

IBM DB2 Information Integrator
OmniFind Edition



Remarques sur le produit

Version 8.2 Fix Pack 1

IBM DB2 Information Integrator
OmniFind Edition



Remarques sur le produit

Version 8.2 Fix Pack 1

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à l'Annexe "Remarques".

mars 2005

LE PRESENT DOCUMENT EST LIVRE "EN L'ETAT". IBM DECLINE TOUTE RESPONSABILITE, EXPRESSE OU IMPLICITE, RELATIVE AUX INFORMATIONS QUI Y SONT CONTENUES, Y COMPRIS EN CE QUI CONCERNE LES GARANTIES DE QUALITE MARCHANDE OU D'ADAPTATION A VOS BESOINS. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.can.ibm.com> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
Tour Descartes
92066 Paris-La Défense Cedex 50*

© Copyright IBM France 2005. Tous droits réservés.

© **Copyright International Business Machines Corporation 2004, 2005. All rights reserved.**

Table des matières

1	A propos des remarques sur le produit	v
1	Informations de support technique sur le Web	v
1	Installation de DB2 II OmniFind Edition,	
1	version 8.2.1	1
1	Désinstallation de DB2 II OmniFind	
1	Edition	3
1	Nouveautés de cette version	5
1	Prise en charge de nouveaux systèmes d'exploitation	
1	et types de sources de données	5
1	Définition d'un ID de collection personnalisé	5
1	Exploration des sites Web protégés par mots de passe	6
1	Exploration d'ensembles de cadres dans des pages	
1	Web	8
1	Configuration des jetons de sécurité des balises	
1	META dans les documents Web	9
1	Indexation du texte d'ancrage dans des liens aux	
1	documents interdits	10
1	Substitution des instructions n'appelant aucun index	
1	et aucune suite dans les pages Web	10
1	Configuration d'un espace d'exploration global pour	
1	les moteurs de balayage Web	11
1	Configuration de plusieurs serveurs Lotus Notes	13
1	Page Spécifier le serveur Notes à explorer	13
1	Améliorations des performances du moteur de	
1	balayage VeniceBridge	15
1	Spécification des informations de connexion pour	
1	les systèmes de fichiers Windows	16
1	Le nom externe est maintenant le nom d'affichage	17
1	Configuration et surveillance du programme	
1	d'écoute de données	17
1	Vue Système : page Programme d'écoute de	
1	données	17
1	Vue Système du moniteur : page Programme	
1	d'écoute de données	18
1	Page Détails du programme d'écoute de données	19
1	Arrêt et démarrage du programme d'écoute de	
1	données	20
1	Estimation du nombre de documents dans une	
1	collection	20
1	Le contrôle des ressources système est automatique	
1	avec l'assistant Collection.	21
1	Contrôle de l'activité de l'analyseur syntaxique	21
1	Page Détails de l'analyseur syntaxique	22
1	Mappage des métadonnées HTML pour la recherche	
1	de zones	24
1	Page de mappages de zone de métadonnées	
1	HTML	24
1	Suppression de nouveaux caractères de changement	
1	de ligne provenant d'espaces dans des documents	
1	japonais ou chinois	26
1	Configuration de types de documents pour un	
1	analyseur syntaxique de collection et une session	
1	Stellent	26
1	Configuration de types de documents pour	
1	l'analyseur syntaxique de collection	27
1	Configuration de types de documents pour la	
1	session Stellent	29
1	Améliorations du traitement des requêtes	33
1	Réduction des résultats issus du même site Web	33
1	Mise en évidence étendue dans les résultats de	
1	recherche	34
1	Prise en charge de délai d'attente dans les	
1	applications de recherche	34
	Limitations connues, problèmes et	
	solutions disponibles	37
	Décompression du fichier tar DB2 UDB pour AIX	37
	Segments de mémoire partagés dans AIX	37
	Redémarrage du système après la désinstallation de	
	Windows	38
	Processus URLFetcher orphelin pour Linux et AIX	38
	Emplacement d'index sur le serveur	39
	Limites de taille de segment.	39
	Tables DB2 UDB avec des objets de grande taille	39
	Le statut de l'index ne s'affiche pas	39
	Autorisations ID utilisateur pour Linux et AIX	39
1	Activation de l'accès au document direct pour	
1	FileNet Content Services	40
1	Limitations de l'exploration de tables de	
1	pseudonymes.	40
	Remarques	41
	Marques	43
	Pour prendre contact avec IBM	45
	Informations sur le produit	45
	Commentaires sur la documentation	45

1

1 A propos des remarques sur le produit

1 Les remarques sur le produit présentent les nouveautés de cette version, les
1 limitations connues, les types d'incident, les solutions disponibles et les mises à
1 jour de la documentation relative à IBM DB2 Information Integrator OmniFind
1 Edition, version 8.2.1 (version 8.2 et Fix Pack 1).

1 IBM DB2 Information Integrator OmniFind Edition, version 8.2.1 n'a pas été
1 traduit. Les modifications apportées aux interfaces du produit et les messages sont
1 en anglais uniquement.

1

1 Informations de support technique sur le Web

1 Le produit DB2 II OmniFind Edition a été récemment renommé WebSphere
1 Information Integrator OmniFind Edition. Il est possible que vous rencontriez des
1 références à WebSphere Information Integrator OmniFind Edition dans les pages
1 Web concernant le produit, mais l'interface, la documentation et les informations
1 concernant le dépannage (remarques techniques) se réfèrent toujours à la marque
1 DB2.

1 Vous pouvez obtenir des informations de support technique sur les sites Web
1 suivants :

1 **IBM DB2 Information Integrator OmniFind Edition**

1 www.ibm.com/software/data/integration/db2ii/supportomni.html

1 **IBM WebSphere Information Integrator**

1 <http://www.ibm.com/software/data/integration/db2ii/support.html>

1 **IBM WebSphere Application Server**

1 <http://www.ibm.com/software/webservers/appserv/was/support/>

1 **IBM DB2 Universal Database pour Linux, UNIX et Windows version 8**

1 www.ibm.com/software/data/db2/udb/winos2unix/support

1 Installation de DB2 II OmniFind Edition, version 8.2.1

1 Avec le tableau de bord d'installation de DB2 II OmniFind Edition, vous pouvez
1 installer le produit complet (version 8.2.1) ou installer uniquement la mise à jour
1 de la version 8.2.1 sur votre installation existante.

