

Вложение страницы:

```
quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498  
?AttNo=0&AttName==QPCons3.ppt
```

Искатели Список заполнения

Формат URI для документов, обработанных искателем Список заполнения:

```
seedlist://URL_страницы?pageID=ID_страницы[&useSSL;=true]
```

Параметры

Ко всем полям применяется кодировка URL.

URL_страницы

URL для документа (уникальный для каждого документа).

ID_страницы

Идентификатор объекта для документа.

useSSL В случае протокола HTTPS в URI добавляется &useSSL=true. В противном случае useSSL не задается.

Пример

Протокол HTTPS:

```
seedlist://quickrserver.ibm.com:10035/lotus/mydoc?uri=dm:bec6090046f1cd52bc5cfcb06e9f4550&verb=view&pageID=N1FSZUR1MkJQNjZSMDZQMUMwM1FPNjZCQzY2SUw2SUhPNk1RQ0M2Uk80Nk9PNjVCRUM2UUs2TDFDMA==&useSSL=true
```

Искатели Файловая система UNIX

Формат URI для документов, обработанных искателем Файловая система UNIX:

`file:///Имя_каталога/Имя_файла`

Параметры

Ко всем полям применяется кодировка URL.

Имя_каталога

Абсолютный путь каталога.

Имя_файла

Имя файла.

Пример

`file:///home/user/test.doc`

Искатели Web Content Management

Формат URI для документов WebSphere Content Management, обработанных искателем Web Content Management:

`wcm://URL_страницы?pageID=ID_страницы[&useSSL=true]`

Параметры

Ко всем полям применяется кодировка URL.

URL_страницы

URL для документа (уникальный для каждого документа).

ID_страницы

Идентификатор страницы.

useSSL В случае протокола HTTPS в URI добавляется useSSL=true. В противном случае useSSL не задается.

Примеры

Протокол HTTP:

```
wcm://wp6server.ibm.com:9081/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0
```

Протокол HTTPS:

```
wcm://wp6server.ibm.com:9444/wps/wcm/myconnect/Web+Content/Site01/SiteArea01/ContentTest01?pageID=6QReDeJ9DI3R0663E03Q06L1E2MR47MHOC3Q862RD6J0863B0GJS86J9E0&useSSL=true
```

Искатели WebSphere Portal: WebSphere Portal версии 5

Формат URI для документов WebSphere Portal версии 5, обработанных искателем WebSphere Portal:

```
wps://URL_страницы?portletDefID=ID_определения_портлета&portletID=ID_портлета  
&pageID=ID_страницы[&useSSL=true]
```

Параметры

Ко всем полям применяется кодировка URL.

URL_страницы

URL для документа (уникальный для каждого документа).

ID_опред_портлета

Идентификатор определения портлета.

ID_портлета

Идентификатор портлета.

ID_страницы

Идентификатор страницы.

useSSL В случае протокола HTTPS в URI добавляется useSSL=true. В противном случае useSSL не задается.

Примеры

Тело документа:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+  
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta  
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!  
?portletDefID=3_0_3S&pageID=6_0_6J
```

Примеры

Включена возможность SSL:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+  
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta  
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!  
?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

Искатели WebSphere Portal: WebSphere Portal версии 6

Формат URI для документов WebSphere Portal версии 6, обработанных искателем WebSphere Portal:

```
wp6://URL_страницы?portletURL=URL_портлета?portletDefID=ID_определения_портлета  
&pageID=ID_страницы[&useSSL=true]
```

Параметры

Ко всем полям применяется кодировка URL.

URL_страницы

URL для документа (уникальный для каждого документа).

URL_портлета

Уникальный URL для документа.

ID_опред_портлета

Идентификатор определения портлета.

ID_страницы

Идентификатор страницы.

useSSL В случае протокола HTTPS в URI добавляется useSSL=true. В противном случае useSSL не задается.

Примеры

Протокол HTTP:

```
wp6://wp6server.ibm.com:9081/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGhgYWjjobuhmaGDka0Bu5uHqFRpkaAAAEisaBQ!!  
?portletUrl=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyBykYGBmS2tcDoDE40MebyRiU-woYGFoaG7oZuhgZGjggbuZh6uXY4ihAYOMGQ7Z0CBDAwDcXPkM  
&portletDefID=6_24LS10811G1F102A0G6HEJUR10  
&pageID=3_24LS108110R6B02A0GGU94LN00
```

Протокол HTTPS:

```
wp6://wp6server.ibm.com:9444/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BkbGBgY1rDwMDEJ1XAwMggYxZvZOITbGhgYWjjobuhmaGDka0Bu5uHq5RtqaAAA50L41Q!!  
?portletUrl=/wps/myportal!/ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz9CP0os3gjE59gQwMLQ0P_IDMnAyNHA3f3UEsTD1NjA6B8pFm8AQ7gaEBAAdjIPrz6_Tzyc1P1C3IjDHQdFRUBTu-saA!!/d12/d0/Y2BiUZnBwMqsyBykYGBmS2tcDoDE40MebyRiU-woYGFoaG7oZuhgZGjggbuZh6uXWZihAYOMGQ5Z31BDAAwAk73P2  
&portletDefID=6_24LS10811G1F102A0G6HEJMU10  
&pageID=3_24LS108110R6B02A0GGU94T410&useSSL=true
```

Искатели Файловая система Windows

Форматы URI для документов, обработанных искателем Файловая система Windows:

file:///Имя_каталога/Имя_файла

file:///Имя_сетевой_папки/Имя_каталога/Имя_файла

Параметры

Ко всем полям применяется кодировка URL.

Имя_каталога

Абсолютный путь каталога.

Имя_файла

Имя файла.

Имя_сетевой_папки

Только для документов на удаленных серверах: имя общей папки в сети Windows.

Примеры

Локальная файловая система:

file:///d:/directory/test.doc

Сетевая файловая система:

file:///filesvr.ibm.com/directory/file.doc

Управление анализаторами

Чтобы усовершенствовать извлечение документов, можно задать опции синтаксического анализа документов и метаданных, лингвистического анализа и категоризации перед добавлением в индекс поисковой системы предприятия.

Доступные опции анализа содержимого документов, необходимые для оптимального извлечения информации:

Задание опций анализа документов на китайском, японском и корейском

Можно задать опции сегментации энграмм для анализа документов на китайском, японском и корейском языках. Можно также удалять символы новой строки из пустого пространства в документах на китайском и японском.

Включение собственного поиска XML

Если собрание содержит документы XML, можно разрешить поиск по этим документам с использованием собственного синтаксиса запроса XML, например, поиск фрагментов XPath и XML. Собственный поиск XML позволяет пользователям задавать запросы на основе взаимосвязей между различными элементами XML.

Конфигурирование категорий

Документы, использующие один и тот же шаблон URI или содержащие определенные слова, можно сгруппировать в категории. При выполнении пользователями поиска в собрании они могут ограничить результаты поиска документами, принадлежащими конкретным категориям.

Конфигурирование полей поиска

Можно отобразить элементы из документов XML на поля поиска в индексе. Можно также отобразить на поля поиска элементы метаданных из документов HTML. Создавая поля поиска в индексе поисковой системы предприятия, вы предоставляете пользователям возможность запроса определенных частей документов XML и HTML и повышаете точность результатов поиска.

Конфигурирование опций обработки текста

Если в поисковую систему предприятия добавлены пользовательские механизмы тестового анализа, можно выбрать механизм анализа для собрания. Связав с собранием механизм анализа, можно задать опции отображения содержимого для его лингвистического анализа и аннотирования. Кроме того, можно задать, как результаты анализа будут отображаться в индексе поисковой системы предприятия или в таблицах базы данных JDBC.

Отображение полей на классы значимости

Можно задать, чтобы в результатах поиска ранг документов, где соответствие терминам запроса найдено в определенных полях, был выше ранга других соответствующих терминам запроса документов. При отображении полей на классы значимости можно задать, какие поля содержимого и метаданных будут значимыми. Можно также сконфигурировать оценки, используемые каждым классом значимости для ранжирования документов.

Понятия, связанные с данным



Лингвистическая поддержка семантического поиска

-  Включение анализа текста в поисковую систему предприятия
 -  Основные понятия текстового анализа
 -  Прикладные программы семантического поиска
 -  Элементы запроса семантического поиска
- “Поддержка языков и кодовых страниц” на стр. 169
- “Определение форматов документов” на стр. 155

Работа с категориями

При помощи категорий можно сгруппировать документы с общими характеристиками и находить и получать только документы, соответствующие критерию, заданному для членов данной группы.

Если вы связали документы с категориями и ваши прикладные программы поиска поддерживают такую возможность, пользователи могут выполнять поиск в части собрания, указав имя категории. При выполнении поиска по всему собранию пользователи могут уточнить результаты поиска и просмотреть только документы, принадлежащие той же категории, что и один из документов результата.

При конфигурировании категории задаются правила, указывающие, что искатель должен связать документы с данной категорией. Документы можно группировать по использованию одинаковых шаблонов URI или по конкретному содержимому (например, документы, содержащие или не содержащие определенные слова или словосочетания).

Для создания и управления обработкой категорий используется консоль администратора поисковой системы предприятия:

- Тип категоризации выбирается при создании собрания. Возможные варианты: не использовать категории или использовать категории на основе правил.
- Если нужно, тип категоризации можно изменить при конфигурировании для собрания правил синтаксического анализа. Если тип категоризации изменен после просмотра документов искателем и их индексации, качество поиска понизится, пока вы снова не просмотрите документы искателем и не перестроите главный индекс.
- Если выбрать использование категорий на основе правил, при помощи консоли администратора можно управлять деревом категорий, категориями и правилами категорий. Если категории или правила категоризации изменены после просмотра документов искателем и их индексации, качество поиска понизится, пока искатель не просмотрит эти документы снова и не будет выполнено полное построение индекса.

Категории на основе правил

Вы можете сконфигурировать правила для управления связями документов с категориями в собрании поисковой системы предприятия.

Можно создать правила категорий для собраний, создаваемых в поисковой системе предприятия, а также для категорий, перенастраиваемых из собраний IBM WebSphere Portal. Чтобы сконфигурировать правила для категоризации документов, при создании собрания или при задании для него опций синтаксического анализа нужно указать, что вы хотите использовать категории на основе правил.

Синтаксический анализатор использует задаваемые вами правила для связи документов с одной или несколькими категориями:

- Если для документа выполняется как минимум одно правило в категории, синтаксический анализатор связывает данный документ с этой категорией.
- Если для документа выполняется как минимум одно правило в нескольких категориях, синтаксический анализатор связывает данный документ со всеми этими категориями.
- Если для документа не выполняется ни одно из правил для категории, синтаксический анализатор не связывает документ с категорией. Пользователи могут найти этот документ и получить его, выполнив поиск в собрании, но при помощи поиска в категории получить этот документ невозможно.

Если вы - администратор дерева категорий (или таксономии) для собрания, вы решаете, в какое место иерархии категорий добавить новую категорию. При помощи дерева категорий вы также выбираете категорию, которую нужно отредактировать, а затем добавляете правила для категоризации документов, удаляете правила или изменяете содержимое отдельных правил.

Конфигурируя правило для категоризации документов, вы выбираете, будет ли в поисковой системе предприятия при определении принадлежности документа данной категории использоваться идентификатор URI документа или содержимого в этом документе.

Правила шаблонов URI

Правило URI применяется к URI документа. Вы задаете неполный URI (шаблон), и правило выполняется для тех документов, идентификаторы URI которых соответствуют заданному шаблону.

Например, если задать текст правила `/hr/`, для первого из следующих URI правило выполняется, а для второго - не выполняется:

```
file:///corporate/hr/medicalform.doc  
http://company.com/human_resources/medicalform.htm
```

Поскольку все URI воспринимаются как шаблоны, система игнорирует все знаки звездочки, указанные в качестве символов подстановки в начале или конце шаблона. Например, `*/hr/*` и `/hr/` соответствуют одному и тому же набору URI.

В правилах шаблонов URI регистр символов не учитывается. Если URI содержит пробелы, шаблон URI должен соответствовать правилам кодирования URI поисковой системы предприятия. В следующем примере показаны правильные и неправильные способы задания URI для пути в файловой системе Windows:

```
Неправильный URI: file:///c:/program files/  
Правильный URI: file:///c:/program+files/
```

Правила содержимого документа

Правила содержимого документа задаются в таком же формате, как запросы. Если документ соответствует запросу, правило для него выполняется. При конфигурировании правила нужно задать слова и фразы, которые должны содержать (или не содержать) документы, а также выбрать язык для применения правил поиска морфологических форм слов.

Например, следующее правило задает, что если документ содержит слово `hr` или фразу `human resources`, для него правило выполняется:

```
hr "human resources"
```

Другой пример: следующее правило задает, что если заголовок документа содержит слово `"health"`, но не содержит словосочетание `"employee benefits"`, для него правило выполняется:

```
+title:health -title:"employee benefits"
```

Правила содержимого подвергаются такой же лингвистической нормализации, как и запросы SI-API (Search and Index API - API поиска и индексации). Однако синтаксис правил содержимого поддерживает лишь подмножество операций, доступных в синтаксисе запросов SI-API. В правилах содержимого разрешаются только следующие операции запроса:

- +** Знак плюс перед термином означает, что этот термин должен присутствовать в документе.
- Знак минус перед термином означает, что этот термин должен отсутствовать в документе.
- " "** Группа терминов, заключенная в кавычки, означает, что в документе должно присутствовать это словосочетание в точности.

имя_поля:

Имя поля перед термином или словосочетанием указывает, что термин или словосочетание должны (или не должны) содержаться в заданном поле документа.

Поддерживаются все поля содержимого и метаданных, сконфигурированные как поля с возможностью поиска в собрании. Не поддерживаются следующие ключевые слова и типы полей SI-API:

- `site:`
- `url:`
- `link:`
- `docID:`
- `samegroupas:`
- параметрические поля
- маркеры защиты
- атрибуты (например, `$source`, `$language`, `$doctype` и так далее)

Деревья категорий

С помощью дерева категорий можно просмотреть все категории на основе правил в собрании. Дерево категорий позволяет создать категории, удалить категории и отредактировать правила, связывающие документы с категориями.

Дерево категорий (другое название - таксономия) организовано иерархически. Дерево начинается с корневой категории, от которой отходят все остальные категории. Категорий и подкатегорий может быть произвольное количество, что обеспечит пользователям широту выбора при просмотре и получении документов.

Например, если документ пропускается через правила для нескольких категорий, он связывается со всеми этими категориями. Когда пользователи выполняют поиск в категории или просматривают в результатах поиска документы, принадлежащие

определенной категории, принадлежность документа нескольким категориям повышает вероятность нахождения его пользователями.

Управляя деревом категорий, можно управлять числом категорий, которым принадлежат документы, вкладывая новые категории в существующие. При создании категории вы задаете, должна ли она быть создана на корневом уровне или как подкатегория другой категории. При помощи дерева категорий можно также удалить категории из собрания или изменить правила связи документов с категориями. При редактировании категории ее можно переименовать, добавить или удалить правила категоризации или изменить содержимое отдельных правил.

При управлении деревом категорий руководствуйтесь следующим описанием поведения поиска и просмотра:

- Если пользователь выполняет поиск в категории высокого уровня, поиск документов, соответствующих критерию поиска, выполняется в этой категории и всех ее подкатегориях. Если пользователь выполняет поиск в категории, не содержащей дополнительных подкатегорий, поиск выполняется только в этой категории.
- Если пользователь просматривает результаты поиска и выбирает опцию для просмотра документов, принадлежащих конкретной категории, выводятся документы только этой категории. В результатах поиска выводятся также имена всех подкатегорий, что позволяет пользователю перейти с одной категории на другую и одновременно просмотреть поднаборы документов.

Выбор типа категоризации

При выборе типа категоризации вы задаете способ связи документов с категориями, используемый в собрании.

Прежде чем приступать к работе...

Изменить тип категоризации может администратор поисковой системы предприятия или администратор изменяемого собрания.

Об этой задаче

Тип категоризации задается при создании собрания. В случае необходимости можно изменить способ категоризации документов для собрания. Можно использовать категории на основе правил, конфигурируемые специально для собрания, или не использовать никаких категорий.

Важное замечание: Если изменить тип категоризации после просмотра искателем данных и создания индекса для собрания, индекс станет несогласованным. Чтобы результаты поиска были точными, повторно запустите искатели для всех документов в собрании и постройте заново полный индекс.

Процедура

Чтобы выбрать тип категоризации:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Выбрать дерево категоризации**.
2. На странице Выбрать тип категоризации выберите одну из следующих опций:
Нет Выберите эту опцию, если не хотите категоризировать документы в этом собрании.

На основе правил

Выберите эту опцию, если вы хотите категоризировать документы в соответствии с правилами, сконфигурированными специально для этого собрания.

Конфигурирование категорий

Для собрания можно создать сколько угодно категорий, в каждой из которых может быть сколько угодно правил. Правила определяют, какие документы автоматически связываются с категорией.

Прежде чем приступить к работе...

Конфигурировать категории может администратор поисковой системы предприятия или администратор собрания, к которому принадлежат эти категории.

В качестве типа категоризации должна быть выбрана опция использования категорий на основе правил.

Чтобы просмотреть примеры задания правил связывания документов с категориями, нажмите кнопку **Справка** во время создания или редактирования категории.

Об этой задаче

Если в прикладных программах поиска включена поддержка категорий, пользователи могут выполнять поиск в подмножестве собрания, задав имя категории. Кроме того, пользователи могут выбрать категорию в результатах поиска и просматривать только документы, относящиеся к выбранной категории.

Важное замечание: Если изменить категории или правила категорий после просмотра искателем данных и создания индекса для собрания, индекс станет несогласованным. Чтобы результаты поиска были точными, повторно запустите искатели для всех документов в собрании и постройте снова полный индекс.

Процедура

Чтобы сконфигурировать категорию:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать дерево категорий**.
2. На странице Дерево категорий выберите место в дереве, куда хотите добавить категорию, затем выберите **Создать категорию**.

Если выбрать корень дерева, новая категория будет создана на уровне корня.

Если выбрать имя категории, новая категория будет вложена в дерево категорий под выбранной категорией.

Откроется мастер, в котором надо задать правила для связывания документов с новой категорией:

- a. На странице Создать категорию введите описательное имя для категории и нажмите кнопку **Далее**.
- b. На странице Создать правила категории нажмите кнопку **Добавить правило**.
- c. На странице Создать правило категории введите в поле **Имя правила** уникальное имя для этого правила. Имя должно быть уникальным для всех категорий в собрании.
- d. Задайте правило, которое нужно использовать для связывания документов с этой категорией, и нажмите кнопку **ОК**.

- Чтобы при определении принадлежности документа к этой категории использовать URI этого документа, выберите **Шаблон URI** и задайте шаблон URI.

Если указанный вами текст существует в URI, документ будет связан с категорией.

Например: `file:///c:/program+files/finance`

- Чтобы определять, принадлежит ли документ к категории, при помощи запроса доступного для поиска содержимого, выберите **Содержимое документа**, выберите язык документов и задайте слова и словосочетания, которые должны или не должны входить в содержимое документов.

Формат выражений для правил такой же, как и формат запроса, при этом разрешаются только операции запроса включение (+), исключение (-), словосочетание (" ") и имя поля (*field_name*:). Правила содержимого не поддерживают сегментацию энграмм.

Если в документе содержатся (или отсутствуют) указанные вами слова, документ включается в данную категорию.

Например: `+финансы -бухгалтерия +title:"финансовый год"`

- е. Нажмите кнопку **Готово**.

Новая категория появится на странице Дерево категорий вместе с другими категориями, относящимися к этому собранию.

Ссылки, связанные с данной

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

Работа с полями поиска XML

Чтобы пользователи могли выполнять поиск в определенных частях документов XML, отобразите элементы XML на поля поиска.

Для отображения элементов XML на поля поиска служит консоль администратора поисковой системы предприятия.

Обычно в документе XML индексируются все текстовые данные. Отображение элементов XML на поля поиска позволяет использовать структурную информацию документов XML для поддержки более конкретных запросов. Например, отобразив имя элемента XML на имя поля, можно обеспечить возможность поиска данных в элементе XML с указанием имени поля и их возврата в результатах поиска. Отображение можно сделать более точным, задав атрибуты элемента XML (и значения этих атрибутов) в качестве критериев, чтобы элемент стал полем поиска.

поля поиска XML

Поля поиска XML позволяют пользователям делать запросы по определенным частям документов XML.

Популярность документов в формате XML растет, так как они одновременно содержат полуструктурированный и неструктурированный текст. Структура XML является инкапсулированной и использует контекст, в явном виде заданный элементами XML, между которыми заключен текст. Например, имя автора может быть указано в следующем виде:

```
<author>John Smith</author>
```

В этом контексте текст John Smith определяет автора документа XML.

Поисковую систему предприятия позволяет связывать (отображать) элементы XML для поиска с заданием имен полей. При конфигурировании опций синтаксического анализа для собрания вы задаете отображение элементов XML на имена полей поиска. Задание отображения элементов XML на поля поиска позволяет пользователям выполнять поиск значений этих элементов, указав в запросе имена полей. Поисковые запросы для именованных полей могут дать более точные результаты поиска, чем запросы в виде свободного текста, которые выполняют поиск по всему содержимому документов.

Например, если в собрании есть документы XML, и задано, что элементы <title> и <author> будут помечены в индексе как поля поиска, пользователи смогут делать запросы непосредственно к этим элементам. При поиске author:Smith будут найдены документы XML, содержащие Smith в элементах, отображенных в поле с именем author.

Другой пример - пусть элемент XML с именем <summary> содержит информацию, которую полезно вывести в результатах поиска. Если отобразить элемент <summary> на поле поиска и задать вывод значения этого элемента в результатах поиска, содержимое этого элемента войдет в состав документа результатов.

При отображении элемента XML в имя поля текст внутри элемента становится доступен для поиска с указанием имени поля. Но если элемент XML включает атрибуты, их значения не индексируются и для непосредственного поиска недоступны. Чтобы запрашивать значения атрибутов, надо сконфигурировать опции синтаксического анализа на консоли администратора поисковой системы предприятия и включить собственный поиск XML. Когда вы включите собственный поиск XML, структура документа XML станет доступна в индексе, и в запросах к этому документу можно будет задавать ограничения XPath. Например:

```
@xml:xp::'//elementName[@attributeName="attributeValue"]
```

Понятия, связанные с данным



Элемент запроса семантического поиска

Задачи, связанные с данной

“Включение поддержки собственного поиска XML” на стр. 153

Отображение элементов XML на поля поиска

При отображении элемента XML в поле поиска вы определяете, в каких элементах XML могут выполнять поиск пользователи, задав в запросе имя поля.

Прежде чем приступать к работе...

Отображать элементы XML на поля поиска может только администратор поисковой системы предприятия или администратор собрания, к которому относятся документы XML.

Ограничения

Для одного собрания может существовать несколько отображений полей XML, но только одно отображение корневого элемента. Для корневого элемента документа XML отображения применяются соответствующим образом.

Об этой задаче

Когда вы создаете отображение поля XML или добавляете, изменяете или удаляете поля в существующем отображении поля XML, внесенные изменения вступают в силу

только после перезапуска анализатора. Новые и измененные отображения применяются к новым данным, проанализированным после перезапуска анализатора, и не влияют на данные, уже внесенные в индекс. Чтобы изменить документы, уже внесенные в индекс, нужно выполнить их повторный просмотр искателем и индексацию.

В этой задаче проиллюстрировано, как отобразить записи о персонале и позволить пользователям выполнять запросы непосредственно по некоторым элементам, на примере следующего документа XML.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<personnel>
  <personnelrecord>
    <phone>5555</phone>
    <email>joe@us.ibm.com</email>
    <jobroles>Manager, architect
      <jobrole>Managing Search Development Group</jobrole>
      <jobrole>Architecting Search Technology</jobrole>
    </jobroles>
    <location>New York</location>
    <section id="expertise">
      <text>Linguistics</text>
    </section>
  </personnelrecord>
</personnel>
```

Процедура

Чтобы отобразить элементы XML в этом примере на поля поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Отобразить элементы XML на поля**.
2. На странице Отображения полей XML выберите **Создать отображение XML**. Откроется страница Создать отображение поля XML.
3. В поле **Имя корневого элемента XML** введите имя корневого элемента: `personnel`.
Заданное здесь имя должно в точности совпадать с именем корневого элемента в документах XML, в которых будет выполняться поиск. При анализе и индексировании документов XML поисковая система предприятия выбирает используемое отображение в соответствии с именем корневого элемента.
4. В поле **Имя отображения XML** введите имя для этого набора правил для отображения полей XML.
После создания набора правил отображения XML это имя выводится на странице Отображения полей XML, и его надо выбирать для добавления, удаления или изменения правил отображения.
5. Отобразите элемент XML `jobrole` на поле поиска `jobrole`:
 - a. В поле **Имя поля** введите `jobrole`.

Совет: Для получения информации о символах и метасимволах ASCII, не поддерживаемых в именах полей, нажмите кнопку **Справка**.

- b. В поле **Имя элемента XML** введите `jobrole`.
- c. Чтобы пользователи могли вызывать запросы по полю `jobrole`, включите переключатель **Поиск по полю**.
- d. Чтобы документы соответствовали критерию поиска только в случае, если терминам поиска соответствует все значение поля `jobrole` (то есть когда в этом поле нет никаких других слов), включите переключатель **Полное соответствие**.

- e. Чтобы пользователи могли сортировать результаты поиска по значениям в поле `jobrole`, включите переключатель **Возможность сортировки**.
Это поле выводится, как сортируемое, только если программа поиска поддерживает эту возможность. Пример программы поиска не содержит в списке доступных для сортировки полей те поля, на которые отображаются элементы XML.
 - f. Чтобы пользователи могли просмотреть значения поля `jobrole` в результатах поиска, включите переключатель **Результаты поиска**.
6. Отобразите на это же поле поиска элемент XML `jobroles`:
- a. Нажмите кнопку **Добавить поле**, чтобы добавить в список правил отображения полей пустую строку.
 - b. В поле **Имя поля** введите `jobrole`.
 - c. В поле **Имя элемента XML** введите `jobroles`.
- Совет:** Имена элементов XML не обязаны совпадать с именами полей поиска, и на одно и то же поле поиска можно отобразить несколько элементов XML.
- d. Чтобы пользователи могли запрашивать поле `jobrole`, задавать, что значение этого поля должно полностью соответствовать терминам поиска, и просматривать поле в результатах поиска, включите соответствующие переключатели.
7. Отобразите элемент XML `section` с атрибутом под именем `ID`, содержащим значение `expertise`, на поле поиска `expertise`:
- a. Нажмите кнопку **Добавить поле**, чтобы добавить в список правил отображения полей пустую строку.
 - b. В поле **Имя поля** введите `expertise`.
 - c. В поле **Имя элемента XML** введите `section`.
 - d. В поле **Имя атрибута XML** введите `id`.
 - e. В поле **Значение атрибута XML** введите `expertise`.
 - f. Чтобы пользователи могли вызывать запросы по полю `expertise`, включите переключатель **Поиск по полю**.
 - g. Чтобы документы соответствовали критерию поиска только в случае, если терминам поиска соответствует все значение поля `expertise` (то есть когда в этом поле нет никаких других слов), включите переключатель **Полное соответствие**.
 - h. Чтобы пользователи могли сортировать результаты поиска по значениям в поле `expertise`, если программа поиска поддерживает эту возможность, включите переключатель **Возможность сортировки**.
 - i. Чтобы пользователи могли просмотреть значения поля `expertise` в результатах поиска, включите переключатель **Результаты поиска**.

Примеры:

Чтобы найти всех сотрудников организации, работающих над поисковыми продуктами, задайте следующий запрос:

```
jobrole:search
```

Чтобы найти всех сотрудников организации, компетентных в области лингвистики, задайте следующий запрос:

```
expertise:linguistics
```

Работа с полями поиска HTML

Чтобы пользователи могли выполнять поиск в особых разделах метаданных документов HTML, надо отобразить элементы метаданных HTML на поля поиска.

Для отображения элементов метаданных HTML на поля поиска используйте консоль администратора поисковой системы предприятия.

Отображение элементов метаданных HTML на поля поиска позволяет пользователям искать документы HTML при помощи более точных запросов.

Поля поиска HTML

Поля поиска HTML позволяют пользователям выполнять запросы по атрибутам документов HTML.

Элементы метаданных в документах HTML похожи на атрибуты документов тем, что они предоставляют информацию о документе, о его форматировании и о том, какой доступ к нему разрешен из Web. Например:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta name="copyright" content="(C) Copyright IBM Corporation 2005" />
<meta name="content.owner" content="(C) Copyright IBM Corporation 2005" />
<meta name="security" content="public" />
<meta name="abstract" content="This topic describes an IBM product." />
<meta name="format" content="XHTML" />
```

Поисковая система предприятия может связывать (отображать) имена элементов метаданных HTML с именами полей поиска. При конфигурировании опций анализа для собрания можно задать отображение элементов метаданных HTML на имена полей поиска. Отображение элементов метаданных HTML на поля поиска позволяет пользователям находить документы с этими элементами, указывая в запросах имена полей поиска. Поисковые запросы для конкретных полей могут дать более точные результаты поиска, чем запросы в виде свободного текста, которые выполняют поиск по всему содержимому документов.

Например, если в собрании есть документы HTML и включено индексирование элементов метаданных copyright и abstract в качестве полей поиска, пользователи могут выполнять запросы по этим конкретным элементам. При поиске copyright:IBM будут найдены документы HTML, которые содержат строку IBM в метаданных copyright.

Отображая элементы метаданных HTML на поля поиска, вы указываете, хотите ли вы отобразить все элементы метаданных HTML, только элементы, принадлежащие в набору элементов метаданных Dublin Core, или только указанные вами элементы метаданных HTML. Описание элементов в наборе элементов Dublin Core смотрите на сайте Dublin Core Metadata Initiative:

<http://dublincore.org/documents/dcmi-terms/#H2>

Если вы выбрали создание отображений для всех элементов метаданных HTML или для всех элементов метаданных Dublin Core, имя поля по умолчанию совпадает с именем элемента метаданных. Можно изменить имя поля поиска по умолчанию и указать другие опции поиска для определенных элементов, добавляя элементы в список элементов, которые вы хотите сконфигурировать индивидуально.

Отображение элементов метаданных HTML на поля поиска

При отображении элемента метаданных HTML в поле поиска вы определяете, в каких элементах метаданных HTML могут выполняться поиск пользователи, задав в запросе имя поля.

Прежде чем приступать к работе...

Чтобы отображать элементы метаданных HTML на поля поиска, надо быть администратором поисковой системы предприятия или администратором того собрания, к которому принадлежит этот документ в формате HTML.

Об этой задаче

Когда вы создаете отображение поля HTML или добавляете, изменяете или удаляете поля в существующем отображении поля HTML, изменения вступают в силу только после перезапуска анализатора. Новые и измененные отображения применяются к новым данным, проанализированным после перезапуска анализатора, и не влияют на данные, уже внесенные в индекс. Чтобы изменить документы, уже внесенные в индекс, нужно выполнить их повторный просмотр искателем и индексацию.

Процедура

Чтобы отобразить элементы метаданных HTML на поля поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Отобразить метаданные HTML на поля**.
2. На странице Отображения полей HTML задайте, какие элементы метаданных HTML надо отобразить на поля поиска:
 - При отображении всех элементов HTML или всех элементов, принадлежащих к набору элементов Dublin Core следует указать, разрешить ли пользователям выполнять поиск с указанием имени поля и следует ли выводить поля в результатах поиска. Следует также указать, должны ли термины поиска пользователей полностью соответствовать всему значению элемента метаданных (чтобы каждое слово в поле соответствовало одному из заданных терминов поиска пользователя) и разрешить ли пользователям сортировку результатов поиска по данному полю.

Система автоматически использует имена элементов метаданных в качестве имен полей. Если вы хотите переопределить имена полей по умолчанию для конкретных элементов или использовать с конкретными элементами другие опции поиска, можно отобразить имя элемента на уникальное имя поля, а затем задать опции поиска для использования с данным элементом.

- Чтобы отобразить конкретные элементы метаданных на поля поиска:
 - a. Нажмите кнопку **Добавить поле**, чтобы добавить в список правил отображения полей пустую строку.
 - b. Введите имя, которое хотите связать с отображаемым элементом метаданных HTML. Пользователи смогут указывать имя этого поля при запросе документов HTML в этом собрании.

Совет: Для получения информации о символах и метасимволах ASCII, не поддерживаемых в именах полей, нажмите кнопку **Справка**.

- c. Введите имя отображаемого элемента метаданных.
- d. Чтобы пользователи могли делать запросы по этому полю, включите переключатель **Поиск по полю**.

Совет: Если этот переключатель выключен, в поле не будет возможен поиск ни для поиска по полям, ни для свободного текстового поиска. Это действие может быть полезным, например, если вы выбрали опцию включить все элементы метаданных HTML или все элементы метаданных Dublin Core, но не хотите разрешать поиск по некоторым полям.

- e. Чтобы пользователи могли вызвать запросы по этому полю и задать, что документ соответствует критерию поиска, только если терминам запроса соответствует все значение поля (то есть, если в поле нет никаких других слов), включите переключатель **Полное соответствие**.
- f. Чтобы пользователи могли сортировать результаты поиска в алфавитном порядке (строчная сортировка) по этому полю, включите переключатель **Возможность сортировки**.
Это поле выводится, как сортируемое, только если программа поиска поддерживает эту возможность. Пример программы поиска не содержит в списке доступных для сортировки полей те поля, на которые отображаются элементы метаданных HTML.
- g. Если тип данных для поля - DECIMAL, DOUBLE, INTEGER, SHORT, TIME или TIMESTAMP и вы хотите дать пользователям возможность использовать параметрические запросы при поиске по этому полю или сортировать результаты в числовом порядке по значениям этого поля, включите переключатель **Параметрический поиск**.
- h. Чтобы пользователи могли просмотреть это поле в результатах поиска, включите переключатель **Результаты поиска**.

Пример:

Пользователи могут включать в запрос имена отображенных полей для поиска документов в формате HTML с конкретными метаданными. Например, если вы отображали элемент метаданных HTML с именем description на поле поиска с именем abstract, пользователи, чтобы найти документы HTML, в которых есть данные о компьютерах Thinkpad, могут ввести примерно такой запрос:

```
abstract:thinkpad
```

Пользовательская обработка текста

Можно улучшить качество и точность результатов поиска при помощи интеграции алгоритмов пользовательской обработки текста с собраниями поисковой системы предприятия.

OmniFind Enterprise Edition поддерживает архитектуру управления неструктурированной информацией IBM (Unstructured Information Management Architecture, UIMA) - структуру для создания, просмотра, составления и внедрения функций анализа текста. Разработчики программ создают и тестируют алгоритмы анализа для содержимого поиска, затем создают архив механизмов обработки (файл .pear), содержащий все ресурсы, необходимые для использования этого архива в поиске предприятия. Для возможности поиска в собраниях с использованием пользовательских алгоритмов анализа этот архив (содержащий механизм анализа текста) нужно добавить в поисковую систему предприятия.

Компонент логики анализа в механизме анализа текста называется *аннотатором*. Каждый аннотатор выполняет конкретные задачи лингвистического анализа. Механизм обработки текста может содержать любое число аннотаторов или состоять из нескольких механизмов анализа текста, каждый из которых содержит свои собственные пользовательские аннотаторы.

Информация, генерируемая аннотаторами, называется *результатами анализа*. Результаты анализа, соответствующие искомой вами информации, записываются в структуру данных, которая называется *общая структура анализа*.

Чтобы сконфигурировать опции обработки текста, выполните следующие задачи:

- Выберите механизм анализа текста, который нужно использовать для создания аннотаций к документам в собрании.
- Если собрание содержит документы XML со смысловой разметкой, и вы хотите использовать эту разметку в пользовательском анализе текста, можно связать с собранием файлы отображения и отобразите вывод отображения на общую структуру анализа.

Например, можно отобразить содержимое элементов <addressee> и <customer> на аннотации Person в общей структуре анализа. Эти аннотации станут доступными вашим пользовательским аннотаторам, которые смогут найти дополнительную информацию (например, пол сотрудника). Можно также отобразить аннотации Person на индекс поисковой системы предприятия, что позволит пользователям выполнять поиск сотрудников, не зная исходные элементы XML.

Если вы хотите разрешить пользователям задавать в запросах исходные элементы XML, отображения XML определять не нужно. Вместо этого можно сконфигурировать опции синтаксического анализа и включить собственный поиск XML для собрания.

- Отобразите общую структуру анализа на индекс поисковой системы предприятия, что позволит выполнять поиск аннотированных документов с применением семантического поиска.

Например, в зависимости от объектов и взаимосвязей, обнаруживаемых аннотаторами, пользователи могут выполнять поиск понятий, встречающихся в одном и том же предложении (таких как конкретное имя человека и название фирмы), или выполнять поиск ключевого слова и понятия (например, найти имя Alex и номер телефона).

- Отобразить общую структуру анализа на реляционную базу данных. Можно отобразить данные на таблицы IBM DB2 или Oracle. Этот тип отображения позволяет использовать результаты анализа в программах баз данных, например, в программе исследования данных. Кроме того, он позволяет при помощи запросов SQL выполнять поиск данных вне поисковой системы предприятия.

Понятия, связанные с данным

-  Интеграция пользовательского анализа текста
-  Основные понятия текстового анализа
-  Рабочий поток для интеграции пользовательского анализа
-  Алгоритмы анализа текста
-  Прикладные программы семантического поиска
-  Элемент запроса семантического поиска

Добавление механизмов анализа текста в систему

Если вы создали пользовательский механизм анализа текста, чтобы использовать его для поисковой системы предприятия, необходимо сначала добавить этот механизм в систему. Собрания могут использовать этот механизм для анализа документов, составления аннотаций, а также для повышения точности результатов поиска.

Прежде чем приступить к работе...

Чтобы добавлять механизмы анализа текста в систему, надо быть администратором поисковой системы предприятия.

Об этой задаче

Разработчики прикладных программ могут создавать архив механизма обработки (файл .pear), который следует при анализе текста шаблону UIMA. Этот архив содержит все ресурсы, необходимые для выполнения поиска в собраниях поисковой системы предприятия. Для возможности поиска в собраниях с использованием пользовательских алгоритмов анализа этот архив (содержащий механизм анализа текста) нужно добавить в поисковую систему предприятия.

После добавления механизма анализа текста в систему можно изменить его имя для вывода и выбрать опцию для просмотра источника XML. Этот источник XML показывает, какая информация генерируется данным анализом.

Если механизм анализа текста связан с собранием, его нельзя удалить из системы.

Процедура

Чтобы добавить пользовательский механизм анализа текста в поисковую систему предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Анализ выберите **Конфигурировать механизмы анализа текста**.
4. На странице Механизмы анализа текста выберите **Добавить механизм анализа текста**.
5. На странице Добавить механизм анализа текста введите описательное имя для нового механизма. Система использует это имя для идентификации механизма анализа текста на консоли администратора.
6. Задайте положение файла .pear. Если размер файла меньше 8 Мбайт, этот файл может находиться в вашей системе, в таком случае его можно найти при помощи просмотра. Если размер этого файла больше 8 Мбайт, он должен быть на сервере индексов; в этом случае задайте полный путь файла.
7. Нажмите кнопку **ОК**. Новый механизм анализа текста появится в списке на странице Механизмы анализа текста.

Понятия, связанные с данным

 Рабочий поток для интеграции пользовательского анализа

 Интеграция пользовательского анализа текста

 Основные понятия текстового анализа

 Разметка XML при анализе и поиске

Задачи, связанные с данной

 Создание файла отображения элементов XML на общую структуру анализа

Связывание механизма текстового анализа с собранием

Если с поисковой системой предприятия связаны пользовательские механизмы анализа текста, можно выбрать один из них для использования с собранием. Это

позволит пользователям делать семантические запросы при поиске в собрании, а также повысит качество и точность результатов поиска.

Прежде чем приступать к работе...

Чтобы связать механизм анализа текста с собранием, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Об этой задаче

Если с этим собранием уже связан один из механизмов анализа текста, при связывании другого механизма анализа произойдет следующее:

- Если выбрать опцию **Без пользовательского анализа**, все отображения анализа текста, ранее определенные для этого собрания, будут сброшены. Собрание станет использовать системные значения по умолчанию.
- Если выбрать имя другого пользовательского механизма анализа текста, все отображения анализа текста, ранее определенные для собрания, будут сохранены. Например, если вы замените механизм_1 на механизм_2, механизм_2 унаследует файлы отображения XML, сконфигурированные для механизм_1.

Процедура

Чтобы связать механизм анализа текста с собранием:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. Выберите **Выбрать механизм анализа текста**. Если в систему поиска предприятия не были добавлены пользовательские механизмы анализа текста или если собрание использует алгоритмы анализа по умолчанию, именем этого механизма будет **По умолчанию**.
3. На странице Выберите механизм анализа текста для этого собрания выберите имя механизма, который будет использоваться в этом собрании. Если нет доступных механизмов анализа текста, или выбрана опция **Без пользовательского анализа**, при подготовке документов для индекса анализатор применяет правила анализа текста по умолчанию.

Понятия, связанные с данным

-  Рабочий поток для интеграции пользовательского анализа
-  Интеграция пользовательского анализа текста
-  Основные понятия текстового анализа

Отображение элементов XML на общую структуру анализа

Если собрание содержит документы XML со смысловой разметкой, и вы хотите при помощи этой разметки разрешить пользователям выполнять поиск в индексе или таблицах реляционных баз данных поисковой системы предприятия с использованием семантического поиска, нужные элементы XML можно отобразить на общую структуру анализа.

Прежде чем приступать к работе...

Чтобы отображать элементы XML на общую структуру анализа, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Ограничения

Максимальный размер файла отображения - 8 Мбайт.

Об этой задаче

Чтобы включить для процессов пользовательского механизма анализа текста доступ к конкретным элементам в документах XML, или отобразить несколько элементов XML на общий тип для использования в семантическом поиске, можно создать пользовательские файлы отображения. Файлы отображения должны следовать шаблону UIMA для анализа текста.

При добавлении файлов отображения в собрание, где используется пользовательский механизм анализа текста, нужно включить отображение элементов XML в исходных документах на аннотации в общей структуре анализа. Затем эти аннотации сможет использовать пользовательский механизм анализа текста. После этого общую структуру анализа можно отобразить на индекс и позволить пользователям делать запросы к аннотациям при поиске в собрании с применением семантического поиска.

Например, можно отобразить содержимое элементов `addressee` и `customer` на аннотации `Person` в общей структуре анализа. Эти аннотации станут доступными вашим пользовательским аннотаторам, которые смогут найти дополнительную информацию (например, пол сотрудника). Можно также отобразить аннотации `Person` на индекс поисковой системы предприятия, что позволит пользователям выполнять поиск сотрудников, не зная исходные элементы XML.

Если вы хотите разрешить пользователям задавать в запросах исходные элементы XML, конфигурировать отображения не нужно. Вместо этого можно сконфигурировать опции синтаксического анализа и включить собственный поиск XML для собрания.

Процедура

Чтобы отобразить элементы XML на общую структуру анализа:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. В области **Отобразить элементы XML на общую структуру анализа**, нажмите кнопку **Добавить отображение**.
3. На странице **Отобразить элементы XML на общую структуру анализа** введите для файла отображения описательное имя для вывода.
4. Задайте положение файла. Если файл отображения находится в локальной системе, можно нажать кнопку **Просмотр** и найти этот файл. Если файл отображения находится на сервере индексов, введите полный путь.
5. Нажмите кнопку **ОК**. Новый файл отображения будет добавлен на страницу **Опции обработки текста**.

Понятия, связанные с данным

-  Рабочий поток для интеграции пользовательского анализа
-  Интеграция пользовательского анализа текста
-  Основные понятия текстового анализа
-  Разметка XML при анализе и поиске

Задачи, связанные с данной

 Создание файла отображения элементов XML на общую структуру анализа

Отображение общей структуры анализа на индекс

Можно задать, какие результаты текстового анализа из общей структуры анализа следует отобразить на индекс и сделать доступными для пользователей, направляющих запросу к собранию с использованием семантического поиска.

Прежде чем приступить к работе...

Чтобы отобразить общую структуру анализа на индекс, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Ограничения

Максимальный размер файла отображения - 8 Мбайт.

Об этой задаче

Отображение общей структуры анализа на индекс поисковой системы предприятия позволяет пользователям задавать семантически точные запросы, что повышает качество результатов поиска.

Например, в зависимости от объектов и взаимосвязей, обнаруживаемых аннотаторами, пользователи могут выполнять поиск понятий, встречающихся в одном и том же предложении (таких как конкретное имя сотрудника и название фирмы), или слова и понятия (например, найти имя Alex и номер телефона).

Процедура

Чтобы отобразить общую структуру анализа на индекс:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. В области **Отобразить общую структуру анализа на индекс** нажмите кнопку **Выбрать файл отображения**.
3. На странице Выбрать файл отображения для этого собрания выберите файл отображения для использования с индексом поисковой системы предприятия:
 - Чтобы использовать для индекса поисковой системы предприятия правила отображения по умолчанию, выберите **По умолчанию**.
 - Чтобы отобразить на индекс общую пользовательскую структуру анализа, задайте положение нужного файла отображения. Если файл находится в вашей локальной системе, его можно найти в окне просмотра. Если этот файл находится на сервере индексов, введите полный путь.
4. Нажмите кнопку **ОК**. Заданный вами файл отображения появится на странице Опции обработки текста.

Понятия, связанные с данным

 Рабочий поток для интеграции пользовательского анализа

 Интеграция пользовательского анализа текста

 Основные понятия текстового анализа

 Отображение индекса для результатов пользовательского анализа

Задачи, связанные с данной

 Создание файла отображения общей структуры анализа на индекс

Отображение общей структуры анализа на реляционную базу данных

Можно задать, какие результаты текстового анализа из общей структуры анализа следует отобразить на реляционную базу данных для использования в прикладных программах базы данных.

Прежде чем приступить к работе...

Чтобы отобразить общую структуру анализа на реляционную базу данных, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Ограничения

Максимальный размер файла отображения - 8 Мбайт.

Об этой задаче

Отображение общей структуры анализа на таблицы реляционной базы данных делает данные доступными для использования прикладными программами базы данных. Например, пользователи могут задавать запросы SQL вне поисковой системы предприятия для выполнения поиска в аннотациях, добавленных при анализе текста. Эту информацию можно также использовать для дополнительной обработки текста, например, в программах исследования данных.

Для отображения общей структуры анализа на реляционную базу данных можно использовать одновременно только один файл отображения. Если загрузить новый файл отображения, текущий файл отображения будет перезаписан.

Процедура

Чтобы отобразить общую структуру анализа на реляционную базу данных:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. В области **Отобразить общую структуру анализа на реляционную базу данных** нажмите кнопку **Добавить отображение**.
3. На странице Отобразить общую структуру анализа на реляционную базу данных введите для файла отображения, который будет использоваться для отображения информации на реляционную базу данных, описательное имя для вывода.
4. Задайте положение файла отображения. Если файл находится в вашей локальной системе, его можно найти в окне просмотра. Если этот файл находится на сервере индексов, введите полный путь.
5. Нажмите кнопку **ОК**. Имя для вывода, заданное вами для файла отображения, появится на странице Опции обработки текста.

Понятия, связанные с данным

 Рабочий поток для интеграции пользовательского анализа

 Интеграция пользовательского анализа текста

-  Основные понятия текстового анализа
-  Отображение базы данных для выбранных результатов анализа
- Задачи, связанные с данной**
-  Создание файла отображения общей структуры анализа на базу данных

Конфигурирование потоков для службы анализатора

Если хватает ресурсов памяти, можно увеличить число потоков, доступных синтаксическому анализатору для анализа документов.

Прежде чем приступить к работе...

Число потоков анализатора может иметь смысл увеличить, если у вас большое число собраний. Убедитесь, что в системе достаточно памяти для поддержки дополнительных потоков. Анализатору с одним потоком требуется 200 Мбайт памяти. Каждому дополнительному потоку требуется 50 Мбайт памяти.

Конфигурировать число потоков, запускаемых для анализатора, может администратор поисковой системы предприятия или администратор собрания.

Процедура

Чтобы сконфигурировать число потоков синтаксического анализатора:

1. Откройте собрание для редактирования, откройте страницу Анализ и нажмите кнопку **Конфигурировать опции анализа**.
2. Задайте число потоков анализатора, запускаемых при запуске анализатора, и нажмите кнопку **ОК**.
3. Перезапустите анализатор.

Включение расширенного анализа для составных терминов

Можно улучшить качество поиска, разрешив синтаксическому анализатору использовать дополнительный анализ сложных терминов. Дополнительный анализ разделяет сложные термины на составные части, и каждая часть обрабатывается как отдельный термин.

Прежде чем приступить к работе...

Чтобы задавать опции для синтаксического анализа сложных терминов, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Об этой задаче

В некоторых языках принято писать термины из нескольких корней в одно слово без пробелов (*сложные термины*). Дополнительный анализ и разделение сложных терминов полезны при поиске в таких языках, как немецкий, и критически важны в языках типа корейского.

Если вы включите дополнительный анализ для сложных терминов, пользователи смогут искать термины в составе сложных терминов без использования символов подстановки в терминах запроса. Например, поиск немецкого *Organ* (орган) может

вернуть документы, содержащие термины *Organspender* (донор органов), но не вернет документы, содержащие *Organisation* (организация). Если при запросе с символом подстановки *Organ** возможен возврат любой строки, начинающейся с *Organ*, при данном поиске признаются только те соответствия, где найдено полное слово - компонент сложного термина.

Другой пример: составной термин *Mustermann* разбивается на два элемента (*muster* и *mann*), которые хранятся в индексе по отдельности. Если в запросе задан терм с символом подстановки *Musterna**, программа поиска не может идентифицировать *Musterna* как префикс разбитого на элементы слова. В результате документы, содержащие *Mustermann*, не будут найдены. Если вы хотите, чтобы пользователи могли использовать символы подстановки в запросах для составных термов, не включайте для них расширенный анализ.

Термины пользовательских словарей, например, словаря синонимов и словаря значимых слов, также применяются к составным частям, используемым как отдельные слова в запросе.

Процедура

Чтобы разрешить дополнительный анализ сложных терминов:

1. Вызовите редактирование собрания, выберите Синтаксический анализ и щелкните по **Конфигурировать опции синтаксического анализа**.
2. Включите переключатель **Разрешить дополнительный анализ сложных терминов** и нажмите кнопку **ОК**.

Понятия, связанные с данным

“Символы подстановки в запросах” на стр. 187



Лингвистическая поддержка семантического поиска



Включение анализа текста в поисковую систему предприятия

Включение поддержки собственного поиска XML

Если собрание содержит документы XML, можно разрешить пользователям использовать при поиске документов разметку XML, включив для собрания поддержку собственного поиска XML.

Прежде чем приступить к работе...

Включить поддержку поиска документов XML с использованием собственного поиска XML может только администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Собственный поиск XML, например, поиск XPath или фрагментов XML, может обеспечить более точные результаты поиска, благодаря использованию разметки XML документов в запросе. Пользователи могут задать, чтобы термин запроса встречался в определенном элементе или атрибуте XML.

Например, счета от продавца компьютеров в формате XML могут содержать записи `<order>` (заказ), включающие элементы `<company>` (компания) и `<computertype>` (тип компьютера). Если вы ищете найти счета, содержащие заказы на портативные компьютеры (notebook) фирмы IBM, поиск по ключевым словам IBM и notebook

можно дать вам документы с информацией о портативных компьютерах Dell и о моделях настольных компьютеров IBM. Поиск XML позволяет задать, чтобы слово IBM присутствовало в элементе <company>, слово notebook - в элементе <computer> и чтобы оба элемента находились в одном и том же элементе <order>. Так вы получите заказы, относящиеся конкретно к портативным компьютерам IBM.

Процедура

Чтобы разрешить пользователям выполнять поиск в собрании, используя собственный поиск XML:

1. Откройте собрание для редактирования, откройте страницу Анализ и нажмите кнопку **Конфигурировать опции анализа**.
2. Выберите **Разрешить пользователям выполнять поиск документов XML с использованием собственного поиска XML**.
3. Нажмите кнопку **ОК**.

Понятия, связанные с данным

-  Лингвистическая поддержка семантического поиска
-  Включение анализа текста в поисковую систему предприятия
-  Прикладные программы семантического поиска
-  Элемент запроса семантического поиска

“поля поиска XML” на стр. 139

Определение форматов документов

В поисковой системе предприятия типы документов и соответствующие им типы анализаторов определяются с помощью отображения по умолчанию расширений URL и типов MIME.

Отредактировав файл конфигурации `parserTypes.cfg`, можно переопределить и дополнить отображения по умолчанию расширений URL и типов MIME на типы анализаторов. В файле `parserTypes.cfg` определены правила назначения расширений файлов и типов MIME типам анализаторов. Например, для расширения файлов `.content` можно задать, что документы этого типа должны анализироваться анализатором HTML.

Разные форматы документов имеют различные внутренние представления. В поисковой системе предприятия для анализа документов используются внутренние и внешние фильтры, а многие документы анализируются с помощью служб анализаторов, специализированных для конкретных форматов.

Определение формата документов и назначение анализатора происходит следующим образом:

1. Алгоритм определения формата документов проверяет расширение URL обрабатываемого документа.
2. Система проверяет тип MIME документа, представляющий собой часть заданных искателем метаданных.
3. Система пытается назначить каждому документу подходящий тип анализатора. Для текстовых документов (TXT), документов HTML и XML система назначает особый тип анализатора для каждого формата.

Для всех остальных форматов документов система использует анализатор Stellent. Обратите внимание на то, что технология фильтрации документов Stellent не принадлежит компании Oracle. Ссылки на Stellent в этой документации - синонимы ссылок на технологию Oracle Outside In Content Access.

Анализатор Stellent поддерживает несколько сот форматов документов, но в поисковой системе предприятия разрешено использовать только часть имеющихся фильтров документов. Однако файлы конфигурации можно отредактировать, чтобы разрешить анализатору Stellent анализировать другие типы документов.

Важное замечание: Фильтры документов, добавляемые вами и не входящие в поднабор фильтров документов в конфигурации системы по умолчанию, не проверены, и поддержка для них не обеспечивается.

4. Если файл `parserTypes.cfg` недоступен, для определения типа документа и соответствующего ему анализатора используется назначение по умолчанию. Чтобы определить тип документа, система выполняет следующие действия:
 - a. Сравнивает расширение URL с пользовательскими правилами расширения в файле `parserTypes.cfg`.
 - b. Сравнивает тип MIME с пользовательскими правилами для типов MIME в файле `parserTypes.cfg`.
 - c. Сравнивает расширение URL с правилами по умолчанию для поисковой системы предприятия.

- d. Сравнивает тип MIME с правилами по умолчанию для типов MIME для поисковой системы предприятия.
5. Если система не в состоянии установить формат документа, этот документ будет отброшен. При этом на экран может быть выведено сообщение об ошибке, извещающее, что данный тип документов не поддерживается.

Если в качестве анализатора назначен Stellent, вы можете увидеть сообщение о том, что он не может опознать тип документа. Эта ошибка возникает, если:

- Документ поврежден.
- Формат документа не поддерживается Stellent. Чтобы устранить ошибку, добавьте в файл `stellentTypes.cfg` форматы отброшенных файлов. Необходимо также изменить файл `parserTypes.cfg`, указав, что типы MIME или расширения форматов отброшенных документов надо связать с анализатором Stellent.

Поддерживаемые типы документов по умолчанию

При определении формата документов оцениваются только документы определенных типов.

Следующие форматы документов - собственные типы; они распознаются и анализируются автоматически встроенными службами анализаторов собраний:

HTML
Plain text
XML

По умолчанию документы следующих форматов анализирует анализатор Stellent:

Adobe Portable Document Format (PDF)
Lotus 1-2-3
Lotus Freelance Graphics
Lotus Word Pro
Just System Ichitaro
Microsoft Excel (версии до 2007)
Microsoft PowerPoint (версии до 2007)
Microsoft Visio
Microsoft Word (версии до 2007)
Rich Text Format (RTF)
StarOffice/OpenOffice Calc
StarOffice/OpenOffice Impress
StarOffice/OpenOffice Draw
StarOffice/OpenOffice Writer

Форматы файлов Office Open XML в Microsoft Office и форматы OpenDocument в OpenOffice обрабатываются без внесения изменений в файлы конфигурации.

Для анализа документов других типов надо изменить файлы конфигурации (`parserTypes.cfg` и `stellenttypes.cfg`), задав в них правила отображения определенных типов документов на службы анализатора собрания или на фильтр Stellent.

Ограничение: Обработка текста с двумя направлениями письма в файлах PDF для соответствия логическому порядку чтения технологией Stellent не предусмотрена. Анализатор Stellent не дает гарантии правильности порядка текста, извлеченного из файлов PDF. Для файлов PDF с двумя направлениями письма порядок анализа текста, скорее всего, будет отличаться от логического порядка чтения. Это

ограничение вызывает проблемы при обработке файлов PDF, написанных на таких языках, как иврит и арабский, где большая часть текста написана справа налево (текст с двумя направлениями письма).

Типы документов, связанные с анализаторами собраний и анализаторами Stellent

Чтобы обеспечить точный и эффективный анализ документов в пространстве искателя можно создать файлы конфигурации и задать в них, какие типы документов будут анализироваться анализатором собрания, а какие - фильтрами документов Stellent.

В собрании поисковой системы предприятия большинство форматов документов обрабатывается встроенными анализаторами HTML или XML. Некоторые типы документов обычно не анализируются (например, документы Postscript), другие типы документов обрабатываются функциями анализа Stellent (например, типы документов Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT и Ichitaro).

Поскольку метаданные бывают неточны, иногда документы в формате простого текста и HTML ошибочно направляются анализаторам Stellent, а оттуда обратно в один из встроенных анализаторов, что может снизить производительность. Для других документов определить их тип может оказаться невозможным, поэтому такие документы будут пропущены. Чтобы избежать такой ситуации, можно создать файлы конфигурации, задающие, где и как должны обрабатываться различные типы документов.

Чтобы связать типы документов с анализатором собрания и анализатором Stellent, надо выполнить следующие задачи:

1. Задание типов документов для анализатора собрания. Этот шаг включает создание файла конфигурации, в котором типы документов отображаются на анализатор, используемый собранием. Можно создать по одному такому файлу конфигурации для каждого собрания.
2. Задание типов документов для анализатора Stellent. Этот шаг включает в себя создание файла конфигурации, в котором типы документов отображаются на фильтры документов Stellent, используемые собранием. Можно создать по одному такому файлу конфигурации для каждого собрания.
3. Остановка и перезапуск анализатора. Чтобы изменения вступили в силу, с консоли управления поисковой системы предприятия включите монитор собрания, для которого задали типы документов, после чего остановите и перезапустите анализатор.

Связывание типов документов с анализатором собраний

Чтобы связать типы документов с анализатором собраний, нужно создать файл конфигурации `parserTypes.cfg`. С консоли администратора поиска эта задача не поддерживается.

Прежде чем приступать к работе...

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Если файл конфигурации не существует, анализатор собрания использует правила службы анализатора по умолчанию. Если файл конфигурации существует, правила в этом файле задают:

- Какое расширение URL и какой тип MIME отображаются на каждый тип анализатора.
- Как анализировать документы, тип которых неизвестен из-за неполных метаданных.

Формат файла `parserTypes.cfg` - последовательность строк, где каждая строка задает одно из следующих правил:

EXTENSION *расширение анализатор*

Все документы, URL которых заканчиваются указанным расширением, будут обрабатываться заданным анализатором. Не включайте в расширение точку; регистр символов при сравнении не учитывается.

CONTENTTYPE *тип/подтип анализатор*

Заданным анализатором будут обрабатываться все документы, тип содержимого которых соответствует указанному типу/подтипу. Тип/подтип содержимого документа будет соответствовать условию поиска, если тип документа эквивалентен значению `тип`, а подтип эквивалентен значению `подтип` (или если в качестве значения `подтип` задан символ подстановки (звездочка, `*`)).

UNKNOWN *анализатор*

Заданным анализатором будут обрабатываться все документы, у которых неизвестны расширение и тип содержимого (то есть документы, не сделанные доступными искателем).

DEFAULT *анализатор*

Заданным анализатором будут обрабатываться все документы, не обрабатываемые другими правилами.

Для всех случаев должен быть задан *анализатор* `text`, `html`, `xml`, `stellent` или `none`, где `none` означает, что документы данного типа обрабатываться не будут.

Если документу соответствует несколько правил, приоритет будет выше у более конкретного правила, независимо от их порядка:

- Правило **EXTENSION** более конкретно, чем правило **CONTENTTYPE**.
- Правило **CONTENTTYPE**, включающее подтип, более конкретно, чем то же правило с символом подстановки. Например, приоритет правила для типа содержимого `application/postscript` выше приоритета правила для типа `application/*`.
- Для одного расширения или типа содержимого не должно существовать двух правил. В этом случае приоритет правил зависит от реализации.

Процедура

Чтобы связать типы документов с анализатором собраний:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Создайте файл конфигурации, как показано ниже (*ID_собрания* это ID конфигурируемого вами собрания):
`КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.parserdriver/parserTypes.cfg`

3. При помощи текстового редактора отредактируйте файл и задайте правила правила служб анализатора, затем сохраните и закройте этот файл.
4. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Пример

В следующем примере встроенный анализатор HTML обрабатывает все документы с расширением txt, htm или html, с типом содержимого, начинающимся с text/, или с неизвестным расширением и типом содержимого. Встроенный анализатор XML обрабатывает все документы с расширением xml или типом содержимого text/xml. Все остальные документы, в том числе с типом содержимого, начинающимся с application/, посылаются анализатору Stellent.

```
EXTENSION doc stellent
EXTENSION txt htm
EXTENSION htm htm
EXTENSION html html
EXTENSION xml xml
EXTENSION ps none
CONTENTTYPE text/xml xml
CONTENTTYPE text/* htm
CONTENTTYPE application/* stellent
UNKNOWN htm
DEFAULT stellent
```

Задачи, связанные с данной

“Изменение правил замены для некоторых тегов HTML” на стр. 161

“Связывание типов документов с анализатором Stellent” на стр. 163

Ссылки, связанные с данной

“Правила по умолчанию для служб анализаторов собрания”

Правила по умолчанию для служб анализаторов собрания

Если не создан файл конфигурации, в котором типы файлов и типы содержимого назначены искателю для собрания, при анализе документов будут использоваться правила по умолчанию.

По умолчанию анализатор собрания использует следующие правила:

```
EXTENSION pdf stellent
EXTENSION ppt stellent
EXTENSION prz stellent
EXTENSION lwp stellent
EXTENSION doc stellent
EXTENSION rtf stellent
EXTENSION xls stellent
EXTENSION 123 stellent
EXTENSION vsd stellent
EXTENSION vdx stellent
EXTENSION jxw stellent
EXTENSION jsw stellent
EXTENSION jtw stellent
EXTENSION jaw stellent
EXTENSION juw stellent
EXTENSION jbw stellent
EXTENSION jvw stellent
EXTENSION jfw stellent
EXTENSION jtt stellent
EXTENSION jtd stellent
EXTENSION jttd stellent
EXTENSION jtdd stellent
```

```
EXTENSION jtdx stellent
EXTENSION ps none
EXTENSION xml xml
EXTENSION txt text
EXTENSION htm html
EXTENSION html html
EXTENSION shtml html
EXTENSION xhtml html
EXTENSION asp html

CONTENTTYPE application/postscript none
CONTENTTYPE application/* stellent
CONTENTTYPE text/rtf stellent
CONTENTTYPE text/richtext stellent
CONTENTTYPE text/xml xml
CONTENTTYPE text/html html
CONTENTTYPE text/plain text

UNKNOWN none
DEFAULT none
```

Задачи, связанные с данной

“Связывание типов документов с анализатором собраний” на стр. 157

Анализ документов неизвестных типов

Если тип документа неизвестен (например, у файла документа нет расширения или же с этим документом не связан тип MIME), можно задать правила, чтобы анализатор не отклонил этот документ.

Об этой задаче

Если анализатор не распознал формат файла, он пытается проанализировать этот документ по умолчанию, как документ HTML. Если содержимое не соответствует формату HTML, анализатор отклоняет документ. Записи обо всех отклоненных документах записываются в следующее место сервера индексов:

```
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/ID_собрания/dropped_doc_logs/
dropped_docs_pd_date.log
```

Процедура

Чтобы избежать этой ситуации, сконфигурируйте анализатор на использование анализа ASCII для документов неизвестных типов:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Создайте или отредактируйте следующий файл конфигурации анализатора, где *ID_собрания* - это ID конфигурируемого вами собрания:
КОРНЕВОЙ_КАТАЛОГ_ES/master_config/*ID_собрания*.parserdriver/parserTypes.cfg
3. Добавьте в этот файл следующее правило:
UNKNOWN text
4. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Изменение правил замены для некоторых тегов HTML

Можно изменить правила замены тегов HTML, используемые анализатором при анализе документов HTML. С консоли администратора поиска эта задача не поддерживается.

Прежде чем приступить к работе...

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Чтобы передать текстовую информацию для индекса поисковой системы предприятия, анализатор HTML заменяет информацию разметки (теги HTML) другими символами, которые моделируют теги подобно тому, как их обрабатывает браузер Web. Например, вместо тега абзаца (<p>) в общей структуре анализа (Common Analysis Structure, CAS) и в индексе появится разделитель абзацев. Эти правила замены влияют на то, какая информация текста будет восприниматься как находящаяся в одном абзаце.

Чтобы приблизить значение определенных тегов HTML к стандарту HTML, можно изменить файл конфигурации анализатора для собрания, заменив некоторые из этих правил.

Совет: Чтобы узнать больше о других способах управления обработкой тегов HTML в документах HTML, посмотрите документ <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011251>.

Процедура

Чтобы изменить правила замены тегов HTML:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.parserdriver/collection.properties` (где *ID_собрания* - ID конфигурируемого вами собрания).
3. Задайте для следующего параметра значение true:
`trevi.tokenizer.newHtmlTagReplacement=true`

Правила замены для следующих тегов HTML изменены на показанные ниже значения. Для всех прочих тегов HTML продолжают использоваться правила по умолчанию.

```
dfn    EMPTY
div    NEW LINE
q      BLANK
span   EMPTY
```

4. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Задачи, связанные с данной

“Связывание типов документов с анализатором собраний” на стр. 157

Ссылки, связанные с данной

Правила замены HTML по умолчанию

Анализатор HTML интерпретирует информацию разметки в документах HTML и заменяет ее на текстовую информацию для индекса поисковой системы предприятия.

Анализатор HTML поисковой системы предприятия использует при анализе тегов HTML следующие правила замены. В первом столбце показан тег HTML. Во втором столбце показано значение замены.

comment	BLANK
a	EMPTY
applet	EMPTY
area	EMPTY
b	EMPTY
base	EMPTY
big	EMPTY
body	EMPTY
br	NEW LINE
caption	EMPTY
center	EMPTY
del	EMPTY
dfn	PARAGRAPH
dir	PARAGRAPH
div	BLANK
dl	PARAGRAPH
em	EMPTY
form	EMPTY
frame	PARAGRAPH
h1	PARAGRAPH
h2	PARAGRAPH
h3	PARAGRAPH
h4	PARAGRAPH
h5	PARAGRAPH
h6	PARAGRAPH
hr	PARAGRAPH
iframe	EMPTY
img	NEW LINE
li	BLANK
meta	EMPTY
object	EMPTY
ol	BLANK
option	EMPTY
p	PARAGRAPH
q	PARAGRAPH
samp	PARAGRAPH
script	EMPTY
select	EMPTY
spacer	BLANK
span	NEW LINE
strike	EMPTY
strong	EMPTY
style	EMPTY
table	PARAGRAPH
td	NEW LINE
title	PARAGRAPH
tr	NEW LINE
ul	BLANK
xmp	PARAGRAPH

Задачи, связанные с данной

“Изменение правил замены для некоторых тегов HTML” на стр. 161

Связывание типов документов с анализатором Stellent

Чтобы задать, какие типы документов должны анализировать фильтры документов Stellent, нужно создать файл конфигурации `stellenttypes.cfg`. С консоли администратора поиска эта задача не поддерживается.

Oracle Outside In Technology:

Технология фильтрации документов Stellent не принадлежит компании Oracle. Ссылки на Stellent в этой документации - синонимы ссылок на технологию Oracle Outside In Content Access.

Ограничения

OmniFind Enterprise Edition поддерживает фильтры Stellent для следующих типов документов:

- Adobe Portable Document Format (PDF)
- Lotus 1-2-3
- Lotus Freelance Graphics
- Lotus Word Pro
- Ichitaro
- Microsoft Excel
- Microsoft PowerPoint
- Microsoft Visio
- Microsoft Word
- Rich Text Format (RTF)

Если вы хотите включить дополнительные типы документов в индекс поисковой системы предприятия, и форматы этих документов поддерживаются фильтром Stellent, можно сконфигурировать правила анализа в файле конфигурации `stellenttypes.cfg`.

Важное замечание: Добавляемые вами форматы документов, не входящие в приведенный выше список, не проверены, и поддержка для них не обеспечивается.

Полный список форматов документов Stellent смотрите в разделе *Outside In Technology: Supported File Formats* документа http://www.oracle.com/technology/products/content-management/oit/ds_oitFiles.pdf.

Прежде чем приступить к работе...

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Файл конфигурации `stellenttypes.cfg` задает:

- Правила принятия, для типов файлов, которые должны быть проанализированы анализатором Stellent. Тип файла соответствует одному из типов файлов, распознаваемых библиотекой Stellent.
- Собственные правила для типов файлов, которые должны быть отправлены назад анализатору собрания для обработки одним из встроенных анализаторов. Эта операция необходима, поскольку анализатор собрания может отправить документ анализатору Stellent по ошибке, из-за неточных метаданных.
- Правила отклонения для типов файлов, которые должны быть отклонены, так как они не поддерживаются системой поиска предприятия.

Если файл конфигурации задан, но не существует, запуск анализатора завершится неудачно. Если файл конфигурации не задан в файле `stellent.properties` для свойства `OutsideInSupportedTypes`, для анализаторов Stellent будут использоваться правила синтаксического анализа по умолчанию.

Файл конфигурации содержит список типов документов и необходимые способы их обработки. Формат этого файла - последовательность строк, где каждая строка задает правило одного из следующих форматов:

```
accept DEFAULT
accept ALL тип документа
accept тип_stellent тип документа
native DEFAULT
native тип_stellent тип документа
reject тип_stellent
```

Где:

тип документа

- значение, которое должно использоваться для маркера запроса `doctype`. Можно выполнять поиск документов по их типам. Например, для поиска документов PDF пользователь может задать `$doctype::pdf`.

тип_stellent

Одно из значений типа фильтра в библиотеке Stellent, например, `FI_123R1`.

DEFAULT

- означает, что в список принимаемых или собственных типов, в зависимости от типа правила, включаются все правила по умолчанию. Эта опция позволяет не заменять конфигурацию по умолчанию, а вместо этого расширить ее.

ALL

- означает, что принимаются все типы с заданным маркером `doctype`, не перечисленные явно в списке.

Правила в файле конфигурации обрабатываются следующим образом:

1. Если для значения *тип* задано правило `reject`, тип не принимается.
2. Если для значения *тип_stellent* существует правило типа `native` (включая правила синтаксического анализа по умолчанию, если задано `native DEFAULT`), документ отсылается назад встроенному анализатору вместе со значением для маркера *тип документа*, задаваемого данным правилом. Значение *тип документа* должно быть `txt`, `htm` или `xml` (задающее соответственно простой текст, HTML или XML).
3. Если для значения *тип* задано правило `accept` (тип принимается (в том числе и типы списка по умолчанию, если задано `accept DEFAULT`)).
4. Если задано правило `accept ALL`, документ принимается.
5. Иначе документ отклоняется и не анализируется.

Если тип документов принимается, используется значение *тип документа*, заданное в примененном правиле. Это значение отсылается назад анализатору собрания вместе с проанализированным содержанием.

Процедура

Чтобы связать типы документов с анализатором Stellent:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.

2. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/ID_собрания.stellent.properties` (где `ID_собрания` - ID конфигурируемого вами собрания).
3. Для свойства `OutsideInSupportedTypes` задайте абсолютный путь создаваемого вами файла конфигурации.

Например, для одного собрания можно задать следующий файл конфигурации (и хранить его с другими файлами для конкретных собраний):

```
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.stellent/stellenttypes.cfg
```

Другой пример: для возможности использования одинаковых параметров со всеми собраниями можно создать следующий файл конфигурации (и хранить его с другими файлами системного уровня). При использовании этого способа обязательно укажите его путь в файле `stellent.properties` для каждого собрания, как указано на шаге 2.

```
КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/stellent/stellenttypes.cfg
```

4. При помощи текстового редактора создайте файл конфигурации и задайте правила анализа Stellent, затем сохраните и закройте этот файл.
5. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Примеры

В следующем файле конфигурации сеанс Stellent принимает, кроме типов документов списка по умолчанию, формат Microsoft Visio.

```
accept DEFAULT
accept FI_VISIO3 visio
accept FI_VISIO4 visio
accept FI_VISIO5 visio
accept FI_VISIO6 visio
```

В следующем файле конфигурации поддерживаются и доступны для поиска документы Postscript с типом документов `ps`, формат X pixmap (XPM) отсылается назад встроенному анализатору текста, формат изображений PNG отклоняется, а все остальные типы файлов принимаются и делаются доступными для поиска с типом документов `other`.

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
reject FI_PNG
accept ALL other
```

Задачи, связанные с данной

“Связывание типов документов с анализатором собраний” на стр. 157

Ссылки, связанные с данной

“Правила анализа по умолчанию для анализаторов Stellent”

Правила анализа по умолчанию для анализаторов Stellent

Если не создан файл конфигурации, в котором типы файлов назначены фильтрам документов Stellent, при анализе документов анализатор использует правила по умолчанию.

Анализатор Stellent принимает и анализирует документы следующих типов:

ACCEPT FI_123R1	123
ACCEPT FI_123R2	123
ACCEPT FI_123R3	123
ACCEPT FI_123R4	123
ACCEPT FI_123R6	123
ACCEPT FI_123R9	123
ACCEPT FI_EXCEL	xls
ACCEPT FI_EXCEL2000	xls
ACCEPT FI_EXCEL2002	xls
ACCEPT FI_EXCEL2003	xls
ACCEPT FI_EXCEL2007	xlsx
ACCEPT FI_EXCEL3	xls
ACCEPT FI_EXCEL4	xls
ACCEPT FI_EXCEL5	xls
ACCEPT FI_EXCEL97	xls
ACCEPT FI_EXTPOWERPOINT4	ppt
ACCEPT FI_EXTPOWERPOINTMAC4	ppt
ACCEPT FI_FREELANCE	prz
ACCEPT FI_FREELANCE3	prz
ACCEPT FI_ICHITAR03	jwx
ACCEPT FI_ICHITAR04	jsw
ACCEPT FI_ICHITAR08	jtd
ACCEPT FI_PDF	pdf
ACCEPT FI_PDFMACBIN	pdf
ACCEPT FI_POWERPOINT2	ppt
ACCEPT FI_POWERPOINT2000	ppt
ACCEPT FI_POWERPOINT2007	pptx
ACCEPT FI_POWERPOINT3	ppt
ACCEPT FI_POWERPOINT4	ppt
ACCEPT FI_POWERPOINT7	ppt
ACCEPT FI_POWERPOINT9597	ppt
ACCEPT FI_POWERPOINT97	ppt
ACCEPT FI_POWERPOINTMAC3	ppt
ACCEPT FI_POWERPOINTMAC4	ppt
ACCEPT FI_POWERPOINTMACB3	ppt
ACCEPT FI_POWERPOINTMACB4	ppt
ACCEPT FI_RTF	rtf
ACCEPT FI_RTFJ	rtf
ACCEPT FI_STAROFFICEWRITER8	odt
ACCEPT FI_STAROFFICEDRAW8	odg
ACCEPT FI_STAROFFICEIMPRESS8	odp
ACCEPT FI_STAROFFICECALC8	ods
ACCEPT FI_STAROFFICECALC6	sxc
ACCEPT FI_STAROFFICEDRAW6	sxd
ACCEPT FI_STAROFFICEIMPRESS6	sxi
ACCEPT FI_STAROFFICEWRITER6	sxw
ACCEPT FI_STAROFFICECALC52	sdc
ACCEPT FI_STAROFFICEIMPRESS52	sdd
ACCEPT FI_STAROFFICEWRITER52	sdw
ACCEPT FI_VISIO3	vsd
ACCEPT FI_VISIO4	vsd
ACCEPT FI_VISIO5	vsd
ACCEPT FI_VISIO6	vsd
ACCEPT FI_VISIO2003	vsd
ACCEPT FI_WINWORD1	doc
ACCEPT FI_WINWORD1COMPLEX	doc
ACCEPT FI_WINWORD1J	doc
ACCEPT FI_WINWORD2	doc
ACCEPT FI_WINWORD2000	doc
ACCEPT FI_WINWORD2002	doc
ACCEPT FI_WINWORD2003	doc
ACCEPT FI_WINWORD2007	docx
ACCEPT FI_WINWORD5J	doc
ACCEPT FI_WINWORD6	doc
ACCEPT FI_WINWORD7	doc
ACCEPT FI_WINWORD97	doc
ACCEPT FI_WORD4	doc

ACCEPT FI_WORD5	doc
ACCEPT FI_WORD6	doc
ACCEPT FI_WORDPRO	lwp
ACCEPT FI_WORDPRO97	lwp

Анализатор Stellant возвращает документы следующих типов анализатору собрания на обработку одним из встроенных анализаторов:

NATIVE FI_7BITTEXT	txt
NATIVE FI_ANSI	txt
NATIVE FI_ANSI8	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_ASCII	txt
NATIVE FI_ASCII8	txt
NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_CYRILLIC1251	txt
NATIVE FI_CYRILLICKO18	txt
NATIVE FI_EBCDIC_1026	txt
NATIVE FI_EBCDIC_273	txt
NATIVE FI_EBCDIC_277	txt
NATIVE FI_EBCDIC_278	txt
NATIVE FI_EBCDIC_280	txt
NATIVE FI_EBCDIC_284	txt
NATIVE FI_EBCDIC_285	txt
NATIVE FI_EBCDIC_297	txt
NATIVE FI_EBCDIC_37	txt
NATIVE FI_EBCDIC_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_OLDCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_HTML	htm
NATIVE FI_HTML_ARABIC_ASM0708	htm
NATIVE FI_HTML_ARABIC_DOS	htm
NATIVE FI_HTML_ARABIC_ISO	htm
NATIVE FI_HTML_ARABIC_MAC	htm
NATIVE FI_HTML_ARABIC_WINDOWS	htm
NATIVE FI_HTML_BALTIC_ISO	htm
NATIVE FI_HTML_BALTIC_WINDOWS	htm
NATIVE FI_HTML_CENTRALEUROPEAN_DOS	htm
NATIVE FI_HTML_CENTRALEUROPEAN_ISO	htm
NATIVE FI_HTML_CENTRALEUROPEAN_MAC	htm
NATIVE FI_HTML_CENTRALEUROPEAN_WINDOWS	htm
NATIVE FI_HTML_CHINESEBIG5	htm
NATIVE FI_HTML_CHINESEEUC	htm
NATIVE FI_HTML_CHINESEGB	htm
NATIVE FI_HTML_CHINESESIMPLIFIED_EUC	htm
NATIVE FI_HTML_CHINESESIMPLIFIED_WINDOWS	htm
NATIVE FI_HTML_CHINESETRADITIONAL_WINDOWS	htm
NATIVE FI_HTML_CYRILLIC_DOS	htm
NATIVE FI_HTML_CYRILLIC_ISO	htm
NATIVE FI_HTML_CYRILLIC_KO18R	htm
NATIVE FI_HTML_CYRILLIC_MAC	htm
NATIVE FI_HTML_CYRILLIC_WINDOWS	htm
NATIVE FI_HTML_CYRILLIC1251	htm
NATIVE FI_HTML_CYRILLICKO18	htm
NATIVE FI_HTML_EBCDIC_1026	htm
NATIVE FI_HTML_EBCDIC_273	htm
NATIVE FI_HTML_EBCDIC_277	htm
NATIVE FI_HTML_EBCDIC_278	htm

NATIVE FI_HTML_EBCDIC_280	htm
NATIVE FI_HTML_EBCDIC_284	htm
NATIVE FI_HTML_EBCDIC_285	htm
NATIVE FI_HTML_EBCDIC_297	htm
NATIVE FI_HTML_EBCDIC_37	htm
NATIVE FI_HTML_EBCDIC_500	htm
NATIVE FI_HTML_EBCDIC_870	htm
NATIVE FI_HTML_EBCDIC_871	htm
NATIVE FI_HTML_GREEK_ISO	htm
NATIVE FI_HTML_GREEK_MAC	htm
NATIVE FI_HTML_GREEK_WINDOWS	htm
NATIVE FI_HTML_HEBREW_DOS	htm
NATIVE FI_HTML_HEBREW_ISO_VISUAL	htm
NATIVE FI_HTML_HEBREW_WINDOWS	htm
NATIVE FI_HTML_JAPANESE_MAC	htm
NATIVE FI_HTML_JAPANESE_SHIFTJIS	htm
NATIVE FI_HTML_JAPANESE_EUC	htm
NATIVE FI_HTML_JAPANESE_JIS	htm
NATIVE FI_HTML_JAPANESE_SJIS	htm
NATIVE FI_HTML_KOREAN_JOHAB	htm
NATIVE FI_HTML_KOREAN_WINDOWS	htm
NATIVE FI_HTML_KOREAN_HANGUL	htm
NATIVE FI_HTML_LATIN2	htm
NATIVE FI_HTML_RUSSIAN_DOS	htm
NATIVE FI_HTML_THAI_WINDOWS	htm
NATIVE FI_HTML_TURKISH_DOS	htm
NATIVE FI_HTML_TURKISH_ISO	htm
NATIVE FI_HTML_TURKISH_MAC	htm
NATIVE FI_HTML_TURKISH_WINDOWS	htm
NATIVE FI_HTML_VIETNAMESE_WINDOWS	htm
NATIVE FI_HTML_WESTERNEUROPEAN_ISO	htm
NATIVE FI_HTML_WESTERNEUROPEAN_MAC	htm
NATIVE FI_HTML_WESTERNEUROPEAN_WINDOWS	htm
NATIVE FI_HTML_UNICODE	htm
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_JAPANESE_JIS	txt
NATIVE FI_LATIN2	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt
NATIVE FI_PP2KHTML	htm
NATIVE FI_SHIFTJIS	txt
NATIVE FI_UNICODE	txt
NATIVE FI_UTF8	txt
NATIVE FI_W2KHTML	htm
NATIVE FI_WML	xml
NATIVE FI_WML_CHINESEBIG5	xml
NATIVE FI_WML_CHINESE_EUC	xml
NATIVE FI_WML_CHINESEGB	xml
NATIVE FI_WML_CYRILLIC1251	xml
NATIVE FI_WML_CYRILLIC018	xml
NATIVE FI_WML_JAPANESE_EUC	xml
NATIVE FI_WML_JAPANESE_JIS	xml
NATIVE FI_WML_JAPANESE_SJIS	xml
NATIVE FI_WML_KOREAN_HANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_XHTML	htm
NATIVE FI_XL2KHTML	htm
NATIVE FI_XML	xml
NATIVE FI_XML_DOCTYPE_HTML	htm

Задачи, связанные с данной

“Связывание типов документов с анализатором Stellent” на стр. 163

Поддержка языков и кодовых страниц

Лингвистическая обработка для поисковой системы предприятия выполняется по-разному анализатором и поисковыми серверами.

При лингвистической обработке анализатор не делает различия между языками и национальными версиями. Даже если пользователь выполняет поиск в собрании, содержащем документы на разных языках, поисковые сервера ограничивают результаты поиска только конкретным языком или национальной версией.

Например, если в метаданных для английского документа указана локаль en_US, этот документ будет проиндексирован и как документ на английском языке (en), и как документ, использующий американскую версию английского языка (en_US). Этот тип индексирования позволяет правильно представить информацию, специфичную для национальной версии, например, числа, даты и время. При выполнении пользователем поиска в собрании, документ может быть найден независимо от того, ведется ли поиск документов en или en_US.

Если документ проиндексирован только по коду языка, например, en, локаль при этом не задается. Например, если пользователи будут искать в собрании документы с кодом en_US, такие документы не будут найдены.

В поисковой системе предприятия обеспечивается лингвистическая поддержка для следующих языков и двухсимвольных языковых кодов, указанных в стандарте ISO 639:

Языки с простым текстом:

- en=английский
- sq=албанский
- az=азербайджанский (латиница)
- bg=болгарский
- be=белорусский
- ca=каталанский
- hr=хорватский
- cs=чешский
- da=датский
- nl=голландский
- et=эстонский
- fi=финский
- fr=французский
- de=немецкий
- el=греческий
- hu=венгерский
- is=исландский
- id=индонезийский
- in=индонезийский
- it=итальянский
- kk=казахский
- lv=латышский
- lt=литовский
- lo=лаосский
- mk=македонский
- ms=малайский
- mt=мальтийский
- no=норвежский
- nb=норвежский (букмол)
- pl=польский
- pt=португальский
- ro=румынский

ru=русский
sr=сербский (кириллица)
sh=сербский (латиница)
sk=словацкий
sl=словенский
es=испанский
sv=шведский
tr=турецкий
uk=украинский
cy=валлийский

Идеографические языки:

Для упрощенного и традиционного китайского вместо двухсимвольных кодов используются расширенные коды.

zh-CN=китайский (упрощенный)
zh-TW=китайский (традиционный)
ja=японский
ko=корейский

Языки со сложным текстом:

ar=арабский
as=ассами
bn=бенгали
gu=гуджарати
iw=иврит
he=иврит
hi=хинди
kn=каннада
ml=малаялам
mr=маратхи
or=ория
pa=панджаби
ta=тамилский
te=телугу
th=тайский
ur=урду
vi=вьетнамский

Поисковая система предприятия может автоматически распознавать многие из этих языков и автоматически определять кодовую страницу, используемую в обычных текстовых документах. Если вы хотите явно задать используемый язык или кодовую страницу, при конфигурировании искателя можно отключить возможность их автоматического определения.

Автоматическое определение языка

Поисковая система предприятия может обрабатывать документы практически на любом языке.

Если документ написан на одном из указанных ниже языков, система может автоматически определить этот язык. Если язык документов известен, можно задать его при конфигурировании искателя, чтобы системе не нужно было автоматически определять язык.

Арабский
Болгарский
Чешский
Упрощенный китайский
Традиционный китайский
Датский
Голландский
Английский
Финский
Французский (Канада)

Французский (Франция)
Немецкий (Германия)
Немецкий (Швейцария)
Греческий
Иврит
Венгерский
Исландский
Итальянский
Японский
Корейский
Норвежский (букмол)
Польский
Португальский (Бразилия)
Португальский (Португалия)
Румынский
Русский
Испанский
Шведский
Тайский
Турецкий

Автоматическое определение кодовой страницы

Поисковая система предприятия поддерживает документы в различных кодовых страницах.

Для текстовых файлов система может автоматически определять указанные ниже кодовые страницы. Для других форматов документов система использует для определения кодовой страницы метаданные в документе (например, элементы метаданных HTML). Если кодовая страница документов известна, можно задать ее при конфигурировании искателя, чтобы системе не нужно было определять кодовую страницу автоматически.

Формы кодировки Unicode:

UTF-8
UTF-16BE
UTF-16LE

Формы многобайтной кодировки:

Shift-JIS
ISO-2022-CN
ISO-2022-JP
ISO-2022-KR
GB18030
EUC-JP
EUC-KR

Формы однобайтной кодировки:

ISO-8859-1: датский, голландский, немецкий, английский, французский, итальянский, норвежский, португальский, испанский, шведский
ISO-8859-2: чешский, венгерский, польский, румынский
ISO-8859-5: русский
ISO-8859-6: арабский
ISO-8859-7: греческий
ISO-8859-8: иврит, иврит в визуальной кодировке
ISO-8859-9: турецкий
Windows-1250: чешский, венгерский, польский, румынский
Windows-1251: русский
Windows-1252: датский, голландский, немецкий, английский, французский, итальянский, норвежский, португальский, испанский, шведский
Windows-1253: греческий

Windows-1254: турецкий
Windows-1255: иврит
Windows-1256: арабский
KOI8-R: русский

Определение набора символов не всегда выполняется точно. Процесс определения кодовой страницы пытается найти набор символов, который лучше всего соответствует характеристикам байтовых данных, но этот процесс в определенной степени статистический и точность результата нельзя гарантировать.

Точность определения повышается, если входные данные в основном содержат текст на одном языке. Требуется как минимум несколько сотен байт текстовых данных на этом языке.

Если обнаруженная кодировка не поддерживается, система использует кодовую страницу по умолчанию для собрания.

Лингвистический анализ документов на китайском, японском и корейском языках

Чтобы улучшить обработку документов, написанных на китайском, японском и корейском языках, можно задать опции лингвистического анализа.

Можно указать анализатору, чтобы для лексического анализа документов на китайском, японском и корейском языках он использовал метод сегментации энграмм. Кроме того, можно сконфигурировать анализатор таким образом, чтобы для документов на китайском и японском языках он удалял символы новой строки из пустого пространства.

Сегментация энграмм

При создании собрания вы выбираете тип лексического анализа для анализа документов на языках, где для разделения слов не используются пробельные символы.

При сегментации на основе пробелов Unicode в качестве разделителей между словами используются пробельные символы. При использовании сегментации энграмм перекрывающиеся последовательности из любого количества символов рассматриваются как отдельные слова. Для языков типа китайского, японского или корейского, в которых не используются пробельные символы в качестве разделителей слов, сегментация энграмм может дать более хорошие результаты при поиске, чем сегментация на основе пробелов Unicode.

Метод сегментации, используемый для анализа документов, задается при создании собрания. После создания собрания можно узнать метод сегментации (в опциях синтаксического анализа), но нельзя изменить его.

Сведения о том, как сконфигурировать поддержку полного анализа и разбиения энграмм в собраниях поисковой системы предприятия и как обрабатываются символы в собраниях с полной поддержкой энграмм, смотрите в документе <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27011088>.

Удаление пробелов из текста

Синтаксический анализатор можно сконфигурировать для удаления из текста пробельных символов.

Прежде чем приступить к работе...

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Если включить эту опцию для собрания, синтаксический анализатор будет удалять последовательности пробельных символов, разделяющие двухбуквенные символы. Например, пробельные символы может потребоваться удалить для документов, написанных на языке, в котором слова не разделяются пробелами, например, на китайском или японском.

При конфигурировании для синтаксического анализатора удаления пробельных символов можно задать, следует ли удалять пробельные символы только между символами набора двухбайтных символов (DBCS) или независимо от контекста символов. Последнюю опцию можно использовать, например, если вы включаете английский текст в документ на японском и хотите удалить пробельные символы также и из английского текста.

Синтаксический анализатор удаляет следующие символы:

- Табуляция (0x09)
- Символ перевода строки, LF (0x0A)
- Символ возврата каретки, CR (0x0D)

Чтобы изменения вступили в силу, остановите и перезапустите анализатор. Чтобы изменения были применены для документов, которые уже сохранены в индексе, выполните их повторный просмотр искателем и полное построение индекса.

Процедура

Чтобы удалить пробелы из текста:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.

2. При помощи текстового редактора отредактируйте следующий файл (*ID_собрания* - это ID, заданный для собрания или назначенный системой при создании собрания):

```
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.parserdriver/  
collection.properties
```

3. Задайте способ удаления пробелов:

- Чтобы удалять пробелы между символами DBCS, задайте для свойства `removeCjNewLineChars` значение `true`:
`removeCjNewLineChars=true`
- Чтобы удалять пробелы из любых мест в документах, задайте для свойства `removeCjNewLineChars` значение `true`, а для свойства `removeCjNewLineCharsMode` значение `all`:

```
removeCjNewLineChars=true  
removeCjNewLineCharsMode=all
```

Управление индексами

Чтобы у пользователей поисковой системы предприятия всегда был доступ к самой свежей информации, для всех собраний создаются индексы, и их содержимое периодически изменяется.

Чтобы в данных, собранных искомателями, можно было выполнять поиск, необходимо создать индексы. При первом создании собрания поисковая система предприятия создает индекс для всех данных, ранее собранных искомателями. Когда искоматели обрабатывают новые и измененные источники данных, создается обновление (оно называется *разностным индексом*) для нового содержимого. В конечном счете эти обновления необходимо объединить с основным индексом. Этот процесс объединения называется построением *полного индекса*. При построении разностного или полного индекса новое содержимое копируется на поисковые серверы и становится доступным для поиска.

Искоматели собирают данные непрерывно или периодически в соответствии с расписанием. Если индекс обновляется часто, это позволяет пользователям выполнять поиск в наиболее свежих данных. Однако постоянно обновляемый индекс надо время от времени строить полностью. Чем больше становится индекс, тем больше он потребляет системных ресурсов. Для поддержания оптимальной производительности выполняйте полное построение регулярно.

На то, как часто необходимо выполнять полное построение индекса, влияют:

- Системные ресурсы (пространство в файловой системе, скорость процессора и оперативная память)
- Сколько документов должно быть просмотрено и перепросмотрено искомателями
- Тип данных, которые будут обрабатываться искомателями
- Как часто вы изменяете правила категорий (изменения вступают в действие только после полного построения индекса)
- Как часто вы запускаете искоматель принудительно (вместо запуска в запланированное время)

Для собраний с несколькими миллионами документов, построенных преимущественно из Web-документов, следует выполнять полное построение индекса примерно раз в день, а дельта-построения - каждые 1-2 часа.

Для поддержания оперативного индекса, в котором можно выполнять поиск, необходимо выполнять следующие задачи:

- Задавать расписания для построения индекса
- Изменять расписания для индексов
- Включать и отключать расписания для индексов
- Конфигурировать одновременное построение индексов

Для задания опций, влияющих на вид индекса для пользователя, можно выполнять следующие задачи:

- Конфигурировать поддержку символов подстановки в запросах
- Конфигурировать области действия для ограничения диапазона документов, в которых пользователи могут выполнять поиск
- Объединять в результатах поиска документы из одного источника

- Удалять URI из индекса

Задачи, связанные с данной

“Отслеживание активности индекса для собрания” на стр. 318

“Слежение за очередью индексации поисковой системы предприятия” на стр. 319

Планирование операций построения индекса

Можно запланировать построение полного индекса и обновление индекса с занесением в него новых данных.

Прежде чем приступать к работе...

Запланировать построение индекса для собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Чтобы быть уверенным, что пользователи всегда обращаются к самой свежей информации в источниках при поиске, запланируйте регулярное построение индекса. При построении полного индекса он строится полностью. Процессы индексирования считывают все данные, собранные искателями и проанализированные анализаторами. При дельта-построении индекса становится доступной для поиска информация, обработанная искателем со времени последнего полного построения индекса.

По умолчанию выбрана опция планирования построения индекса. Эта опция сообщает процессу планировщика, что задачи полного построения и дельта-построения индексов надо запланировать на момент запуска поисковой системы предприятия. Можно отключить переключатель **Включать при запуске системы** в любой момент, если вы хотите отменить запуск запланированного построения индекса. Например, может потребоваться отключить планирование для диагностики возникших проблем.

Для экономии ресурсов системы и повышения производительности система автоматически определяет, не произошли ли изменения, требующие внесения поправок в индексы. Если нет изменений, влияющих на индекс, запланированное требование построения отклоняется.

Процедура

Чтобы запланировать построение индекса:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Планировать построения индекса**.
2. Чтобы задать, как часто надо обновлять индекс, добавляя в него новое содержимое, задайте на странице Планирование построения индексов в области **Задайте расписание дельта-построения индекса** следующие опции:
 - a. В области **Начиная с** в полях **Год**, **Месяц**, **День**, **Час** и **Минута** задайте время первого дельта-построения индекса.
 - b. В области **Интервал обновления** в полях **дни**, **часы** и **минуты** задайте частоту дельта-построения индексов.

Как правило, дельта-построения индексов следует выполнять часто, например, каждый час или два. В зависимости от того, как часто изменяется содержание

источника, задайте больший или меньший промежуток времени. Например, можно задать обновление каждый час (0 дней и 1 час) или каждые 12 часов (0 дней и 12 часов).

3. Чтобы задать, как часто следует выполнять полное построение индекса, определите в области **Задайте расписание полного построения индекса** следующие опции:
 - a. В области **Начиная с** в полях **Год, Месяц, День, Час и Минута** задайте время первого полного построения индекса.
 - b. В области **Интервал обновления** в полях **дни, часы и минуты** задайте частоту полного построения индекса.Как правило, полное построение следует выполнять регулярно, например, каждые 24 часа. В зависимости от того, как часто изменяется содержание источника, задайте больший или меньший промежуток времени. Например, можно задать построение каждые 12 часов (0 дней и 12 часов) или каждые два с половиной дня (2 дня и 12 часов).
4. Нажмите кнопку **ОК**.

Изменение расписания индексирования

Можно изменить расписание построения индекса.

Прежде чем приступать к работе...

Изменить расписание индексирования может администратор поисковой системы предприятия или администратор данного собрания.

Процедура

Чтобы изменить расписание индексирования:

1. Вызовите редактирование собрания, выберите страницу **Индекс** и измените соответствующие значения в полях **Месяц, День, Год и Час**. Укажите, как часто будет выполняться дельта-построение индекса и полное построение индекса.
2. Нажмите кнопку **Применить**.

Включение и выключение расписаний индекса

Можно включать и выключать расписания для построения индекса.

Прежде чем приступать к работе...

Включать или выключать расписание индексирования может администратор поисковой системы предприятия или администратор данного собрания.

Об этой задаче

Выключить расписание для индекса можно, если требуется, чтобы запланированное построение этого индекса не запускалось. Например, может понадобиться отключить расписание, чтобы построение индекса не запускалось в запланированные дату и время, если в это время требуется исправление ошибок.

Включать и выключать расписание можно при редактировании собрания, а также при слежении за собранием.

Процедура

1. Чтобы включить или выключить расписание для индекса при редактировании собрания, выполните следующие действия:
 - a. Вызовите редактирование собрания, которое вы хотите изменить.
 - b. На странице Индекс включите или выключите переключатель **Включать при запуске системы**, чтобы включить или выключить расписание для обновления индекса.
 - c. Включите или выключите переключатель **Включать при запуске системы**, чтобы включить или выключить расписание для полного построения индекса.
 - d. Нажмите кнопку **Применить**.
2. Чтобы включить или выключить расписание для индекса при слежении за собранием, выполните следующие действия:
 - a. Включите слежение за собранием, которое вы хотите изменить.
 - b. На странице Индекс, если индекс внесен в расписание и вы не хотите запускать построение в указанные дату и время, нажмите кнопку **Отключить расписание**. Построение индекса не будет выполняться до тех пор, пока вы не включите расписание или не нажмете кнопку **Пуск** для запуска построения индекса.
 - c. Если индекс внесен в расписание, но расписание для построения отключено, нажмите кнопку **Включить расписание**.
Индекс будет поставлен в очередь на построение на дату и время, указанные в расписании индексирования.

Конфигурирование одновременных операций построения индекса

Управляя использованием ресурсов индексации, вы задаете, для скольких собраний могут одновременно обрабатываться требования построения индекса. При достаточных ресурсах системы можно повысить качество поиска, разрешив операции по дельта-построению индексов во время полного построения индекса.

Прежде чем приступить к работе...

Задавать опции построения индексов для системы может только администратор поисковой системы предприятия.

Об этой задаче

Поисковая система предприятия может создавать несколько индексов одновременно, разделяя ресурсы между собраниями, что позволяет параллельную обработку требований построения индекса для нескольких собраний. Путем совместного использования процессов можно добиться того, что построение очень большого индекса не будет блокировать доступность других индексов, ожидающих построения в очереди.

Когда затребовано или запланировано построение индекса, этот индекс ставится в очередь индексирования, где и дожидается обработки. Так как у каждого собрания есть собственный индекс, одновременно в очереди индексирования может находиться несколько запросов на построение индекса. При настройке опций индексации для системы вы задаете, сколько собраний может совместно использовать ресурсы индексации для параллельной обработки требований.

Кроме того, можно задать обработку требований обновления индексов на время полного построения индекса для собрания. Если включить эту опцию, серверы поиска будут получать последние документы (посредством дельта-обновления индекса) во

время более длительного процесса полного построения индекса. С другой стороны, построение индекса требует большого количества ресурсов. Во время построения индекса используется большой объем системной памяти и дискового пространства. Если включить эту опцию при нехватке дискового пространства или памяти, возможно снижение производительности системы в целом.

Если увеличить число одновременных операций построения индекса, требования построения индексов, уже помещенные в очередь, не будут запущены автоматически. Это изменение затрагивает новые операции построения индексов, помещенные в очередь после изменения данного значения.

Если уменьшить число одновременных операций построения индекса, текущие операции построения индексов не будут остановлены автоматически. Это изменение вступит в силу после остановки текущих операций построения индексов, что позволит запустить операции построения индексов, помещенные в очередь.

Процедура

Чтобы задать опции построения индекса для системы:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Индекс нажмите кнопку **Конфигурировать опции индексирования**.
4. На странице Опции индексации уровня системы введите число собраний, для которых разрешено совместное использование ресурсов системы и параллельная обработка требований построения индекса.

Число собраний, совместно использующих ресурсы индексации, не может превышать число собраний в вашей поисковой системе предприятия. Например, если у вас есть пять собраний, нужно ввести число не больше пяти.

5. Если ресурсов системы хватает для поддержки нескольких параллельных операций построения индексов для отдельных собраний, можно включить опцию, разрешающую выполнение операций дельта-построения индексов параллельно полному построению индекса.

Ограничение: Если вы выбрали эту опцию в момент, когда выполнялось построение индекса для какого-либо собрания, это построение может закончиться неудачно. Прежде чем разрешать одновременное построение разностного и главного индексов, отключите все запланированные построения индексов и дождитесь окончания уже запущенных построений (или прервите их до завершения, если возможно). После включения этой опции снова включите отключенные запланированные построения индексов.

Построение индексов только при обнаружении изменений

Если построение индекса запланировано, оно выполняется только при обнаружении изменений, которые необходимо внести в индекс. С помощью команды **startIndexBuild** можно запустить построение индекса и выбрать, должна ли система проверить наличие изменений.

Прежде чем приступать к работе...

Чтобы вручную задать, что к построению индекса следует приступить только при обнаружении изменений, нужно зарегистрироваться как администратор поисковой системы предприятия.

Ограничения

Система не проверяет наличие изменений, которые могут повлиять на индекс, пока минимум дважды не будет выполнено полное построение индекса. Эта реализация вносит изменения (такие как изменения определений области действия или правил объединения результатов поиска), которые нельзя обнаружить, пока полное построение индекса не будет выполнено хотя бы один раз.

Об этой задаче

Построения индексов расходуют системные ресурсы, такие как процессорное время, память и пропускная способность ввода-вывода. В целях сбережения системных ресурсов и улучшения производительности поисковая система предприятия может определить наличие изменений, которые необходимо внести в индекс, и выполнять построение индекса, только если оно необходимо для внесения изменений. Система может определить следующие события наличия изменений для индекса:

Только полные построения индекса

- Удаление URI по шаблону, при котором удаляются документы из индекса.
- Удаление из системы искателя, при котором удаляются документы из индекса.

Полные построения и дельта-построения индекса

- Изменение определений области действия или правил объединения результатов поиска.
- Добавление или удаление документов непосредственно при помощи параметров искателя или API push-обработки.
- Изменение коэффициентов значимости, влияющих на оценки статического ранжирования документов в индексе.
- Изменение правил использования символов подстановки в запросах.

Обнаружение изменений применяется ко всем запланированным требованиям построений индекса. Построение индекса выполняется только при наличии отложенных изменений, которые необходимо внести в индекс. Если изменения не обнаружены, требование на построение индекса отклоняется.

Принудительное построение индекса, независимо от наличия изменений, можно запустить с консоли администратора поисковой системы предприятия. При помощи команды **startIndexBuild** можно также запустить построение индекса вручную и указать, должна ли система проверить наличие изменений перед тем, как перейти к обработке требования.

Процедура

Чтобы запустить построение индекса из командной строки и затребовать, чтобы оно было запущено, только если будут обнаружены изменения:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. В случае конфигурации с несколькими серверами зарегистрируйтесь на сервере индексов.
2. Введите команду:

```
esadmin controller startIndexBuild -опции
```

Опции:

-cid *ID_собрания*

ID собрания, для которого вы хотите запустить построение индекса.

-buildType *тип_построения*

Задаёт тип построения индекса. Допустимые значения: `main` (полное) и `delta` (дельта-построение).

-detectChanges

Необязательный параметр. Указывает подвергнуть требование построения индекса проверке на обнаружение изменений. Обработка построения индекса начнется только при обнаружении изменений, которые необходимо внести в индекс.

В следующем примере требование запустить полное построение индекса для собрания `col_1` будет обработано только при обнаружении изменений, которые необходимо внести в индекс:

```
esadmin controller startIndexBuild -cid col_1 -buildType main
-detectChanges
```

Остановка построения индексов

Для остановки построения главного индекса и дельта-индексов вместо консоли администратора поисковой системы предприятия можно использовать команду **stopIndex**.

Процедура

Чтобы остановить построение не с консоли администратора, а при помощи команды:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. Для установки с несколькими серверами зарегистрируйтесь на сервере индексов.
2. Введите команду:

```
esadmin stopIndex -cid id_собрания -buildType тип Где:
```

-cid *ID_собрания*

Задаёт ID собрания, к которому относится индекс.

тип

Задаёт тип построения индекса, которое надо остановить. Допустимые значения - `main` или `delta`.

Пример

```
esadmin stopIndex -cid col1 -buildType delta
```

Опции, влияющие на представление индекса для поиска

После индексации документов вы можете задавать опции, управляющие возможностями пользователей по поиску документов и просмотру найденных документов.

Чтобы задать опции, влияющие на представление индекса для пользователя, можно выполнить следующие задачи:

- Сконфигурировать поддержку терминов запроса с символами подстановки. Можно встроить поддержку запросов с символами подстановки в индекс, а можно задать опции расширения терминов запроса при обработке запроса.

- Задать области действия, чтобы ограничить диапазон документов, в которых пользователи могут выполнять поиск. Когда пользователи выполняют поиск в этом собрании, поиск выполняется только в документах из области действия, а не по всему индексу.
- Объединять найденные документы из одного источника. Можно группировать документы, соответствующие некоторому URI или шаблону URI в индексе, и показывать только первые найденные документы (пользователи смогут задавать опции, чтобы увидеть объединенные документы).
- Удалять URI из индекса. Иногда нужно предотвратить поиск по некоторым документам в индексе.

Для некоторых типов искателей и для собраний, в которых не включена защита, используется обнаружение дубликатов документов, чтобы не показывать пользователям в результатах поиска одинаковые или почти одинаковые документы.

Индексированные опции для поиска документов

При конфигурировании опций для поиска в данных искателя или при отображении элементов метаданных XML и HTML на поля поиска можно задать, как выполнять поиск документов и как выводить документы в результатах поиска.

Заданные вами опции поиска хранятся в индексе с документами. Они позволяют ограничить данные, доступные пользователям для запроса, и данные, доступные им для просмотра в результатах поиска.

Опции искателя:

При конфигурировании искателей для поиска в источниках данных, содержащих поля, доступны следующие опции управления, задающие, можно ли выполнить поиск для поля, как его можно выполнить и может ли это поле быть возвращено в результатах поиска:

- Свободный текстовый поиск
- Поиск по полям
- Полное соответствие
- Возможность сортировки
- Параметрический поиск
- Результаты поиска
- Содержимое документа

Опции отображения полей XML и HTML:

Если при конфигурировании синтаксического анализатора задается отображение элементов XML и элементов метаданных HTML на доступные для поиска поля в индексе, можно задать следующие опции:

- Поиск по полям
- Полное соответствие
- Возможность сортировки
- Результаты поиска

Если вы конфигурируете опции поиска для конкретных элементов метаданных HTML, в противоположность всем элементам или же элементам, входящим в набор элементов метаданных Dublin Core, можно задать также возможность параметрического поиска для полей, содержащих числовые значения.

Свободный текстовый поиск

Индекс поисковой системы предприятия - это полнотекстовый индекс с содержимым из разных источников данных. Можно выполнить поиск по содержимому, введя простой запрос на естественном языке. Процесс поиска выполнит поиск по полям и содержимому, чтобы найти документы, соответствующие запросу.

Чтобы разрешить выполнение для полей простого текстового поиска, включите переключатель **Свободный текстовый поиск** при конфигурировании искателя. Для поиска по полям заголовков, ключевых слов и описаний включите этот переключатель и переключатель **Поиск по полям**.

Пример 1:

Свободный текстовый поиск может быть очень простым, например как следующий запрос:

велосипедная цепь

Чтобы указать, какие слова должны и какие не должны встречаться в документе, можно использовать специальную запись. Например, перед словом можно поставить знак плюс (+), это означает, что найденный документ должен содержать это слово. Если перед словом поставить знак минус (-), содержащие это слово документы будут исключены из результатов поиска. Для поиска точных словосочетаний заключите два или несколько слов в двойные кавычки ("").

Пример 2:

Следующему свободному текстовому запросу будут соответствовать только документы, содержащие точную фразу научная фантастика и не содержащие слова робот:

+ "научная фантастика" -робот

Поиск по полям

Поиск по полям позволяет ограничить область поиска определенными полями данных и полями метаданных в документе. Например, можно потребовать, чтобы определенные слова содержались в заголовке документа.

Чтобы разрешить поиск с указанием имен полей, включите переключатель **Поиск по полям** при конфигурировании искателя или при конфигурировании опций отображения полей для элементов XML и HTML. Для поиска по полям заголовков, ключевых слов и описаний включите этот переключатель и переключатель **Свободный текстовый поиск**.

Пример:

Чтобы воспользоваться поиском по полям в поиске предприятия, включите в ваш запрос имя поля и слово или словосочетание, которые должны содержаться в этом поле.

Следующий запрос ищет документы, содержащие слово `ibm` и словосочетание поисковая система предприятия в поле заголовка:

`title:ibm title:"поисковая система предприятия"`

Полное соответствие

Поиск полных соответствий может повысить качество результатов поиска, позволяя задавать точные запросы. Поиск полных соответствий позволяет выполнять запросы по полям и элементам XML и получать только те документы, в которых значение поля или значение элемента XML полностью соответствует условиям запроса. Если в значении поля содержится меньше или больше информации, соответствие не фиксируется.

Совет: Определяя полноту соответствия, система преобразует условия запроса в нижний регистр, удаляет лишние пробелы в строке запроса и выполняет поиск по шаблону с символами подстановки. Однако при этом не выполняется лемматизация и поиск синонимов, а стоп-слова не удаляются. Для полного соответствия элемента XML нужно, чтобы имя элемента (без вложенных элементов) и его значение точно соответствовали терминам запроса.

Поиск полных соответствий в полях выполняется с помощью синтаксиса запросов поисковой системы предприятия (SIAP) или синтаксиса запросов XMLFrag2 (запросы XPath не поддерживаются). Знак равенства (=) перед терминами запроса означает, что нужно выполнить поиск полных соответствий.

Чтобы разрешить поиск полных соответствий, включите переключатель **Полное соответствие** при конфигурировании искателя или при конфигурировании опций отображения полей для элементов XML и HTML. Если при конфигурировании опций синтаксического анализа для собрания разрешить пользователям выполнять поиск документов XML собственными средствами поиска XML, для всех элементов XML возможен поиск полных соответствий условиям запросов.

Пример 1:

Поле с именем color (цвет) содержит значение охра желтая.

- Это значение удовлетворяет следующему запросу полного соответствия, так как в этом запросе нет других терминов:

```
color:"охра желтая"
```

- Это значение не удовлетворяет следующему запросу полного соответствия, так как, кроме слов охра желтая, этот запрос содержит слово светлая:

```
color:"охра желтая светлая"
```

- Это значение не удовлетворяет следующему запросу полного соответствия, поскольку поле color, кроме содержащегося в запросе слова "охра", оно содержит также слово желтая:

```
color:охра
```

Пример 2:

Без полного соответствия следующий запрос XMLFrag2 может вернуть документы, содержащих термины <диагноз>сосудистый паркинсонизм акинетико-ригидного типа</диагноз> или другие термины, не полностью соответствующие условиям запроса:

```
@xmlf2: '<диагноз>сосудистый паркинсонизм</диагноз>'
```

При полном соответствии следующий запрос XMLFrag2 возвращает только документы, в которых все содержимое значения элемента XML соответствует условиям запроса:

```
@xmlf2: '<диагноз>сосудистый паркинсонизм</диагноз>'
```

Возможность сортировки

Если источник данных содержит поля или если пользователи выполняют поиск документов XML или HTML, вам может понадобиться сортировать результаты поиска по значениям в определенном поле. В примере программы поиска для поисковой системы предприятия перечислены имена всех полей, сконфигурированных как поля с возможностью сортировки. Пользователи могут сортировать результаты не только по релевантности или дате документов, но и в алфавитном порядке (при строчной сортировке), выбрав одно из полей в этом списке. При этом можно выбрать сортировку документов в восходящем либо нисходящем порядке.

Документы результатов без поля сортировки выводятся после результатов поиска. Документы результатов, содержащие поле сортировки, но проиндексированные до того, как это поле было сконфигурировано как поле сортировки, также выводятся в конце списка результатов.

Чтобы разрешить пользователям сортировать результаты поиска в алфавитном порядке по значениям поля, включите переключатель **Возможность сортировки** при конфигурировании искателя или при конфигурировании опций отображения полей для элементов XML и HTML. Если поле содержит числовые значения, включите переключатель **Параметрический поиск**, чтобы задать, что значения этого поля можно использовать для числовой сортировки результатов.

Параметрический поиск

Параметрический поиск - это тип поиска по полям, позволяющий выполнять сравнительные и вычислительные запросы для числовых полей, полей даты и метаданных. Например, можно выполнить поиск документов определенного размера или документов, созданных после определенной даты. Кроме того, можно выполнить поиск документов с атрибутами большими, меньшими или равными определенному значению.

Для числовой сортировки результатов по значениям поля надо разрешить для этого поля параметрический поиск.

Чтобы выполнить параметрический поиск по определенному полю или иметь возможность числовой сортировки результатов, включите переключатель **Параметрический поиск** при конфигурировании искателя или при конфигурировании опций отображения полей для конкретных элементов метаданных HTML.

Пример 1:

Следующий запрос находит объекты, которые стоят ровно 50 долларов (или 50 единиц той валюты, которая используется для поля цена):

```
#price:=50
```

Пример 2:

Следующий запрос находит документы, у которых размер файла превосходит 1024, но не превосходит 2048:

```
#filesize:>1024<=2048
```

Результаты поиска

Можно выполнить поиск в некоторых полях, но не выводить эти поля в результатах поиска, или вывести поле в результатах поиска, даже если это поле не запрашивалось. Например, вам может понадобиться запрос к финансовым данным, чтобы получить содержательный отчет, но вы не хотите выводить зарплаты сотрудников в результатах рядом с их фамилиями.

Чтобы разрешить вывод поля в результатах поиска, включите переключатель **Результаты поиска** при конфигурировании искателя или при конфигурировании опций отображения полей для элементов XML и HTML.

Содержимое документа

Для некоторых типов документов, например, документов Web, весь документ считается содержимым. Для других типов документов, например, содержащих поля, можно задать, в каких полях содержится полезный контент, в противоположность метаданным.

Чтобы задать, что в поле представлено содержимое документа, включите переключатель **Содержимое документа** при конфигурировании искателя. Если включены оба переключателя - **Содержимое документа** и **Свободный текстовый поиск**, значение поля будет использовано для выявления дубликатов документов и станет частью динамической области сводки документа в результатах поиска.

Понятия, связанные с данным

 Синтаксис запросов

Обнаружение дубликатов документов

Обнаружение дубликатов документов - технология, используемая для предотвращения получения документов с одинаковым или очень похожим содержимым.

Если результаты поиска будут содержать несколько копий одинаковых (или очень похожих) документов, качество поиска может ухудшиться. Дубликаты документов анализируются, только если выполняются оба следующих условия:

- Для собрания используется модель ранжирования на основе ссылок. Эта модель применяется к искателям, обрабатывающим сайты, например, к искателю Web или WebSphere Portal.
- Отключена защита собрания.

При глобальном анализе процесс индексации обнаруживает дубликаты путем просмотра содержимого каждого документа. При наличии двух документов с одинаковым содержимым они воспринимаются как дубликаты.

Если вы хотите, чтобы при анализе обнаруженных повторений рассматривались также и метаданные документов, при конфигурировании искателей для собрания нужно включить переключатель **Содержимое документов** и задать опции для просмотра искателем метаданных. Тогда искатель будет обрабатывать поля метаданных как содержимое документов и включать метаданные в анализ содержимого повторяющихся документов. Похожий анализ выполняется после конфигурирования вами опций синтаксического анализа документов HTML и XML и включения переключателя **Содержимое документов**.

Если указать, что поле (или поле метаданных) составляет содержимое документа, содержимое этого поля будет добавлено в динамическую сводку документа в результатах поиска, что может повлиять на включение документа в вывод результатов. Если включить в программе поиска определение близких по содержимому документов (задать для свойства `NearDuplicateDetection` в методе `setProperty` значение `Yes`), вывод документов со сходными названиями и сводками будет подавляться при просмотре пользователем результатов поиска. Чтобы просмотреть близкие по содержанию документы, вывод которых был подавлен, пользователи могут щелкнуть по соответствующей ссылке.

В группе повторяющихся документов один документ считается главным, а остальные - дубликатами. У всех документов в группе повторяющихся документов одно и то же каноническое представление содержимого. При индексации обрабатывается содержимое (маркеры) главного документа. Для дубликатов индексируются только маркеры метаданных. При удалении главного документа из индекса новым главным становится следующий дубликат. При выполнении пользователями поиска в собрании возвращается только главный документ.

Понятия, связанные с данным

“Анализ повторных документов и защита собрания” на стр. 264

Символы подстановки в запросах

Можно разрешить пользователям включать в термины поиска символы подстановки и искать слова, соответствующие определенным шаблонам.

Термин запроса с символами подстановки - это термин, содержащий символ звездочки (*). Когда пользователь передает запрос с символами подстановки, в результаты поиска включаются все документы в индексе, в которых есть термин запроса, и все документы в индексе, в которых есть слова, соответствующие шаблону с символом подстановки. Например, символ подстановки на конце термина запроса `мор*` может соответствовать словам `море`, `морковь` и `мораль`.

При конфигурировании опций символов подстановки для индекса выбирают, разрешать ли пользователям задавать символы подстановки в запросах, и, если символы подстановки разрешены, как должна осуществляться поддержка символов подстановки:

- Можно разрешить поиск слов по шаблону с символами подстановки во всех частях документа, а можно ограничить поиск по шаблону только полями.
- Можно разрешить поддержку запросов с символами подстановки для всех полей, а можно ограничить поиск по шаблону заданными полями.
- Можно задать, что символ подстановки разрешается только в последней позиции термина запроса (конечный символ подстановки), а можно разрешить символ подстановки в любом месте термина запроса. (В имени поля символ подстановки не допускается.)
- В зависимости от разрешенных мест для символов подстановки можно выбрать варианты расширения терминов запроса (термины запросов с символами подстановки расширяются до всех соответствующих терминов из индекса). Можно хранить все возможные версии расширения терминов в индексе, а можно выполнять расширение терминов при обработке запроса.

Любые изменения параметров символов подстановки начинают действовать при следующем полном построении индекса.

Расширение в индексе

Чтобы включать расширения терминов в индекс, нужно задать, сколько начальных символов слова должны совпадать с символами шаблона в термине запроса, чтобы слово считалось соответствующим этому шаблону. Результативными будут только те термины запроса, которые содержат не меньшее число символов (не считая *). Например, если задать 4, для фиксации совпадения термин запроса должен содержать минимум четыре символа.

Если ввести 4, слово технология будет соответствовать терминам запроса техн* и техно*, но не те*.

При построении полного индекса или дельта-индекса индексируются не только исходные термины, но и все возможные расширения каждого термина в документах. Преимущество этого подхода в том, что при обработке запроса дополнительное время на расширение терминов не требуется. Однако при этом увеличивается размер индекса, поэтому нужно иметь достаточный объем системных ресурсов для индекса большего размера.

Такой подход лучше использовать для собраний относительно небольшого размера или когда пространство и время, затрачиваемые при построении индекса, не столь важны, как время ответа на запрос. Например, этот подход можно выбрать для поиска в каталоге или списке работников.

Данный подход применим только тогда, когда выбрана поддержка символов подстановки на конце слова. Если включить поддержку символов подстановки в любом месте термина запроса, опция, которая вставляет варианты расширения терминов в индекс, будет недоступна.

Расширение в запросе

Чтобы расширение терминов запросов и применение правил соответствия шаблону происходили при передаче пользователем запроса, содержащего символы подстановки, нужно задать, сколько вариантов термина запроса образуют множество терминов для поиска. Например, если задать число 50, в качестве терминов для поиска могут использоваться до 50 расширений термина запроса.

Так термину запроса техн* будут соответствовать слова технический, техника, технология и другие слова (не более 50 в нашем примере), начинающиеся с техн.

Расширение в запросе слабо влияет на размер индекса, но может снизить производительность запросов. Процессы поиска должны проверить все возможные расширения термина запроса с символами подстановки вплоть до предельного числа таких расширений, заданных в параметрах символов подстановки.

Такой подход лучше использовать для собраний относительно большого размера, а также когда необходимо минимизировать пространство и время, затрачиваемые при построении индекса. Например, этот подход можно выбрать для репозитория электронной почты, для которых индекс должен постоянно соответствовать быстро меняющемуся набору документов, а время ответа на запрос не столь критично.

Данный подход применим независимо от того, включена ли поддержка только конечных символов подстановки или же символов подстановки в любом месте термина запроса.

Поддержка символов подстановки в запросах

Набор расширений термина запроса с символами подстановки содержит все термины индекса, которые можно получить заменой символа подстановки на произвольную последовательность символов. Этот набор определяется так:

- Если собрание поддерживает символы подстановки в любом месте термина запроса, всякий термин запроса со звездочкой интерпретируется как термин с символом подстановки.
- Набор содержит не больше максимального числа расширений, сконфигурированного администратором поисковой системы предприятия. Если в индексе есть расширения сверх этого числа, они игнорируются. (В результатах поиска указывается, если некоторые расширения символов подстановки были проигнорированы.)
- Если поддержка символов подстановки ограничивается набором полей, набор расширений содержит только термины, которые есть в одном из заданных полей. Достаточно присутствия термина в одном поле одного документа данного индекса.
- Если термин запроса - термин поля, символ подстановки должен находиться после спецификатора поля (например, имя_поля:*сфера). Имя поля не может содержать двоеточие (:).
- Если поддержка символов подстановки ограничивается набором полей, имя поля в термине запроса с символами подстановки должно быть одним из имен, заданных на консоли администратора поисковой системы предприятия. В противном случае для этого термина никакие расширения найдены не будут.
- Символы подстановки поддерживаются только в терминах с простым текстом, но не в именах элементов XML, именах атрибутов или значениях атрибутов. Не поддерживается термин, состоящий из одного лишь символа подстановки.

Как символы подстановки влияют на индекс

Если поддержка символов подстановки выполняется путем расширения индекса, размер индекса и время его построения увеличивается.

При использовании расширения индекса в индекс вносятся все начальные подстроки термина и сам термин. Например, для термина культура индексируются следующие термины:

к ку кул куль культ культу культур культура

Число терминов, хранящихся в индексе, увеличивается пропорционально средней длине слова. Сжатие индекса сокращает его размер, но незначительно. Время, необходимое на построение индекса, увеличивается соответственно средней длине слов.

Размер индекса для документов на английском языке увеличивается примерно в четыре раза, поскольку средняя длина английских слов составляет пять-шесть символов. Размер индекса с лексическим анализом энграмм увеличивается приблизительно в два раза, поскольку энграмма состоит из двух символов.

Расширение индекса рекомендуется для поддержки символов подстановки в следующих ситуациях:

- Собрание так мало, что увеличение пространства и времени, связанное с расширением индекса, не приводит к проблемам с производительностью.
- Включение всех возможных расширений символов подстановки в результаты поиска необходимо для выполнения требований пользователя (или предприятия).

Влияние поддержки символов подстановки на индекс можно смягчить, задав для расширений символов подстановки минимальную длину начальной подстроки. Например, если задать минимальную длину префикса три, префиксы k и ku для слова культура не будут вноситься в индекс. Для английских документов такое ограничение приводит к росту объема не в четыре, а в три раза.

Если для поддержки символов подстановки используется метод расширения запросов, никакие подстроки в индекс не вносятся. Термины расширяются при передаче запроса, и размер индекса увеличивается только за счет небольшой структуры данных, необходимой для поддержки расширения. Как правило, размер индекса с расширением запросов на 10 - 20% больше размера индекса без поддержки символов подстановки, а время его построения увеличивается не более, чем на 10%. Задание максимального числа расширений ни на размер, ни на время построения индекса не влияет.

Конфигурирование опций для символов подстановки в запросах

При конфигурировании опций индексирования для собрания поисковой системы предприятия можно указать, разрешить ли пользователям включать в условия запросов символы подстановки.

Прежде чем приступить к работе...

Чтобы сконфигурировать опции для символов подстановки, надо быть администратором поисковой системы предприятия или администратором собрания, к которому принадлежит индекс.

Об этой задаче

Изменения, вносимые при задании опций для символов подстановки, вступают в силу при следующем полном построении индекса.

Процедура

Чтобы сконфигурировать поддержку символов подстановки в запросах:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Конфигурировать опции для символов подстановки**.
2. На странице Опции для символов подстановки включите переключатель **Поддерживать символы подстановки в запросах**.
3. Необязательно: Можно включить поддержку символов подстановки в запросах с поиском в свободном тексте. Например, свободный текстовый запрос техн*, не выполняющий поиск в явно заданном поле, будет возвращать развернутые результаты (такие как технология или техника), только если включен этот переключатель.
4. Задайте, в каких полях будут поддерживаться символы подстановки:
 - Чтобы отключить обработку символов подстановки для запросов с поиском в полях, выберите **Без полей**.
 - Чтобы включить поддержку запросов, содержащих символы подстановки, для всех полей в документе, выберите **Все поля**.
 - Чтобы ограничить поддержку символов подстановки для некоторых полей, выберите **Конкретные поля**, а затем введите нужные имена полей. Развернутые результаты будут возвращаться только для заданных здесь полей. Например, запрос author:john* будет возвращать развернутые результаты, только если задать поддержку символов подстановки для поля author.

5. Задайте, должен ли символ подстановки стоять только в конце термина запроса (концевой символ подстановки) или же он может применяться без ограничений, в любом месте термина запроса.

При выборе позиции и типа символа подстановки нужно также задать способ включения поддержки символов подстановки. Для вывода подробностей нажмите на консоли администратора кнопку **Справка**.

Области действия

Если нужно показать пользователям сокращенное представление собрания, сконфигурируйте область действия.

Область действия - это группа связанных URI в индексе. При конфигурировании области действия вы ограничиваете видимые для пользователей документы собрания. Когда пользователи выполняют поиск в этом собрании, поиск выполняется только в документах из области действия, а не по всему индексу. Чтобы использовать данную возможность, ваши программы поиска должны включать поддержку областей поиска.

Создавая область действия, вы задаете диапазон URI в индексе, в котором пользователи смогут выполнять поиск. Ограничение документов, в которых пользователи могут выполнять поиск, повышает вероятность того, что документы в результатах поиска будут соответствовать запросам пользователей.

Например, можно создать одну область действия, включающую в себя URI для отдела технической поддержки, и другую область действия с URI для отдела кадров. Если ваши поисковые программы поддерживают области действия, пользователи в отделе технической поддержки будут получать документы из области действия для отдела технической поддержки, а пользователи из отдела кадров - из области действия для отдела кадров.

Можно создать любое число областей действия, хотя создание слишком большого числа областей действия может повлиять на производительность. Конфигурируйте области действия таким образом, чтобы большинство требований поиска фильтровалось только по одной или двум областям действия. Поскольку области действия могут содержать целые URI или шаблоны URI, один и тот же документ может принадлежать к нескольким областям действия.

После конфигурирования областей действия может потребоваться дважды выполнить полное построение индекса, чтобы изменения вступили в силу. Если сконфигурировать области действия перед первым полным построением индекса для собрания, пользователи смогут выполнить поиск в собрании, но не смогут увидеть данные областей действия в результатах поиска. Чтобы результаты поиска гарантированно отражали диапазон URI в области действия, нужно еще раз выполнить полное построение индекса.

Если сконфигурировать области действия после полного построения индекса, изменения вступят в силу после следующего полного построения индекса.

Конфигурирование областей действия

При конфигурировании области действия для собрания поисковой системы предприятия задаются URI или шаблоны URI, определяющие документы в индексе, для которых пользователям разрешено выполнять поиск.

Прежде чем приступить к работе...

Конфигурировать области действия может только администратор поисковой системы предприятия или администратор собрания, к которому принадлежат эти области действия.

Об этой задаче

Если ваши прикладные программы поиска поддерживают области действия, пользователи смогут выполнять поиск только в документах, соответствующих заданным URI (эти URI определяют границы области действия для поиска в собрании).

После конфигурирования областей действия может потребоваться дважды выполнить полное построение индекса, чтобы изменения вступили в силу. Если сконфигурировать области действия перед первым построением индекса для собрания, пользователи смогут выполнить поиск в собрании, но не смогут увидеть данные областей действия в результатах поиска. Чтобы результаты поиска гарантированно отражали диапазон URI в области действия, нужно еще раз выполнить полное построение индекса.

Если сконфигурировать области действия после полного построения индекса, изменения вступят в силу после следующего полного построения индекса.

Процедура

Чтобы сконфигурировать область действия:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать области действия**.
2. На странице Области действия выберите **Создать область действия**.
3. Задайте имя для области действия и URI и шаблоны URI, определяющие границы этой области действия. Можно также задать URI и шаблоны URI, исключаемые из области действия.
4. Нажмите кнопку **ОК**.

Новая область действия появится в списке на странице Области действия вместе с другими областями действия для этого собрания.

Ссылки, связанные с данной

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

Сворачивание URI

Поисковая система предприятия может расположить результаты поиска так, чтобы документы из источников с одним префиксом URI выводились в результатах поиска одной свернутой группой.

При объединении результатов первый результат обычно выводится по левому краю страницы. Другие результаты с более низким рангом группируются и выводятся под первым результатом со сдвигом вправо.

Для объединения в одну группу документов результатов с разными префиксами URI можно связать префиксы URI с именем созданной вами группы. Например, если у вас есть три сервера для управления финансовыми данными, можно сгруппировать в результатах поиска документы со всех трех серверов и разместить результаты с более низким рангом под результатами с более высоким рангом.

Поисковые прикладные программы могут объединять документы в результатах поиска с помощью префиксов URI или имен групп. В поисковой программе примера

поисковой системы предприятия показаны два верхних документа - результата поиска. Если возвращено более двух документов с одним префиксом URI (или относящихся к одной группе URI), можно выбрать опцию для просмотра объединенных результатов.

Пользователи могут использовать синтаксис запросов поисковой системы предприятия (`samegroupas:префикс URI`) для поиска всех документов из группы с указанным в запросе префиксом URI.

Как организовать префиксы URI и имена групп

Используя консоль администратора для конфигурирования правил объединения результатов поиска, вы задаете префиксы URI документов, которые хотите объединять; кроме того, можно связать префиксы URI с именем группы.

Порядок, в котором заданы префиксы URI, имеет значение. Сервер индекса использует порядок префиксов URI при вычислении значения каждого URI в собрании. Для каждого URI:

1. Сервер индекса последовательно просматривает префиксы URI в правилах объединения результатов поиска.
2. Когда сервер индекса находит первый префикс URI, соответствующий префиксу документа в индексе, он добавляет имя группы (или префикс URI, если в правиле не задано имя группы) в качестве дополнительного термина поиска для документа.

Если для web-документа не найден соответствующий префикс URI, сервер индекса использует в качестве префикса URI имя хоста URL. Если для документа NNTP не найден соответствующий префикс URI, сервер индекса использует в качестве префикса URI ID первого сообщения в значении заголовка ссылки.

Добавив префикс URI в список префиксов для объединения результатов поиска, нужно поместить этот префикс URI в место списка, соответствующее желаемому порядку, в котором сервер индекса будет проверять его и, возможно, задавать в качестве дополнительного термина поиска для документов в индексе:

- Если вы добавили префикс URI и не связали его с именем группы, можно выбрать этот отдельный префикс URI и переместить его вверх или вниз в списке.
- Если вы добавили префикс URI и связали его с именем группы, при перемещении этого префикс URI вверх или вниз в списке перемещается вся группа префиксов URI. Порядок префиксов URI внутри группы не имеет значения; при выборе отдельного префикса URI автоматически выбирается вся группа.

Объединение URI в результатах поиска

Можно задать опции для группировки и объединения документов результатов из источников с одинаковым префиксом URI. Можно также задать имя группы для объединения полученных документов с разными префиксами URI.

Прежде чем приступить к работе...

Чтобы задавать опции для объединения результатов поиска, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Об этой задаче

Изменения, внесенные в объединении результатов поиска, вступят в действие только после следующего полного построения индекса.

Процедура

Чтобы задать опции для объединения результатов поиска:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Объединить результаты поиска**.
2. На странице Объединить результаты поиска нажмите кнопку **Добавить префикс URI**.
3. На странице Добавить префикс URI для объединения результатов введите префикс URI для документов, которые вы хотите объединять в результатах поиска. Например:

```
http://finance/ROI/  
http://server1.com/finance/  
db2://LOCALDB/SCHEMA1.TABLE1/  
exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/
```
4. Можно ввести описательное имя группы, которое вы хотите связать с этим префиксом URI. Чтобы объединить в одну группу документы, найденные в разных источниках, при добавлении каждого из префиксов URI вводите одно и то же имя группы.
5. Нажмите кнопку **ОК**.
6. На странице Объединить результаты поиска разместите новое правило в соответствии с требуемым порядком его просмотра сервером индексов:
 - Если вы добавили префикс URI и не связали его с именем группы, этот новый префикс URI появится в нижней части списка. Переместите его в необходимое положение при помощи клавиш со стрелками.
 - Если вы связали новый префикс URI с именем группы, этот новый префикс URI появится в нижней части набора префиксов URI, принадлежащих к этой группе. Используйте клавиши со стрелками, чтобы переместить всю группу префиксов URI в необходимое положение.
7. Чтобы изменить префикс URI или имя группы, выберите этот префикс URI и нажмите кнопку  **Редактировать**.
8. Чтобы удалить префикс URI из списка, выберите префикс URI и нажмите кнопку  **Удалить**.

Удаление URI из индекса

Чтобы запретить пользователям поиск некоторых документов в собрании, можно удалить URI этих документов из индекса.

Прежде чем приступить к работе...

Для удаления URI из индекса надо быть администратором поисковой системы предприятия или администратором этого собрания.

Об этой задаче

Если указать полный URI, пользователи перестанут видеть этот URI в результатах поиска. Однако если пользователь повторяет один из предыдущих запросов, и результаты поиска для этого запроса находятся в кэше поиска, результаты с удаленным URI будут продолжать возвращаться в результатах поиска. Кэш поиска не обновляется, и этот URI не удаляется из индекса до следующей операции дельта-построения или полного построения.

Если для удаления нескольких URI вы указали шаблон URI, пользователи будут продолжать видеть URI, соответствующие этому шаблону, в результатах поиска до следующего полного построения индекса.

При удалении URI из индекса он не удаляется из пространства искателя. Когда искатель в следующий раз обнаружит этот документ, он снова впишет его в индекс, и документ станет доступен для поиска опять. Чтобы удалить URI из пространства для искателя, надо изменить правила искателя, исключив из них источник этого документа, после этого остановить и перезапустить искатель.

Процедура

Чтобы удалить из индекса URI для определенных документов:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Удалить URI из индекса**.
2. На странице Удалить URI из индекса введите URI (или шаблоны URI), которые хотите удалить из индекса.

Например:

```
http://domain.org/ru/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

Ссылки, связанные с данной

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

Управление сервером поиска

Для серверов поиска можно задать опции, управляющие: использованием пространства кэша для возвращаемых результатов поиска, максимальной длиной имен при выводе сводок документов в результатах поиска, связыванием пользовательских словарей для улучшения результатов поисков и возвращением в результатах поиска предварительно заданных URI при появлении в запросе определенных термов.

Когда пользователь передает запрос, серверы поиска используют индекс для быстрого обнаружения нужных документов. Чтобы получить для нужных документов метаданные, серверы поиска используют склад поисковой системы предприятия, содержащий данные синтаксического и лексического анализа. Метаданные могут содержать идентификатор URI, заголовков, описание, дату, тип данных документов, а также другую информацию.

При конфигурировании серверов поиска для собрания можно задать опции, управляющие обработкой запросов, включая опции, которые могут повлиять на производительность запросов:

Конфигурирование кэша поиска

Для оптимизации производительности запросов можно указать, что результаты поиска (ответы на запросы) должны помещаться в кэш, и сконфигурировать объем пространства, выделяемый для результатов поиска в кэше.

Конфигурирование максимальной длины вывода для сводок документов

Большинство документов результатов содержат сводку содержимого документа, которая помогает пользователям решить, нужен ли им полученный документ. Можно указать, какой объем пространства использовать в результатах поиска для вывода этой информации сводки.

Задание другого языка по умолчанию

Когда создается собрание, задается язык по умолчанию для поиска документов в этом собрании, но при необходимости можно указать другой язык.

Связывание пользовательских словарей

Если ваши разработчики прикладных программ создали пользовательские словари синонимов, стоп-слов или значимых слов, можно задать использование этих словарей при выполнении пользователями поиска в собрании.

Конфигурирование прямых ссылок

Можно заранее задать идентификаторы URI, возвращаемые для определенных ключевых слов и фраз. Если пользователи укажут эти ключевые слова или фразы в запросе, в результатах поиска будет возвращен заранее заданный URI. Эти URI прямых ссылок возвращаются в добавление к URI, возвращаемым серверами поиска при поиске в индексе.

Понятия, связанные с данным

“Ранжирование документов” на стр. 209

“Пользовательские словари значимых слов” на стр. 213

Кэши поиска

Если число документов в результатах потенциально велико, можно повысить производительность, кэшируя результаты поиска.

Когда серверы поиска обрабатывают требования поиска, они сначала проверяют, нет ли в кэше результатов для такого же запроса. Если поисковые серверы обнаружат подходящий ответ на запрос, они могут быстро вернуть результаты поиска пользователю. Если поисковые серверы не обнаружат подходящий ответ на запрос, они будут выполнять поиск в индексе.

При заполнении кэша поиска из него удаляются самые старые результаты поиска, а также результаты редко обрабатываемых запросов, тем самым освобождая место для новых результатов поиска.

С консоли администратора поисковой системы предприятия можно включить поддержку кэширования и задать емкость кэша (число ответов на запросы, которые могут кэшироваться одновременно).

При изменении параметров кэша поиска необходимо перезапустить поисковые серверы, чтобы изменения вступили в действие.

Конфигурирование кэша поиска

Вы можете включить или отключить кэш поиска для собрания. Кроме того, можно задать опции, управляющие размером кэша поиска.

Прежде чем приступить к работе...

Конфигурировать кэш поиска для собрания может пользователь с ролью администратора поисковой системы предприятия или администратор собрания.

Процедура

Чтобы сконфигурировать кэш поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. На странице Опции сервера поиска включите переключатель **Использовать кэш поиска**.
3. В поле **Максимальное число элементов в кэше** задайте максимальное число ответов на запрос, которые могут находиться в кэше поиска.
4. Нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, запустите мониторинг серверов поиска и перезапустите процессы сервера.

Пользовательские словари синонимов

Для повышения качества результатов поиска можно разрешить пользователям при поиске в собрании искать также синонимы терминов запроса.

Если вы создали словарь синонимов, добавили его в систему поисковой системы предприятия и связали с собранием, пользователи могут при поиске в этом собрании искать документы, содержащие синонимы терминов их запросов. Расширяя запросы таким образом, пользователи могут с большей вероятностью найти все

интересующие их документы, а не только документы, точно соответствующие заданным терминам запроса. Поскольку при создании словаря синонимов вы определяете, какие слова являются синонимами, пользователи смогут найти нужные документы, не задавая всех вариантов терминов запроса.

Например, в организации могут использоваться акронимы и сокращения для обозначения отделов организации, оборудования и т.п., или же документы в собраниях могут содержать специальные термины вашей области деятельности. Создав словарь синонимов, можно сделать, чтобы для запросов, содержащих акроним (например, ACL), возвращались документы, содержащие другие формы этого термина (например, ACLs, списки управления доступом, элементы управления доступом и т.д.).

Язык запросов поисковой системы предприятия поддерживает использование синонимов, позволяя пользователям добавлять символ тильды (~) перед термином запроса. Например, запрос ~WAS может возвращать документы, содержащие "WebSphere Application Server". Разработчики прикладных программ могут также сделать синонимы доступными, используя свойства запроса (при этом не потребуется специальный синтаксис).

Словари синонимов содержат различные варианты слов; эти словари обладают следующими особенностями:

- Слова не относятся к определенному языку и могут использоваться в различных языках. У собрания есть только один словарь синонимов.
- Слова задаются в основной словарной форме. Все возможные формы нужно добавить в список синонимов. Например, такими формами могут быть формы различных падежей или множественного числа.

Большинство терминов, добавляемых в словарь синонимов - это точные семантические эквиваленты, то есть если термин А - синоним термина В, то и термин В - синоним термина А. В любом запросе вместо термина А можно использовать термин В и наоборот.

Однако можно также добавить термины, соответствующие различным использованиям термина, включая общие и более частные варианты термина. Например, можно создать одну группу синонимов, содержащую слова строение и дом, и другую группу, содержащую слова банки, сберкассы и тара.

Чем менее строго соответствуют друг другу термины, тем шире набор результатов поиска, но при этом некоторые из результатов поиска могут оказаться не относящимися к делу документами. API поиска и индексации содержит методы, позволяющие пользователю выбирать подходящие синонимы при задании требования поиска, и методы, показывающие пользователям, какие термины запроса были расширены при помощи синонимов.

Для создания словаря синонимов специалист по предметной области, для которой предназначено собрание, должен создать список синонимов в формате XML или попросить разработчика прикладной программы, чтобы он создал файл XML. Для преобразования этого файла XML в двоичный файл (.dic) надо использовать инструмент поисковой системы предприятия **essyndictbuilder**.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь синонимов для поиска документов в собрании, когда конфигурируют опции серверов поиска для собрания.

Ограничение: Добавленный в систему пользовательский словарь синонимов нельзя редактировать. Чтобы исправить синонимы, доступные для собрания, нужно:

1. Изменить исходный файл XML.
2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить старый словарь синонимов из собраний, в которых он используется.
4. Удалить старый словарь синонимов из системы.
5. Добавить в систему новый словарь синонимов.
6. Связать новый словарь синонимов с собраниями, которые будут его использовать.

Можно написать сценарий, содержащий эти действия, после чего с его помощью внедрить словарь повторно в поисковой системе предприятия.

Понятия, связанные с данным



Поддержка синонимов в программах поиска

Задачи, связанные с данной



Создание файла XML для синонимов



Создание словаря синонимов

“Повторное внедрение пользовательских словарей” на стр. 203

Добавление словарей синонимов в систему

Если вы создаете пользовательские словари синонимов для поиска документов в собрании, необходимо связать эти словари с поисковой системой предприятия. После этого можно выбрать, какой из словарей синонимов надо использовать для поиска в собрании.

Прежде чем приступить к работе...

Чтобы добавлять пользовательские словари синонимов для использования с запросами поисковой системы предприятия, надо быть администратором поисковой системы предприятия.

Ограничения

Максимальный размер словаря синонимов - 8 Мбайт.

Процедура

Чтобы связать словари синонимов с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Поиск выберите **Конфигурировать словари синонимов**.
4. На странице Конфигурировать словари синонимов нажмите кнопку **Добавить словарь синонимов**.
5. На странице Добавить словарь синонимов введите для этого словаря синонимов уникальное имя для вывода и (необязательно) описание.
6. Задайте положение файла .dic. Если файл находится в вашей локальной системе, его можно найти в окне просмотра. Если этот файл находится на сервере индексов, введите полный путь.

7. Нажмите кнопку **ОК**. Пользовательский словарь синонимов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря синонимов с собранием

Если с поисковой системой предприятия связаны словари синонимов, можно выбрать один из них для использования при поиске в собрании. Если условие запроса совпадает с термином в этом словаре, в результатах поиска будут возвращаться также документы, содержащие синонимы.

Прежде чем приступить к работе...

Чтобы выбрать словарь синонимов для собрания, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Процедура

Чтобы связать словарь синонимов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. В поле **Имя словаря синонимов** на странице Опции сервера поиска выберите словарь синонимов, который надо использовать при запросах пользователей к данному собранию.

Список доступных словарей синонимов включает в себя все словари синонимов, добавленные в поисковую систему предприятия.

3. Нажмите кнопку **ОК**.

Пользовательские словари стоп-слов

Для повышения качества результатов поиска можно указать, что определенные слова будут автоматически удаляться из условий запроса во время его обработки.

Словарь стоп-слов содержит специфичные для предприятия слова, которые используются часто и поэтому бесполезны в качестве условий поиска. Исключая эти слова из запросов, можно добиться того, что пользователи не будут завалены найденными документами с очень малой релевантностью (возвращены будут только документы, соответствующие остальным терминам запроса). Во время обработки запросов серверы поиска удаляют стоп-слова из запросов. Удаляются стоп-слова из вашего пользовательского словаря и стоп-слова, предварительно определенные для поисковой системы предприятия (такие как обычные предлоги и артикли).

В поисковой системе предприятия по умолчанию выполняется опознание стоп-слов соответствующего языка. Этот процесс удаляет из запроса обычные часто встречающиеся слова, такие как *a* и *the*. В ваш пользовательский словарь стоп-слов надо вносить только стоп-слова, специфичные для предприятия или домена.

При обработке запроса стоп-слова удаляются до того, как предлагаются орфографические варианты написания. Если запрос состоит только из стоп-слов, при его обработке стоп-слова не удаляются. В этом случае удаление стоп-слов отключается, чтобы запрос дал какие-либо результаты. Например, если слово *машина* задано как стоп-слово, а вы ищете *машина*, результаты поиска будут содержать документы со словом *машина*. Но если вы ищете *швейная машина*, результатом поиска будут только документы со словом *швейная*.

Для создания словаря стоп-слов специалист по предметной области, для которой предназначено собрание, должен создать список стоп-слов в формате XML, или попросить сделать это разработчика прикладной программы. Для преобразования файла XML в двоичный файл (.dic) нужно использовать инструмент поисковой системы предприятия **esstopworddictbuilder**.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь стоп-слов для поиска документов в собрании, когда конфигурируют опции серверов поиска для собрания.

Ограничение: После добавления систему пользовательский словарь стоп-слов нельзя редактировать. Чтобы изменить стоп-слова, доступные при обработке запроса, необходимо:

1. Изменить исходный файл XML.
2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить старый словарь стоп-слов из собраний, в которых он используется.
4. Удалить старый словарь стоп-слов из системы.
5. Добавить в систему новый словарь стоп-слов.
6. Связать новый словарь стоп-слов с собраниями, которые будут его использовать.

Можно написать сценарий, содержащий эти действия, после чего с его помощью внедрить словарь повторно в поисковой системе предприятия.

Понятия, связанные с данным

 Пользовательские словари стоп-слов

Задачи, связанные с данной

 Создание файла XML для стоп-слов

 Создание словаря стоп-слов

“Повторное внедрение пользовательских словарей” на стр. 203

Добавление словарей стоп-слов в систему

Если вы создаете пользовательские словари стоп-слов для удаления слов из запросов, необходимо добавить эти словари в поисковую систему предприятия. После этого можно выбрать, какой из словарей стоп-слов вы хотите использовать для поиска в собрании.

Прежде чем приступить к работе...

Чтобы добавлять пользовательские словари стоп-слов в систему, надо быть администратором поисковой системы предприятия.

Ограничения

Максимальный размер словаря стоп-слов - 8 Мбайт.

Процедура

Чтобы связать пользовательские стоп-слова с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.

3. На странице Поиск нажмите кнопку **Конфигурировать словари стоп-слов**.
4. На странице Конфигурировать словари стоп-слов нажмите кнопку **Добавить словарь стоп-слов**.
5. На странице Добавить словарь стоп-слов введите для этого словаря уникальное имя для вывода.
6. Задайте положение файла .dic. Если файл находится в вашей локальной системе, его можно найти в окне просмотра. Если этот файл находится на сервере индексов, введите полный путь.
7. Нажмите кнопку **ОК**. Пользовательский словарь стоп-слов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря стоп-слов с собранием

Если с поисковой системой предприятия связаны словари стоп-слов, можно выбрать один из них для использования при поиске в собрании. Если условие запроса совпадает с термином в этом словаре, такой термин удаляется из запроса перед его обработкой.

Прежде чем приступать к работе...

Чтобы выбрать словарь стоп-слов для собрания, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Процедура

Чтобы связать словарь стоп-слов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. В поле **Имя словаря стоп-слов** на странице Опции сервера поиска выберите словарь стоп-слов, который вы хотите использовать при запросах пользователей к данному собранию.
Список доступных словарей включает в себя все словари стоп-слов, добавленные в поисковую систему предприятия.
3. Нажмите кнопку **ОК**.

Повторное внедрение пользовательских словарей

При помощи консоли администратора поисковой системы предприятия нельзя вносить изменения в словари синонимов, стоп-слов или значимых слов. Однако можно включить соответствующие шаги в сценарий и затем использовать этот сценарий для повторного внедрения словаря.

Прежде чем приступать к работе...

Чтобы выполнить повторное внедрение пользовательского словаря, нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Процедура

Для повторного внедрения пользовательского словаря в поисковую систему предприятия:

1. На сервере индексов откройте один из следующих файлов, чтобы найти имя нужного вам словаря:

- Чтобы найти имя словаря синонимов, откройте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/SynonymConfiguration.xml.
- Чтобы найти имя словаря стоп-слов, откройте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/StopWordDictionaryConfiguration.xml.
- Чтобы найти имя словаря значимых слов, откройте файл КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/master_config/BoostingWordDictionaryConfiguration.xml.

В следующем примере показан файл SynonymConfiguration.xml с двумя словарями синонимов, которые были внедрены под именами hello и goodbye:

```
% cat $ES_NODE_ROOT/master_config/SynonymConfiguration.xml
```

```
<SynonymConfiguration>
<Synonyms>
<Synonym Name="hello" ID="SynonymId_1">
<Filename>synonym_hello1.dic</Filename>
<Timestamp>1169766691776</Timestamp>
</Synonym>
<Synonym Name="goodbye" ID="SynonymId_2">
<Filename>synonym_goodbye2.dic</Filename>
<Timestamp>1169767224839</Timestamp>
</Synonym>
</Synonyms>
</SynonymConfiguration>
```

При загрузке словарей система назначает уникальные ID файлов. В предыдущем примере эти ID - synonym_hello1.dic и synonym_goodbye2.dic. Абсолютный путь к этим файлам на сервере индексов:

- Путь для словаря синонимов - КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/custom_dictionary/synonym_*.dic.
- Путь для словаря стоп-слов - КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/custom_dictionary/stopwordDictionary_*.dic.
- Путь для словаря значимых слов - КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/custom_dictionary/boostingwordDictionary_*.dic.

В системах AIX, Linux и Solaris можно использовать команду **ls** для получения списка доступных словарей. Например:

```
% ls -l $ES_NODE_ROOT/data/custom_dictionary/synonym_*.dic
```

```
-rw-rw-r-- 1 esuser users 9 Jan 25 15:11 /home/esuser/node/data/custom_dictionary/synonym_hello1.dic
-rw-rw-r-- 1 esuser users 9 Jan 25 15:11 /home/esuser/node/data/custom_dictionary/synonym_goodbye2.dic
```

2. Найдя словарь, который вы хотите изменить, остановите поисковую систему предприятия.
3. Запишите вместо словаря, который вы хотите изменить, новый файл словаря. Например, перезапишите synonym_hello1.dic или synonym_goodbye2.dic.
4. Если ваша поисковая система предприятия работает на двух или на четырех серверах, вручную перезапишите файл словаря и на серверах поиска. Файлы словарей располагаются по тому же пути, что и на сервере индексов (КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data/custom_dictionary/).
5. Перезапустите поисковую систему предприятия.

Понятия, связанные с данным

“Пользовательские словари синонимов” на стр. 198

“Пользовательские словари стоп-слов” на стр. 201

“Пользовательские словари значимых слов” на стр. 213

Динамическое составление сводок

Динамическое составление сводок - технология, которая определяет, какие словосочетания в результирующем документе наиболее полно соответствуют понятиям, поиск которых выполняет пользователь.

Для поисковой системы предприятия динамическое составление сводок пытается найти в документе предложения, содержащие много поисковых терминов. Выбранные несколько предложений или частей предложений показываются в результатах поиска. Поисковые термины при выводе результатов поиска в формате HTML выделяются.

При конфигурировании опций сервера поиска для собрания можно задать максимальную длину для вывода сводок документов в результатах поиска. Поскольку сводка содержит символы выделения, буфер, возвращаемый прикладной программе поиска, будет больше заданного максимума. Однако число выводимых символов не будет превышать заданного максимума, хотя сводка может оказаться и короче (в зависимости от сводных данных, извлеченных из исходного документа).

Настройка сводок документов на консоли администратора

Можно настроить объем информации, выводимой в сводках документов, задав опции для поискового сервера с консоли администратора поисковой системы предприятия.

Прежде чем приступать к работе...

Задать размер выводимой информации сводок собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Значение, заданное для максимального размера выводимой информации сводок документов, действует совместно со значением, заданным для числа предложений, которое может содержать сводка каждого документа. Применяется то ограничение, которое дает меньшую длину сводки документа.

Например, если задать, что сводка может содержать не более четырех предложений, в сводке будет только четыре предложения, даже если максимальная длина сводки в символах больше общего числа символов в этих предложениях. Другой пример - предельное значение из 10 предложений в сочетании с пределом длины для вывода в 500 символов может привести к получению сводки документа, содержащей менее 10 предложений.

Процедура

Чтобы сконфигурировать длину выводимой информации сводок документов:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. На странице Опции сервера поиска задайте максимальную длину выводимой информации сводок документов. При просмотре результатов поиска пользователи будут видеть сводки, длина которых не превышает заданного значения.
3. Задайте, сколько предложений может содержаться в сводке каждого документа (сводки могут содержать до 10 предложений).

4. Нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, запустите мониторинг серверов поиска и перезапустите процессы сервера.

Настройка сводок документов при помощи редактирования свойств

Для каждого документа в результатах поиска для запроса поисковой системы предприятия выводится сводка. Отредактировав файл свойств, можно настроить объем информации, выводимой в каждой из этих сводок. by editing a properties file.

Об этой задаче

Можно настроить описания в результатах поиска, изменив значения следующих свойств в файле `ES_NODE_ROOT/master_config/ID_собрания.runtime.node1/runtime-generic.properties`:

MinWordsPerSentence

Минимальное число слов в каждом предложении сводки. Более короткие предложения включаются в сводку, только если предложений, в которых больше `MinWordsPerSentence` слов, недостаточно. Значение по умолчанию - 4.

MaxWordsPerSentence

Максимальное число слов в каждом предложении сводки. Если в предложении больше слов, в сводку включается только часть предложения (это часть, содержащая термы запроса и не превышающая `MaxWordsPerSentence` слов). Остаток предложения исключается. Значение по умолчанию - 20.

Предложения для сводок документов выбираются по специальному внутреннему алгоритму, который определяет релевантность всех предложений, в которые входят термы поиска. Отбор по релевантности выполняется до фильтрации предложений по их длине.

NumberOfReturnedSentences

Число предложений в описании документа. Значение по умолчанию - 5.

MaxSentencesPerDocument

Максимальное число предложений в документе, которые при создании описания рассматриваются как кандидаты для внесения в описание. Значение по умолчанию - 1000.

Процедура

Чтобы настроить сводки документов в результатах поиска:

1. Зарегистрируйтесь на серверах поиска в качестве администратора поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. При помощи текстового редактора отредактируйте следующий файл (`ID_собрания` - это ID, заданный для собрания или назначенный системой при создании собрания):
`ES_NODE_ROOT/master_config/ID_собрания.runtime.node1/runtime-generic.properties`

Совет: Чтобы узнать, какой ID у собрания с данным именем, посмотрите файл `ES_NODE_ROOT/master_config/collections.ini`.

3. Измените свойства, которые хотите настроить, затем сохраните файл и закройте редактор.

4. Остановите и перезапустите серверы поиска, чтобы изменения вступили в силу.

Работа с прямыми ссылками

Прямые ссылки - это документы, возвращаемые в результатах поиска, когда пользователь вводит запрос, включающий в себя определенные слова и словосочетания.

Для конфигурирования прямых ссылок для собрания используется консоль администратора поисковой системы предприятия.

Прямые ссылки

Прямые ссылки позволяют выдавать пользователям ссылки на те документы, которые заранее определены как соответствующие данным терминам запроса.

Прямая ссылка - это URI, которое поисковая система предприятия автоматически включает в результаты поиска, если запрос содержит определенные слова или фразы. Обычно URI прямых ссылок появляются в верхней части списка результатов, чтобы пользователи увидели документы, которые вы заранее определили как соответствующие данному запросу.

Прямые ссылки связи возвращаются в дополнение к другим результатам поиска. Процессы поиска ищут в индексе документы, удовлетворяющие условиям запроса, и возвращают URI для этих документов вместе с URI прямых ссылок.

При конфигурировании прямой ссылки можно указать для URI содержательный заголовок и резюме, чтобы пользователи могли быстро определить, тот ли это, который они хотели получить.

Например, для URI <http://www.ibm.com/education/us/> можно использовать заголовок Обучение IBM в США, и резюме Решения, продукты и ресурсы для профессионалов, обучающего персонала и студентов в США.

Чтобы можно было использовать прямые ссылки в собраниях поисковой системы предприятия, в программе поиска должна быть задана опция вывода прямых ссылок. В некоторых поисковых программах у пользователей может быть возможность включать и отключать возвращение прямых ссылок при поиске в собрании.

Конфигурирование прямых ссылок

Чтобы создать прямую ссылку для собрания поисковой системы предприятия, свяжите URI документа с ключевыми словами, для которых этот документ будет включаться в результаты поиска.

Прежде чем приступить к работе...

Для конфигурирования прямых ссылок надо быть администратором поисковой системы предприятия или администратором собрания, к которому принадлежит эта прямая ссылка.

Об этой задаче

Чтобы увидеть примеры задания ключевых слов и URI для прямых ссылок, при создании или редактировании прямой ссылки выберите **Справка**.

Чтобы изменения вступили в силу, перезапуск серверов поиска не требуется.

Процедура

Чтобы сконфигурировать прямую ссылку:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать прямые ссылки**.
2. На странице Прямые ссылки выберите **Создать прямую ссылку**.
3. Задайте ключевые слова и фразы, для которых в результатах поиска будет возвращаться эта прямая ссылка, URI документа, который вы заранее определяете как документ для такого запроса, и другие опции для этой прямой ссылки.

Можно задать одно ключевое слово, несколько ключевых слов или одну фразу (два или более слов в кавычках) на строке. Разделяйте ключевые слова пробелами (не запятыми!). Чтобы начать новую строку, нажмите клавишу Enter.

4. Нажмите кнопку **ОК**.

Новая прямая ссылка появится в списке на странице Прямые ссылки вместе с другими прямыми ссылками для этого собрания.

Ссылки, связанные с данной

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

Ранжирование документов

Когда пользователь выполняет поиск в собрании, процессы поиска возвращают наиболее релевантные результаты для терминов и условий запроса.

Серверы поиска поддерживают сложный синтаксис запросов и используют несколько способов получения наиболее релевантных результатов поиска. Можно дополнить поведение ранжирования по умолчанию, сконфигурировав опции, управляющие важностью документов в результатах поиска:

- Можно создать пользовательские словари значимых слов для управления ранжированием документов, содержащих заданные слова, в результатах поиска.
- Можно управлять оценками документов, соответствующих заданному шаблону URI.
- Можно управлять оценками документов с полями, отображенными на классы значимости.

Понятия, связанные с данным

“Ранжирование документов на основе классов значимости” на стр. 217

“Пользовательские словари значимых слов” на стр. 213

Назначение оценок на основе текста

Поисковая система предприятия динамически вычисляет балл для каждого документа, соответствующего условиям запроса.

Оценка текста для документа выражает важность термов запроса в документе. Для вычисления оценки текста каждого соответствующего запросу документа поисковая система предприятия учитывает много факторов, таких как:

- Позволяют ли термы отличить документ от других документов. Например, если термы запроса встречаются в одном документе и отсутствуют в других, это означает, что они важны для этого документа и позволяют отличить его от других документов. Термы, встречающиеся во многих документах, дают меньший вклад в оценку документов, чем термы запроса, встречающиеся в меньшем наборе документов.
- Число вхождений термов запроса в документе. При оценке терма учитывается каждое его вхождение в документе. Чем больше вхождений терма запроса в документе, тем выше оценка этого документа. При выполнении вами поиска слова (например, *thinkpad*) документ с большим числом вхождений этого слова ранжируется в результатах поиска выше, чем другие документы с меньшим числом вхождений.
- Для документов HTML рассматриваются атрибуты терма запроса (то есть контекст терма). Оценка терма выполняется на основе его атрибутов, таких как *location*, *bold*, *italic*, *anchor* и т.д. Обычно термы, встречающиеся в заголовке документа, оцениваются выше, чем термы, встречающиеся в абзацах текста. Выделенные термы (например, жирным шрифтом) оцениваются выше, чем термы в простом тексте. В параметрах синтаксического анализатора можно сконфигурировать важность атрибутов.
- Вес термов запроса. Чтобы настроить важность термов в документе, для них можно сконфигурировать показатели значимости. Для этого надо связать с собранием словарь значимых слов (содержащий термы и их показатели значимости). Этот словарь будет использоваться при поиске, а показатели

значимости термов в этом словаре будут влиять на оценку документа. Чем выше показатель значимости, тем выше вклад термина в оценку документа.

- Близость термов запроса в документе. Если термины запроса встречаются в документе рядом друг с другом, их контекстная близость используется для вычисления оценки текста. Например, у вас есть два документа. В одном документе говорится об автостоянке в городе (car park in the city); здесь слова car и park встречаются вместе. В другом документе говорится об автомобильном демонстрационном зале в городе недалеко от парка (car showroom in a city near to the park); здесь слова car и park встречаются порознь. Если вы ищете словосочетание car park (автостоянка), близость этих слов друг к другу в первом документе внесет больший вклад в ранг этого документа по сравнению со вторым.
- При вычислении оценки документа учитывается также его длина и разнообразие слов в документе (например, число уникальных слов).

Статическое ранжирование

Для некоторых типов документов можно назначить фактор статического ранжирования, увеличивающий важность этих документов относительно других результатов поиска.

При создании собрания вы задаете опции **Важности документов**. Выбранный вами тип важности документов определяет, будет ли с документами в собрании связан коэффициент статического ранжирования. При выполнении пользователями поиска в собрании, где используется статическое ранжирование, коэффициент статического ранжирования влияет на способ ранжирования документов, возвращаемых в результатах поиска.

Для содержимого Web основой коэффициента статического ранжирования служат ссылки. Число ссылок на документ из других документов и сами документы, ссылающиеся на этот документ, могут увеличить релевантность документа относительно других результатов поиска.

Для документов, содержащих поля или метаданные с датами, в качестве основы для коэффициента статического ранжирования используется дата документа. Поле даты документа, предоставляемое искателем, может содержать дату последнего изменения документа или дату его последнего просмотра искателем, в зависимости от заданной вами конфигурации искателя.

Дата документа может увеличить его релевантность. Например, недавние статьи в группах новостей NNTP могут быть более релевантны, чем старые. Если источник данных содержит несколько полей с датами, при конфигурировании искателя можно выбрать, какое из них лучше всего определяет релевантность документов.

Если в собрании используется статическое ранжирование, не следует смешивать в одном собрании источники данных, использующие разные типы ранжирования. Например, для использования в качестве фактора статического ранжирования ссылок на документ необходимо, чтобы собрание содержало только документы Web. При наличии в одном собрании документов с разными моделями ранжирования точность ранжирования документов будет ниже, и порядок результатов поиска может оказаться неожиданным.

Кроме того следует убедиться, что документы в собрании содержат поля и значения, позволяющие применить статическое ранжирование. Например, представим себе собрание, сконфигурированное для использования статического ранжирования на основе дат документов, и искатель в этом собрании, сконфигурированный для

использования дат документов, содержащихся в конкретном поле. Если документ не содержит этого поля, ранг важности для него может быть определен неправильно и порядок результатов поиска может оказаться неожиданным.

Влияние ранжирования на основе ссылок

Статическое ранжирование наряду с такими факторами, как назначение оценки для шаблонов значимых URI, вносит вклад в статическую оценку документа и влияет на его важность. Модель ранжирования на основе ссылок обычно применяется для собраний Web, поскольку эта модель вычисляет статический ранг документа на основе числа ссылок на этот документ. Документ, на который ссылаются больше других документов, ранжируется как более релевантный.

Из этого следует, что при конфигурировании данной модели для другого собрания (не Web) или собрания смешанного типа (содержащего и документы Web, и документы других типов) качество поиска может ухудшиться, поскольку для документов других типов (не Web) отсутствует механизм ссылок.

При включении ранжирования на основе ссылок включается также обнаружение повторений документов. Статический ранг дубликатов будет тем же, что и у главного документа. Если в группе повторяющихся документов для каких-либо документов не сконфигурированы коэффициенты значимости шаблонов URI, у всех повторяющихся документов будет одна и та же статическая оценка.

Восстановление значений по умолчанию для статического ранжирования документов

Если при создании собрания вы конфигурируете опцию статического ранжирования документов, можно вернуть значения свойств по умолчанию, отредактировав файл `runtime.properties` для этого собрания.

Прежде чем приступать к работе...

Восстановить значения по умолчанию ранжирования документов в собрании может только администратор поисковой системы предприятия.

Об этой задаче

Чтобы восстановить значения по умолчанию ранжирования документов для собрания, надо изменить файлы `runtime.properties` для этого собрания и всех поисковых серверов в поисковой системе предприятия. В конфигурации с несколькими серверами файл `runtime.properties` находится на сервере индексов в каталоге `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.runtime.ID_узла`, где `ID_собрания` - ID для собрания, а `ID_узла` - ID для поисковых серверов.

Например, чтобы изменить собрание `coll` в поисковой системе предприятия с несколькими серверами, надо изменить файл `runtime.properties` для этого собрания и оба поисковых сервера (`node3` и `node4`):