1 Avant d'installer la mise à jour version 8.2.1, vous devez sauvegarder votre
1 installation existante. Voir *Administering Enterprise Search* pour plus d'informations
1 sur la sauvegarde et la restauration de votre système. Pendant l'installation,
1 spécifiez le chemin vers le répertoire de données existant. Le tableau de bord
1 d'installation utilisera les informations d'installation existantes pour installer la
1 mise à jour.

1 Si vous souhaitez installer la version complète 8.2.1 du produit et que vous avez
1 précédemment installé DB2 II OmniFind Edition version 8.2, vous devez d'abord
1 désinstaller DB2 II OmniFind Edition. Vous ne pouvez pas installer le produit
1 complet sur une installation existante. De même, vous ne pouvez pas sauvegarder
1 votre système existant, installer la version 8.2.1, puis restaurer les paramètres de
1 configuration de la version 8.2 sur un système équipé de la version 8.2.1.

1 Pour installer la version 8.2.1 :

- 1 1. Lancez le tableau de bord de l'installation.

Option	Description
Linux	Connectez-vous en tant que superutilisateur et tapez <code>./install-linux.bin</code>
AIX	Connectez-vous en tant que superutilisateur et tapez <code>./install-aix.bin</code>
Windows	Vous devez disposer des droits d'administrateur. Cliquez deux fois sur le fichier <code>install-win32.exe</code> ou tapez <code>install-win32.exe</code>

- 1 2. Après l'ouverture du tableau de bord d'installation, cliquez sur **Installation du**
1 **produit**.
- 1 3. Pour installer la mise à niveau de la version 8.2.1 :
 - 1 a. Sélectionnez l'option permettant de **mettre à niveau une installation DB2 II**
1 **OmniFind Edition existante** dans la fenêtre de mise à niveau d'installation.
1 Si vous sélectionnez cette option, votre configuration système de la version
1 8.2 existante sera migrée vers un système de version 8.2.1.
 - 1 b. Tapez le chemin dans le répertoire de données existant. Cliquez sur
1 **Suivant**.
 - 1 c. Vérifiez les paramètres dans la fenêtre de récapitulatif. Cliquez sur **Suivant**
1 pour démarrer l'installation.
- 1 4. Pour installer le produit complet de la version 8.2.1 :
 - 1 a. Sélectionnez l'option permettant d'**installer une nouvelle version de DB2 II**
1 **OmniFind Edition**. Si vous avez déjà installé DB2 II OmniFind Edition
1 version 8.2, assurez-vous de la désinstaller avant de sélectionner cette
1 option.

1
1

- b. Suivez les procédures figurant dans le document *Installation Guide for Enterprise Search* pour installer le produit.

1 Désinstallation de DB2 II OmniFind Edition

1 Le programme de désinstallation supprimera les composants de recherche
1 d'entreprise. Vous pouvez également supprimer la configuration de toutes les
1 données système.

1 Vous devez sauvegarder votre système avant de désinstaller DB2 II OmniFind
1 Edition.

1 Pour une configuration à multiples serveurs, exécutez le programme de
1 désinstallation sur chaque serveur. Pour désinstaller DB2 II OmniFind Edition :

- 1 1. Pour Linux et AIX, connectez-vous en tant que superutilisateur.
- 1 2. Allez dans le répertoire _uninst dans le répertoire d'installation
1 (ES_INSTALL_ROOT), puis démarrez le programme de désinstallation :

1

Option	Description
Linux ou AIX	Tapez ./uninstall.bin
Windows	Cliquez deux fois sur le fichier ou tapez uninstall.exe

1

1

1

1

1

1

1

1

1

1

3. Facultatif : Pour supprimer toutes les informations de configuration
d'administrateur de recherche d'entreprise, cochez la case permettant de
supprimer toutes les données et tous les fichiers de configuration.

Avertissement : Si vous cochez cette case, vous supprimerez toutes les
données de toutes les collections du système.

Sous Windows, vous devez redémarrer l'ordinateur avant de réinstaller DB2 II
OmniFind Edition.

1

1 Nouveautés de cette version

1 Les nouvelles fonctions et les nouveaux dispositifs de DB2 II OmniFind Edition
1 incluent des améliorations apportées à l'administration de la recherche, à la qualité
1 de la recherche et à la performance, et la prise en charge de Microsoft Windows
1 2003.

1 Prise en charge de nouveaux systèmes d'exploitation et types de 1 sources de données

1 DB2 II OmniFind Edition prend en charge le système d'exploitation Windows
1 Server 2003 32 bits (Standard Server, Advanced Server et Datacenter Server) et peut
1 explorer les bases de données Microsoft SQL Server.

1 Pour installer Windows Server 2003, suivez les mêmes instructions que pour
1 Windows Server 2000 figurant dans le document *Installation Guide for Enterprise
1 Search*.

1 Pour explorer les bases de données SQL Server, utilisez la console d'administration
1 de recherche d'entreprise pour créer un moteur de balayage DB2 et configurer les
1 sources de données que vous souhaitez explorer.

1 Définition d'un ID de collection personnalisé

1 Lorsque vous créez une collection, vous pouvez choisir d'utiliser l'identificateur
1 interne que le système crée pour vous, ou définir un ID de collection personnalisé.

1 Si vous indiquez un identificateur de collection personnalisé, vos applications de
1 recherche peuvent utiliser les interfaces de programmation d'application de
1 recherche d'entreprise pour appeler la collection avec cet identificateur au lieu de
1 l'identificateur potentiellement chiffré que le système crée.

1 Dans ce cas, vous devez effectuer cette opération au moment de créer la collection.
1 Il est impossible d'éditer ultérieurement la collection et de changer cet
1 identificateur.