```
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/coll.runtime.node3/runtime.properties
КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/coll.runtime.node4/runtime.properties
```

Процедура

Чтобы восстановить значения по умолчанию ранжирования документов для собрания:

1. Зарегистрируйтесь на сервере индексов как администратор поисковой системы предприятия.
2. Найдите ID собрания, для которого хотите восстановить значения ранжирования по умолчанию. ID собрания находится в файле `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/collections.ini`. Отсортируйте этот файл для удобства просмотра. В следующем примере `coll` - ID собрания:

```
% sort $ES_NODE_ROOT/master_config/collections.ini | more
collection1.configfile=coll_config.ini
collection1.datadir=/home/esearch/node/data/coll
collection1.description=
collection1.displayname=Collection1
collection1.flags=0
collection1.id=coll
collection1.sectiontype=collection
collection1.type=1
...
```

3. Отредактируйте файл `runtime.properties` для собрания, где нужно восстановить значения, внося в него изменения:

- a. Удалите следующие свойства:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- b. Если в файле `runtime.properties` задано `trevi.sourcetype=1`, что означает ранжирование документов по ссылкам, отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/runtime.1/runtime.properties` и затем скопируйте и вставьте следующие свойства по умолчанию в файл `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- c. Если в файле `runtime.properties` задано `trevi.sourcetype=2`, что означает ранжирование документов по дате, отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/runtime.2/runtime.properties` и затем скопируйте и вставьте следующие свойства по умолчанию в файл `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

- d. Если в файле `runtime.properties` задано `trevi.sourcetype=3`, что означает, что коэффициент статического ранжирования не используется при ранжировании документов в собрании, отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/runtime.0/runtime.properties` и затем скопируйте и вставьте следующие свойства по умолчанию в файл `runtime.properties`:

```
trevi.autorank.dfthreshold1
trevi.autorank.dfthreshold2
trevi.autorank.dfthreshold3
trevi.autorank.rc0.*
trevi.autorank.rc1.*
```

4. В конфигурации с несколькими серверами повторите шаг 3 на стр. 212, чтобы изменить файл `runtime.properties` для того же собрания и второго поискового сервера.
5. Посмотрите состояние страницы Поиск и перезапустите процессы поиска для этого собрания.

Повторите эти шаги для каждого собрания, в котором нужно восстановить значения параметров ранжирования документов по умолчанию.

Пользовательские словари значимых слов

Чтобы улучшить качество результатов поиска, можно изменить порядок ранжирования документов в результатах поиска, создав пользовательский словарь значимых слов.

Если в запросе задано слово, включенное в словарь значимых слов, важность документов, содержащих это слово, возрастет или уменьшится в зависимости от коэффициента значимости, назначенного в словаре этому слову.

С помощью словаря значимых слов можно гарантировать, что при задании пользователем определенных терминов запроса будут возвращаться определенные документы. Допустим, что у вас есть собрание, содержащее множество документов об автомобилях. Для таких документов можно предположить важность определенных ключевых слов, относящихся к моделям автомобилей, имени производителя и так далее. Для управления ранжированием результатов поиска можно назначить коэффициент значимости, связав значимые значения с ключевыми словами (модель, производитель и т.п.) в словаре значимых слов. При выполнении пользователями поиска в собрании и задании ими запроса, содержащего любое из этих ключевых слов, документы об автомобилях будут ранжироваться в результатах поиска выше, чем остальные документы.

Коэффициенты значимости принимают значения от -10 до 10. При обработке запросов серверы поиска рассматривают как более важные документы с положительными коэффициентами значимости и как менее важные - документы с отрицательными коэффициентами значимости.

Например, документ, соответствующий терминам запроса с высокими коэффициентами значимости, получает более высокий ранг, чем без использования коэффициентов значимости. (Коэффициент значимости - единственная величина, влияющая на ранг документа.)

При создании словаря можно задать для любого числа слов одинаковые коэффициенты значимости. В словарь можно включать термины из одного или нескольких слов. Термины из нескольких слов выступают в поиске в качестве словосочетаний.

Если слово с весовым вкладом, заданным через коэффициент значимости, входит в запрос с операцией OR (например: этот | тот), для терминов этого запроса вычисляется взвешенное среднее. Полученная в итоге сводная оценка используется для всех вхождений операндов запроса OR. Частные оценки для отдельных операндов запроса OR не вычисляются.

Оценка значимости, основанная на словарях значимых слов, не поддерживается для терминов поиска по полям. При анализе терминов запроса для численной оценки

важности документа используется только текст запроса, но не имя поля. Чтобы применить коэффициенты значимости к терминам запроса в полях, надо отобразить имена полей на классы значимости.

Для создания словаря значимых слов специалист по предметной области, для которой предназначено собрание, должен создать список значимых слов в формате XML или попросить разработчика прикладной программы, чтобы он создал файл XML. Для преобразования этого файла XML в двоичный файл (.dic) надо использовать инструмент поисковой системы предприятия **esboosttermdictbuilder**.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь значимых слов, используемый для поиска документов в собрании, когда конфигурируют опции поискового сервера для собрания.

Ограничение: Добавленный в систему пользовательский словарь значимых слов нельзя редактировать. Чтобы исправить значимые слова, доступные для обработки запросов, надо:

1. Изменить исходный файл XML.
2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить словарь значимых слов из собраний, в которых он используется.
4. Удалить старый словарь значимых слов из системы.
5. Добавить в систему новый словарь значимых слов.
6. Связать новый словарь значимых слов с собраниями, в которых он должен использоваться.

Можно написать сценарий, содержащий эти действия, после чего с его помощью внедрить словарь повторно в поисковой системе предприятия.

Понятия, связанные с данным

“Ранжирование документов” на стр. 209



Пользовательские словари значимых слов

Задачи, связанные с данной



Создание файла XML для значимых слов



Создание словаря значимых слов

“Повторное внедрение пользовательских словарей” на стр. 203

Добавление словарей значимых слов в систему

Если вы создаете пользовательские словари значимых слов, необходимо связать эти словари с поисковой системой предприятия. После этого можно выбрать, какой из словарей значимых слов использовать для поиска в собрании.

Прежде чем приступить к работе...

Добавлять пользовательские словари значимых слов в систему может администратор поисковой системы предприятия.

Ограничения

Максимальный размер словаря значимых слов - 8 Мбайт.

Процедура

Чтобы связать пользовательские словари значимых слов с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Поиск выберите **Конфигурировать словари значимых слов**.
4. На странице Конфигурировать словари значимых слов нажмите кнопку **Добавить словарь значимых слов**.
5. На странице Добавить словарь значимых слов введите для этого словаря значимых слов уникальное имя для вывода и (необязательно) описание.
6. Задайте положение файла .dic. Если файл находится в вашей локальной системе, его можно найти в окне просмотра. Если этот файл находится на сервере индексов, введите полный путь.
7. Нажмите кнопку **ОК**. Пользовательский словарь значимых слов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря значимых слов с собранием

Если с поисковой системой предприятия связаны словари значимых слов, можно выбрать один из них для использования при поиске в собрании. Если термин запроса совпадает с термином в словаре, важность документов, содержащих этот термин, возрастет или уменьшится, в зависимости от коэффициента значимости, назначенного в словаре этому термину.

Прежде чем приступать к работе...

Выбрать словарь значимых слов для собрания может администратор поисковой системы предприятия или администратор этого собрания.

Процедура

Чтобы связать словарь значимых слов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. В поле **Имя словаря значимых слов** на странице Опции сервера поиска выберите словарь значимых слов, который надо использовать при запросах пользователей к данному собранию.
Список доступных словарей включает в себя все словари значимых слов, добавленные в поисковую систему предприятия.
3. Нажмите кнопку **ОК**.

Ранжирование документов на основе шаблонов URI

Можно повысить или понизить важность документов, назначив коэффициенты значимости для шаблонов URI.

Всем документам при их добавлении в индекс назначается оценка статического ранжирования по умолчанию. Эта оценка зависит от того, включено ли для собрания статическое ранжирование, и, если это так, от типа статического ранжирования - по дате документов или (для документов Web) по числу ссылок на данный документ из других документов.

Относительной важностью документов можно управлять, назначая коэффициенты значимости для шаблонов URI. Коэффициент значимости используется с оценкой статического ранжирования по умолчанию и другими коэффициентами при определении конечной статической оценки документа.

Порядок, в котором сконфигурированы шаблоны URI, имеет значение. Когда сервер индексов вычисляет значение для каждого документа в собрании, он оценивает шаблоны URI в порядке их перечисления в списке. Для каждого URI:

1. Сервер индекса последовательно просматривает шаблоны URI.
2. Когда сервер индекса находит первый шаблон URI, соответствующий документу в индексе, он применяет к этому документу коэффициент значимости, сконфигурированный для данного URI.
3. Если для документа не найден соответствующий шаблон URI, используется оценка статического ранжирования по умолчанию.

Сконфигурировав коэффициент значимости для шаблона URI, разместите его в соответствии с требуемым порядком просмотра шаблонов сервером индексов.

Влияние на оценки документов, соответствующих шаблонам URI

Можно повысить или понизить важность документов, соответствующих шаблону URI, применив коэффициент значимости к оценке статического ранжирования по умолчанию.

Прежде чем приступить к работе...

Управлять важностью документов, соответствующих URI, может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Сконфигурированный коэффициент значимости используется совместно с оценкой статического ранжирования при вычислении новой статической оценки для всех документов, соответствующих заданному шаблону URI.

Коэффициенты значимости изменяют только статические оценки, и эти коэффициенты являются единственными участниками вычисления, определяющими конечный ранг документа. Например, если на один документ много ссылок (что выражается в его высокой начальной оценке), то другой документ, на который ссылки отсутствуют, всегда будет иметь более низкий ранг, чем первый.

Процедура

Чтобы повлиять на оценки документов, соответствующих заданному шаблону URI:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Управление оценками на основе соответствия шаблонам URI**.
2. На странице Управление оценками на основе соответствия шаблонам URI выберите **Добавить шаблон URI**.
3. Введите шаблон для документов, важность которых в результатах поиска вы хотите повысить или понизить. Например:

```
http://domain.org/ru/*  
db2://*ROI*  
*/afs/*
```

4. Введите для коэффициента значимости значение от -10 до 10. На основе этого коэффициента значимости будет вычисляться конечная статическая оценка для всех документов, соответствующих шаблону URI.
5. Нажмите кнопку **ОК**.
6. На странице Управление оценками на основе соответствия шаблонам URI разместите новый шаблон URI в соответствии с требуемым порядком его просмотра сервером индексов.
Сервер индексов вычисляет оценки статического ранжирования в порядке перечисления URI. Для достижения лучших результатов в начало списка ставьте более точно заданные URI. В следующем примере подкаталог /forms соответствует шаблону URI `http://www.ibm.com/ru/*`. Чтобы добиться правильного вычисления оценки для документов в подкаталоге /forms, поставьте шаблон URI для подкаталога /forms первым в списке:
`http://www.ibm.com/ru/forms/* 8`
`http://www.ibm.com/ru/* -2`
7. Чтобы изменить шаблон URI или коэффициент значимости, выберите шаблон URI и нажмите кнопку  **Редактировать**.
8. Чтобы удалить шаблон URI из списка, выберите шаблон URI и нажмите кнопку  **Удалить**.
9. Чтобы применить коэффициенты значимости к ранее проиндексированным документам, выполните полное построение индекса.

Ранжирование документов на основе классов значимости

Отображая поля на классы значимости, можно влиять на порядок ранжирования документов в результатах поиска.

При анализе документов анализатор назначает *классы значимости* маркерам документов в соответствии с полями, к которым принадлежат эти маркеры. Эти классы значимости включаются в индекс и используются для вычисления оценок, оказывающих влияние на ранжирование документов результатов.

Чтобы воздействовать на то, как вычисляются оценки, можно сконфигурировать численные коэффициенты значимости для классов значимости. Если термин запроса совпадает с маркером в поле, отображаемом на класс значимости, вклад этого вхождения маркера будет влиять на итоговую оценку значимости документа. В вычислении этой оценки участвует коэффициент значимости, сконфигурированный в классе значимости.

Пусть, например, вам нужно оценить значимость полей заголовка. Если термин запроса встречается в заголовке, это вхождение будет вносить большой вклад в оценку значимости документа и повысит его ранг в результатах поиска.

Изменить ранжирование документа можно при помощи консоли управления поисковой системой предприятия, задав коэффициенты значимости для классов значимости и отобразив поля на классы значимости. Для поисковой системы предприятия предварительно сконфигурировано шестнадцать классов значимости. Восемь из этих классов значимости предназначены для использования с полями содержимого, а другие восемь - для полей метаданных. Коэффициенты, связанные с классами значимости по умолчанию, можно изменить, кроме того, можно связать с классами значимости другие поля или дополнительные поля.

Если вы изменили отображения полей, надо снова обработать документы искателем и проанализировать их, чтобы внесенные вами изменения можно было применить к

ранее проиндексированным документам. Если вы изменили коэффициенты, заданные для классов значимости, включите слежение за серверами поиска, остановите и перезапустите процессы на серверах поиска, чтобы эти изменения вступили в силу.

Обнаружение дубликатов документов и сводки документов

При отображении поля на класс значимости надо указать, используется ли это поле для обнаружения дубликатов документов и можно ли включать содержимое поля в сводки документов в результатах поиска.

- Если поле используется для поиска дубликатов документов, это поле считается полем содержимого, и для выбора остаются только классы значимости, предназначенные для полей содержимого. Содержимое полей этих типов можно использовать в результатах поиска в динамических сводках документов.
- Если поле не используется для поиска дубликатов документов, это поле считается полем метаданных, и для выбора остаются только классы значимости, предназначенные для полей метаданных. В этом случае два документа, идентичные во всех отношениях, кроме указанного поля, считаются дубликатами друг друга, и поле не используется в динамических сводках документов.

Высокие и низкие значения возврата

Во время обработки запроса процесс поиска оценивает число возвращаемых документов результатов. Чтобы определить, характеризуется ли запрос высоким значением возврата, низким значением возврата или же промежуточным значением между низким и высоким значениями возврата, задаются пороговые значения:

Низкое значение возврата

Если оцениваемое число полученных документов ниже нижнего порогового значения, запрос считается обладающим низким значением возврата.

Высокое значение возврата

Если оцениваемое число полученных документов выше верхнего порогового значения, запрос считается обладающим высоким значением возврата.

Среднее значение возврата

Если оцениваемое число документов находится между этими двумя пороговыми значениями, значение возврата считается средним.

Для каждого класса значимости задаются коэффициенты значимости, которые при обработке запросов связываются с высокими и низкими значениями возврата. Нижний коэффициент значимости влияет на относительную важность запросов с низкими значениями возврата, а верхний коэффициент значимости - на относительную важность запросов с высокими значениями возврата. Сочетание этих двух коэффициентов влияет на относительную важность запросов со средним значением возврата.

Значения коэффициентов значимости управляют относительной важностью каждого из входящих термина запроса в документе. Каждое найденное в документе вхождение термина запроса считается с учетом соответствующего коэффициента значимости.

При конфигурировании классов значимости для собрания можно отредактировать коэффициенты значимости по умолчанию. Например, можно задать такие коэффициенты значимости, что термины запроса, найденные в полях заголовков, будут учитываться с пятикратным преимуществом по сравнению с обычным текстом.

Понятия, связанные с данным

Отображение полей на классы значимости

Можно изменять относительную важность полей, отображая имена полей на классы значимости.

Прежде чем приступить к работе...

Отображать поля на классы значимости может администратор поисковой системы предприятия или администратор данного собрания.

Об этой задаче

Через коэффициенты значимости система влияет на ранжирование документов, содержащих термины запроса в полях, отображенных на классы значимости.

Поисковая система предприятия резервирует некоторые отображения для внутренних полей и обычного текста, у которых нет каких-либо других отличительных характеристик. Вы можете отображать другие поля на классы значимости, которые используют эти зарезервированные поля, но сами зарезервированные поля редактировать или удалять нельзя.

Процедура

Чтобы отобразить поля на классы значимости:

1. Вызовите редактирование собрания, перейдите на страницу Анализ и нажмите кнопку **Отобразить поля на классы значимости**.
2. На странице Отобразить поля на классы значимости нажмите кнопку **Добавить поле**.
3. На странице Добавить поле в класс значимости введите имя поля, которое нужно отобразить на класс значимости.

Вы можете указать имя поля, существующего в источнике, обработанном искателем, или во внешнем источнике, имя поля, отображенного из элемента XML, имя поля, отображенного из элемента метаданных HTML, или же одно из заранее определенных имен полей.

4. Укажите, используется ли это поле для обнаружения дубликатов документов. Если этот переключатель включен, список доступных классов значимости будет содержать классы, относящиеся к полям содержимого. Если в результаты поиска попадает документ с этим полем, содержимое поля выводится в области сводки документа.
Если этот переключатель выключен, список доступных классов значимости будет содержать классы, относящиеся к полям метаданных. Содержимое поля не будет выводиться в области сводки документов результатов поиска.
5. Выберите класс значимости и нажмите кнопку **ОК**.
Добавленное вами поле будет выведено на странице Отобразить поля на классы значимости. Вы можете выбрать опцию редактирования классов значимости и задать другие коэффициенты значимости для оценивания документов, содержащих это поле.
6. Чтобы задать, используется ли поле для обнаружения дубликатов документов или отобразить это поле на другой класс значимости, выберите  **Редактировать**. (Редактировать поля, зарезервированные поисковой системой предприятия, нельзя.)

7. Чтобы удалить поле из класса значимости, выберите  **Удалить**. (Удалять поля, зарезервированные поисковой системой предприятия, нельзя.)
8. Чтобы применить изменения к ранее индексированным документам, выполните их повторный просмотр искателем и индексацию.

Понятия, связанные с данным

“Ранжирование документов” на стр. 209

Конфигурирование коэффициентов значимости для классов значимости

Коэффициенты значимости, сконфигурированные вами для классов значимости, отражают вашу оценку вклада конкретных полей в найденных документах в выполнение запроса. Классы значимости с высокими коэффициентами значимости могут повысить важность найденных документов с полями, отображаемыми на эти классы значимости.

Прежде чем приступить к работе...

Конфигурировать коэффициенты значимости для классов значимости может администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Для перерасчета важности найденных документов с полями, отображаемыми на классы значимости, в системе используются коэффициенты значимости, сконфигурированные для класса значимости, коэффициент статического ранжирования по умолчанию и другие коэффициенты.

Процедура

Чтобы сконфигурировать коэффициенты значимости для классов значимости:

1. Вызовите редактирование собрания, перейдите на страницу Анализ и нажмите кнопку **Отобразить поля на классы значимости**.
2. На странице Отобразить поля на классы значимости нажмите кнопку **Редактировать классы значимости**.
3. На странице Классы значимости найдите класс значимости, который нужно изменить, и выберите  **Редактировать**.
4. На странице Редактировать класс значимости задайте новые значения для верхних и нижних коэффициентов значимости. Для обоих коэффициентов можно ввести одинаковые значения.
5. Нажмите кнопку **ОК**.
6. Чтобы изменения вступили в силу, включите слежение за серверами поиска и выберите значки для остановки и перезапуска процессов поиска. При выполнении пользователями запросов относительная важность найденных документов с полями, отображаемыми на данный класс значимости, будет определяться новыми коэффициентами значимости.

Понятия, связанные с данным

“Ранжирование документов” на стр. 209

Значения классов значимости по умолчанию

В системе поиска предприятия есть 16 классов значимости, с помощью которых можно влиять на порядок ранжирования документов в результатах поиска.

Для вычисления оценок для полей и текста, у которых нет каких-либо других отличительных характеристик, для использования в поисковой системе предприятия зарезервированы следующие поля:

```
es_special_field.regular_text
es_special_field.default_field
es_special_field.default_metadata_field
```

Вы можете отображать другие поля на классы значимости, которые используют эти зарезервированные поля, но сами зарезервированные поля редактировать или удалять нельзя.

Для всех остальных полей коэффициенты значимости, используемые системой для вычисления рангов документов, можно редактировать. Можно также отобразить любое число полей на любые классы значимости, включая те из них, которые используются зарезервированными полями.

Ниже в таблице перечислены имена классов значимости, коэффициенты значимости по умолчанию для запросов с низкими значениями возврата, коэффициенты значимости по умолчанию для запросов с высокими значениями возврата и имена заранее определенных полей, отображаемых на классы значимости в конфигурации по умолчанию.

Коэффициенты значимости по умолчанию изменяются в соответствии с методом статического ранжирования, выбранным для собрания во время создания данного собрания. При этом возможны опции без статического ранжирования, с ранжированием по числу ссылок на документ (для источников Web) или с ранжированием по датам документов.

Таблица 6. Значения классов значимости по умолчанию

Нижние и верхние коэффициенты значимости по умолчанию				
Имя класса значимости	Без статического ранжирования	По ссылкам на документы	По датам документов	Предопределенные для отображения поля
Класс содержимого A	Нижний: 4 Верхний: 2	Нижний: 6 Верхний: 1	Нижний: 4 Верхний: 2	es_special_field.regular_text
Класс содержимого B	Нижний: 5 Верхний: 4	Нижний: 7 Верхний: 3	Нижний: 5 Верхний: 4	es_special_field.html_emphasized_text Включает элементы HTML: b, big, caption, dfn, em, h4, h5, h6, strong
Класс содержимого C	Нижний: 7 Верхний: 4	Нижний: 9 Верхний: 3	Нижний: 7 Верхний: 4	es_special_field.html_headers Включает элементы HTML: h1, h2, h3
Класс содержимого D	Нижний: 2 Верхний: 5	Нижний: 1 Верхний: 5	Нижний: 2 Верхний: 5	title
Класс содержимого E	Нижний: 1 Верхний: 1	Нижний: 5 Верхний: 10	Нижний: 1 Верхний: 1	es_special_field.anchor
Класс содержимого F	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.anchor_same_dir
Класс содержимого G	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.anchor_same_host
Класс содержимого H	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.default_field

Таблица 6. Значения классов значимости по умолчанию (продолжение)

Нижние и верхние коэффициенты значимости по умолчанию				
Имя класса значимости	Без статического ранжирования	По ссылкам на документы	По датам документов	Предопределенные для отображения поля
Класс метаданных А	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.default_metadata_field
Класс метаданных В	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных С	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных D	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных E	Нижний: 1 Верхний: 1	Нижний: 5 Верхний: 1	Нижний: 1 Верхний: 1	keywords
Класс метаданных F	Нижний: 1 Верхний: 1	Нижний: 3 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.urlhost
Класс метаданных G	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.urlpath
Класс метаданных H	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	описание

Понятия, связанные с данным

“Ранжирование документов” на стр. 209

Прикладные программы поиска для поисковой системы предприятия

Прикладная программа поиска позволяет выполнять поиск в собраниях и внешних источниках поисковой системы предприятия. Можно создать любое число прикладных программ поиска, каждая из которых будет выполнять поиск в любом числе собраний и внешних источников.

Пример прикладной программы поиска

Этот пример программы поиска демонстрирует многие функции поиска и получения, доступные для поисковой системы предприятия. Кроме того, пример прикладной программы поиска - работающий пример, иллюстрирующий использование API поиска и индексации IBM (SI-API, Search and Index API) для построения пользовательских интерактивных прикладных программ поиска, приспособленных для задач вашего предприятия.

Если не изменить свойства в файле конфигурации по умолчанию, пример прикладной программы поиска позволяет выполнять поиск во всех активных собраниях и внешних источниках в поисковой системе предприятия. Пример прикладной программы поиска можно использовать для тестирования новых собраний и внешних источников перед тем, как делать их доступными для пользователей.

Пример программы поиска автоматически связывается со всеми собраниями и внешними источниками. В производственной среде администраторы поисковой системы предприятия решают, каким программам поиска разрешено выполнять поиск в тех или иных собраниях.

Пользовательские прикладные программы поиска

Созданные вами прикладные программы поиска можно выполнять как самостоятельные прикладные программы Web в среде IBM WebSphere Application Server или как портлеты в среде IBM WebSphere Portal. При помощи API поиска и индексации можно разработать прикладные программы поиска, которые, как и пример программы поиска, будут одинаково работать в обеих средах.

Настроить программу поиска вам поможет настройщик программ поиска. Эта программа позволяет выбирать опции в графическом интерфейсе и видеть результаты сделанных изменений по мере их внесения. При сохранении этих изменений будет обновлен файл конфигурации программы поиска.

Совет:

Подробные примеры способов использования настройщика программы поиска и установки программы поиска как портлета в WebSphere Portal смотрите в документе IBM Redbook, IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Понятия, связанные с данным

“Индексированные опции для поиска документов” на стр. 182

“Идентификаторы программ поиска” на стр. 264

“Защита на уровне документов” на стр. 265

-  Обзор API поиска и индексирования
-  Синтаксис запросов
-  Задание свойств запроса

Задание прикладных программ поиска для собраний

Чтобы использовать новую прикладную программу поиска, сначала надо связать ее с собраниями, в которых она может выполнять поиск.

Прежде чем приступать к работе...

Задать для прикладных программ поиска собрания, в которых они могут выполнять поиск, может пользователь с ролью администратора поисковой системы предприятия.

Процедура

Чтобы связать прикладную программу поиска с одним или несколькими собраниями:

1. Выберите **Защита** на панели инструментов консоли администратора.
2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать прикладные программы поиска**.
3. На странице Конфигурировать прикладные программы поиска нажмите кнопку **Добавить прикладную программу поиска**.
4. Введите имя прикладной программы поиска.
5. Выберите собрания, в которых прикладная программа поиска сможет выполнять поиск:
 - Нажмите кнопку **Все собрания и внешние источники**, если вы хотите, чтобы прикладная программа поиска работала со всеми собраниями, которые вы добавили в систему.
 - Чтобы разрешить программе поиска доступ только к указанным собраниям, выберите **Отдельные собрания и внешние источники**.
Если выбрана эта опция, появляется список имен собраний и имен внешних источников. Включите переключатели **Выбрать** для всех собраний, в которых может выполнять поиск эта прикладная программа.
6. Нажмите кнопку **ОК**.

Функции примера программы поиска

Пример программы поиска для поисковой системы предприятия иллюстрирует большинство функций поиска, которые можно встраивать в пользовательские программы поиска.

Пример программы поиска можно использовать для поиска по всем собраниям и внешним источникам одновременно. Если не изменены свойства программы по умолчанию, эту программу можно использовать для поиска во всех собраниях и внешних источниках в поисковой системе предприятия.

Функции запроса

Эти функции позволяют:

- Задавать простые запросы в виде свободного текста.

- Задавать более сложные запросы для улучшения точности результатов поиска. Например, можно выполнить поиск по определенным полям или элементам XML или при помощи синтаксиса запроса найти документы, содержащие или не содержащие конкретные слова и словосочетания.
- Задавать, в каких собраниях и внешних источниках выполнять поиск.
- Выполнять поиск в источниках конкретных типов или же всех типов.
- Выполнять поиск в документах определенных типов. Например, можно выполнить поиск только в документах Microsoft Word или только в документах PDF.
- Указывать, на каком языке заданы термины запроса. Можно также задать языки документов, в которых будет выполняться поиск.
- Выполнять поиск в определенных подмножествах собрания. Например, программа поиска может ограничить видимость до заранее определенного диапазона документов (области действия), или передать запрос, выполняющий поиск только в документах названной категории.
- Расширять запрос, добавляя синонимы терминов запроса. Если с собранием связан словарь синонимов, документы, содержащие синонимы терминов запроса, возвращаются в результатах поиска.

Функции результатов поиска

Эти функции позволяют:

- Посмотреть результаты поиска, соответствующие вашим терминам запроса.
- Задавать число найденных документов на каждой странице и прокручивать набор результатов вперед и назад.
- Скрывать или выводить подробности о документах результатов. Например, можно просмотреть краткие описания документов или такие подробности, как имена полей в каждом документе результатов.
- Объединять документы из одного источника. Например, если из одного источника возвращено 100 документов, в наборе результатов выводится группа из двух максимально соответствующих документов. Чтобы просмотреть остальные 98 документов, надо выбрать опцию для просмотра большего числа документов из того же источника.
- Сортировать документы по релевантности, по дате документа или по значениям в конкретном поле. При сортировке по дате или по полю можно задать просмотр результатов в восходящем или нисходящем порядке.
- Выбирать из вариантов исправлений орфографии в случае обнаружения в строке запроса слов с возможными ошибками.
- Просматривать информацию о категориях, к которым принадлежит найденный документ (если собрание использует категории), и просматривать только документы, принадлежащие к определенной категории.
- Задавать дополнительные термины запросов в результатах поиска.

Функции получения документов

Эти функции позволяют:

- Получать документ, щелкнув по URI документа и открыв его в браузере. Если искатель Примечания или Domino Document Manager сконфигурирован для использования протокола DIIOP, документы, просмотренные этими искателями, можно вывести на экран с помощью клиентской программы просмотра Lotus Notes вместо браузера.

Если сконфигурирована защита уровня документов для искателя, получать документы могут только пользователи с правами доступа к защищенному содержимому.

- Получать документ, щелкнув по прямой ссылке. Прямая ссылка связывает ключевые слова с определенными URI. Если запрос содержит заданные ключевые слова, соответствующие им URI (их высокая релевантность для этих ключевых слов определена заранее) выводятся в начале списка результатов поиска.

Свойства программы поиска

Чтобы задать опции для вашей среды, изменить внешний вид программы и выбрать опции, доступные пользователям после ее запуска, можно отредактировать файл конфигурации для программы поиска.

Свойства также можно отредактировать при помощи настройщика программ поиска. При использовании настройщика результат внесенных изменений можно увидеть немедленно. Если вы удовлетворены заданными опциями для поиска в собраниях и просмотра результатов поиска, можно сохранить эти опции, изменив файл конфигурации для программы поиска.

Важное замечание: Если вы запустили программу поиска как портлет в WebSphere Portal, нельзя использовать Настройщик программы поиска для интерактивного внесения изменения в программу. Редактировать свойства и конфигурировать экземпляр портлета надо через интерфейс управления WebSphere Portal.

Файл конфигурации примера программы поиска для поисковой системы предприятия - `config.properties`. в данном разделе обсуждаются свойства, задаваемые этим файлом, и описываются их значения по умолчанию. Если вы создаете файлы конфигурации для пользовательских программ поиска, свойства, задаваемые этими файлами и значения этих свойств будут другими.

Параметры среды

Вы можете задать опции, управляющие работой программы поиска.

applicationName

Задает имя допустимой программы поиска. Значение по умолчанию - Default.

Измените это значение, если хотите в качестве программы по умолчанию использовать другую программу.

Совет: Оставив имя программы Default, можно использовать программу примера поиска, выполняющую поиск во всех собраниях и внешних источниках при помощи одного запроса.

timeout

Задает срок ожидания ответа от поискового сервера в секундах для требования поиска. Надо ввести целое число (например, 60, но не 60,5 и не "шестьдесят"). Если срок ожидания не задан, по умолчанию используется значение 30 секунд.

hostname

Задает полное имя хоста для Web-сервера, сконфигурированного для поддержки вашего экземпляра WebSphere Application Server. По умолчанию используется значение localhost.

Чтобы программа поиска работала правильно, измените значение по умолчанию на полное имя хоста, которое задано в конфигурации WebSphere Application Server. Например, если имя хоста локального компьютера - MyMachine, а имя хоста Web-сервера - www.ibm.com, задайте `www.ibm.com`.

protocol

Задает протокол для связи с Web-сервером: `http` или `https`. Если поле оставлено пустым, значение по умолчанию - `http`.

port Задает номер порта Web-сервера, который сконфигурирован для поддержки вашего экземпляра WebSphere Application Server. Значение по умолчанию - 80 (обычное значение при использовании протокола HTTP). Для протокола HTTPS обычно используется порт 443.

trustStore

Если вы используете протокол HTTPS, задайте полный путь для файла склада ключей (файл базы данных, содержащий общедоступные ключи). Информация *trustStore* позволяет использовать протокол Secure Sockets Layer (SSL) для доверенной связи. Указывая путь в Windows, вместо одной обратной дробной черты используйте две. Например: `x:\Application Server\webserver.key`

trustPassword

Если вы используете протокол HTTPS, укажите пароль для заданного файла склада ключей.

username

Программа поиска автоматически устанавливает для этого значения имя пользователя, введенное пользователем при регистрации в этой программе поиска. Указывайте здесь имя пользователя, только если вы хотите изменить поведение по умолчанию при аутентификации пользователей. Это поле используется, только если в WebSphere Application Server включена глобальная защита.

password

Программа поиска автоматически устанавливает для этого значения пароль, введенный пользователем при регистрации в этой программе поиска. Пароль надо задавать, только если введено имя пользователя. Это поле используется, только если в WebSphere Application Server включена глобальная защита.

ssoCookieName

Задает имя файла cookie, содержащего строку маркера единой регистрации (SSO). Значение по умолчанию - `LtraToken`.

proxyHost

Задает полное имя хоста прокси-сервера, если для доступа к поисковому серверу требуется прокси-сервер.

proxyPort

Задает порт для заданного хоста прокси-сервера.

proxyUser

Задает имя пользователя, используемое для регистрации на прокси-сервере, если прокси-сервер требует базовой аутентификации.

proxyPassword

Задает пароль для заданного имени пользователя прокси-сервера.

filter

Задает класс для получения документов, перечисленных в списке результатов поиска. Класс по умолчанию - `com.ibm.es.api.filters.SetDocumentURIFilterFetch`. Это значение надо

изменять, только если у вас есть пользовательский класс, который надо использовать для получения документов вместо класса по умолчанию.

logging.level

Задает степень подробности журнала:

OFF В журнал не записывается никаких сообщений.

SEVERE

В журнал заносятся только сообщения о серьезных ошибках. Это значение по умолчанию.

INFO В журнал записываются информационные сообщения.

FINE Сообщения трассировки заносятся в журнал с низкой степенью детализации. (Эта опция соответствует уровню записи в журнал FINE в классе Java `java.util.logging.Level`.)

ALL В журнал записываются все сообщения.

Значки типов источников

Можно настроить изображения для типов источников данных, к которым принадлежат документы - результаты поиска. Перечисленные ниже значки типов источников (соответствующие искомателям и внешним источникам, поддерживаемым OmniFind Enterprise Edition после установки) предопределены в файле `config.properties`.



documentSource.vbr.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем Content Edition. Значок по умолчанию - `/images/sourceVBR.gif`.



documentSource.db2.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем DB2. Значок по умолчанию - `/images/sourceDB2.gif`.



documentSource.cm.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем DB2 Content Manager. Значок по умолчанию - `/images/sourceCM.gif`.



documentSource.dominodoc.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем Domino Document Manager. Значок по умолчанию - `/images/sourceDominoDoc.gif`.



documentSource.exchange.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем Exchange Server. Значок по умолчанию - `/images/sourceExchange.gif`.



documentSource.database.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искомателем База данных JDBC. Значок по умолчанию - `/images/sourceJDBC.gif`.

**documentSource.nntp.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем NNTP. Значок по умолчанию - /images/sourceNNTP.gif.

**documentSource.notes.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Примечания. Значок по умолчанию - /images/sourceNotes.gif.

**documentSource.quickplace.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем QuickPlace. Значок по умолчанию - /images/sourceWorkplace.gif.

**documentSource.seedlist.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Список заполнения. Значок по умолчанию - /images/sourceSeedlist.gif.

**documentSource.unixfs.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Файловая система UNIX. Значок по умолчанию - /images/sourceUnixFS.gif.

**documentSource.web.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Web. Значок по умолчанию - /images/sourceWeb.gif.

**documentSource.wcm.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Web Content Management. Значок по умолчанию - /images/sourceWorkplace.gif.

**documentSource.wps.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем WebSphere Portal. Значок по умолчанию - /images/sourceWPS.gif.

**documentSource.winfs.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Файловая система Windows. Значок по умолчанию - /images/sourceWindowsFS.gif.

**documentSource.ldap.icon**

Задает путь и имя файла изображения, означающего, что документ относится к внешнему источнику, созданному для сервера LDAP. Значок по умолчанию - /images/sourceLDAP.gif.

**documentSource.jdbc.icon**

Задает путь и имя файла изображения, означающего, что документ относится к внешнему источнику - таблице базы данных JDBC (Java Database Connectivity). Значок по умолчанию - /images/sourceJDBC.gif.

Значки клиентских программ просмотра

Документы - результаты поиска могут выводиться в браузере. Документы, обработанные искателями Примечания crawlers или Domino Document Manager, сконфигурированными для использования протокола DIIOP, можно также просматривать в клиентской программе просмотра Lotus Notes.

Чтобы документы можно было просматривать в клиентской программе просмотра, для следующего свойства должно быть установлено значение true:

```
clientViewer.show=true
```

Можно настроить изображения для клиентских программ просмотра. В следующем примере значок Lotus Notes указывает, что документ можно вывести клиентской программой просмотра:

```
client.notes.icon=/images/notes.gif  
client.dominodoc.icon=/images/notes.gif
```

В списке результатов поиска этот значок и ссылка на клиентскую программу просмотра выводятся так:



Клиентская программа просмотра

Поля документа

Для типов источников данных, в которых есть поля, можно задать, какие поля выводятся в списке документов - результатов поиска.

fields.URI prefix=*список_имен_полей_через_пробел*

Перед символом двоеточия (:) в префиксе URI необходимо указывать символ обратной косой черты (\). Чтобы продолжить список имен полей на другой строке, введите в конце предыдущей строки символ обратной косой черты (\). Например:

```
fields.db2\://=databasename tablename  
fields.domino\://=databasetitle filename creator  
fields.dominodoc\://=librarydbtitle documentdbtitle filename author  
fields.exchange\://=from creator  
fields.file\://=directory filename  
fields.https\://=documentID  
fields.http\://=documentID  
fields.jdbc\://=databasename tablename  
fields.news\://=group from  
fields.quickplace\://=placetitle roomtitle creator  
fields.seedlist\://=author  
fields.vbr\://=itemname repositorytype revisionuser  
fields.wcm\://=author owner modifier  
fields.web\://=  
fields.wp6\://=  
fields.wps\://=
```

Значки полей

Для типов источников данных и документов, в которых есть поля, можно настроить изображения, соответствующие полям. Все поля выше сводной информации о документе содержат идентифицирующие их изображения. В файле `config.properties` предопределены следующие значки полей.

**field.icon.databasetitle**

Задает путь и имя файла изображения, означающего, что поле содержит название документа. Значок по умолчанию - /images/notesdb.gif.

**field.icon.databasename**

Задает путь и имя файла изображения, означающего, что поле содержит имя базы данных, к которой принадлежит документ. Значок по умолчанию - /images/db2.gif.

**field.icon.tablename**

Задает путь и имя файла изображения, означающего, что поле содержит имя таблицы, к которой принадлежит документ. Значок по умолчанию - /images/table.gif.

**field.icon.directory**

Задает путь и имя файла изображения, означающего, что поле содержит имя каталога, к которому принадлежит документ. Значок по умолчанию - /images/closedFolder.gif.

**field.icon.filename**

Задает путь и имя файла изображения, означающего, что поле содержит имя файла документа. Значок по умолчанию - /images/document.gif.

field.icon.documentID

Задает путь и имя файла изображения, означающего, что поле содержит идентификатор документа. Это пустое изображение можно использовать с документами Web, например, чтобы задать изображение для URL, но не показывать это изображение пользователю. Значок по умолчанию - /images/dot.gif.

**field.icon.group**

Задает путь и имя файла изображения, означающего, что поле содержит идентификатор документа. Это пустое изображение можно использовать с документами Web, например, чтобы задать изображение для URL, но не показывать это изображение пользователю. Значок по умолчанию - /images/document.gif.

**field.icon.from**

Задает путь и имя файла изображения, означающего, что поле идентифицирует отправителя документа. Значок по умолчанию - /images/author.gif.

**field.icon.creator**

Задает путь и имя файла изображения, означающего, что поле идентифицирует создателя документа. Значок по умолчанию - /images/author.gif.

**field.icon.author**

Задает путь и имя файла изображения, означающего, что поле идентифицирует автора документа. Значок по умолчанию - /images/author.gif.



field.icon.revisionuser

Задает путь и имя файла изображения, означающего, что поле идентифицирует того, кто пересмотрел документ. Значок по умолчанию - /images/author.gif.



field.icon.owner

Задает путь и имя файла изображения, означающего, что поле идентифицирует владельца документа. Значок по умолчанию - /images/author.gif.



field.icon.modifier

Задает путь и имя файла изображения, означающего, что поле идентифицирует того, кто изменил документ. Значок по умолчанию - /images/author.gif.

Значок поля по умолчанию

Можно задать изображение, используемое, если для полей, выводимых в списке результатов поиска, не заданы значки полей. В файле `config.properties` предопределен следующий значок поля по умолчанию.



field.defaultIcon

Задает путь и имя файла изображения, содержащего значок по умолчанию для полей в списке результатов поиска. Значок по умолчанию - /images/database.gif.

Поля дат

Можно задать, какие поля являются полями дат. Данные из полей, имена которых здесь заданы, форматируются в списке результатов поиска как даты. Формат даты соответствует параметрам национальной версии, заданным в браузере.

date.fields=*список_имен_полей_через_пробел*

Чтобы продолжить список имен полей на другой строке, введите в конце предыдущей строки символ обратной косой черты (\).

Пример:

```
date.fields=modifieddate createddate
```

Названия документов

Вы можете задать альтернативные названия документов путем замены текста названия более информативными данными (то есть названия можно *корректировать*). Например, вместо вывода названий документов с неинформативной меткой Слайд 1 можно задать, что в списке результатов поиска название Слайд 1 должно подавляться. Для идентификации документа можно использовать более осмысленное поле, например, имя файла.

Кроме того, можно задать альтернативные названия документов путем удаления неинформативных слов из названий документов (то есть названия можно *сокращать*). Например, если названия некоторых документов - результатов поиска начинаются с Microsoft Word, можно улучшить читаемость результатов поиска путем подавления повторяющегося начала текста.

titles.clean=*список_названий_через_запятую*

titles.truncatePrefix=*список_префиксов_через_запятую*

В этих списках через запятую задаются значения, которые могут содержать пробелы и другие символы, кроме запятой. Чтобы продолжить список на другой строке, введите в конце предыдущей строки символ обратной косой черты (\).

Например:

```
titles.clean=Slide 1, Layout 1, untitled, \
Untitled Document, PowerPoint Presentation, \
(страница без имени)
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

Значения по умолчанию для предпочтений пользователя

Можно задать значения по умолчанию для страницы программы поиска Предпочтения. Если пользователь изменяет эти предпочтения, новые значения действуют только в течение текущего сеанса пользователя. В файле `config.properties` предопределены следующие предпочтения.

preferences.resultsRange=10

Задаёт, что на каждой странице списка результатов поиска может выводиться до 10 документов - результатов поиска.

preferences.siteCollapsing=Yes

Указывает, что URI из одного и того же источника объединяются в результатах поиска. Объединение сайтов доступно, только когда результаты отсортированы по релевантности. Для источников данных Web и NNTP все URI, соответствующие URI корневого сайта (например, `www.ibm.com`) объединяются автоматически. Для других типов источников данных и для сайтов с более глубокими уровнями путей (такими как `www.ibm.com/hr`) правила объединения сайтов должны быть сконфигурированы на консоли администратора поисковой системы предприятия.

preferences.spellCorrections=Yes

Включает вывод вариантов исправления ввода, когда пользователь отправляет запрос с возможной ошибкой в написании. Обратите внимание, что перед вычислением возможных вариантов написания всегда удаляются стоп-слова.

preferences.extendedHighlighting=No

Задаёт, что в дополнительных полях (например, в названии документа) не должны выделяться термины запроса (кроме поля сводки документа).

Собрания и внешние источники по умолчанию

Можно задать, какие собрания и внешние источники изначально выбраны на страницах Предпочтения и Сложный поиск. Пользователи могут отредактировать набор по умолчанию, чтобы выполнять поиск не во всех собраниях и внешних источниках, сделанных доступными по умолчанию. Если здесь ограничить набор собраний и внешних источников, пользователи могут выбрать любое собрание или внешний источник, доступные для программы поиска, при изменении предпочтений или опций сложного поиска.

preferences.defaultCollections=*

preferences.defaultCollections=*список_ID_собраний_через_пробел*

Укажите символ звездочки (*), чтобы разрешить поиск во всех собраниях и внешних источниках. (Собрания и внешние источники надо связать с

программой поиска на консоли администратора поисковой системы предприятия.) Это значение по умолчанию, заданное в файле `config.properties`.

Чтобы ограничить область поиска, доступную пользователям, если они не изменят предпочтения или расширенные опции поиска, укажите ID собраний для собраний и внешних источников, в которых пользователи будут выполнять поиск по умолчанию.

Например:

```
preferences.defaultCollections=*  
preferences.defaultCollections=coll_id1 coll_id2
```

Дополнительная информация для результатов поиска

Можно настроить объем информации, выводимой в списке результатов поиска, а также включить или выключить возможность фильтрации результатов поиска пользователями. По умолчанию в файле `config.properties` заданы следующие значения.

refreshButton.show=false

Определяет, выводится ли кнопка **Обновить** на странице простого поиска. Для сложного поиска кнопка **Обновить** доступна всегда. Если задать для этой опции значение `true`, пользователи смогут обновлять список собраний и внешних источников, доступных для поиска.

Если вы используете настройщик программ поиска, кнопка **Обновить** не нужна.

Если вы не используете настройщик программ поиска, возможно, вы захотите включить вывод кнопки **Обновить** для проверки изменений, сделанных в файле конфигурации. Сохранив изменения, можно нажать кнопку **Обновить**, чтобы посмотреть, как эти изменения повлияют на программу поиска. Если кнопка **Обновить** не выводится, надо перезапустить прикладную программу `ESSearchServer` в `WebSphere Application Server`, чтобы изменения вступили в силу.

Если для поиска не доступны никакие собрания или внешние источники (например, если указано неправильное имя хоста, не запущены серверы поиска или в `WebSphere Application Server` не запущена прикладная программа `ESSearchServer`), кнопка **Обновить** выводится автоматически, чтобы помочь исправить эту проблему.

builtQueryString.show=false

Управляет выводом полностью развернутого синтаксиса запроса в области перед списком документов - результатов поиска. Включите эту опцию, чтобы был виден фактически обработанный запрос.

extraQueryData.show=false

Управляет выводом дополнительной информации о запросе. Установите для этой опции значение `true`, если вы хотите увидеть информацию об ограничениях списков управления доступом, именах собраний и внешних источников, в которых будет выполняться поиск, а также о языке запроса.

refineResults.show=true

Разрешает или запрещает пользователям уточнять результаты поиска путем задания дополнительных терминов запроса. Если задать для этой опции значение `true`, в нижней части страницы результатов поиска выводится окно ввода запроса с меткой **Поиск в результатах**.

sorting.show=true

Управляет выводом опции для сортировки результатов поиска. Установите для этой опции значение false, чтобы подавить вывод опций **Сортировать по** и **Порядок сортировки** для сортировки результатов поиска.

sourceTypeFilter.show=true

Управляет выводом в результатах поиска опции для фильтрации результатов по типам источников. Задайте для этой опции значение false, если не хотите позволить пользователям фильтровать результаты по типам источников.

Чтобы пользователи не могли фильтровать результаты по типу документа, удалите отдельные или все записи типов документов в файле конфигурации (`documentType.метка=типы_документов`).

filter.showOnTwoLines=true

Управляет выводом опций фильтрации результатов по типу источника и по типу файла на одной или на двух строках в списке результатов поиска. При просмотре результатов поиска пользователи могут выбрать тип источника и тип файла для вывода только документов - результатов поиска, соответствующих выбранным фильтрам.

Чтобы максимизировать пространство, доступное для вывода результатов поиска, задайте для этого свойства значение false. Чтобы улучшить читаемость фильтров, в особенности если доступные фильтры занимают больше одной строки, можно задать для этого свойства значение true, тогда каждый фильтр выводится на отдельной строке.

clientViewer.show=true

Определяет, будет ли использоваться для вывода полученного документа клиентская программа просмотра Lotus Notes. Задайте для этой опции значение false, если не хотите использовать эту программу просмотра для документов Domino.

showDetails.show=true

Управляет выводом ссылок Показать подробности и Скрыть подробности в результатах поиска. Задайте для этой опции значение false, если не хотите, чтобы пользователи могли просматривать дополнительные подробности о полученных документах.

showDetailsImage.show=true

Управляет выводом ссылок подробностей полученных документов в окне. Задайте для этой опции значение false, если не хотите, чтобы пользователи могли просматривать дополнительные подробности о полученных документах, помещая курсор на URI документа.

numberSearchResultsReturned.show=true

Определяет, выводится ли общее число результатов поиска. Задайте для этой опции значение false, если не хотите, чтобы пользователи видели, сколько документов возвращено в результатах поиска.

showMessage.error=true

Управляет выводом сообщений об ошибках. Задайте для этой опции значение false, если не хотите, чтобы сообщения об ошибках выводились в верхней части окна программы поиска.

showMessage.warning=true

Управляет выводом предупреждений. Задайте для этой опции значение false, если не хотите, чтобы предупреждения выводились в верхней части окна программы поиска.

showMessage.info=true

Управляет выводом информационных сообщений. поисковой этой опции значение false, если не хотите, чтобы информационные сообщения выводились в верхней части окна программы поиска.

showMessage.success=true

Управляет выводом сообщений об успешном выполнении. поисковой этой опции значение false, если не хотите, чтобы сообщения об успешном выполнении действий выводились в верхней части окна программы поиска.

Настройка баннера и логотипа

Можно настроить изображения, которые выводятся в области баннера в верхней части программы поиска. Например, можно заменить изображения по умолчанию для OmniFind Enterprise Edition на изображения с символикой вашего предприятия. Чтобы не выводить баннер, прокомментируйте одну или обе этих строки. Свойство banner.icon задает рисунок, который выводится в левой части области баннера. Свойство banner2.icon задает рисунок, который выводится в правой части области баннера.

```
banner.icon=/images/WS_II_0FEdition.gif  
banner2.icon=/images/WS_II_mosaic.gif
```

Настройка фонового изображения

Можно настроить изображения, которые выводятся в качестве фона страниц в программе поиска. Например, можно заменить изображения по умолчанию для поисковой системы предприятия на изображения с символикой вашего предприятия. Чтобы не выводить фоновое изображение, прокомментируйте одну или несколько из этих строк.

```
search.backgroundImage=/images/II0F_search.gif  
preferences.backgroundImage=/images/II0F_options.gif  
advanced.backgroundImage=/images/II0F_advanced.gif  
browse.backgroundImage=/images/II0F_tree.gif  
myProfile.backgroundImage=/images/II0F_profile.gif  
logoff.backgroundImage=/images/II0F_logout.gif
```

Ссылки

Свойства в области Ссылки файла config.properties позволяют выводить имена страниц программы поиска в виде ссылок на каждой странице вместо вывода на панели инструментов и на страницах со вкладками. Вывод ссылок полезен, если программа поиска работает как портлет, и нужно минимизировать пространство, которое она занимает на странице портала.

Если вы предпочитаете перемещаться по программе поиска при помощи выбора опций на панели инструментов и на страницах со вкладками, прокомментируйте эти строки.

Вкладки поиска

Свойства в области файла config.properties Вкладки поиска задают имена страниц JSP (Java Server Pages), используемых для страниц со вкладками в окне программы поиска Поиски (Простой поиск, Сложный поиск и Дерево категорий). Не редактируйте эти страницы, если у вас нет опыта программирования на Java и работы с JSP.

Вот некоторые примеры возможной настройки этой области:

- Направление программы поиска на пользовательские страницы JSP, придающие другой внешний вид страницам со вкладками.
- Комментирование записей Дерево категорий. Например, если для собраний не сконфигурированы категории, нет смысла выводить в программе поиска страницу Дерево категорий.
- Копирование записей страниц со вкладками в область файла `config.properties` Панели инструментов и комментирование этих строк. Например, можно включить только вывод панели инструментов и совсем не выводить страницы со вкладками.

Панели инструментов

Свойства в области файла `config.properties` Панели инструментов задают имена страниц JSP (Java Server Pages), используемых для панели инструментов программы поиска. Не редактируйте эти страницы, если у вас нет опыта программирования на Java и работы с JSP.

Вот некоторые примеры возможной настройки этой области:

- Направление программы поиска на пользовательские страницы JSP, придающие другой внешний вид панели инструментов.
- Комментирование записей панели инструментов для ссылок, которые вы не хотите выводить. Например, можно отключить на панели инструментов вывод ссылки на страницу О программе.
- Перемещение функции вывода страницы Сложный поиск из области файла `config.properties` вкладки, чтобы эта опция была доступна только на панели инструментов.

Содержательные метки типов документов

Можно улучшить читаемость фильтра по типу документов, если вместо фактических имен типов документов выводить более сжатые и выразительные названия. Типы документов, доступные программам поиска, определяются классом `SIAPI` (Search and Index API - интерфейс прикладного программирования для поиска и индексов) `AvailableDocumentTypes`. Кроме того, в целях удобства доступные типы документов перечислены в конце файла `config.properties`.

documentType.метка=список_типов_документов_через_пробел

Задает имя, которое выводится в списке результатов поиска в строке фильтра по типу документов, и список фактических типов документов, которые выводятся, если пользователь выберет этот фильтр.

Например, можно задать метку `html` и назначить этому имени расширения файлов и типы MIME для различных документов Web. Если пользователь щелкнет по **html**, чтобы отфильтровать результаты поиска, будут выведены только документы с заданными расширениями и типами MIME.

В файле `config.properties` predefinedены следующие отображения типов документов:

```
documentType.html=shtml text/html html xhtml htm
documentType.doc=doc application/msword
documentType.ppt=application/mspowerpoint ppt
documentType.xls=xls application/x-excel application/msexcel \
application/x-msexcel application/excel application/vnd.ms-excel
documentType.xml=xml text/xml
documentType.txt=txt text/plain
documentType.pdf=pdf application/pdf
```

Если значение, заданное вами для метки типа документов, совпадает со значением для свойства в файле `application.properties`, выводится не

заданное вами здесь значение, а значение свойства в файле `application.properties`. Например, если в качестве метки для фильтра типов документов задать значение `documentType.unixfs`, в качестве имени фильтра типов файлов в виде ссылки будет выводиться значение свойства `unixfs` в файле `application.properties` (**файловая система UNIX**).

Пользовательские фильтры

Можно задать пользовательские запросы для фильтрации вывода документов - результатов поиска.

filterCustom.метка=термины_запроса

Задаёт имя, которое выводится в списке результатов поиска в строке пользовательского фильтра, и запрос, уточняющий результаты поиска, если пользователь выберет этот фильтр. (При просмотре результатов поиска пользователи могут выбрать пользовательский фильтр для вывода только документов - результатов поиска, соответствующих предопределённому запросу.)

В следующем примере результаты поиска фильтруются для вывода только документов, относящихся к базе данных отдела кадров (`hr` - human resources):

```
filterCustom.HR_database_only=databasename::hr
```

Если пользователь нажмет **HR_database_only**, чтобы отфильтровать результаты поиска, будет обработан запрос `databasename::hr`. При выводе результатов поиска будут выведены только документы из базы данных `hr`.

Если значение, заданное вами для метки пользовательского фильтра, совпадает со значением для свойства в файле `application.properties`, выводится не заданное вами здесь значение, а значение свойства в файле `application.properties`. Например, если в качестве метки пользовательского фильтра задать значение `filterCustom.hostData`, в качестве имени пользовательского фильтра в виде ссылки будет выводиться значение свойства `hostData` в файле `application.properties` (**Параметры сервера**).

Несколько пользовательских фильтров закомментированы и содержатся в файле `config.properties` в качестве примера.

Обнаружение дубликатов

При добавлении документов в индекс поисковой системы предприятия выполняется анализ с целью удаления дубликатов, чтобы пользователи не видели повторяющихся документов в результатах поиска. Для дальнейшей фильтрации результатов поиска можно задать опцию, подавляющую вывод документов, почти идентичных друг другу, и предотвращающую их вывод в результатах поиска.

preferences.nearDuplicateDetection=No

Задаёт, что очень похожие документы не будут фильтроваться при обработке запросов.

Если задать значение `Yes`, вывод документов со сходными названиями и сводками при просмотре пользователем результатов поиска будет подавляться. Сообщение информирует пользователей, что некоторые документы пропущены, поскольку они похожи на другие документы в наборе результатов. Чтобы отключить подавление и просмотреть все документы в наборе результатов, пользователь может щёлкнуть по соответствующей ссылке.

Для подавления почти одинаковых документов в объекте SI-API Query (Search and Index API - API поиска и индексации) для программы поиска должен быть задан метод `setProperty` со строкой `NearDuplicateDetection`, в которой указано значение `Yes` (например:
`query.setProperty("NearDuplicateDetection", "Yes");`).

Анализ первых результатов (столбчатые диаграммы для полей метаданных)

Можно задать опции для вывода столбчатых диаграмм, представляющих анализ первых результатов. Опция по умолчанию - анализировать первые 500 результатов. Каждая диаграмма соответствует одному полю метаданных, а каждый столбик в ней соответствует значению поля. Длина столбика показывает частоту встречаемости значения поля. Чем длиннее столбик, тем чаще встречается это значение в поле.

При конфигурировании пространства для искателя можно задать опции полей метаданных. Чтобы сконфигурировать диаграмму первых результатов для поля метаданных, нужно задать, что это поле допускает поиск и его можно возвращать в результатах поиска.

Для свойств, конфигурируемых вами для столбчатой диаграммы, используется следующий формат (где *номер* - номер, уникально идентифицирующий столбчатую диаграмму в программе поиска, *опция* - опция столбчатой диаграммы, а *значение* - значение этой опции):

`topResultsChartsномер.опция=значение`

topResultsChartsномер.titleKey=ключ_программы

Задаёт заголовок для диаграммы, где *номер* - номер, уникально идентифицирующий столбчатую диаграмму в программе поиска, а *ключ_программы* - метка для заголовка диаграммы. Этой меткой может быть ключ в файле `application.properties` для программы поиска или заданное вами здесь значение. В следующем примере метку для заголовка диаграммы задаёт значение ключа `topResults.mostRecentDocuments` в файле `application.properties`:
`topResultsCharts3.titleKey=topResults.mostRecentDocuments.`

Другой пример: заголовок диаграммы представляет собой точное значение, заданное вами здесь: `topResultsCharts3.titleKey=Organizations`

topResultsChartsномер.enable=true

Задаёт, будет ли эта столбчатая диаграмма выводиться при просмотре результатов поиска пользователями. Если задать значение `false`, диаграмма выводиться не будет.

topResultsChartsномер.fieldName=имя_поля

Задаёт имя поля метаданных, значения которого следует анализировать для данной диаграммы. Например, `datasSetTitle`. Для каждой добавляемой вами диаграммы нужно задать уникальное имя поля (использовать одно и то же поле в нескольких диаграммах нельзя).

topResultsChartsномер.maxValues.collapsed=число

Задаёт число свернутых элементов для вывода в этой диаграмме. Например, задайте 5, чтобы в этом поле выводились первые пять наиболее часто встречающихся значений.

topResultsChartsномер.maxValues.expanded=число

Задаёт число элементов, выводящихся на полностью развернутой диаграмме. Например, если задать 10, в развернутой диаграмме будет показано не более 10 различных значений поля.

topResultsChartsномер.fieldValueSeparator=символ

Задает символ, разделяющий в этом поле анализируемые значения. Например, если поле содержит несколько значений, разделенных точкой с запятой (например, agent;seller;broker), это свойство позволяет определить в качестве разделителя значений поля точку с запятой (;), чтобы каждое значение можно было добавить в столбчатую диаграмму в виде отдельного элемента. Без этой опции все значение поля будет добавлено в столбчатую диаграмму как единый элемент.

topResultsChartsномер.canUserChangeFieldName=true

Задает, сможет ли пользователь при просмотре результатов поиска выбрать другое поле и вывести для него первые результаты. Если задать значение false, при просмотре результатов поиска пользователи не смогут выбрать другое поле для анализа.

Если задать значение true, помимо результатов для текущего поля будет выводиться список всех полей, найденных в первых 500 результатах. Если пользователь выбирает поле из этого списка, надпись диаграммы меняется на **Диаграмма динамических полей** и выводится столбчатая диаграмма результатов для выбранного поля, пока пользователь не выберет другое поле или не закроет браузер. При следующем запуске программы поиска снова выводятся результаты для исходного поля.

topResultsChartsномер.width=число

Задает ширину вывода столбчатой диаграммы (в пикселях). Например, 300.

topResultsChartsномер.barheight=число

Задает высоту каждого столбика в диаграмме (в пикселях). Например, 10.

topResultsChartsномер.color=#код_цвета

Задайте шестнадцатеричный код основного цвета столбика в столбчатой диаграмме. Значение по умолчанию - синий (#0309C0).

topResultsChartsномер.color.gradient=#код_цвета

Задает цвет перехода для столбика в столбчатой диаграмме. Значение по умолчанию - бирюзовый (#00FFFF). Цвет столбика с переходом от значений низкой частоты к значениям высокой частоты будет меняться от цвета, заданного опцией color до цвета, заданного опцией color.gradient.

Пример: .

topResultsChartsномер.sortKey=frequency

Задает способ сортировки элементов в столбчатой диаграмме.

Поддерживаемые значения:

нет Элементы на столбчатой диаграмме не сортируются.

метка Элементы на столбчатой диаграмме сортируются по значениям поля. Если с помощью свойства `topResultsCharts.numberListOfLabels.prefixKey=имя_поля` сконфигурированы пользовательские метки, элементы в столбчатой диаграмме будут также сортироваться по имени метки.

частота

Элементы на столбчатой диаграмме сортируются по числу возвращенных результатов с данным значением.

topResultsChartsномер.sortOrder=descending

Для элементов, сортируемых по имени метки или частоте, задает порядок сортировки. Поддерживаемые значения:

восходящий

При сортировке по меткам элементы сортируются в восходящем

алфавитном порядке. При сортировке по частотам значения поля сортируются так, что более частые значения будет ниже в списке, чем менее частые.

нисходящий

При сортировке по меткам элементы сортируются в нисходящем алфавитном порядке. При сортировке по частотам значения поля сортируются так, что более частые значения будет выше в списке, чем менее частые.

topResultsChartsномер.listOfLabels.prefixKey=имя_поля

Необязательный параметр. Позволяет указать информацию, которую всегда следует выводить для определенных значений поля (где *ИМЯ_ПОЛЯ* задает поле, значение которого анализируется для этой диаграммы). Метки, заданные для этого свойства, всегда выводятся в столбчатой диаграмме, даже если вхождения значения поля отсутствуют в результатах поиска.

Для поля можно сконфигурировать любое число пользовательских меток. Для каждой метки задаются два свойства в следующем формате:

*имя_поля*номер.value=значение

*имя_поля*номер.displayValue=значение_для_вывода

где:

имя_поля

Имя поля, для которого конфигурируются метки.

номер Номер, однозначно идентифицирующий метку.

значение

Значение, которое будет выводиться в результатах анализа.

значение_для_вывода

Метка для вывода в столбчатой диаграмме. Этой меткой может быть ключ в файле `application.properties` для программы поиска или заданное вами здесь значение.

Например:

```
topResultsCharts1.listOfLabels.prefixKey=databasetitle
```

```
databasetitle1.value=JK Enterprises Articles & Papers  
databasetitle1.displayValue=Articles & papers
```

```
databasetitle2.value=JK Enterprises Blank Forms  
databasetitle2.displayValue=Blank forms
```

```
databasetitle3.value=JK Enterprises Bulletins & Guidelines  
databasetitle3.displayValue=Bulletins & Guidelines
```

Анализ первых результатов (пользовательский HTML)

Можно задать опции для вывода диаграмм первых результатов, расширив существующий `API com.ibm.es.searchui.charts.servlet.AbstractDynamicChart` при помощи пользовательского класса Java. Если для вывода результатов используется этот подход, при помощи HTML можно отформатировать любые возвращаемые результаты поиска, а не только результаты запросов, выполняющих поиск в полях метаданных. Результаты могут быть получены для документов собраний поисковой системы предприятия либо возвращены из операций поиска во внешних репозиториях и сайтах.

Для программы поиска можно сконфигурировать любое число диаграмм. Назначьте для каждой диаграммы заголовок при помощи следующего свойства:

topResultsCharts*номер*.titleKey=*ключ_программы*

Задаёт заголовок для диаграммы, где *номер* - номер, уникально идентифицирующий столбчатую диаграмму в программе поиска, а *ключ_программы* - метка для заголовка диаграммы. Этой меткой может быть ключ в файле application.properties для программы поиска или заданное вами здесь значение. В следующем примере метку для заголовка диаграммы задаёт значение ключа topResults.mostRecentDocuments в файле application.properties:
topResultsCharts3.titleKey=topResults.mostRecentDocuments.

Каждый элемент в диаграмме соответствует одному значению результатов поиска и составляет набор свойств, для которых используется следующий формат (где *номер* - номер, уникально идентифицирующий диаграмму, *опция* - опция столбчатой диаграммы, а *значение* - значение этой опции):

topResultsCharts*номер*.опция=*значение*

topResultsCharts*номер*.enable=true

Задаёт, будет ли эта диаграмма выводиться при просмотре пользователями результатов поиска. Если задать значение false, диаграмма выводиться не будет.

topResultsCharts*номер*.maxValues.collapsed=*число*

Задаёт число свернутых элементов для вывода на этой диаграмме. Например, задайте 5 для вывода в этом поле первых пяти наиболее часто встречающихся результатов, которые соответствуют критериям.

topResultsCharts*номер*.width=*число*

Задаёт ширину вывода диаграммы (в пикселях). Например, 400.

topResultsCharts*номер*.dynamicChartClass=*пользовательский_класс*

Задаёт имя пользовательского класса Java, расширяющего API com.ibm.es.searchui.charts.servlet.AbstractDynamicChart и определяющего способ вывода элементов на диаграмме. Например:
topResultsCharts.3.dynamicChartClass=com.ibm.es.searchui.charts.servlet.DynamicMostRecentDocuments

В пользовательском классе Java DogearSearchResults примера используются следующие свойства (они включаются в файл config.properties по умолчанию примера программы поиска для примера). Информацию о параметрах, которые можно включить в пользовательскую программу поиска, смотрите в документации к API Dogear.

Следующие свойства нельзя задать при помощи настройщика программ поиска:

topResultsCharts*номер*.xsl.fileName=*файл_стилей*

Задаёт путь и имя таблицы стилей, которую следует использовать для форматирования вывода первых результатов на диаграмме. Например: /styles/dogear.xsl.

topResultsCharts*номер*.url=*url*

Задаёт URL, в котором будет выполняться поиск.

topResultsCharts*номер*.url.parameters=ps=*число*

Задаёт размер страницы. Например, при значении ps=3 максимальный размер страницы будет равен 3.

Это просто строка параметра, определяемая в API Dogear REST.

Анализ первых результатов (максимальное число результатов)

`topResult.resultSize=число`

Значение по умолчанию и максимальное число результатов, которые возвращает анализ первых результатов, равно 500. Это значение можно уменьшить, но нельзя увеличить. Например, это значение можно уменьшить если при анализе первых результатов происходят ошибки из-за истечения сроков ожидания запросов. Это свойство нельзя задать с помощью настройщика программ поиска.

Понятия, связанные с данным



Задание свойств запроса



Классы Java для вывода лучших результатов

Редактирование свойств примера программы поиска

Пример прикладной программы поисковой системы предприятия может искать информацию во всех активных собраниях и внешних источниках вашей системы. Можно редактировать файл конфигурации для задания опций среды вашего Web-сервера, для использования другой программы поиска в качестве программы поиска по умолчанию или для управления составом опций, выводимых при запуске программы поиска.

Об этой задаче

Программа установки внедряет пример программы поисковой системы предприятия в IBM WebSphere Application Server на поисковых серверах для поисковой системы предприятия. Для конфигурирования этой программы поиска надо редактировать файл конфигурации `config.properties`, который внедряется вместе с ней.

Чтобы изменения вступили в силу, остановите и перезапустите программу `ESSearchServer` на сервере прикладных программ `WebSphere`.

Процедура

Чтобы отредактировать пример свойств программы поиска:

1. Зарегистрируйтесь на сервере поиска как администратор поисковой системы предприятия.
2. Редактируйте файл `config.properties` обычным текстовым редактором.
Файл `config.properties` установлен в следующий каталог (КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES здесь - каталог установки OmniFind Enterprise Edition на поисковом сервере:
`КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF/config.properties`
3. Отредактируйте свойства, задав информацию о среде Web-сервера и предпочтениях поиска, затем сохраните и закройте файл. В этом файле знак решетки # указывает на строку комментария.
4. Остановите и перезапустите программу `ESSearchServer`.

AIX, Linux или Solaris

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

Windows

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Эти сценарии находятся в каталоге WAS_INSTALL_ROOT/AppServer/bin:

- Для WebSphere Application Server Версии 5 путь установки по умолчанию - /usr/WebSphere в системах AIX, /opt/WebSphere в системах Linux и Solaris, C:\Program Files\WebSphere в системах Windows.
- Для WebSphere Application Server Версии 6 путь установки по умолчанию - /usr/IBM/WebSphere в системах AIX, /opt/IBM/WebSphere в системах Linux и Solaris, C:\Program Files\IBM\WebSphere в системах Windows.

Понятия, связанные с данным



Задание свойств запроса



Классы Java для вывода лучших результатов

Задачи, связанные с данной

“Конфигурирование серверов поиска для приема только защищенных требований (по протоколу SSL)” на стр. 252

Настройка программы поиска

Настройщик программ поиска - это графический интерфейс, который можно использовать для настройки программ поиска поисковой системы предприятия или пользовательских программ поиска.

Ограничения

Настройщик программы поиска доступен как автономная программа. Его нельзя запустить в WebSphere Portal для настройки программ поиска, запускаемых как портлеты. Для настройки программ поиска, запускаемых как портлеты, надо редактировать свойства и конфигурировать экземпляр портлета через интерфейс управления WebSphere Portal.

Об этой задаче

Настройщик программ поиска позволяет наглядно представить изменения, которые вы хотите внести, и изменить программу поиска, не изменяя файл конфигурации. Например, можно изменить рисунки для шапки и фона, вид интерфейса поиска и задать опции для работы с результатами поиска.

Выбирая те или иные опции в настройщике программ поиска, вы сразу видите результаты такого выбора. Сохраняя изменения, вы тем самым изменяете файл конфигурации для программы поиска.

Чтобы изменения вступили в силу, остановите и перезапустите программу ESSearchServer на сервере прикладных программ WebSphere.

Совет:

Подробные примеры способов использования настройщика программ поиска смотрите в документе IBM Redbook, IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Процедура

Чтобы настроить программу поиска:

1. Чтобы настроить пример программы поиска для поисковой системы предприятия, введите URL для настройщика программ поиска в браузере. Например:

```
http://SearchServer.com/ESSearchApplication/palette.do
```

SearchServer.com - имя хоста поискового сервера.

Кроме того, если для вашего Web-сервера не задано использование порта 80, надо задать допустимый номер порта. Например:

```
http://SearchServer.com:9080/ESSearchApplication/palette.do
```

Совет: Если вы - администратор поисковой системы предприятия, можно также открыть настройщик программ поиска, выбрав опцию **Настройщик поиска** на консоли администратора поисковой системы предприятия.

2. Чтобы настроить пользовательскую программу поиска, введите URL для настройщика программ поиска и добавьте имя файла конфигурации для вашей программы поиска. Например:

```
http://SearchServer.com/ESSearchApplication/palette.do?configFile=/WEB-INF/myConfig.properties
```

Если заданный вами файл не существует, будут выведены значения из файла `config.properties` для примера программы поиска.

Совет: Можно также задать файл конфигурации, который будет использоваться с программой поиска, нажав кнопку **Загрузить** после запуска настройщика программ поиска и задав имя файла.

3. Если в WebSphere Application Server включена глобальная защита, зарегистрируйтесь с допустимым ID пользователя и паролем.
4. Выберите опции, которые хотите изменить, например, информацию о сервере поиска, типы выводимых сообщений, опции запросов и результатов поиска и имена образов, по которым будут определяться различные типы документов в результатах поиска.

Чтобы получить справку по заданию опций, выберите **Справка по настройщику**.

Чтобы увидеть действие некоторых изменений, таких как представление результатов поиска, введите запрос и нажмите кнопку **Поиск**.

5. Добившись желаемого результата, нажмите кнопку **Сохранить**, чтобы внести изменения в файл конфигурации. Если вы нажмете кнопку **Сброс**, в настройщике программ поиска будут восстановлены значения опций последней сохраненной версии файла конфигурации.
6. На сервере поиска зарегистрируйтесь как администратор поисковой системы предприятия, остановите и перезапустите программу `ESSearchServer`.

AIX, Linux или Solaris

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

Windows

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Эти сценарии находятся в каталоге `WAS_INSTALL_ROOT/AppServer/bin`:

- Для WebSphere Application Server Версии 5 путь установки по умолчанию - `/usr/WebSphere` в системах AIX, `/opt/WebSphere` в системах Linux и Solaris, `C:\Program Files\WebSphere` в системах Windows.

- Для WebSphere Application Server Версии 6 путь установки по умолчанию - /usr/IBM/WebSphere в системах AIX, /opt/IBM/WebSphere в системах Linux и Solaris, C:\Program Files\IBM\WebSphere в системах Windows.

Задачи, связанные с данной

“Конфигурирование серверов поиска для приема только защищенных требований (по протоколу SSL)” на стр. 252

Клонирование примера программы поиска

Чтобы использовать пример программы поиска как модель для создания собственных программ поиска, можно скопировать файл `config.properties`, либо использовать настройщик программ поиска.

Об этой задаче

Чтобы упростить создание пользовательских программ поиска, можно клонировать опции конфигурации, заданные для примера программы поиска, а затем выбрать опции, которые нужно изменить.

После клонирования примера программы поиска задайте имя файла конфигурации, чтобы запустить новую программу поиска. Необходимо также задать имя нового файла конфигурации, чтобы настроить новую программу поиска с помощью настройщика программ поиска.

Клонировав пример программы поиска, можно быстро создавать программы поиска для конкретных целей или конкретных пользователей. Например, можно создать одну программу поиска для работников отдела управления персоналом, а другую - для продавцов.

Чтобы изменения вступили в силу, остановите и перезапустите программу `ESSearchServer` на сервере прикладных программ `WebSphere`.

Процедура

Чтобы клонировать пример программы поиска:

1. Если вы хотите изменить файл конфигурации:
 - a. Скопируйте файл `config.properties` примера программы поиска, переименовав его.
Файл `config.properties` установлен в следующий каталог (КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES здесь - каталог установки `OmniFind Enterprise Edition` на поисковом сервере:
КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF/config.properties

Создаваемый файл должен находиться в подкаталоге `WEB-INF`.
 - b. Измените свойства, которые вы хотите использовать для пользовательской программы поиска, и сохраните этот файл. По меньшей мере, необходимо изменить свойство `applicationName`, чтобы задать имя пользовательской программы поиска.
2. Если вы хотите клонировать пример программы поиска, используя настройщик программ поиска:
 - a. Запустите настройщик программ поиска, добавив имя файла конфигурации, который вы хотите создать. В следующем примере создается файл с именем `myNewFile.properties`:

<http://ESServer.com/ESSearchApplication/palette.do?configFile=/WEB-INF/myNewFile.properties>

Так как этот файл пока не существует, используются значения, заданные в файле конфигурации примера программы поиска `config.properties`.

Совет: Можно также создать файл конфигурации для программой поиска, после запуска настройщика программ поиска нажав кнопку **Загрузить** и задав имя файла. Файл будет создан, когда вы нажмете кнопку **Сохранить**, чтобы сохранить опции настройки.

- b. Если в WebSphere Application Server включена глобальная защита, зарегистрируйтесь с допустимым ID пользователя и паролем.
 - c. Задайте имя вашей программы поиска, задайте опции для поиска в собраниях и нажмите кнопку **Сохранить**. Заданные изменения сохраняются в новом файле конфигурации в подкаталоге WEB-INF.
Чтобы получить справку по заданию опций, выберите **Справка по настройщику**. Чтобы увидеть действие некоторых изменений, таких как представление результатов поиска, введите запрос и нажмите кнопку **Поиск**.
3. На сервере поиска зарегистрируйтесь как администратор поисковой системы предприятия, остановите и перезапустите программу ESSearchServer.

AIX, Linux или Solaris

```
./stopServer.sh ESSearchServer  
./startServer.sh ESSearchServer
```

Windows

```
stopServer ESSearchServer  
startServer ESSearchServer
```

Эти сценарии находятся в каталоге `WAS_INSTALL_ROOT/AppServer/bin`:

- Для WebSphere Application Server Версии 5 путь установки по умолчанию - `/usr/WebSphere` в системах AIX, `/opt/WebSphere` в системах Linux и Solaris, `C:\Program Files\WebSphere` в системах Windows.
- Для WebSphere Application Server Версии 6 путь установки по умолчанию - `/usr/IBM/WebSphere` в системах AIX, `/opt/IBM/WebSphere` в системах Linux и Solaris, `C:\Program Files\IBM\WebSphere` в системах Windows.

Анализ частых результатов

Вы можете помочь пользователям уточнить набор результатов поиска, сконфигурировав опции для анализа частых результатов.

По существу, анализ частых результатов классифицирует текущий набор результатов поиска в соответствии с частотой встречаемости анализируемых значений. Пользователи могут фильтровать результаты поиска, выбрав одно из анализируемых значений, после чего выбранное значение будет автоматически добавлено в качестве нового ключевого слова в критерии поиска. При анализе частых результатов это значение позволяет пользователям точно настроить результаты поиска без использования полного или расширенного синтаксиса поиска.

В программе поисковой системы предприятия можно использовать столбчатые диаграммы, позволяющие увидеть, какие значения полей метаданных наиболее часто встречаются в результатах поиска. Кроме того, создав пользовательский класс Java, можно просмотреть частые результаты (в том числе и результаты из внешних для поисковой системы предприятия источников) в любом формате HTML.

Ограничения

Убедитесь, что на компьютере, где запускается WebSphere Application Server и программа поиска, установлены шрифты, соответствующие используемому вами языку. Этот шаг необходим, поскольку при генерации столбчатой диаграммы гарантирует выбор шрифта, обеспечивающий поддержку символов в метке диаграммы. Особенно это критично при использовании азиатских языков. Если вы устанавливаете шрифты после установки WebSphere Application Server, нужно перезапустить WebSphere Application Server, чтобы изменения вступили в силу.

Если программа поиска запускается как автономная программа, вы можете сконфигурировать диаграммы частых результатов, применив настройщик программы поиска или отредактировав файл конфигурации для программы поиска. Если программа поиска запускается вами как портлет в WebSphere Portal, нужно отредактировать свойства и сконфигурировать экземпляр этого портлета при помощи интерфейса администратора WebSphere Portal. Использовать настройщик программы поиска для конфигурирования опций анализа частых результатов нельзя.

Об этой задаче

Вы можете просмотреть частые результаты в графическом виде, если выведете чаще всего встречающиеся значения полей метаданных в столбчатой диаграмме. Кроме того, программу поиска можно расширить, применив класс Java, и вывести частые результаты в другом формате, например, вывести их при помощи HTML в неупорядоченном списке.

Столбчатые диаграммы для полей метаданных

Можно задать опции для анализа полей метаданных и вывода результатов анализа в столбчатых диаграммах. Эти диаграммы будут выводиться дополнительно к результатам запроса пользователя. Каждая диаграмма соответствует одному полю метаданных (такому как размер документа, автор, дата и т.п.), а каждый столбик диаграммы соответствует конкретному значению поля.

Длина столбика представляет число документов, содержащих конкретное значение поля, относительно других документов, содержащих другие значения в этом же поле. Чем длиннее столбик, тем чаще встречается это значение в поле. Пользователи могут точно настроить результаты, щелкнув по нужному столбику в столбчатой диаграмме. Значение поля, представленное выбранным столбиком, будет добавлено в качестве дополнительного термина запроса, и набор новых результатов поиска сузятся в соответствии с дополнительными критериями поиска.

При конфигурировании пространства для искателя можно выбрать опцию, задающую опции для полей метаданных. Например, можно задать допустимость поиска в поле метаданных как в свободном тексте, по имени поля (представленном в результатах поиска), как в параметрических данных и так далее. Чтобы в результатах поиска выводились диаграммы для полей метаданных, для искателя нужно сконфигурировать опции полей метаданных. Как минимум, нужно указать, что в поле допускается поиск и что его можно показать в результатах поиска.

Если при конфигурировании опций для поля метаданных выбрать опцию **Полное соответствие**, то при выборе пользователем в столбчатой диаграмме этого столбика будет запущен запрос полного соответствия. Поиск полных соответствий задает, что результаты должны быть возвращены, только если терму запроса соответствуют все значение поля. Если в поле содержится меньше или больше информации, соответствие не фиксируется.

Если при конфигурировании опций для поля метаданных не выбрать опцию **Полное соответствие**, то при выборе пользователем в столбчатой диаграмме этого столбика будет запущен запрос поиска по полям. В этом случае в поле должен встретиться дополнительный терм запроса, но ему не обязательно соответствовать все значение поля.

Результаты, отформатированные при помощи HTML

Можно расширить API

`com.ibm.es.searchui.charts.servlet.AbstractDynamicChart` при помощи пользовательского класса Java. Если для вывода результатов используется этот подход, с помощью HTML можно отформатировать вывод любых результатов поиска, а не только результатов запросов, выполняющих поиск в полях метаданных. Результаты могут быть получены для документов собраний поисковой системы предприятия либо возвращены из операций поиска во внешних репозиториях и сайтах.

Каждая диаграмма соответствует одному классу Java, а класс задает способ представления частых результатов на диаграмме. Например, `com.ibm.es.searchui.charts.servlet.DynamicMostRecentDocuments` (класс примера, поставляемый с кодом примера поисковой системы предприятия) представляет результаты в виде неупорядоченного списка. Документы отсортированы по датам, выводятся только заголовки и даты документов.

Другой класс примера, поставляемый с поисковой системой предприятия, `com.ibm.es.searchui.charts.servlet.DogearSearchResults`, показывает, как можно обеспечить пользователей списком закладок от Lotus Connections Dogear, относящихся к пользовательскому запросу.

Если пользователь щелкнет по результату поиска в отформатированном пользовательском выводе, документ будет выведен в новом окне браузера.

Процедура

Следующая процедура показывает, как с помощью настройщика программы поиска задать поля метаданных для анализа. Чаще всего встречающиеся значения полей метаданных выводятся в графическом виде на столбчатой диаграмме.

1. Откройте настройщик программы поиска. Если файл конфигурации для программы поиска, которую вы хотите настроить, еще не выводится, загрузите его и нажмите кнопку **Применить**.
2. Найдите **Диаграммы частых результатов** и выберите **Добавить диаграмму**.
3. Когда появится список опций диаграммы, оставьте переключатель **Пользовательская диаграмма** выключенным (включите его, только если для анализа и возврата частых результатов вами был создан пользовательский класс Java) и включите переключатель **Включить диаграмму**, чтобы эта столбчатая диаграмма выводилась при просмотре пользователями результатов поиска.
4. В поле **Заголовок диаграммы** задайте для диаграммы описательную метку. Заданное вами здесь значение заменит текст-заглушку **Строка для новой диаграммы**.
5. Решите, хотите ли вы включить переключатель **Разрешить динамический выбор полей**. Если включить этот переключатель пользователи при просмотре результатов поиска смогут выбрать другие поля и вывести для них частые результаты.

Возможно, вы захотите использовать эту опцию, чтобы позволить пользователям настраивать набор результатов поиска. Например, пользователь может выполнить поиск в базе данных сотрудников-женщин. После ввода запроса с поиском по полю пола (например, `sex:F`) значения для первых

результатов выводятся как столбики на этой столбчатой диаграмме. Далее пользователь выбирает из списка полей в начальном наборе результатов job. На столбчатой диаграмме будут показаны должности, которые чаще всего занимают женщины. После этого пользователь может выбрать конструктор из списка доступных полей. Теперь в результатах будет показана информация о сотрудниках-женщинах, занимающих должность конструктора.

6. В поле **Имя поля метаданных** введите имя поля метаданных, значения которого будут анализироваться для этой диаграммы.
7. При наличии в поле нескольких значений, заданных через разделительный символ, задайте этот символ в поле **Разделитель значений поля**. Например, если поле содержит два значения, например, имя и фамилию заказчика, разделенных точкой с запятой, здесь можно задать точку с запятой, чтобы каждое значение было добавлено в диаграмму по отдельности. Если не указать символ-разделитель, все значение поля будет анализироваться как единый элемент.
8. Задайте опции для вывода столбчатой диаграммы, такие как размер и цвет столбиков или число столбиков, которые будут выводиться в свернутой или развернутой диаграмме.
9. Задайте опции для сортировки результатов анализа. Например, сортировка может выполняться по значениям поля или по частоте встречаемости значений.
10. Если вы хотите, чтобы информация для определенных значений поля выводилась всегда (даже при отсутствии вхождений данного значения поля в результатах поиска), нажмите в области **Пользовательские метки** кнопку **Добавить строку**. Задайте значение поля, которое должно выводиться на столбчатой диаграмме даже при отсутствии возвращенных вхождений этого значения в результатах поиска, и задайте метку для соответствующего столбика диаграммы.
Например, если вы хотите, чтобы в результатах поиска всегда выводилась фирма-конкурент (например, JK Enterprises), даже если частота вхождений этого значения не позволяет квалифицировать его для вывода в частых результатах, задайте JK Enterprises как исходное значение для анализа, после чего введите описательную метку для идентификации соответствующего столбика в столбчатой диаграмме.
11. Нажмите кнопку **Далее** рядом с заголовком диаграммы, чтобы применить заданные для нее опции.
12. Если в области программы поиска уже есть результаты поиска, к текущему поиску будет применен анализ частых результатов и появится вновь созданная столбчатая диаграмма. В противном случае введите запрос, чтобы проверить ее вывод на экране.
13. Если созданная диаграмма вам подходит, нажмите кнопку **Сохранить**, чтобы изменения были внесены в файл конфигурации для программы поиска.

Понятия, связанные с данным



Задание свойств запроса



Классы Java для вывода лучших результатов

Обращение к программе поиска

Вы обращаетесь к программе поиска, задавая URL в браузере.

Прежде чем приступать к работе...

Программу поиска надо сконфигурировать для среды вашего Web-сервера.

Об этой задаче

Пример программы поиска установлен на поисковые сервера для поисковой системы предприятия. Можно использовать эту прикладную программу, как есть, для тестирования новых собраний и внешних источников перед тем, как сделать их доступными для пользователей. Можно также использовать эту программу в качестве модели для создания своих собственных программ поиска.

Процедура

Чтобы запустить программу поиска:

1. Введите URL программы поиска в браузере. Например:

`http://SearchServer.com/ESSearchApplication/`

`SearchServer.com` - имя хоста поискового сервера.

Кроме того, если для вашего Web-сервера не задано использование порта 80, надо задать допустимый номер порта. Например:

`http://SearchServer.com:9080/ESSearchApplication/`

2. Чтобы запустить пользовательскую программу поиска, введите URL для примера программы поиска и добавьте имя файла конфигурации для вашей прикладной программы поиска. Например:

`http://SearchServer.com/ESSearchApplication/search.do?configFile=/WEB-INF/myConfig.properties`

Если заданный вами файл не существует, выводится пример программы поиска для поисковой системы предприятия.

3. Если в WebSphere Application Server включена глобальная защита, зарегистрируйтесь в программе с допустимым ID пользователя и паролем.

Если в каких-либо собраниях, доступных для программы поиска, включена защита, и защищенные собрания содержат искатели, сконфигурированные для проверки параметров аутентификации пользователя при обработке запросов, можно сконфигурировать профиль пользователя. На странице Мой профиль задайте параметры аутентификации для доступа к защищенным доменам. После этого можно будет выполнять поиск в таких доменах, не регистрируясь в них.

Если искатель поддерживает защиту с единой регистрацией (SSO), вы можете просматривать защищенные домены, не создавая профиль пользователя.

4. Передайте запрос со страницы Поиск. Будет выполнен просмотр всех собраний и внешних источников, выбранных на странице Параметры.

Конфигурирование серверов поиска для приема только защищенных требований (по протоколу SSL)

Можно отключить интерфейс HTTP на серверах поиска и сконфигурировать эти серверы для приема требований поиска только через SSL и защищенный интерфейс HTTPS.

Об этой задаче

Чтобы сконфигурировать серверы поиска для использования только протокола Secure Sockets Layer (SSL) при обработке требований поиска, надо отключить протокол HTTP. Надо обеспечить также, чтобы один и тот же файл склада ключей хранился на обоих серверах поиска и на всех клиентских компьютерах, таких как сервер WebSphere Portal, где установлен портлет поиска для поисковой системы предприятия. Файл склада ключей, называемый также *truststore*, содержит открытые ключи, которые позволяют использовать SSL для доверенной связи.

Процедура

Чтобы сконфигурировать серверы поиска для приема только защищенных требований: only secure requests:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. Для установки с несколькими серверами выполните следующие действия на сервере индексов:
 - a. Остановите поисковую систему предприятия:
`esadmin system stopall`
 - b. Отредактируйте файл `ES_NODE_ROOT/master_config/nodes.ini`.
 - c. Измените значение `ID_узла.searchserverport` с порта HTTP (обычно 80) на порт HTTPS (обычно 443) и сохраните этот файл. Для установки на нескольких серверах измените оба значения `ID_узла.searchserverport` (по одному на каждом сервере поиска).
2. Выполните следующие действия на сервере поиска. Для установки с несколькими серверами выполните следующие действия на каждом сервере поиска:
 - a. Для установки с несколькими серверами зарегистрируйтесь на сервере поиска как администратор поисковой системы предприятия.
 - b. Отредактируйте файл `ES_NODE_ROOT/nodeinfo/es.cfg`.
 - c. Измените свойство `TrustStore`, указав в нем полный путь к файлу склада ключей SSL.
 - d. Измените свойство `HTTPProtocol`, задав в нем HTTPS, и сохраните этот файл.
 - e. Введите следующую команду, где *пароль_склада_ключей* - пароль склада ключей. Эта команда шифрует значения пароля и изменяет значение `TrustStorePassword` в файле `es.cfg`.

AIX, Linux или Solaris

```
eschangetrustpw.sh пароль_склада_ключей
```

Windows

```
eschangetrustpw пароль_склада_ключей
```

- f. Убедитесь, что значения свойств `trustStore` и `trustPassword` в файле `config.properties` для программы поиска задают правильный полный путь и пароль для файла склада ключей. Проверить или изменить эту информацию можно, открыв файл `config.properties` в редакторе, или при помощи настройщика программы поиска.

- g. Кроме того, при использовании портлета поиска поисковой системы предприятия убедитесь, что для файла склада ключей задан правильный полный путь и пароль в параметрах портлета trustStore и trustPassword. Для проверки и изменения этой информации используйте возможности управления портлетами из интерфейса управления WebSphere Portal.
3. Перезапустите поисковую систему предприятия:
- ```
esadmin system startall
```
- Задачи, связанные с данной**
- “Редактирование свойств примера программы поиска” на стр. 243
  - “Настройка программы поиска” на стр. 244
  - “Конфигурирование поисковой системы предприятия в WebSphere Portal версии 5.1” на стр. 349
  - “Конфигурирование поисковой системы предприятия в WebSphere Portal версии 6” на стр. 354

---

## Конфигурирование серверов поиска для приема требований через прокси-сервер

Можно сконфигурировать серверы поиска для приема требований через прокси-сервер

### Процедура

Чтобы можно было передавать требования на серверы поиска через прокси-сервер:

1. Зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему. Для установки с несколькими серверами зарегистрируйтесь на сервере индексов.  

```
esadmin system stopall
```
2. Отредактируйте файл `ES_NODE_ROOT/master_config/nodes.ini`.
  - a. Измените свойство `ProxyServer`, задав в нем полное имя хоста для прокси-сервера.
  - b. Измените свойство `ProxyServerPort`, задав в нем порт для прокси-сервера.
  - c. Необязательно: Если на прокси-сервере требуется аутентификация для всех требований, измените свойство `ProxyServerUserName`, задав в нем допустимое имя пользователя для прокси-сервера.
  - d. Необязательно: Если на прокси-сервере требуется аутентификация для всех требований, введите следующую команду, где *пароль\_прокси-сервера* - пароль для заданном вами имени пользователя прокси-сервера. Эта команда шифрует значение пароля и изменяет значение `ProxyServerUserPassword` в файле `es.cfg`.

#### AIX, Linux или Solaris

```
eschangeпрохурw.sh пароль_прокси-сервера
```

#### Windows

```
eschangeпрохурw пароль_прокси-сервера
```

3. Убедитесь, что в свойствах `прохуHost` и `прохуPort` в файле конфигурации (`.properties`) для программы поиска заданы правильное полное имя хоста и номер порта для прокси-сервера.

Если на прокси-сервере требуется аутентификация для всех требований, убедитесь, что свойства `прохуUser` и `прохуPassword` задают допустимое имя пользователя и

пароль для прокси-сервера. Эту информацию можно проверить или изменить, отредактировав файл конфигурации или воспользовавшись настройщиком программы поиска.

4. Кроме того, при использовании портлета поиска поисковой системы предприятия убедитесь, что в параметрах портлета `прохуHost` и `прохуPort` заданы правильное полное имя хоста и номер порта для прокси-сервера.

Если на прокси-сервере требуется аутентификация для всех требований, убедитесь, что свойства `прохуUser` и `прохуPassword` задают допустимое имя пользователя и пароль для прокси-сервера. Эту информацию можно проверить или изменить с помощью опций управления в интерфейсе администратора WebSphere Portal.

5. Перезапустите поисковую систему предприятия:  
`esadmin system startall`

---

## Поддержка внешних источников

*Внешний источник* - это источник данных, на котором программе поисковой системы предприятия разрешен поиск документов без обработки искателем, синтаксического анализа и индексации.

Типы источников данных, на которых возможен поиск как на внешних источниках:

- Базы данных, которые поддерживают протокол Java database connectivity (JDBC). Поддерживаются только базы данных IBM DB2, Oracle, Microsoft SQL Server 2000 и Microsoft SQL Server 2005. Для каждой таблицы, в которой вы разрешаете поиск, создается отдельный внешний источник данных.

**Ограничение:** Поддержка баз данных SQL Server 2005 ограничена таблицами, не содержащими данных типа Variant. Драйвер JDBC для SQL Server 2005 не поддерживается в системах AIX.

- Серверы LDAP (Lightweight Directory Access Protocol). Для каждого сервера LDAP создается один внешний источник.

Сконфигурировав информацию о внешнем источнике, свяжите его хотя бы с одной программой поиска. После этого поиск на внешнем источнике будет доступен пользователям наравне с запросами к собраниям, созданным при помощи искателя, синтаксического анализа и индексации данных.

### Понятия, связанные с данным



Федераторы API поиска и индексации

---

## Добавление внешних источников в систему

Добавляя внешний источник в поисковую систему предприятия, вы задаете тип этого источника. Мастер помогает задать информацию об источнике данных и допустимых типах поиска.

### Прежде чем приступить к работе...

Чтобы добавить внешний источник в систему, надо быть администратором поисковой системы предприятия.

### Ограничения

Для поиска в базе данных Oracle как во внешнем источнике необходимо, чтобы на поисковых серверах поисковой системы предприятия была установлена программа клиента Oracle.

Драйвер JDBC для Microsoft SQL Server 2005 не поддерживается в системах AIX.

### Об этой задаче

Добавляя в систему информацию о внешнем источнике, вы разрешаете пользователям запрашивать источник при помощи программ поисковой системы предприятия. Можно разрешить поиск на серверах LDAP (Lightweight Directory Access Protocol) и в таблицах баз данных JDBC (Java database connectivity).

При конфигурировании сервера LDAP запускается мастер, помогающий задать информацию, которая позволяет системе соединиться с сервером, а также опции поиска на сервере.

При конфигурировании базы данных JDBC запускается мастер, помогающий задать информацию, которая позволяет системе соединиться с базой данных, выбрать таблицы, на которых надо обеспечить возможность поиска, и задать опции поиска данных в этих таблицах. Для каждой таблицы, которую вы добавите в систему, создается отдельный внешний источник для поиска.

Информацию о драйверах SQL Server 2000 смотрите по адресу <http://www.microsoft.com/downloads/details.aspx?familyid=07287B11-0502-461A-B138-2AA54BFDC03A&displaylang=en>. Информацию о драйверах SQL Server 2005 смотрите по адресу <http://www.microsoft.com/downloads/details.aspx?familyid=e22bc83b-32ff-4474-a44a-22b6ae2c4e17&displaylang=en>.

## Процедура

Чтобы добавить внешний источник в систему:

1. Чтобы включить базы данных JDBC в поисковой системе предприятия, перед добавлением внешнего источника выполните следующие действия. Это действие требуется, чтобы система находила необходимые драйверы JDBC; его нужно выполнить только один раз.
  - a. На сервере искателя зарегистрируйтесь как администратор поисковой системы предприятия.
  - b. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/configurations/interfaces/discovery__interface.ini` так, чтобы переменные среды `CLASSPATH` и `LD_LIBPATH` включали путь класса к драйверам JDBC и путь к файлам библиотек.
  - c. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/configurations/interfaces/customcommunication__interface.ini` так, чтобы переменные среды `CLASSPATH` и `LD_LIBPATH` включали путь класса к драйверам JDBC и путь к файлам библиотек.
  - d. Необязательно: Чтобы использовать драйвер JDBC Oracle для локальных или внесенных в каталог баз данных, добавьте путь к библиотеке Oracle в переменную среды `LD_LIBPATH` (например, `LD_LIBPATH=.../home/oracle/OraHome1/lib32`) и отредактируйте файл `esrset.sh`, чтобы указать путь к библиотеке, и экспортируйте переменную каталога установки Oracle.  
Например:

```
ORACLE_HOME=/home/oracle/OraHome1
export ORACLE_HOME
```
  - e. Перезапустите поисковую систему предприятия, включая CCL (common communication layer):

### **AIX, Linux или Solaris**

```
esadmin stop
stopccl.sh
startccl.sh
esadmin start
```

### **Окно командной строки Windows**

```
esadmin stop
stopccl
startccl
esadmin start
```

## Администрирование - Службы Windows

Чтобы запустить SCL в фоновом режиме:

- 1) Введите `esadmin stop`.
  - 2) Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
  - 3) Щелкните правой кнопкой по **IBM OmniFind Enterprise Edition** и выберите **Остановить**. После остановки службы выберите **Пуск**.
  - 4) Введите `esadmin start`.
2. Выберите **Внешние источники** - откроется окно Внешние источники.
  3. Нажмите кнопку **Добавить внешний источник**.
  4. Выберите тип внешнего источника, который хотите добавить - сервер LDAP или базу данных JDBC.

5. Нажмите кнопку **Далее**, чтобы начать конфигурирование внешнего источника. Откроется мастер для создаваемого типа источника. Чтобы сконфигурировать внешний источник, следуйте подсказкам мастера. Чтобы узнать больше о доступных опциях, выберите **Помощь** на любой странице мастера.

Приведенные ниже имена и положения драйвера JDBC по умолчанию могут оказаться полезны при конфигурировании информации о соединении для баз данных DB2 и Oracle:

### DB2: Драйвер JDBC Legacy

Имя драйвера: `COM.ibm.db2.jdbc.app.DB2Driver`

Пример положения:

*корневой\_каталог\_установки\_db2/java/db2java.zip*

### DB2: Universal JDBC Driver

Имя драйвера: `com.ibm.db2.jcc.DB2Driver`

Примеры положения:

*корневой\_каталог\_установки\_db2/java/db2jcc.jar*

*корневой\_каталог\_установки\_db2/java/db2jcc\_license\_cu.jar*

**Oracle** Имя драйвера: `oracle.jdbc.driver.OracleDriver`

Пример положения: *домашний\_каталог\_oracle/jdbc/lib/ojdbc14.jar*

### Microsoft SQL Server 2000

Имя драйвера: `com.microsoft.jdbc.sqlserver.SQLServerDriver`

Примеры положения:

*домашний\_каталог\_jdbc\_mssql/lib/mssqlserver.jar*

*домашний\_каталог\_jdbc\_mssql/lib/msbase.jar*

*домашний\_каталог\_jdbc\_mssql/lib/msutil.jar*

### Microsoft SQL Server 2005

Имя драйвера: `com.microsoft.sqlserver.jdbc.SQLServerDriver`

Пример положения:

*каталог\_установки/sqljdbc\_1.0/локаль/sqljdbc.jar*

Например: *каталог\_установки/sqljdbc\_1.0/enu/sqljdbc.jar*

6. Задав опции поиска на внешнем источнике, нажмите **Готово**.

Новый внешний источник появится в окне Внешние источники вместе с другими внешними источниками, добавленными в систему.

## Понятия, связанные с данным



Федераторы API поиска и индексации

---

## Задание прикладных программ поиска для внешних источников

Поиск на внешнем источнике станет доступным только после того, как вы свяжете его хотя бы с одной программой поиска.

### Прежде чем приступать к работе...

Задать для прикладных программ поиска внешние источники, в которых они могут выполнять поиск, может администратор поисковой системы предприятия.

### Процедура

Чтобы связать прикладную программу поиска с одним или несколькими внешними источниками:

1. Выберите **Защита** на панели инструментов консоли администратора.
2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать прикладные программы поиска**.
3. На странице Конфигурировать прикладные программы поиска нажмите кнопку **Добавить прикладную программу поиска**.
4. Введите имя прикладной программы поиска.
5. Выберите внешние источники, в которых прикладная программа поиска сможет выполнять поиск:
  - Чтобы программа поиска обращалась ко всем внешним источникам, которые вы добавляете в систему, выберите **Все собрания и внешние источники**.
  - Чтобы разрешить программе поиска доступ только к указанным вами внешним источникам, выберите **Отдельные собрания и внешние источники**.  
Если выбрана эта опция, появляется список имен собраний и имен внешних источников. Включите переключатели **Выбрать** для всех внешних источников, в которых может выполнять поиск эта прикладная программа.
6. Нажмите кнопку **ОК**.

### Понятия, связанные с данным



Федераторы API поиска и индексации

---

## Защита поисковой системы предприятия

Механизмы защиты в поиске предприятия позволяют защитить источники от поиска неавторизованными пользователями и разрешить выполнение функций управления только определенным пользователям.

Поисковая система предприятия позволяет пользователям выполнять поиск в широком диапазоне источников данных. Чтобы к содержимому обращались только пользователи, авторизованные для доступа к нему, а к консоли администратора - только авторизованные для доступа к ней, поисковая система предприятия координирует и применяет защиту на нескольких уровнях.

### Совет:

Подробные примеры способов конфигурирования защиты для поисковой системы предприятия смотрите в документе IBM Redbook, IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios. В сценариях показано, как включить глобальную защиту в WebSphere Application Server с репозиторием LDAP, сконфигурировать компонент управления идентификацией и различные искатели, чтобы обеспечить защиту на уровне документов.

### Web-сервер

Первый уровень защиты - это Web-сервер. Если в WebSphere Application Server включена глобальная защита, можно назначить пользователям роли администраторов и выполнять аутентификацию пользователей, управляющих системой. Когда пользователь регистрируется на консоли администратора, этому пользователю доступны только те функции и собрания, для управления которыми он авторизован.

Кроме того, для аутентификации пользователей, выполняющих поиск в собраниях, прикладные программы поиска могут использовать механизмы защиты WebSphere Application Server.

### Защита уровня собрания

При создании собрания можно включить защиту на уровне собрания. Эту опцию нельзя изменить после создания собрания. Если защита на уровне собрания не включена, позже нельзя будет задать элементы управления защитой на уровне документов.

Когда включена защита уровня собрания:

- Процессы глобального анализа поисковой системы предприятия применяют другие правила для индексации повторяющихся документов.
- Можно сконфигурировать опции для применения защиты уровня документов, например, связать маркеры защиты с документами при просмотре документов искателем, потребовать проверки текущих параметров аутентификации при обработке запросов и задать, нужно ли индексировать текст привязки в документах Web.
- Можно применить защиту, задав для программы поиска (не для отдельных пользователей) собрания и внешние источники, в которых ей разрешен поиск. Затем можно при помощи стандартных механизмов управления доступом разрешать или запрещать пользователям доступ к прикладным программам поиска.

Платой за включение защиты собрания является снижение качества поиска. Включение защиты собрания уменьшает объем информации, индексируемой

для каждого документа. Побочное действие состоит в том, что для некоторых запросов будет найдено меньше результатов.

### **Защита на уровне документов**

При конфигурировании искателей для собрания можно включить защиту уровня документов. Например, можно задать опции для задания для данных маркеров защиты в процессе сбора данных искателем. Ваши программы поиска могут использовать эти маркеры (хранящиеся в индексе вместе с документами) для управления доступом, чтобы только пользователи с соответствующими параметрами аутентификации могли запрашивать данные и просматривать результаты поиска.

Для некоторых типов источников данных можно сконфигурировать опции выполняемой при обработке запросов проверки параметров аутентификации пользователя на соответствие с текущими элементами управления доступом. Этот дополнительный уровень защиты обеспечивает проверку привилегий пользователей в реальном времени на источнике данных. Эта возможность позволяет защититься от ситуаций, в которых параметры аутентификации пользователя изменились после внесения в индекс документа и его маркеров защиты.

Во время фазы глобального анализа, выполняющей обработку текста привязки, обычно связывает текст из одного документа (документа-источника) с другим документом (документом назначения), в котором не обязательно есть этот текст. При конфигурировании искателя Web можно задать, нужно ли исключить текст привязки из индекса, если этот текст ссылается на документ, которые искателю Web не разрешено просматривать.

Защита собраний - это не только механизмы аутентификации и управления доступом, при помощи которых поисковая система предприятия защищает индексированное содержимое. Кроме этого, есть способы, защищающие данные в процессе передачи от перехвата злонамеренным неавторизованным пользователем. Например, поисковые серверы используют для связи с сервером индексов и с программами поиска такие протоколы, как SSL (Secure Sockets Layer), SSH (Secure Shell) и Secure Hypertext Transfer Protocol (HTTPS).

Дополнительная защита обеспечивается шифрованием. Например, пароль администратора поисковой системы предприятия, заданный при установке продукта, хранится в зашифрованном виде. Пароли, которые пользователи задают в профилях пользователей, также хранятся в зашифрованном формате.

Для повышенной защиты нужно, чтобы аппаратное обеспечение сервера было должным образом изолировано и защищено от неавторизованного вторжения. Установив брандмауэр, можно защитить серверы поисковой системы предприятия от вторжения из другой части вашей сети. Кроме того, на серверах поисковой системы предприятия не должно быть открыто никаких портов. Сконфигурируйте систему так, чтобы она ожидала требований только на портах, в явном виде назначенных для работы поисковой системы предприятия и прикладных программ поиска.

---

## **Защита при установке**

Программа установки OmniFind Enterprise Edition создает среду для применения защиты в ситуациях, когда пользователи выполняют задачи администратора или выполняют поиск в собраниях поисковой системы предприятия.

## ID администратора поисковой системы предприятия

В процессе установки программа установки запрашивает ID пользователя и пароль, которые будут использоваться для администратора поисковой системы предприятия. Система использует эти параметры аутентификации для:

- Аутентификации администратора поисковой системы предприятия при выполнении задач администратора.
- Создания внутренней базы данных поисковой системы предприятия.
- Запуска всех сеансов или процессов поисковой системы предприятия.

Заданный при установке ID пользователя должен быть допустимым ID пользователя операционной системы с привилегиями системного администратора. Программа установки сохраняет эти параметры аутентификации (в зашифрованном виде) в файле свойств на каждом сервере поисковой системы предприятия.

## Шифрование

Для защиты конфиденциальных данных используется шифрование раздела данных аутентификации всех сообщений, передаваемых в поисковой системе предприятия. Этот процесс требует лишь небольших дополнительных затрат, так как шифруются только ID и пароли аутентификации. Все пароли, сохраняемые системой (в файлах конфигурации, в базе данных поисковой системы предприятия и других местах) также шифруются.

## Защита WebSphere Application Server

Если на сервере поиска ранее не был установлен WebSphere Application Server, программа установки автоматически устанавливает этот продукт с отключенной глобальной защитой. Если впоследствии в WebSphere Application Server будет включена глобальная защита, аутентификацию администратора поисковой системы предприятия будет выполнять WebSphere Application Server.

Включая глобальную защиту, нужно добавить заданные при установке ID и пароль администратора поисковой системы предприятия в реестр пользователей WebSphere Application Server, например, в каталог Lightweight Directory Access Protocol (LDAP).

Если глобальная защита включена после установки OmniFind Enterprise Edition, нужно изменить значения конфигурации и ввести команду **eschangewaspw**, чтобы зашифровать параметры аутентификации WebSphere Application Server и сохранить их в файле свойств поисковой системы предприятия.

---

## Сравнение аутентификации и управления доступом

Для защиты содержимого от неавторизованных пользователей и для управления доступом к функциям управления поисковая система предприятия поддерживает аутентификацию пользователей и авторизацию (управление доступом).

### Аутентификация

Аутентификация - это любой процесс, при котором система проверяет идентификацию пользователя, пытающегося обратиться к системе. Поскольку управление доступом обычно основано на идентификации пользователя, обращающегося к ресурсу, аутентификация необходима для эффективной защиты.

Для аутентификации пользователей поисковой системы предприятия используются параметры аутентификации, которые состоят, как минимум, из ID пользователя и пароля.

Для аутентификации пользователей, обращающихся к консоли администратора, поисковая система предприятия использует поддержку аутентификации, входящую в состав WebSphere Application Server.

## **Авторизация (управление доступом)**

Авторизация - это любой механизм, посредством которого система предоставляет и отзывает права доступа к некоторым данным или права выполнения некоторого действия. Обычно пользователь должен зарегистрироваться в системе, используя некоторый вид аутентификации. Механизмы управления доступом определяют, какие операции может или не может выполнять данный пользователь, сравнивая идентификатор пользователя с списком управления доступом (ACL). Для управления доступом используются:

- Разрешения для файлов (такие, как права создавать, читать, редактировать или удалять файл).
- Разрешения для программ (такие, как право запускать программу).
- Разрешения для данных (такие, как право получать или изменять информацию в базе данных).

---

## **Административные роли**

Поисковая система предприятия использует для управления доступом к разным функциям на консоли администратора понятие ролей.

При установке OmniFind Enterprise Edition (OmniFind Enterprise Edition) программа установки конфигурирует ID пользователя и пароль для администратора поисковой системы предприятия. При первом обращении к консоли администратора надо зарегистрироваться под именем этого пользователя. Если в WebSphere Application Server не включена глобальная защита, этот ID пользователя - единственный ID пользователя, который можно использовать для обращения к консоли администратора поисковой системы предприятия.

Если в WebSphere Application Server включена глобальная защита, можно назначить администраторами поисковой системы предприятия других пользователей. Назначая пользователям роли, можно ограничивать доступ к отдельным собраниям и управлять функциями, выполняемыми отдельными пользователями-администраторами. ID пользователей, назначаемых для ролей администраторов поисковой системы предприятия, должны существовать в реестре пользователей WebSphere Application Server.

При регистрации пользователя-администратора поисковая система предприятия выполняет аутентификацию ID пользователя. На консоли будут доступны только те собрания и функции, которыми разрешено управлять этому пользователю.

Пользователям можно назначать следующие административные роли:

### **Администратор поисковой системы предприятия**

Этот пользователь создает собрания и обладает полномочиями на управление всеми аспектами поисковой системы предприятия. При установке OmniFind Enterprise Edition вы задаете ID пользователя и пароль для первого пользователя - администратора поисковой системы предприятия. После

первой регистрации этот пользователь может назначить других пользователей для роли администратора поисковой системы предприятия.

#### **Администратор собрания**

Этот пользователь может изменять определенные собрания, отслеживать их и управлять их работой. Эти пользователи не могут создавать собрания. Администраторы собраний могут отслеживать и выполнять действия на уровне системы, только если администратор поисковой системы предприятия предоставил им такие полномочия.

#### **Оператор**

Этот пользователь может отслеживать определенные собрания и управлять их работой. Например, эти пользователи могут запускать и останавливать работу собраний, но не могут создавать или изменять собрания. Оператор может отслеживать и выполнять действия на уровне системы, только если администратор поисковой системы предприятия предоставил ему такие полномочия.

#### **Наблюдатель**

Этот пользователь может отслеживать определенные собрания. Эти пользователи не могут управлять работой собраний (например, запускать или останавливать серверы), создавать и изменять собрания. Наблюдатель может отслеживать (но не выполнять) действия на уровне системы, только если администратор поисковой системы предприятия предоставил ему такие полномочия.

## **Конфигурирование пользователей-администраторов**

Конфигурируя административные роли, можно ограничивать доступ к собраниям и управлять функциями, выполняемыми отдельными пользователями-администраторами.

#### **Прежде чем приступить к работе...**

Прежде чем назначать пользователя для административной роли, убедитесь, что в WebSphere Application Server включена защита. Убедитесь также, что этот ID пользователя внесен в реестр пользователей WebSphere Application Server.

Для конфигурирования пользователей-администраторов надо быть администратором поисковой системы предприятия.

#### **Процедура**

Чтобы назначить пользователям административные роли:

1. Нажмите кнопку **Защита** - откроется окно **Защита**.
2. На странице **Административные роли** нажмите кнопку **Добавить пользователя**.
3. Введите ID пользователя для административной роли и выберите соответствующую административную роль.
4. Если этот пользователь - не администратор поисковой системы предприятия, задайте, разрешен ли ему доступ к страницам с панели инструментов **Система**.  
Например, вам может понадобиться разрешить некоторым операторам или администраторам собраний отслеживать файлы журнала на уровне системы.
5. Если этот пользователь - не администратор поисковой системы предприятия, выберите собрания и внешние источники, которыми он может управлять.

Можно включить переключатели для отдельных собраний и внешних источников или позволить пользователю управлять всеми собраниями и внешними источниками.

---

## Защита уровня собрания

Для обеспечения защиты на уровне собрания надо сконфигурировать опции индексирования содержимого и опции, позволяющие прикладным программам поиска выполнять поиск в определенных собраниях.

При создании собрания можно выбрать опцию, включающую защиту собрания. Если эта опция выбрана, позже можно сконфигурировать элементы управления защитой на уровне документов. Если защита собрания включена, процессы глобального анализа поисковой системы предприятия будут также применять различные правила для индексации повторных документов.

После создания прикладной программы поиска ее ID позволяет задать собрания и внешние источники, в которых она может выполнять поиск, и пользователей, имеющих к ней доступ.

## Анализ повторных документов и защита собрания

Если включена защита собрания, процессы глобального анализа не определяют повторения документов в собрании.

При глобальном анализе процессы индексирования определяют документы, являющиеся повторениями (точными или приблизительными) друг друга. Затем они связывают все эти документы с одним каноническим представлением содержимого. Если позволить определять повторения документов, в результатах поиска не будет множества документов с одинаковым (или почти одинаковым) содержимым.

Если при создании собрания включить защиту собрания, повторения документов не определяются и, соответственно, не связываются с общим каноническим представлением. Вместо этого каждый документ индексируется отдельно. Это гарантирует оценку параметров защиты для каждого документа, что позволяет пользователям выполнять поиск только в документах с маркерами защиты, соответствующими данным регистрации пользователей. Содержимое двух документов может быть почти одинаковым, но для защиты они могут использовать разные списки управления доступом.

Предположим, что у нас есть два документа - document\_A и document\_B, и у пользователя есть права доступа только для документа document\_B. Если document\_B будет исключен как дублирующий, пользователь не увидит ничего в результатах поиска, так как для документа document\_A действуют ограничения доступа.

Отключение анализа повторений документов может улучшить защиту документов в собрании, но если пользователи будут получать в результатах поиска несколько копий одного документа, качество поиска может ухудшиться.

## Идентификаторы программ поиска

Возможность поиска в различных собраниях определяется отображением программы поиска на собрания и внешние источники, в которых ей разрешен поиск. Если для поставляемого примера программы поиска используется имя программы Default, эту программу поиска можно использовать для поиска во всех собраниях и внешних источниках.

Каждая прикладная программа поиска должны передавать интерфейсу прикладного программирования (API) поисковой системы предприятия правильный идентификатор прикладной программы (APPID). Программа поиска может выполнять поиск только в собраниях и внешних источниках, связанных с этим APPID.

Чтобы программа поиска могла обращаться к собранию или внешнему источнику, администратор поисковой системы предприятия должен связать эту программу поиска с конкретными собраниями и источниками, в которых она сможет выполнять поиск. Программа поиска может выполнять поиск во всех собраниях и внешних источниках в поисковой системе предприятия или же только в заданных вами собраниях и внешних источниках.

У примера программы поиска (ESSearchApplication) есть файл свойств, в котором задается имя этой программы. Положение по умолчанию этого файла свойств: `KORНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES\installedApps\ESSearchApplication.ear\ESSearchApplication.war\WEB-INF\config.properties`.

Начальное значение имени этой программы - Default. Если изменить это значение, изменится список собраний и внешних источников, в которых может выполнять поиск ESSearchApplication.

Чтобы задать, какие пользователи могут выполнять поиск в данных собраниях, нужно связать пользователей (или группы пользователей) с клиентскими программами при помощи стандартных средств управления доступом WebSphere Application Server, аналогично тому, как с помощью этих средств можно ограничить доступ к некоторому URL. Например, можно разрешить доступ только к URL запуска прикладной программы поиска.

Дополнительную информацию об ID прикладных программ поиска и о том, как добавить управление защитой в пользовательские программы поиска, смотрите в главе API поиска и индексирования для поисковой системы предприятия.

#### Понятия, связанные с данным



Обзор API поиска и индексирования



Защита API поиска и индексирования

---

## Защита на уровне документов

Если при создании собрания для него включена защита, можно сконфигурировать элементы управления защитой уровня документов. Элементы управления защитой уровня документов обеспечивают, что в результатах поиска выводятся только документы, которые разрешено видеть пользователю, передавшему требование поиска.

Поисковая система предприятия поддерживает различные способы конфигурирования элементов управления защитой уровня документов:

- Документы можно подвергать предварительной фильтрации и связывать с маркерами защиты перед добавлением в индекс.
- Для некоторых типов данных для результатов поиска может выполняться последующая фильтрация, при которой параметры аутентификации пользователя сверяются с текущими данными управления доступом. Компонент управления идентификацией поисковой системы предприятия может шифровать различные параметры аутентификации, которые требуются пользователям для доступа к разным репозиториям, и сохранять зашифрованные параметры аутентификации в

профилях. Если источники для поиска защищены продуктом, который использует защиту с единой регистрацией (SSO), компонент управления идентификацией может управлять доступом к документам, не требуя от пользователей создания профилей.

- Для большинства типов искателей для связи маркеров защиты с документами в индексе можно использовать пользовательский класс Java (подключаемый модуль).
- Для документов, обрабатываемых искателем Web, можно исключить из индекса текст привязки из этих документов, содержащий ссылки на документы, доступ к которым запрещен.

#### Понятия, связанные с данным

 Защита программ

## Предварительная фильтрация и постфильтрация результатов поиска

Есть два различных способа выполнять фильтрацию документов, чтобы в результатах поиска выводились только документы, которые разрешено просматривать пользователю, передавшему требование поиска.

- Первый способ - во время просмотра искателем скопировать исходные списки управления доступом (ACL) для документов в индекс и позволить механизму поиска сравнивать параметры аутентификации пользователей с информацией в скопированных ACL документов. Предварительная фильтрация документов и управление добавлением документов в индекс повышают производительность. Однако сложно смоделировать в индексе все политики защиты различных источников данных и реализовать однотипные алгоритмы сравнения. Кроме того, при этом способе не учитываются изменения, которые могут быть внесены в исходные списки управления доступом.
- Второй способ - постфильтрация документов в наборе результатов на основе текущих данных защиты, полученных от источников данных. При таком способе именно источники данных определяют, будет ли документ возвращен пользователю, гарантируя, что набор результатов соответствует текущим элементам управления доступом. Однако при этом способе снижается производительность поиска, поскольку требуются соединения со всеми источниками данных. Если какой-либо источник недоступен, ссылки на его документы должны быть отфильтрованы из набора результатов вместе с документами, на просмотр которых у пользователя нет прав.

**Важное замечание:** В конфигурации с несколькими серверами постфильтрация для некоторых типов источников выполняется на сервере искателя. Если сервер искателя не активен (например, остановлен для обслуживания), пользователи не получают никаких результатов при запросах к собраниям поисковой системы предприятия. Кроме того, никакие результаты не будут возвращены, если недоступны серверы источников данных, необходимые для управления доступом.

В поисковой системе предприятия для применения элементов управления доступом используется сочетание этих двух способов. Такое решение дает оптимальную производительность и обеспечивает точное применение политик защиты исходных репозиториях документов. Сохраняя высокоуровневые данные управления доступом в индексе, система может создать промежуточный набор результатов (потенциально меньшего размера), к которому можно затем применить постфильтрацию для проверки текущих элементов управления доступом. При этом используется такое предположение: если у пользователя есть доступ к репозиторию, в котором содержится документ, вероятно, у него есть доступ и к документу.

Сохраняемые в индексе данные управления доступом зависят от типа искателя. Например, искатель Примечания может сохранять элементы управления доступом на уровне базы данных и сервера, а искатель QuickPlace - элементы управления доступом для серверов, площадок и комнат.

Все типы источников данных поисковой системы предприятия поддерживают способность по время просмотра искателем вносить в индекс исходные списки управления доступом. Кроме того, некоторые типы источников данных поддерживают способность выполнять постфильтрацию набора результатов и проверять текущие параметры аутентификации пользователя (для этой поддержки используются исходные механизмы защиты или компонент управления идентификацией поисковой системы предприятия).

Эта двойная система защиты включает следующие задачи:

- Извлечение информации исходных списков управления доступом во время просмотра искателем.
- Сохранение в индексе информации списков управления доступом для серверов и баз данных.
- Создание для пользователя контекста защиты при регистрации пользователя в системе или при инициализации сеанса. В этой задаче должны учитываться различные идентификаторы, которые один пользователь может применять для доступа к различным источникам данных.
- Выполнение поиска с использованием контекста защиты данного пользователя и создание промежуточного набора результатов, содержащего только документы, к которым у пользователя есть доступ на уровне репозитория.
- Постфильтрация промежуточного набора результатов на основе информации о текущих списках управления доступом, полученной с источников данных для документов в наборе результатов.

## Проверка на основе хранимых маркеров защиты

Если при создании собрания для него включена защита, можно сконфигурировать элементы управления защитой уровня документов, сохранив данные защиты в индексе.

По умолчанию каждый документ связан с маркером общедоступности, делающим документ доступным каждому. Если для собрания включена защита, этот маркер можно заменить на значение, заданное администратором или извлеченное из некоторого поля в просмотренном искателем документе. При конфигурировании искателя вы задаете, что хотите при помощи маркеров защиты указать, каким пользователям разрешен доступ к документам, найденным этим искателем.

Когда администратор собрания конфигурирует искатель, он может задать опции защиты для отдельных таблиц, файловых систем и т.п. Это значит, что для разных источников данных в пространстве для искателя можно задать разные правила защиты. Администратор может:

- Задать, что эти документы - общедоступные (все пользователи могут выполнять поиск этих документов)
- Назначить каждому документу пользовательские маркеры защиты
- Извлечь маркеры защиты из определенного поля в просмотренных искателем данных и назначить извлеченный маркер для каждого документа

Маркеры защиты (кроме используемого по умолчанию маркера общедоступности) полностью определяются пользователем. Маркер защиты может представлять ID

пользователя, ID группы, роль пользователя или любое другое значение, определенное вами как действительное для источника данных.

Например, администратор может задать, что для управления доступом к документам, обработанным искателем Notes, нужно использовать поле `hrDeptName`. Кроме того, администратор может задать, что если в документе нет этого поля или если это поле не содержит данных защиты, для доступа к документам нужно использовать два определенных пользователем маркера `hrgroup1` и `hrgroup2`.

Маркеры защиты передаются искателю через его файл конфигурации. Для каждого документа искатель задает значение маркера защиты в метаданных этого документа. Компонент индексации читает маркер защиты и применяет его, когда помещает в индекс информацию для документа. Если администратор источника данных изменяет список управления доступом, измененные элементы управления доступом становятся доступными после следующей операции полного построения или дельта-построения индекса.

Можно применять пользовательские бизнес-правила для определения значения маркеров защиты путем шифрования правил в классе Java. При конфигурировании свойств искателя вы задаете имя подключаемого модуля, который будет использовать искатель при просмотре документов. Маркеры защиты, добавляемые этим подключаемым модулем, сохраняются в индексе и могут использоваться для управления доступом к документам.

## Как программы поиска используют маркеры защиты

Клиентская программа поиска должна во время поиска предоставить маркеры защиты, чтобы можно было правильным образом отфильтровать документы. Если маркеры защиты не предоставлены, при выполнении поиска автоматически применяется используемый по умолчанию маркер общедоступности.

Пример программы поиска для поисковой системы предприятия демонстрирует, как можно реализовать защиту на уровне документов. В этом примере предполагается, что администратор назначил значение маркера защиты группе документов (вместо того, чтобы извлекать маркер защиты из поля в просматриваемом документе). Программа поиска использует ID регистрации пользователя, чтобы определить, к каким документам может обращаться пользователь. Вместо самого ID пользователя эта программа поиска использует ID группы, в которую входит пользователь. Когда в качестве маркера защиты используется ID группы, можно добавлять пользователей в эту группу и удалять пользователей из нее; перестроение индекса для этого не потребуется.

Маркер защиты, назначенный администратором для набора документов, представляет допустимый ID группы операционной системы. Другим документам в пространстве для искателя назначаются другие ID групп. Например:

Документы 1-5: Маркер защиты = Группа1  
Документы 6-10: Маркер защиты = Группа2

## Проверка текущих параметров аутентификации при обработке запроса

Если при создании собрания для него включена защита, в некоторых типах доменов можно выполнять проверку параметров аутентификации текущего пользователя при передаче запроса этим пользователем.

Перед ответом на запрос серверы поиска обращаются к исходным репозиториям для проверки разрешений текущего пользователя, а затем удаляют из результатов поиска все документы, для которых у пользователя нет разрешений на просмотр.

При конфигурировании перечисленных ниже типов искателей можно выбрать опцию проверки параметров регистрации пользователя путем сравнения их с элементами управления доступом, управляемыми исходным репозиторием. Когда документы просмотрены искателем и внесены в индекс, для проверки пользователей, пытающихся выполнять поиск в защищенных собраниях, используется компонент управления идентификацией поисковой системы предприятия.

- Искатель Content Edition (только репозитории типов Documentum, FileNet Panagon Content Services, Hummingbird DM, Portal Document Manager и SharePoint)
- Искатель DB2 Content Manager
- Искатель Domino Document Manager
- Искатель Примечания
- Искатель QuickPlace
- Искатель Файловая система Windows

Для следующих типов искателей проверка параметров аутентификации пользователя может быть выполнена при использовании портлета поиска в WebSphere Portal для поиска в собраниях поисковой системы предприятия.

- Искатель Web Content Management
- Искатель WebSphere Portal

#### **Понятия, связанные с данным**

“Обеспечение защиты уровня документов для документов Lotus Domino” на стр. 286

“Обеспечение защиты уровня документов для документов файловых систем Windows” на стр. 289

#### **Задачи, связанные с данной**

“Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей” на стр. 286

## **Управление идентификацией в поисковой системе предприятия**

Управление параметрами аутентификации множества пользователей - это общая проблема для систем уровня предприятия. Для решения этой проблемы в поисковую систему предприятия включен дополнительный компонент управления идентификацией.

### **Совет:**

Подробные примеры способов конфигурирования компонента управления идентификацией в поисковой системе предприятия смотрите в документе IBM Redbook, IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

Информация на предприятии может существовать в различных форматах. Она может находиться в разных местах, и для работы с ней могут использоваться различные программы. Например, пользователи предприятия могут использовать прикладную программу SQL для доступа к реляционным базам данных или систему управления документами для доступа к документам, связанным с их работой.

Обычно для управления доступом к критической информации в этих репозиториях используется управляющее программное обеспечение. Идентификация пользователей

выполняется на хост-системе при помощи ID пользователя и пароля. После того, как система аутентифицирует пользователя, управляющее программное обеспечение определяет на основе заданных прав доступа пользователя, какие документы разрешено просматривать этому пользователю и с какими документами он может работать.

Обычно у пользователя для каждого репозитория свои ID пользователя и пароли. Когда пользователи обращаются напрямую к репозиториям предприятия, они должны пройти аутентификацию; аналогичным образом, пользователи должны предоставить параметры аутентификации перед просмотром документов в собрании поисковой системы предприятия, для которого требуется проверка текущих параметров аутентификации. Пользователи, у которых несколько идентификаторов, должны предоставить соответствующие параметры аутентификации для каждого идентификатора.

Если на консоли администратора задано, что для управления идентификацией нужно использовать поисковую систему предприятия, серверы поиска могут использовать следующие способы проверки параметров аутентификации пользователей при обработке запросов:

- Программы поиска могут выводить приглашение пользователю зарегистрировать параметры аутентификации, необходимые для доступа к различным доменам в профиле пользователя. Профиль пользователя хранится в зашифрованном виде на защищенном складе данных и позволяет пользователю выполнять поиск в защищенных доменах. Если для домена, для которого требуется проверка параметров аутентификации, не заданы параметры аутентификации, документы этого домена исключаются из результатов поиска.
- Если для просмотра документов в собрании используется искатель, поддерживающий защиту с единой регистрацией (SSO), и вы задали, что нужно использовать защиту SSO для управления доступом к документам, система может применять методы защиты SSO для аутентификации пользователей во время сеанса поиска. Пользователю не придется создавать профиль, содержащий параметры аутентификации, или вводить ID пользователя и пароль при поиске в защищенных доменах.

Когда пользователи выполняют поиск в собраниях, для которых требуется проверка текущих параметров аутентификации при передаче запроса, система может использовать профиль или методы защиты SSO, чтобы запретить или разрешить доступ к документам.

## Получение информации о группах пользователя

Для проверки параметров аутентификации пользователя компонент управления идентификацией должен получить информацию о группах пользователя для каждого идентификатора пользователя и добавить эту информацию в строку контекста защиты пользователя (USC). Эта информация о группах используется для фильтрации результатов в соответствии с данными управления доступом, хранимыми в индексе поисковой системы предприятия, или данными аутентификации SSO. Компонент управления идентификацией использует маркеры SSO или параметры аутентификации пользователя для соединения с системой источника данных и запрашивает информацию о группах, в которые входит этот пользователь.

При конфигурировании опции управления идентификацией на панели администратора можно задать, как часто нужно обновлять эту информацию о группах. Можно получать новую информацию о группах каждый раз, когда пользователь регистрируется в программе поиска, или же через определенные

промежутки времени, например, раз в три дня.

## Защита без компонента управления идентификацией

Не во всех предприятиях хотят управлять идентификаторами пользователей с помощью компонента управления идентификацией поисковой системы предприятия. Если на консоли администратора поисковой системы предприятия отключить компонент управления идентификацией, ваша программа поиска должна будет сама генерировать строку контекста защиты пользователя (USC). Сгенерированная строка USC используется для задания значения ограничений списка управления доступом для каждого запроса. Например:

```
Query q = factory.createQuery("IBM");
q.setACLConstraints("User's Security Context in XML");
```

**Совет:** Чтобы помочь вам писать собственные функции управления идентификацией, в расширение для API поиска и индексирования (SI-API) включены функции для работы с базой данных управления идентификацией. Эти API позволяют генерировать строку USC с помощью объектов Java - эта строка в формате XML строится автоматически.

Строка запроса XML должна быть в показанном ниже формате, где ... содержит полностью сформированную строку XML:

```
@SecurityContext::'...'
```

Формат этой строки XML:

```
<identities id="имя_пользователя_для_регистрации">
 <ssoToken>значение_маркера</ssoToken>
 <identity id="домен_защиты">
 <type>Notes</type>
 <username>имя_пользователя_домена</username>
 <password encrypt="no">пароль_пользователя_домена</password>
 <groups>
 <group id="g1" />
 <group id="g2" />
 </groups>
 <properties>
 <property name="имя_свойства">значение_свойства</property>
 ...
 </properties>
 </identity>
 ...
</identities>
```

### **identities**

Значение атрибута id - ID пользователя, который вводится пользователем при регистрации в системе.

### **ssoToken**

Необязательно: Задайте маркер LTPA (Lightweight Third-Party Authentication), созданный для этого пользователя на время сеанса браузера. Этот параметр используется, только если в домене назначения разрешена SSO и искатель сконфигурирован для использования защиты SSO.

### **identity**

Содержит параметры аутентификации пользователя для конкретного источника данных. Значение атрибута id - домен, в котором хранятся параметры аутентификации пользователя (в случае Domino это имя домена Domino).

**type**

Указывает тип данных и соответствующий тип искателя (Notes, DB2, Exchange Server и т.д.).

**username**

Задает имя пользователя для поиска в домене.

**password**

Задает пароль для заданного имени пользователя. Для атрибута encrypt нужно задать значение no (поисковая система предприятия не поддерживает методы шифрования вне компонента управления идентификацией).

**groups**

Задает имена групп, в которые входит пользователь. Для каждого имени группы используется отдельный элемент group.

**properties**

Задает список свойств для данного соединения, например, ID администратора и зашифрованный пароль, использованные для создания искателя, или свойство, указывающее, включена ли для источника SSO.

**имя\_свойства**

Имя свойства

**значение\_свойства**

Значение свойства.

## Проверка пользователей с профилями пользователей

Прикладные программы поиска могут предложить пользователям сохранить параметры аутентификации, которые требуются для доступа к различным доменам, в профиле пользователя.

При поиске в домене, требующем проверки параметров аутентификации пользователей при передаче запроса, пользователи должны передать программе поиска свои данные регистрации для входа в этот домен. При помощи управления идентификацией поисковой системы предприятия пользователи могут сохранять в своих профилях параметры аутентификации для любого числа доменов. Эти параметры шифруются для защиты и в таком виде хранятся в поисковой системе предприятия.

Если для домена, для которого требуется проверка параметров аутентификации, не заданы параметры аутентификации, документы этого домена исключаются из результатов поиска.

Пользователи могут создать профили пользователя и зарегистрировать свои параметры аутентификации во время работы с программой поиска. В примере программы поиска для поисковой системы предприятия эта возможность обеспечивается опцией **Мой профиль**. В пользовательских программах поиска эта возможность может быть реализована иначе.

Собрания могут содержать документы из различных типов источников. Например, собрание может содержать документы, найденные в файловой системе Windows и нескольких базах данных Lotus Notes. Компонент управления идентификацией различает разные типы источников и запрашивает только параметры аутентификации для доступа к доменам, для которых требуется проверка.

По умолчанию каждому пользователю разрешено выполнять поиск (при этом пользователь должен предоставить ID пользователя и пароль, соответствующие защищенным доменам). Если пользователь забыл ID пользователя или пароль для

определенного домена, можно отключить поиск в этом домене, выключив соответствующий переключатель. Если поиск в домене отключен, защищенные документы из этого домена не будут возвращены в наборе результатов.

После создания профиля пользователь может передать требование поиска. У компонента управления идентификацией есть информация, необходимая для построения строки контекста защиты пользователя (USC), которая будет использоваться в последующих требованиях поиска. Если вы не используете компонент управления идентификацией, программа поиска должна передать строку USC при выполнении пользователем поиска в доменах, требующих проверки текущих параметров аутентификации.

При последующей попытке пользователя выполнять поиск в собраниях поисковой системы предприятия компонент управления идентификацией повторяет процесс проверки параметров аутентификации, но теперь компонент управления идентификацией сможет найти профиль пользователя. Если ничего не изменилось, для пользователя автоматически будет получена информация о том, куда можно передавать требования поиска, и пользователю не будет выводиться запрос создания профиля.

Если компонент управления идентификацией обнаружит, что какой-либо из параметров аутентификации пользователя изменен, пользователю автоматически выводится страница профиля при обращении к программе поиска. Например, это происходит, если изменен пароль для какого-либо из доменов, для которых разрешен поиск, или если в собрание добавлен домен, для которого требуется аутентификация.

Пользователь может проигнорировать рекомендацию обновить профиль, но в этом случае соответствующие документы будут исключены из результатов поиска.

В поставляемом с поисковой системой предприятия примере программы поиска пользователи могут в любой момент обновить профили, выбрав **Мой профиль** на панели инструментов.

## **Проверка пользователей с использованием защиты SSO**

Если для просмотра документов в собрании используется искатель, поддерживающий защиту с единой регистрацией (SSO), при задании опций управления идентификацией можно задать, что для управления доступом к документам нужно использовать защиту SSO.

## **Включение SSO**

Аутентификация с единой регистрацией позволяет пользователю один раз пройти аутентификацию и получить доступ ко многим ресурсам без необходимости еще раз вводить параметры аутентификации. В поисковой системе предприятия аутентификация SSO облегчает управление большим числом имен и паролей пользователей, которые пользователи должны задавать для доступа к документам в защищенным собраниям.

IBM WebSphere Application Server и Lotus Domino поддерживают вариант SSO, который называется Lightweight Third-Party Authentication (LTPA). Когда пользователь пытается обратиться к одному из этих продуктов, ему предлагается пройти аутентификацию, введя имя пользователя и пароль. Эти имя пользователя и пароль проверяются по репозиторию LDAP, который совместно используют оба эти продукта. После аутентификации пользователя создается опознавательный код (cookie) сеанса, содержащий маркер LTPA. После этого пользователь может обращаться к другим ресурсам на любом сервере с той же конфигурацией

аутентификации, и ему не будет выводиться запрос повторного ввода параметров аутентификации. Этот маркер сохраняется на время действия сеанса браузера.

Чтобы включить поддержку SSO для собраний поисковой системы предприятия:

- Убедитесь, что на серверах поиска поисковой системы предприятия включены глобальная защита WebSphere Application Server и реестр LDAP. В качестве реестра LDAP можно использовать любой допустимый продукт LDAP, который поддерживается продуктом WebSphere Application Server.
- Убедитесь, что механизм аутентификации WebSphere сконфигурирован для использования активного механизма аутентификации LTPA. При конфигурировании LTPA задайте такое допустимое имя домена, которое допускает настройку, например, your.server.com.
- Убедитесь, что ключ LTPA экспортирован из WebSphere Application Server и импортирован в другие продукты в том же домене, для которых нужно включить поддержку LTPA.

С помощью браузера убедитесь, что описанная выше конфигурация защиты работает правильно; после этого можно с помощью консоли администратора поисковой системы предприятия сконфигурировать искатели для поддержки аутентификации SSO.

## SSO и управление идентификацией

Когда пользователи выполняют поиск в собраниях, для которых требуется проверка текущих параметров аутентификации, система может использовать методы защиты SSO, чтобы запретить или разрешить доступ к документам. При поиске в источниках, поддерживающих аутентификацию SSO, пользователям не выводится запрос параметров аутентификации. Компонент управления идентификацией используется, если выполнены все приведенные ниже условия:

- SSO правильно включена в WebSphere Application Server и доменах назначения.
- Включена защита как минимум в одном из собраний, в которых программа поиска может выполнять поиск.
- На консоли администратора поисковой системы предприятия включены опции для использования компонента управления идентификацией и защиты SSO.
- При конфигурировании следующих типов искателей выбрана опция для использования защиты SSO и опции для применения защиты уровня документов (например, опции для включения параметров управления доступом в индекс или опции для проверки текущих параметров аутентификации при обработке запросов):
  - Content Edition (доступно только для репозитория Portal Document Manager)
  - Domino Document Manager (доступно только для искателей, использующих протокол DIIOP)
  - Примечания (доступно только для искателей, использующих протокол DIIOP)
  - QuickPlace (доступно только для искателей, использующих протокол DIIOP)

## Защита портлета поиска

Если пользователи используют портлет поисковой системы предприятия для поиска в собраниях из WebSphere Portal, защита поддерживается также для документов, обработанных искателями Список заполнения, Web Content Management и WebSphere Portal.

Для этих типов источников защита поиска поддерживается, только если вы используете портлет, а не запускаете программу поиска извне WebSphere Portal. В

WebSphere Portal параметры аутентификации пользователя передаются портлетом поиска. После того, как пользователь зарегистрировался на WebSphere Portal, во все требования поиска включаются данные защиты этого пользователя (имя пользователя, членство в группах и т.п.). Поскольку эта информация всегда доступна, SSO на основе маркера LTPA не требуется.

Но если вы используете другой продукт для защиты сайтов и документов на сервере WebSphere Portal, при конфигурировании искателя надо задать опции SSO. Например, если вы используете IBM Tivoli Access Manager WebSEAL или CA SiteMinder SSO Agent for PeopleSoft, надо задать параметры аутентификации, которые позволят искателю обращаться к документам на сервере с использованием единой регистрации. В таком случае вы включаете SSO, чтобы искатель мог обращаться к защищенному содержимому, а не для защищенного поиска.

## Конфигурирование управления идентификацией

Компонент управления идентификацией для поисковой системы предприятия позволяет задать, как проверять текущие параметров аутентификации при обработке запроса.

### Прежде чем приступить к работе...

Для конфигурирования опций управления идентификацией надо быть администратором поисковой системы предприятия.

### Об этой задаче

Когда пользователи выполняют поиск в собраниях, для которых требуется проверка текущих параметров аутентификации при обработке запроса, система компонент управления идентификацией для поисковой системы предприятия может использовать профиль пользователя или методы защиты с единой регистрацией (SSO), чтобы запретить или разрешить доступ к документам.

### Процедура

Чтобы сконфигурировать управление идентификацией:

1. Нажмите кнопку **Защита** - откроется окно Защита.
2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать управление идентификацией**.
3. На странице Сконфигурировать управление идентификацией включите переключатель для использования компонента управления идентификацией поисковой системы предприятия для управления проверкой параметров аутентификации пользователей при обработке запросов. Если этот переключатель не включен, прикладная программа поиска может предоставить строку контекста защиты пользователя (user's security context - USC), когда пользователи выполняют запрос к доменам, в которых во время обработки запроса требуется проверка текущих учетных данных.
4. Задайте, с какой частотой компонент управления идентификацией должен извлекать параметры аутентификации пользователя из записей группы в реестре пользователей WebSphere Application Server. Данные параметров аутентификации можно обновлять при каждом обращении пользователя к программе поиска или через определенный интервал времени.
5. Для типов искателей, поддерживающих аутентификацию SSO, задайте, должен ли компонент управления идентификацией использовать для проверки пользователей

маркеры защиты SSO вместо профилей пользователей. Можно включить переключатель для использования аутентификации SSO для всех типов искателей или же для отдельных типов искателей.

**Важное замечание:** Компонент управления идентификацией использует механизмы защиты SSO, только если защита SSO правильно сконфигурирована в WebSphere Application Server и доменах назначения.

## Анализ привязок текста

Если включена защита собрания, процессы глобального анализа применяют специальные правила индексации текста привязки в документах, просматриваемых искателями Web. Если не включить защиту для собрания, можно при конфигурировании отдельных искателей Web выбрать, индексировать ли текст привязки в ссылках на запрещенные документы.

Текст привязки - это информация в гипертекстовой ссылке, описывающая страницу назначения ссылки. Например, в следующей ссылке текст Синтаксис запросов - это текст привязки в ссылке на страницу `syntax.htm`:

```
Синтаксис запросов
```

Обычно искатель Web переходит по ссылкам в документах для просмотра дополнительных документов и включает в индекс страницы назначения ссылок. Во время глобального анализа процессы индекса связывают текст привязки не только с документом, в котором он находится (документом-источником), но и с документом назначения. В приведенном примере текст привязки Синтаксис запросов связывается со страницей назначения `syntax.htm` и с исходной страницей, содержащей конструкцию привязки. Такая связь позволяет получать документ назначения по запросам, содержащим текст из документа-источника. Эта связь, однако, представляет некоторый риск для защиты, если пользователям разрешен просмотр документа назначения, но запрещен просмотр документа-источника.

Если при создании собрания включить защиту собрания, обработка текста привязки выключается. Текст привязки не будет проиндексирован с документом, если только он не встречается в самом документе или в метаданных документа. Такое управление защитой гарантирует, что пользователям не будет видна информация в документах, к которым у них нет доступа; документ возвращается в результатах поиска, только если запросу соответствуют его собственное содержимое или метаданные.

Включение защиты собрания может улучшить защиту документов Web, позволяя пользователям выполнять поиск только в документах с маркерами защиты, соответствующими данным регистрации пользователей. Однако при отключенной обработке текста привязки среди результатов поиска могут оказаться не все документы, потенциально релевантные для запроса.

Если не включить защиту для собрания, можно при конфигурировании дополнительных свойств искателей Web выбрать, индексировать ли текст привязки в ссылках на запрещенные документы.

## Индексация текста привязки в ссылках на запрещенные документы

Если некоторые документы содержат ссылки на документы, которые искателю Web запрещено просматривать, при конфигурировании искателя Web можно указать, нужно ли сохранять в индексе текст привязки для таких ссылок.

**Прежде чем приступить к работе...**

Для конфигурирования опций индексации текста привязки нужно быть администратором поисковой системы предприятия или администратором собрания для искателя Web, который вы хотите сконфигурировать.

### Об этой задаче

Директивы в файле robots.txt или метаданные документов Web могут запрещать искателю Web доступ к документам с некоторого сайта. Если документ, который разрешено просматривать искателю Web, содержит ссылки на запрещенные документы, можно задать, как нужно обрабатывать текст привязки для таких ссылок.

При конфигурировании искателя Web можно выбрать, индексировать ли текст привязки в ссылках на запрещенные документы. Для максимальной защиты укажите, что текст привязки в ссылках на запрещенные документы не должен индексироваться. Однако если текст привязки не индексируется, результаты поиска могут содержать не все документы, потенциально релевантные для запроса.

### Процедура

Чтобы включить или отключить индексацию текста привязки в ссылках на запрещенные документы:

1. Редактируя собрание, на странице Искатель найдите искатель Web, который хотите сконфигурировать, и выберите  **Свойства искателя**.
2. Нажмите кнопку **Изменить дополнительные свойства искателя Web**.
3. Чтобы индексировать текст привязки во всех документах, просматриваемых этим искателем, включите переключатель **Индексировать текст привязки в ссылках на запрещенные документы**. Пользователи смогут узнать о страницах, которые искателю Web не разрешено просматривать, выполняя поиск текста в тексте привязки ссылок, указывающих на такие страницы.  
Чтобы исключить из индекса текст привязки в ссылках на запрещенные документы, выключите этот переключатель. Пользователи не смогут узнать о страницах, которые искателю Web не разрешено просматривать. Текст привязки будет исключен из индекса вместе с запрещенными документами.
4. Нажмите кнопку **ОК** и затем, на странице Свойства искателя Web еще раз нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, остановите и перезапустите искатель.

Чтобы эти изменения были применены для ранее проиндексированных документов, нужно выполнить повторный просмотр документов искателем и затем заново проиндексировать их. Если предыдущие операции просмотра добавили в индекс информацию о запрещенных документах, эта информация будет удалена из индекса.

---

## Включение защиты для поисковой системы предприятия

Если вы хотите усилить защиту в то время, когда пользователи управляют поисковой системой предприятия или выполняют в ней поиск, необходимо сконфигурировать глобальную защиту в WebSphere Application Server. Необходимо также сконфигурировать информацию о защите в поисковой системе предприятия, консоли администратора и поисковых программах.

### Совет:

Подробные примеры о том, как включить глобальную защиту в WebSphere Application Server с репозиторием LDAP и сконфигурировать различные искатели для

использования защиты уровня документов приведены в документе IBM Redbook, IBM OmniFind Enterprise Edition Configuration and Implementation Scenarios.

## Процедура

Чтобы включить защиту в поисковой системе предприятия:

1. Определите, какой тип реестра пользователей вы хотите использовать для аутентификации пользователей. Например, многие администраторы WebSphere Application Server предпочитают использовать реестр пользователей Lightweight Directory Access Protocol (LDAP).
2. На консоли администратора поисковой системы предприятия выберите **Защита** и назначьте по крайней мере одному из пользователей в реестре пользователей WebSphere Application Server административную роль **Администратор поисковой системы предприятия**.

Вместо этого можно добавить ID администратора поисковой системы предприятия, заданный при установке OmniFind Enterprise Edition в реестре пользователей WebSphere Application Server.

**Важное замечание:** После включения глобальной защиты только те ID пользователей, которые занесены в реестр пользователей и которым назначена административная роль поисковой системы предприятия, будут иметь доступ к консоли администратора и смогут управлять поисковой системой предприятия.

3. Чтобы включить глобальную защиту и сконфигурировать реестр пользователей, следуйте процедурам, описанным в разделе “Конфигурирование глобальной защиты и реестра пользователей LDAP в WebSphere Application Server”.
4. Если вы включили глобальную защиту после установки OmniFind Enterprise Edition, надо сообщить поисковой системе предприятия ID пользователя и пароль WebSphere Application Server. Для этого служит команда **eschangewaspw**:
  - Если OmniFind Enterprise Edition установлена на одном сервере, действуйте согласно разделу “Включение защиты для поисковой системы предприятия с одним сервером” на стр. 280.
  - Если OmniFind Enterprise Edition установлена на нескольких серверах, действуйте согласно разделу “Включение защиты для поисковой системы предприятия с несколькими серверами” на стр. 281.
5. Выполните задачи, относящиеся к тем типам документов, которые вы собираетесь обрабатывать искателем и выполнять в них поиск. Подробности смотрите в разделе “Требования к конфигурации искателя для поддержки защиты” на стр. 282.

## Конфигурирование глобальной защиты и реестра пользователей LDAP в WebSphere Application Server

Чтобы включить защиту в OmniFind Enterprise Edition, необходимо включить глобальную защиту в WebSphere Application Server.

### Об этой задаче

В качестве составной части глобальной защиты необходимо сконфигурировать реестр пользователей для аутентификации ID пользователей. В этом разделе описывается, как сконфигурировать реестр пользователей Lightweight Directory Access Protocol (LDAP) в WebSphere Application Server одновременно со включением глобальной защиты. Несмотря на то, что WebSphere Application Server поддерживает другие типы реестров пользователей, вы не можете использовать реестр локальной операционной системы для аутентификации пользователей поисковой системы

предприятия. Чтобы использовать реестр локальной операционной системы, учетные записи операционной системы для всех пользователей на вашем предприятии должны существовать на серверах поиска поисковой системы предприятия.

В описании действий этой задачи предполагается, что вы используете WebSphere Application Server версии 6. Если у вас более ранняя версия WebSphere Application Server, пути по умолчанию и некоторые метки пользовательского интерфейса могут быть другими. Предполагается также, что используется IBM Tivoli Directory Server для реестра LDAP. При использовании другого типа реестра или пользовательского реестра следует ввести информацию, необходимую для этого реестра.

В данной задаче просуммированы шаги, необходимые для конфигурирования глобальной защиты при использовании с поисковой системой предприятия. Подробные указания смотрите в Информационном центре *WebSphere Application Server, Версия 6.0.x* по следующему адресу: <http://publib.boulder.ibm.com/infocenter/wasinfo/v6r0/index.jsp>

### Процедура

Чтобы включить глобальную защиту в WebSphere Application Server:

1. На сервере поиска поисковой системы предприятия перейдите к следующему URL, чтобы открыть консоль администратора WebSphere Application Server, где *локальный\_хост* - либо localhost, либо имя сервера, например, omnifind.search.xyz.com.  
`http://локальный_хост:9060/ibm/console`
2. Выберите **Защита**, а затем - **Глобальная защита**.
3. Настройте WebSphere для использования реестра LDAP:
  - a. Под надписью **Реестры пользователей** выберите **LDAP**.
  - b. Укажите ID пользователя и пароль сервера, которые используются для запуска сервера прикладных программ.
  - c. Для типа реестра выберите IBM Tivoli Directory Server.
  - d. Укажите имя хоста сервера LDAP как IP-адрес или как имя хоста службы доменных имен (DNS). По умолчанию используется номер порта 389.
  - e. Укажите базовое уникальное имя (distinguished name, DN), которое будет начальной точкой для поиска в реестре, например, ou=sales,o=ibm,c=us.
  - f. Поскольку некоторые серверы LDAP не поддерживают анонимное связывание при поиске в реестре, укажите DN для сервера прикладных программ, например, cn=searchuser,o=ibm,c=us, а затем укажите пароль для этого сервера прикладных программ. Сервер прикладных программ использует эти DN и пароль для связи с реестром.
  - g. Чтобы использовать протокол SSL для связи между WebSphere и LDAP, включите переключатель **SSL включен**.
  - h. Выберите **Применить**, а затем нажмите кнопку **ОК**.
4. Под **Общие свойства** включите переключатели **Включить глобальную защиту** и **Применить защиту Java 2**.
5. В качестве активного механизма аутентификации выберите Simple WebSphere Authentication Mechanism (SWAM).
6. В качестве активного реестра пользователей выберите реестр пользователей Lightweight Directory Access Protocol (LDAP).
7. Нажмите кнопку **ОК**.
8. Щелкните по ссылке **Сохранить** в верхней части страницы. После вопроса о сохранении сделанных изменений нажмите кнопку **Сохранить**.

9. На панели инструментов выберите **Выйти из программы**.
10. Остановите и перезапустите программу ESSearchServer.

#### **AIX, Linux или Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Эти сценарии находятся в каталоге WAS\_INSTALL\_ROOT/AppServer/bin:

- Для WebSphere Application Server Версии 5 путь установки по умолчанию - /usr/WebSphere в системах AIX, /opt/WebSphere в системах Linux и Solaris, C:\Program Files\WebSphere в системах Windows.
  - Для WebSphere Application Server Версии 6 путь установки по умолчанию - /usr/IBM/WebSphere в системах AIX, /opt/IBM/WebSphere в системах Linux и Solaris, C:\Program Files\IBM\WebSphere в системах Windows.
11. Перезапустите консоль администратора WebSphere Application Server.
  12. Так как сервер теперь запускается в защищенном режиме, введите ID пользователя и пароль сервера, указанные вами при конфигурировании реестра пользователей LDAP для регистрации в консоли (смотрите шаг 3b на стр. 279).

## **Включение защиты для поисковой системы предприятия с одним сервером**

Если вы включили глобальную защиту WebSphere Application Server после установки OmniFind Enterprise Edition, надо при помощи команды **eschangewaspw** обновить пароль пользователя WebSphere Application Server в файле конфигурации поисковой системы предприятия, `es.cfg`.

### **Прежде чем приступать к работе...**

Убедитесь, что в файле `config.properties` для программы ESSearchApplication указаны правильные имя пользователя и пароль WebSphere Application Server. По умолчанию положение этого файла - `ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF`.

### **Об этой задаче**

Команда **eschangewaspw** шифрует пароль перед его сохранением в файле `es.cfg`.

### **Процедура**

Чтобы включить использование глобальной защиты в существующей поисковой системе предприятия с одним сервером:

1. На сервере поисковой системы предприятия зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему.  
`esadmin system stopall`
2. Убедитесь, что в файле `ES_NODE_ROOT/nodeinfo/es.cfg` в записи `WASUser` указано правильное имя пользователя.
3. Запустите следующий сценарий, где *пароль\_WAS* - пароль для имени пользователя WebSphere Application Server, указанного в файле `ES_NODE_ROOT/nodeinfo/es.cfg` (смотрите шаг 2).

#### AIX, Linux или Solaris

`eschangewaspw.sh пароль_WAS`

#### Windows

`eschangewaspw пароль_WAS`

4. В Windows выберите **Панель управления** → **Администрирование** → **Службы** и добавьте то же имя пользователя и пароль, что и для WebSphere Application Server, для служб WebSphere Application Server и ESSearchServer.
5. Перезапустите поисковую систему предприятия:  
`esadmin system startall`

## Включение защиты для поисковой системы предприятия с несколькими серверами

Если вы включили глобальную защиту WebSphere Application Server после установки OmniFind Enterprise Edition, надо при помощи команды **eschangewaspw** обновить пароль пользователя WebSphere Application Server в файле конфигурации поисковой системы предприятия, `es.cfg`.

### Прежде чем приступать к работе...

Убедитесь, что в файле `config.properties` для программы `ESSearchApplication` указаны правильные имя пользователя и пароль WebSphere Application Server. По умолчанию положение этого файла на серверах поиска - `ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/ESSearchApplication.war/WEB-INF`.

### Об этой задаче

Команда **eschangewaspw** шифрует пароль перед его сохранением в файле `es.cfg`.

### Процедура

Чтобы включить использование глобальной защиты в существующей поисковой системе предприятия с несколькими серверами:

1. Выполните следующие шаги на сервере индексов поисковой системы предприятия:
  - a. Зарегистрируйтесь как администратор поисковой системы предприятия и остановите поисковую систему:  
`esadmin system stopall`
  - b. Убедитесь, что в файле `ES_NODE_ROOT/nodeinfo/es.cfg` в записи `WASUser` указано правильное имя пользователя.
  - c. Запустите следующий сценарий, где `пароль_WAS` - пароль для пользователя WebSphere Application Server, указанного в файле `ES_NODE_ROOT/nodeinfo/es.cfg` (смотрите шаг 1b).

#### AIX, Linux или Solaris

`eschangewaspw.sh пароль_WAS`

#### Windows

`eschangewaspw пароль_WAS`

2. Выполните следующие действия на втором сервере поиска (для конфигурации с двумя серверами), или на сервере искателей и на обоих серверах поиска (для конфигурации с четырьмя серверами):
  - a. Зарегистрируйтесь как администратор поисковой системы предприятия.

- b. Запустите следующий сценарий, где *пароль\_WAS* - пароль для пользователя WebSphere Application Server, указанного в файле `ES_NODE_ROOT/nodeinfo/es.cfg` (смотрите шаг 1b на стр. 281).

**AIX, Linux или Solaris**

```
eschangewaspw.sh пароль_WAS
```

**Окно командной строки Windows**

```
eschangewaspw пароль_WAS
```

3. В Windows выберите **Панель управления** → **Администрирование** → **Службы** и добавьте то же имя пользователя и пароль, что и для WebSphere Application Server, для служб WebSphere Application Server и ESSearchServer.
4. Перезапустите поисковую систему предприятия на сервере индексов поисковой системы предприятия:

```
esadmin system startall
```

---

## Требования к конфигурации искателя для поддержки защиты

Чтобы можно было включить сбор информации, необходимой для защиты уровня документа, у искателя должно быть разрешение на доступ к исходной информации защиты. Для некоторых типов данных для конфигурирования защищенной среды нужно выполнить дополнительные действия.

Таблица 7. Требования к конфигурации искателя для поддержки защиты

---

### Искатели Content Edition

Перед созданием искателя для доступа к репозиториям в непосредственном режиме сконфигурируйте систему WebSphere Information Integrator Content Edition для работы в непосредственном режиме и сконфигурируйте соединитель на сервере искателя.

Перед созданием искателя для доступа к репозиториям в режиме сервера запустите сценарий (`escrvbr.sh` в AIX, Linux и Solaris или `escrvbr.vbs` в Windows) для конфигурирования сервера искателя.