1 Pour spécifier le type d'identificateur de collection à utiliser, procédez comme suit :

- 1 1. Lancez la console d'administration de recherche d'entreprise. Si la sécurité est
1 activée, connectez-vous en tant qu'utilisateur disposant de droits
1 d'administrateur de recherche d'entreprise.
- 1 2. Cliquez sur **Créer une collection** ou sur **Assistant de collection**, selon que vous
1 souhaitez ou non utiliser l'assistant pour créer une collection.
- 1 3. Sélectionnez l'une des options suivantes dans la zone **ID de collection** :

1 **ID de collection par défaut**

1 Choisissez cette option afin de permettre au système de créer un
1 identificateur interne pour la collection.

1 **ID de collection personnalisé**

1 Choisissez cette option si vous souhaitez spécifier votre propre
1 identificateur pour la collection. La valeur indiquée peut contenir
1 uniquement des caractères alphanumériques et le trait de soulignement.

Exploration des sites Web protégés par mots de passe

Les moteurs de balayage Web DB2 II OmniFind Edition peuvent explorer les sites Web protégés par mots de passe au moyen de l'authentification standard HTTP et en conservant les informations dans les cookies.

Prise en charge d'authentification standard HTTP

En tant qu'administrateur de recherche d'entreprise, vous pouvez identifier les sites Web protégés par mots de passe que vous souhaitez explorer et spécifier les données d'identification dans un fichier de configuration appelé `httpauth.ini` avant de démarrer l'exploration.

Le fichier `httpauth.ini` doit suivre les règles de formatage suivantes :

- Le fichier `httpauth.ini` est constitué de texte, il ne peut pas comporter de caractères non imprimables, espaces mis à part. Les entrées de données sont séparées par de nouvelles lignes.
- Le fichier peut contenir des entrées d'autorisation, des commentaires et des lignes vierges dans n'importe quel ordre.
- Un commentaire est une ligne dont le premier caractère autre qu'un espace est #.
- Les commentaires ne peuvent pas apparaître sur une ligne avec du texte hors commentaire.
- Chaque entrée d'autorisation doit comporter les informations suivantes, dans l'ordre suivant, sur une ligne unique :
 - une adresse URL cible (la racine du système de fichiers serveur pour lequel une autorisation est nécessaire),
 - un nom de domaine significatif pour le serveur,
 - un nom d'utilisateur pour la connexion,
 - un mot de passe pour la connexion,
 - une adresse URL de réauthentification,
 - une liste de jetons de sécurité séparés par des points-virgules, qui peut être vide, mais la zone doit toujours être présente, avec une virgule en tant que séparateur après la cinquième zone.
- Les zones d'une ligne sont séparées par des virgules. Des espaces sont autorisés autour des virgules.
- Le mot de passe de chaque entrée est chiffré avec la méthode de chiffrement utilisée dans le reste de DB2 II OmniFind Edition.
- L'autorisation pour une adresse URL est supposée être valide pour toutes les URL situées au-dessous de l'URL spécifiée dans le système de fichiers du serveur.

Vous pouvez modifier le fichier `httpauth.ini` en utilisant un utilitaire de ligne de commande appelé `httpauthedit`. À l'aide de cet utilitaire, vous pouvez taper des mots de passe dans le texte (déchiffrés) lorsque vous créez ou modifiez une entrée, et l'utilitaire chiffre les mots de passe avant de les écrire sur le disque. Pour changer un mot de passe, vous devez supprimer une entrée puis la rajouter.

Pour démarrer l'utilitaire `httpauthedit`, tapez la commande suivante sur une ligne :

```
java -classpath ES_INSTALL_ROOT/lib/URLFetcher.jar:esinstall/lib/es.oss.jar  
com.ibm.es.wc.uf.HTTPAuthEdit configdir
```

1 *ES_INSTALL_ROOT* est le répertoire d'installation de DB2 II OmniFind Edition et
1 *configdir* est le répertoire de configuration maître de l'instance du moteur de
1 balayage où le fichier *httpauth.ini* doit être créé (ou modifié).

1 Le répertoire de configuration maître est dans
1 *ES_NODE_ROOT/master_config/CR_INST_DIR*, où *ES_NODE_ROOT* représente
1 la valeur de *NodeDirectory* dans le fichier *ES_CFG*, et *CR_INST_DIR* représente le
1 répertoire créé pour une nouvelle instance du moteur de balayage Web (à qui le
1 logiciel d'administration donne un nom au hasard). Assurez-vous que vous
1 travaillez sur la bonne configuration d'instance. Les modifications n'affecteront pas
1 le moteur de balayage jusqu'au prochain démarrage.

1 L'utilitaire *httpauthedit* accepte les commandes suivantes :

1 *Tableau 1. Commandes acceptées par l'utilitaire HTTPAUTHEDIT*

Commande	Description
?	Aide : Décrit l'ensemble des commandes.
b	Parcourir : Imprime le contenu du fichier.
a	Ajouter : Ajoute une nouvelle entrée d'autorisation au fichier. Cette commande nécessite cinq ou six arguments séparés par des virgules sur une ligne. Par exemple : <code>http://www.ibm.com/solutions/, solutions, wwwsearch, foobar, xyz123, http://www.ibm.com/solutions/, group1;group2;admin3</code> <ul style="list-style-type: none">• Adresse URL cible : <code>http://www.ibm.com/solutions/</code>• Nom du domaine : <code>solutions</code>• Nom d'utilisateur : <code>wwwsearch, foobar</code>• Mot de passe : <code>xyz123</code>• Adresse URL de réauthentification : <code>http://www.ibm.com/solutions/</code>• Facultatif : Jetons de sécurité : <code>group1;group2;admin3</code>
d	Supprimer : Supprime une entrée d'autorisation du fichier. Cette commande nécessite un argument : l'adresse URL cible de l'entrée d'authentification supprimée du fichier, comme <code>http://www.ibm.com/solutions/</code> .
q	Quitter : Quitte le programme.