При конфигурировании искателя задайте ID пользователя и пароль, с помощью которых искатель сможет обращаться к каждому просматриваемому репозиторию. При необходимости для каждого репозитория в пространстве для искателя можно задать отдельные ID пользователя и пароль.

Темы, связанные с данной:

- “Доступ к репозиториям Content Edition в непосредственном режиме” на стр. 47
- “Доступ к репозиториям WebSphere II Content Edition в режиме сервера” на стр. 48
- “Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition” на стр. 49
- “Конфигурирование сервера искателей в Windows для WebSphere II Content Edition” на стр. 50

---

### Искатели DB2

Перед созданием искателя запустите сценарий (`escrdb2.sh` в AIX, Linux и Solaris или `escrdb2.vbs` в Windows) для конфигурирования сервера искателя.

При конфигурировании искателя для просмотра удаленных некаталогизированных баз данных задайте ID пользователя и пароль, с помощью которых искатель сможет просматривать каждую базу данных на сервере баз данных назначения. При необходимости для каждой базы данных в пространстве для искателя можно задать отдельные ID пользователя и пароль.

Темы, связанные с данной:

- “Конфигурирование сервера искателей в UNIX для искателей DB2” на стр. 53
- “Конфигурирование сервера искателей в Windows для искателей DB2” на стр. 54

---

### Искатели DB2 Content Manager

---

Таблица 7. Требования к конфигурации искателя для поддержки защиты (продолжение)

---

Перед созданием искателя запустите сценарий (escrcm.sh в AIX, Linux и Solaris или escrcm.vbs в Windows) для конфигурирования сервера искателя.	Темы, связанные с данной:
При конфигурировании искателя задайте ID пользователя и пароль, с помощью которых искатель сможет обращаться к каждому просматриваемому серверу. При необходимости для каждого сервера в пространстве для искателя можно задать отдельные ID пользователя и пароль.	<ul style="list-style-type: none"><li>• “Конфигурирование сервера искателей в UNIX для искателей DB2 Content Manager” на стр. 61</li><li>• “Конфигурирование сервера искателей в Windows для искателей DB2 Content Manager” на стр. 63</li></ul>

---

### Искатели Domino Document Manager, Примечания и QuickPlace

---

Для просмотра искателем серверов Lotus Domino, использующих протокол NRPC (Notes Remote Procedure Call):	Темы, связанные с данной:
<ul style="list-style-type: none"><li>• В системе AIX убедитесь, что на сервере искателя установлен и доступен модуль IOCP (I/O Completion Port, порт завершения ввода-вывода).</li><li>• Перед созданием искателя запустите сценарий (escrnote.sh в AIX, Linux и Solaris или escrnote.vbs в Windows) для конфигурирования сервера искателя.</li><li>• На сервере искателя поисковой системы предприятия должен быть установлен сервер Domino, и этот сервер Domino должен входить в домен Domino для просмотра.</li><li>• Чтобы при передаче пользователем требования поиска проверялись текущие параметры аутентификации пользователя, просматриваемый сервер Domino нужно сконфигурировать как доверенный сервер Lotus Domino.</li><li>• При конфигурировании искателя задайте путь для файла ID пользователя Lotus Notes с правами доступа к серверу, например, c:\Program Files\lotus\notes\data\name.id или /local/notesdata/name.id, и пароль для этого файла ID.</li></ul>	<ul style="list-style-type: none"><li>• “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 84</li><li>• “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 79</li><li>• “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 81</li><li>• “Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей” на стр. 286</li><li>• “Конфигурирование серверов, использующих протокол DIIOP” на стр. 83</li><li>• “Конфигурирование сервера QuickPlace для использования защиты Local User” на стр. 288</li><li>• “Конфигурирование Directory Assistance на сервере QuickPlace” на стр. 289</li></ul>

Для просмотра искателем серверов Lotus Domino, использующих протокол DIIOP (Domino Internet Inter-ORB Protocol):
<ul style="list-style-type: none"><li>• В системе AIX убедитесь, что на сервере искателя установлен и доступен модуль IOCP (I/O Completion Port, порт завершения ввода-вывода).</li><li>• Сконфигурируйте сервер искателя, чтобы он мог использовать этот протокол.</li><li>• При конфигурировании искателя задайте полный ID пользователя Lotus Notes с правами доступа к серверу, например, User Name/Any Town/My Company, и пароль для этого ID пользователя.</li></ul>

Для просмотра искателем серверов QuickPlace нужно сконфигурировать сервер QuickPlace для поддержки защиты локального пользователя (Local User) или Directory Assistance, в зависимости от типа требуемой защиты.

---

### Искатели Exchange Server

---

Таблица 7. Требования к конфигурации искателя для поддержки защиты (продолжение)

При конфигурировании искателя задайте ID пользователя с правами доступа к общедоступным папкам на просматриваемом сервере Exchange Server и пароль для этого ID пользователя.	Тема, связанная с данной: <ul style="list-style-type: none"><li>• “Проверка доступа к защищенным документам Exchange Server” на стр. 285</li></ul>
Чтобы искатель мог при просмотре данных использовать управление ключами Exchange Server и протокол Secure Sockets Layer (SSL), нужно также задать полный путь к файлу хранилища ключей и пароль, с помощью которого искатель сможет обращаться к этому файлу. Заданный файл хранилища ключей должен существовать на сервере искателя поисковой системы предприятия.	
<b>Искатели База данных JDBC</b>	
При конфигурировании искателя можно задать ID пользователя и пароль, с помощью которых искатель сможет просматривать таблицы в базе данных назначения. При необходимости для каждой базы данных в пространстве для искателя можно задать отдельные ID пользователя и пароль.	
<b>Искатели NNTP</b>	
Просматриваемые серверы NNTP должны разрешать серверу искателя читать данные.	
<b>Искатели Файловая система UNIX</b>	
Просматриваемые подкаталоги AIX, Linux и Solaris должны разрешать серверу искателя читать данные.	
<b>Искатели Web</b>	
Искатель Web следует протоколу Robots Exclusion. Если на Web-сервере на верхнем уровне каталогов сервера есть файл robots.txt, искатель анализирует этот файл и просматривает на этом сервере только те сайты, которые ему разрешено просматривать. Информацию об этом протоколе смотрите по адресу <a href="http://www.robotstxt.org/wc/exclusion.html">http://www.robotstxt.org/wc/exclusion.html</a> .	Темы, связанные с данной: <ul style="list-style-type: none"><li>• “Сайты, защищенные базовой аутентификацией HTTP” на стр. 104</li><li>• “Сайты, защищенные аутентификацией на основе форм” на стр. 105</li></ul>
При конфигурировании искателя Web:	
<ul style="list-style-type: none"><li>• Нужно задать имя пользовательского агента для искателя. В правилах в файлах robots.txt на серверах для просмотра можно задать это имя, чтобы разрешить или запретить доступ.</li><li>• Необязательно: Если для ограничения доступа к сайтам Web-сервер использует базовую аутентификацию HTTP, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.</li><li>• Необязательно: Если для ограничения доступа к сайтам Web-сервер использует формы HTML, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.</li></ul>	
<b>Искатели Список заполнения, Web Content Management и WebSphere Portal</b>	

Таблица 7. Требования к конфигурации искателя для поддержки защиты (продолжение)

---

<p>Перед созданием искателя нужно запустить сценарий установки для интеграции поисковой системы предприятия с сервером WebSphere Portal. Для различных версий WebSphere Portal поставляются разные сценарии.</p> <p>При конфигурировании искателя задайте полное уникальное имя (DN), с помощью которого искатель сможет получать страницы с просматриваемого сервера, например, uid=admin,cn=RegularEmployees,ou=Software Group,o=IBM,c=US, и задайте пароль для этого DN. Это DN должно совпадать с DN, заданным в WebSphere Portal.</p> <p>Убедитесь, что для заданного вами DN пользователя определены разрешения в компоненте Portal Access Control (PAC) WebSphere Portal. Искатель использует PAC для получения данных управления доступом для обрабатываемых документов.</p>	<p>Тема, связанная с данной:</p> <ul style="list-style-type: none"><li>• “Сценарии конфигурирования для интеграции поисковой системы предприятия с WebSphere Portal” на стр. 348</li></ul>
<hr/> <b>Искатели Файловая система Windows</b>	
<p>Просматриваемые подкаталоги должны разрешать серверу искателя читать данные. При конфигурировании искателя для просмотра удаленных файловых систем задайте ID пользователя, с помощью которого искатель сможет обращаться к удаленным данным, и задайте пароль для этого ID пользователя.</p> <p>Чтобы при передаче пользователем требования поиска проверялись текущие параметры аутентификации пользователя, нужно правильно сконфигурировать учетные записи домена. Требования к конфигурированию учетных записей доменов для файлов, найденных искателем на локальном компьютере, отличаются от требований для файлов, найденных искателем на удаленном сервере Windows.</p>	<p>Темы, связанные с данной:</p> <ul style="list-style-type: none"><li>• “Обеспечение защиты уровня документов для документов файловых систем Windows” на стр. 289</li><li>• “Защищенный поиск для доверенных доменов Windows” на стр. 291</li></ul>

---

## Проверка доступа к защищенным документам Exchange Server

Чтобы использовать искатель Exchange Server для документов, защищенных брандмауэром, убедитесь, что у сервера искателя есть доступ к серверу общедоступной папки Microsoft Exchange Server.

### Об этой задаче

Если у сервера искателя нет доступа к защищенному серверу Exchange Server, вы получите от сервера HTTP code 501 (Not Implemented). Кроме того, вы можете получить сообщения о неожиданном ответе HTTP.

### Процедура

Чтобы обеспечить доступ сервера искателя к документам за брандмауэром:

1. Запустите браузер на сервере искателя.

2. Перейдите на URL сервера общедоступной папки Exchange Server, который должен просматривать искатель. Например: `http://exchange.yourCompany.com/public/`
3. Убедитесь, что можете открыть страницу Exchange Server.  
Если у вас нет доступа к серверу Exchange Server, обратитесь к администратору сервера вашей организации.

## Обеспечение защиты уровня документов для документов Lotus Domino

Если просматриваемый искателем сервер Domino использует протокол вызова удаленных процедур (Notes remote procedure call, NRPC), необходимо сконфигурировать сервер искателя так, чтобы можно было применить управление доступом на уровне документов.

Чтобы применить защиту уровня документов для документов на сервере Domino, который использует протокол NRPC, надо установить сервер Domino на сервере искателя. Этот сервер Domino должен быть членом вашего домена Domino. Чтобы установить и сконфигурировать сервер Domino, следуйте инструкциям в документации Lotus Domino.

Чтобы серверы поиска могли проверять, есть ли у пользователя, выполняющего поиск в защищенном собрании, права просмотра документов, удовлетворяющих критериям поиска, необходимо выполнить также следующие задачи:

- “Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей”.
- “Конфигурирование глобальной защиты и реестра пользователей LDAP в WebSphere Application Server” на стр. 278.

### Понятия, связанные с данным

“Проверка текущих параметров аутентификации при обработке запроса” на стр. 268

“Искатели Notes” на стр. 76

## Конфигурирование доверенных серверов Lotus Domino для проверки параметров аутентификации пользователей

Чтобы применить защиту для документов, просмотренных искателем Примечания, который использует протокол вызова удаленных процедур Notes (Notes remote procedure call, NRPC), серверы Domino надо сконфигурировать, сделав их доверенными серверами Lotus Domino.

### Прежде чем приступить к работе...

Эта процедура обязательна, если нужно применить защиту на уровне документов при поиске в удаленных базах данных. Для поиска в базах данных, локальных по отношению к серверу искателя, эта процедура необязательна.

Для конфигурирования доверенных серверов на сервере искателя должен быть установлен сервер Domino. Этот сервер Domino должен быть членом вашего домена Domino.

### Об этой задаче

При конфигурировании опций защиты на уровне документов для искателя Примечания вы задаете, нужно ли применить управление доступом путем проверки

параметров аутентификации текущего пользователя при передаче им запроса. Для применения этого типа защиты просматриваемые искателем серверы Domino должны быть доверенными серверами Lotus Domino.

Когда пользователи выполняют поиск в доменах, требующих проверки их текущих данных регистрации, доверенный сервер позволяет переключить контекст с ID сервера Domino на ID текущего пользователя. База данных Domino открывается так, как если бы ее открыл текущий пользователь, и для этого пользователя применяется вся информация списка управления доступом к этой базе данных.

Как правило, возможность такого переключения контекстов доступна только для баз данных, хранящихся в каталоге data локального сервера Domino. Начиная с Lotus Domino Версии 6.5.1, эта возможность реализуется через доверенный сервер. При конфигурировании доверенного сервера администратор Domino задает, каким серверам Domino доверяется выполнение ответственных операций, таких как действия от имени другого пользователя при обращении к базе данных с удаленного компьютера.

### Процедура

Чтобы сконфигурировать доверенный сервер, выполните следующие действия на всех серверах Domino, просматриваемых искателем Примечания:

1. На сервере Domino откройте с помощью файла ID администратора домена Domino клиент администратора Lotus Domino.
2. Выберите **Файл**, затем **Открыть сервер**.
3. Введите имя сервера Domino, для которого нужно включить функциональные возможности доверенного сервера.
4. Выберите вкладку **Конфигурация**.
5. Раскройте объект **Сервер**, выберите документ **Текущий сервер** и нажмите кнопку **Редактировать сервер**.
6. Выберите вкладку **Защита**, прокрутите до конца документа, найдите запись **Доверенные серверы** и щелкните по стрелке вниз.
7. Выберите одну из опций:

#### **LocalDomainServers**

Выберите эту опцию, если все серверы в домене Domino должны считаться доверенными серверами.

#### *имя\_сервера*

Задайте имя сервера Domino, который вы хотите просматривать искателем как доверенный сервер.

Если просматриваемый сервер Domino находится в другом домене Domino, надо указать имя этого сервера или выбрать нужную группу **OtherDomainServers**. Надо также выполнить процедуры Domino для перекрестной сертификации файла ID сервера Domino с другими доменами Domino. Информацию об этих процедурах смотрите в документации по серверу Domino.

8. Нажмите кнопку **Сохранить и закрыть**, чтобы сохранить внесенные изменения.
9. Остановите и перезапустите удаленные серверы Domino, сконфигурированные вами как доверенные серверы.

#### **Понятия, связанные с данным**

“Проверка текущих параметров аутентификации при обработке запроса” на стр. 268

## Конфигурирование сервера QuickPlace для использования защиты Local User

Если вы планируете конфигурировать искатель QuickPlace для использования опции Local User при реализации защиты, перед созданием искателя надо сконфигурировать Domino Directory на сервере Lotus QuickPlace.

### Об этой задаче

При конфигурировании искателя QuickPlace вы выбираете режим защиты, который искатель использует на уровне документов. Если выбрать режим Local User, надо убедиться, что все ID локальных пользователей и локальные группы зарегистрированы в Domino Directory (иерархия Domino Directory должна соответствовать иерархии QuickPlace).

Кроме того, надо убедиться, что ID пользователя и пароль, под которыми будет работать искатель, зарегистрированы в Domino Directory и наделены правами чтения просматриваемой базы данных.

Для использования QuickPlace требуется только имя пользователя. Но для просмотра источников QuickPlace требуется полностью развернутый ID пользователя. Формат развернутого ID пользователя:

*имя\_пользователя/имя\_площадки/QR/имя\_домена*

Определите при помощи этой процедуры полностью развернутую версию ID пользователя, убедитесь, что у этого ID пользователя есть права чтения базы данных QuickPlace, и добавьте ID пользователя в Domino Directory. В Domino Directory должен быть ID пользователя, под которым искатель будет просматривать базы данных QuickPlace, и все локальные пользователи и группы QuickPlace (иерархия Domino Directory должна соответствовать иерархии QuickPlace).

### Процедура

Чтобы сконфигурировать сервер QuickPlace для использования защиты Local User:

1. Подтвердите разрешения ID пользователя:
  - a. Откройте документ Server на сервере QuickPlace.
  - b. Откройте страницу Файлы, затем откройте список управления доступом (список ACL) для базы данных, которую будет просматривать искатель.
  - c. Убедитесь, что ID локального пользователя, под которым будет работать искатель, есть в ACL, и что у этого ID пользователя есть разрешение на чтение базы данных.  
Полностью развернутую форму этого ID пользователя вы должны были задать на шаге 2.
2. Добавьте пользователя в Domino Directory:
  - a. Откройте документ Server на сервере QuickPlace.
  - b. На странице People and Groups в объекте дерева людей добавьте полностью развернутый ID пользователя, который вы подтвердили на шаге 1.
  - c. В поле **Internet password** (Пароль Интернет) задайте пароль для этого ID пользователя.

## Конфигурирование Directory Assistance на сервере QuickPlace

Если вы планируете конфигурировать искатель QuickPlace для использования каталога LDAP при реализации защиты, сначала надо создать базу данных Directory Assistance на сервере Lotus QuickPlace.

### Ограничения

На сервере QuickPlace, который вы хотите просматривать, должны быть запущены задачи DPOP и HTTP.

### Процедура

Чтобы сконфигурировать LDAP Directory Assistance на сервере QuickPlace:

1. Создайте базу данных Directory Assistance:
  - a. Откройте документ Server на сервере QuickPlace.
  - b. Создайте базу данных при помощи шаблона **Directory Assistance(6)**. Этот шаблон находится на сервере.
  - c. Выберите **Добавить Directory Assistance**, чтобы создать документ в базе данных.
  - d. Откройте вкладку Основные и в поле **Тип домена** выберите **LDAP**.
  - e. Откройте вкладку Контексты именования и убедитесь, что переключатель **Доверенные для параметров аутентификации** включен.
  - f. Откройте вкладку LDAP и задайте информацию о сервере LDAP.
  - g. Сохраните и закройте документ Server.
2. Сконфигурируйте сервер QuickPlace для использования базы данных Directory Assistance:
  - a. Откройте документ Server на сервере QuickPlace.
  - b. Откройте вкладку Основные и в поле **Имя базы данных Directory Assistance** задайте имя базы данных, созданной на шаге 1.
  - c. Сохраните и закройте документ Server.

Теперь сервер QuickPlace может использовать сервер LDAP как вторичный каталог Domino.

## Обеспечение защиты уровня документов для документов файловых систем Windows

Чтобы включить проверку текущих параметров аутентификации пользователя при поиске в документах, просмотренных искателем Файловая система Windows, надо сконфигурировать информацию учетной записи домена и на сервере искателя, и на сервере Microsoft Windows.

При конфигурировании искателя Файловая система Windows вы задаете, просматривать ли подкаталоги на локальном компьютере или же на удаленном компьютере. Если для собрания включена защита, можно задать также опции управления доступом к документам в просматриваемых подкаталогах.

Если вы решили применить управление доступом путем проверки параметров аутентификации текущего пользователя при передаче им запроса, надо проверить, что учетные записи домена правильно сконфигурированы. Требования к конфигурированию учетных записей доменов для файлов, найденных искателем на локальном компьютере, отличаются от требований для файлов, найденных искателем на удаленном сервере Windows.

**Важное замечание:** Нельзя выполнять проверку параметров аутентификации пользователей при обработке запросов, если выполняются оба следующих условия:

- Север Windows для просмотра искателем не входит ни в какой домен.
- Каталог для просмотра искателем - удаленный каталог, например, \\servername\hostname.

## Проверка с помощью локальных данных управления доступом

Для проверки параметров аутентификации текущего пользователя система использует информацию как его локальной учетной записи, так и учетной записи домена (если компьютер принадлежит к домену Windows). Для проверки параметров аутентификации при обработке запроса информация защиты должна включать оба имени пользователя - локальное и доменное.

### Локальные учетные записи

В локальной учетной записи имя пользователя записывается в формате:

*ИМЯ\_КОМПЬЮТЕРА\ИМЯ\_ПОЛЬЗОВАТЕЛЯ*

Для входа в систему пользователи указывают только имя пользователя, но для правильного назначения прав пользователя Windows используется полное имя. Например, если в локальной учетной записи имя пользователя - abcuser, в полной учетной записи оно может быть WINSERVER1\abcuser.

Работая с программами поиска и конфигурируя профили для поиска защищенных документов в локальной системе, пользователи должны указывать имя для регистрации в Windows (например, abcuser).

### Доменные учетные записи

В доменной учетной записи имя пользователя записывается в формате:

*ИМЯ\_ДОМЕНА\ИМЯ\_ПОЛЬЗОВАТЕЛЯ*

Для входа в систему пользователи вводят эту информацию в формате:

*ИМЯ\_ПОЛЬЗОВАТЕЛЯ@ИМЯ\_ДОМЕНА*

Например, если вы сконфигурировали назначения прав пользователей для файла и выбрали домен WIN1\abcuser, учетная запись будет иметь вид abcuser@win1.company.com.

Работая с программами поиска и конфигурируя профили для поиска документов в защищенном домене, пользователи должны указывать имя для регистрации в Windows (например, abcuser@win1.company.com).

Чтобы применить проверку текущих параметров аутентификации на локальных компьютерах, в учетных записях пользователей Windows, используемых сервером искателя, должны быть следующие права. Назначить права пользователям можно с помощью окна Windows Администрирование: **Администрирование** → **Локальная политика безопасности** → **Локальные политики** → **Назначение прав локальным пользователям**.

- У ID пользователя, от имени которого запущен сервер искателя, должно быть право **Действовать как часть операционной системы**. Это право конфигурируется для пользователя - администратора поисковой системы предприятия на сервере искателя при установке OmniFind Enterprise Edition.
- У пользователей должно быть право **Входить локально**.

## Проверка с помощью данных управления доступом к удаленному домену

Для операционной системы Windows любой каталог, начинающийся с записи `\\имя_сервера` считается удаленным каталогом. Например:  
`\\software\utilities\IBM`

Для доступа к удаленному каталогу пользователи должны задать свои имена в формате:

`ИМЯ_ПОЛЬЗОВАТЕЛЯ@ИМЯ_ДОМЕНА`

Работая с программами поиска и конфигурируя профили для поиска защищенных документов в удаленной системе, пользователи должны указывать имя для регистрации в удаленной системе Windows (например, `abcuser@win1.company.com`).

Чтобы применить проверку текущих параметров аутентификации на удаленных компьютерах, в учетных записях пользователей Windows должны быть следующие права. Назначить права пользователям можно с помощью окна Windows Администрирование: **Администрирование** → **Политика безопасности домена**.

- Сервер искателя и просматриваемый сервер Windows должны входить в один домен.
- У ID пользователя, от имени которого запущен сервер искателя, должно быть право **Действовать как часть операционной системы**. Это право конфигурируется для пользователя - администратора поисковой системы предприятия на сервере искателя при установке OmniFind Enterprise Edition.
- У пользователей должно быть право **Входить в качестве пакетного задания**.

### Понятия, связанные с данным

“Проверка текущих параметров аутентификации при обработке запроса” на стр. 268

“Искатели файловой системы Windows” на стр. 117

“Защищенный поиск для доверенных доменов Windows”

## Защищенный поиск для доверенных доменов Windows

Чтобы обеспечить защиту уровня документов для удаленных файловых систем Windows, поисковая система предприятия поддерживает верификацию по спискам управления доступом (ACL) в доверенных доменах.

## Конфигурирование искателя

Чтобы сконфигурировать искатель Файловая система Windows для поддержки доверенных доменов, надо задать опции в новом файле конфигурации. С консоли администратора поиска сконфигурировать эту возможность нельзя.

1. Чтобы включить поддержку защиты на уровне документов в доверенных доменах Windows, отредактируйте следующий файл:

`КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_сеанса/winfscrawler_ext.xml`

**Совет:** Чтобы определить ID сеанса для искателя Файловая система Windows, который вы хотите сконфигурировать, можно задать слежение за искателем с консоли администратора поисковой системы предприятия или использовать команду **esadmin report collections**.

2. Задайте имя домена Windows и имя NETBIOS для Active Directory. Например:

```
<ExtendedProperties>
 <SetAttribute XPath="/Crawler/DataSources/Server/Target"
 Name="Domain">jk.enterprises.com
 </SetAttribute>
 <SetAttribute XPath="/Crawler/DataSources/Server/Target"
 Name="NetBIOSDomain">JKE1
 </SetAttribute>
</ExtendedProperties>
```

3. Остановите и перезапустите искатель, чтобы изменения вступили в силу.

### Ограничения

- Документы не могут содержать списки управления доступом из нескольких доменов Windows. Пользователи и группы домена для собрания должны принадлежать одному домену Windows.
- Для поддержки верификации доступа к удаленной файловой системе серверы Windows должны работать в одном домене Windows или в доверенных доменах Windows.
- Искатель файловых систем Windows читает имя NETBIOS для Active Directory, связанного с просматриваемым сервером Windows, и использует это имя NETBIOS для фильтрации файла ACL. Active Directory, к которому подсоединяется сервер искателей, доверяет другому Active Directory, который определяет учетные записи пользователей и групп.
- Учетная запись пользователя, которую вы задали для искателя, чтобы он мог обращаться к удаленному серверу Windows, должна принадлежать к домену Windows, для которого вам требуется управление доступом.
- Операционная система Windows разрешает только одной учетной записи связываться с сетевыми папками на одном файл-сервере. Прочие учетные записи не могут соединиться с тем же файл-сервером в то же время. Поэтому вы не можете сконфигурировать различные учетные записи для различных искателей, просматривающих один сервер Windows, даже если это искатели из разных собраний.

#### Понятия, связанные с данным

“Обеспечение защиты уровня документов для документов файловых систем Windows” на стр. 289

---

## Отключение защиты для поисковой системы предприятия

Можно отключить защиту для прикладной программы поисковой системы предприятия в WebSphere Application Server. Если ранее были сконфигурированы элементы управления защитой уровня документов, можно задать, чтобы эти элементы управления игнорировались. Параметры защиты влияют также на вывод объединенных результатов в результатах поиска.

## Отключение защиты для прикладной программы предприятия в WebSphere Application Server

Для управления аутентификацией пользователей в процессах поисковой системы предприятия WebSphere Application Server можно отключить глобальную защиту для отдельных прикладных программ предприятия.

### Об этой задаче

Программа установки OmniFind Enterprise Edition внедряет в WebSphere Application Server три прикладных программы предприятия:

- Программа ESAdmin содержит интерфейс для консоли администратора поисковой системы предприятия.
- Программа ESSearchApplication содержит интерфейс для консоли примера программы поиска.
- Программа ESSearchServer обеспечивает всю удаленную связь для реализации SIAPИ поисковой системы предприятия и позволяет интерфейсам SIAPИ взаимодействовать с серверами поиска.

По умолчанию все три прикладных программы предприятия поддерживают глобальную защиту WebSphere Application Server. Если эти программы определяют, что глобальная защита включена, они начинают аутентифицировать все получаемые ими требования.

В некоторых организациях может потребоваться включить или отключить защиту для отдельных прикладных программ предприятия. Например, можно аутентифицировать всех пользователей с доступом к консоли администратора, но не аутентифицировать пользователей, использующих интерфейсы SIAPИ или пример программы поиска.

### Процедура

Чтобы отключить защиту для прикладной программы предприятия:

1. На сервере поиска запустите консоль администратора WebSphere Application Server.  
Чтобы открыть консоль администратора, можно:
  - Выбрать эту программу в меню **Пуск** Windows.
  - В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя\_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
  - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя\_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
2. В окне ввода ID пользователя и пароля введите ID администратора и пароль, заданные при включении в WebSphere Application Server глобальной защиты.
3. После регистрации на консоли администратора выберите **Applications** (Прикладные программы), а затем **Enterprise Applications** (Прикладные программы предприятия).
4. Включите переключатель рядом с именем прикладной программы предприятия, для которой надо отключить защиту.
5. Прокрутите список вниз и щелкните по ссылке **Map security roles to users/groups** (Отобразить роли защиты для пользователей/групп).
6. Найдите роль **AllAuthenticated** и включите переключатель под столбцом **Everyone?** (Все?).
7. Нажмите кнопку **OK**.
8. Выберите **Save** (Сохранить), чтобы сохранить внесенные изменения.
9. При использовании WebSphere Network Deployment включите переключатель **Synchronize changes with Nodes** (Синхронизировать изменения с Nodes).
10. Нажмите кнопку **Save** (Сохранить).

11. Остановите и перезапустите программу ESSearchServer.

#### **AIX, Linux или Solaris**

```
./stopServer.sh ESSearchServer
./startServer.sh ESSearchServer
```

#### **Windows**

```
stopServer ESSearchServer
startServer ESSearchServer
```

Эти сценарии находятся в каталоге WAS\_INSTALL\_ROOT/AppServer/bin:

- Для WebSphere Application Server Версии 5 путь установки по умолчанию - /usr/WebSphere в системах AIX, /opt/WebSphere в системах Linux и Solaris, C:\Program Files\WebSphere в системах Windows.
- Для WebSphere Application Server Версии 6 путь установки по умолчанию - /usr/IBM/WebSphere в системах AIX, /opt/IBM/WebSphere в системах Linux и Solaris, C:\Program Files\IBM\WebSphere в системах Windows.

## **Отключение защиты на уровне документов**

Можно разрешить пользователям выполнять поиск в собрании независимо от того, связано ли с документами в индексе какое-либо управление доступом. Для искателей, поддерживающих проверку текущих параметров аутентификации, можно также разрешить пользователям поиск в собрании без проверки текущих элементов управления доступом во время обработки запросов.

### **Прежде чем приступить к работе...**

Включать или отключать защиту на уровне документов для всех документов в собрании может администратор поисковой системы предприятия или администратор собрания.

### **Ограничения**

Опции защиты на уровне документов можно указывать только тогда, когда защита была включена во время создания собрания.

### **Об этой задаче**

Можно сконфигурировать искатели для связывания маркеров защиты с документами, просматриваемыми этими искателями. Поисковые программы могут использовать эти маркеры, хранящиеся в индексе, для обеспечения управления доступом при выполнении пользователями поиска в собрании. Для некоторых искателей можно также указать, что при передаче пользователями запросов надо проверять текущие элементы управления доступом, связанные с документами в их исходных репозиториях.

Чтобы снять эти ограничения защиты, можно указать, что серверы поиска должны игнорировать все маркеры защиты, переданные с запросом. Можно также разрешить пользователям запрашивать документы без сравнения их параметров аутентификации с текущими элементами управления доступом.

Временное отключение защиты на уровне документов может потребоваться при тестировании нового собрания или при устранении ошибки поисковой программы.

### **Процедура**

Чтобы отключить управление доступом на уровне документов:

1. Отредактируйте собрание, перейдите на страницу **Общие** и нажмите кнопку **Включить или выключить защиту на уровне документов**.
2. На странице **Защита на уровне документов** для всех документов включите переключатель **Игнорировать управление доступом на уровне документов в индексе**, если вы не хотите, чтобы маркеры защиты, которые искатели связывают с документами, использовались при запросах пользователей к собранию.  
При этом искатели продолжают добавлять маркеры защиты к документам, но поисковые серверы игнорируют эти маркеры, что позволяет пользователям находить ранее защищенные документы.
3. Включите переключатель **Не проверять текущие параметры аутентификации во время обработки запроса**, если вы не хотите проверять текущие элементы управления доступом, связанные с документами в исходных репозиториях при передаче запросов пользователями. Этот переключатель доступен только для документов, найденных искателями, которые поддерживают эту возможность.  
Если этот переключатель включен, другие опции защиты на уровне документов продолжают действовать. Например, если при конфигурировании искателя вы задали опции для сохранения элементов управления доступом в индексе, это управление защитой будет продолжать действовать, пока не будет включен также переключатель **Игнорировать управление доступом на уровне документов в индексе**.

## Отключение защиты для объединенных результатов поиска

Когда включена защита собрания, результаты поиска будут объединены с другими результатами поиска на том же сайте, если не указать, что вы не хотите проверять параметры аутентификации пользователя во время обработки запроса.

### Прежде чем приступить к работе...

Включать или отключать проверку параметров аутентификации для всех документов в собрании может администратор поисковой системы предприятия или администратор собрания.

### Ограничения

Опции защиты на уровне документов можно указывать только тогда, когда защита была включена во время создания собрания.

### Об этой задаче

При проверке текущих параметров аутентификации источники всех документов изучаются и, возможно, направляются на проверку. Необходимо отключить проверку текущих параметров аутентификации, если вы хотите, чтобы документы, имеющие один и тот же префикс URI или принадлежащие к ранее сконфигурированной объединенной группе URI, были объединены в результатах поиска.

### Процедура

Для отключения проверки текущих параметров аутентификации, чтобы документы были объединены в результатах поиска:

1. Отредактируйте собрание, перейдите на страницу **Общие** и нажмите кнопку **Включить или выключить защиту на уровне документов**.

2. На странице Защита на уровне документов для всех документов включите переключатель **Не проверять текущие параметры аутентификации во время обработки запроса**.
3. В режиме слежения за собранием откройте страницу Поиск и остановите, а затем перезапустите процессы серверов поиска.

При запросе пользователями собрания документы, имеющие один и тот же префикс URI или принадлежащие к сайтам, сконфигурированным для объединения, объединяются в результатах поиска. В примере поисковой программы пользователи могут просматривать объединенные результаты, щелкнув по ссылке **Еще результаты из того же источника**.

---

## Запуск и остановка поисковой системы предприятия

После создания собрания надо запустить серверы искателей, анализа и индексации (серверы поиска запускаются автоматически). После внесения изменений в собрание остановите и перезапустите эти серверы.

Большинство серверов поисковой системы предприятия могут работать непрерывно или в соответствии с заданными расписаниями. Например, можно задать расписания для построения полных и дельта-индексов. После запуска поисковой системы предприятия обычно останавливать и перезапускать процессы сервера нужно, только если изменены параметры конфигурации (например, изменены категории или увеличен размер кэша поиска).

При изменении содержимого собрания или правил сбора искателями данных из источников в вашем предприятии для вступления изменений в силу нужно остановить и снова запустить искатели. Если не изменять правила для искателей, искатели Web работают непрерывно, а остальные - в соответствии с заданными расписаниями.

Чтобы повысить доступность серверов поиска в случае, когда сервер индексов и консоль администратора недоступны, можно задать команды запуска серверов поиска для собрания в автономном режиме.

---

## Запуск поисковой системы предприятия

Чтобы пользователи могли выполнять поиск в собрании, нужно запустить сначала процессы системы, а затем серверы искателя, анализатора и индексации (серверы поиска запускаются автоматически).

### Прежде чем приступать к работе...

Сконфигурируйте источники данных, с которыми должен работать искатель, и задайте опции, определяющие, как выполнять синтаксический анализ, индексацию и поиск данных. Например, если вы хотите, чтобы пользователи могли просматривать в результатах поиска подробную информацию о категориях, сконфигурируйте категории перед запуском анализатора.

Запустить серверы поисковой системы предприятия может только администратор поиска предприятия, администратор собрания или оператор с полномочиями управления этим собранием.

Серверы поисковой системы предприятия нужно запускать для собрания в правильном порядке. Например, сначала необходимо запустить искатель и просмотреть данные; только после этого можно будет индексировать документы.

### Ограничения

Чтобы запустить поисковую систему предприятия, у вас должна быть учетная запись пользователя, позволяющая пройти локальную аутентификацию. При попытке запустить эту систему с учетной записью Andrew File System (AFS) произойдет ошибка.

### Процедура

Чтобы запустить поисковую систему предприятия:

1. Если поисковая система предприятия используется в конфигурации с двумя или четырьмя серверами, зарегистрируйтесь как администратор поисковой системы предприятия и запустите на каждом сервере CCL (common communication layer - общий слой связи):

#### **AIX, Linux или Solaris**

```
startccl.sh -bg
```

#### **Окно командной строки Windows**

```
startccl
```

#### **Администрирование - Службы Windows**

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск → Программы → Инструменты управления → Службы.**
  - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск.**
2. Запустите компоненты поисковой системы предприятия:
    - a. Зарегистрируйтесь на любом из серверов поисковой системы предприятия как администратор поисковой системы предприятия.
    - b. Запустите все компоненты системы: `esadmin system startall`  
Эта команда запускает следующие процессы и программы:
      - Web-сервер (в конфигурации с несколькими серверами Web-сервер запускается на обоих серверах поиска)
      - Программы ESSearchServer и ESAdmin на сервере WebSphere Application Server (в конфигурации с несколькими серверами эти программы запускаются на обоих серверах поиска)
      - Сеанс ESAdmin на сервере индексов
      - CCL на компьютере, с которого запускается команда (если CCL еще не запущен)
      - Сетевой сервер баз данных для поисковой системы предприятия
      - Информационный центр поисковой системы предприятия (в конфигурации с несколькими серверами информационный центр запускается на обоих серверах поиска)
  3. Запустите консоль администратора поисковой системы предприятия и зарегистрируйтесь как администратор поисковой системы предприятия. Если используются административные роли, можно зарегистрироваться как администратор собрания или как оператор с полномочиями управления собранием, которое нужно запустить.
  4. В окне Собрания найдите нужное собрание и нажмите кнопку  **Слежение.**
  5. На странице Искатель нажмите для каждого запускаемого искателя кнопку  **Запустить.**
    - Если вы запускаете искатель Web, этот искатель сразу начинает просмотр данных. Искатели этих типов работают непрерывно и выполняют просмотр и повторный просмотр документов Web.
    - Если вы запускаете искатель другого типа, запускается сеанс искателя. Искатель начнет просмотр данных согласно дате и времени в расписании. Если вы не задали расписание или хотите запустить искатель раньше, включите мониторинг искателя и щелкните по значку запуска для каждого источника данных, который должен обработать искатель.

После запуска искателя можно позволить ему работать непрерывно. Если искатель есть в расписании, он будет снова запускаться в заданные в расписании даты и моменты времени.

6. После того, как искатель выполнит просмотр данных, откройте страницу Анализ и нажмите кнопку  **Запустить**, чтобы запустить анализатор.

Можно позволить анализатору работать непрерывно. Обычно анализатор нужно останавливать только после внесения изменений в обработку им данных (например, после изменения категорий или отображения полей XML).

7. Необязательно: Чтобы принудительно запустить процесс индексации, не дожидаясь запланированных для этого даты и времени, откройте страницу Индекс и в области **Главный** нажмите кнопку  **Запустить**.

Можно позволить процессу индексации работать непрерывно. Индекс будет построен в запланированные даты и моменты времени.

**Совет:** Серверы поиска запускаются автоматически, и можно позволить им работать непрерывно. Обычно серверы поиска не требуется останавливать, за исключением случаев изменения параметров кэша поиска или сводки документов. Чтобы перезапустить серверы поиска, откройте страницу Поиск, нажмите кнопку  **Остановить**, а затем  **Запустить**.

Чтобы повысить доступность серверов поиска в случае, когда консоль администратора недоступна, можно задать команды запуска серверов поиска для собрания в автономном режиме. Если сервер индексов не запущен, консоль администратора недоступна.

#### **Понятия, связанные с данным**

“Административные роли” на стр. 262

#### **Задачи, связанные с данной**

“Регистрация на консоли администратора” на стр. 19

“Управление набором запускаемых или останавливаемых компонентов” на стр. 301

“Управление серверами поиска в автономном режиме” на стр. 302

---

## **Остановка поисковой системы предприятия**

Остановка и перезапуск сервера поисковой системы предприятия может потребоваться, если вы внесли изменения в его конфигурацию или хотите исправить ошибки.

### **Прежде чем приступить к работе...**

Остановить серверы поисковой системы предприятия может только администратор поисковой системы предприятия, администратор собрания или оператор с полномочиями управления этим собранием.

### **Об этой задаче**

Серверы поисковой системы предприятия можно остановить независимо друг от друга. Например, если вы останавливаете и перезапускаете искатель, чтобы он использовал внесенные в его конфигурацию изменения, не нужно останавливать и перезапускать серверы поиска.

### **Процедура**

Чтобы остановить серверы поисковой системы предприятия:

1. В окне Собрания найдите нужное собрание и выберите  **Слежение**.
2. На странице Искатель найдите нужный искатель и остановите или приостановите его.

Если вы изменили свойства искателя, остановите и перезапустите искатель, чтобы он использовал эти изменения. Если вы изменили пространство искателя и хотите применить изменения к документам, которые уже проиндексированы, нужно также выполнить их повторный просмотр искателем.

**Совет:** Сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажмите кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Процесс завершен, если значок состояния для искателя указывает, что он остановлен.

3. На странице Анализ нажмите кнопку  **Остановить**, чтобы остановить анализатор.  
Если вы изменили правила обработки анализатором данных, остановите и перезапустите анализатор, чтобы он использовал эти изменения. Изменения применяются только к новым просматриваемым искателем документам. Чтобы применить изменения к документам, которые уже внесены в индекс, нужно запустить полный повторный просмотр искателем этих документов, после чего проанализировать и проиндексировать их заново.
4. На странице Индекс нажмите кнопку  **Остановить**, чтобы остановить выполняемое построение индекса.  
Можно также остановить построение индекса при отслеживании очереди заданий. Для этого на панели инструментов выберите **Система**, откройте страницу Индекс и нажмите кнопку  **Остановить** для индекса, построение которого вы хотите остановить.
5. На странице Поиск нажмите кнопку  **Остановить**, чтобы остановить серверы поиска. Обычно остановка и перезапуск серверов поиска требуется только при изменении параметров кэша или сводки документов.
6. Чтобы остановить не отдельные серверы, а всю поисковую систему предприятия:
  - a. Зарегистрируйтесь на любом из серверов поисковой системы предприятия как администратор поисковой системы предприятия.
  - b. Остановите все компоненты системы: `esadmin system stopall`

Эта команда останавливает следующие процессы и программы:

- Web-сервер (в конфигурации с несколькими серверами Web-сервер будет остановлен на обоих серверах поиска)
- Программы ESSearchServer и ESAdmin на сервере WebSphere Application Server (в конфигурации с несколькими серверами эти программы будут остановлены на обоих серверах поиска)
- Сеанс ESAdmin на сервере индексов
- CCL (common communication layer - общий слой связи) для поисковой системы предприятия на компьютере, с которого запускается команда:
- Сетевой сервер баз данных для поисковой системы предприятия
- Информационный центр поисковой системы предприятия (в конфигурации с несколькими серверами информационный центр будет остановлен на обоих серверах поиска)

#### **Задачи, связанные с данной**

“Регистрация на консоли администратора” на стр. 19

---

## Управление набором запускаемых или останавливаемых компонентов

Управлять набором запускаемых и останавливаемых компонентов можно при помощи команд **esadmin system startall** и **esadmin system stopall**.

### Об этой задаче

Компоненты системы поиска предприятия, которые можно запустить или остановить при помощи команд **esadmin system startall** и **esadmin system stopall**, перечислены в файле `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/AutoRunComponents.properties`. По умолчанию эти команды запускают и останавливают все компоненты в этом списке.

Если вы хотите предотвратить запуск или остановку определенных компонентов, отредактируйте этот файл.

### Процедура

Чтобы задать, какие компоненты следует запустить или остановить при запуске или остановке поисковой системы предприятия:

1. Зарегистрируйтесь как администратор поисковой системы предприятия на сервере, где планируется запускать команду **esadmin system startall** или **esadmin system stopall**.
2. Отредактируйте файл `КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/default_config/AutoRunComponents.properties`.
3. Чтобы отключить запуск какого-либо компонента, добавьте для него поле `Component.startable.ID_компонента=false`, где *ID\_компонента* - компонент, который вы не хотите запускать.
4. Чтобы отключить остановку какого-либо компонента, добавьте для него поле `Component.storable.ID_компонента=false`, где *ID\_компонента* - компонент, который вы не хотите останавливать.
5. Сохраните файл и закройте текстовый редактор.

При следующем вводе команды **esadmin system startall** или **esadmin system stopall** этот компонент не будет запущен или остановлен в соответствии с внесенными для него изменениями в файле свойств.

**Пример:** В следующем примере сервер HTTP будет запущен командой **esadmin system startall** (заданной по умолчанию), но не будет остановлен командой **esadmin system stopall** (управляемой в этом примере выделенной строкой):

```

Подробности для компонента 3

Component.name.3=IBM HTTP server
Component.impl.class.3=com.ibm.es.control.util.component.impl.HTTPControlImpl
Component.nodes.3=search
По умолчанию запускаются все компоненты.
Component.storable.3=false
```

### Задачи, связанные с данной

“Запуск поисковой системы предприятия” на стр. 297

## Управление серверами поиска в автономном режиме

Для повышения доступности серверов поиска их можно запустить для отдельных собраний, даже если сервер индексов не работает.

### Ограничения

Для остановки и запуска серверов поиска нужно быть администратором поисковой системы предприятия.

Запускать и останавливать серверы поиска в автономном режиме с консоли администратора поисковой системы предприятия нельзя. Если сервер индексов не запущен, консоль администратора недоступна.

Прежде чем вы сможете запускать и останавливать серверы в автономном режиме, убедитесь, что серверы искателей, анализатора, индексов и поиска для данного собрания запускались про крайней мере однажды. Это необходимо для синхронизации нужных файлов на серверах поиска.

Если для собрания включена защита на уровне документов, убедитесь также, что запущен сервер искателей. Это необходимо для работы элементов управления защитой на уровне документов. Если сервер искателей не запущен, в результатах поиска будут возвращены только документы, не требующие аутентификации.

### Об этой задаче

Если сервер индексов доступен, можно позволить пользователям продолжать выполнять поиск в системе, запустив серверы поиска в автономном режиме. В установке с несколькими серверами ввести команды запуска и остановки серверов можно с любого сервера поисковой системы предприятия. Эти команды попытаются запустить или остановить все доступные серверы поиска для указанного собрания.

### Процедура

Чтобы запустить или остановить серверы поиска в автономном режиме:

1. Чтобы запустить серверы поиска для собрания, если сервер индексов не работает:

- a. На сервере искателей и на серверах поиска зарегистрируйтесь как администратора поисковой системы предприятия и запустите службу CCL:

#### AIX, Linux или Solaris

```
startccl.sh -bg
```

#### Окно командной строки Windows

```
startccl
```

#### Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- 1) Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
  - 2) Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
- b. На серверах поиска запустите сценарий **startServer**, который расположен в каталоге *КАТАЛОГ\_УСТАНОВКИ\_WAS/AppServer/bin*, чтобы запустить программу **ESSearchServer** в **WebSphere Application Server**:

### **AIX, Linux или Solaris**

```
./startServer.sh ESSearchServer
```

### **Windows**

```
startServer ESSearchServer
```

- c. Введите следующую команду, где *id\_собрания* идентифицирует собрание, к которому относятся запускаемые серверы поиска:  

```
esadmin startSearch -cid id_собрания
```
2. Чтобы остановить серверы поиска для собрания, если сервер индексов не работает:
  - a. Зарегистрируйтесь на любом из серверов поисковой системы предприятия как администратор поисковой системы предприятия.
  - b. Введите следующую команду, где *id\_собрания* идентифицирует собрание, к которому относятся останавливаемые серверы поиска:  

```
esadmin stopSearch -cid id_собрания
```

### **Задачи, связанные с данной**

“Запуск поисковой системы предприятия” на стр. 297

“Остановка поисковой системы предприятия” на стр. 299



---

## Слежение за активностью поисковой системы предприятия

При слежении за активностью системы и собраний можно просмотреть состояния различных процессов, отследить возможные ошибки и настроить параметры конфигурации для улучшения производительности.

С помощью консоли администратора поисковой системы предприятия можно следить за системой и при необходимости настроить ее работу. Можно просмотреть подробную статистику по каждому из основных процессов (таких как просмотр искателем, анализ, индексация, поиск). В статистику входит среднее время отклика и информация о ходе выполнения, например, число документов, просмотренных или проиндексированных за сеанс.

Щелкнув по соответствующим значкам, можно остановить и запустить большинство процессов. Эти операции позволяют приостановить процесс, внести изменения в его конфигурацию или исправить ошибку и перезапустить обработку, когда вы будете готовы ее продолжить.

### Задачи, связанные с данной

“Запуск поисковой системы предприятия” на стр. 297

“Остановка поисковой системы предприятия” на стр. 299

---

## Оценка числа документов в собрании

Создавая или изменяя собрание поисковой системы предприятия необходимо задавать предположительное число документов в этом собрании. Система использует это число для оценки необходимых ресурсов памяти и диска; это число не используется для ограничения размера собрания.

### Прежде чем приступать к работе...

Изменить оценку размера собрания может администратор поисковой системы предприятия или администратор этого собрания.

### Об этой задаче

Когда размер собрания достигнет оцененного вами значения, система не прекратит добавлять документы в индекс. Если вы сконфигурировали оповещения для собрания и включили опцию уведомления о том, что число документов в индексе превысило ограничение, по умолчанию это ограничение соответствует значению, заданному как оценка числа документов в собрании. Система следит за этой оценкой и заданным процентным порогом оповещения и будет отправлять сообщения электронной почты, когда число документов собрания приблизится к максимальному.

### Процедура

Чтобы задать оценку размера собрания:

1. Откройте собрание для редактирования, откройте страницу Анализ и нажмите кнопку **Конфигурировать общие опции**.
2. В поле **Примерное число документов** введите ожидаемое число документов в собрании. Значение по умолчанию - 1000000 документов.

---

## Слежение за собранием

Вы можете просмотреть общую информацию о состоянии каждого компонента в собрании или выбрать опции для просмотра подробной информации об отдельных компонентах и URI.

### Прежде чем приступить к работе...

Все пользователи-администраторы поисковой системы предприятия могут наблюдать за собраниями. Запускать или останавливать компоненты, а также включать и отключать расписания может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

### Процедура

Чтобы наблюдать за собранием:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**. Появится информация о текущем состоянии каждого из компонентов собрания.

**Совет:** Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Слежение**, чтобы перейти в окно слежения за собранием.

2. Чтобы просмотреть подробную информацию о каком-либо URI, выберите  **Подробности URI**.

Например, вам может понадобиться узнать, находится ли определенный URI в индексе или скопирован ли на поисковые серверы индекс, в котором находится URI.

3. Для наблюдения за отдельным компонентом и просмотра подробной статистической информации о работе этого компонента щелкните по значку **Состояние**.

---

## Просмотр подробностей об URI

Можно просматривать подробную информацию об URI. Можно просмотреть текущую информацию и хронологическую информацию о том, как выполняется просмотр, индексирование и поиск для документа, представленного этим URI.

### Прежде чем приступить к работе...

Прежде чем отправить требование на просмотр отчета об URI или послать отчет по адресу электронной почты, убедитесь, что компонент, от которого вы хотите получить информацию, активен. Например, чтобы увидеть подробности о просмотре, индексации и поиске документа, убедитесь, что искатель Web, сервер индексов и серверы поиска запущены. Чтобы отслеживать отброшенный документ, убедитесь, что сконфигурированы опции записи в журналы для отслеживания документов.

### Об этой задаче

Процесс сбора информации об URI отнимает много времени. Можно выбрать опцию просмотра затребованной информации и затем подождать ее вывода. Более эффективна опция отправки отчета по указанному вами электронному адресу.

Чтобы можно было получить отчет, нужно сконфигурировать в поисковой системе предприятия информацию о почтовом сервере. Эта информация задается при конфигурировании опций электронной почты на странице Журнал окна Система.

Сервер индексов и серверы поиска могут давать информацию обо всех URI (например, находится ли некоторый URI в индексе и был ли он скопирован на серверы поиска). Чтобы увидеть информацию о том, как просматривался документ, надо задать URI для документа, который просматривался искателем Web.

## Процедура

Чтобы просмотреть подробности об URI:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.

**Совет:** Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Слежение**, чтобы перейти в окно слежения за собранием.

2. Выберите  **Подробности URI**.
3. На странице Подробности URI введите URI, для которого нужно просмотреть информацию.
4. Включите переключатели для типов информации, которую вы хотите получить:

### **Подробности искателя (доступно только для искателей Web)**

Включите этот переключатель, чтобы увидеть информацию о том, как документ просматривался искателем Web, и информацию о его текущем состоянии в пространстве для искателя.

### **Подробности индекса**

Включите этот переключатель, чтобы увидеть, был ли документ индексирован и скопирован на серверы поиска.

### **Подробности поиска**

Включите этот переключатель, чтобы увидеть информацию о том, доступен ли документ для поиска и как можно выполнить такой поиск.

### **Документы, отброшенные синтаксическим анализатором**

Включите этот переключатель, чтобы увидеть, был ли документ отброшен из поисковой системы предприятия во время его синтаксического анализа, и, если это так, почему он был отброшен.

### **Документы, отброшенные из индекса**

Включите этот переключатель, чтобы увидеть, был ли документ отброшен из поисковой системы предприятия во время его индексирования и лексического анализа, и, если это так, почему он был отброшен.

5. Чтобы вывести на экран отчет, нажмите кнопку **Просмотреть отчет**.
6. Чтобы послать отчет по электронной почте для последующего просмотра, нажмите кнопку **Послать отчет**.
  - a. На странице Послать подробный отчет об URI введите электронный адрес для получения отчета в поле **Адрес электронной почты для уведомления**.
  - b. Нажмите кнопку **Послать отчет**.

### **Понятия, связанные с данным**

“Отслеживание документов” на стр. 323

### **Задачи, связанные с данной**

“Просмотр отчетов об отброшенных документах” на стр. 324

### **Ссылки, связанные с данной**

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

---

## Отслеживание работы искателей

Вы можете просматривать общую информацию о состоянии каждого искателя в собрании или выбрать опции для просмотра подробной информации о работе искателей.

### Прежде чем приступить к работе...

Если ваша административная роль разрешает вам только следить за собранием, вы можете просматривать статистику искателя, но не можете изменять поведение искателя (например, запускать и останавливать его).

### Процедура

Чтобы следить за искателем:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель работает или приостановлен, и вы хотите увидеть подробную информацию о состоянии этого искателя, выберите  **Подробности**. Типы выводимых при этом статистических показателей зависят от типа искателя.

Если ваша административная роль позволяет управлять процессами для собрания, вы можете запускать, останавливать и приостанавливать искатель при просмотре подробностей о его действиях. Если для искателя можно задать расписание, вы можете также включить или отключить расписание просмотра.

4. Если искатель остановлен или приостановлен, и вам нужно запустить сеанс искателя, выберите  **Запустить** или **Возобновить**.

#### Для искателей Web:

Если искатель был остановлен, он начнет просмотр снова по всему пространству для искателя. Если искатель был приостановлен, он возобновит просмотр, начиная с того назначения, где он был приостановлен.

Если вы хотите, чтобы искатель начал полный просмотр немедленно, щелкните по значку **Подробности**, затем по значку **Начать полный повторный просмотр**. Искатель начинает просмотр всего пространства искателя, включая страницы, которые не изменились со времени последнего просмотра. Повторный просмотр всех документов может потребоваться, например, после изменения правил синтаксического анализа документов, если нужно применить эти правила к проиндексированным ранее документам.

#### Для остальных типов искателей:

Если искатель был остановлен, он начнет просмотр в момент времени, предусмотренный расписанием. При первом просмотре источника данных искатель выполняет полный просмотр. При повторном запуске просмотра по расписанию искатель просматривает либо все изменения в источнике данных (добавления, удаления и изменения документов), либо только добавления и изменения документов. Тип просмотра конфигурируется в расписании искателя.

Если вы не задали расписания искателя или хотите запустить искатель раньше, щелкните по значку **Подробности**. Затем в области подробностей

пространства искателя щелкните по значку типа просмотра, который хотите запустить: полный просмотр, все изменения или только новые и измененные документы. Необходимо выбрать соответствующий значок запуска для каждого источника данных, который нужно просматривать (например, сервера, базы данных, подпапки).

5. Если искатель запущен и нужно остановить его, выберите  **Остановить** или  **Приостановить**. Искатель прекращает просмотр данных, пока вы его не перезапустите или не возобновите его работу.

Если вы возобновляете работу приостановленного искателя, просмотр начинается с того назначения, где он был приостановлен. Например, искатель DB2 возобновит просмотр с первой строки таблицы, которую просматривал в момент приостановки.

---

## Просмотр подробностей об активности искателя Web

Просматривая подробности об активности искателя Web, можно оценить общую производительность и нужным образом настроить свойства искателя Web и определения пространства для искателя.

### Прежде чем приступить к работе...

Все пользователи-администраторы могут следить за действиями искателя. Запустить или остановить искатель может администратор поисковой системы предприятия, администратор собрания или оператор этого собрания.

### Процедура

Чтобы просмотреть подробности о работе искателя Web:

1. В окне Собрания найдите собрание, владеющее искателем Web, за которым вы хотите следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель Web, который нужно отслеживать, запущен или приостановлен, выберите  **Подробности**.
4. На странице подробностей для искателя Web просмотрите или выберите указанные ниже опции, чтобы увидеть подробную статистику о текущих и прошлых действиях искателя.
  - Нажмите кнопку **Подробности потока**, чтобы увидеть, сколько потоков активно просматривают сайты, и сколько потоков неактивно.
  - Нажмите кнопку **Активные сайты**, чтобы увидеть информацию о сайтах, в данный момент просматриваемых искателем.
  - Нажмите кнопку **Недавно просмотренные URL**. Эта информация показывает последние просмотренные искателем объекты. Если элементы в этом списке не изменяются после обновления окна, искатель не выполняет просмотр.
  - Нажмите кнопку **Хронология искателя**, чтобы посмотреть отчеты о предыдущих действиях искателя.
  - В области **Состояние URL** введите URL, о котором хотите увидеть информацию.
    - a. Нажмите кнопку **Подробности URL**, чтобы увидеть информацию о его состоянии. Подробности URL можно затребовать только для тех URL, которые уже просматривались искателем.

- b. Нажмите кнопку **Подробности сайта**, чтобы задать информацию для включения в отчет о сайте, которому принадлежит URL. Можно затребовать подробности сайта для ранее просмотренного искателем сайта или для сайта, который искатель еще не просматривал.

Например, с помощью этой опции можно узнать, находится ли данный URL в пространстве для искателя, был ли он просмотрен или только найден, когда его следует просматривать снова, а также информацию о последней попытке просмотра сайта. Кроме того, можно заказать вывод содержимого файла `robots.txt` для сайта - это бывает полезно при выяснении, почему искатель пропустил этот сайт.

## Подробности потока искателя Web

При наблюдении за работой искателя Web можно видеть число потоков, активно просматривающих сайты Web, и число потоков в состоянии неактивности.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать состояние потоков искателя. Наиболее часто выводятся следующие состояния:

### Ожидание

Означает, что у потока нет URL для просмотра. Это состояние возможно, если поток завершил просмотр URL, а искатель не успевает достаточно быстро найти другие URL. Например, если значение свойства искателя, управляющего его временем ожидания перед получением другой страницы с того же сайта, слишком велико, URL могут поступать с недостаточной скоростью.

### Выборка

Означает, что поток загружает страницу с сайта.

### Завершен

Означает, что поток отправляет остальному искателю просмотренные страницы, но еще не готов к просмотру нового URL.

### Приостановлен

Указывает на то, что искатель приостановлен.

В идеальном случае все потоки все время получают страницы. Если потоки часто находятся в состоянии "Завершен", возможно, у базы данных есть проблемы с пропускной способностью.

Если потоки часто находятся в состоянии ожидания, измените значение, заданное в поле **Максимальное число активных хостов** в свойствах искателя. Если это значение невелико, возможно, в пространстве для искателя недостаточно сайтов, чтобы занять все потоки, или недостаточно URL, пригодных для просмотра. Низкая активность может быть вызвана ошибками поиска DNS и ошибками поиска роботов.

## Активные сайты искателя Web

За работой искателя Web можно наблюдать, чтобы получать информацию о сайтах, которые в данный момент обрабатывает искатель.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать статистическую информацию об активных сайтах. В статистике показано:

- Сколько URL для просмотра искатель получил в память из внутренней базы данных в этот раз

- Сколько URL искатель попытался просмотреть на данный момент
- Сколько времени остается до деактивации сайта и его удаления из памяти для этой итерации искателя
- Сколько времени сайт уже находится в памяти

Эта информация постоянно изменяется по мере работы искателя по сконфигурированным для него правилам искателя. В идеале число активированных URL должно приближаться к значению, сконфигурированному в поле **Максимальное число активных хостов** в свойствах памяти искателя.

Если число активированных URL близко к нулю, искатель не находит нужных URL. Такая низкая активность может быть вызвана ошибками поиска DNS, проблемами со связью в сети, ошибками в базах данных и ошибками в определении пространства для искателя. Например:

- Если много сайтов долго находятся в памяти, но обработано мало URL, проверьте наличие проблем со связью в сети.
- Если в списке недостаточно сайтов, проверьте наличие ошибок в определении пространства для искателя или ошибок поиска DNS.
- Если искатель просматривает сайты с приемлемой скоростью, но оставляет в памяти много непросмотренных URL, отредактируйте свойства памяти искателя и исправьте значение срока ожидания в поле **Максимальное время хранения URL в памяти**, чтобы сайты дольше оставались в памяти.

## Скорость просмотра искателя Web

За работой искателя Web можно наблюдать, чтобы получать информацию о том, с какой скоростью искатель загружает страницы с сайтов.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать статистическую информацию о скорости просмотра данных искателем (скорости работы искателя). Кроме того, можно просматривать статистику о числе URL, просмотренных искателем с начала текущего сеанса.

Скорость работы искателя - это число страниц, просматриваемых искателем за секунду. Это число связано с несколькими свойствами, которые можно сконфигурировать для искателя Web:

- Число потоков искателя
- Число активных сайтов
- Период времени, в течение которого искатель должен ожидать перед получением следующей страницы с того же сервера Web.

Если у искателя есть по одному активному сайту на каждый поток искателя, и искатель должен делать паузу в две секунды перед получением очередной страницы с того же сервера Web, он не может просматривать больше одной страницы за две секунды для каждого потока. Например, если искатель использует число потоков по умолчанию (200), эти 200 потоков могут обрабатывать 100 страниц в секунду.

Если активных сайтов в два раза больше, чем потоков искателя, и искатель должен делать паузу в две секунды перед получением очередной страницы с того же сервера Web, он может просматривать одну страницу в секунду для каждого потока. Однако в этом случае ограничивающими факторами становятся скорости загрузки из сети и пропускная способность базы данных. Индикатором хорошей производительности

искателя является скорость работы искателя, соответствующая числу потоков искателя, активных сайтов и времени ожидания искателя.

Еще один параметр, на который следует обратить внимание при слежении за производительностью искателя Web - число URL, просмотренных искателем с начала текущего сеанса искателя. Разделите это число на общее время работы искателем, чтобы получить среднюю пропускную способность за длительное время. Если это число не растет, искатель либо завершил работу, либо не может ее продолжать. Например, ходу работы искателя могут помешать ошибки связи в сети, ошибки баз данных и неудачные попытки поиска DNS.

## Создание отчетов искателя Web

Просматривая отчеты о предыдущей активности искателя Web, можно оценить общую производительность и нужным образом настроить свойства искателя Web и определения пространства для искателя.

### Прежде чем приступать к работе...

Если ваша административная роль позволяет вам только следить за собраниями, вы можете просматривать статистику искателя и создавать отчеты о его активности, но не можете изменять поведение искателя (например, запускать и останавливать его).

### Об этой задаче

Информация об активности искателя Web может содержаться в разных типах отчетов. В некоторых типах отчетов информация возвращается сразу же после ее сбора во внутренней базе данных искателя. На создание отчета по сайту и отчета по кодам состояния HTTP требуется время. При создании отчетов этого типа можно указать для их получения электронный адрес, если вы не хотите ждать возвращения результатов на консоли администратора поисковой системы предприятия.

Информацию о том, как интерпретировать статистические показатели в отчетах, можно посмотреть, нажав кнопку **Справка** во время отслеживания искателя Web и создания отчетов.

### Процедура

Чтобы создать отчеты искателя Web:

1. В окне Собрания найдите собрание, владеющее искателем Web, за которым вы хотите следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель Web, для которого нужно создать отчет, работает или приостановлен, выберите  **Подробности**.
4. На странице подробностей для искателя Web выберите опцию для типа отчета, который хотите создать:
  - В области **Сводка состояния искателя** нажмите кнопку **Хронология искателя**, чтобы создать отчеты об искателе и всех найденных или просмотренных им сайтах.
  - В области **Состояние URL** задайте URL конкретного сайта, для которого нужно создать отчет, и нажмите кнопку **Подробности сайта**.

5. Для каждого из отчетов по хронологии искателя и сайту можно включить переключатель напротив каждого статистического показателя, который вы хотите видеть в отчете, и затем нажать кнопку **Просмотреть отчет**.

Для этих видов статистик искатель возвратит отчет на консоль администратора сразу после получения информации из своей внутренней базы данных.

6. Если вы создаете отчет о хронологии искателя, можно задать опции для создания отчета по сайту, затем нажать кнопку **Запустить отчет**.

Этот отчет включает при его создании выбранные вами статистики и сохраняется в указанном вами файле (имя файла должно быть абсолютным). Вы можете задать, чтобы созданный отчет пересылался вам по электронной почте.

7. Если вы создаете отчет о хронологии искателя, можно задать опции для создания отчета по кодам состояния HTTP, затем нажать кнопку **Запустить отчет**.

Этот отчет содержит информацию о распределении числа кодов состояния HTTP по сайтам. Он сохраняется в указанном вами файле (имя файла должно быть абсолютным). Вы можете задать, чтобы созданный отчет пересылался вам по электронной почте.

Этот отчет позволяет увидеть, например, какие сайты возвращают много кодов состояния 4xx (означающих, что страницы не найдены), кодов состояния 5xx (указывающих на проблемы на сервере), кодов состояния bxx (указывающих на проблемы со связью) и так далее.

Он наиболее полезен в случае искателей, уже находящихся некоторое время в активном состоянии (например, несколько недель). С его помощью можно выявлять исчезнувшие сайты, вновь появившиеся сайты, сайты со слишком большим количеством URL (что может означать избыточный поиск в базе данных Lotus Notes) и сайты с рекурсивной файловой системой, обслуживаемые сервером HTTP. Если сайты с большим количеством кодов состояния HTTP не попадают в индекс, производительность искателя можно улучшить, удалив эти сайты из пространства для искателя.

## Код состояния HTTP, возвращаемые искателем Web

При наблюдении за искателем Web можно просматривать информацию о кодах состояния HTTP, возвращаемых искателю со страниц, которые он пытается просмотреть.

### Сводка таблицы

При наблюдении хронологии искателя Web или состояния определенного адреса URL выводится информация о кодах состояния HTTP, возвращенных искателю. Эту информацию можно использовать при управлении пространством для искателя и при оптимизации производительности искателя. Например, если искатель получает для URL множество кодов состояния HTTP, которые говорят о том, что заданные страницы невозможно просмотреть, можно улучшить производительность, удалив этот URL из пространства искателя.

В следующей таблице перечислены коды состояния HTTP и то, как их обрабатывает искатель Web. Значения от 100 до 505 - стандартные коды состояния HTTP (дополнительную информацию смотрите по адресу <http://www.w3.org/Protocols/rfc2616/rfc2616.html>). Остальные коды состояния HTTP - собственные коды возврата HTTP поисковой системы предприятия и искателя Web.

Таблица 8. Код состояния HTTP от искателя Web

Код	Описание	Код	Описание	Код	Описание	Код	Описание
NULL	Не просмотрен	400	Неверное требование	500	Внутренняя ошибка сервера	693	Ошибка выбора (URLFetcher)
100	Продолжить	401	Не авторизован	501	Не реализован	694	Ошибка записи (URLFetcher)
101	Переключение протоколов	402	Требуется оплата	502	Неверный шлюз	695	Неполный блок заголовка (URLFetcher)
200	Успешное завершение	403	Запрещен	503	Служба недоступна	699	Неожиданная ошибка (URLFetcher)
201	Создан	404	Не найден	504	Истек срок ожидания шлюза	700	Ошибка анализа (нет конца заголовка)
202	Принят	405	Неразрешенный метод	505	Версия HTTP не поддерживается	710	Ошибка анализа (заголовок)
203	Ненадежная информация	406	Неприемлем	611	Ошибка чтения	720	Ошибка анализа (нет кода HTTP)
204	Нет содержимого	407	Необходима аутентификация прокси	612	Ошибка соединения	730	Ошибка анализа (тело)
205	Сброс содержимого	408	Истек срок ожидания требования	613	Истек срок ожидания при чтении	740 или 4044	Исключен файлом robots.txt
206	Частичное содержимое	409	Конфликт	614	Ошибка при установлении соединения SSL	741	Роботы временно недоступны
300	Несколько вариантов выбора	410	Исчез	615	Другая ошибка чтения	760	Исключен определением пространства для искателя
301	Перемещен постоянно	411	Необходима длина	616	Ненормальный FBA	761	Недопустим для локального пространства для искателя; допустим для глобального
302	Найден	412	Не выполнено предварительное условие	617	Ошибка кодировки	770	Неверный протокол или нестандартный порт системы
303	Смотрите другие	413	Слишком большой объект требования	618	Перенаправление без URL перенаправления	780	Исключен по типу файла
304	Не изменен	414	Слишком длинный URI требования	680	Ошибка поиска DNS	786	Недопустимый URL
305	Используйте прокси	415	Неподдерж. тип мультимедиа	690	Неверный формат URL	2004	Нет тега META для индекса

Таблица 8. Код состояния HTTP от искателя Web (продолжение)

Код	Описание	Код	Описание	Код	Описание	Код	Описание
306	(Не используется)	416	Диапазон в требовании превышает объем ресурса	691	Соединение утеряно (URLFetcher)	3020	Программное перенаправление
307	Временное перенаправление	417	Ожидания не оправдались	692	Истек срок ожидания записи (URLFetcher)		

## Примечания к таблице

### Коды состояния 4xx

Код 400 (неверное требование) встречается редко. В соответствии со стандартом кодов состояния HTTP коды 4xx означают, что ошибка произошла на стороне клиента (то есть искателя). Однако обычно проблема заключается в сервере или в URL, полученном искателем в качестве ссылки. Например, некоторые серверы Web не принимают URL, пытающиеся перейти на верхний уровень по отношению к корневому каталогу сайта (например, <http://xyz.ibm.com/../../sales>). Другие серверы Web нормально обрабатывают такие URL, игнорируя указание родительского каталога, если искатель уже находится в корневом каталоге.

Некоторые серверы воспринимают требования корневого каталога сайта как ошибочные, и некоторые устаревшие ссылки могут требовать выполнения операций, которые уже не распознаются или не поддерживаются. При запросе страницы, которая больше не обслуживается, сервер прикладных программ генерирует исключительную ситуацию, и Web-сервер возвращает код состояния HTTP 400, так как требование уже не считается допустимым.

**615** Означает, что сервер искателя, загружающий данные с сайтов, неожиданно столкнулся с исключительной ситуацией. Большое число кодов состояния этого типа может означать, что возникла проблема с искателем.

### Коды состояния 61x

За исключением 615, коды состояния 61x сообщают о возможных проблемах при просмотре сайтов, например, об истечении срока ожидания. Для следующих кодов состояния могут потребоваться действия по исправлению:

#### 611, 612 и 613

Причиной этих ошибок могут быть медленные сайты или низкая производительность сети.

- 611** Ошибка при получении документа искателем.
- 612** Ошибка при попытке искателя соединиться с Web-сервером.
- 613** Истечение срока ожидания при получении документа искателем.

**614** Означает, что искатель не может просматривать защищенные сайты (HTTPS). Если вы считаете, что эти сайты должны быть доступны, убедитесь, что на сервере искателя и на Web-сервере назначения установлены правильные сертификаты. Например, если сайт сертифицирован известными сертифицированными (certificate authorities, CA), можно добавить новые CA на склад доверенных сертификатов, используемый искателем.

Кроме того, проверьте, как сконфигурированы самоподписанные сертификаты на сайтах, которые пытается просматривать искатель. Искатель сконфигурирован на принятие самоподписанных сертификатов. Некоторые сайты создают самоподписанный сертификат для корневого URL (например, <http://sales.ibm.com/>), а затем пытаются использовать этот сертификат на поддоменах (например, <http://internal.sales.ibm.com/>). Искатель не принимает сертификаты, используемые таким образом. Он принимает самоподписанные сертификаты, только если имя домена субъекта ([sales.ibm.com](http://sales.ibm.com/)) и лица, подписавшего сертификат, совпадает с именем домена требуемой страницы.

- 616** Указывает, что в загрузке после переаутентификации по-прежнему есть форма регистрации для аутентификации на основе форм (form-based authentication, FBA).

Если информация из файла конфигурации FBA (форма регистрации плюс данные аутентификации - имя пользователя, пароль и тому подобное) не обеспечит успешной аутентификации искателя, код состояния 616 присваивается всем страницам, зависящим от аутентификации на основе форм. Администратор должен будет выяснить, по какой причине конфигурация FBA не сработала.

- 617** Указывает на невозможность создания строки из байтового содержания документа из-за неверной кодировки (набора символов) или наличия в документе неверных байтов.

- 618** Указывает, что был обнаружен недопустимый URL перенаправления, когда искатель получил приведенные ниже коды состояния HTTP. Возможно, у заголовка ответа HTTP недопустимое положение.
- 301 Moved Permanently (Перемещен постоянно)
  - 302 Found (Найден)

- 680** Означает, что искатель не смог получить IP-адреса для хостов в пространстве искателя, возможно, из-за проблем с доступом в сеть. Эта ошибка означает, что искатель не смог обработать целые сайты, а не только некоторые URL. Большое число кодов состояния этого типа может значительно снизить пропускную способность.

#### **Коды состояния 69x**

Коды состояния с 690 по 699 не записываются в постоянную базу данных искателя. Эти коды представляют итоги, отражающие не истинный итог загрузки с удаленного хоста, а временную ситуацию внутри искателя, например, когда один из компонентов был закрыт, пока другой ожидал результата или отправки результата. Эти коды состояния появляются в некоторых журналах, но не в постоянных записях, и поэтому их не следует использовать как значения набора выбора.

#### **Коды состояния 7xx**

Коды 7xx преимущественно связаны с правилами пространства для искателя:

##### **710 - 730**

Означает, что из-за проблем искатель не смог выполнить загрузку полностью или обнаружил на сайте недопустимые данные HTML. Если выводится большое число кодов состояния этого типа, обратитесь за помощью к представителю службы поддержки поисковой системы предприятия.

#### **740 или 4044**

Означает, что содержимое файла нельзя индексировать, так как этот документ исключен в соответствии с ограничениями файла `robots.txt`.

**740** Означает, что в индекс можно включать ссылки привязки, указывающие на исключенный документ.

**4044** Означает, что ссылки привязки в документах, указывающие на исключенный документ, также исключаются из индекса.

**741** Означает, что на сайте есть файл `robots.txt`, разрешающий просмотр, но его загрузка не удалась. Если искатель несколько раз не может просмотреть URL, этот URL удаляется из пространства искателя. Если коды состояния этого типа попадают часто, проверьте, временно или постоянно недоступен сайт назначения. Если сайт назначения недоступен, удалите его из пространства искателя.

Остальные коды состояния 7xx обычно возвращаются в случае внесения изменений в пространство искателя, когда искатель уже проработал некоторое время. Обычно эти коды состояния не свидетельствуют о проблемах, требующих вашего внимания.

**3020** Указывает, что документ с кодом состояния 200 содержит заголовок положения, который перенаправляет пользовательский агент на другой URL.

---

## **Отслеживание работы анализатора**

Отслеживайте работу анализатора, когда хотите просмотреть информацию о документах, проанализированных анализатором, перед добавлением их в индекс поисковой системы предприятия. Вы можете просматривать статистические данные и управлять активностью анализатора.

### **Прежде чем приступить к работе...**

Если ваша административная роль позволяет только следить за собраниями, вы можете просматривать состояние анализатора, но не можете запускать или останавливать анализатор.

### **Об этой задаче**

При отслеживании подробностей анализатора вы видите снимок активности, содержащий статистическую информацию об активности анализатора в конкретный момент времени. Эта статистическая информация показывает число просмотренных, просматриваемых или ожидающих просмотра документов и число документов, которые проанализированы и ожидают сохранения в индексе.

Когда анализатор активен, сообщения дают дополнительную информацию о состоянии анализатора. Например:

- Синтаксический анализатор, возможно, активно работает с документами.
- Синтаксический анализатор, возможно, бездействует. Анализатор бездействует, пока нет новых документов для анализа. Если возникли ошибки, анализатор ожидает перезапуска. Анализатор сам перезапускает себя, если нет доступных служб синтаксического анализа (например, автоматический перезапуск происходит, если не удалось установить соединение со службой анализа или если все виртуальные Java-машины анализатора заняты работой с другими собраниями).

- Анализатор, возможно, приостановлен (например, до завершения построения индекса).

### Процедура

Чтобы следить за анализатором для собрания:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Анализатор.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Анализатор, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.

3. Если анализатор работает, и вы хотите просмотреть подробную информацию о его работе, нажмите кнопку  **Подробности**.

Если ваша административная роль позволяет управлять процессами для собрания, можно запускать и останавливать анализатор при просмотре подробностей о его активности.

4. Если анализатор остановлен и нужно запустить его, выберите  **Запустить**.

При первом создании собрания запускайте анализатор только после того, как искатель начнет просмотр данных. Это дает анализатору данные для анализа и категоризации. Если вы не вносите изменения в правила анализа, можно позволить анализатору работать непрерывно.

5. Если анализатор запущен и нужно остановить его, выберите  **Остановить**.

Если в правила анализа внесены изменения, анализатор нужно остановить и перезапустить. Например, если вы изменили конфигурацию анализатора, нужно остановить и перезапустить анализатор, чтобы эти изменения вступили в силу.

---

## Отслеживание активности индекса для собрания

Слежение за индексом для собрания может понадобиться, если вам нужно следить за ходом построения индекса, включить или отключить расписание индекса или же запустить или остановить индексацию.

### Прежде чем приступать к работе...

Все пользователи-администраторы могут следить за активностью индекса. Запускать или останавливать построение индекса, а также включать и отключать расписание индексирования может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

### Процедура

Чтобы следить за индексом для собрания:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Индекс.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Индекс, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.

3. Если индекс включен в расписание, но вы хотите не начинать построение в запланированный момент времени, выберите  **Отключить расписание**. Построение индекса не начнется, пока вы не включите расписание или не запустите процесс индексирования.

4. Если индекс внесен в расписание, но расписание для построения отключено, нажмите кнопку  **Включить расписание**. Построение индекса будет поставлено в очередь в дату и время, которые вы задали в расписании индексации.
5. Если индексация остановлена и нужно запустить ее, выберите  **Запустить**.  
Обычно индексация выполняется регулярно по расписанию. Если вы остановили индексацию во время процесса построения или отключили расписание индексации, нажмите кнопку **Запустить**, чтобы принудительно возобновить построение индекса.
6. Если построение индекса выполняется и нужно остановить его, выберите  **Остановить**.  
Остановить дельта-построение индекса может понадобиться, например, для принудительного полного построения индекса после изменения используемого в собрании типа категоризации.
7. Если при построении индекса были ошибки, щелкните по  **Ошибка**.  
Откроется страница Содержимое файла журнала, где можно просмотреть дополнительную информацию об ошибках индексации. На этой странице можно выбрать отдельные сообщения об ошибках и посмотреть подробности, относящиеся к ошибке.

---

## Слежение за очередью индексации поисковой системы предприятия

Вы можете просматривать состояние всех процессов индексации в очереди индексации, останавливать построение индекса, а также удалять индекс из очереди.

### Прежде чем приступать к работе...

Управлять очередью индексации может пользователь с ролью администратора поисковой системы предприятия.

### Об этой задаче

Одновременно можно создавать несколько индексов, но только один индекс для собрания может находиться в очереди в каждый момент времени. При настройке опций индексации для системы вы задаете, сколько индексов может параллельно использовать одну очередь и общие ресурсы индексации.

### Процедура

Чтобы следить за очередью индексации:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите страницу Индекс.  
Появится список собраний, у которых есть индексы в очереди индексации. Для каждого создаваемого индекса можно посмотреть его тип (полный или дельта), время помещения в очередь индексации и время начала построения индекса (если процесс еще выполняется).
3. Чтобы управлять отдельным индексом, щелкните по значку **Состояние**.  
Например, вам может понадобиться узнать, насколько индексация близка к завершению, посмотреть, сколько документов содержится в индексе, или отключить расписание индексации.
4. Чтобы остановить процесс построения индекса, выберите  **Остановить**.

Например, если вы изменили правила категорий, может понадобиться остановить дельта-построение индекса, чтобы иметь возможность вместо этого принудительно начать полное построение индекса.

Чтобы возобновить остановленное построение индекса, дождитесь помещения этого индекса в очередь в очередное запланированное время запуска или щелкните по значку **Состояние**, чтобы следить за этим индексом, а затем выберите  **Запустить**, чтобы запустить построение индекса.

5. Чтобы удалить индекс из очереди, выберите  **Удалить**.

---

## Отслеживание работы серверов поиска

Вы можете просматривать подробную информацию о состоянии, касающуюся активности сервера поиска, для конкретного собрания или же подробную информацию о состоянии для серверов поиска по всей поисковой системе вашего предприятия.

### Прежде чем приступить к работе...

Все пользователи-администраторы поисковой системы предприятия могут следить на поисковых серверах за собраниями, на управление которыми у них есть полномочия. Для слежения за всеми серверами поиска в поисковой системе предприятия вы должны быть администратором поисковой системы предприятия.

Запустить или остановить сервер поиска может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

### Процедура

1. Чтобы следить за серверами поиска для одного собрания:
  - a. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
  - b. Откройте страницу Поиск.

**Совет:** Если при редактировании собрания вы уже находитесь на странице Поиск, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.
2. Для слежения за всеми серверами поиска в поисковой системе предприятия:
  - a. Выберите **Система**, чтобы открыть окно Система.
  - b. Выберите страницу Поиск.
3. Если сервер поиска остановлен и нужно запустить его, выберите  **Запустить**.
4. Если сервер поиска запущен и нужно остановить его, выберите  **Остановить**.

Если вы включили или отключили кэш поиска, внесли изменения в размер кэша поиска или в прямые ссылки, нужно остановить и перезапустить серверы поиска, чтобы эти изменения вступили в силу.
5. Чтобы посмотреть сводку о длительности обработки сервером поиска поисковых требований, нажмите кнопку **Время ответа хронология**.

Этот отчет содержит среднее время в миллисекундах, затраченное сервером поиска на ответы на поисковые требования в конкретный день.

Среднее время ответа - это показатель производительности системы, соответствующий качеству обслуживания. Увеличение времени ответа может указывать на перегрузку системы. Например, в системе может быть превышено число собраний, в которых выполняется поиск, а также размер этих собраний.
6. Чтобы посмотреть список часто выполняемых запросов, нажмите кнопку **Частые запросы**.

Этот отчет содержит ключевые слова 50 наиболее частых запросов с указанием частоты каждого запроса.

Просматривая самые частые запросы, можно определить, для каких из них можно задать прямые ссылки. Создание прямых ссылок может повышать качество поиска для многих пользователей. Это позволяет всегда возвращать в результатах поиска документы с хорошим соответствием его условиям.

Можно также создать ссылки на ресурсы, где эти запросы могут выполняться с портала предприятия. Например, если пользователям часто требуется информация о счетах подотчетных сумм предприятия, добавьте ссылку на страницу с обсуждением процедур составления таких счетов на домашнюю страницу внутренней сети предприятия.

#### 7. Чтобы посмотреть список самых последних запросов, нажмите кнопку **Последние запросы**.

Этот отчет содержит ключевые слова 50 последних выполненных запросов.

Просматривая последние запросы, можно выявлять текущие тенденции в вашей организации, а также ситуации, требующие срочных действий. Например, можно обнаружить всплеск интереса к определенной теме. Эта волна интереса может указывать на необходимость создать прямую ссылку на эту тему или другими способами сделать ее доступной пользователям (например, с помощью ссылки на портале предприятия).

## Изменение способа подсчета статистики

Можно изменить способ подсчета числа популярных запросов и текущих запросов.

### Об этой задаче

При слежении за серверами поиска можно выбрать опции просмотра списка 50 самых популярных запросов и списка 50 последних обработанных запросов. В конфигурации сервера поиска по умолчанию запросы с эквивалентными строками запроса, но разными параметрами ранжирования результатов считаются различными запросами. Это значит, например, что счетчик запросов увеличивается, когда пользователь щелкает по опции просмотра следующей страницы результатов.

Можно изменить способ подсчета статистики запросов в системе, отредактировав файл `runtime-generic.properties` для сервера поиска. Если задать для параметра **distinctRecentQueryCheck** значение `true`, система будет считать отдельным запросом только тот запрос, который вернул начальную страницу результатов.

Чтобы изменить свойства сервера поиска, нужно зарегистрироваться в качестве администратора поисковой системы предприятия. Запустить или остановить сервер поиска может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

### Процедура

Чтобы изменить способ подсчета статистики запросов:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition. В конфигурации с несколькими серверами зарегистрируйтесь на сервере поиска.
2. Отредактируйте следующий файл, где *ID\_собрания* - ID собрания, которое вы хотите изменить, а *ID\_узла* - сервер поиска, на котором вы вносите изменения:  
`КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config/ID_собрания.runtime.ID_узла/runtime-generic.properties`

3. Задайте для параметра **distinctRecentQueryCheck** значение true, сохраните этот файл и закройте его.
4. В конфигурации с несколькими серверами повторите эти действия на втором сервере поиска.
5. Если вы хотите изменить способ подсчета статистики запросов для других собраний, повторите описанные выше действия для каждого собрания, которое вы хотите изменить.
6. Чтобы изменения вступили в силу, зарегистрируйтесь на консоли администратора поисковой системы предприятия и перезапустите серверы поиска:
  - a. Выберите **Система**, чтобы открыть окно Система.
  - b. Выберите страницу Поиск.
  - c. Для каждого измененного вами сервера поиска выберите  **Остановить**.
  - d. Для каждого измененного вами сервера поиска выберите  **Запустить**.

---

## Слежение за работой системы приема данных

Слежение за работой системы приема данных выполняется с целью определения ее состояния и просмотра подробностей работы клиентской программы приема данных.

### Прежде чем приступить к работе...

Чтобы следить за системой приема данных, необходимо быть администратором поисковой системы предприятия.

**Важное замечание:** Система приема данных в будущих выпусках не будет поддерживаться. Используйте при разработке клиентских программ для поисковой системы предприятия вместо API системы приема данных API поиска и индексирования (SI-API). Информация ниже приводится для тех пользователей, кто ранее создавал программы системы приема данных.

### Процедура

Чтобы следить за системой приема данных:

1. Выберите **Система**, чтобы открыть окно Система.
2. На странице Система приема данных просмотрите значки состояния, чтобы определить, работает система приема данных или остановлена.
3. Если система приема данных работает и вы хотите просмотреть подробную информацию о работе клиентской программы, нажмите кнопку  **Подробности**.  
Значки состояния на странице Подробности системы приема данных показывают, работает система приема данных или остановлена. Статистические параметры показывают, сколько требований ожидают обработки, текущее состояние каждого из потоков, обрабатывающих требования клиентской программы, и сколько потоков активно для данного состояния потоков.
4. Если вы изменили номер порта для системы приема данных или максимальное число документов, которые можно хранить во временном хранилище, нажмите кнопку  **Перезапустить**.

Система приема данных запускается при запуске поисковой системы предприятия. Нет необходимости перезапускать систему приема данных, если вы не изменили ни одного из этих параметров конфигурации.

---

## Отслеживание документов

Документы можно отбрасывать из системы на разных стадиях обработки. Вы можете задать опции уведомлений об отбрасывании документов и его причинах.

Если синтаксический анализатор обнаруживает ошибку, препятствующую анализу документа, в журнал записывается сообщение об отброшенном документе с кодом причины. (Ошибки этого типа не приводят к удалению из индекса старых версий документа.)

Документы могут отбрасываться на стадиях индексации, и информация об этом также записывается в журнал. Например, возможно явное удаление шаблонов URI и URI. Документ мог быть просмотрен искателем, который затем был удален. Возможно, что документ источника больше не существует (с документом связан отрицательный код HTTP), или с ним связан неизвестный код HTTP. Кроме того, документы могут быть отброшены, если отсутствует информация о ранге для документа, для которого требуется глобальный анализ.

Если известно, что документ был просмотрен искателем, но не появился в индексе, проследить путь обработки документа в системе можно при помощи консоли администратора поисковой системы предприятия. Подробные отчеты могут рассказать, когда, где и почему документ был отброшен. Например, в отчете может указываться, что документ был неожиданно отброшен при глобальном анализе, или что администратор удалил URI из индекса.

### Задачи, связанные с данной

“Просмотр подробностей об URI” на стр. 306

## Конфигурирование файлов журнала для отслеживания документов

Чтобы определять, когда, где и почему тот или иной документ был отброшен из системы, можно сконфигурировать запись в файлы журналов информации об отбрасывании документов.

### Прежде чем приступить к работе...

Конфигурировать опции для отслеживания отброшенных документов может только администратор поисковой системы предприятия или администратор собрания.

### Об этой задаче

Чтобы файлы журналов не занимали слишком много дискового пространства, система всегда переходит к новому файлу журнала при изменении текущей даты. Если дата не изменилась, но размер файл журнала достиг максимального разрешенного размера, система создает новый файл журнала. Когда достигнуто максимальное число файлов журнала, самый старый файл журнала отбрасывается, чтобы можно было создать новый файл журнала.

### Процедура

Чтобы сконфигурировать файлы журнала для отслеживания документов:

1. Вызовите редактирование собрания, выберите страницу Журнал и щелкните по **Конфигурировать отслеживание документов**.
2. На странице Отслеживание документов убедитесь, что включен переключатель отслеживания документов.

3. Задайте число файлов журнала для записи информации о документах, отброшенных из системы. Эти файлы журналов совместно используются всеми сеансами, в которых возможно отбрасывание документов.

## Просмотр отчетов об отброшенных документах

Можно просматривать подробную информацию о документах, отброшенных из поисковой системы предприятия. Эта информация доступна только в том случае, если включено отслеживание документов для этого собрания.

### Прежде чем приступать к работе...

Прежде чем отправить требование на просмотр отчета об отброшенных документах или послать отчет по адресу электронной почты, убедитесь, что сеансы, от которых вы хотите получить информацию, активны. Например, чтобы узнать о документах, отброшенных при синтаксическом анализе или индексации, убедитесь, что запущены сеансы анализатора или индексации для этого собрания.

Чтобы можно было получить отчет, сконфигурируйте в поисковой системе предприятия информацию о почтовом сервере. Эта информация задается при конфигурировании опций электронной почты на странице Журнал окна Система.

### Об этой задаче

Процесс сбора информации об отброшенных документах отнимает много времени. Можно выбрать опцию просмотра информации и затем подождать ее вывода. Более эффективна опция отправки отчета по указанному вами электронному адресу.

Если документ был отброшен, в отчете указывается дата и время, когда был отброшен документ, уровень серьезности ошибки, компонент и сеанс, в которых возникла ошибка, и сообщение об ошибке.

### Процедура

Чтобы просмотреть подробности об отброшенных документах:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.

**Совет:** Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Слежение**, чтобы перейти в окно слежения за собранием.

2. Выберите  **Подробности URI**.
3. На странице Подробности URI введите URI, для которого нужно просмотреть информацию.
4. Включите переключатели для типов информации, которую вы хотите получить:

#### Документы, отброшенные синтаксическим анализатором

Включите этот переключатель, чтобы видеть, был ли документ отброшен при синтаксическом анализе, а если был, то по какой причине.

#### Документы, отброшенные из индекса

Включите этот переключатель, чтобы видеть, был ли документ отброшен при индексации или анализе, а если был, то по какой причине.

5. Укажите, в каком виде хотите просматривать отчет:
  - Чтобы вывести на экран отчет, нажмите кнопку **Просмотреть отчет**.
  - Чтобы послать отчет по электронной почте для последующего просмотра, нажмите кнопку **Послать отчет**.

На странице **Послать** подробный отчет об URI введите электронный адрес для получения отчета в поле **Адрес электронной почты для уведомления** и включите **Послать отчет**.

#### **Задачи, связанные с данной**

“Просмотр подробностей об URI” на стр. 306

#### **Ссылки, связанные с данной**

“Форматы URI для индекса поисковой системы предприятия” на стр. 123

## **Просмотр файлов журналов отброшенных документах**

Можно просматривать сообщения в журнале о документах, отброшенных из поисковой системы предприятия. Эта информация доступна только в том случае, если включено отслеживание документов для этого собрания.

### **Об этой задаче**

Чтобы посмотреть отчет об отброшенном документе, нужно знать URI этого документа. Просматривая файлы журнала отброшенных документов, можно увидеть дату и время, когда был отброшен документ, уровень серьезности ошибки, компонент и сеанс, в которых возникла ошибка, и подробное сообщение об ошибке.

### **Процедура**

Чтобы просмотреть файлы журнала об отброшенных документах:

1. На странице **Собрания** найдите собрание, за которым нужно следить, и выберите  **Слежение**.

**Совет:** Если вы редактируете собрание и уже находитесь на странице **Общие**, можно выбрать  **Слежение**, чтобы перейти в окно слежения за собранием.

2. Выберите  **Файлы журнала отброшенных документов**.
3. На странице **Файлы журнала отброшенных документов** выберите файл журнала для просмотра. Имя каждого файла журнала указывает, был ли документ отброшен анализатором (pd) или при построении индекса (in), и дату создания файла. Если создается несколько журналов с одним типом и одной датой, последовательность создания в пределах этой даты указывается числовым суффиксом. Например:

```
dropped_doc_in_20060525.log
dropped_doc_pd_20060524.log (содержит более поздние значения для этой даты)
dropped_doc_pd_20060524.log.1
dropped_doc_pd_20060524.log.2 (содержит самые ранние значения для этой даты)
```

4. Нажмите кнопку **Просмотреть журнал**.  
Для каждого сообщения на странице **Содержимое** файлов журнала выводится дата и время выдачи сообщения, его уровень серьезности, имя сеанса, выдавшего сообщение, ID сообщения и текст ошибки.

При помощи кнопок можно переходить на первую, последнюю, предыдущую или следующую страницу файла журнала. Кроме того, можно задать номер страницы для перехода сразу на эту страницу.

5. Чтобы просмотреть более подробную информацию о каком-либо сообщении, выберите  **Подробности**.

На странице **Подробности** сообщения журнала выводится имя хоста сервера поисковой системы предприятия, на котором выдано сообщение, имя файла, сгенерировавшего ошибку, имя функции и номер строки, где произошла ошибка, ID процесса и ID потока.

Можно использовать кнопки перемещения к следующему и предыдущему сообщению в файле журнала.

---

## Файлы журналов и оповещения

Можно выбрать типы сообщений, записываемых в журнал для собрания и для системы, задать опции для создания и просмотра файлов журналов, получения оповещений и получения электронной почты об этих сообщениях.

В нормальном режиме компоненты поисковой системы предприятия записывают сообщения журнала в общий файл журнала. Этот файл журнала находится в каталоге `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/logs` на сервере индексов. Для просмотра данных этого общего журнала можно использовать консоль администратора.

В случае ошибки, например, неисправности сетевой связи, каждый компонент поисковой системы предприятия записывает сообщения журнала в каталог `logs` на сервере, где этот компонент установлен. Для просмотра файлов журнала используйте программу просмотра файлов на этом компьютере, например, утилиту `tail` в системе UNIX. Для просмотра этих типов файлов нельзя использовать консоль администратора.

При конфигурировании файлов журнала можно выбрать типы сообщений для записи в журнал (например, сообщения об ошибках или предупреждения), задать частоту отбрасывания старых файлов журнала для освобождения места под новые файлы журнала, задать максимальный размер файлов журнала и выбрать язык сообщений. Можно также задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов.

При слежении за файлами журнала можно выбрать, какой файл журнала открыть. Можно отфильтровать содержимое файла журнала, чтобы просмотреть сообщения только определенного уровня серьезности (например, только сообщения об ошибках) или сообщения, сгенерированные определенным сеансом поисковой системы предприятия. При просмотре файла журнала можно просмотреть подробности об отдельных сообщениях. Например, можно вывести имя функции, сгенерировавшей сообщение, и другую информацию, которая может быть необходима для устранения причины ошибки.

### Понятия, связанные с данным



Сообщения для поисковой системы предприятия

---

## Оповещения

Поисковую систему предприятия можно сконфигурировать так, чтобы при наступлении определенных событий записывать в файл журнала сообщения.

Сообщения, запускаемые событиями (их называют оповещениями), сообщают вам о ситуациях, которые требуют вмешательства, например, исчерпание ресурсом свободного пространства. При конфигурировании оповещений для поисковой системы предприятия вы задаете условия, за которыми должна следить система. Если система обнаруживает такое условие, она автоматически записывает в файл журнала сообщение.

Чтобы получать непосредственное уведомление о каком-либо условии, можно задать опции отправки сообщения по электронной почте при записи сообщения слежения в журнал.

Можно сконфигурировать оповещения для событий уровня собрания и для событий, происходящих на уровне системы. На уровне собрания система может:

- Отслеживать число документов, просматриваемых каждым искателем, и генерировать сообщение оповещения, что вскоре будет достигнуто максимально допустимое число документов.
- Отслеживать число документов, добавляемых в индекс для собраний, и генерировать сообщение оповещения, что вскоре будет достигнуто максимально допустимое число документов.
- Сообщать, что время, необходимое для ответов на требования поиска, превышает заданный вами предел.

На уровне системы система может отслеживать место на диске для каждого сервера поисковой системы предприятия и генерировать сообщение оповещения в случае малого объема свободного пространства.

## Конфигурирование оповещений уровня собрания

Конфигурирование оповещений позволяет записывать сообщения в файл журнала при возникновении определенных событий уровня собрания. Кроме того, можно включить отправку сообщений по электронной почте при записи в журнал сообщений об этих событиях.

### Прежде чем приступить к работе...

Конфигурировать оповещения может пользователь с ролью администратора поисковой системы предприятия или администратор собрания.

### Процедура

Чтобы сконфигурировать оповещения уровня собрания:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать оповещения**.
2. Чтобы система отслеживала число документов, просматриваемых каждым искателем, выполните следующие действия:
  - a. Включите переключатель **Когда число документов, просмотренных любым искателем, достигнет определенного процента от разрешенного максимума**.
  - b. В поле **Проценты** задайте, когда надо записывать сообщение в журнал. Это число задается в виде процентов от максимального числа документов, которые могут просмотреть искатели (**Максимальное число документов для просмотра** задается при конфигурировании свойств искателя). Значение по умолчанию - 90 процентов.

Так как для разных искателей можно настроить разные максимальные значения, для каждого искателя в журнал записываются отдельные сообщения. Например, если вы используете значение порога уведомления по умолчанию, искателю DB2 разрешено просмотреть 2000000 документов, а искателю Примечания разрешено просмотреть 1000000 документов; одно сообщение будет записано, когда искатель DB2 просмотрит 1800000 документов, а другое - когда искатель Примечания просмотрит 900000 документов.
3. Чтобы система отслеживала число документов, добавляемых в индекс, выполните следующие действия:
  - a. Включите переключатель **Когда число документов в собрании достигнет определенного процента от их предполагаемого числа**.

- b. В поле **Проценты** задайте, когда надо записывать сообщение в журнал. Это число задается в виде процентов от оценки числа документов, которые, как вы ожидаете, будут в этом собрании. Значение по умолчанию - 85 процентов.

В поле **Предел** выводится текущее значение предполагаемого размера собрания. Чтобы изменить это значение, откройте страницу **Общие** для этого собрания, выберите опцию для конфигурирования общих опций и задайте новое значение в поле **Предполагаемое число документов**

**Внимание:** Это предельное значение и предполагаемое число документов, заданные для собрания, используются только для отслеживания роста размера собрания. Они не задают абсолютного предельного значения размера индекса.

- 4. Чтобы система информировала вас в случае превышения предела времени ответа на требования поиска, выполните следующие действия:
  - a. Включите переключатель **Когда время ответа при поиске превысит предел**.
  - b. В поле **Предел** введите число секунд, которое вы считаете приемлемым в качестве максимального времени ответа при поиске.

Если это значение будет превышено, система запишет сообщение об этом событии в журнал. Например, если вы оставите значение по умолчанию, система будет записывать сообщение в журнал каждый раз, когда среднее время ответа сервера поиска на запросы поиска будет больше или равно пяти секундам.

Типичное время ответа не превышает полсекунды. Если средние значения превышают секунду, это может означать, что операционная система требует настройки для лучшей производительности, или что в настройках конфигурации сервера поиска что-то задано неправильно. Например, может быть нужно увеличить объем, выделяемый для кэша поиска.

- 5. Нажмите кнопку **ОК**.

Чтобы получать сообщение по электронной почте, когда система записывает в журнал сообщения об этих событиях, откройте страницу **Журнал** и выберите **Конфигурировать опции электронной почты для сообщений** и задайте адрес электронной почты. ID сообщений для включенных оповещений автоматически добавляются в список ID сообщений, для которых нужно отправлять сообщение по электронной почте.

Чтобы получать электронную почту, необходимо также сконфигурировать информацию о почтовом сервере. Для этого администратор поисковой системы предприятия должен выбрать на панели инструментов **Система**, открыть страницу **Журнал** и нажать кнопку **Конфигурировать опции электронной почты для сообщений**.

#### **Задачи, связанные с данной**

“Получение сообщений журнала по электронной почте” на стр. 332

## **Конфигурирование оповещений уровня системы**

Конфигурирование оповещений позволяет записывать сообщения в файл журнала при возникновении определенных событий уровня системы. Кроме того, можно включить отправку сообщений по электронной почте при записи в журнал сообщений об этих событиях.

### **Прежде чем приступать к работе...**

Конфигурировать оповещения уровня системы может только администратор поисковой системы предприятия.

### **Процедура**

Чтобы сконфигурировать оповещения уровня системы:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Журнал выберите **Конфигурировать оповещения**.
4. Чтобы система отслеживала объем свободного места на серверах поисковой системы предприятия, включите переключатель **Когда объем свободного пространства файловой системы достигнет определенного процента от общего объема**.
5. В поле **Проценты** задайте, когда система должна уведомлять вас о том, что на сервере осталось мало свободного места. Это число указывается в виде процентной доли от общего пространства файловой системы. Значение по умолчанию - 80 процентов.

Если поисковая система предприятия установлена на нескольких серверах, система создает для каждого сервера отдельный журнал сообщений. Например, одно сообщение информирует вас; когда становится мало пространство на сервере искателя, а другие отдельные сообщения - об ограничениях пространства на сервере индекса и поисковых серверах.

6. Нажмите кнопку **ОК**.

Чтобы получать сообщение по электронной почте каждый раз, когда система записывает в журнал сообщение об этом событии, откройте страницу Журнал и выберите **Конфигурировать опции электронной почты для сообщений**, чтобы задать адрес электронной почты и информацию о почтовом сервере.

#### **Задачи, связанные с данной**

“Получение сообщений журнала по электронной почте” на стр. 332

---

## **Конфигурирование файлов журнала**

Можно задать типы сообщений, которые надо записывать в журнал, и опции для создания файлов журнала.

### **Прежде чем приступить к работе...**

Сконфигурировать файлы журнала уровня собрания может администратор поисковой системы предприятия или администратор этого собрания. Сконфигурировать файлы журнала уровня системы может только администратор поисковой системы предприятия.

### **Об этой задаче**

Чтобы файлы журналов не занимали слишком много дискового пространства, система всегда переходит к новому файлу журнала при изменении текущей даты. Если дата не изменилась, но размер файл журнала достиг максимального разрешенного размера, система создает новый файл журнала. Когда достигнуто максимальное число файлов журнала, самый старый файл журнала отбрасывается, чтобы можно было создать новый файл журнала.

Чтобы получать электронную почту о событиях, записываемых в журнал, надо предварительно указать информацию о доставке электронной почты. Затем надо задать сообщения, для которых надо отправлять электронную почту.

### **Процедура**

Чтобы сконфигурировать файлы журнала поисковой системы предприятия:

1. Если надо сконфигурировать опции создания файлов журнала уровня системы и перехода к новым файлам журнала:
  - a. Выберите **Система**, чтобы открыть окно Система.
  - b. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
  - c. На странице Журнал выберите **Конфигурировать опции файлов журнала**. Откроется страница Опции файлов журнала уровня системы.
2. Если надо сконфигурировать опции создания файлов журнала уровня собрания и перехода к новым файлам журнала:
  - a. В окне Собрания найдите собрание, для которого нужно задать опции, и выберите  **Редактировать**.
  - b. На странице Журнал выберите **Конфигурировать опции файлов журнала**. Откроется страница Опции файлов журнала уровня собрания.
3. В поле **Тип информации для записи в журнал** выберите тип сообщений, которые надо записывать в журнал:

#### **Только сообщения об ошибках**

Сообщения об ошибках указывают, что возникла нежелательная ситуация или непредвиденное поведение и процесс не может быть продолжен. Вы должны исправить эту ошибку.

#### **Сообщения об ошибках и предупреждения**

Предупреждения указывают возможные конфликты или несовместимости, которые не приводят к остановке процесса. Это - опция по умолчанию.

#### **Все сообщения**

Информационные сообщения содержат общую информацию о системе и текущей задаче и не требуют никаких действий по исправлению.

**Важное замечание:** Выбор этой опции может снизить производительность системы. Записывайте в журнал все сообщения, только если это необходимо для диагностики ошибок или если вас просит сделать это служба программной поддержки IBM.

4. В поле **Максимальный размер каждого файла журнала** введите максимальный размер в мегабайтах для каждого файла журнала. Значение по умолчанию - 10 Мбайт.

Когда размер файла журнала достигает этого значения, создается новый файл журнала, пока не будет достигнуто разрешенное вами максимальное число файлов журнала. Сохраняя относительно небольшие файлы журнала, можно облегчить их просмотр.
5. В поле **Максимальное число файлов журнала** введите максимальное число создаваемых файлов журнала. Значение по умолчанию - 16.

Если вы хотите, чтобы можно было просматривать более старые сообщения журнала, увеличьте эти значения. Если вас больше интересуют последние сообщения и вам не нужно хранить длинную хронологию активности, уменьшите это значение.
6. В поле **Национальная версия по умолчанию** выберите язык для записи сообщений в журнал. Значение по умолчанию - Английский.
7. Нажмите кнопку **ОК**.
8. Чтобы изменения вступили в силу, введите следующие команды для остановки и перезапуска поисковой системы предприятия.

```
esdmin system stopall
esdmin system startall
```

## Конфигурирование информации о SMTP-сервере

Чтобы получать электронные письма о работе поисковой системы предприятия, необходимо сконфигурировать информацию о сервере почтового протокола Simple Mail Transfer Protocol (SMTP-сервере).

### Прежде чем приступать к работе...

Чтобы сконфигурировать информацию о SMTP-сервере, надо быть администратором поисковой системы предприятия.

### Об этой задаче

Получать электронную почту позволяют несколько функций администратора поисковой системы предприятия. Для получения электронной почты в рамках любой из этих функций необходимо указать сведения о вашем SMTP-сервере:

- Если вы сконфигурировали оповещения уровня собрания или уровня системы, можно получать сообщения по электронной почте при каждой записи этих сообщений в журнал. Можно также получать сообщения по электронной почте при записи других сообщений, (не только сообщений, генерируемых событиями, за которыми ведется слежение).
- Если вы хотите просмотреть подробную информацию об URI в индексе или о документе, отброшенном из поисковой системы предприятия, можно получить отчет по электронной почте.
- Если вы выполняете слежение за искателем Web и указали, что хотите создавать хронологические отчеты о работе этого искателя Web, можно получать уведомление по электронной почте после создания отчета.

### Процедура

Чтобы сконфигурировать информацию о SMTP-сервере:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
4. На странице Опции электронной почты для сообщений системы в поле **Почтовый SMTP-сервер для доставки электронной почты**, введите полное имя хоста или IP-адрес используемого SMTP-сервера.  
Система будет использовать этот сервер для отправки электронной почты по указанным вами адресам.
5. В поле **Частота проверки электронной почты** задайте, как часто система должна проверять наличие заданных сообщений и отправлять о них уведомления по электронной почте.  
Система с заданной частотой объединяет все сообщения на один адрес электронной почты в одно сообщение и отправляет это сообщение.
6. Нажмите кнопку **ОК**.

## Получение сообщений журнала по электронной почте

Можно задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов.

### Прежде чем приступать к работе...

Конфигурировать опции электронной почты для сообщений уровня системы может только администратор поисковой системы предприятия. Конфигурировать опции электронной почты для сообщений уровня собрания может только администратор поисковой системы предприятия или администратор этого собрания.

Чтобы получать электронную почту, необходимо задать информацию о почтовом сервере Simple Mail Transfer Protocol (SMTP-сервере).

### Об этой задаче

При конфигурировании оповещений можно выбрать опцию записи сообщений в журнал при определенных событиях. Если эти опции включены, можно сконфигурировать опции для автоматического получения электронной почты при записи в журнал этих сообщений. Кроме того, можно задать опции для получения электронной почты при записи других сообщений, а не только сообщений, вызванных событиями.

### Процедура

Чтобы сконфигурировать опции электронной почты для сообщений:

1. Если вы хотите получать оповещения по электронной почте о сообщениях системы:
  - a. Выберите **Система**, чтобы открыть окно Система.
  - b. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
  - c. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
  - d. На странице Опции электронной почты для сообщений системы включите переключатель **Посылать оповещения по электронной почте о сообщениях уровня системы**.
  - e. В поле **Адрес электронной почты для получения оповещений** введите один или несколько адресов электронной почты. Обычно информация о сообщениях системы должен получать администратор поисковой системы предприятия. Указывайте адреса через запятую. Например:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
  - f. Чтобы получать электронную почту обо всех записываемых в журнал сообщениях об ошибках, включите переключатель **Посылать оповещения обо всех сообщениях об ошибках**.
  - g. Если вы хотите получать оповещения по электронной почте только при записи в журнал определенных сообщений уровня системы, укажите ID этих сообщений в области **Посылать оповещения об определенных сообщениях**. Вводите по одному ID сообщения в строке. Например:  
FFQC4819E  
FFQ00005E  
Несколько ID сообщений входят в список по умолчанию (выберите **Справка**, чтобы посмотреть описания этих сообщений).
  - h. Нажмите кнопку **ОК**.
2. Если вы хотите получать оповещения по электронной почте о сообщениях собрания:
  - a. Выберите **Собрания**, чтобы открыть окно Собрания.
  - b. В списке собраний найдите собрание, которое нужно сконфигурировать, и выберите  **Редактировать**.

- c. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
- d. На странице Опции электронной почты для сообщений собрания включите переключатель **Посылать оповещения по электронной почте о сообщениях уровня собрания**.
- e. В поле **Адрес электронной почты для получения оповещений** введите один или несколько адресов электронной почты. Обычно информацию о сообщениях уровня собрания должен получать администратор собрания.  
Указывайте адреса через запятую. Например:  
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
- f. Чтобы получать электронную почту обо всех записываемых в журнал сообщениях об ошибках, включите переключатель **Посылать оповещения обо всех сообщениях об ошибках**.
- g. Если вы хотите получать оповещения по электронной почте только при записи в журнал определенных сообщений уровня собрания, укажите ID этих сообщений в области **Посылать оповещения об определенных сообщениях**. Вводите по одному ID сообщения в строке. Например:  
FFQC4819E  
FFQ00005E  
Несколько ID сообщений входят в список по умолчанию (выберите **Справка**, чтобы посмотреть описания этих сообщений).
- h. Нажмите кнопку **ОК**.

#### Понятия, связанные с данным



Сообщения для поисковой системы предприятия

#### Задачи, связанные с данной

“Конфигурирование оповещений уровня собрания” на стр. 328

“Конфигурирование оповещений уровня системы” на стр. 329

## Изменение размера журнала запросов

Увеличить или уменьшить размер файлов журналов, создаваемых для обработки запросов, можно путем редактирования файла конфигурации. С консоли администратора поиска эта задача не поддерживается.

### Об этой задаче

При обработки запросов данные журнала записываются в файл *ID\_собрания\_OmniFindQueryLog\_дата.log*, где *ID\_собрания* указывает собрание, которое вы хотите конфигурировать, а *дата* - дата создания файла журнала. До создания нового файла журнала можно увеличить или уменьшить размер этого файла в зависимости от того, сколько данных вы хотите в него записывать.

### Процедура

Чтобы изменить размер файлов журнала обработки запросов:

1. Зарегистрируйтесь как администратор поисковой системы предприятия. На установке поисковой системы предприятия с несколькими серверами зарегистрируйтесь на сервере индексов.
2. Откройте файл the КОРНЕВОЙ\_КАТАЛОГ\_УСТАНОВКИ\_ES/master\_config/*ID\_собрания.runtime.node1/runtime-generic.properties*.
3. Найдите свойство **MaxFileSize**. Увеличьте или уменьшите его значение, чтобы увеличить или уменьшить размер файлов журнала, и сохраните изменения.

4. Для поисковой системы предприятия с одним сервером:
  - a. При помощи консоли администратора поисковой системы предприятия задайте слежение за собранием, которое вы изменяете, и остановите сервер поиска.
  - b. Перезапустите программу ESSearchServer на WebSphere Application Server.
  - c. С консоли администратора перезапустите остановленные серверы поиска.
  - d. Откройте программу поиска в новом браузере.
5. Для поисковой системы предприятия с несколькими серверами:
  - a. Зарегистрируйтесь на сервере индексов как администратор поисковой системы предприятия.
  - b. Введите команды для перезапуска поисковой системы предприятия:

```
esadmin system stopall
esadmin system startall
```

---

## Просмотр файлов журнала

Можно просматривать сообщения журнала, которые компоненты системы и собрания записывают в общий файл журнала. Кроме того, можно задать фильтры, чтобы видеть только сообщения некоторого уровня серьезности и только от отдельных сеансов поисковой системы предприятия.

### Прежде чем приступить к работе...

Все пользователи-администраторы поисковой системы предприятия могут просматривать файлы журнала для собраний, которыми им разрешено управлять. Просматривать файлы журнала уровня системы может только администратор поисковой системы предприятия или пользователь, которому разрешен доступ к панели инструментов **Система**.

### Процедура

1. Чтобы просмотреть файлы журнала для одного собрания:
  - a. Выберите **Собрания**, чтобы открыть окно Собрания.
  - b. Найдите в списке собраний собрание, которое нужно просмотреть, выберите  **Слежение** и откройте страницу Журнал.

**Совет:** Если вы редактируете собрание и уже находитесь на странице Журнал, можно выбрать  **Наблюдатель**, чтобы перейти в окно слежения за собранием.

2. Для просмотра файлов журнала уровня системы:
  - a. Выберите **Система**, чтобы открыть окно Система.
  - b. Выберите страницу Журнал.
3. В поле **Файл журнала** выберите файл журнала для просмотра.

Имя каждого файла журнала состоит из типа файла журнала (например, имени системы или собрания) и даты создания файла. Если для одной даты создается несколько файлов журнала одного типа, порядок создания этих файлов указывается цифровым суффиксом. Например:

```
тип_файла_журнала_20060526.1log (содержит последние записи для данной даты)
тип_файла_журнала_20060526.1log.1
тип_файла_журнала_20060526.1log.2 (содержит самые первые записи для данной даты)
тип_файла_журнала_20060525.1log (содержит последние записи для данной даты)
```

тип\_файла\_журнала\_20060525.1ог.1  
тип\_файла\_журнала\_20060525.1ог.2  
тип\_файла\_журнала\_20060525.1ог.3 (содержит первые записи для данной даты)

4. Чтобы просматривать только сообщения определенных уровней серьезности, включите соответствующие переключатели в поле **Серьезность**.
5. Чтобы просматривать только сообщения от определенных сеансов, включите соответствующие переключатели в поле **Сеанс**.
6. Нажмите кнопку **Просмотреть журнал**.

Для каждого сообщения на странице Содержимое файлов журнала выводится дата и время выдачи сообщения, его уровень серьезности, имя сеанса, выдавшего сообщение, ID сообщения и текст ошибки.

При помощи кнопок можно переходить на первую, последнюю, предыдущую или следующую страницу файла журнала. Кроме того, можно задать номер страницы для перехода сразу на эту страницу.

7. Чтобы просмотреть более подробную информацию о каком-либо сообщении, выберите  **Подробности**.

На странице Подробности сообщения журнала выводится имя хоста сервера поисковой системы предприятия, на котором выдано сообщение, имя файла, сгенерировавшего ошибку, имя функции и номер строки, где произошла ошибка, ID процесса и ID потока.

Можно использовать кнопки перемещения к следующему и предыдущему сообщению в файле журнала.

---

## Резервное копирование и восстановление поисковой системы предприятия

Резервное копирование и восстановление поисковой системы предприятия можно выполнить при помощи сценариев резервного копирования и восстановления.

### Что копируют сценарии

Сценарии выполняют резервное копирование и восстановление следующих файлов:

- Файлов конфигурации из каталога `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config`
- Файлов баз данных для искателей, включая все метаданные искателей, такие как время последнего просмотра источников данных
- Всех файлов в каталоге `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/data`
- Файлов индексов для собраний, для которых сконфигурированы каталоги данных, отличные от каталога по умолчанию

### Структура каталогов резервного копирования

Сценарий резервного копирования создает в каталоге, указанном вами при его запуске, следующие подкаталоги. У ID администратора поисковой системы предприятия должно быть разрешение на запись в указанный вами каталог.

#### **master\_config**

Содержит файлы конфигурации из каталога `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/master_config`

#### **database**

Содержит файлы баз данных с сервера искателя

#### **data**

Содержит файлы индексов с сервера индексов

### Рекомендации по использованию

- Вы можете создать резервную копию данных с одного компьютера и восстановить ее на другом компьютере. Однако:
  - Файлы, резервная копия которых была создана из одной версии OmniFind Enterprise Edition, нельзя восстановить в системе, где используется другая версия OmniFind Enterprise Edition.
  - Данные нужно восстанавливать в системе с тем же или большим числом серверов поисковой системы предприятия. Например, если создается резервная копия поисковой системы предприятия, работающей на одном сервере, можно восстановить данные в системе, где используется два или четыре сервера поисковой системы предприятия. Данные, резервная копия которых была создана в четырехсерверной системе, нельзя восстановить в системе, где используется два или один сервер.
  - Файлы, резервная копия которых создавалась в одной операционной системе, нельзя восстановить для поисковой системы в другой операционной системе. Например, если вы установили поисковую систему предприятия в системе AIX и теперь хотите запускать в системе Linux, нужно установить новую поисковую систему предприятия на используемых серверах Linux.
- Чтобы была создана резервная копия наиболее свежих индексированных данных, перед резервным копированием следует выполнить полное построение индекса.

- В системе, для которой создается резервная копия, и системе, в которой восстанавливаются данные, все параметры для каталога установки (КОРНЕВОЙ\_КАТАЛОГ\_УСТАНОВКИ\_ES), каталога данных (КОРНЕВОЙ\_КАТАЛОГ\_УЗЛА\_ES), а также ID администратора и пароль поисковой системы предприятия должны быть одинаковы.
- Для многосерверной конфигурации резервное копирование и восстановление системы нужно выполнять с сервера индексов поисковой системы предприятия. Поскольку все данные искателя находятся на сервере искателя в базах данных, для резервного копирования и восстановления данных искателя в сценариях запускаются удаленные команды.
- Для резервного копирования файлов поисковой системы предприятия в другой каталог у вас должно быть достаточно доступного дискового пространства. Сценарии резервного копирования и восстановления не проверяют файлы.
- На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.
- Если в системе происходит сбой из-за неисправимой ошибки, необходимо переустановить OmniFind Enterprise Edition и затем запустить сценарий восстановления.

---

## Резервное копирование поисковой системы предприятия

Резервная копия поисковой системы предприятия создается при помощи сценариев `esbackup.sh` для AIX, Linux и Solaris или `esbackup.bat` для Microsoft Windows.

### Ограничения

У ID администратора поисковой системы предприятия должно быть разрешение на запись в каталог, задаваемый вами при запуске сценария резервного копирования.

На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.

**Внимание:** Если нажать `Ctrl+C`, чтобы прервать сценарий резервного копирования, система перейдет в несогласованное состояние. Чтобы запустить все сценарии служб и любые сценарии, запущенные, но остановленные в процессе резервного копирования, нужно ввести команду:

```
esadmin system startall
```

### Процедура

Чтобы создать резервную копию поисковой системы предприятия:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Если CCL (common communication layer - общий слой связи) для поисковой системы предприятия не запущен, запустите его:

**AIX, Linux или Solaris**  
`startccl.sh -bg`

### Окно командной строки Windows

```
startccl
```

### Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск** → **Программы** → **Инструменты управления** → **Службы**.
  - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск**.
3. В многосерверной конфигурации убедитесь, что CCL запущен на каждом сервере. При необходимости повторите предыдущие действия.
  4. Запустите резервное копирование, где *каталог\_резервной\_копии* - каталог для данных резервной копии:

### AIX, Linux или Solaris

```
esbackup.sh каталог_резервной_копии
```

### Окно командной строки Windows

```
esbackup.bat каталог_резервной_копии
```

### Ссылки, связанные с данной

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373

---

## Восстановление поисковой системы предприятия

После переустановки OmniFind Enterprise Edition можно с помощью сценария `esrestore.sh` для AIX, Linux и Solaris или сценария `esrestore.bat` для Microsoft Windows восстановить поисковую систему предприятия.

### Ограничения

На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.

Файлы, резервная копия которых была создана из одной версии OmniFind Enterprise Edition, нельзя восстановить в системе, где используется другая версия OmniFind Enterprise Edition. Кроме того, число поисковых серверов предприятия в системе, в которую восстанавливаются данные, должно быть не меньше, чем в системе, из которой выполнялось резервное копирование данных.

### Процедура

Чтобы восстановить поисковую систему предприятия:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
2. Если CCL (common communication layer - общий слой связи) для поисковой системы предприятия не запущен, запустите его:

### AIX, Linux или Solaris

```
startccl.sh -bg
```

### Окно командной строки Windows

```
startccl
```

### Администрирование - Службы Windows

Чтобы запустить CCL в фоновом режиме:

- a. Запустите службы Windows: **Пуск → Программы → Инструменты управления → Службы.**
  - b. Щелкните правой кнопкой мыши по **IBM OmniFind Enterprise Edition** и выберите **Пуск.**
3. В многосерверной конфигурации убедитесь, что CCL запущен на каждом сервере. При необходимости повторите предыдущие действия.
  4. Остановите контроллер:  
`esadmin stop`
  5. Восстановите данные поисковой системы предприятия, введя следующую команду (где *каталог\_резервной\_копии* - каталог созданной вами резервной копии файлов):

### AIX, Linux или Solaris

`esrestore.sh каталог_резервной_копии`

### Окно командной строки Windows

`esrestore.bat каталог_резервной_копии`

### Ссылки, связанные с данной

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373

---

## Экспорт и импорт конфигураций собраний

Можно экспортировать и импортировать отдельные конфигурации собраний. Экспортируются и импортируются только файлы конфигурации собрания, но не данные собрания.

Экспортировав собрание из одной поисковой системы предприятия, его можно после этого импортировать в другую поисковую систему предприятия. Однако экспортировать и импортировать собрания можно только между системами, в которых запускается одна и та же версия поисковой системы предприятия. Например, можно экспортировать собрание из системы версии 8.4, после чего импортировать его в систему версии 8.4 или версии 8.4.0.150, но экспорт собрания из системы версии 8.3 и импорт его в систему версии 8.4 невозможен.

После экспорта собрание можно импортировать только в поисковую систему предприятия, где используется та же операционная система. Например, нельзя экспортировать собрание из поисковой системы предприятия, установленной в Linux, и импортировать его в поисковую систему предприятия, установленную в Windows.

### Экспорт собраний

Для экспорта собрания используется команда **esadmin export**, экспортирующая одно собрание за один раз. С консоли администратора поисковой системы предприятия экспорт собраний не поддерживается.

1. Зарегистрируйтесь как администратор поисковой системы предприятия. При конфигурации с несколькими серверами нужно зарегистрироваться на сервере индексов. (При попытке запустить эту команду с любого другого сервера поисковой системы предприятия команда завершается неудачно.)
2. Введите следующие команды:

```
esadmin export -help
esadmin export -cid ID_собрания [-fname имя_файла_экспорта] [-verbose]
```

Где:

**-help**

Выводит для команды информацию справки.

**-cid** *ID\_собрания*

Задаёт ID экспортируемого собрания.

**Совет:** Определить ID собрания, которое вы хотите экспортировать, можно при помощи консоли администратора поисковой системы предприятия или команды **esadmin report collections**.

**-fname** *имя\_файла\_экспорта*

Задаёт путь к файлу экспорта. Если не указать абсолютное имя файла, подразумевается каталог `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/dump`. Если опустить эту опцию, в каталоге `КОРНЕВОЙ_КАТАЛОГ_УЗЛА_ES/dump` будет создан файл с соблюдением следующего соглашения об именовании:

```
_export_ггггММдд_ЧЧммссz.zip
```

где:

*ггггММдд*

- текущий год, месяц и день запуска команды экспорта.

*ЧЧммсс*

- текущий час, минута и секунда запуска команды экспорта.

*z* - разница со временем по Гринвичу для часового пояса на момент запуска команды экспорта. Например, зимнее московское время - это время по Гринвичу + 0300.

**-verbose**

Выводит информацию, помогающую устранить неисправности в операции экспорта.

## Импорт собраний

Для импорта собрания используется команда **esadmin import**, импортирующая одно собрание за один раз. С консоли администратора поисковой системы предприятия импорт собраний не поддерживается.

1. Зарегистрируйтесь как администратор поисковой системы предприятия. При конфигурации с несколькими серверами нужно зарегистрироваться на сервере индексов. (При попытке запустить эту команду с любого другого сервера поисковой системы предприятия команда завершается неудачно.)
2. Введите следующие команды:

```
esadmin import -help
esadmin import -fname имя_файла_импорта
 [-cid ID_нового_собрания]
 [-name имя_нового_собрания]
 [-colDataDir каталог_данных_нового_собрания]
 [-force]
 [-verbose]
```

Где:

**-help**

Выводит для команды информацию справки.

**-fname** *имя\_файла\_импорта*

Задает путь к файлу импорта. Если не указать абсолютное имя файла, подразумевается каталог КОРНЕВОЙ\_КАТАЛОГ\_УЗЛА\_ES/dump.

**-cid** *ID\_нового\_собрания*

Задает ID нового собрания, если собрание надо импортировать с другим ID.

**Совет:** Определить ID для собрания, которое вы хотите импортировать, можно с помощью консоли администратора поисковой системы предприятия или команды **esadmin report collections**.

**-name** *имя\_нового\_собрания*

Задает имя нового собрания, если собрание надо импортировать с другим именем.

**-colDataDir** *каталог\_данных\_нового\_собрания*

Задает каталог данных для собрания. Если не задан, подразумевается каталог по умолчанию.

**-force**

Принудительный импорт собрания, уже существующего в системе назначения. При помощи ID импортируемого собрания система определяет, какое собрание заменить в системе назначения.

**Важное замечание:**

- После импорта собрания нельзя импортировать собрание с тем же ID повторно, если только не применить эту опцию, позволяющую импортировать это собрание принудительно.
- Импортируемое принудительно собрание не сохраняет ни механизмы и словари анализа текста, ни подключаемые модули искателей. В них перезаписывается информация, содержащаяся в импортируемом собрании.
- Все искатели, связанные с собранием, удаляются и заменяются искателями, заданными в файле импорта. Поскольку удаляются искатели, удаляются также все их метаданные и документы, синтаксис которых еще не был проанализирован. После импорта собрания нужно выполнить повторный просмотр искателями всех документов, чтобы заново сгенерировать эти данные для импортированных искателей.

**-verbose**

Выводит информацию, помогающую устранить неисправности в операции импорта.

## Рекомендации по использованию

### Механизмы и словари анализа текста

Никакие механизмы и словари анализа текста (например, словари синонимов, стоп-слов и значимых слов), связанные с собранием в системе экспорта, с данными конфигурации собрания не экспортируются. Экспортируются только связи имен с механизмами и словарями анализа текста.

При импорте, если в системе назначения существует механизм или словарь анализа текста с тем же именем, он связывается с импортируемым собранием. Иначе выводится предупреждение, указывающее, что связи установить не удалось. По этой причине механизмы и словари анализа текста

следует выгрузить в систему назначения и использовать те же имена, которые использовались в импортируемом собрании.

Если в системе назначения не существует механизм или словарь анализа текста с тем же именем, связь разрывается. При разорванной связи собрание будет работать в системе назначения правильно, но соответствующие механизмы и словари анализа текста использовать не будет.

#### **Модели памяти**

Модель памяти, сконфигурированная для системы экспорта, не сохраняется. Модель памяти, сконфигурированная для системы экспорта остается той, какой она была сконфигурирована. При попытке импортировать собрание в систему, сконфигурированную для использования модели памяти меньшего размера, выводится предупреждение. Поведение работы собрания может оказаться неожиданным и оказать серьезное влияние на использование ресурсов системы меньшего размера. Чтобы избежать ошибок, импортируйте собрание в систему, где используется такая же модель памяти или большего размера, чем в системе экспорта.

#### **Искатели**

Нужно убедиться, что в системе назначения установлено и сконфигурировано необходимое программное обеспечение.

Если обрабатываемые искателями источники данных недоступны, импортированные искатели работать не будут. Например, если в собрание включен искатель файловой системы, обрабатывающий локальную файловую систему, он не сможет выполнять просмотр, если та же структура каталогов не будет существовать в системе назначения.

Подключаемые модули искателей не экспортируются. При импорте собрания выводится предупреждение, после чего операция импорта продолжается. После импорта собрания измените свойства искателей и свяжите с ними подключаемые модули искателей. Для правильной работы искателя необходимо убедиться, что его подключаемые модули внедрены в систему назначения.

#### **Ссылки, связанные с данной**

“Команды, коды возврата и ID сеансов поисковой системы предприятия” на стр. 373



---

## Интеграция с Lotus Notes Версии 8

Можно расширить поисковые возможности внедренных клиентов IBM Lotus Notes Версии 8, внедрив подключаемый модуль поиска OmniFind Enterprise Edition Lotus Notes. Этот подключаемый модуль позволяет пользователям выполнять поиск в собраниях поисковой системы предприятия с панели поиска клиента Lotus Notes.

Чтобы интегрировать возможности поисковой системы предприятия с Lotus Notes, надо создать сайт обновления подключаемых модулей. После этого пользователи смогут установить это подключаемый модуль в свои локальные клиентские системы Lotus Notes Версии 8.

---

### Создание сайта обновления подключаемых модулей для поисковой системы предприятия

Чтобы интегрировать поисковую систему предприятия с Lotus Notes Версии 8, надо внедрить сайт обновления Eclipse поисковой системы предприятия на Web-сервер вашей организации. Сайт обновления позволяет пользователям внедрить подключаемый модуль поисковой системы предприятия в их локальные установки клиента Lotus Notes.

#### Об этой задаче

Файлы, требуемые для создания сайта обновления поисковой системы предприятия, содержатся в архиве `com.ibm.es.notes.search.plugin_8.4.0.150.zip`.

#### Процедура

Для создания сайта обновления Eclipse поисковой системы предприятия:

1. Скопируйте файл `com.ibm.es.notes.search.plugin_8.4.0.150.zip` на Web-сервер вашей организации.
2. Распакуйте содержимое этого файла zip в подкаталог общего каталога пути установки вашего Web-сервера.

Например, чтобы внедрить сайт обновления на Web-сервер, который используется для поисковой системы предприятия, распакуйте содержимое файла zip в каталог `C:\Program Files\IBM\HTTP Server\htdocs\en_US\omnifind`.

---

### Установка подключаемого модуля поисковой системы предприятия в клиенте Lotus Notes версии 8

Чтобы использовать функции поисковой системы предприятия для обработки запросов, можно добавить подключаемый модуль поисковой системы предприятия на панель поиска Lotus Notes.

#### Об этой задаче

В верхнем правом углу экрана клиента Lotus Notes находится панель поиска. На этой панели есть несколько подключаемых модулей поиска, установленных при базовой установке клиента Lotus Notes. После установки подключаемого модуля поисковой системы предприятия вы сможете выполнять поиск в собраниях поисковой системы предприятия с этой панели поиска.

## Процедура

Чтобы установить подключаемый модуль поисковой системы предприятия в клиенте Lotus Notes:

1. Откройте файл КАТАЛОГ\_УСТАНОВКИ\_NOTES\notes.ini, где КАТАЛОГ\_УСТАНОВКИ\_NOTES - обычно c:\lotus\notes.
2. Добавьте в конец файла свойство под именем OMNIFIND\_ENTERPRISE\_EDITION\_SEARCH\_SERVER\_URL и задайте для него значение `http://имя_хоста:порт/ESSearchApplication/search.do?q=`, где *имя\_хоста* - имя хоста сервера поиска в поисковой системе предприятия, а *порт* - порт Web-сервера.
3. Сохраните и закройте этот файл.
4. Откройте клиентскую программу Lotus Notes.
5. Выберите **Файл** → **Программа** → **Установить**.
6. Нажмите радиокнопку **Искать новые возможности для установки**.
7. Выберите **Добавить удаленное положение**.
8. Задайте имя для этого нового положения, например, enterprise search. В поле URL введите корневой URL сайта обновления поисковой системы предприятия. Например, если вы распаковали файл com.ibm.es.notes.search.plugin\_8.4.0.150.zip в подкаталог omnifind корневого английского каталога IBM HTTP Server, надо будет задать URL `http://имя_хоста:порт/omnifind/`.
9. Нажмите кнопку **Закончить**, чтобы сохранить новый удаленный сайт.
10. Включите переключатель рядом с именем сайта и нажмите кнопку **Далее**.
11. Прочтите лицензионное соглашение, выберите **Я принимаю условия лицензионного соглашения** и нажмите кнопку **Далее**.
12. Нажмите кнопку **Закончить**, чтобы установить подключаемый модуль. Если вы получите предупреждение об установке неподписанного модуля, выберите **Установить этот подключаемый модуль** и нажмите кнопку **ОК**.
13. Когда появится предложение перезапустить клиент Lotus Notes, выберите **Да**.

Теперь можно вызывать меню рядом с панелью поиска Lotus Notes и выбирать опцию поиска **OmniFind Enterprise Edition**. Если вы введете термины запроса и нажмете кнопку **Поиск**, требование будет направлено серверу поисковой системы предприятия, а результаты будут выведены в окне браузера Web Lotus Notes.

---

## Интеграция с WebSphere Portal

Возможности поиска IBM WebSphere Portal можно расширить, внедрив портлет поиска для портлетов поисковой системы предприятия в WebSphere Portal и сконфигурировав WebSphere Portal, чтобы в качестве механизма поиска по умолчанию использовался этот портлет поиска.

### Точки интеграции

С программой установки OmniFind Enterprise Edition поставляются сценарии установки для интеграции поисковой системы предприятия с WebSphere Portal. После выполнения этих сценариев можно интегрировать поисковую систему предприятия с WebSphere Portal различными способами:

#### Портлет поиска для поисковой системы предприятия

WebSphere Portal предоставляет пользователям единую точку доступа для взаимодействия с прикладными программами, содержанием, процессами и людьми. Структура WebSphere Portal позволяет интегрировать и размещать новые программы, называемые портлетами, не оказывая влияния на другие программы в этом портале.

Внедрив в WebSphere Portal портлет поиска для поисковой системы предприятия, можно использовать интерфейс WebSphere Portal для выполнения поиска в собраниях поисковой системы предприятия и работы с результатами поиска. Конфигурируя параметры WebSphere Portal, можно добиться того, что портлет поисковой системы предприятия будет иметь такие же вид и поведение, как и остальные портлеты в среде WebSphere Portal.

#### Центр поиска WebSphere Portal

Центр поиска WebSphere Portal служит в качестве центральной отправной точки для поиска во всех источниках, которые сделаны доступными для поиска через WebSphere Portal. Центр поиска и портлет универсального поиска позволяют выполнять поиск в содержимом WebSphere Portal и во всех других собраниях, зарегистрированных в Центре поиска.

При выполнении сценариев установки для интеграции поисковой системы предприятия с WebSphere Portal Версии 5.1 к странице в интерфейсе Центра поиска добавляется страница поисковой системы предприятия. Можно либо выбрать эту страницу для поиска только в собраниях предприятия, либо ввести запрос на поиск в собраниях поисковой системы предприятия и других собраниях, доступных в Центре поиска.

При выполнении сценариев установки для интеграции поисковой системы предприятия с WebSphere Portal Версии 6 функции поисковой системы предприятия интегрируются в виде службы объединения, которую можно использовать для поиска в собраниях поисковой системы предприятия и других собраниях, доступных в Центре поиска.

#### Панель поиска WebSphere Portal

Во всех темах интерфейса WebSphere Portal в верхнем правом углу находится панель поиска. По умолчанию эта панель направляет все требования поиска применяемому по умолчанию механизму поиска Центра поиска. Чтобы использовать для обработки запросов более мощные функции поисковой системы предприятия, можно изменить это поведение по умолчанию, чтобы все требования поиска перенаправлялись портлету поиска поисковой системы предприятия.

### **Искатели WebSphere Portal и Web Content Management**

Чтобы включить сайты WebSphere Portal и сайты IBM Workplace Web Content Management в индекс поисковой системы предприятия, можно с помощью консоли управления поисковой системой предприятия сконфигурировать искатели WebSphere Portal и Web Content Management. Затем можно использовать портлет поисковой системы предприятия или программу поиска для поиска проиндексированного содержимого.

Искатель WebSphere Portal может просматривать сайты WebSphere Portal Версии 5.1 и WebSphere Portal Версии 6. Искатель Web Content Management может просматривать сайты на сервере WebSphere Portal Версии 6.

### **Документы IBM Lotus Quickr**

Чтобы включить документы IBM Lotus Quickr в индекс поисковой системы предприятия, можно с помощью консоли администратора поисковой системой предприятия сконфигурировать искатель Список заполнения. Затем можно использовать портлет поисковой системы предприятия в WebSphere Portal или автономную программу поиска для поиска проиндексированного содержимого.

Искатель Список заполнения может просматривать библиотеки с содержимым (документами) Lotus Quickr на сервере WebSphere Portal Версии 6.

## **Преимущества интеграции**

Поисковая система предприятия расширяет возможности среды поиска WebSphere Portal, предоставляя поддержку для поиска в большем числе типов источников данных. При помощи портлета поиска для поисковой системы предприятия можно выполнять поиск на сайтах и во всех других типах источников данных, поддерживаемых поисковой системой предприятия.

Поисковая система предприятия дает также преимущества в масштабируемости. Portal Search Engine полезен для малого и среднего бизнеса, в котором для обеспечения рабочей нагрузки по поиску и получению документов достаточно одного поискового сервера. Для крупного предприятия можно распределить рабочую нагрузку поисковой системы по нескольким серверам, два из которых будут обеспечивать поддержку операций поиска и получения данных.

---

## **Сценарии конфигурирования для интеграции поисковой системы предприятия с WebSphere Portal**

Для интеграции поисковой системы предприятия с IBM WebSphere Portal можно использовать сценарии, поставляемые с программой установки OmniFind Enterprise Edition.

Нужно скопировать файл JAR, содержащий сценарии установки для вашей версии WebSphere Portal, с сервера поисковой системы предприятия на сервер, где установлен WebSphere Portal. Сценарии установки:

- Внедряют файлы EAR, позволяющие использовать поисковую систему предприятия в WebSphere Portal и создавать искатели для добавления содержимого WebSphere Portal и IBM Workplace Web Content Management в собрания поисковой системы предприятия.
- Внедряют файлы WAR, необходимые портлету поисковой системы предприятия.
- Создают страницы в WebSphere Portal и связывают файлы портлета поисковой системы предприятия с этими страницами.

- Копируют все необходимые файлы JAR в каталоги установки WebSphere Portal (перед копированием файлов JAR, используемых поисковой системой предприятия, создаются резервные копии файлов JAR, которые уже есть в каталогах установки).
- Обеспечивают точку интеграции для WebSphere Information Integrator Content Edition для поиска документов Portal Document Manager.

После выполнения этих сценариев нужно с помощью интерфейса управления WebSphere Portal изменить свойства портлета поиска и задать информацию о сервере поиска для поисковой системы предприятия.

## Рекомендации по использованию

- Эти сценарии устанавливают все точки интеграции между поисковой системой предприятия и WebSphere Portal. Например, нельзя выборочно установить портлет и не устанавливать файлы EAR для поддержки искателей WebSphere Portal и Web Content Management.
- Если у вас не был установлен WebSphere Information Integrator Content Edition, а впоследствии вы решили использовать портлет для поиска документов Portal Document Manager, нужно запустить сценарий для удаления поисковой системы предприятия из WebSphere Portal. Затем можно вновь запустить сценарий установки, задав путь установки WebSphere Information Integrator Content Edition.
- Эти сценарии останавливают и перезапускают WebSphere Portal. Возможно, лучше запустить эти сценарии вне обычного рабочего времени, чтобы недоступность служб портала не помешала работе пользователей.
- Если при выполнении сценариев установки возникли ошибки, еще раз запустите сценарий установки. Для задач, успешно выполненных при первой попытке, могут выводиться сообщения об ошибках, но процесс установки будет продолжен и будут выполнены оставшиеся задачи.
- При первом обращении к странице портлета поисковой системы предприятия после выполнения сценария установки эта страница может открываться медленно, потому что система должна скомпилировать Java Server Pages (файлы JSP) для портлета.

---

## Конфигурирование поисковой системы предприятия в WebSphere Portal версии 5.1

Для интеграции поисковой системы предприятия с WebSphere Portal Версии 5.1.0 или новее используется сценарий `wp5_install`.

### Об этой задаче

Файлы, необходимые для интеграции поисковой системы предприятия с WebSphere Portal, находятся в архиве `es.wp5.install.jar`. При распаковке этого файла из него извлекаются следующие файлы:

- `ESSearchPortlet.war`
- `ESSearchAdapterPortlet.war`
- `ESSearchAdapter.ear`
- `ESPACTServer.ear`
- `esapi.jar`
- `siapi.jar`
- `es.security.jar`
- Файлы сценариев, пакетные файлы, файлы XML и JACL, необходимые для установки

## Процедура

Для интеграции поисковой системы предприятия с системой WebSphere Portal Версии 5.1:

1. Скопируйте файл `es.wp5.install.jar` с сервера поисковой системы предприятия на сервер WebSphere Portal и распакуйте этот файл при помощи команды Java **JAR** (или **TAR**).
2. Необязательно: Если требуется поддержка интеграции с WebSphere Portal Document Manager (PDM), выполните один из следующих шагов:
  - Запустите программу установки WebSphere Information Integrator Content Edition, выберите опцию для установки только соединителя и установите соединитель PDM на сервере WebSphere Portal.
  - Создайте на сервере WebSphere Portal структуру каталогов WebSphere Information Integrator Content Edition и скопируйте на сервер WebSphere Portal следующие файлы из существующей установки WebSphere Information Integrator Content Edition, где `CE_ROOT` - корневой каталог установки WebSphere Information Integrator Content Edition:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. В командной строке введите команду **wp5\_install.bat** (в Windows) или **wp5\_install.sh** (в AIX, Linux или Solaris). В команде нужно задать следующие параметры (для удобства чтения они показаны ниже в отдельных строках):

```
wp5_install.bat
-WASDir "C:\\Program Files\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
-IICEDir "C:\\IICE"
```

### **WASDir**

Полный путь каталога установки WebSphere Application Server.

### **WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

### **WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

### **WPSDir**

Полный путь каталога установки WebSphere Portal.

### **WPSUser**

Имя пользователя - администратора WebSphere Portal.

### **WSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

### **WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

### **IICEDir**

Полный путь каталога установки WebSphere Information Integrator Content Edition; требуется, только если на сервере WebSphere Portal установлен соединитель Portal Document Manager.

4. После выполнения этого сценария (WebSphere Portal будет остановлен и перезапущен) измените портлет поисковой системы предприятия, задав сервер поиска:
  - a. Зарегистрируйтесь в WebSphere Portal, используя ID администратора WebSphere Portal и соответствующий пароль.
  - b. В верхнем правом углу выберите **Управление**.
  - c. В области навигации слева выберите **Управление портлетами** и затем выберите **Портлеты**.
  - d. Измените значение опции **Поиск по** на **Заголовок содержит**.
  - e. В поле **Поиск** введите enterprise search и нажмите кнопку **Поиск**.
  - f. Когда справа появятся новые значки, выберите значок гаечного ключа, чтобы сконфигурировать портлет поиска для поисковой системы предприятия.
  - g. В списке параметров портлета измените следующие параметры:

**hostname**

Задайте полное имя хоста сервера поиска для поисковой системы предприятия.

**port**

Задайте номер порта, который WebSphere Application Server использует на сервере поиска для поисковой системы предприятия. Значение по умолчанию - 80 (значение по умолчанию для связи SSL - 443).

**username**

Если в WebSphere Application Server на сервере поиска включена глобальная защита, задайте допустимое имя пользователя из реестра пользователей WebSphere Application Server.

**password**

Если задано имя пользователя WebSphere Application Server, задайте соответствующий пароль.

**protocol**

Задайте протокол для связи между WebSphere Portal и сервером поиска. Протокол по умолчанию - HTTP. Если используется SSL, задайте протокол HTTPS.

**trustStore**

Если используется SSL, задайте полный путь и имя файла склада сертификатов SSL.

**trustPassword**

Если используется SSL, задайте пароль для заданного файла склада сертификатов.

**ssoCookieName**

Задайте имя опознавательного файла (cookie), содержащего строку маркеров для единой регистрации (SSO). Значение по умолчанию - LtraToken.

**proxyHost**

Если для доступа к серверу поиска поисковой системы предприятия требуется прокси-сервер, задайте полное имя хоста прокси-сервера.

**proxyPort**

Если задан прокси-сервер, задайте для него номер порта.

**proxyUser**

Если на прокси-сервере требуется базовая аутентификация, задайте имя пользователя для регистрации на прокси-сервере.

### proxyPassword

Если задано имя пользователя для прокси-сервера, задайте соответствующий пароль.

- h. Нажмите кнопку **ОК**, чтобы сохранить внесенные изменения.

## Конфигурирование панели поиска WebSphere Portal Версии 5.1 для использования поисковой системы предприятия

WebSphere Portal Версии 5.1.0 или новее можно сконфигурировать так, чтобы при вводе пользователем запроса на панели поиска использовалась поисковая система предприятия, а не механизм поиска WebSphere Portal по умолчанию.

### Прежде чем приступить к работе...

Чтобы можно было перенаправить требования поиска поисковой системе предприятия, нужно сначала запустить сценарий установки **wp5\_install** для интеграции поисковой системы предприятия с WebSphere Portal. Кроме того, нужно изменить параметры портлета поисковой системы предприятия, задав имя хоста, порт и другую информацию о сервере поиска для поисковой системы предприятия.

### Об этой задаче

Во всех темах интерфейса WebSphere Portal в верхнем правом углу находится панель поиска. По умолчанию эта панель направляет все требования поиска портлету Центра поиска. Чтобы использовать для обработки запросов более мощные функции поисковой системы предприятия, можно изменить это поведение по умолчанию, чтобы все требования поиска перенаправлялись портлету поиска поисковой системы предприятия.

Если перенаправить требования с панели поиска, это изменение повлияет на страницы, которые используют ту же тему WebSphere Portal, что и портлет поиска для поисковой системы предприятия, и вызывают файл `SearchBarInclude.jsp`. Страницы, использующие другую тему, или не вызывающие файл `SearchBarInclude.jsp` будут продолжать использовать применяемый по умолчанию портлет Центра поиска.

После выполнения этой задачи уже нельзя будет использовать Центр поиска, если только не отменить внесенные изменения (например, можно восстановить исходный файл `SearchBarInclude.jsp`).

### Процедура

Чтобы для запросов пользователей, введенных на панели поиска WebSphere Portal, использовался портлет поисковой системы предприятия:

1. На сервере WebSphere Portal перейдите в каталог `WAS_INSTALL_ROOT/AppServer/installedApps/eswin1/wps.ear/wps.war/themes/html`.
2. Создайте резервную копию файла `SearchBarInclude.jsp`, для чего скопируйте этот файл под другим именем (например, `SearchBarInclude.jsp.BACKUP`).
3. Отредактируйте файл `SearchBarInclude.jsp`, заменив содержимое между тегами `<wps:if loggedIn="yes">` и `</wps:if>` на следующий текст. В атрибуте `action=` элемента `form` замените `localhost:9081` на имя хоста и номер порта вашего сервера WebSphere Portal.

```
<wps:if loggedIn="yes">
<wps-internal:adminlinkinfo name="%=AdminUniqueNamesMappingService.SEARCH_CENTER%">
 <td nowrap="nowrap" align="right" valign="middle" class="wpsLinkBar">
 <form style="margin: 0px;" method="post"
```



**WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

**WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

**WPSDir**

Полный путь каталога установки WebSphere Portal.

**WPSUser**

Имя пользователя - администратора WebSphere Portal.

**WPSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

**WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

---

## Конфигурирование поисковой системы предприятия в WebSphere Portal версии 6

Для интеграции поисковой системы предприятия с WebSphere Portal Версии 6 используется сценарий `wrb_install`.

**Об этой задаче**

Файлы, необходимые для интеграции поисковой системы предприятия с WebSphere Portal, находятся в архиве `es.wrb.install.jar`. При распаковке этого файла из него извлекаются следующие файлы:

- `ESSearchPortlet.war`
- `ESPACTServer.ear`
- `esapi.jar`
- `es.search.provider.jar`
- `es.security.jar`
- Значки типа исходных файлов программы поиска, используемые на странице результатов провайдера поиска
- Файлы сценариев, пакетные файлы, файлы XML и JACL, необходимые для установки

**Процедура**

Для интеграции поисковой системы предприятия с системой WebSphere Portal Версии 6:

1. Скопируйте файл `es.wrb.install.jar` с сервера поисковой системы предприятия на сервер WebSphere Portal и распакуйте этот файл при помощи команды Java **JAR** (или **TAR**).
2. Необязательно: Если требуется поддержка интеграции с WebSphere Portal Document Manager (PDM), выполните один из следующих шагов:
  - Запустите программу установки WebSphere Information Integrator Content Edition, выберите опцию для установки только соединителя и установите соединитель PDM на сервере WebSphere Portal.
  - Создайте на сервере WebSphere Portal структуру каталогов WebSphere Information Integrator Content Edition и скопируйте на сервер WebSphere Portal

следующие файлы из существующей установки WebSphere Information Integrator Content Edition, где CE\_ROOT - корневой каталог установки WebSphere Information Integrator Content Edition:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. В командной строке введите команду **wp6\_install.bat** (в Windows) или **wp6\_install.sh** (в AIX, Linux или Solaris). В команде нужно задать следующие опции (для удобства чтения они показаны ниже в отдельных строках):

```
wp6_install.bat
-WSPProfileDir "C:\\Program Files\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Program Files\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\\Program Files\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
-IICEDir "C:\\IICE"
```

#### **WSPProfileDir**

Полный путь каталога профилей WebSphere Portal. Путь по умолчанию - /usr/IBM/WebSphere/AppServer/profiles/wp\_profile в системах AIX, /opt/IBM/WebSphere/AppServer/profiles/wp\_profile в системах Linux и Solaris и C:\Program Files\IBM\WebSphere\profiles\wp\_profile в системах Windows.

#### **WASDir**

Полный путь корневого каталога установки WebSphere Application Server, требуется только в системах AIX, Linux и Solaris. Путь для корневого каталога по умолчанию - /usr/IBM/WebSphere/AppServer в системах AIX, /opt/IBM/WebSphere/AppServer в системах Linux и Solaris, и C:\Program Files\IBM\WebSphere\AppServer в системах Windows.

#### **WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

#### **WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

#### **WPSDir**

Полный путь каталога установки WebSphere Portal.

#### **WPSUser**

Имя пользователя - администратора WebSphere Portal.

#### **WSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

#### **WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

#### **IICEDir**

Полный путь каталога установки WebSphere Information Integrator Content Edition; требуется, только если на сервере WebSphere Portal установлен соединитель Portal Document Manager.

4. После выполнения этого сценария, остановки и перезапуска WebSphere Portal измените портлет поисковой системы предприятия, задав сервер поиска:

- a. Зарегистрируйтесь в WebSphere Portal, используя ID администратора WebSphere Portal и соответствующий пароль.
- b. В нижнем левом углу выберите **Управление**.
- c. В области навигации слева выберите **Управление портлетами** и затем выберите **Портлеты**.
- d. Измените значение опции **Поиск по** на **Заголовок содержит**.
- e. В поле **Поиск** введите enterprise search и нажмите кнопку **Поиск**.
- f. Когда справа появятся новые значки, выберите значок гаечного ключа, чтобы сконфигурировать портлет поиска для поисковой системы предприятия.
- g. В списке параметров портлета измените следующие параметры:

**hostname**

Задайте полное имя хоста сервера поиска для поисковой системы предприятия.

**port**

Задайте номер порта, который WebSphere Application Server использует на сервере поиска для поисковой системы предприятия. Значение по умолчанию - 80 (значение по умолчанию для связи SSL - 443).

**username**

Если в WebSphere Application Server на сервере поиска включена глобальная защита, задайте допустимое имя пользователя из реестра пользователей WebSphere Application Server.

**password**

Если задано имя пользователя WebSphere Application Server, задайте соответствующий пароль.

**protocol**

Задайте протокол для связи между WebSphere Portal и сервером поиска. Протокол по умолчанию - HTTP. Если используется SSL, задайте протокол HTTPS.

**trustStore**

Если используется SSL, задайте полный путь и имя файла склада сертификатов SSL.

**trustPassword**

Если используется SSL, задайте пароль для заданного файла склада сертификатов.

**ssoCookieName**

Задайте имя опознавательного файла (cookie), содержащего строку маркеров для единой регистрации (SSO). Значение по умолчанию - LtprToken.

**proxyHost**

Если для доступа к серверу поиска поисковой системы предприятия требуется прокси-сервер, задайте полное имя хоста прокси-сервера.

**proxyPort**

Если задан прокси-сервер, задайте для него номер порта.

**proxyUser**

Если на прокси-сервере требуется базовая аутентификация, задайте имя пользователя для регистрации на прокси-сервере.

**proxyPassword**

Если задано имя пользователя для прокси-сервера, задайте соответствующий пароль.

h. Нажмите кнопку **ОК**, чтобы сохранить внесенные изменения.

## Конфигурирование Центра поиска WebSphere Portal Версии 6 для поисковой системы предприятия

WebSphere Portal Версии 6 можно сконфигурировать так, чтобы при вводе пользователями запросов в Центре поиска WebSphere Portal выполнялся поиск в собраниях поисковой системы предприятия.

### Ограничения

Если требуется выполнять поиск в защищенных собраниях поисковой системы предприятия, пользователи должны запускать портлет поиска для поисковой системы предприятия и сконфигурировать профиль пользователя. Профиль шифруется и сохраняется в защищенном хранилище поисковой системы предприятия. Этот профиль должен существовать, чтобы пользователи могли передавать из Центра поиска WebSphere Portal запросы для поиска в защищенных собраниях.

### Об этой задаче

Центр поиска в WebSphere Portal Версии 6 поддерживает возможности объединенного поиска для нескольких собраний. В собраниях может быть содержимое разных типов, например, библиотеки документов или содержимое (страницы и портлеты) WebSphere Portal. После выполнения сценариев установки для интеграции поисковой системы предприятия с WebSphere Portal можно сконфигурировать Центр поиска, чтобы в нем можно было выполнять поиск в собраниях поисковой системы предприятия.

### Процедура

Чтобы сконфигурировать Центр поиска для поиска в собраниях поисковой системы предприятия:

1. Зарегистрируйтесь в WebSphere Portal, используя ID администратора WebSphere Portal и соответствующий пароль.
2. В нижнем левом углу выберите **Управление**.
3. В области навигации слева выберите **Управление поисковой системой** и затем выберите **Управление поиском**.
4. Выберите **Службы поиска**, затем **Новая служба поиска**.
5. В поле **Реализация службы поиска** выберите **Служба поиска поисковой системы предприятия** и введите имя для этой службы в поле **Имя службы**.
6. В списке параметров измените следующие параметры:

#### hostname

Задайте полное имя хоста сервера поиска для поисковой системы предприятия.

#### port

Задайте номер порта, который WebSphere Application Server использует на сервере поиска для поисковой системы предприятия. Значение по умолчанию - 80 (значение по умолчанию для связи SSL - 443).

#### username

Если в WebSphere Application Server на сервере поиска включена глобальная защита, задайте допустимое имя пользователя из реестра пользователей WebSphere Application Server.

**password**

Если задано имя пользователя WebSphere Application Server, задайте соответствующий пароль.

**protocol**

Задайте протокол для связи между WebSphere Portal и сервером поиска. Протокол по умолчанию - HTTP. Если используется SSL, задайте протокол HTTPS.

**trustStore**

Если используется SSL, задайте полный путь и имя файла склада сертификатов SSL.

**trustPassword**

Если используется SSL, задайте пароль для заданного файла склада сертификатов.

**ssoCookieName**

Задайте имя опознавательного файла (cookie), содержащего строку маркеров для единой регистрации (SSO). Значение по умолчанию - LtraToken.

**proxyHost**

Если для доступа к серверу поиска поисковой системы предприятия требуется прокси-сервер, задайте полное имя хоста прокси-сервера.

**proxyPort**

Если задан прокси-сервер, задайте для него номер порта.

**proxyUser**

Если на прокси-сервере требуется базовая аутентификация, задайте имя пользователя для регистрации на прокси-сервере.

**proxyPassword**

Если задано имя пользователя для прокси-сервера, задайте соответствующий пароль.

7. Нажмите кнопку **ОК**, чтобы сохранить внесенные изменения.

## Конфигурирование панели поиска WebSphere Portal Версии 6 для использования поисковой системы предприятия

WebSphere Portal Версии 6 можно сконфигурировать так, чтобы при вводе пользователем запроса на панели поиска использовалась поисковая система предприятия, а не механизм поиска WebSphere Portal по умолчанию.

### Прежде чем приступать к работе...

Чтобы можно было перенаправить требования поиска поисковой системе предприятия, нужно сначала запустить сценарий установки **wp6\_install** для интеграции поисковой системы предприятия с WebSphere Portal. Кроме того, нужно изменить параметры портлета поисковой системы предприятия, задав имя хоста, порт и другую информацию о сервере поиска для поисковой системы предприятия.

### Об этой задаче

Во всех темах интерфейса WebSphere Portal в верхнем правом углу находится панель поиска. По умолчанию эта панель направляет все требования поиска портлету Центра поиска. Чтобы использовать для обработки запросов более мощные функции

поисковой системы предприятия, можно изменить это поведение по умолчанию, чтобы все требования поиска перенаправлялись портлету поиска поисковой системы предприятия.

Если перенаправить требования с панели поиска, это изменение повлияет на страницы, которые используют ту же тему WebSphere Portal, что и портлет поиска для поисковой системы предприятия, и вызывают файл `banner_searchControl.jspf`. Страницы, использующие другую тему, или не вызывающие файл `banner_searchControl.jspf`, будут продолжать использовать применяемый по умолчанию портлет Центра поиска.

После выполнения этой задачи уже нельзя будет использовать Центр поиска, если только не отменить внесенные изменения (например, можно восстановить исходный файл `banner_searchControl.jspf`).

### Процедура

Чтобы для запросов пользователей, введенных на панели поиска WebSphere Portal, использовался портлет поисковой системы предприятия:

1. Остановите экземпляр сервера прикладных программ WebSphere Portal.
2. На сервере WebSphere Portal перейдите в каталог `КОРНЕВОЙ_КАТАЛОГ_ПРОФИЛЯ_WPS/installedApps/имя_узла/wps.ear/wps.war/themes/html/имя_текущей_темы`, где `имя_узла` - имя узла сервера WebSphere Portal, а `текущее_имя_темы` - текущая примененная схема для сервера WebSphere Portal. По умолчанию имя темы для сервера WebSphere Portal - IBM.
3. Создайте резервную копию файла `banner_searchControl.jspf`, для чего скопируйте этот файл под другим именем (например, `banner_searchControl.jspf.BACKUP`).
4. Отредактируйте файл `banner_searchControl.jspf`, заменив его содержимое следующим текстом. В атрибуте `action=` элемента `form` замените `localhost:10038` на имя хоста и номер порта вашего сервера WebSphere Portal.

```
<%@ taglib uri="/WEB-INF/tld/SearchMenuControl.tld" prefix="searchmenu" %>
<%String ic = (bidiImageRTL == null) ? "icons/scope_search_submit.gif" :
"icons/scope_search_submit"+bidiImageRTL+".gif";%>

<searchmenu:adminlinkinfo name="SEARCH_CENTER">
 <div class="searchControl">
 <form name="SearchForm" style="margin: 0px;" method="GET"
 action="http://localhost:10038/wps/omnifind/portalSearchBar.jsp">
 <table border="0" cellpadding="0" cellspacing="0">
 <tr>
 <td>
 <portal-fmt:text key="search.theme.control.label" bundle="nls.engine"/></td>
 <td valign="middle" style="padding: 0px 4px 0px 4px;">
 <input type="text" name="q"></input></td>
 <td valign="middle"><input tabIndex="4" valign="middle"
 title="<portal-fmt:text key='search.theme.searchresultsicon.alttext' bundle='nls.engine' />"
 alt="<portal-fmt:text key='search.theme.searchresultsicon.alttext' bundle='nls.engine' />"
 src="<portal-logic:urlFindInTheme file=">"/>" type="image"></input></td>
 </tr>
 </table>
 </form>
 </div>
</searchmenu:adminlinkinfo>
```

5. Откройте файл `banner.jspf` и сохраните этот файл. Если вместо темы WebSphere Portal по умолчанию вы используете свою собственную тему, это действие выполнять не обязательно; оно изменяет дату файла, чтобы гарантировать, что файл будет перекомпилирован.
6. Откройте файл `Default.jsr` и сохраните этот файл.
7. Перезапустите экземпляр сервера прикладных программ WebSphere Portal.

## Конфигурирование портлета поисковой системы предприятия для Lotus Quickr

Можно сконфигурировать портлет поисковой системы предприятия в WebSphere Portal версии 6 для поиска в источниках Lotus Quickr.

### Прежде чем приступать к работе...

Запустите команду **wp6\_install.bat** (в Windows) или **wp6\_install.sh** (в AIX, Linux или Solaris) и следуйте процедурам конфигурирования поисковой системы предприятия в WebSphere Portal версии 6.

### Процедура

Чтобы сконфигурировать портлет поисковой системы предприятия в WebSphere Portal версии 6 для поиска в источниках Lotus Quickr:

1. Измените параметры портлета в конфигурации WebSphere Portal:
  - a. Зарегистрируйтесь в WebSphere Portal, используя ID администратора Lotus Quickr и соответствующий пароль.
  - b. Выберите **Управление узлом**, а затем **Расширенное управление**.
  - c. В области навигации слева выберите **Управление портлетами** и затем выберите **Портлеты**.
  - d. Измените значение опции **Поиск по** на **Заголовок содержит**.
  - e. В поле **Поиск** введите `enterprise search` и нажмите кнопку **Поиск**.
  - f. Когда справа появятся новые значки, выберите значок гаечного ключа, чтобы сконфигурировать портлет поиска для поисковой системы предприятия.
  - g. В списке параметров портлета измените следующие параметры:

#### **hostname**

Задайте полное имя хоста сервера поиска для поисковой системы предприятия.