1 L'adresse URL de réauthentification peut être identique à l'adresse URL cible :

- L'adresse URL de réauthentification doit être une adresse URL qui, en cas de requête sans autorisation, entraîne une réponse 401 du serveur, avec un protocole HTTP, pour le domaine pour lequel le client souhaite soumettre l'autorisation.
- L'adresse URL de réauthentification doit entraîner cette réponse sans redirection (l'adresse URL de réauthentification doit représenter la véritable page de destination, et non une page 301, 302 ou autre qui redirige le client vers une autre page). Cette adresse URL de réauthentification est parfois différente de l'URL racine du domaine l'autorisant. Le fait que l'adresse URL de réauthentification soit identique à l'adresse URL racine dépend de la logique interne du site Web. L'administrateur du site Web peut fournir des détails.

1 Les espaces de début et fin de ligne seront supprimés des valeurs de zones. Si
1 vous tapez une saisie incorrecte, *httpauthedit* affiche un message d'erreur.

1 Les modifications sont enregistrées lorsque vous tapez la commande de sortie. Il
1 n'existe pas de commande d'enregistrement explicite. Pour annuler une

1 modification, vous pouvez annuler le programme en le quittant sans utiliser la
1 commande de sortie.

1 Cookies

1 Le moteur de balayage Web conservera les cookies reçus des serveurs et les
1 utilisera pour la vie de l'instance du moteur de balayage. Lorsque le moteur de
1 balayage s'interrompt, il enregistre sur le disque tous les cookies qui n'ont pas
1 expiré et les recharge au début de la session d'exploration suivante. Le fichier
1 cookie est appelé cookies.ini, il est situé dans le répertoire de données du moteur
1 de balayage.

1 Le cookie du moteur de balayage est indépendant de l'authentification HTTP
1 (fichier httpauth.ini), mais si les cookies sont utilisés par les serveurs Web pour
1 protéger le contenu, l'authentification peut également être nécessaire, même avec
1 des cookies fournis manuellement. Les administrateurs de site Web peuvent fournir
1 des informations concernant la configuration requise pour explorer leurs sites.

1 En tant qu'administrateur de recherche d'entreprise, vous pouvez ajouter des
1 cookies à ceux qui ont été trouvés par le moteur de balayage, en utilisant un
1 éditeur de texte pour les ajouter au fichier cookies.ini. Ce fichier est réécrit par le
1 moteur de balayage chaque fois qu'il s'arrête : vous devez donc conserver les
1 cookies entrés manuellement dans un autre fichier et les ajouter si nécessaire avant
1 de démarrer une exploration. (Le moteur de balayage n'élimine pas les cookies non
1 expirés, mais en cas d'incident lors de l'écriture de l'ensemble des cookies, il ne
1 faut pas perdre les cookies spécifiés manuellement.)

1 Pour spécifier les cookies pour un moteur de balayage Web, vous devez modifier
1 le fichier cookies.ini. Dans le fichier cookies.ini, chaque ligne est une entrée. Les
1 lignes vides et les commentaires sont autorisés, mais ne seront pas conservés.
1 Chaque entrée a le format suivant. Chaque entrée doit être sur une seule ligne :
1 Cookie(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSIDNFLSNL;path=/,
1 https://www.ibm.com:443/help/solutions/

1 *Cookie* est un mot clé obligatoire, indiquant le début d'une entrée. Le nombre 53
1 représente la longueur en caractères du nom du cookie et la paire de valeurs, ainsi
1 que tout autre contenu (chemin, paramètre de sécurité, etc) à envoyer avec le
1 cookie. Le nombre 40 représente la longueur de l'adresse URL associée, pour
1 validation.

1 Pour ASPSESSION. . . , ces 53 caractères représentent le contenu du cookie à
1 envoyer au serveur d'origine. Cette chaîne est suivie par une virgule (,) qui fait
1 office de séparateur.

1 L'adresse URL https://www.ibm.com:443/help/solutions/ est utilisée pour valider
1 le cookie en fournissant un nom de domaine par exemple. Avec des cookies ajoutés
1 manuellement, vous devez inclure une adresse URL de validation répondant aux
1 critères de restrictions de sécurité et de confidentialité des cookies.

1 Exploration d'ensembles de cadres dans des pages Web

1 Le moteur de balayage Web peut maintenant explorer des pages Web créées avec
1 des cadres HTML.

Dans DB2 II OmniFind Edition version 8.2, le moteur de balayage Web ne pouvait pas analyser les ensembles de cadres, ni extraire des liens à partir des cadres dans d'autres pages HTML.

Configuration des jetons de sécurité des balises META dans les documents Web

Les nouvelles options de sécurité du moteur de balayage Web vous permettent de spécifier les noms des balises META dans un fichier de configuration. Le contenu de ces balises est utilisé en tant que jetons de sécurité pour contrôler l'accès aux documents dans lesquels se trouvent les balises.

Dans DB2 II OmniFind Edition, version 8.2, les jetons de sécurité ont été définis de manière globale, ils s'appliquent à tous les documents ou ne s'appliquent à aucun d'entre eux, selon que le niveau de sécurité au niveau du document a été activé ou non pour le moteur de balayage.

Vous pouvez dorénavant contrôler l'accès aux documents en spécifiant les balises META en tant que jetons de sécurité. Pour utiliser la sécurité basée sur les balises META, vous devez activer le niveau de sécurité pour le moteur de balayage Web, puis vous devez modifier un fichier de configuration pour spécifier les noms de balises META (vous ne pouvez pas spécifier de noms de balises META dans la console d'administration).