#### **port**

Задайте номер порта, который WebSphere Application Server использует на сервере поиска для поисковой системы предприятия. Значение по умолчанию - 80 (значение по умолчанию для связи SSL - 443).

#### **username**

Если в WebSphere Application Server на сервере поиска включена глобальная защита, задайте допустимое имя пользователя из реестра пользователей WebSphere Application Server.

#### **password**

Если задано имя пользователя WebSphere Application Server, задайте соответствующий пароль.

#### **protocol**

Задайте протокол для связи между WebSphere Portal и сервером поиска. Протокол по умолчанию - HTTP. Если используется SSL, задайте протокол HTTPS.

#### **trustStore**

Если используется SSL, задайте полный путь и имя файла склада сертификатов SSL.

#### **trustPassword**

Если используется SSL, задайте пароль для заданного файла склада сертификатов.

**ssoCookieName**

Задайте имя опознавательного файла (cookie), содержащего строку маркеров для единой регистрации (SSO). Значение по умолчанию - LtpaToken.

**proxyHost**

Если для доступа к серверу поиска поисковой системы предприятия требуется прокси-сервер, задайте полное имя хоста прокси-сервера.

**proxyPort**

Если задан прокси-сервер, задайте для него номер порта.

**proxyUser**

Если на прокси-сервере требуется базовая аутентификация, задайте имя пользователя для регистрации на прокси-сервере.

**proxyPassword**

Если задано имя пользователя для прокси-сервера, задайте соответствующий пароль.

- h. Нажмите кнопку **ОК**, чтобы сохранить внесенные изменения.
2. Чтобы обратиться к портлету после его конфигурирования:
  - a. Зарегистрируйтесь на сервере Lotus Quickr.
  - b. В окне браузера измените URL на следующий: `http://имя_хоста:порт/lotus/myquickr/ESSearchPortlet`

## Удаление поисковой системы предприятия из WebSphere Portal Версии 6

Для удаления поисковой системы предприятия из системы WebSphere Portal Версии 6 используется сценарий `wp6_uninstall`.

**Об этой задаче**

При удалении поисковой системы предприятия из WebSphere Portal параметры портлета, заданные для портлета поисковой системы предприятия в процессе установки, не сохраняются.

Когда вы запустите этот сценарий, он остановит сервер WebSphere Portal. После удаления программ поисковой системы предприятия этот сценарий перезапустит сервер WebSphere Portal.

**Процедура**

Чтобы удалить поисковую систему предприятия из системы WebSphere Portal Версии 6:

В командной строке введите команду **wp6\_uninstall.bat** (в Windows) или **wp6\_uninstall.sh** (в AIX, Linux или Solaris). В команде нужно задать следующие параметры (для удобства чтения они показаны ниже в отдельных строках):

```
wp6_uninstall.bat
-WPSProfileDir "C:\Program Files\IBM\WebSphere\AppServer\profiles\wp_profile"
-WASDir "C:\Program Files\IBM\WebSphere\AppServer"
-WASUser wpsbind -WASPassword wpsbind
-WPSDir "C:\Program Files\IBM\WebSphere\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com:9081"
```

**WPSProfileDir**

Полный путь каталога профилей WebSphere Portal.

**WASDir**

Полный путь корневого каталога установки WebSphere Application Server, требуется только в системах AIX, Linux и Solaris.

**WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

**WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

**WPSDir**

Полный путь каталога установки WebSphere Portal.

**WPSUser**

Имя пользователя - администратора WebSphere Portal.

**WPSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

**WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

---

## Интеграция поисковой системы предприятия с кластеризованной системой WebSphere Portal

Можно сконфигурировать портлет поисковой системы предприятия для работы в кластерной системе WebSphere Portal версии 6.

### Рекомендации по использованию

- Прежде чем запускать сценарии конфигурирования для поисковой системы предприятия, убедитесь, что менеджер сетевого внедрения WebSphere Application Server работает и все узлы кластера также работают.
- Эти сценарии устанавливают все точки интеграции между поисковой системой предприятия и WebSphere Portal. Например, нельзя выборочно установить портлет и не устанавливать файлы EAR для поддержки искателей WebSphere Portal и Web Content Management.
- Эти сценарии останавливают и перезапускают все экземпляры сервера WebSphere Portal в кластере. Возможно, лучше запустить эти сценарии вне обычного рабочего времени, чтобы недоступность служб портала не помешала работе пользователей.
- Если при выполнении сценариев установки возникли ошибки, еще раз запустите сценарий установки. Для задач, успешно выполненных при первой попытке, могут выводиться сообщения об ошибках, но процесс установки будет продолжен и будут выполнены оставшиеся задачи.
- При первом обращении к странице портлета поисковой системы предприятия после выполнения сценария установки эта страница может открываться медленно, потому что система должна скомпилировать Java Server Pages (файлы JSP) для портлета.

## Конфигурирование поисковой системы предприятия в кластерной системе WebSphere Portal

Для интеграции поисковой системы предприятия с кластерной системой WebSphere Portal версии 6 служит сценарий `wrb_cluster_install`.

### Об этой задаче

Файлы, необходимые для интеграции поисковой системы предприятия с WebSphere Portal, находятся в архиве `es.wpb.install.jar`. При распаковке этого файла из него извлекаются следующие файлы:

- `ESSearchPortlet.war`
- `ESPACTServer.ear`
- `esapi.jar`
- `es.search.provider.jar`
- `es.security.jar`
- Файлы сценариев, пакетные файлы, файлы XML и JACL, необходимые для установки

## Процедура

Для интеграции поисковой системы предприятия с кластерной системой WebSphere Portal Версии 6:

1. Скопируйте файл `es.wpb.install.jar` с сервера поисковой системы предприятия на каждый узел кластера, где установлен WebSphere Portal, и распакуйте этот файл при помощи команды Java **JAR** (или **TAR**).
2. Необязательно: Если требуется поддержка интеграции с WebSphere Portal Document Manager (PDM), выполните на каждом узле кластера одно из следующих действий:
  - Запустите программу установки WebSphere Information Integrator Content Edition, выберите опцию для установки только соединителя и установите соединитель PDM на сервере WebSphere Portal.
  - Создайте на сервере WebSphere Portal структуру каталогов WebSphere Information Integrator Content Edition и скопируйте на сервер WebSphere Portal следующие файлы из существующей установки WebSphere Information Integrator Content Edition, где `CE_ROOT` - корневой каталог установки WebSphere Information Integrator Content Edition:

```
CE_ROOT/lib/vbr.jar
CE_ROOT/ejb/vbr_pdm.jar
CE_ROOT/war/services.war
CE_ROOT/vbr_services.properties
```

3. В командной строке введите команду **wp6\_cluster\_copyFiles.bat** command (в Windows) или **wp6\_cluster\_copyFiles.sh** (в AIX, Linux или Solaris) на каждом узле кластера, где установлен WebSphere Portal. В команде нужно задать следующие опции (для удобства чтения ниже в примерах они показаны на отдельных строках):

```
wp6_cluster_copyFiles.bat
-WPSDir "C:\Program Files\IBM\WebSphere\PortalServer"
-WSPProfileDir "C:\Program Files\IBM\WebSphere\AppServer\profiles\wp_profile"

wp6_cluster_copyFiles.sh
-WASDir /opt/IBM/WebSphere/AppServer
-WPSDir /opt/IBM/WebSphere/PortalServer
-WSPProfileDir /opt/IBM/WebSphere/AppServer/profiles/wp_profile
```

4. В командной строке введите команду **wp6\_cluster\_install.bat** (в Windows) или **wp6\_cluster\_install.sh** (в AIX, Linux или Solaris). В команде нужно задать следующие опции (для удобства чтения они показаны ниже в отдельных строках):

```
wp6_cluster_install.bat
-WPSClusterName MyCluster
-WSPProfileDir "C:\Program Files\IBM\WebSphere\profiles\wp_profile"
-WASDir "C:\Program Files\IBM\WebSphere\AppServer"
-WASUser wpsbind -WASPassword wpsbind
```

```
-WPSDir "C:\\Program Files\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin -WSPassword wpsadmin
-WPSHost "portalserver.ibm.com"
-webServerName webserver1
-webServerNodeName node1
-IICEDir "C:\\Program Files\\IBM\\Content Edition"
```

**WPSClusterName**

Имя кластера, в котором установлен WebSphere Portal.

**WPSProfileDir**

Полный путь каталога профилей WebSphere Portal. Путь по умолчанию - /usr/IBM/WebSphere/AppServer/profiles/wp\_profile в системах AIX, /opt/IBM/WebSphere/AppServer/profiles/wp\_profile в системах Linux и Solaris и C:\\Program Files\\IBM\\WebSphere\\profiles\\wp\_profile в системах Windows.

**WASDir**

Полный путь корневого каталога установки WebSphere Application Server, требуется только в системах AIX, Linux и Solaris. Путь для корневого каталога по умолчанию - /usr/IBM/WebSphere/AppServer в системах AIX, /opt/IBM/WebSphere/AppServer в системах Linux и Solaris.

**WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

**WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

**WPSDir**

Полный путь каталога установки WebSphere Portal.

**WPSUser**

Имя пользователя - администратора WebSphere Portal.

**WPSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

**WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

**webServerName**

Имя определения Web-сервера, на котором установлен WebSphere Portal.

**webServerNodeName**

Имя узла сервера WebSphere Application Server, на котором находится определение Web-сервера.

**IICEDir**

Полный путь каталога установки WebSphere Information Integrator Content Edition; требуется, только если на сервере WebSphere Portal установлен соединитель Portal Document Manager.

5. Когда сценарий завершит работу, откройте браузер и зарегистрируйтесь на консоли администратора WebSphere вашего сервера Network Deployment. Типовой адрес - <http://hostname:9060/ibm/console>.
6. Разверните раздел **Серверы** и выберите **Web-серверы**.
7. Включите переключатель **Выбрать** рядом с вашим Web-сервером, а затем нажмите кнопку **Сгенерировать подключаемый модуль**.
8. Включите переключатель **Выбрать** рядом с вашим Web-сервером, а затем нажмите кнопку **Распространить подключаемый модуль**.

9. Закройте консоль администратора.
10. Измените портлет поисковой системы предприятия, чтобы указать сервер поиска:
  - a. Зарегистрируйтесь в WebSphere Portal, используя ID администратора WebSphere Portal и соответствующий пароль.
  - b. В нижнем левом углу выберите **Управление**.
  - c. В области навигации слева выберите **Управление портлетами** и затем выберите **Портлеты**.
  - d. Измените значение опции **Поиск по** на **Заголовок содержит**.
  - e. В поле **Поиск** введите enterprise search и нажмите кнопку **Поиск**.
  - f. Когда справа появятся новые значки, выберите значок гаечного ключа, чтобы сконфигурировать портлет поиска для поисковой системы предприятия.
  - g. В списке параметров портлета измените следующие параметры:

**hostname**

Задайте полное имя хоста сервера поиска для поисковой системы предприятия.

**port**

Задайте номер порта, который WebSphere Application Server использует на сервере поиска для поисковой системы предприятия. Значение по умолчанию - 80 (значение по умолчанию для связи SSL - 443).

**username**

Если в WebSphere Application Server на сервере поиска включена глобальная защита, задайте допустимое имя пользователя из реестра пользователей WebSphere Application Server.

**password**

Если задано имя пользователя WebSphere Application Server, задайте соответствующий пароль.

**protocol**

Задайте протокол для связи между WebSphere Portal и сервером поиска. Протокол по умолчанию - HTTP. Если используется SSL, задайте протокол HTTPS.

**trustStore**

Если используется SSL, задайте полный путь и имя файла склада сертификатов SSL.

**trustPassword**

Если используется SSL, задайте пароль для заданного файла склада сертификатов.

**ssoCookieName**

Задайте имя опознавательного файла (cookie), содержащего строку маркеров для единой регистрации (SSO). Значение по умолчанию - LtpaToken.

**proxyHost**

Если для доступа к серверу поиска поисковой системы предприятия требуется прокси-сервер, задайте полное имя хоста прокси-сервера.

**proxyPort**

Если задан прокси-сервер, задайте для него номер порта.

**proxyUser**

Если на прокси-сервере требуется базовая аутентификация, задайте имя пользователя для регистрации на прокси-сервере.

**proxyPassword**

Если задано имя пользователя для прокси-сервера, задайте соответствующий пароль.

- h. Нажмите кнопку **ОК**, чтобы сохранить внесенные изменения.

## Удаление поисковой системы предприятия из кластерной системы WebSphere Portal

Для удаления поисковой системы предприятия из кластерной системы WebSphere Portal Версии 6 используется сценарий `wp6_cluster_uninstall`.

**Об этой задаче**

При удалении поисковой системы предприятия из WebSphere Portal параметры портлета, заданные для портлета поисковой системы предприятия в процессе установки, не сохраняются.

Когда вы запустите этот сценарий, он остановит сервер WebSphere Portal. После удаления программ поисковой системы предприятия этот сценарий перезапустит сервер WebSphere Portal.

**Процедура**

Чтобы удалить поисковую систему предприятия из кластерной системы WebSphere Portal Версии 6:

1. В командной строке введите команду **wp6\_cluster\_uninstall.bat** (в Windows) или **wp6\_cluster\_uninstall.sh** (в AIX, Linux или Solaris) на одном из узлов кластера. В команде нужно задать следующие параметры (для удобства чтения они показаны ниже в отдельных строках):

```
wp6_cluster_uninstall.bat
-WPSClusterName MyCluster
-WSPProfileDir "C:\\Program Files\\IBM\\WebSphere\\profiles\\wp_profile"
-WASDir "C:\\Program Files\\IBM\\WebSphere\\AppServer"
-WASUser wpsbind
-WASPassword wpsbind
-WPSDir "C:\\Program Files\\IBM\\WebSphere\\PortalServer"
-WPSUser wpsadmin
-WSPassword wpsadmin
-WPSHost "portalserver.ibm.com"
-webServerName webserver1
-webServerNodeName node1
```

**WPSClusterName**

Имя кластера, в котором установлен WebSphere Portal.

**WSPProfileDir**

Полный путь каталога профилей WebSphere Portal. Путь по умолчанию - `/usr/IBM/WebSphere/AppServer/profiles/wp_profile` в системах AIX, `/opt/IBM/WebSphere/AppServer/profiles/wp_profile` в системах Linux и Solaris и `C:\\Program Files\\IBM\\WebSphere\\profiles\\wp_profile` в системах Windows.

**WASDir**

Полный путь корневого каталога установки WebSphere Application Server, требуется только в системах AIX, Linux и Solaris. Путь для корневого

каталога по умолчанию - /usr/IBM/WebSphere/AppServer в системах AIX, /opt/IBM/WebSphere/AppServer в системах Linux и Solaris.

**WASUser**

Имя пользователя - администратора WebSphere Application Server; требуется, только если в WebSphere Application Server включена глобальная защита.

**WASPassword**

Пароль для пользователя - администратора WebSphere Application Server (если он задан).

**WPSDir**

Полный путь каталога установки WebSphere Portal.

**WPSUser**

Имя пользователя - администратора WebSphere Portal.

**WPSPassword**

Пароль для заданного пользователя - администратора WebSphere Portal.

**WPSHost**

Имя хоста и номер порта сервера WebSphere Portal.

**webServerName**

Имя определения Web-сервера, на котором установлен WebSphere Portal.

**webServerNodeName**

Имя узла сервера WebSphere Application Server, на котором находится определение Web-сервера.

2. Когда сценарий завершит работу, откройте браузер и зарегистрируйтесь на консоли администратора WebSphere вашего сервера Network Deployment. Типовой адрес - <http://hostname:9060/ibm/console>.
3. Разверните раздел **Серверы** и выберите **Web-серверы**.
4. Включите переключатель **Выбрать** рядом с вашим Web-сервером, а затем нажмите кнопку **Сгенерировать подключаемый модуль**.
5. Включите переключатель **Выбрать** рядом с вашим Web-сервером, а затем нажмите кнопку **Распространить подключаемый модуль**.
6. Закройте консоль администратора.



---

## Перенастройка из WebSphere Portal в поисковую систему предприятия

В поисковой системе предприятия есть мастер по перенастройке, который можно использовать для перенастройки собраний и таксономий на основе правил из IBM WebSphere Portal в поисковую систему предприятия.

В поисковой системе предприятия таксономия называется *деревом категорий*. После перенастройки таксономии для управления деревом категорий и правилами категорий можно использовать консоль администратора поисковой системы предприятия. После перенастройки собрания управлять им можно с консоли администратора.

Для перенастройки таксономий и собраний нужно запустить мастер по перенастройке на сервере индексов поисковой системы предприятия.

---

## Перенос собрания из WebSphere Portal

Чтобы перенастроить собрания и таксономии на основе правил из WebSphere Portal в поисковую систему предприятия, подготовьте эти собрания в WebSphere Portal, затем перенастройте их при помощи мастера по перенастройке.

### Прежде чем приступать к работе...

Если вы планируете перенастроить таксономии и собрания, перед использованием данной процедуры для перенастройки собраний перенастройте файлы таксономии на основе модели. Такой подход гарантирует, что перенастроенные правила категоризации будут работать с перенастроенными собраниями.

### Процедура

Чтобы перенастроить собрание (и, при желании, таксономию) из WebSphere Portal в поисковую систему предприятия:

1. В WebSphere Portal Search Engine остановите в перенастраиваемых собраниях все процессы искателей и примите или отклоните все приостановленные документы. (В поисковой системе предприятия понятие приостановленный документ не поддерживается.)
2. Для каждого перенастраиваемого собрания при помощи портлетов Portal Search Engine экспортируйте параметры в файлы XML.
3. Если сервер индексации поисковой системы предприятия установлен на отдельном сервере, скопируйте на него экспортированные файлы XML.
4. На этом сервере зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке OmniFind Enterprise Edition.
5. Перейдите в каталог установки поисковой системы предприятия:

```
UNIX: cd $ES_INSTALL_ROOT/bin
Windows: cd %ES_INSTALL_ROOT%\bin
```

6. Для перенастройки собрания с включенной защитой введите следующую команду, чтобы запустить мастер по перенастройке, затем нажмите кнопку **Далее**.

UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat

7. Для отключения защиты уровня собрания перенастраиваемых собраний введите следующую команду, чтобы запустить мастер по перенастройке, затем нажмите кнопку **Далее**.

UNIX: ./eswpsmigrate.sh disable.security  
Windows: eswpsmigrate.bat disable.security

8. Выберите **Перенастроить параметры поиска из Portal Search Engine в WebSphere Portal**, затем нажмите кнопку **Далее**.
9. Найдите каталог, содержащий экспортированные файлы конфигурации Portal Search Engine, выберите файлы для перенастройки, затем нажмите кнопку **Далее**. Выбранные файлы конфигурации будут проанализированы и проверены.
10. Введите для каждого собрания приведенную ниже информацию, затем нажмите кнопку **Далее**, чтобы перенастроить собрания в поисковую систему предприятия:
  - Имя собрания для использования в поисковой системе предприятия.
  - Критерий, по которому будет определяться важность документов для собрания. Можно ничего не выбрать, выбрать коэффициент статического ранжирования на основе состояний документов или коэффициент статического ранжирования на основе числа ссылок на документы Web из других документов Web.
  - Тип категоризации, используемый для данного собрания. Если ничего не задать, информация о таксономиях не перенастраивается в поисковую систему предприятия. Если выбрать категории на основе правил, вместе с собранием в поисковую систему предприятия будет перенастроена таксономия.

Если при перенастройке происходят ошибки, посмотрите файл MigrationWizard.log, находящийся в каталоге установки мастера по перенастройке.

Теперь можно использовать консоль администратора поисковой системы предприятия, чтобы сконфигурировать дополнительные параметры для перенастроенных собраний.

**Требование:** При конфигурировании свойств искателя Web для перенастроенного собрания нужно задать адрес электронной почты, на который будут отправляться комментарии об искателе, и имя пользовательского агента (если при конфигурировании свойств искателя Web потребуется помощь, нажмите кнопку **Справка**).

11. Запустите для перенастроенных собраний процессы искателей, синтаксического анализа и индексации с консоли администратора поисковой системы предприятия.
12. Определив, что перенастроенное собрание доступно для поиска в поисковой системе предприятия, удалите исходное собрание из Portal Search Engine.
13. Необязательно: Как администратор WebSphere Portal, выполните следующие действия, если хотите разрешить пользователям выполнять поиск в перенастроенном собрании с портала в WebSphere Portal.
  - a. Внедрите портлет поисковой системы предприятия в WebSphere Portal.  
В кластере сервера WebSphere Portal это можно сделать с сервера, где установлен менеджер внедрения WebSphere Application Server. Менеджер внедрения распространяет портлет поисковой системы предприятия на другие серверы в кластере сервера WebSphere Portal.
  - b. Добавьте портлет поисковой системы предприятия на подходящие страницы портала.

В WebSphere Portal управление доступом для портлета поиска моделируется по доступу к конкретным страницам и портлетам. Параметры собрания перенастраиваются, но портлет поиска должен быть установлен вручную администратором сервера WebSphere Portal.

---

## Перенос параметров собраний

При перенастройке собраний из IBM WebSphere Portal мастер по перенастройке создает параметры по умолчанию для собраний и искателей.

Если один и тот же параметр существует и в собраниях Portal Search Engine, и в собраниях поисковой системы предприятия, при перенастройке собрания в поисковую систему предприятия мастер использует параметр из Portal Search Engine. Для параметров, существующих только в поисковой системе предприятия, мастер использует значения, задаваемые вами при перенастройке собрания или значения, используемые для собраний поисковой системы предприятия по умолчанию.

### Параметры, существующие в Portal Search Engine и в поисковой системе предприятия

Мастер по перенастройке переносит для каждого перенастраиваемого собрания следующие параметры:

- Сайты Portal Search Engine в собрании Portal Search Engine
- Язык собрания
- Таксономию (или дерево категорий) и правила для категорий на основе правил, если в собрании поисковой системы предприятия используется категоризация на основе правил

Каждый сайт Portal Search Engine в собрании объединяется в искатель Web поисковой системы предприятия. Мастер по перенастройке переносит следующие параметры искателя:

- Начальные URL
- Число параллельных процессов искателя
- Глубину просмотра
- Срок ожидания (в секундах) при получении документа
- Набор символов по умолчанию
- Правила для просмотра сайтов искателем (правила включений и исключений)

### Параметры, существующие только в поисковой системе предприятия

При перенастройке собрания нужно задать информацию о нем. Мастер по перенастройке перенастраивает эти параметры и использует параметры по умолчанию для собраний в поисковой системе предприятия, конфигурируя каждое перенастраиваемое вами собрание.

Конфигурации собраний и искателей Web можно изменить при помощи консоли администратора поисковой системы предприятия. В круглых скобках () показаны значения по умолчанию для перенастраиваемых данных.

- Имя собрания
- Стратегия статического ранжирования документов
- Тип используемой категоризации (основе правил или ничего)

- Используется ли кэш поиска и сколько ответов на запросы может в нем храниться (да, 5000)
- Выполняется ли мониторинг времени отклика поиска и генерируется ли оповещение при превышении предела (да, 5 секунд)
- Используется ли управление доступом (нет)
- Расписание для дельта-построений индексов
- Расписание для полного построения индекса
- Уровень детализации журнала (все сообщения)

Кроме того, мастер по перенастройке создает для каждого искателя Web следующие параметры:

- Имя искателя
- Описание искателя
- MPL (maximum page length - максимальная длина страницы)
- Параметры защиты документов
- Типы MIME (Multipurpose Internet Mail Extension) для документов, с которыми должен работать искатель (если это применимо к типу источника данных)

Прежде чем запускать перенастроенный искатель Web, проверьте все свойства этого искателя и параметры пространства для искателя и убедитесь, что заданы все необходимые значения (обязательные поля отмечены красной звездочкой). В частности, убедитесь, что заданы адрес электронной почты для получения комментариев об искателе и имя пользовательского агента для искателя. Если при конфигурировании свойств искателя Web потребуется помощь, нажмите кнопку **Справка**.

---

## Файл журнала мастера по перенастройке

Мастер по перенастройке записывает все сообщения в файл WpsMigratorLog.log, находящийся в его каталоге установки.

Для каждого перенастроенного собрания файл журнала WpsMigratorLog.log содержит значения для всех параметров, прочитанных из WebSphere Portal Search Engine, и указывает, в какое место собраний поисковой системы предприятия они были импортированы.

---

## Команды, коды возврата и ID сеансов поисковой системы предприятия

Команды можно использовать для диагностики ошибок, определения состояния различных частей системы, запуска и остановки сеансов, а также запуска и остановки системы.

В установке с несколькими серверами можно запускать команды с любого сервера системы. Однако следует запускать команды с сервера индексов. Сервер индексов или сервер контроллера может обращаться к информации всех серверов в системе.

Формат большинства команд:

```
esadmin имя_команды аргументы
esadmin ID_сеанса действие -опция
```

Чтобы получить дополнительную информацию обо всех командах, введите `esadmin help`. Чтобы получить дополнительную информацию конкретной команде, введите `esadmin действие help`.

### Команды `esadmin` поисковой системы предприятия

Вводите следующие команды в одну строку.

Таблица 9. Команды `esadmin` поисковой системы предприятия

Команда	Описание
<code>esadmin system startall</code>	Запускает компоненты поисковой системы предприятия на всех ее серверах, включая Web-сервер, прикладную программу <code>ESSearchServer</code> и информационный центр на серверах поиска; сеансы искателя на сервере искателя и сеансы индексации на сервере индексов. Запускает слой общей связи (CCL) только на локальном сервере. Для перезапуска CCL необходимо вручную остановить и перезапустить CCL на каждом удаленном сервере поисковой системы предприятия.  <b>Пример команды:</b> <code>esadmin system startall</code>
<code>esadmin system stopall</code>	Останавливает компоненты поисковой системы предприятия на всех ее серверах, включая информационный центр, прикладную программу <code>ESSearchServer</code> и Web-сервер на серверах поиска; сеансы искателя на сервере искателя и сеансы индексации на сервере индексов. Останавливает CCL только на локальном сервере. Для перезапуска CCL необходимо вручную остановить и перезапустить CCL на каждом удаленном сервере поисковой системы предприятия.  <b>Пример команды:</b> <code>esadmin system stopall</code>
<code>esadmin system checkall</code>	Проверяет состояние всех компонентов поисковой системы предприятия на всех ее серверах.  <b>Пример команды:</b> <code>esadmin system checkall</code>

Таблица 9. Команды **esadmin** поисковой системы предприятия (продолжение)

Команда	Описание
<code>esadmin id_сеанса_искателя start</code>	<p>Запускает сеанс искателя. Эта команда не запускает никаких действий искателя.</p> <p><b>Пример команды:</b>  <code>esadmin col1.WEB1.esadmin start</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5310I WEB1Crawler1 (sid: col1.WEB1.esadmin)            не работает.            FFQC5314I Результат: 0</p>
<code>esadmin id_сеанса_искателя startCrawl</code>	<p>Запускает просмотр.</p> <p><b>Пример команды:</b>  <code>esadmin col3.DB21.esadmin startCrawl</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)            уже работает. PID: 23650            FFQC5314I Результат: 0</p>
<code>esadmin id_сеанса_искателя pause</code>	<p>Приостанавливает просмотр.</p> <p><b>Пример команды:</b>  <code>esadmin col3.DB21.esadmin pause</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)            уже работает. PID: 23650            FFQC5314I Результат: 0</p>
<code>esadmin id_сеанса_искателя resume</code>	<p>Возобновляет просмотр.</p> <p><b>Пример команды:</b>  <code>esadmin col3.DB21.esadmin resume</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)            уже работает. PID: 23650            FFQC5314I Результат: 0</p>
<code>esadmin id_сеанса_искателя stopCrawl</code>	<p>Останавливает просмотр.</p> <p><b>Пример команды:</b>  <code>esadmin col3.DB21.esadmin stopCrawl</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)            уже работает. PID: 23650            FFQC5314I Результат: 0</p>
<code>esadmin id_сеанса_искателя stop</code>	<p>Останавливает сеанс искателя.</p> <p><b>Пример команды:</b>  <code>esadmin col3.DB21.esadmin stop</code></p> <p><b>Пример сообщений и кодов возврата:</b>            FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)            уже работает. PID: 23650            FFQC5314I Результат: 0</p>

Таблица 9. Команды **esadmin** поисковой системы предприятия (продолжение)

Команда	Описание
esadmin <i>id_сеанса_искателя</i> getCrawlerStatus	<p>Получает состояние искателя. Возвращаемая информация зависит от того, является ли искатель искателем Web или же иным искателем.</p> <p><b>Пример для искателя Web:</b></p> <pre>esadmin col1.WEB1.esadmin getCrawlerStatus</pre> <p><b>Возможные коды возврата и сообщения для искателя Web:</b></p> <pre>FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) уже работает. PID: 23650</pre> <p><b>Пример для другого искателя (не Web):</b></p> <pre>esadmin col3.DB21.esadmin getCrawlerStatus</pre> <p><b>Возможные коды возврата и сообщения для другого искателя (не Web):</b></p> <pre>FFQC5303I db2crawler (sid: db2col.DB2_96945) уже работает. PID: 5936</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 378.</p>
esadmin <i>id_сеанса_искателя</i> getCrawlSpaceStatus  esadmin <i>id_сеанса_искателя_web</i> getCrawlStatus -selections <i>значение</i>	<p>Получает общее состояние пространства для любого искателя, за исключением искателя Web.</p> <p><b>Пример команды:</b></p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatus</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650</pre> <p>Получает общее состояние пространства для искателя Web.</p> <p><b>Пример команды:</b></p> <pre>esadmin col1.WEB1.esadmin getCrawlStatus</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 378.</p>

Таблица 9. Команды **esadmin** поисковой системы предприятия (продолжение)

Команда	Описание
<pre>esadmin id_сеанса_искателя getCrawlSpaceStatusDetail -ts id_сервера_назначения esadmin id_сеанса_искателя_Web getCrawlDetailsPerSite -url строка -selections число -threshold число</pre>	<p>Получает подробное состояние пространства для любого искателя, за исключением искателя Web. Если не задать сервер назначения, возвращаются данные для всех серверов назначения. Например, если искатель DB2 просматривает базы данных FOUNTAIN и SAMPLE и вы не задали опцию сервера назначения, возвращается состояние всех таблиц баз данных FOUNTAIN и SAMPLE.</p> <p><b>Пример команды:</b></p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650</pre> <p>Получает подробное состояние пространства для искателя Web.</p> <p><b>Пример команды:</b></p> <pre>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 378.</p>
<pre>esadmin monitor getCollectionParserMonitorStatus -cid ID_собрания</pre>	<p>Получает состояние анализатора.</p> <p><b>Пример команды:</b></p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12543</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 378.</p>
<pre>esadmin startMain -cid collection_id</pre>	<p>Запускает полное построение индекса.</p> <p><b>Пример команды:</b></p> <pre>esadmin startMain -cid coll</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 25917 FFQC5314I Результат: 1117671147056</pre>
<pre>esadmin startDelta -cid id_собрания</pre>	<p>Запускает дельта-построение индекса.</p> <p><b>Пример команды:</b></p> <pre>esadmin startDelta -cid coll</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 4548 FFQC5314I Результат: 1117670603408</pre>

Таблица 9. Команды **esadmin** поисковой системы предприятия (продолжение)

Команда	Описание
<pre>esadmin controller startIndexBuild -cid ID_собрания -buildType тип -detectChanges</pre>	<p>Запустить построение главного или разностного индекса и задать, что построение должно выполняться только тогда, когда не обнаружено изменений, которые должны быть включены в индекс.</p> <p><b>Пример команды:</b></p> <pre>esadmin controller startIndexBuild -cid col_1 -buildType main -detectChanges</pre>
<pre>esadmin monitor getCollectionIndexMonitorStatus -cid ID_собрания -buildType [main   delta] -numrecords последние_N_записей</pre>	<p>Получает состояние полного или дельта-построения индекса. Опция numrecords выводит последние <i>N</i> записей состояния построения индекса. Если опция numrecords опущена, выводится 20 последних записей построения индекса.</p> <p><b>Пример команды:</b></p> <pre>esadmin monitor getCollectionIndexMonitorStatus -cid col1 -buildType main -numrecords 4</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12649</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 378.</p>
<pre>esadmin startSearch -cid id_собрания</pre>	<p>Запускает процессы сервера поиска.</p> <p><b>Пример команды:</b></p> <pre>esadmin startSearch -cid col1</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 25917 FFQC5314I Результат: 0</pre>
<pre>esadmin stopSearch -cid id_собрания</pre>	<p>Останавливает процессы сервера поиска.</p> <p><b>Пример команды:</b></p> <pre>esadmin stopSearch -cid col1</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 15292 FFQC5314I Результат: 0</pre>

Таблица 9. Команды **esadmin** поисковой системы предприятия (продолжение)

Команда	Описание
<pre>esadmin monitor getCollectionSearchMonitorStatus -cid ID_собрания esadmin id_сеанса_сервера_поиска getStatus -cid id_собрания</pre>	<p>Получает состояние сервера поиска.</p> <p><b>Пример команды:</b></p> <pre>esadmin monitor getCollectionSearchMonitorStatus -cid coll</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12649</pre> <p>Возвращает подробную информацию о состоянии индексов поиска для собрания на заданном сервере поиска. На каждом сервере поиска есть один сеанс менеджера поиска. Каждый сеанс менеджера поиска отвечает за отслеживание и работу индексов поиска на определенном сервере поиска.</p> <p><b>Пример команды:</b></p> <pre>esadmin searchmanager.node1 getStatus -cid coll</pre> <p><b>Пример сообщений и кодов возврата:</b></p> <pre>FFQC5303I Менеджер поиска (node1) (sid: searchmanager.node1) уже работает. PID: 15711 FFQC5314I Результат: PID CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния”.</p>

## Подробная информация для команд состояния

Некоторые команды возвращают обширную информацию. В этом разделе описана информация о состоянии искателя и пространства искателя, которая может быть возвращена. Таблица в разделе “Команды **esadmin** поисковой системы предприятия” на стр. 373 содержит возможную возвращаемую информацию для каждой команды **esadmin**. В этом разделе описана возвращаемая информация для следующих команд:

- Состояние искателя Web
- Состояние других искателей (кроме искателя Web)
- Состояние пространства для искателя Web
- Состояние пространства для других искателей (кроме искателя Web)
- Подробное состояние пространства для искателя Web
- Подробное состояние пространства для других искателей (кроме искателя Web)
- Состояние анализатора
- Состояние построения индекса
- Состояние сервера поиска
- Подробное состояние сервера поиска

**Состояние искателя Web:** Когда вы вводите команду для получения состояния искателя Web, эта команда возвращает информацию в формате документа XML. Команда состояния искателя Web возвращает следующую информацию:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<CrawlerStatus>
<CrawlerRunLevel Value="Running"/>
<CrawlerThreadStateDist Count="4" Total="200">
<CrawlerThreadState State="FETCHING" Count="100"/>
. . .
<CrawlerThreadState State="FETCHING" Count=100>
<ActiveBucketList Count="500">
<ActiveBucket URL="http://w3.ibm.com/"
NumActURLs="355"
NumProcURLs="350"
TimeRem="5" Duration="1195"/>
. . .
</ActiveBucketList>
<CrawlRate Value="75"/>
<RecentlyCrawledURLList Count="40">
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<NumURLsThisSession Value="160000"/>
</CrawlerStatus>
```

Ниже в таблице описаны все элементы XML и их возможные атрибуты, возвращаемые командой состояния искателя Web:

Таблица 10. Информация состояния искателя Web

Элемент	Атрибуты	Описание
CrawlerStatus	<ul style="list-style-type: none"> <li>CrawlerThreadStateDist</li> <li>ActiveBucketList</li> <li>CrawlRate</li> <li>RecentlyCrawledURLList</li> <li>NumURLsThisSession</li> </ul>	Состояние искателя.
Значение CrawlerRunLevel	<ul style="list-style-type: none"> <li>Строка (английский язык) “Not started”: Сеанс искателя существует, но он еще не получил сообщения запуска, начинающее обработку документов.</li> <li>“Started”: Искатель запущен.</li> <li>“Running”: Искатель завершил инициализацию и запуск, и теперь активно просматривает документы.</li> <li>“Paused”: Искатель получил указание приостановить активный просмотр, но не завершать работу.</li> <li>“Stopping”: Искатель получило сигнал остановки и собирается остановиться.</li> <li>“Error”: Искатель в невосстановимом состоянии; его надо остановить и перезапустить, чтобы продолжить просмотр.</li> </ul>	Информация о текущих действиях искателя.
Состояние CrawlerThreadState	Строка (английский язык)	Действия потока искателя. В это поле показано, что делают потоки.

Таблица 10. Информация состояния искателя Web (продолжение)

Элемент	Атрибуты	Описание
ActiveBucket	<ul style="list-style-type: none"> <li>• URL: Строка (спецификация URL) Протокол, хост и номер порта для просматриваемых URL.</li> <li>• NumActURLs: Целое (положительное) Число URL в комплекте в момент, когда он стал доступным для просмотра (был активирован).</li> <li>• NumProcURLs: Целое (неотрицательное) Число обработанных URL в комплекте (просмотренных или отвергнутых)</li> <li>• TimeRem: Целое Оставшийся срок в секундах из времени, отведенного для просмотра комплекта.</li> <li>• Duration: Целое (неотрицательное) Время в секундах, прошедшее после активации комплекта.</li> </ul>	Текущее действие на указанном сайте.
CrawlRate	<p>Значение: Целое (неотрицательное)</p> <p>Скорость просмотра в страницах в секунду (для всех комплектов)</p>	Показатель пропускной способности искателя.
RecentlyCrawledURL	<p>URL: Строка (спецификация URL)</p> <p>Строка задает протокол, хост, порт и имя просмотренного файла.</p>	Последняя просмотренная страница.
NumURLsThisSession	Значение: Целое (неотрицательное)	Число URL, просмотренных с момента начала просмотра этим экземпляром искателя (процессом).

**Состояние другого искателя (кроме искателя Web):** Когда вы вводите команду для получения состояния другого искателя (кроме искателя Web), эта команда возвращает информацию в формате документа XML. Для других искателей (кроме искателя Web) команда **getCrawlerStatus** возвращает следующую информацию:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<GeneralStatus>
<Status>0</Status>
<StatusMessage>Idle</StatusMessage>
<NumberOfServers>1</NumberOfServers>
<NumberOfCompletedServers>1</NumberOfCompletedServers>
<NumberOfTargets>3</NumberOfTargets>
<NumberOfCompletedTargets>3</NumberOfCompletedTargets>
<NumberOfCrawledRecords>115</NumberOfCrawledRecords>
<RunningThreads>0</RunningThreads>
</GeneralStatus>
```

Ниже в таблице описаны элементы XML и атрибуты для каждого из искателей поисковой системы предприятия, кроме искателя Web. Эту информацию возвращает команда состояния искателя.

Таблица 11. Информация о состоянии для искателей NNTP, DB2, баз данных JDBC и Notes

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)

Таблица 11. Информация о состоянии для искателей NNTP, DB2, баз данных JDBC и Notes (продолжение)

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число серверов NNTP в пространстве для искателя.	Число баз данных в пространстве для искателя.	Число баз данных в пространстве для искателя.
NumberOfCompletedServers	Число просмотренных серверов NNTP.	Число просмотренных баз данных.	Число просмотренных баз данных.
NumberOfTargets	Число групп новостей в пространстве для искателя.	Число баз данных в пространстве для искателя.	Число представлений и папок в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных групп новостей.	Число просмотренных таблиц.	Число просмотренных представлений и папок.
NumberOfCompletedRecords	Число просмотренных статей.	Число просмотренных записей.	Число просмотренных документов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Таблица 12. Информация о состоянии для искателей Exchange Server, DB2 Content Manager и Content Edition

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число серверов Exchange Server в пространстве для искателя.	Число серверов Content Manager в пространстве для искателя.	Число репозиторий в пространстве для искателя.
NumberOfCompletedServers	Число просмотренных серверов Exchange Server.	Число просмотренных серверов Content Manager.	Число просмотренных репозиторий.
NumberOfTargets	Число подпапок в пространстве для искателя.	Число типов элементов в пространстве для искателя.	Число классов в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных подпапок.	Число просмотренных типов элементов.	Число просмотренных классов элементов.
NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных документов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Таблица 13. Информация о состоянии для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка

Таблица 13. Информация о состоянии для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
NumberOfServers	Число площадок в пространстве для искателя.	Число библиотек в пространстве для искателя.	Постоянное значение - 1.
NumberOfCompletedServers	Число просмотренных площадок.	Число просмотренных библиотек.	0 или 1, если просмотрены все подкаталоги.
NumberOfTargets	Число баз данных на площадках и в комнатах в пространстве для искателя.	Число шкафов в пространстве для искателя.	Число подкаталогов в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных баз данных на площадках и в комнатах в пространстве для искателя.	Число просмотренных шкафов.	Число просмотренных подкаталогов.
NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных файлов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Таблица 14. Информация о состоянии искателей WebSphere Portal и Web Content Management

Имя элемента и атрибута	Искатель WebSphere Portal	Искатель Web Content Management
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число серверов в пространстве для искателя.	Число сайтов в пространстве для искателя.
NumberOfCompletedServers	Число просмотренных серверов.	Число просмотренных сайтов.
NumberOfTargets	Число серверов в пространстве для искателя.	Число сайтов в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных серверов.	Число просмотренных сайтов.
NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.
RunningThreads	Число потоков искателя.	Число потоков искателя.

**Состояние пространства для искателя Web:** Когда вы вводите команду для получения состояния пространства для искателя Web, эта команда возвращает информацию в формате документа XML. Команда состояния пространства для искателя Web возвращает следующую информацию:

Таблица 15. Значения маски выбора для команды состояния пространства для искателя Web

Бит маски	Выбирает
1	Число страниц на складе необработанных данных.
2	Число обнаруженных сайтов.
4	Число сайтов с записями DNS.
8	Число сайтов без записей DNS.
16	Число обнаруженных URL.
32	Число сохраненных уникальных страниц.
64	Число просмотренных URL.

Таблица 15. Значения маски выбора для команды состояния пространства для искателя Web (продолжение)

Бит маски	Выбирает
128	Число непросмотренных URL.
256	Число просроченных URL.
512	Распределение кодов состояния HTTP.

Все значения представляют собой накопленные суммы для всех сеансов, использующих текущую внутреннюю базу данных:

```
<CrawlStatus>
 <NumPagesInRDS Value="5422386"/>
 <NumSitesDiscovered Value="15332"/>
 <NumSitesWithDNS Value="14832"/>
 <NumSitesWithoutDNS Value="500"/>
 <NumURLsDiscovered Value="15222999"/>
 <NumUniquePagesSaved Value="6234789"/>
 <NumURLsCrawled Value="7800422"/>
 <NumURLsUncrawled Value="7422577"/>
 <NumURLsOverdue Value="14000"/>
 <HTTPCodeDist Count="4" Total="1031000"/>
 <HTTPCode Code="200" Count="1000000"/>
 <HTTPCode Code="301" Count="1000"/>
 <HTTPCode Code="404" Count="10000"/>
 <HTTPCode Code="780" Count="20000"/>
 </HTTPCode Code="780" Count="20000">
</CrawlStatus>
```

Возвращаемые данные могут содержать следующие элементы (возможно, все или же никакие из них):

Таблица 16. Информация о состоянии пространства для искателя Web

Элемент	Атрибут	Описание
CrawlerStatus	<ul style="list-style-type: none"> <li>NumPagesInRDS</li> <li>NumSitesDiscovered</li> <li>NumSitesWithDNS</li> <li>NumSitesWithoutDNS</li> <li>NumURLsDiscovered</li> <li>NumUniquePagesSaved</li> <li>NumURLsCrawled</li> <li>NumURLsUncrawled</li> <li>NumURLsOverdue</li> <li>HTTPCodeDist</li> </ul>	Информация, которую можно быстро получить из суммарного состояния искателя (по всем сеансам).
NumPagesInRDS	<p>Значение: Неотрицательное целое</p> <p>Сколько страниц сейчас находятся в кэше склада необработанных данных (raw data store, RDS) - только для данного искателя.</p>	Насколько заполнен склад необработанных данных (RDS) - только за счет вклада данного искателя.
NumSitesDiscovered	<p>Значение: Неотрицательное целое</p> <p>Сколько хостов обнаружено при просмотре (или из начальных URL).</p>	Мера покрытия искателем просматриваемого домена (число хостов)

Таблица 16. Информация о состоянии пространства для искателя Web (продолжение)

Элемент	Атрибут	Описание
NumSitesWithDNS	Значение: Неотрицательное целое  У скольких хостов есть связанные с ними IP-адреса (разрешенные искателем в фоновом режиме).	Показатель того, насколько эффективно искатель может получать IP-адреса для хостов, обнаруженных по именам DNS в URL.
NumSitesWithoutDNS	Значение: Неотрицательное целое  У скольких хостов нет связанных с ними IP-адресов (разрешенных искателем в фоновом режиме).	Показатель того, насколько эффективно искатель может получать IP-адреса для хостов, обнаруженных по именам DNS в URL.
NumURLsDiscovered	Значение: Неотрицательное целое  Сколько уникальных URL посетил искатель.	Мера покрытия искателем просматриваемого домена (число URL).
NumUniquePagesSaved	Значение: Неотрицательное целое  Сколько уникальных страниц записано на RDS для дальнейшей обработки другими компонентами поисковой системы предприятия.	Вклад данного искателя в размер индекса.
NumURLsCrawled	Значение: Неотрицательное целое  Сколько уникальных URL просмотрел искатель.	Показатель способности искателя обрабатывать данные от начала до конца. Это число отличается от числа страниц, записанных на RDS, поскольку не все просмотренные страницы записываются на RDS.
NumURLsOverdue	Значение: Неотрицательное целое  Сколько уникальных URL подлежат повторному просмотру.	Показатель способности искателя обрабатывать пространство Web.

**Состояние пространства для других искателей (кроме искателя Web):** Когда вы вводите команду для получения состояния пространства для другого искателя (не искателя Web), эта команда возвращает информацию в формате документа XML. Для других искателей (кроме искателя Web) команда **getCrawlSpaceStatus** возвращает следующую информацию:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<ServerStatus>
 <Server Name ="FOUNTAIN">
 <Status>5</Status>
 <StatusMessage>Scheduled</StatusMessage>
 <NumberOfTargets>1</NumberOfTargets>
 <NumberOfCompletedTargets>1</NumberOfCompletedTargets>
 <NumberOfErrors>0</NumberOfErrors>
 <StartTime>1118354510512</StartTime>
 <EndTime>1118354514386</EndTime>
 <ScheduleConfigured>2</ScheduleConfigured>
 <ScheduleTime>1118393377000</ScheduleTime>
 <TotalTime>3874</TotalTime>
 </Server>
</ServerStatus>
```

Ниже в таблице описаны элементы XML и атрибуты для каждого из искателей поисковой системы предприятия, кроме искателя Web. Эту информацию возвращает команда состояния пространства искателя. Для искателей Notes при уровне сводки 0 Server@Name - это имя сервера + имя базы данных. Если уровень сводки - 1, Server@Name - это имя сервера + имя каталога.

Таблица 17. Информация о состоянии пространства для искателей NNTP, DB2, баз данных JDBC и Notes

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
Server@Name	Имя сервера News	Имя базы данных	Имя базы данных или каталога
Server/Status	Состояние: (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>
Server/NumberOfTargets	Число групп новостей в пространстве для искателя.	Число баз данных в пространстве для искателя.	Число представлений и папок или каталогов в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных групп новостей.	Число просмотренных таблиц.	Число просмотренных представлений и папок или каталогов.
Server/NumberOfErrors	Число ошибок.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>• 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>• 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>• 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>• 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>• 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>• 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.

Таблица 17. Информация о состоянии пространства для искателей NNTP, DB2, баз данных JDBC и Notes (продолжение)

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.
Server/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0, 1: <ul style="list-style-type: none"> <li>0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)</li> <li>1: Искатель Notes просматривает документы в режиме каталогов.</li> </ul>

Таблица 18. Информация о состоянии пространства для искателей Exchange Server, DB2 Content Manager и Content Edition

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Server@Name	Имя сервера Exchange Server.	Серверы DB2 Content Manager.	Имя репозитория.
Server/Status	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>
Server/NumberOfTargets	Число подпапок в пространстве для искателя.	Число типов элементов в пространстве для искателя.	Число классов элементов в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных подпапок.	Число просмотренных типов элементов.	Число просмотренных классов элементов.
Server/NumberOfErrors	Число ошибок.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.

Таблица 18. Информация о состоянии пространства для искателей Exchange Server, DB2 Content Manager и Content Edition (продолжение)

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.
Server/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.

Таблица 19. Информация о состоянии пространства для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Server@Name	Каталог площадки.	База данных библиотеки.	Постоянное значение - localhost.
Server/Status	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>5: Запланирован</li> <li>-1: Ошибка</li> </ul>

Таблица 19. Информация о состоянии пространства для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Server/NumberOfTargets	Число баз данных на площадках и в комнатах в пространстве для искателя.	Число шкафов в пространстве для искателя.	Число подкаталогов в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных баз данных на площадках и в комнатах в пространстве для искателя.	Число просмотренных шкафов.	Число подкаталогов в пространстве для искателя.
Server/NumberOfErrors	Число ошибок.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.
Server/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.

Таблица 20. Информация о состоянии пространства для искателей WebSphere Portal и Web Content Management

Имя элемента и атрибута	Искатель WebSphere Portal	Искатель Web Content Management
Server@Name	Сервер WebSphere Portal	Начальный URL поиска Web Content Management

Таблица 20. Информация о состоянии пространства для искателей WebSphere Portal и Web Content Management (продолжение)

Имя элемента и атрибута	Искатель WebSphere Portal	Искатель Web Content Management
Server/Status	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• 5: Запланирован</li> <li>• -1: Ошибка</li> </ul>
Server/NumberOfTargets	Число серверов в пространстве для искателя.	Число сайтов в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных серверов.	Число просмотренных сайтов.
Server/NumberOfErrors	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>• 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>• 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>	0, 1, 2 <ul style="list-style-type: none"> <li>• 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию.</li> <li>• 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено.</li> <li>• 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.</li> </ul>
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.
Server/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.

**Подробное состояние пространства для искателя Web:** Когда вы вводите команду для получения подробного состояния пространства для искателя Web, эта команда возвращает информацию в формате документа XML. Команда подробного состояния пространства для искателя возвращает следующую информацию:

Таблица 21. Значения маски выбора для команды подробного состояния пространства для искателя Web

Бит маски	Выбирает
1	Число страниц на складе необработанных данных.
2	Число обнаруженных сайтов.
4	Число сайтов с записями DNS.
8	Число сайтов без записей DNS.
16	Число обнаруженных URL.
32	Число сохраненных уникальных страниц.
64	Число просмотренных URL.
128	Число непросмотренных URL.
256	Число просроченных URL.
512	Распределение кодов состояния HTTP.

Пример возвращаемой информации:

```
<CrawlDetailsPerSite>
 <Site URL=http://w3.ibm.com/">
 <NumURLsDiscovered Value="5422386"/>
 <NumURLsOverdue Value="15332"/>
 <NumURLsCrawled Value="15332"/>
 <NumURLsUncrawled Value="15332"/>
 <NumURLsOverdueBy Threshold="604800" Value="14832"/>
 <NumURLsActivated Value="2200"/>
 <LastActivationTime Value="1076227340"/>
 <LastActivationDuration Value="4300"/>
 <IPAddressList Count="1"/>
 <IPAddress Value="9.205.41.33"/>
 </IPAddressList>
 <RobotsContent>
 содержимое файла robots. . .
 </RobotsContent>
 <HTTPCodeDist Count="4" Total="1031000"/>
 <HTTPCode Code="200" Count="1000000"/>
 <HTTPCode Code="301" Count="1000"/>
 <HTTPCode Code="404" Count="10000"/>
 <HTTPCode Code="780" Count="20000"/>
 </HTTPCodeDist>
</CrawlDetailsPerSite>
```

Ниже в таблице описано каждое поле, возвращаемой командой подробного состояния пространства для искателя Web:

Таблица 22. Подробная информация о состоянии пространства для искателя Web

Элемент	Атрибуты	Описание
CrawlDetailsPerSite	<ul style="list-style-type: none"> <li>LastActivationTime:</li> <li>LastActivationDuration:</li> <li>IPAddressList:</li> <li>RobotsContent:</li> <li>HTTPCodeDist:</li> </ul>	Информация, которую можно быстро получить для подробного состояния одного сайта.
Сайт	URL	URL корневой страницы сайта.
NumURLsDiscovered	Значение	Число URL, обнаруженных на сайте.
NumURLsOverdue	Значение	Число URL, подлежащих повторному просмотру на этом сайте.

Таблица 22. Подробная информация о состоянии пространства для искателя Web (продолжение)

Элемент	Атрибуты	Описание
NumURLsCrawled	Значение	Число URL, просмотренных на сайте.
NumURLsUncrawled	Значение	Число URL, еще не просмотренных на этом сайте.
NumURLsOverdueBy	Порог, Значение: Целое (положительное или отрицательное)  Это значение относится к числу URL, подлежащих повторному просмотру. Порог задает интервал времени, который URL ожидают повторного просмотра. Порог измеряется в секундах от настоящего времени. Если порог отрицателен, это означает, что повторный просмотр URL просрочен. Если порог положителен, это означает, что повторный просмотр URL должен быть выполнен в будущем.	Число URL, для которых повторный просмотр затребован в некотором интервале до настоящего момента или будет затребован в некотором интервале после настоящего момента.
NumURLsActivated	Значение	Число URL, помещенных в память во время последнего просмотра сайта и объявленных доступными для потоков искателя.
LastActivationTime	Значение	Интервал в секундах с момента, когда URL данного сайта последний раз заносились в память.
LastActivationDuration	Значение	Интервал в секундах, когда URL данного сайта последний раз находились в памяти и были доступными для потоков искателя.
IPAddressList	IP-адрес	Все известные IP-адреса с хоста сервера данного сайта.
IP-адрес	Значение	Адрес в формате IPv4 с точками для хоста сервера данного сайта.
RobotsContent	Текст	Текст из файла robots, если он есть.
HTTPCodeDist	Код HTTP	Распределение кодов HTTP для попыток загрузки с данного сайта.
Код HTTP	Код: Целое  Код состояния HTTP или иной внутренний код.	Сколько раз определенный код состояния HTTP был получен при просмотре данного сайта.

**Подробное состояние пространства для других искателей (кроме искателя Web):** Когда вы вводите команду для получения подробного состояния пространства для другого искателя (не искателя Web), эта команда возвращает информацию в формате документа XML. Для других искателей (кроме искателя Web) команда **getCrawlSpaceStatusDetail** возвращает следующую информацию:

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<TargetStatus>
 <Target Name ="escmgr.crawlerinstances">
 <Status>2</Status>
 <StatusMessage>Completed</StatusMessage>
 <NumberOfRecords></NumberOfRecords>
 <NumberOfCrawledRecords>117</NumberOfCrawledRecords>
 <NumberOfInsertedRecords>21</NumberOfInsertedRecords>
```

```

<NumberOfUpdatedRecords>45</NumberOfUpdatedRecords>
<StartTime>1118354510727</StartTime>
<EndTime>1118354514386</EndTime>
<AggregationLevel>0<AggregationLevel>
<Target>
</TargetStatus>

```

Таблица 23. Подробная информация о состоянии пространства для искателей NNTP, DB2, баз данных JDBC и Notes

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
Target@Name	Имя группы новостей	Имя таблицы	Имя представления или папки
Target@CrawlType	Неприменимо.	0,1 (DB2); 0 (база данных JDBC) <ul style="list-style-type: none"> <li>0: Активный просмотр (нормальный режим)</li> <li>1: Пассивный просмотр (публикация событий DB2)</li> </ul>	0
Target/Status	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>
Target/NumberOfRecords	Номер последней статьи на сервере.	Число просмотренных записей.	Число просмотренных документов.
Target/NumberOfCompletedRecords	Число просмотренных статей.	Число просмотренных записей.	Число просмотренных документов.
Target/NumberOfInsertedRecords	Число новых опубликованных статей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOfUpdatedRecords	Неприменимо.	Число измененных записей.	Число измененных записей.
Target/NumberOfDeletedRecords	Неприменимо.	Число удаленных записей.	Число удаленных записей.
Target/StartTime	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.

Таблица 23. Подробная информация о состоянии пространства для искателей NNTP, DB2, баз данных JDBC и Notes (продолжение)

Имя элемента и атрибута	Искатель NNTP	Искатели DB2 и JDBC	Искатель Notes
Target/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0, 1: <ul style="list-style-type: none"> <li>0: Искатель просматривает документы в нормальном режиме.</li> <li>Искатель просматривает документы в режиме каталога.</li> </ul>
Target/LastUpdatedTime	Неприменимо.	Время последнего изменения: <ul style="list-style-type: none"> <li>0: Активный просмотр (нормальный режим)</li> <li>1: Пассивный просмотр (публикация событий DB2)</li> </ul>	Неприменимо.
Target/LastResetTime	Неприменимо.	Время последнего сброса статистики: <ul style="list-style-type: none"> <li>0: Активный просмотр (нормальный режим)</li> <li>1: Пассивный просмотр (публикация событий DB2)</li> </ul>	Неприменимо.

Таблица 24. Подробная информация о состоянии пространства для искателей Exchange Server, DB2 Content Manager и Content Edition

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Target@Name	Имя подпапки	Имя типа элементов	Имя класса элементов
Target@CrawlType	0	0	0
Target/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>0: Не просмотрен</li> <li>1: Идет просмотр</li> <li>2: Завершен (и не запланирован)</li> <li>3: Ожидание</li> <li>4: Приостановлен</li> <li>-1: Ошибка</li> </ul>
Target/NumberOfRecords	Неприменимо.	Неприменимо.	Неприменимо.
Target/NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных документов.

Таблица 24. Подробная информация о состоянии пространства для искателей Exchange Server, DB2 Content Manager и Content Edition (продолжение)

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Target/NumberOf InsertedRecords	Число вставленных записей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOf UpdatedRecords	Неприменимо.	Число измененных записей.	Число измененных записей.
Target/NumberOf DeletedRecords	Неприменимо.	Число удаленных записей.	Число удаленных записей.
Target/StartTime	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.
Target/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.
Target/LastUpdatedTime	Неприменимо.	Неприменимо.	Неприменимо.
Target/LastResetTime	Неприменимо.	Неприменимо.	Неприменимо.

Таблица 25. Подробная информация о состоянии пространства для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Target@Name	Имя базы данных площадки или комнаты	Имя базы данных шкафа	Имя подкаталога
Target@CrawlType	0	0	0
Target/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>
Target/NumberOf Records	Неприменимо.	Неприменимо.	Неприменимо.
Target/NumberOf CompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных файлов.

Таблица 25. Подробная информация о состоянии пространства для искателей QuickPlace, Domino Document Manager, файловых систем UNIX и файловых систем Windows (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Target/NumberOf InsertedRecords	Число вставленных записей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOf UpdatedRecords	Число измененных записей.	Число измененных записей.	Число измененных записей.
Target/NumberOf DeletedRecords	Число удаленных записей.	Число удаленных записей.	Число удаленных записей.
Target/StartTime	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.
Target/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.
Target/LastUpdatedTime	Неприменимо.	Неприменимо.	Неприменимо.
Target/LastResetTime	Неприменимо.	Неприменимо.	Неприменимо.

Таблица 26. Подробная информация о состоянии пространства для искателей WebSphere Portal и Web Content Management

Имя элемента и атрибута	Искатель WebSphere Portal	Искатель Web Content Management
Target@Name	Имя сервера WebSphere Portal	Начальный URL поиска, представляющий этот сайт
Target@CrawlType	0	0
Target/Status	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>	<ul style="list-style-type: none"> <li>• 0: Не просмотрен</li> <li>• 1: Идет просмотр</li> <li>• 2: Завершен (и не запланирован)</li> <li>• 3: Ожидание</li> <li>• 4: Приостановлен</li> <li>• -1: Ошибка</li> </ul>
Target/NumberOf Records	Неприменимо.	Неприменимо.
Target/NumberOf CompletedRecords	Общее число просмотренных записей.	Общее число просмотренных записей.
Target/NumberOf InsertedRecords	Число вставленных записей.	Число вставленных записей.
Target/NumberOf UpdatedRecords	Число измененных записей.	Число измененных записей.
Target/NumberOf DeletedRecords	Число удаленных записей.	Число удаленных записей.

Таблица 26. Подробная информация о состоянии пространства для искателей WebSphere Portal и Web Content Management (продолжение)

Имя элемента и атрибута	Искатель WebSphere Portal	Искатель Web Content Management
Target/StartTime	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.
Target/AggregationLevel	0: Искатель просматривает документы в нормальном режиме.	0: Искатель просматривает документы в нормальном режиме.
Target/LastUpdatedTime	Неприменимо.	Неприменимо.
Target/LastResetTime	Неприменимо.	Неприменимо.

**Состояние анализатора:** Когда вы вводите команду для получения состояния анализатора, эта команда возвращает информацию в формате документа XML. Команда состояния анализатора возвращает следующую информацию:

```
FFQC5314I Result:
<Monitor Type="Parser">
 <ParserStatus>
 <Status>1</Status>
 <State>Parsing</State>
 <NumberOfDocsToBeIndexed>231974</NumberOfDocsToBeIndexed>
 <ParseRate>0</ParseRate>
 <ParseRateMBPerHour>0</ParseRateMBPerHour>
 <NumberOfCpmThreads>3</NumberOfCpmThreads>
 <ParserServiceSession>parserservice.1</ParserServiceSession>
 </ParserStatus>
 <CrawlerStatus>
 <Name>WEBCrawler1</Name>
 <Crawlerid>coll.WEB1.esadmin</Crawlerid>
 <Type>WEB</Type>
 <ParserStatus>1</ParserStatus>
 <NumberOfDocsAlreadyParsed>29</NumberOfDocsAlreadyParsed>
 </CrawlerStatus>
</Monitor>
```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния анализатора:

Таблица 27. Элементы для команды состояния анализатора

Элемент	Описание
Состояние	<ul style="list-style-type: none"> <li>0: Сеанс анализатора для этого собрания остановлен.</li> <li>1: Сеанс анализатора для этого собрания запущен.</li> </ul>

Таблица 27. Элементы для команды состояния анализатора (продолжение)

Элемент	Описание
State	<p>Возможные состояния: Initializing (инициализируется), Idle (бездействует), Restart (перезапущен), Parsing (идет анализ), Stopped (остановлен), Paused (приостановлен), Resuming (возобновлен), NoParserServiceIsAvailable (служба анализатора недоступна).</p> <p>Состояние Initializing указывает, что анализатор запускается и инициализирует состояние.</p> <p>Состояние Idle указывает, что анализатор бездействует <i>N</i> минут в ожидании поступления документов от искателей в данном собрании. Время бездействия по умолчанию - 300 секунд.</p> <p>Состояние Restart указывает, что анализатор ждет перезапуска виртуальной Java-машины для анализа. JVM для анализа работает в отдельном сеансе и выполняет собственно обработку документов.</p> <p>Состояние Parsing указывает, что анализатор обрабатывает документы.</p> <p>Состояние Paused указывает, что анализатор приостановлен сеансом построения индекса для данного собрания.</p> <p>Состояние Resuming указывает, что анализатор переведен из состояния Paused в состояние Parsing сеансом построения индекса для данного собрания.</p> <p>Состояние NoParserServiceIsAvailable указывает, что виртуальная Java-машина для обработки документов данного собрания недоступна. Это состояние означает, что все JVM анализатора используются для других собраний.</p>
NumberOfDocsToBeIndexed	Число документов на складе для данного собрания. В это число входят также документы, помеченные для удаления при следующем построении индекса.
ParseRate	Скорость анализа в документах в секунду.
ParseRateMBPerHour	Скорость анализа в Мбайтах в час.
NumberOfCpmThreads	Число потоков CPM, используемых JVM анализатора для обработки документов данного собрания.
ParserServiceSession	Имя JVM анализатора, обрабатывающей документы для данного собрания. Это поле доступно, только если анализатора находится в состоянии Parsing.
Название	Имя искателя.
Crawlerid	ID, созданный для этого искателя системой.
Type	Тип искателя (Web, NNTP, DB2 и т.п.)
ParserStatus	<ul style="list-style-type: none"> <li>0: Документы от данного искателя не анализируются (сеанс анализатора остановлен).</li> <li>1: Документы от данного искателя анализируются (сеанс анализатора запущен).</li> </ul>
NumberofDocsAlreadyParsed	Число уже проанализированных документов от данного искателя.

**Состояние построения индекса:** Когда вы вводите команду для получения состояния построения индекса, эта команда возвращает информацию в формате документа XML. Команда состояния построения индекса возвращает следующую информацию:

```
<?xml version="1.0"?>
<Monitor Type="MainIndexHistory" Count="1">
 <IndexStatus Id="1">
 <StartTime>1131987633901<StartTime>
 <Progress>0</Progress>
 <CurrentPhase>0</CurrentPhase>
```

```

<TotalPhase>3</TotalPhase>
<IndexCopyTime>49822</IndexCopyTime>
<CurrentServer>0</CurrentServer>
<TotalServer>0</TotalServer>
<IndexBuildTime>46158</IndexBuildTime>
<Status>0</Status>
<JobID>1131987633899</JobID>
<MessagesAvailable>false</MessagesAvailable>
<StopTime>1131987734199</StopTime>
<TotalTime>100298</TotalTime>
<NumberOfDocuments>43</NumberOfDocuments>
</IndexStatus>
<CurrentIndexWildcardSupport/>
<NextIndexWildcardSupport Type="None" Limit="0"/>
<ScheduleStatus>
 <Status>1</Status>
 <ScheduledTime Enabled="false"></ScheduledTime>
</ScheduleStatus>
</Monitor>

```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния построения индекса:

Таблица 28. Элементы для команды состояния построения индекса

Элемент	Описание
IndexStatusId	ID состояния индекса.
StartTime	Момент запуска построения индекса в секундах, истекших с 01.01.1970. Момент настоящего времени из этого значения можно получить по формуле January 1, 1970 %B StartTime. Дополнительную информацию об исчислении времени смотрите на сайте <a href="http://en.wikipedia.org/wiki/Unix_epoch">http://en.wikipedia.org/wiki/Unix_epoch</a> .
Progress	Процент выполнения для данного построения индекса.
CurrentPhase	<ul style="list-style-type: none"> <li>• 1: стадия перезаписи склада</li> <li>• 2: стадия глобального анализа</li> <li>• 3: стадия построения индекса</li> </ul>
TotalPhase	Общее число фаз построения индекса. В настоящее время это значение - всегда 3.
IndexCopyProgress	Процент выполнения для копирования индекса. Процесс копирования индекса копирует построенный индекс с сервера построения индексов на серверы поиска.
CurrentServer	Сервер поиска, на который в данный момент копируется индекс.
TotalServer	Число серверов поиска, на которые копируется индекс.
IndexCopyTime	Общее время копирования индекса на все серверы поиска.
IndexBuildTime	Общее время для всех фаз построения индекса
Состояние	<ul style="list-style-type: none"> <li>• 0: построение и копирование индекса</li> <li>• -1: ошибка требования построения индекса</li> <li>• 1: идет построение индекса, копирование или оба процесса</li> </ul>
JobID	Уникальный ID, связанный с каждым требованием построения индекса.
MessagesAvailable	Логическое значение, указывающее на доступность сообщений об ошибках (если возникли ошибки).
StopTime	Время завершения всех фаз построения и копирования индекса.
TotalTime	Интервал между временем запуска и временем завершения.
NumberOfDocuments	Число документов в индексе.

Таблица 28. Элементы для команды состояния построения индекса (продолжение)

Элемент	Описание
CurrentIndexWildcardSupport	Значение параметра символов подстановки для следующего построения индекса. Возможные значения - None (не используются), QueryExpansion (расширение в запросе) и IndexExpansion (расширение в индексе).
ScheduleStatus	<ul style="list-style-type: none"> <li>0, если расписание для данного собрания и типа индекса не включено.</li> <li>1, если расписание для данного собрания и типа индекса включено.</li> </ul>
ScheduledTimeEnabled	Момент следующего построения индекса для данного собрания и типа индекса в секундах, прошедших с 01.01.1970. Момент настоящего времени из этого значения можно получить по формуле January 1, 1970 %2B ScheduledTimeEnabled. Дополнительную информацию об исчислении времени смотрите на сайте <a href="http://en.wikipedia.org/wiki/Unix_epoch">http://en.wikipedia.org/wiki/Unix_epoch</a> .

**Состояние сервера поиска:** Когда вы вводите команду для получения состояния сервера поиска, эта команда возвращает информацию в формате документа XML. Команда состояния сервера поиска возвращает следующую информацию:

```
FFQC5314I Result: <?xml version="1.0"?>
<Monitor Type="Search" Count="1">
<SearchStatus Name="Search Manager (node1)" SearchID=
"searchmanager.node1" HostName="myComputer.svl.ibm.com">
<Status>1</Status>
</SearchStatus>
</Monitor>
```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния сервера поиска:

Таблица 29. Элементы для команды состояния сервера поиска

Элемент	Описание
SearchStatusName	Имя и ID сеанса менеджера поиска, отслеживающего индексы поиска для данного собрания и управляющего ими.
HostName	Имя хоста для сервера, на котором расположен индекс поиска.
Состояние	<ul style="list-style-type: none"> <li>0, если индекс поиска для данного собрания не работает.</li> <li>1, если индекс поиска для данного собрания работает.</li> </ul>

**Подробное состояние сервера поиска:** Команда состояния сервера поиска может возвращать следующую информацию:

```
FFQC5303I Менеджер поиска (node1) (sid: searchmanager.node1)
уже работает. PID: 15711
FFQC5314I Результат: PID
CacheHits=3
QueryRate=1
Port=44008
SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1
```

Ниже в таблице описаны элементы информации, возвращаемой командой подробного состояния сервера поиска:

Таблица 30. Элементы для команды подробного состояния сервера поиска

Элемент	Описание
CacheHits	Число результатов, полученных из кэша поиска.
QueryRate	Число запросов, полученных за последний интервал времени. По умолчанию этот интервал равен 5 минутам.
Port	Номер порта, используемого индексом поиска для ожидания или получения запросов.
SessionId	ID сеанса для индекса поиска этого собрания.
CacheHitRate	Число результатов, полученных из кэша поиска, в виде процентной доли от общего числа результатов.
ResponseTime	Среднее время ответа в миллисекундах за заданный интервал времени. (Интервал по умолчанию - 5 минут.)
Состояние	<ul style="list-style-type: none"> <li>• 0, если индекс поиска для данного собрания не работает.</li> <li>• 1, если индекс поиска для данного собрания работает.</li> </ul>
SessionName	Имя сеанса для индекса поиска этого собрания.

## Коды возврата для команд esadmin

Команды **esadmin** могут возвращать следующие коды:

Таблица 31. Коды возврата для команд esadmin

Код	Название	Описание
0	CODE_ERROR_NONE	Команда выполнена успешно.
102	CODE_ERROR_INSTANTIATION_EXCEPTION	Ошибка при создании экземпляра обработчика команды.
103	CODE_ERROR_ACCESS_EXCEPTION	Ошибка недопустимого доступа при создании экземпляра обработчика команды.
104	CODE_ERROR_EXECUTE_EXCEPTION	
105	CODE_ERROR_THROWABLE	
106	CODE_ERROR_NO_SUCH_METHOD	
107	CODE_ERROR_INVALID_SESSION	
108	CODE_ERROR_INVALID_PARAMETER	
109	CODE_ERROR_SESSION_NOT_RUNNING	

## Получение ID сеансов

Команда **esadmin check** выводит список компонентов поисковой системы предприятия и соответствующих ID сеансов. В следующей таблице показан список общих сеансов, их ID, серверы, на которых они выполняются, и состояния сеансов.

Таблица 32. Примеры имен сеансов, исходных серверов, ID сеансов и состояний сеансов

Сеанс	Сервер, на котором он выполняется	ID сеанса	Состояние сеанса
configmanager	сервер индексов	10433	Started
controller	сервер индексов	10464	Started
customcommunication	сервер индексов	Неприменимо	Неприменимо
обнаружение	сервер индексов	10649	Started

Таблица 32. Примеры имен сеансов, исходных серверов, ID сеансов и состояний сеансов (продолжение)

Сеанс	Сервер, на котором он выполняется	ID сеанса	Состояние сеанса
наблюдатель	сервер индексов	10682	Started
parserservice	сервер индексов	10718	Started
resource.node1	сервер индексов	10759	Started
samplecpp	сервер индексов	10827	Started
sampletest	сервер индексов	10857	Started
scheduler	сервер индексов	10889	Started
searchmanager.node1	сервер индексов	10927	Started
utilities.node1	сервер индексов	10384	Started

#### Понятия, связанные с данным

“Слежение за активностью поисковой системы предприятия” на стр. 305

“Резервное копирование и восстановление поисковой системы предприятия” на стр. 337



Сообщения для поисковой системы предприятия



Сообщения для поисковой системы предприятия

#### Задачи, связанные с данной

“Отслеживание работы искателей” на стр. 308

“Запуск поисковой системы предприятия” на стр. 297

“Остановка поисковой системы предприятия” на стр. 299

“Управление серверами поиска в автономном режиме” на стр. 302



---

## Регистрозависимость в поисковой системе предприятия

В различных компонентах поисковой системы предприятия - синтаксисе запросов, прямых ссылках, именах полей и так далее - регистрозависимость понимается по-разному.

### Синтаксис запросов

При поиске регистр игнорируется во всех случаях, кроме следующих:

#### Имена элементов и атрибутов XML

Регистрозависимы. Условия и значения атрибутов регистронезависимы даже в запросах XML. Например, в этом документе:

```
<book>
 <Author>
 <Name>Ferdinand</Name>
 <Contact Type="eMail">ferdi@nand.org</Contact Type>
 <Contact Type="Phone">+1 408 876 4242</Contact Type>
 </Author>
</book>
```

Следующие запросы не вернут этот документ:

- @xmlxp::'author[Name ftcontains ("Ferdinand")]'
- @xmlxp::'//contact[@type="eMail"]'
- @xmlf2::'<author><name>Ferdinand</name></author>'
- @xmlf2::'<CONTACT TYPE="email">ferdi</contact>'

А приведенные ниже запросы этот документ вернут:

- @xmlxp::'Author[Name ftcontains ("ferdinand")]'
- @xmlxp::'//Contact[@Type="email"]'
- @xmlf2::'<Author><Name>ferdinand</Author><Name>'
- @xmlf2::'<Contact Type="email">ferdi</Contact>'

#### Списки управления доступом (ACL)

Регистрозависимы.

#### URL в условиях docid: и samegroupas:

Регистрозависимы. Однако части URL в условиях site: и url: регистронезависимы. Пусть, например, URL документа - <http://www.here.com/HR/>:

- Запрос docid:<http://www.here.com/hr> не вернет этот документ.
- Запросы url:hr и url:HERE вернут этот документ.
- Запрос site:HERE.com вернет этот документ.

#### ID категорий и таксономии

Регистронезависимы. Например, в запросе taxonomy\_id::category\_id ни регистр taxonomy\_id, ни регистр category\_id не важны. Запрос RuleBased::c42 соответствует категории c42 в таксономии на основе правил, а также соответствует rulebased::C42. Поиск по имени категории недоступен, но доступен поиск по ID категории.

#### Области действия

Регистронезависимы. Например, и Scope:RESEARCH, и scope::research вернут документы из области с именем Research.

**Условия с символами подстановки**

Регистронезависимы. Например, условие `Feг*n*d` эквивалентно `feг*n*d`.

**Имена полей**

Регистронезависимы. Например, все запросы `Title:Expenses`, `TITLE:expenses` и `title:expenses` эквивалентны. Все имена полей регистронезависимы, даже если они взяты из файла отображений XML. Однако внешние источники со своей семантикой могут обрабатывать имена полей как регистрозависимые.

**Прямые ссылки**

Регистронезависимы.

**Категоризатор на основе правил**

Правила URL регистрозависимы, а правила содержимого документов регистронезависимы.

**Сворачивание URI и определения значимости на основе шаблонов URI**

URI регистрозависимы, но свернутые имена групп URI регистронезависимы. Нельзя задать два свернутых имени групп URI, различающихся только регистром.

**Имена полей**

Регистронезависимы. Это правило применимо к именам полей в запросах, отображениям XML, определениям классов значимости и остальным интерфейсам администратора, в которых задаются имена полей. Если в интерфейсе вы сконфигурируете имя поля в верхнем или смешанном регистре, система может перевести его в нижний регистр, и при очередном просмотре конфигурации вы, возможно, увидите его в нижнем регистре. Кроме того, внешний источник может воспринимать имя поля как регистрозависимое.

**Словари**

Регистронезависимы. Это правило применимо к словарям синонимов, словарям стоп-слов, словарям предложений по правописанию и словарям значащих слов.

**Имена и ID собраний**

Регистрозависимы. Если вы задаете имя собрания или ID собрания в API администратора или поиска, они должны в точности соответствовать регистру собрания с этим именем. Но хотя ID собраний регистрозависимы, все равно запрещено задавать два ID собраний, различающиеся только регистром. Это же ограничение применяется к именам и ID искателей и источников данных.

**Прикладные программы поиска**

Регистрозависимы. Нельзя задать два имени или ID программ поиска, которые различаются только регистром.

---

## Документация поисковой системы предприятия

Документацию по OmniFind Enterprise Edition можно читать в форматах PDF и HTML.

Программа установки OmniFind Enterprise Edition автоматически устанавливает информационный центр с версиями HTML документации поисковой системы предприятия. При установке на несколько серверов информационный центр устанавливается на оба сервера поиска. Если информационный центр не установлен, при нажатии кнопки Справка открывается информационный центр на сайте IBM.

Для просмотра установленных версий документов PDF перейдите в каталог КОРНЕВОЙ\_КАТАЛОГ\_УСТАНОВКИ\_ES/docs/локаль/pdf. Например, документы на английском языке находятся в каталоге КОРНЕВОЙ\_КАТАЛОГ\_УСТАНОВКИ\_ES/docs/en\_US/pdf.

Получить PDF-версии документации на всех доступных языках можно на сайте документации OmniFind Enterprise Edition Версии 8.5.

Кроме того, на сайте поддержки OmniFind Enterprise Edition доступны загружаемые материалы, исправления, технические замечания и информационный центр.

В следующей таблице перечислена доступная документация с указанием имен файлов и их местоположения.

Таблица 33. Документация для поисковой системы предприятия

Название	Имя файла	Положение
Информационный центр		<a href="http://publib.boulder.ibm.com/infocenter/discover/v8r5/">http://publib.boulder.ibm.com/infocenter/discover/v8r5/</a>
<i>Руководство по установке поисковой системы предприятия</i>	iiysi.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/локаль/pdf/
<i>Руководство Начинаем работу</i> (Этот документ доступен также в печатном виде на английском, французском и японском языках.)	OmniFindEE850_qsg_двухбукв_код_локали.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/локаль/pdf/
<i>Управление поисковой системой предприятия</i>	iiysa.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/локаль/pdf/
<i>Programming Guide and API Reference for Enterprise Search</i> (Руководство по программированию и справочник по API поисковой системы предприятия)	iiysp.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/en_US/pdf/
<i>Troubleshooting Guide and Messages Reference</i> (Руководство по устранению неисправностей и Справочник по сообщениям)	iiysm.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/локаль/pdf/
<i>Интеграция анализа текста</i>	iiyst.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/docs/локаль/pdf/

Таблица 33. Документация для поисковой системы предприятия (продолжение)

Название	Имя файла	Положение
Подключаемый модуль для Google Desktop Search	iiysg.pdf	КОРНЕВОЙ_КАТАЛОГ_УСТАНОВКИ_ES/ docs/локаль/pdf/

---

## Возможности доступности

Возможности доступности помогают пользователям с физическими недостатками, такими как ограниченная подвижность или слабое зрение, с успехом пользоваться программными продуктами.

IBM стремится создавать продукты, которые были бы доступны для каждого, независимо от возраста или физических возможностей.

### Возможности доступности

Ниже в списке перечислены основные возможности доступности в OmniFind Enterprise Edition:

- Работа с использованием только клавиатуры
- Интерфейсы с возможностью применения программ чтения с экрана

Информационный центр OmniFind Enterprise Edition и связанные с ним публикации реализуют возможности доступности. Эти возможности описаны в [http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/topic/com.ibm.classify.nav.doc/dochome/accessibility\\_info.htm](http://publib.boulder.ibm.com/infocenter/discover/v8r5m0/topic/com.ibm.classify.nav.doc/dochome/accessibility_info.htm).

### Перемещения по интерфейсу с помощью клавиатуры

Этот продукт использует стандартные клавиши навигации Microsoft Windows.

В программе установки OmniFind Enterprise Edition можно использовать следующие быстрые клавиши.

*Таблица 34. Быстрые клавиши для программы установки*

Действие	Клавиши
Выделить радиокнопку	Клавиша со стрелкой
Выбрать радиокнопку	Клавиша Tab
Выделить кнопку	Клавиша Tab
Выбрать кнопку	Клавиша Enter
Перейти в следующее или предыдущее окно или отменить операцию	Выделите кнопку при помощи клавиши Tab и нажмите Enter
Сделать активное окно неактивным	Ctrl + Alt + Esc

### Информация об интерфейсе

Пользовательские интерфейсы для консоли администратора, примера программы поиска и настройщика программ поиска построены на основе браузера; вы можете использовать Microsoft Internet Explorer или Mozilla FireFox. Список быстрых клавиш и других функций доступности для вашего браузера смотрите в электронной справке по Internet Explorer или FireFox.

### Информация о доступности

Публикации по OmniFind Enterprise Edition в формате Adobe Portable Document Format (PDF) можно просматривать при помощи Adobe Acrobat Reader. Эти файлы PDF

содержатся на компакт-диске, прилагаемом к продукту; они доступны также по адресу <http://www.ibm.com/support/docview.wss?rs=63&uid=swg27010938>.

## **IBM и доступность**

Дополнительную информацию о том, как IBM работает над доступностью своих продуктов, смотрите на сайте IBM Human Ability and Accessibility Center.

---

## Глоссарий терминов поисковой системы предприятия

В этом глоссарии определяются термины, используемые в интерфейсах и документации поисковой системы предприятия.

### **IP-адрес (IP address)**

Уникальный адрес устройства или логической единицы в сети, соответствующий стандарту IP.

### **Java Database Connectivity (JDBC)**

Промышленный стандарт независимой от баз данных связи между платформой Java и различными базами данных. Интерфейс JDBC содержит API уровня вызовов для доступа к базам данных на основе SQL.

### **JavaScript**

Язык сценариев Web, используемый в браузерах и Web-серверах.

### **JavaServer Pages (JSP)**

Технология создания сценариев на сервере, позволяющая динамически встраивать код Java в страницы Web (файлы HTML) и выполнять его при загрузке страницы для возвращения клиенту динамического содержимого.

### **Portal Document Manager (PDM)**

Предоставляет пользователям единый центральный репозиторий документов для обеспечения взаимодействия между рабочими группами.

Администраторы могут эффективно управлять своими документами и могут контролировать способы доступа пользователей к информации.

### **Secure Sockets Layer (SSL)**

Протокол защиты, обеспечивающий конфиденциальность связи. При помощи SSL программы структуры "клиент-сервер" могут связываться друг с другом с исключением возможностей подслушивания, подделки и искажения сообщений.

### **Unstructured Information Management Architecture (UIMA)**

Архитектура IBM, определяющая структуру для реализации систем анализа неструктурированных данных.

### **административная роль (administrative role)**

Классификация пользователя, определяющая его права доступа.

### **администратор поисковой системы предприятия (enterprise search administrator)**

Административная роль, позволяющая пользователю управлять всей поисковой системой предприятия.

### **анализатор (parser)**

Программа, интерпретирующая документы, которые добавлены на склад данных поисковой системы предприятия. Анализатор извлекает из документов информацию и подготавливает эти документы для индексации, поиска и получения.

### **анализ ссылок (link analysis)**

Метод, основанный на анализе гиперссылок между документами и помогающий определить, какие страницы в собрании важны для пользователей.

**анализ текста (text analysis)**

Процесс извлечения из текста семантических признаков и другой информации для повышения возможностей получения данных в сборании. Смотрите также семантический поиск.

**аннотатор (annotator)**

Компонент программного обеспечения, выполняющий конкретные задачи лингвистического анализа, генерирующий и записывающий аннотации. Аннотатор - это компонент логики анализа в механизме анализа.

**аннотатор регулярных выражений (regular expression annotator)**

Программный компонент, который выделяет в текстовом документе блоки или единицы информации, такие как номера продуктов, на основе регулярных выражений, описывающих шаблоны для поиска в тексте документа. Если одно из регулярных выражений соответствует части текста документа, аннотатор регулярных выражений создает соответствующие аннотации для этого соответствия или его части. Затем эти аннотированные выражения сохраняются либо в индексе поисковой системы предприятия с помощью файла отображения индекса, либо в базе данных с поддержкой JDBC с помощью файла отображения базы данных.

**аннотация (annotation)**

Информация о характере текста. Например, в аннотации может быть указано, что текст представляет собой название компании. В архитектуре UIMA (Unstructured Information Management Architecture - архитектура управления неструктурированной информацией) аннотация - это особая структура признаков.

**аннотирование (summarization)**

Процесс включения в результаты поиска предложений кратко без повторов описывающих содержимое документа. Смотрите также динамическое аннотирование и статическое аннотирование.

**базовые аннотаторы поисковой системы предприятия (enterprise search base annotators)**

Набор стандартных механизмов анализа текста, по умолчанию используемых поисковой системой предприятия для анализа документов.

**библиотека Domino Document Manager (Domino Document Manager library)**

База данных Domino Document Manager, служащая входной точкой для Domino Document Manager.

**библиотека (library)**

Объект системы, служащий каталогом для других объектов. Смотрите также библиотека Domino Document Manager.

**Виртуальная Java-машина (Java virtual machine, JVM)**

Программная реализация процессора, на котором выполняется скомпилированный код Java (апплеты и прикладные программы).

**внешний источник данных (external data source)**

Источник данных для объединения, который не просматривается искателем, не анализируется и не индексируется OmniFind Enterprise Edition. Поиск во внешних источниках данных делегируется API запросов для этих источников данных.

**выделение основ (stemming)**

Смотрите выделение основ слов.

**выделение основ слов (word stemming)**

Процесс лингвистического нормирования, при котором различные формы слова приводятся к общей форме. Например, слова *connections*, *connective*, и *connected* приводятся к форме *connect*.

**гибридный поиск (hybrid search)**

Сочетание логического поиска и свободного текстового поиска.

**дельта-построение индекса (delta index build)**

В поисковой системе предприятия - процесс добавления новой информации в существующий индекс. Сравните с полным построением индекса.

**дерево категорий (category tree)**

Иерархия категорий.

**диакритический знак (diacritic)**

Пометка, меняющая фонетическое значение символа или сочетания символов.

**динамическое ранжирование (dynamic ranking)**

Тип ранжирования, при котором ранг результата определяется в соответствии с вхождением терминов запроса в анализируемые документы. Смотрите также оценка на основе текста. Сравните со статическим ранжированием.

**динамическое составление сводок (dynamic summarization)**

Тип создания краткого описания документа, при котором выделяются термины поиска и в результаты поиска включаются фразы, в которых наиболее полно представлены искомые понятия документа. Противоположное понятие - статическое аннотирование.

**директива no-follow (no-follow directive)**

Директива на странице Web, указывающая роботам (например, искателю Web) не переходить по ссылкам с этой страницы.

**директива no-index (no-index directive)**

Директива на странице Web, указывающая роботам (например, искателю Web) не включать содержимое этой страницы в индекс.

**драйвер анализатора (parser driver)**

В поисковой системе предприятия - служба, передающая документы службе анализатора. Для каждого собрания есть один драйвер анализатора. Служба драйвера анализатора собрания соответствует анализатору этого собрания на консоли управления поисковой системой предприятия.

**значимое слово (boost word)**

Слово, влияющее на относительный ранг документа в результатах поиска. При обработке запроса важность документа, содержащего значимые слова, может быть повышена или снижена в зависимости от коэффициента, назначенного этому слову.

**извлечение информации (information extraction)**

Тип извлечения понятий, при котором в тексте документа автоматически распознаются важные словарные элементы, такие как имена, термины и выражения.

**извлечение понятий (concept extraction)**

Функция анализа текста, которая находит в текстовых документах важные словарные элементы (например, имена людей, названия мест или продуктов) и генерирует список таких элементов. Смотрите также извлечение темы.

**извлечение темы (theme extraction)**

Тип извлечения понятий, при котором в тексте документа автоматически распознаются важные словарные элементы для определения темы документа. Смотрите также извлечение понятий.

**индекс (index)**

Смотрите полнотекстовый индекс.

**искатель (crawler)**

Программа, получающая документы из источника данных и собирающая информацию, которая может использоваться для создания индексов поиска.

**искатель Web (Web crawler)**

Тип искателя, исследующего Web путем получения документов Web и следования по содержащимся в них ссылкам.

**источник данных (data source)**

Любой репозиторий данных, из которого можно получать документы, например, Web, реляционные и нереляционные базы данных и системы управления содержимым.

**катакана (Katakana)**

Набор символов в одном из двух японских фонетических алфавитов, применяемом главным образом для фонетической записи иностранных слов.

**категория на основе правил (rule-based category)**

Категория, созданная по правилам, которые задают, как документы связываются с определенными категориями. Например, можно определить правила, связывающие с конкретными категориями документы, содержащие определенные слова или не содержащие такие слова, или документы, соответствующие шаблону универсального идентификатора ресурса (URI).

**класс значимости (boost class)**

Объект, который содержит спецификации, влияющие на относительный ранг документа в результатах поиска.

**клитика (clitic)**

Синтаксически самостоятельное слово, которое произносится слитно с другим словом. Клитика может писаться как слитно, так и раздельно со словом, с которым она связана фонетически. Типичные примеры клитик в английском языке дает стяжение (*wouldn't* или *you're*).

**кольцо ключей (key ring)**

В компьютерной безопасности - файл, содержащий общие ключи, секретные ключи, корневые доверенные сертификаты и сертификаты серверов. Смотрите также файл склада ключей.

**комната Lotus QuickPlace (Lotus QuickPlace room)**

Часть площадки Lotus QuickPlace с разделами, предназначенная только для авторизованных пользователей с общими интересами и потребностью в коллективной работе.

**комната (room)**

Программа, с помощью которой пользователи могут создавать документы для чтения другими пользователями, отвечать на комментарии других пользователей и просматривать состояние проектов и контрольные сроки их выполнения. Пользователи могут также обмениваться сообщениями в реальном времени с другими посетителями комнаты. Смотрите также комната Lotus QuickPlace.

**конечный символ (trailing character)**

Символ, занимающий последнюю позицию в слове.

**контекстный поиск (proximity search)**

Текстовый поиск, возвращающий результаты, в которых заданные термины находятся на определенном расстоянии друг от друга, например, в одном предложении или абзаце.

**кэш поиска (search cache)**

Буфер, в котором хранятся данные и результаты предыдущих требований поиска.

**лексическая близость (lexical affinity)**

Отношение взаимосвязи в документе слов поиска с близкими значениями. Лексическая близость используется для вычисления релевантности результата.

**лексический анализ (tokenization)**

Процесс разбиения входной информации на элементы.

**лексический анализатор (tokenizer)**

Программа сегментации текста, просматривающая текст и определяющая, какие последовательности символов можно выделить в качестве элементов текста.

**лемма (lemma)**

Базовая форма слова. Леммы важны в языках с развитым словоизменением, например, в русском.

**лемматизация (lemmatization)**

Процесс определения базовой формы и различных грамматических форм слова. Например, при поиске слова mouse будут найдены также документы, содержащие слово mice, а поиск слова go должен вернуть документы, содержащие слова going, gone или went.

**лигатура (ligature)**

Два или несколько символов, объединяемых при печати в один. Например, сочетания символов ff и ffi при печати могут быть изображены как лигатуры.

**лингвистический поиск (linguistic search)**

Тип поиска, при котором просматриваются, получают и индексируются документы, содержащие термины поиска, которые преобразованы в их базовые лексические формы (например, слово дни индексируется как день) или дополнены до базовых форм (как в случае сложных слов).

**локальный федератор (local federator)**

В поисковой системе предприятия - клиентский объект, создаваемый API поиска и индексации и позволяющий пользователям выполнять поиск в наборе разнородных собраний и получать единообразные результаты.

**маркер защиты (security token)**

Информация идентификации и защиты, используемая для авторизации доступа к документам в собрании. Различные типы источников данных поддерживают разные типы маркеров защиты. Примеры маркеров защиты: роли, ID пользователей, ID групп и другая информация, которая может использоваться для управления доступом к содержимому.

**механизм анализа (analysis engine)**

Смотрите механизм анализа текста.

**механизм анализа текста (text analysis engine)**

Компонент программного обеспечения, ответственный за поиск и представление контекстного и смыслового содержимого текста.

**механизм поиска (search engine)**

Программа, принимающая требование поиска и возвращающая пользователю список документов.

**наблюдатель (monitor)**

Пользователь поисковой системы предприятия с полномочиями наблюдать за процессами на уровне собрания.

**начальный URL (start Uniform Resource Locator (URL))**

Начальная точка для работы искателя.

**нормализация символов (character normalization)**

Процесс, при котором различные формы символа, например, символы в верхнем регистре или с диакритическими знаками, приводятся к общей форме.

**обработка архива механизма (processing engine archive)**

Файл zip-архива .pear, включающий механизм анализа UIMA (Unstructured Information Management Architecture) и все необходимые ресурсы для пользовательского анализа в поисковой системе предприятия.

**общая структура анализа (common analysis structure, CAS)**

Структура, в которой хранятся содержимое и метаданные документа и все результаты анализа, созданные механизмом анализа текста. Все операции обмена данными в процессе анализа документа выполняются с помощью общей структуры анализа.

**общий уровень связи (common communication layer, CCL)**

Инфраструктура связи, объединяющая различные компоненты (контроллер, анализатор, искатель, сервер индексов) OmniFind Enterprise Edition.

**объединенный поиск (federated search)**

Возможность поиска, позволяющая выполнять его в нескольких поисковых службах, возвращая объединенный список результатов поиска.

**объектная модель документа (Document Object Model)**

Система, в которой структурированный документ, например, файл XML, представлен в виде дерева объектов. Программы могут обращаться к этому дереву и изменять его.

**оператор (operator)**

Пользователь поисковой системы предприятия с полномочиями наблюдать за процессами на уровне собрания, а также запускать и останавливать эти процессы.

**определение языка (language identification)**

В поисковой системе предприятия функция поиска, определяющая язык документа.

**оценка на основе текста (text-based scoring)**

Процесс назначения документу целочисленного значения, указывающего степень соответствия документа терминам запроса (релевантность). Большее значение указывает на большую степень соответствия запросу. Смотрите также динамическое ранжирование.

**очередь индексации (index queue)**

Список ожидающих обработки требований построения индекса (полного или дельта).

**параметрический поиск (parametric search)**

Тип поиска, при котором ищутся объекты, содержащие числовые значения

или атрибуты, например, даты, целочисленные значения или числовых данные других типов в заданном диапазоне.

**параметры аутентификации (credential)**

Запрашиваемая при аутентификации детальная информация, описывающая пользователя, его связи с группами и другие атрибуты идентичности, важные для защиты. Параметры аутентификации позволяют выполнять разнообразные операции, например, аутентификацию, аудит и делегирование. Например, регистрационные данные (ID пользователя и пароль) для пользователя - это параметры аутентификации, разрешающие пользователю обращаться с своему счету.

**площадка Lotus QuickPlace (Lotus QuickPlace place)**

Узел Web, предоставляемый Lotus QuickPlace и позволяющий географически разобщенным участникам совместно работать над проектами и оперативно общаться в структурированном и защищенном рабочем пространстве.

**площадка (place)**

Виртуальная область, которая видна с портала, служащего для совместной работы индивидуальных пользователей и групп. У каждого пользователя в портале есть персональная площадка для индивидуальной работы, кроме того, у индивидуальных пользователей и групп есть доступ к различным совместно используемым площадкам, которые могут быть как общедоступными, так и защищенными. Смотрите также площадка Lotus QuickPlace.

**поиск неполных соответствий (fuzzy search)**

Поиск, возвращающий слова, написание которых сходно с термином поиска.

**поиск по полю (fielded search)**

Запрос, применяемый только к конкретному полю.

**поиск с логическими операциями (Boolean search)**

Поиск, в котором один или несколько терминов поиска объединяются при помощи логических операций AND, NOT или OR.

**поиск со взвешенными терминами (weighted term search)**

Запрос, в котором некоторые термины заданы как более важные.

**поле (field)**

Область, в которой вводятся данные определенной категории или управляющая информация.

**полное построение индекса (main index build)**

В поисковой системе предприятия - процесс построения индекса целиком. Противоположное понятие - дельта-построение индекса.

**полнотекстовый индекс (full-text index)**

Структура данных, содержащая ссылки на элементы данных и позволяющая при поиске быстро находить документы, содержащие термы запроса.

**пользовательский агент (user agent)**

Прикладная программа, просматривающая данные в Web и оставляющая информацию о себе на сайтах, которые она посетила. В поисковой системе предприятия пользовательским агентом является искатель Web.

**пользовательский механизм анализа текста (custom text analysis engine)**

Механизм анализа текста, созданный с помощью комплекта разработки программ (SDK) UIMA (Unstructured Information Management Architecture); этот механизм анализа текста можно добавить в набор стандартных механизмов анализа текста поисковой системы предприятия (их также

называют базовыми аннотаторами поисковой системы предприятия).  
Смотрите также механизм анализа текста.

**поместить в очередь (enqueue)**

Поставить сообщение или элемент в очередь.

**потребитель общей структуры анализа, потребитель CAS (common analysis structure consumer, CAS consumer)**

Потребитель, который выполняет итоговую обработку результатов анализа, сохраненных в общей структуре анализа. Например, этот потребитель индексирует содержимое общей структуры анализа в механизме поиска или вносит конкретные результаты анализа в реляционную базу данных.

**прикладная программа поиска (search application)**

В поисковой систем предприятия - программа, обрабатывающая запросы, выполняющая поиск в индексе, возвращающая результаты поиска и получающая исходные документы.

**произвольная страница ошибок (soft error page)**

Тип страницы Web с информацией о том, почему нельзя передать запрошенную страницу. Например, вместо простого возврата кода сервер HTTP может вернуть страницу с подробными объяснениями этого кода.

**прокси-сервер (proxy server)**

Сервер, играющий роль посредника для требований HTTP Web, сгенерированных прикладной программой или сервером Web. Прокси-сервер подменяет внешние контент-серверы внутри предприятия.

**пространство для искателя (crawl space)**

Набор источников, соответствующих заданным шаблонам (таким как имена баз данных, пути файловой системы, имена доменов и IP-адреса), который искатель читает для получения элементов для индексации.

**протокол DIIOP (Domino Internet Inter-ORB Protocol)**

Задача сервера, запущенная на сервере и взаимодействующая с Domino Object Request Broker для обеспечения связи между апплетами Java, созданными классами Notes Java и сервером Domino. Пользователи браузеров и серверы Domino используют DIIOP для связи и обмена данными объектов.

**протокол NRPC (Notes Remote Procedure Call)**

Механизм связи Lotus Notes, служащий для связи всех компонентов Notes друг с другом.

**протокол Robots Exclusion (Robots Exclusion Protocol)**

Протокол, позволяющий администраторам сайтов указывать посещающим эти сайты роботам, какие участки сайтов роботы не должны посещать.

**протокол упрощенного доступа к каталогам (Lightweight Directory Access Protocol, LDAP)**

Открытый протокол, который использует TCP/IP для доступа к каталогам, поддерживающим модель X.500. Этот протокол не предъявляет таких требований к ресурсам, как более сложный протокол доступа к каталогам X.500 Directory Access Protocol (DAP). LDAP можно использовать, например, для поиска людей, организаций и прочих ресурсов в Интернете или в каталоге интранета.

**прямая ссылка (quick link)**

Связь между универсальным идентификатором ресурса (URI) и ключевыми словами или словосочетаниями.

**путь признака (feature path)**

Путь для доступа к значению признака в структуре признаков UIMA (Unstructured Information Management Architecture).

**ранжирование (ranking)**

Назначение каждому документу в результатах запроса целочисленного значения. Порядок документов в результатах поиска определяется степенью их соответствия запросу. Более высокий ранг означает более точное соответствие. Смотрите также динамическое ранжирование и статическое ранжирование.

**ранжирование по популярности (popular ranking)**

Тип ранжирования, повышающая ранг документа на основе степени его популярности.

**результаты анализа (analysis results)**

Информация, генерируемая аннотаторами. Результаты анализа записываются в структуру данных под названием общая структура анализа. Результаты анализа, созданные пользовательскими механизмами анализа текста (аннотаторами), можно сделать доступными для поиска, включив их в индекс поисковой системы предприятия.

**результаты поиска (search results)**

Список документов, удовлетворяющих требованию поиска.

**свободный текст (free-form text)**

Неструктурированный текст, состоящий из слов или предложений.

**свободный текстовый поиск (free text search)**

Поиск, при котором термины поиска выражены в виде текста в свободном формате.

**сегментация (segmentation)**

Деление текста на отдельные лексические единицы. При обработке без использования словаря производится сегментация с помощью пробелов и сегментация энграмм, а при обработке на основе словаря может выполняться сегментация по словам, предложениям и абзацам, а также лемматизация.

**сегментация с помощью пробелов на основе Unicode (Unicode-based white space segmentation)**

Метод лексического анализа, использующий для различения символов элементов и разделителей свойства символов Unicode.

**сегментация текста (text segmentation)**

Смотрите сегментация.

**сегментация энграмм (n-gram segmentation)**

Метод анализа, при котором перекрывающиеся последовательности из определенного числа символов рассматриваются как одно слово, а не разбиваются на отдельные слова с помощью пробелов, как при сегментации на основе Unicode.

**семантический поиск (semantic search)**

Тип поиска по ключевым словам с использованием лингвистического и контекстного анализа. Смотрите также анализ текста.

**сервлет (servlet)**

Программа Java, выполняемая на Web-сервере и расширяющая его функциональные возможности посредством генерирования динамического содержимого в ответ на требования клиента Web. Сервлеты широко используются для подключения баз данных к Web.

**сертификат (certificate)**

В компьютерной безопасности - цифровой документ, связывающий общий ключ с личностью владельца сертификата, обеспечивая тем самым его аутентификацию. Сертификат выдается сертифицированным и подписан его цифровой подписью.

**сертификатор (certificate authority)**

Доверенная независимая организация или компания, выдающая цифровые сертификаты для создания цифровых подписей и пар общих ключ-секретный ключ. Сертификатор гарантирует идентификацию тех, кто получает от него уникальные сертификаты.

**символ маски (masking character)**

Символ, представляющий необязательные символы в начале, середине или конце аргумента поиска. Обычно символы маски применяются при поиске вариантов написания аргумента поиска в точном индексе. Смотрите также символ подстановки.

**символ новой строки (newline character)**

Управляющий символ, вызывающий перемещение позиции печати или вывода на одну строку вниз.

**символ подстановки (wildcard character)**

Символ, представляющий необязательные символы в начале, середине или конце аргумента поиска.

**система объединения (federation)**

Процесс объединения систем имен, в результате которого программы могут обрабатывать составные имена, расширяющие эти системы имен.

**система типов (type system)**

Система типов определяет типы объектов (структур признаков), которые механизм анализа текста может обнаруживать в документе. Система типов определяет все возможные структуры признаков в терминах типов и признаков. В системе типов можно определить любое число различных типов. Система типов зависит от домена и прикладной программы.

**склад данных (data store)**

Структура данных, в которой документы хранятся в проанализированной форме.

**склад необработанных данных (raw data store)**

Структура данных, в которой хранятся обработанные поиском документы перед передачей их анализатору. Искатели записывают документы на склад необработанных данных, а анализатор читает их с этого склада. После анализа документов они удаляются со склада необработанных данных. Не смешивайте со складом данных.

**словарь синонимов (synonym dictionary)**

Словарь, дающий возможность при поиске в собрании искать также синонимы терминов запроса.

**служба анализатора (parser service)**

Служба поисковой системой предприятия, выполняющая все операции анализа текста для собраний документов. В каждый момент времени работает как минимум одна служба анализатора.

**собрание (collection)**

Набор источников данных и опций для просмотра, анализа, индексации и поиска для этих источников данных.

**список управления доступом (access control list, ACL)**

В компьютерной безопасности - список, связанный с объектом и идентифицирующий всех, кто может обращаться к этому объекту, с указанием их прав доступа.

**статическое аннотирование (static summarization)**

Тип создания краткого описания, при котором в результаты поиска включается заданное и сохраненное краткое описание документа. Противоположное понятие - динамическое аннотирование.

**статическое ранжирование (static ranking)**

Тип ранжирования, при котором ранг определяется такими характеристиками документа, как дата, число ссылок на этот документ и т.д. Сравните с динамическим ранжированием.

**стоп-слово (stop word)**

Часто используемое слово (например, *в*, *на* или *и*), которое прикладная программа поиска игнорирует.

**страница начального списка (seed list page)**

В WebSphere Portal - страница XML, содержащая ссылки на доступные страницы на портале. Искатели используют начальный список, чтобы определить документы для просмотра. Кроме того, страница начального списка содержит метаданные, сохраняемые с обработанными документами в индексе поисковой системы предприятия.

**структура признаков (feature structure)**

Базовая структура данных, представляющая результат анализа текста. Структура признаков включает атрибуты с их значениями. Каждая структура признаков относится к некоторому типу, а у каждого типа есть заданный набор допустимых признаков или атрибутов, что напоминает классы Java.

**таксономия (taxonomy)**

Классификация объектов по группам на основе сходства характеристик. В поисковой системе предприятия таксономия организует данные в категории и подкатегории. Смотрите также дерево категорий.

**тип MIME (MIME type)**

Стандарт Интернета для идентификации типа объекта, передаваемого по Интернету.

**тип источников данных (data source type)**

Группировка источников данных в соответствии с протоколом, используемым для доступа к данным.

**удаление стоп-слов (stop word removal)**

Процесс удаления незначущих слов из запроса для повышения релевантности результатов поиска.

**удаленный федератор (remote federator)**

Федератор на сервере, объединяющий набор объектов поиска.

**удалить из очереди (dequeue)**

Удалить элементы из очереди.

**универсальный идентификатор ресурса (Uniform Resource Identifier, URI)**

Краткая строка символов, идентифицирующая логический или физический ресурс.

**универсальный указатель ресурса (Uniform Resource Locator, URL)**

Уникальный адрес информационного ресурса, доступного в сети, например, в Интернете. URL содержит краткое имя протокола, используемого для

доступа к информационному ресурсу, и информацию, используемую этим протоколом для определения положения данного информационного ресурса.

**уникальное имя (distinguished name)**

Имя, уникальным образом идентифицирующее запись в каталоге. Уникальное имя состоит из разделенных запятыми пар атрибут:значение. Это также набор пар "имя-значение" (например CN=имя лица и C=страна или регион), уникально идентифицирующий объект в цифровом сертификате.

**управление идентификацией (identity management)**

Набор API поисковой системы предприятия, управляющих доступом к защищенным данным и позволяющим пользователям выполнять поиск в собраниях без ввода ID пользователя и пароля для каждого репозитория в собрании.

**файл базы данных ключей (key database file)**

Смотрите кольцо ключей. кольцо ключей.

**файл склада ключей (keystore file)**

Кольцо ключей, содержащее как общие ключи, которые хранятся как сертификаты подписавшего, так и секретные ключи, хранимые в персональных сертификатах.

**файлы индекса поиска (search index files)**

Набор файлов, в которых индекс поиска хранится в механизме поиска.

**функция обнаружения (discoverer)**

Функция искателя, определяющая, из каких источников искатель может получать информацию.

**хэш-образ (shingle)**

Строка из идущих последовательно элементов (слов), взятых из некоторой последовательности. Например, из фразы "Это вот такое очень короткое предложение" можно получить следующие хэш-образы из трех слов (триграммы):

Это вот такое  
вот такое очень  
такое очень короткое  
очень короткое предложение

Хэш-образы можно использовать для статистического лингвистического анализа. Например, если у двух разных текстов много общих хэш-образов, эти тексты, вполне возможно, связаны друг с другом.

**шкаф Domino Document Manager (Domino Document Manager cabinet)**

База данных Domino Document Manager, которая служит для организации документов. В шкафах находятся базы данных Domino.

**элемент (token)**

Базовая единица текста, индексируемая поисковой системой предприятия. Элементы могут быть словами языка или другими единицами текста, допускающими индексирование.

**эскейп-символ (escape character)**

Символ, изменяющий значений одного или нескольких символов, следующих за ним.

**язык XML Path (XML Path Language, XPath)**

Язык для уникальной идентификации и указания адресов частей исходных

данных XML; используется в технологиях, связанных с XML, таких как XSLT, XQuery и анализаторы XML. XPath - стандарт World Wide Web Consortium.



---

## Замечания и товарные знаки

---

### Замечания

Это информация о продуктах и услугах, предлагаемых в США.

IBM может предлагать описанные продукты, услуги и возможности не во всех странах. Сведения о продуктах и услугах, доступных в настоящее время в вашей стране, можно получить в местном представительстве IBM. Любые ссылки на продукты, программы или услуги IBM не означают явным или неявным образом, что можно использовать только продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любых продуктов, программ и услуг других фирм лежит на пользователе.

Фирма IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Получение этого документа не означает предоставления каких-либо лицензий на эти патенты. Запросы по поводу лицензий следует направлять в письменной форме по адресу:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive Armonk, NY  
10504-1785  
U.S.A.

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

IBM World Trade Asia Corporation Licensing  
2-31 Roppongi 3-chome,  
Minato-ku  
Tokyo 106-0032, Japan

Следующий абзац не применяется в Великобритании или в любой другой стране, где подобные заявления противоречат местным законам: КОРПОРАЦИЯ INTERNATIONAL BUSINESS MACHINES ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ" БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОВМЕСТИМОСТИ, РЫНОЧНОЙ ПРИГОДНОСТИ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ, НО НЕ ОГРАНИЧИВАЯСЬ ИМИ. В некоторых странах для определенных сделок подобные оговорки не допускаются, таким образом, это утверждение может не относиться к вам.

Данная информация может содержать технические неточности и типографские опечатки. Периодически в информацию вносятся изменения, они будут включены в новые издания этой публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые ссылки в данной информации на сайты, не принадлежащие IBM, приводятся только для удобства и никоим образом не означают поддержки IBM этих сайтов. Материалы этих сайтов не являются частью данного продукта IBM, и вы можете использовать их только на собственную ответственность.

IBM может использовать или распространять присланную вами информацию любым способом, как фирма сочтет нужным, без каких-либо обязательств перед вами.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003  
U.S.A.

Такая информация может быть предоставлена на определенных условиях (в некоторых случаях к таким условиям может относиться оплата).

Лицензированная программа, описанная в данном документе, и все лицензированные материалы, доступные с ней, предоставляются IBM на условиях IBM Customer Agreement (Соглашения IBM с заказчиком), Международного соглашения о лицензиях на программы IBM или эквивалентного соглашения.

Приведенные данные о производительности измерены в контролируемой среде. Таким образом, результаты, полученные в других операционных средах, могут существенно отличаться от них. Некоторые показатели измерены получены в системах разработки и нет никаких гарантий, что в общедоступных системах эти показатели будут теми же. Более того, некоторые результаты могут быть получены путем экстраполяции. Реальные результаты могут отличаться от них. Пользователи должны проверить данные для своих конкретных сред.

Информация о продуктах других фирм получена от поставщиков этих продуктов, из их опубликованных объявлений или из других общедоступных источников. Фирма IBM не проверяла эти продукты и не может подтвердить точность измерений, совместимость или прочие утверждения о продуктах других фирм. Вопросы о возможностях продуктов других фирм следует направлять поставщикам этих продуктов.

Все утверждения о будущих планах и намерениях IBM могут быть изменены или отменены без уведомлений, и описывают исключительно цели фирмы.

Все указанные цены IBM - рекомендуемые розничные цены IBM на сегодняшний день и могут быть изменены без предупреждения. Цены дилеров могут отличаться.

Эта информация приводится только для целей планирования. Приведенная информация может быть изменена до того, как описанные продукты станут доступными.

Эта информация содержит примеры данных и отчетов, иллюстрирующие типичные деловые операции. Чтобы эти примеры были правдоподобны, в них включены имена лиц, названия компаний и товаров. Все эти имена и названия вымышлены и любое их сходство с реальными именами и адресами полностью случайно.

## ЛИЦЕНЗИЯ НА КОПИРОВАНИЕ:

Эта информация содержит примеры прикладных программ на языках программирования, иллюстрирующих приемы программирования для различных операционных платформ. Разрешается копировать, изменять и распространять эти примеры программ в любой форме без оплаты фирме IBM для целей разработки, использования, сбыта или распространения прикладных программ, соответствующих интерфейсу прикладного программирования операционных платформ, для которых эти примеры программ написаны. Эти примеры не были всесторонне проверены во всех возможных условиях. Поэтому IBM не может гарантировать их надежность, пригодность и функционирование.

Каждая копия программ примеров или программ, созданных на их основе, должна содержать следующее замечание об авторских правах:

© (Название вашей компании) (год). Части этого кода получены из Примеров Программ IBM Corp. © Copyright IBM Corp. \_введите год или годы\_. Все права защищены.

Части этого продукта:

- Oracle® Outside In Content Access, Copyright © 1992, 2008, Oracle. Все права защищены.
- IBM XSLT Processor Лицензированные материалы - собственность IBM © IBM Corp., 1999-2008. Все права защищены.

## Товарные знаки

Информацию о товарных знаках IBM смотрите по адресу <http://www.ibm.com/legal/copytrade.shtml>.

Следующие термины являются товарными знаками или зарегистрированными товарными знаками других компаний:

Adobe, Acrobat, Portable Document Format (PDF), PostScript и все товарные знаки на основе Adobe - зарегистрированные товарные знаки или товарные знаки Adobe Systems Incorporated в Соединенных Штатах и/или других странах.

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium - товарные знаки Intel Corporation или ее дочерних фирм в Соединенных Штатах и/или других странах.

Java и все товарные знаки на основе Java - товарные знаки Sun Microsystems, Inc. в Соединенных Штатах и/или в других странах.

Linux - зарегистрированный товарный знак Линуса Торвальдса в Соединенных Штатах и/или других странах.

Microsoft, Windows, Windows NT и логотип Windows - товарные знаки корпорации Microsoft в США и/или других странах.

UNIX зарегистрированный товарный знак The Open Group в США и других странах.

Названия других компаний, продуктов и услуг могут быть товарными знаками или марками сервиса других фирм.



# Индекс

## A

- API
  - описание 11
  - поиска и индексации 11, 223
- API поиска и индексации 11, 223

## E

- esbackup.bat, сценарий 338
- esbackup.sh, сценарий 338
- escrcm.sh, сценарий 61
- escrcm.vbs, сценарий 63
- esrestore.bat, сценарий 339
- esrestore.sh, сценарий 339

## H

- HTTPS, конфигурация сервера поиска 252

## I

- ID прикладных программ 265
- ID сеансов, поисковая система предприятия 373
- ID собраний 38
- ID собрания, правила синтаксиса 35
- IOCP, конфигурация искателя 84
- IP-адреса
  - адаптер обратной связи 27
  - поддержка IPv6 28
  - поддержка двух 27
- IP-адреса, изменение 26

## L

- Lotus Notes
  - интеграция с поисковой системой предприятия 345
  - панель поиска 345
  - сайт обновления подключаемых модулей 345
  - установка подключаемого модуля 345
- Lotus Quickr
  - интеграция с WebSphere Portal 347, 360
  - конфигурация искателя QuickPlace 86
  - конфигурация искателя с начальным списком 89

## O

- OmniFind Enterprise Edition
  - API 11
  - ID сеансов 373
  - анализаторы 5
  - диаграмма потока данных 12
  - изменение IP-адресов 26

## OmniFind Enterprise Edition (продолжение)

- изменение пароля на нескольких серверах 21
- изменение пароля на одном сервере 20
- интеграция с Lotus Notes 345
- интеграция с WebSphere Portal 347
- коды возврата 373
- команды 373
- компоненты 3
- консоль администратора 9
- конфигурирование номера порта 24
- обзор 1
- поддержка двух IP-адресов 27
- поддержка протокола IPv6 28
- программы поиска 12
- серверы индексов 7
- серверы искателей 4
- серверы поиска 8

## R

- removeCjNewlineChars option 173

## S

- SIAPI (API поиска и индексации) 11, 223
- SSL, конфигурация сервера поиска 252
- startccl, сценарий 339

## U

- UIMA
  - добавление механизмов анализа текста в систему 147
  - общие структуры анализа 150, 151
  - описание 145
  - отображение общей структуры анализа на индекс 150
  - отображение общей структуры анализа на реляционные базы данных 151
  - отображение результатов анализа в индексе 150
  - отображение результатов анализа на реляционные базы данных 151
  - отображение элементов XML 148
  - связывание с собраниями 148
- URI
  - влияние на статические оценки 215, 216
  - области действия 191
  - объединенные в результатах поиска 192, 193
  - правила категорий 134, 138
  - просмотр подробностей 306
  - прямые ссылки 207
  - удаление из индекса 194
  - форматы в поисковой системе предприятия 123

## W

- Web Content Management
  - интеграция с WebSphere Portal 347
- WebSphere II Classic Federation 59
- WebSphere II Event Publisher Edition, конфигурация искателя DB2 55
- WebSphere MQ, конфигурация искателя DB2 57
- WebSphere MQ, конфигурация сервера искателя 53, 54
- WebSphere Portal
  - версии 5.1, конфигурирование панели поиска 352
  - версии 5.1, удаление поисковой системы предприятия 353
  - версии 6, конфигурирование панели поиска 358
  - версии 6, конфигурирование Центра поиска 357
  - версии 6, удаление поисковой системы предприятия 361
  - версия 5.1, сценарии интеграции 349
  - версия 6, сценарии интеграции 354
  - интеграция с поисковой системой предприятия 347
  - панель поиска, описание 347
  - параметры перенастройки по умолчанию 371
  - перенастройка дерева категорий 369
  - перенастройка собрания 369
  - перенастройка таксономии 369
  - системы с кластерами 362
  - сценарии установки для поисковой системы предприятия 348
  - Центр поиска, описание 347
- Windows
  - поддержка протокола IPv6 28

## X

- XPath, собственный поиск XML 153

## A

- автоматическое определение
  - кодированные страницы 171
  - языки 170
- авторизация, описание 261
- административные роли
  - администратор поисковой системы предприятия (enterprise search administrator) 262, 263
  - администратор собрания 262, 263
  - конфигурирование 263
  - наблюдатель 262, 263
  - оператор (operator) 262, 263
  - описание 262

- администратор поисковой системы предприятия (enterprise search administrator)
  - изменение пароля на нескольких серверах 21
  - изменение пароля на одном сервере 20
  - конфигурирование роли 263
  - описание 262
- администратор собрания
  - конфигурирование роли 263
  - описание 262
- активные сайты, слежение 309, 310
- анализ текста (text analysis)
  - механизмы анализа текста 147, 148
  - общие структуры анализа 150, 151
  - отображение элементов XML 148
- анализ текста привязки
  - глобальный анализ 276
  - защита собрания 276
  - индексация документов 276
  - описание 259
- анализатор ASCII 160
- анализатор Stellent
  - анализ типов документов 157
  - описание 155
  - связывание типов документов 163
  - типы документов по умолчанию 165
- анализаторы
  - анализ сложных терминов 152
  - анализ типов документов 157
  - анализатор ASCII 160
  - выбор типа анализатора 155
  - документы неизвестных типов 160
  - задачи анализа данных 5
  - запуск 317
  - лингвистическая обработка 169
  - описание 5, 133
  - определение кодовой страницы 171
  - определение форматов документов 155
  - определение языка 170
  - остановка 317
  - поддерживаемые типы документов Stellent 165
  - поддерживаемые языки 169
  - потоки 152
  - правила замены HTML 161, 162
  - сегментация энграмм (n-gram segmentation) 172
  - слежение 317
  - собственный поиск XML 153
  - состояние системы 317
  - типы документов для анализаторов Stellent 163
  - типы документов для служб анализатора 157, 159
  - удаление пробельных символов 173
  - удаление символов новой строки 173
  - файлы без расширений 160
- аннотаторы 145
- архивные файлы
  - поддерживаемые форматы 122
  - просмотр искателем 122
  - форматы URI 123
- аутентификация
  - описание 261

- аутентификация (*продолжение*)
  - отключение для прикладных программ предприятия 292
- аутентификация на основе форм 103, 105

## Б

- базовая аутентификация HTTP 103, 104
- базы данных DB2
  - доступ как ко внешнему источнику 255
  - обращение при помощи искателей DB2 51
  - обращение при помощи искателей баз данных JDBC 67
- базы данных Oracle
  - доступ как ко внешнему источнику 255
  - обращение при помощи искателей DB2 51
  - обращение при помощи искателей баз данных JDBC 67
- базы данных SQL Server
  - обращение при помощи искателей DB2 51
  - обращение при помощи искателей баз данных JDBC 67
- брандмауэры, работа искателя с документами Exchange Server 285

## В

- важность документов
  - в перенастроенных собраниях 369
  - включение для собрания 35
  - восстановить значения по умолчанию 211
  - классы значимости 217, 221
  - словари значимых слов 215
  - статическое 210
  - шаблоны URI 215, 216
- включение расписаний индекса 177
- внедрение портлета поиска
  - wp6\_cluster\_uninstall 366
  - сценарий wp5\_install 349
  - сценарий wp5\_uninstall 353
  - сценарий wp6\_cluster\_install 362
  - сценарий wp6\_install 354
  - сценарий wp6\_uninstall 361
  - файл es.wp5.install.jar 349
  - файл es.wp6.install.jar 354, 362
- внешние искатели
  - конфигурирование 119
  - программы приема данных 119
- внешние источники
  - защита ID прикладных программ 265
  - конфигурация 255
  - описание 255
  - связывание с прикладными программами поиска 258
- внешние источники JDBC
  - драйверы JDBC 255
  - конфигурация 255
  - редактирование 255
  - удаление 255

- внешние источники LDAP
  - конфигурирование 255
  - редактирование 255
  - удаление 255
- возможности доступности данного продукта 407
- восстановление поисковой системы предприятия 337
- восстановление системы 337, 339
- время ответа при поиске
  - оповещения 328
  - слежение 320
- выключение расписаний индекса 177

## Г

- глобальная защита WebSphere
  - настройщик программ поиска 244
  - отключение 292
  - свойства программы поиска 243
- глобальное пространство для искателя Web 108
- глобальный анализ
  - анализ текста привязки 259, 276
  - обнаружение повторений документов 186, 259, 264
  - описание 7
- глубина путей URL 96

## Д

- даты просмотренных искателем документов
  - конфигурирование для искателей Web 111
- дельта-индексы
  - описание 7
- дельта-индексы indexes
  - выявление изменений 179
  - одновременное построение 178
  - описание 175
  - планирование 176
- деревья категорий
  - описание 136
  - перенастройка из WebSphere Portal 369
- динамическое ранжирование (dynamic ranking) 209
- динамическое составление сводок (dynamic summarization) 205
- директивы no-follow
  - конфигурирование 111
  - описание 110
- директивы no-index
  - конфигурирование 111
  - описание 110
- доверенные серверы Lotus Domino 286
- документация
  - HTML 405
  - PDF 405
  - поиск 405
- документация по поисковой системе предприятия в формате HTML 405
- Документация по поисковой системе предприятия в формате PDF 405

документы HTML  
 поиск 143, 144  
 правила замены 161, 162  
 синтаксический анализ 161, 162

Документы XML  
 поиск 140  
 собственный поиск XML 153

документы неизвестных типов 160

домены Lotus Domino 286

домены Windows 289

драйверы JDBC  
 для внешних источников JDBC 255  
 для искателей баз данных JDBC 67

### 3

запросы с высоким значением возврата  
 коэффициенты значимости по  
 умолчанию 221  
 описание 217

запросы с низким значением возврата  
 коэффициенты значимости по  
 умолчанию 221  
 описание 217

запуск  
 анализаторы 317  
 мастер по перенастройке 369  
 настройщик программ поиска 244  
 построение индекса 318  
 программа приема данных 322  
 программы поиска 251  
 серверы искателей 308  
 серверы поиска 302, 320  
 серверы поисковой системы  
 предприятия 297

зарезервированное поле es\_special\_  
 field.default\_metadata\_field 221

зарезервированное поле  
 es\_special\_field.default\_field 221

зарезервированное поле  
 es\_special\_field.regular\_text 221

защита  
 ID прикладных программ поиска 265  
 административные роли 263  
 анализ текста привязки 276  
 аутентификация 261, 292  
 включение для поисковой системы  
 предприятия 277  
 включение для собрания 35, 259  
 глобальная защита WebSphere 292  
 глобальная, WebSphere Application  
 Server 277, 278  
 документы Lotus Domino 286  
 домены Windows 289  
 конфигурация HTTPS для поиска 252  
 конфигурация SSL для поиска 252  
 обнаружение повторов  
 документов 264  
 обход проверки управления доступом  
 на уровне документов 294  
 объединенные результаты поиска 295  
 описание 259  
 отключение для прикладной  
 программы предприятия 292  
 поддержка единой регистрации 273  
 подключаемые модули искателей 121  
 профили пользователей 272

защита (*продолжение*)  
 реестр пользователей LDAP 278  
 управление доступом 261  
 управление идентификацией (identity  
 management) 269, 275  
 уровня документов 265, 266, 267, 269,  
 275, 294  
 уровня собрания 264, 294  
 установка на нескольких серверах 281  
 установка на одном сервере 280

защита Local User, искатели  
 QuickPlace 288

защита с единой регистрацией (SSO)  
 конфигурация 275  
 управление идентификацией (identity  
 management) 273

защита уровня документов  
 для документов Lotus Domino 286  
 для файловых систем Windows 289  
 документы Lotus Domino 286  
 конфигурация искателей 41  
 маркеры защиты 267  
 описание 259, 265  
 поддержка единой регистрации 273  
 подключаемые модули искателей 121  
 постфильтрация документов 266  
 предварительная фильтрация  
 результатов 266  
 проверка в реальном времени 269  
 проверка параметров аутентификации  
 текущего пользователя 269  
 профили пользователей 272  
 управление доступом через  
 индекс 267  
 управление идентификацией (identity  
 management) 269, 275

защита уровня собрания  
 ID прикладных программ 265  
 анализ текста привязки 276  
 включение 35  
 обнаружение повторов  
 документов 264  
 описание 259, 264

## I

идеографические языки 169

индексы  
 включение расписания 177, 318  
 влияние символов подстановки 189  
 выявление изменений 179  
 изменение расписания 177  
 команда startIndexBuild 179  
 области действия 181, 191  
 объединенные URI 181, 192, 193  
 одновременное построение 178  
 описание 7, 175  
 оповещения 328  
 отключение расписания 177, 318  
 параллельное построение 178  
 планирование 176  
 символы подстановки 181, 187, 190  
 слежение 318, 319  
 текст привязки 276  
 удаление URI 181, 194  
 удаление из очереди 319  
 форматы URI 123

интеграция с WebSphere Portal  
 Lotus Quickr 347  
 Web Content Management 347  
 описание 347  
 системы с кластерами 362  
 сценарии установки 348  
 сценарий wp5\_install 349  
 сценарий wp6\_cluster\_install 362  
 сценарий wp6\_install 354  
 файл es.wp5.install.jar 349  
 файл es.wp6.install.jar 354, 362

интервалы повторного просмотра для  
 искателей Web 101

искатели  
 Content Edition 46, 47, 48  
 DB2 51  
 DB2 Content Manager 60  
 Domino Document Manager 64  
 Exchange Server 66, 285  
 NNTP 75  
 QuickPlace 86  
 Web 91  
 Web Content Management 113, 116  
 WebSphere Portal 115, 116  
 архивные файлы 122  
 база данных JDBC 67, 69, 71  
 базовые значения для 41  
 включение защиты на уровне  
 документов 41  
 защита уровня документов 265  
 комбинирование типов искателей 41  
 начальные значения для 43  
 обзор конфигурации 41  
 описание 4  
 параметры перенастройки по  
 умолчанию 371  
 планирование 41, 45  
 поддержка для внешних 119  
 подключаемые модули 121  
 Примечания 76, 78  
 программы приема данных 119  
 редактирование пространств для  
 искателей 44  
 редактирование свойств искателя 44  
 с начальным списком 89  
 слежение 308  
 создание 43  
 состояние системы 308  
 удаление 45  
 файловая система UNIX 91  
 файловая система Windows 117  
 форматы URI 123

Искатели Content Edition  
 конфигурация 46  
 конфигурирование в Windows 50  
 конфигурирование в операционной  
 системе AIX 49  
 конфигурирование в операционной  
 системе Linux 49  
 конфигурирование в операционной  
 среде Solaris 49  
 непосредственный режим 47  
 режим сервера 48  
 форматы URI 123

Искатели DB2  
 WebSphere II Classic Federation 59  
 задание в Linux 53

- Искатели DB2 *(продолжение)*
    - конфигурация 51
    - конфигурация WebSphere II Event Publisher Edition 55
    - конфигурация WebSphere MQ 57
    - конфигурирование в AIX 53
    - конфигурирование в Solaris 53
    - конфигурирование в Windows 54
    - конфигурирование публикации событий 53, 54
    - публикация событий 51
    - установка WebSphere MQ в AIX 53
    - установка WebSphere MQ в Linux 53
    - установка WebSphere MQ в Solaris 53
    - установка WebSphere MQ в Windows 54
    - форматы URI 123
  - Искатели DB2 Content Manager
    - конфигурация 60
    - конфигурирование в Windows 63
    - конфигурирование в операционной системе AIX 61
    - конфигурирование в операционной системе Linux 61
    - конфигурирование в операционной среде Solaris 61
    - форматы URI 123
  - Искатели Domino Document Manager
    - конфигурация 64
    - конфигурация протокола DIIOP 83
    - конфигурирование IOCP 84
    - конфигурирование в Windows 81
    - конфигурирование в операционной системе AIX 79
    - конфигурирование в операционной системе Linux 79
    - конфигурирование в операционной среде Solaris 79
    - протокол NRPC 79, 81
    - форматы URI 123
  - Искатели Exchange Server
    - защищенные документы 285
    - конфигурация 66
    - форматы URI 123
  - искатели NNTP, конфигурирование 75
  - Искатели Notes
    - доверенный сервер Lotus Domino 286
    - конфигурация 76
    - конфигурация протокола DIIOP 83
    - конфигурирование IOCP 84
    - конфигурирование в Windows 81
    - конфигурирование в операционной системе AIX 79
    - конфигурирование в операционной системе Linux 79
    - конфигурирование в операционной среде Solaris 79
    - конфигурирование защиты уровня документов 286
    - правила отображения полей 78
    - проверка параметров аутентификации текущего пользователя 286
    - протокол NRPC 79, 81
    - советы по использованию 78
    - форматы URI 123
  - Искатели QuickPlace
    - защита Local User 288
  - Искатели QuickPlace *(продолжение)*
    - конфигурация 86
    - конфигурация протокола DIIOP 83
    - конфигурирование Directory Assistance 289
    - конфигурирование IOCP 84
    - конфигурирование в Windows 81
    - конфигурирование в операционной системе AIX 79
    - конфигурирование в операционной системе Linux 79
    - конфигурирование в операционной среде Solaris 79
    - конфигурирование пользователя Domino 288
    - протокол NRPC 79, 81
    - форматы URI 123
  - искатели Web
    - активные сайты 309, 310
    - глобальное пространство для искателя 108
    - даты просмотренных искателем документов 111
    - директивы по-follow 110, 111
    - директивы по-index 110, 111
    - интервалы повторного просмотра 101
    - конфигурация 91
    - конфигурирование опознавательных файлов 108
    - начальные URL 96, 101
    - недавно просмотренные URL 309
    - ограничение пространства для искателя 96
    - опознавательные файлы (cookie) 107
    - поддержка JavaScript 95
    - подробности потока 309, 310
    - подробности сайта 309
    - пользовательские агенты 92
    - посещение URL при первой возможности 101
    - правила для искателя 96
    - произвольные страницы ошибок 102
    - прокси-серверы 106
    - сайты, защищенные паролями 103, 104, 105
    - скорость работы искателя 311
    - слежение 309
    - создание отчетов 312
    - состояние URL 309
    - состояние системы 309
    - файл followindex.rules 110, 111
    - файлы robots.txt 92, 93
    - формат опознавательных файлов 107
    - хронология искателя 309
  - Искатели Web Content Management
    - конфигурация 113
    - копирование URL сайтов 116
    - форматы URI 123
  - Искатели WebSphere Portal
    - конфигурация 115
    - форматы URI 123
  - Искатели баз данных JDBC
    - конфигурация 67
    - поддерживаемые драйверы 67
    - подключаемый модуль для просмотра нескольких таблиц 69, 71
    - просмотр нескольких таблиц 69, 71
  - Искатели баз данных JDBC *(продолжение)*
    - форматы URI 123
  - искатели с начальным списком форматы URI 123
  - Искатели с начальным списком интеграция с WebSphere Portal 347, 360
    - конфигурация 89
  - Искатели файловой системы UNIX
    - конфигурация 91
    - форматы URI 123
  - Искатели файловой системы Windows
    - конфигурация 117
    - конфигурирование защиты уровня документов 289
    - форматы URI 123
- ## К
- категории
    - вложенные подкатегории 136
    - деревья категорий 136
    - на основе правил 134, 137
    - описание 134
    - перенастройка из WebSphere Portal 369
    - поиск 134
    - создание 138
    - тип категоризации 137
    - форматы URI 123
  - категории на основе правил
    - выбор типа категоризации 137
    - описание 134
    - создание 138
  - китайский
    - сегментация энграмм (n-gram segmentation) 172
    - удаление символов новой строки 173
  - классы значимости
    - запросы с высоким значением возврата 217, 221
    - запросы с низким значением возврата 217, 221
    - значения по умолчанию 221
    - конфигурация 219, 220
    - обнаружение повторений документов 217
    - описание 217
    - отображение полей на 219
  - кластеры
    - WebSphere Portal 362
  - кластеры WebSphere Portal
    - сценарии интеграции 362
    - удаление поисковой системы предприятия 366
    - указания по интеграции 362
  - клонирование
    - искатели 41
    - программы поиска 246
  - ключевые слова в прямых ссылках 207
  - кодировка Unicode 171
  - кодировка страниц
    - автоматическое определение 171
    - поддерживаемые 171
  - коды возврата, поисковая система предприятия 373

коды состояния HTTP  
отчет искателя Web 312  
полученные искателями Web 312  
команда esadmin 373  
команда esadmin startSearch 302  
команда esadmin stopIndex 181  
команда esadmin stopSearch 302  
команда esadmin system startall 373  
команда esadmin system stopall 373  
команда esexchangeproxpw 253  
команда esexchangeproxpw 252  
команда esexchangeproxpw  
конфигурация с несколькими  
серверами 281  
конфигурация с одним сервером 280  
команда startIndexBuild 179  
команды, поисковая система  
предприятия 373  
консоль администратора  
интерфейс 15  
описание 9  
регистрация 19  
сводка задач 15  
конфигурация SMTP-сервера 332  
конфигурация адаптера обратной  
связи 27  
конфигурация доверенного сервера 286  
конфигурирование Directory  
Assistance 289  
конфигурирование журналов  
запросов 334  
конфигурирование пользователя Domino,  
искатели QuickPlace 288  
корейский  
анализ сложных терминов 152  
сегментация энграмм (n-gram  
segmentation) 172  
КОРНЕВОЙ\_КАТАЛОГ\_  
УСТАНОВКИ\_ES, описание 20  
КОРНЕВОЙ\_КАТАЛОГ\_УЗЛА\_ES,  
описание 20, 21  
КОРНЕВОЙ\_КАТАЛОГ\_УСТАНОВКИ\_ES,  
описание 21  
коэффициенты значимости  
для классов значимости 219, 221  
для словарей значимых слов 213  
для шаблонов URI 215, 216  
конфигурация класса значимости 217,  
220  
кэш поиска (search cache)  
конфигурирование 198  
описание 198

**Л**  
лингвистическая поддержка  
коды языков 169  
национальные версии 169  
определение кодовой страницы 171  
определение языка 170  
пользовательский анализ текста 145  
сегментация энграмм (n-gram  
segmentation) 172  
семантический поиск 145, 153  
словари значимых слов 213  
словари синонимов 198  
словари стоп-слов 201

лингвистическая поддержка  
(продолжение)  
собственный поиск XML 153  
удаление пробельных символов 173

**М**  
максимальный интервал повторного  
просмотра 101  
маркеры защиты  
защита уровня документов 267  
как запретить для собрания 294  
конфигурация искателей 267  
мастер по перенастройке  
запуск 369  
описание 369  
параметры искателей по  
умолчанию 371  
параметры собраний по  
умолчанию 371  
собрания 369  
таксономии на основе правил 369  
файл журнала 372  
мастер по собраниям 33  
механизмы анализа текста  
добавление в систему 147  
описание 145  
отображение общей структуры анализа  
на реляционные базы данных 151  
отображение результатов анализа в  
индексе 150  
отображение результатов анализа на  
реляционные базы данных 151  
отображение элементов XML 148  
связывание с собраниями 148  
минимальный интервал повторного  
просмотра 101  
многобайтная кодировка 171  
модуль bos.iochr.te 84  
модуль IOCP, конфигурирование  
искателя 84

**Н**  
наблюдатель  
конфигурирование роли 263  
описание 262  
настройка программы поиска 243, 244  
настройщик программ поиска  
запуск 244  
файл config.properties 244  
национальные версии  
поиск 169  
синтаксический анализ 169  
начальные URL для искателей Web 96,  
101  
недавно просмотренные URL,  
слежение 309  
непосредственный режим, репозитории  
Content Edition 47  
номер порта, поисковая система  
предприятия 24

**О**  
области действия  
описание 191  
поиск 191  
создание 191  
форматы URI 123  
шаблоны URI 191  
обнаружение 4  
обнаружение повторений документов  
включение защиты 264  
глобальный анализ 186, 264  
конфигурация класса значимости 217  
описание 186, 259  
обработка текста  
аннотаторы 145  
механизмы анализа текста 145  
общие структуры анализа 145  
общие структуры анализа  
описание 145  
отображение на индекс 150  
отображение на реляционные базы  
данных 151  
отображение элементов XML на 148  
объединенные URI  
конфигурирование 193  
ограничения защиты 295  
описание 192  
объединенные результаты поиска  
конфигурирование 193  
ограничения защиты 295  
описание 192  
ограничение пространства для искателя  
Web 96  
однобайтная кодировка 171  
Одновременное построение индексов 178  
окно Защита, описание 15  
окно Система, описание 15  
окно Слежение, описание 15  
Окно Собрания  
описание 15  
создание собраний 35  
оператор (operator)  
конфигурирование роли 263  
описание 262  
операционная система AIX  
конфигурация искателя Content  
Edition 49  
конфигурация искателя DB2 53  
конфигурация искателя DB2 Content  
Manager 61  
конфигурация искателя Domino  
Document Manager 79  
конфигурация искателя Notes 79  
конфигурация искателя QuickPlace 79  
конфигурирование публикации  
событий 53  
операционная система Linux  
конфигурация искателя Content  
Edition 49  
конфигурация искателя DB2 53  
конфигурация искателя DB2 Content  
Manager 61  
конфигурация искателя Domino  
Document Manager 79  
конфигурация искателя Notes 79  
конфигурация искателя QuickPlace 79

- операционная система Linux *(продолжение)*
    - конфигурирование публикации событий 53
    - операционная среда Solaris
      - конфигурирование публикации событий 53
  - операционная система Solaris
    - конфигурация искателя DB2 53
  - операционная система Windows
    - конфигурация искателей 81
    - конфигурация искателя Content Edition 50
    - конфигурация искателя DB2 54
    - конфигурация искателя DB2 Content Manager 63
    - конфигурирование публикации событий 54
  - операционная среда Solaris
    - конфигурация искателя Content Edition 49
    - конфигурация искателя DB2 Content Manager 61
    - конфигурация искателя Domino Document Manager 79
    - конфигурация искателя Notes 79
    - конфигурация искателя QuickPlace 79
  - оповещения
    - время ответа при поиске 328
    - документы, просматриваемые искателем 328
    - индексируемые документы 328
    - конфигурация SMTP-сервера 332
    - ограничения на индекс 305
    - описание 327
    - опции электронной почты 328, 329
    - получение электронной почты 332
    - свободное место на серверах 329
    - уровня системы 329
    - уровня собрания 305, 328
  - оповещения о свободном месте 329
  - опознавательные файлы для искателя Web
    - конфигурирование 108
    - описание 107
    - формат 107
  - опции поиска
    - для результатов поиска 182
    - параметрический поиск (parametric search) 182
    - поиск по полю (fielded search) 182
    - полное соответствие 182
    - поля сортировки 182
    - свободный текстовый поиск (free text search) 182
    - содержимое документов 182
  - опция removeCjNewlineCharsMode 173
  - остановка
    - анализаторы 317
    - построение индекса 318, 319
    - серверы искателей 308
    - серверы поиска 302, 320
    - серверы поисковой системы предприятия 297, 299
  - отброшенные документы
    - в файлах журналов 325
    - конфигурирование файлов журналов 323
  - отброшенные документы *(продолжение)*
    - описание 323
    - отчеты 324
  - отображение
    - общей структуры анализа на реляционные базы данных 151
    - общие структуры анализа в индексе 150
    - отображение элементов XML на общую структуру анализа 148
    - полей на классы значимости 219
    - поля поиска HTML 144
    - поля поиска XML 140
    - результатов анализа на реляционные базы данных 151
  - отслеживание документов
    - включение 323
    - конфигурирование файлов журналов 323
    - описание 323
    - отключение 323
    - отчеты 324
    - файлы журнала 325
  - оценка на основе текста (text-based scoring) 209
  - оценка системных ресурсов 305
  - очередь индексации (index queue) 319
- П**
- панель поиска WebSphere Portal
    - версии 5.1, перенаправление на поисковую систему предприятия 352
    - версии 6, перенаправление на поисковую систему предприятия 358
  - параллельное построение индексов 178
  - параметр distinctRecentQueryCheck 321
  - параметр newHtmlTagReplacement 161
  - параметрические поля
    - описание 182
    - числовая сортировка 182
  - пароль администратора
    - изменение на нескольких серверах 21
    - изменение на одном сервере 20
  - пароль, администратор поисковой системы предприятия 20, 21
  - перенастройка
    - собрания 369
    - таксономия на основе правил 369
  - планирование
    - искатели 41, 45
    - построение индекса 176, 177
  - повторное посещение URL при первой возможности 101
  - поддержка JavaScript в искателях Web 95
  - подключаемые модули
    - Искатели баз данных JDBC 69, 71
    - просмотр нескольких структурированных таблиц 69, 71
  - подключаемые модули искателей 121
  - подключаемые модули, для искателей 121
  - подключаемый модуль для Lotus Notes сайт обновления 345
  - установка 345
  - подключаемый модуль нескольких структурированных таблиц 69, 71
  - подробности об URI
    - отброшенные документы 324
    - слежение 306
  - подробности потока, отслеживание 309
  - подробные отчеты по сайтам
    - описание 309
    - создание 312
  - поиск
    - документы HTML 143, 144
    - Документы XML 139, 140, 148
    - категории 134
    - прямые ссылки 207
    - собрания 182
  - поиск по полю
    - описание 182
    - сортировка строк 182
  - поисковая система предприятия
    - API 11
    - ID сеансов 373
    - административные роли 262
    - анализаторы 5, 133
    - восстановление из резервной копии 339
    - диаграмма потока данных 12
    - запуск серверов 297
    - запуск серверов поиска 302
    - защита 259
    - защита уровня документов 265
    - защита уровня собрания 264
    - интеграция с Lotus Notes 345
    - интеграция с WebSphere Portal 347
    - коды возврата 373
    - команды 373
    - компоненты 3
    - консоль администратора 9
    - конфигурирование номера порта 24
    - обзор 1
    - остановка серверов 297, 299
    - остановка серверов поиска 302
    - программы поиска 12
    - резервное копирование 338
    - серверы индексов 7, 175
    - серверы искателей 4, 41
    - серверы поиска 8, 197
    - слежение 305
    - сценарии восстановления 337
    - сценарии резервного копирования 337
    - файлы журнала 327
    - форматы URI 123
  - поисковые поля точного соответствия, описание 182
  - полные индексы
    - выявление изменений 179
    - одновременное построение 178
    - описание 7, 175
    - планирование 176
  - пользователь WebSphere Application Server
    - пароль для конфигурации с несколькими серверами 281
    - пароль для конфигурации с одним сервером 280
  - пользовательские агенты 92
  - пользовательский анализ текста
    - механизмы анализа текста 147, 148
    - описание 145

- пользовательский анализ текста
  - (продолжение)
  - отображение общей структуры анализа на индекс 150
  - отображение общей структуры анализа на реляционную базу данных 151
  - отображение результатов анализа в индексе 150
  - отображение результатов анализа на реляционную базу данных 151
  - отображение элементов XML 148
- поля метаданных, столбчатые диаграммы для частей результатов 247
- поля поиска HTML
  - описание 143
  - отображение элементов 143, 144
  - создание 144
  - элементы Dublin Core 144
- поля поиска XML
  - описание 139, 148
  - отображение элементов 139, 140, 148
  - создание 140
- поля результатов поиска, описание 182
- поля сортировки
  - сортировка строк 182
  - числовая сортировка 182
- поля, отображение на классы значимости 219
- портлеты
  - интеграция с WebSphere Portal 5.1 349
  - интеграция с WebSphere Portal 6 354, 360
  - интеграция с кластерами WebSphere Portal 362
  - конфигурирование для источников Lotus Quickr 360
  - описание 347
  - поисковая система предприятия 347
  - удаление из WebSphere Portal 5.1 353
  - удаление из WebSphere Portal Версии 6 361
  - удаление из кластеров WebSphere Portal 366
- посещение URL при первой возможности 101
- последние запросы
  - подсчет 321
- последние запросы, слежение 320
- построение индекса
  - выявление изменений 179
  - запуск 318
  - команда startIndexBuild 179
  - одновременное 178
  - описание 175
  - остановка 181, 318, 319
  - параллельное 178
  - планирование 177
  - состояние системы 319
- построение индексов 175
- поток данных, поисковая система предприятия 12
- потоки
  - анализатор 152
  - искатель Web 310
- правила IP-адресов для искателей Web 96
- правила адресов для искателей Web 96
- правила доменов для искателей Web 96
- правила замены HTML 161, 162
- правила категорий
  - конфигурирование 138
  - содержимое документов 134, 138
  - шаблоны URI 134, 138
- правила префиксов для искателей Web 96
- прикладные программы предприятия
  - программа ESAdmin 292
  - программа ESSearchApplication 292
  - программа ESSearchServer 292
- пример программы поиска
  - включение обязательного HTTPS 252
  - включение обязательного SSL 252
  - внедрение по умолчанию 243
  - клонирование 246
  - описание 12, 224
  - отключение защиты 292
  - файл config.properties 226, 243
  - функции поиска 223, 224
- проверка для запроса 269
- проверка параметров аутентификации текущего пользователя 269, 286, 289
- программа ESAdmin
  - отключение защиты 292
  - регистрация 19
- программа ESSearchApplication
  - запуск 251
  - отключение защиты 292
  - файл config.properties 226, 243
- программа ESSearchServer
  - отключение защиты 292
  - перезапуск 243, 244
- программа поиска по умолчанию 243
- программа приема данных
  - конфигурирование 119
  - перезапуск 119, 322
  - слежение 322
- программы поиска
  - ID прикладных программ 265
  - доступ 251
  - запуск 251
  - защита уровня собрания 265
  - настройка 244
  - описание 12
  - пользовательские 223
  - пример 223, 224
  - связывание с внешними источниками 258
  - связывание с собраниями 224
- произвольные страницы ошибок, искатели Web 102
- прокси-серверы 106
  - конфигурация сервера поиска 253
- прокси-серверы HTTP 106
- просмотр
  - подробности об URI 306
  - файлы журнала 335
  - файлы журналов отброшенных документов 325
- пространство для искателя (crawl space)
  - конфигурация искателя Web 96
  - описание 4
  - оповещения 328
  - редактирование 44
- протокол DPOP, конфигурация искателя 83
- протокол IPv6 28
- протокол NRPC, конфигурирование искателя 79, 81
- протокол Robots Exclusion
  - идентификация пользовательского агента 92
  - соответствие искателя Web 93
- профили пользователей
  - конфигурация 275
  - описание 272
- прямые ссылки
  - описание 207
  - поиск 207
  - создание 207
  - форматы URI 123
- публикация событий
  - конфигурация искателя DB2 55, 57
  - конфигурирование в Windows 54
  - конфигурирование в операционной системе AIX 53
  - конфигурирование в операционной системе Linux 53
  - конфигурирование в операционной среде Solaris 53
  - описание 51

## Р

- ранжирование документов
  - восстановить значения по умолчанию 211
- ранжирование результатов поиска
  - восстановить значения по умолчанию 211
  - динамическое 209
  - классы значимости 217, 219, 220, 221
  - описание 209
  - оценка на основе текста (text-based scoring) 209
  - словари значимых слов 215
  - статическое 210, 211
  - файл runtime.properties 211
  - шаблоны URI 215, 216
- расширение запроса
  - влияние на время построения индекса 189
  - влияние на размер индекса 189
  - описание 187
- расширение индекса
  - влияние на время построения индекса 189
  - влияние на размер индекса 189
  - описание 187
- расширения файлов
  - исключение из пространств для искателей Web 96
  - поддерживаемые анализаторами Stellent 163, 165
  - поддерживаемые анализатором собраний 157, 159
- регистрация на консоли администратора 19
- редактирование
  - программы приема данных 119
  - пространства для искателей 44
  - свойства искателя 44
  - свойства программы поиска 226, 243
  - собрания 36

- реестр пользователей LDAP 278
- режим сервера, репозитории Content Edition 48
- резервная копия системы 337, 338
- резервное копирование поисковой системы предприятия 337, 338
- результаты поиска
  - настройка сводок 206
  - сводки 206
- результаты поиска (search results)
  - группировка 192, 193
  - динамическое ранжирование (dynamic ranking) 209
  - динамическое составление сводок (dynamic summarization) 205
  - конфигурация класса значимости 217, 219, 220, 221
  - конфигурирование шаблонов URI 215
  - настройка сводок 205
  - объединение 192, 193, 295
  - описание 209
  - оценка на основе текста (text-based scoring) 209
  - постфильтрация 266
  - предварительная фильтрация 266
  - ранжирование (ranking) 216
  - расширение символов
    - подстановки 190
  - сводки 205
  - символы подстановки 187
  - статическое ранжирование (static ranking) 210

**С**

- сайты, защищенные паролями 103
  - аутентификация на основе форм 105
  - базовая аутентификация HTTP 104
- свободный текстовый поиск,
  - описание 182
- сводка задач, консоль управления 15
- сводки
  - динамическое 205
  - настройка 205, 206
- сводки документов
  - настройка 205
  - редактирование свойств для 206
- свойства искателя
  - описание 4
  - редактирование 44
- сегментация энграмм (n-gram segmentation) 172
- семантический поиск 145, 148, 153
- серверы анализаторов
  - конфигурация потоков 152
- серверы индексов
  - запуск 297
  - остановка 299
- серверы искателей
  - запуск 297, 308
  - остановка 299, 308
- серверы поиска
  - запуск 297, 302, 320
  - конфигурация HTTPS 252
  - конфигурация SSL 252
  - конфигурация с прокси-сервером 253
  - кэш поиска (search cache) 198

- серверы поиска (*продолжение*)
  - описание 8, 197
  - остановка 299, 302, 320
  - повторное внедрение словарей 203
  - подсчет числа запросов 321
  - последние запросы 320, 321
  - связанные с ними словари
    - синонимов 201
  - связывание значимых слов со словарями 214, 215
  - связывание словарей стоп-слов 202, 203
  - слежение 320, 321
  - словари значимых слов 213
  - словари синонимов 198, 200
  - словари стоп-слов 201
  - состояние системы 320
  - хронология времени ответа 320
  - частые запросы 320, 321
- серверы поисковой системы предприятия
  - изменение IP-адресов 26
  - поддержка двух IP-адресов 27
  - поддержка протокола IPv6 28
- символы подстановки
  - в запросах 187
  - расширение запроса 187, 190
  - расширение индекса 187, 189, 190
- синтаксис запроса XML, собственный 153
- системные ресурсы
  - оценка 305
- скорость работы искателя, слежение 311
- слежение
  - Активные сайты искателя Web 310
  - анализаторы 317
  - искатели 308
  - искатели Web 309
  - отброшенные документы 324, 325
  - подробности об URI 306
  - Подробности потока искателя Web 310
  - поисковая система предприятия 305
  - последние запросы 320, 321
  - программа приема данных 322
  - серверы поиска 320, 321
  - Скорость просмотра искателя Web 311
  - собрания 306
  - файлы журнала 335
  - хронология времени ответа 320
  - частые запросы 320, 321
- словари значимых слов
  - добавление в систему 214
  - описание 213
  - повторное внедрение 203
  - связывание с собранием 215
- словари синонимов
  - добавление в систему 200
  - описание 198
  - повторное внедрение 203
  - связывание с собранием 201
- словари стоп-слов
  - добавление в систему 202
  - описание 201
  - повторное внедрение 203
  - связывание с собранием 203
- сложные термины, анализ 152

- собрания
  - защита 264
  - защита ID прикладных программ 265
  - защита повторений документов 264
  - защита текста привязки 276
  - обнаружение повторений документов 186
  - обход проверки управления доступом на уровне документов 294
  - описание 3
  - определение ID 38
  - оценка размера 35
  - оценка ресурсов 305
  - параметры перенастройки по умолчанию 371
  - перенастройка из WebSphere Portal 369
  - поиск 182
  - редактирование 36
  - связывание с прикладными программами поиска 224
  - серверы поиска 197
  - синтаксический анализ 133
  - система объединения (federation) 33
  - слежение 306
  - создание в окне Собрания 35
  - создание с помощью мастера по собраниям 33
  - состояние системы 306
  - способы создания 33
  - удаление 38
  - файл MigrationWizard.log 369
  - черновое состояние 33
- собрания объединения 33
- собственный поиск XML 153
- содержимое документов, описание 182
- соединитель Java для DB2 Content Manager 61, 63
- создание
  - искатели 43
  - категории на основе правил 138
  - области действия 191
  - отчеты искателя Web 312
  - поля поиска HTML 144
  - поля поиска XML 140
  - прямые ссылки 207
  - собрания 33, 35
- сообщения об ошибках
  - конфигурация SMTP-сервера 332
  - получение электронной почты 330, 332
  - просмотр файлов журнала 335
  - просмотр файлов журналов отброшенных документов 325
- состояние системы
  - анализаторы 317
  - искатели 308
  - искатели Web 309
  - построение индекса 319
  - серверы поиска 320
  - собрания 306
- статическое ранжирование (static ranking)
  - в перенастроенных собраниях 369
  - включение для собрания 35
  - описание 210

- столбчатые диаграммы
  - классы Java для частных результатов 247
  - частые результаты 247
- страница Анализ, описание 15
- страница Журнал, описание 15
- страница Индекс, описание 15
- страница Искатель, описание 15
- страница Поиск, описание 15
- строка USC 269
- строка контекста защиты пользователя 269
- сценарии
  - esbackup.bat 338
  - esbackup.sh 338
  - escrcm.sh 61
  - escrcm.vbs 63
  - escrdb2.sh 53
  - escrdb2.vbs 54
  - escrmote.sh 79
  - escrmote.vbs 81
  - escrvbr.sh 49
  - escrvbr.vbs 50
  - esrestore.bat 339
  - esrestore.sh 339
  - startccl 339
- сценарии восстановления
  - запуск 339
  - описание 337
- сценарии резервного копирования
  - запуск 338
  - описание 337
- сценарии установки
  - WebSphere Portal 348
- сценарий eschangepw 20, 21
- сценарий escrdb2.sh 53
- сценарий escrdb2.vbs 54
- сценарий escrmote.sh 79
- сценарий escrmote.vbs 81
- сценарий escrvbr.sh 49
- сценарий escrvbr.vbs 50
- сценарий wp5\_install 349
- сценарий wp5\_uninstall 353
- сценарий wp6\_cluster\_install 362
- сценарий wp6\_cluster\_uninstall 366
- сценарий wp6\_install 354
- сценарий wp6\_uninstall 361

**Т**

- таксономии, перенастройка из WebSphere Portal 369
- таксономия на основе правил, перенастройка из WebSphere Portal 369
- тип категоризации
  - выбор 35, 137
  - на основе правил 134
- типы MIME, включение в пространства для искателей Web 96
- типы документов
  - для анализаторов Stellent 163
  - для служб анализатора 157, 159
  - определение 155
  - поддерживаемые анализаторами Stellent 165
  - синтаксический анализ 157

- типы искателей
  - базовые значения для 41
  - объединение в собрании 41
- типы источников данных
  - DB2 для z/OS 51
  - Lotus Quickr for Domino 86
  - Lotus Quickr для WebSphere Portal 89
  - базы данных CA-Datcom 51
  - базы данных DB2 51, 67, 255
  - базы данных DB2 для iSeries 51
  - базы данных Domino Document Manager 64
  - базы данных IMS 51
  - базы данных Informix 51
  - базы данных JDBC 67, 69, 71, 255
  - базы данных Notes 76, 78
  - базы данных Oracle 51, 67, 255
  - базы данных QuickPlace 86
  - базы данных Software AG Adabas 51
  - базы данных SQL Server 51, 67
  - базы данных Sybase 51
  - базы данных VSAM 51
  - группы новостей NNTP 75
  - общедоступные папки Exchange Server 66
  - поддерживаемые поисковой системой предприятия 2
  - поддержка для внешних 2, 11
  - реляционные базы данных 51
  - репозитории Content Edition 46, 47, 48
  - сайты 91
  - сайты Web Content Management 113
  - сайты WebSphere Portal 115
  - типы элементов DB2 Content Manager 60
  - файловые системы UNIX 91
  - файловые системы Windows 117

## У

- уведомления по электронной почте
  - конфигурация SMTP-сервера 332
  - о сообщениях 332
  - об оповещениях 332
- удаление
  - индексов из очереди 319
  - искатели 45
  - собрания 38
- удаление URI из индекса 194
- удаление пробельных символов 173
- удаление символов новой строки 173
- управление доступом
  - защита с единой регистрацией (SSO) 273
  - защита уровня документов 267
  - как запретить для собрания 294
  - описание 261
  - проверка текущего пользователя 269
  - требования для Lotus Domino 286
  - требования для файловых систем Windows 289
  - управление идентификацией (identity management) 269, 272
- управление идентификацией (identity management)
  - контекст защиты пользователя 269
  - конфигурация 275

- управление идентификацией (identity management) (*продолжение*)
  - описание 269
  - отключение 269
  - поддержка единой регистрации 273
  - получение информации о группах профили пользователей 272
  - строка запроса XML 269

## Ф

- файл AdminLinkBarInclude.jsp 352
- файл banner.jspf 358
- файл banner\_searchControl.jspf 358
- файл ccl.properties 26
- файл CCLServer\_date.log 24
- файл config.properties 280, 281
  - клонирование 246
  - настройка 244
  - описания свойств 226
  - редактирование 243
- файл cookies.ini
  - конфигурирование 108
  - описание 107
  - формат 107
- файл crawl.rules 108
- файл crawler\_rdb\_plugin.xml 71
- файл Default.jsp 352, 358
- файл es.cfg 20, 21, 26, 27, 252, 280, 281
- файл es.search.provider.jar 354, 362
- файл es.security.jar 349, 354, 362
- файл es.wp5.install.jar 349
- файл es.wp6.install.jar 354, 362
- файл esapi.jar 349, 354, 362
- файл ESPACServer.ear 349, 354, 362
- файл ESSearchAdapter.ear 349
- файл ESSearchPortlet.war 349, 354, 362
- файл ESSearchRegistrationPortlet.war 349
- файл followindex.rules
  - конфигурирование 111
  - описание 110
- файл global.rules 108
- файл nodes.ini 26, 252
- файл parserdriver.collection.properties 161
- файл parserTypes.cfg 155, 157, 160
- файл runtime-generic.properties 206, 321
- файл SearchBarInclude.jsp 352
- файл siapi.jar 349
- файл stellent.properties 163
- файл stellenttypes.cfg 163
- файл stellentTypes.cfg 155
- файл vbr\_access\_services.jar 49, 50
- файл WpsMigratorLog.log 372
- файлы EAR
  - программа ESAdmin 292
  - программа ESSearchApplication 292
  - программа ESSearchServer 292
- файлы robots.txt
  - идентификация пользовательского агента 92
  - соответствие искателя Web 93
- файлы tar
  - просмотр искателем 122
  - форматы URI 123
- файлы zip
  - просмотр искателем 122
  - форматы URI 123

файлы журнала  
для отслеживания документов 323  
журналы запросов 334  
конфигурация SMTP-сервера 332  
конфигурирование размеров 334  
максимальный размер 330  
мастер по перенастройке 372  
описание 327  
опции электронной почты 332  
переход к новым 330  
положение по умолчанию 327  
просмотр 335  
просмотр отброшенных  
документов 325  
слежение 325, 335  
уровни серьезности 330  
фильтрация 335  
файлы склада ключей 252  
фрагменты XML, собственный поиск  
XML 153

японский  
сегментация энграмм (n-gram  
segmentation) 172  
удаление символов новой строки 173

## Х

хронологические отчеты искателя  
описание 309  
Отчет о кодах состояния HTTP 312  
отчет по сайту 312  
создание 312  
хронология времени ответа,  
слежение 320

## Ц

Центр поиска, WebSphere Portal  
версии 6, интеграция с поисковой  
системой предприятия 357  
описание 347

## Ч

частые запросы  
подсчет 321  
частые запросы, слежение 320  
частые результаты  
столбчатые диаграммы 247

## Э

элементы Dublin Core 144  
элементы XML  
отображение на общую структуру  
анализа 148  
отображение на поля поиска 140  
поиск 139, 148

## Я

языки  
автоматическое определение 170  
двухсимвольные коды 169  
поддерживаемые 169, 170  
поиск 169  
языки с простым текстом 169  
языки со сложным текстом 169



**IBM**



**Java**<sup>™</sup>  
**COMPATIBLE**

GC43-0404-04



Spine information:

OmniFind Enterprise Edition    **Версия 8.5**

**Управление поисковой системой предприятия**