Pour spécifier les balises META que vous souhaitez utiliser en tant que jetons de sécurité :

1. Connectez-vous au serveur du moteur de balayage en tant qu'administrateur de recherche d'entreprise (ces ID utilisateur et mot de passe ont été configurés pendant l'installation de DB2 II OmniFind Edition).
2. Utilisez un éditeur de texte pour modifier le fichier `crawl.properties` dans l'arborescence `master_config`.
3. Si nécessaire, insérez le paramètre suivant : `security_meta_tags`. (Si le paramètre `security_meta_tags` existe déjà, la valeur par défaut est nil.)
4. Spécifiez les noms de balises META dans une liste séparée par des virgules. Par exemple :

```
security_meta_tags=security,acl,access
```

Règles pour les noms de balises :

- Les noms de balises lus depuis le fichier de configuration sont convertis en minuscules.
 - Les noms de balises affichés dans les documents explorés sont convertis en minuscules avant la comparaison avec les noms du fichier de configuration.
 - Le seul caractère de séparation dans la liste des noms est la virgule (les espaces ne délimitent pas un nom).
 - Les espaces imbriqués dans les noms sont conservés ; les espaces de début et de fin sont supprimés.
5. Enregistrez et fermez le fichier.

Si les noms de balises sont configurés pour le paramètre `security_meta_tags`, pendant le traitement de chaque page téléchargée, les balises META du document seront vérifiées par rapport à la liste `security_meta_tags`.

Pour plus d'informations sur l'utilisation de jetons de sécurité pour contrôler l'accès aux documents, voir la rubrique *Document-level security* (sécurité au niveau du document) dans *Administering Enterprise Search*.

Indexation du texte d'ancrage dans des liens aux documents interdits

Si un document inclut des liens à des documents que le moteur de balayage du Web n'est pas autorisé à explorer, vous pouvez choisir de conserver le texte d'ancrage pour ces liens dans l'index.

Les paramètres que contient un fichier robots.txt ou les spécifications de balise méta peuvent empêcher le moteur de balayage sur le Web d'accéder à certains documents sur un site Web. Si un document que le moteur de balayage sur le Web est autorisé à explorer inclut des liens à des documents interdits, vous pouvez choisir le mode de gestion du texte d'ancrage pour ces liens sur la page Propriétés avancées du moteur de balayage Web :

1. Pour inclure le texte d'ancrage pour les liens aux documents interdits dans l'index, cochez la case permettant d'**indexer le texte d'ancrage dans des liens aux documents interdits**. (Le texte d'ancrage est le texte descriptif qui apparaît entre les codes <A> et des documents Web.) Si vous cochez cette case, des pages interdites peuvent être trouvées en recherchant du texte compris dans du texte d'ancrage pour des liens qui pointent vers elles.
2. Pour exclure le texte d'ancrage associé aux liens interdits de l'index (tout en excluant les documents interdits), désélectionnez la case permettant d'**indexer le text d'ancrage pour les liens aux documents interdits**.

Vous pouvez spécifier les propriétés avancées du moteur de balayage Web lors de la création d'un moteur de balayage ou en éditant l'espace d'exploration d'un moteur existant.

Substitution des instructions n'appelant aucun index et aucune suite dans les pages Web

Vous pouvez spécifier des règles dans un fichier de configuration pour décider si le moteur de balayage Web suit les liens aux pages ou indexe les pages contenant des instructions n'appelant aucun index et aucune suite. Cette fonctionnalité n'est pas prise en charge dans la console d'administration de recherche d'entreprise.

Certaines pages Web comportent des instructions n'appelant aucun index et aucune suite, qui chargent les robots (tels que le moteur de balayage Web) de ne pas inclure le contenu de ces pages dans un index, de ne pas suivre les liens de ces pages ou de ne faire aucune de ces opérations.

Dans DB2 II OmniFind Edition version 8.2, ces restrictions sont décidées par l'administrateur du site Web ou le créateur du contenu. Le contrôle de ces paramètres, cependant, peut améliorer la qualité de l'exploration. Par exemple, il existe des ensembles de pages de répertoires contenant des milliers de liens mais aucun autre contenu utile : ces pages peuvent être explorées, leurs liens peuvent être suivis, mais les pages en question ne doivent pas être indexées.

Parfois, l'administrateur du moteur de balayage ne souhaite pas continuer son exploration plus bas dans une hiérarchie, mais les pages de feuilles souhaitées contiennent des liens et aucune instruction n'appelant aucune suite. Il est parfois impossible de demander aux propriétaires du contenu de ces pages d'insérer

l'instruction nécessaire car certaines de ces pages ont été générées automatiquement et n'ont pas de propriétaire.

Le moteur de balayage de la version 8.2.1 recherche le fichier de configuration, appelé `followindex.rules`, dans son répertoire de configuration. Ce fichier peut contenir n'importe quel nombre de règles concernant les suites et les index :

```
forbid follow URLprefix
allow follow URLprefix
forbid index URLprefix
allow index URLprefix
```

Il est impossible de spécifier les adresses IP et les noms d'hôtes DNS des règles de suite et d'index : vous pouvez spécifier les préfixes URL (y compris les caractères génériques) uniquement. L'ordre est important, comme lors de la spécification de règles d'exploration. Si une règle autorisant ou interdisant explicitement la suite ou l'indexation est définie, elle annule les autres paramètres, y compris ceux du document cible.

Configuration d'un espace d'exploration global pour les moteurs de balayage Web

Vous pouvez spécifier des règles dans un fichier de configuration pour créer un espace d'exploration global pour les moteurs de balayage Web et mieux contrôler la suppression d'adresses URL de l'index. Cette fonctionnalité n'est pas prise en charge dans la console d'administration de recherche d'entreprise.

Chaque moteur de balayage est configuré avec un espace d'exploration définissant les adresses URL qui doivent ou non être explorées. Les adresses URL découvertes dans l'espace d'exploration sont conservées (dans une base de données) pour une exploration ultérieure ; les adresses qui ne sont pas dans l'espace d'exploration sont éliminées. Si le moteur de balayage démarre avec une base de données vide, la définition d'espace d'exploration et la base de données restent cohérentes pendant l'exécution du moteur de balayage.

Parfois un moteur de balayage s'arrête, et son espace d'exploration est réduit (en ajoutant de nouvelles règles interdisant des pages). Lorsque le moteur de balayage redémarre, sa définition d'espace d'exploration et la base de données seront incohérentes : la base de données comportera des adresses URL - certaines explorées, d'autres non - qui ne sont pas dans le nouvel espace d'exploration dont la taille est plus petite. Dans DB2 II OmniFind Edition version 8.2, le moteur de balayage Web restaure la cohérence en modifiant les codes de retour HTTP pour ces adresses URL à 760 (exclusion d'espace d'exploration) et en écrivant des instructions pour supprimer les pages dorénavant exclues de l'index. Cette opération a lieu parce qu'il pense être le seul moteur de balayage Web de la collection, et que son espace d'exploration identifie tous les documents Web autorisés dans l'index.

Parfois, lorsque l'exécution d'un moteur de balayage sur un espace d'exploration volumineux prend un certain temps, l'administrateur décide de répartir l'espace d'exploration entre deux moteurs de balayage ou plus (pour s'assurer peut-être que certaines parties sont explorées plus souvent). De nouveaux moteurs de balayage sont créés, chacun d'entre eux disposant de ses propres tables de bases de données indépendantes (vides à l'origine) et d'une partie distincte de l'espace d'exploration original. L'espace d'exploration du moteur de balayage original est alors réduit à ce qu'il reste après la suppression des parties qui vont être explorées par les nouveaux moteurs de balayage.

1 Si le moteur de balayage original est un moteur de balayage de la version 8.2, il
1 rétablit la cohérence de sa base de données de la manière décrite précédemment.
1 Or, l'administrateur ne souhaite pas cette opération : les adresses URL déplacées ne
1 doivent pas être supprimées de l'index ; elles sont explorées par d'autres moteurs
1 de balayage.

1 DB2 II OmniFind Edition version 8.2.1 fournit un second espace d'exploration, de
1 niveau supérieur, représentant les adresses URL qui ne doivent être ni explorées
1 par le moteur de balayage en cours, ni supprimées. (Les adresses URL qui ne sont
1 dans aucun espace sont toujours rejetées pendant l'exploration, et elles seront
1 supprimées de l'index lors du balayage, comme auparavant.)

1 Le second espace d'exploration global est défini par le contenu d'un fichier appelé
1 global.rules, situé dans le répertoire de configuration du moteur de balayage. S'il
1 existe, le fichier est lu pendant l'initialisation du moteur de balayage. S'il n'existe
1 aucun fichier global.rules, le moteur de balayage opère avec un espace
1 d'exploration à un seul niveau (comme dans la version 8.2). Si un espace global est
1 défini, le moteur de balayage gère les adresses URL comme auparavant, mais il
1 envoie des instructions pour supprimer une adresse URL de l'index uniquement si
1 l'adresse URL n'est pas dans l'espace.

1 Le fichier global.rules a la même syntaxe que le fichier local crawl.rules, mais
1 seules les règles de nom de domaine sont significatives. (Un espace d'exploration
1 peut ainsi être partitionné entre les moteurs de balayage uniquement en utilisant
1 les noms d'hôtes DNS, et non les adresses IP ni les masques de préfixes d'URL.)

1 La présence d'un fichier global.rules active la fonctionnalité d'espace d'exploration
1 global (et son absence la désactive). Les adresses URL exclues par un préfixe URL
1 ou une adresse IP dans l'espace d'exploration local (tel que défini dans le fichier
1 crawl.rules) ne sont pas affectées par l'espace d'exploration global ; elles sont
1 toujours exclues. L'espace d'exploration global peut être utilisé uniquement pour
1 prévenir la suppression de l'index des adresses URL exclues d'un espace
1 d'exploration de moteur de balayage par une règle de domaine local.

1 Voici comment fonctionnent les règles :

- 1 1. Si l'adresse URL d'une base de données de moteur de balayage est exclue par
1 une règle de préfixe local ou une règle d'adresse, un code de retour 760 lui est
1 affecté, et elle est supprimée de l'index. C'est alors que la gestion de l'espace
1 d'exploration prend fin. L'adresse URL ne sera plus explorée.
- 1 2. Après l'étape 1 : Si une adresse URL de la base de données du moteur de
1 balayage est exclue par une règle de domaine local, et qu'il n'existe pas
1 d'espace d'exploration global, un code de retour 760 lui est affecté et elle est
1 supprimée de l'index. C'est la fin de la gestion de l'espace d'exploration.
1 L'adresse URL ne sera plus explorée.
- 1 3. Après l'étape 1 : Si une adresse URL de la base de données du moteur de
1 balayage est exclue par une règle de domaine local, mais autorisée
1 explicitement par une règle de l'espace d'exploration global, un code de retour
1 761 lui est affecté. Le moteur de balayage ne l'explorera plus, mais elle n'est
1 pas supprimée de l'index (elle est supposée être dans l'espace d'exploration
1 local d'un autre moteur de balayage).
- 1 4. Après l'étape 1 : Si une adresse URL de la base de données du moteur de
1 balayage est exclue par une règle de domaine local, et n'est pas autorisée
1 explicitement par une règle de l'espace d'exploration global, un code de retour
1 760 lui est affecté, elle est supprimée de l'index. C'est la fin de la gestion de
1 l'espace d'exploration.

1 Remarque importante : L'espace d'exploration global étant uniquement consulté
1 pour empêcher la suppression des adresses URL ayant déjà été exclues par
1 l'espace d'exploration local, le résultat par défaut de l'espace d'exploration
1 global peut donc être une interdiction, si aucune règle ne s'applique à l'adresse
1 URL candidate. Il s'agit de la valeur inverse de la valeur par défaut pour
1 l'espace d'exploration local.

1 Remarquez que le fichier global.rules doit être présent dans le répertoire
1 master_config de chaque moteur de balayage partageant un espace d'exploration
1 global, et que toutes les copies du fichier global, ainsi que les fichiers crawl.rules
1 locaux individuels doivent être modifiés avec précaution pour que leur cohérence
1 soit garantie.

1 Configuration de plusieurs serveurs Lotus Notes

1 Vous pouvez ajouter des bases de données à partir de plusieurs serveurs Lotus
1 Notes dans le même espace d'exploration sans configurer de moteurs de balayage
1 supplémentaires. Dans DB2 II OmniFind Edition version 8.2, chaque moteur de
1 balayage Notes était limité à l'exploration d'un seul serveur Lotus Notes.

1 Après la configuration d'un moteur de balayage Notes pour explorer les bases de
1 données sur un serveur Lotus Notes, effectuez les opérations suivantes pour
1 ajouter des bases de données depuis d'autres serveurs Lotus Notes dans le même
1 espace d'exploration :

- 1 1. Dans la page d'exploration, sélectionnez le moteur de balayage Notes que vous
1 souhaitez configurer, puis cliquez sur l'icône **Espace d'exploration**.
- 1 2. Dans la page Espace d'exploration Notes, cliquez sur **Ajouter base de données**.
- 1 3. Utilisez la fenêtre suivante pour spécifier les informations concernant le serveur
1 Lotus Notes que vous souhaitez explorer : «Page Spécifier le serveur Notes à
1 explorer»

1 Si vous avez besoin d'aide concernant les pages suivantes affichées par l'assistant
1 du moteur de balayage, cliquez sur **Aide**.

1 **Conseil :** Pour changer le mot de passe utilisé par le moteur de balayage pour
1 accéder à un serveur Lotus Notes, sélectionnez le serveur dans la page
1 Espace d'exploration Notes, puis cliquez sur **Editer le serveur**. (Quand le
1 moteur de balayage Notes était limité à l'exploration d'un seul serveur
1 Lotus Notes, cette opération était disponible à partir du haut de la page
1 Espace d'exploration Notes.)

1 Page Spécifier le serveur Notes à explorer

1 La page Spécifier le serveur Notes à explorer permet d'identifier un serveur Lotus
1 Notes que vous souhaitez configurer pour un espace d'exploration existant. Si vous
1 ajoutez un serveur à l'espace d'exploration, indiquez les informations permettant
1 au moteur de balayage d'accéder aux bases de données de ce serveur.

1 Pour plus d'informations sur la configuration de votre environnement Notes, de
1 sorte qu'il puisse être exploré par le moteur de balayage Notes, voir *Administering*
1 *Enterprise Search*.

1 **Important :** Une fois que vous avez cliqué sur **Suivant**, le système valide sa
1 capacité à accéder au serveur Notes que vous spécifiez. S'il est

1 incapable d'accéder au serveur, un message d'erreur s'affiche. Vérifiez
1 et corrigez le nom du serveur et les informations de protocole et
1 cliquez à nouveau sur **Suivant**.

1 **Zones et commandes**

1 **Serveur existant**

1 Indique que vous souhaitez ajouter des bases de données d'un serveur
1 Lotus Notes qui existe déjà dans l'espace d'exploration.

1 Si vous choisissez cette option, sélectionnez le serveur Lotus Notes dont
1 vous souhaitez ajouter les bases de données.

1 **Nouveau serveur**

1 Indique que vous souhaitez ajouter des informations relatives à un
1 nouveau serveur Lotus Notes dans l'espace d'exploration, et sélectionner
1 des bases de données de ce serveur à explorer.

1 Si vous choisissez cette option, spécifiez les options qui permettent au
1 moteur de balayage Notes d'accéder aux bases de données sur le serveur.

1 **Nom du serveur Lotus Notes**

1 Entrez le nom de système hôte complet du serveur Domino qui
1 héberge les bases de données Notes (fichiers .nsf) que ce moteur de
1 balayage doit explorer en tant que server1.ibm.com.

1 Le port par défaut pour les serveurs Notes qui utilisent le
1 protocole NRPC est 1352. Le port par défaut pour les serveurs
1 Notes qui utilisent le protocole DIIOP est 80. Si le serveur n'est pas
1 en mode écoute sur le port par défaut, spécifiez le numéro de port
1 (par exemple, server1.ibm.com:1353 ou server1.ibm.com:81).

1 **Protocole**

1 Spécifiez des informations sur le protocole de transmission utilisé
1 par le serveur :

1 **NRPC (Notes Remote Procedure Call)**

1 Sélectionnez cette option si le serveur Domino utilise le
1 protocole NRPC (Notes remote procedure call protocol)
1 pour échanger des informations entre le serveur et les
1 navigateurs Web.

1 **Important :** Si vous avez préalablement configuré des
1 moteurs de balayage Notes pour cette
1 collection qui utilisent le protocole NRPC,
1 arrêtez-les avant de configurer les informations
1 sur ce moteur de balayage Notes. Les
1 procédures de reconnaissance de recherche
1 d'entreprise qui utilisent le protocole NRPC, et
1 les processus du moteur de balayage qui
1 utilisent NRPC, ne peuvent pas s'exécuter
1 simultanément (les procédures de
1 reconnaissance fournissent au moteur de
1 balayage des informations sur les sources
1 disponibles pour l'exploration).

1 Si vous sélectionnez l'option NRPC, spécifiez les
1 informations suivantes :

1 **Fichier d'ID utilisateur Lotus Notes**

1 Entrez le chemin d'accès absolu à un fichier d'ID

1 utilisateur Notes ayant les droits d'accès aux bases
1 de données de ce serveur. Par exemple, sur une
1 machine UNIX, vous pouvez taper
1 /home/username/username.id.

1 Chaque utilisateur de Notes possède un fichier
1 d'identification qui l'identifie de manière unique.
1 Assurez-vous qu'un exemplaire du fichier d'ID que
1 vous spécifiez ici existe sur le moteur de balayage.
1 L'ID de l'administrateur de recherche d'entreprise
1 doit avoir accès en lecture au fichier d'ID
1 utilisateur Notes et au répertoire où le fichier est
1 stocké. (L'ID utilisateur de l'administrateur de
1 recherche d'entreprise est spécifié lors de
1 l'installation de DB2 II OmniFind Edition.)

1 **Mot de passe**

1 Entrez le mot de passe du fichier d'ID utilisateur
1 Notes spécifié.

1 **DIIOP (Domino Internet Inter-ORB Protocol)**

1 Sélectionnez cette option si le serveur Domino utilise le
1 protocole DIIOP (Domino Internet inter-ORB protocol)
1 pour échanger des informations entre le serveur et les
1 navigateurs Web. Pour utiliser ce protocole, vous devez
1 activer les tâches HTTP et DIIOP sur le serveur Domino.

1 Si vous sélectionnez cette option, spécifiez les informations
1 ci-dessous :

1 **ID utilisateur Lotus Notes**

1 Entrez un ID utilisateur Notes ayant l'autorisation
1 d'accéder aux bases de données de ce serveur, par
1 exemple user1/Hometown/IBM.

1 **Mot de passe**

1 Entrez le mot de passe Notes Internet qui est
1 spécifié pour cet ID utilisateur dans le répertoire
1 Domino. (Si DIIOP est configuré pour un serveur
1 Notes, une des étapes consiste à créer des mots de
1 passe Internet pour les utilisateurs autorisés à
1 accéder au serveur.)

1 **Améliorations des performances du moteur de balayage VeniceBridge**

1 Ventica VeniceBridge est maintenant un produit IBM appelé DB2 Information
1 Integrator Content Edition (DB2 II Content Edition). L'interface du produit et la
1 documentation de DB2 II OmniFind Edition se réfèrent toujours à VeniceBridge,
1 ces remarques sur le produit utilisent donc également le terme VeniceBridge pour
1 présenter les améliorations apportées à l'interface utilisateur.

1 Pour améliorer la performance d'une nouvelle exploration, vous pouvez
1 maintenant sélectionner une correspondance de données lorsque vous configurez
1 des règles pour des classes d'articles VeniceBridge individuelles. Une
1 correspondance de données est une construction DB2 II Content Edition qui fournit
1 les informations de moteur de balayage VeniceBridge à propos de la classe
1 d'article, telle que la date de création, la date de révision, la clé de tri, et la taille
1 du contenu. En associant une correspondance de données à une classe d'article

1 dans la configuration du moteur de balayage, vous pouvez vérifier que le moteur
1 de balayage peut rapidement effectuer une deuxième exploration du contenu dans
1 cette classe d'article.

1 Pour Documentum et FileNet Content Services, vous pouvez tirer parti de cette
1 fonctionnalité sans spécifier de correspondance de données, car le moteur de
1 balayage VeniceBridge dispose de la correspondance de données par défaut.

1 **Important :**

1 Pour profiter de ces améliorations des performances, vous devez exécuter DB2 II
1 Content Edition Fix Pack 1.

1 Pour associer une correspondance de données à une classe d'article dans un espace
1 d'exploration VeniceBridge existant :

- 1 1. Modifiez la collection à laquelle le moteur de balayage VeniceBridge appartient.
- 1 2. Cliquez sur la page d'exploration.
- 1 3. Localisez le nom du moteur de balayage VeniceBridge, puis cliquez sur le
1 bouton **Espace d'exploration**.
- 1 4. Dans la page Espace d'exploration VeniceBridge, localisez la classe d'article que
1 vous souhaitez configurer, puis cliquez sur **Edition**.
- 1 5. Dans la zone **de correspondance de données** de la page Configurer des règles
1 pour une classe d'articles VeniceBridge, sélectionnez le nom de la
1 correspondance de données que vous souhaitez utiliser avec la classe d'article.
- 1 6. Cliquez sur **OK**.

1 Si vous créez un moteur de balayage VeniceBridge ou ajoutez une classe d'article à
1 un espace d'exploration existant, vous sélectionnez le nom de correspondance de
1 données lors de la spécification des autres options de la page Configurer des règles
1 pour une classe d'articles VeniceBridge.

1 **Spécification des informations de connexion pour les systèmes de 1 fichiers Windows**

1 Lorsque vous configurez des règles d'exploration pour un sous-répertoire
1 Windows, vous pouvez spécifier un ID utilisateur et un mot de passe permettant
1 au moteur de balayage d'accéder aux données dans le sous-répertoire.

1 Pour configurer les données d'identification de connexion pour un sous-répertoire
1 dans un espace d'exploration de système de fichiers Windows existant :

- 1 1. Modifiez la collection à laquelle le moteur de balayage de système de fichiers
1 Windows appartient.
- 1 2. Cliquez sur la page d'exploration.
- 1 3. Localisez le nom du moteur de balayage du système de fichiers Windows, puis
1 cliquez sur le bouton **Espace d'exploration**.
- 1 4. Dans la page d'espace d'exploration Windows, localisez le sous-répertoire que
1 vous souhaitez configurer, puis cliquez sur **Edition**.
- 1 5. Dans la page Configurer des règles pour un sous-répertoire Windows, tapez un
1 ID utilisateur et un mot de passe valides pour ce sous-répertoire dans les zones
1 **ID utilisateur** et **Mot de passe**.
- 1 6. Cliquez sur **OK**.

1 Si vous créez un moteur de balayage de système de fichiers Windows ou ajoutez
1 un sous-répertoire à un espace d'exploration existant, vous spécifiez les données
1 d'identification de connexion lors de la spécification des autres options de la page
1 Configurer des règles pour un sous-répertoire Windows.

1 Le nom externe est maintenant le nom d'affichage

1 Dans la console d'administration de recherche d'entreprise, l'intitulé **Nom externe**
1 a été modifié en **Nom d'affichage** dans les assistants de configuration du moteur
1 de balayage pour insister sur le fait que les valeurs spécifiées dans cette zone le
1 sont à des fins d'affichage uniquement.

1 Lorsque vous affectez un nom d'affichage, vous ne modifiez pas le nom interne de
1 l'objet. Par exemple, vous pouvez affecter à une zone de base de données appelée
1 lastUpdateDateTime le nom d'affichage plus explicite Last Updated.

1 Des noms d'affichage significatifs facilitent la recherche des zones souhaitées des
1 utilisateurs. Ils permettent également aux utilisateurs d'effectuer une recherche sur
1 une seule zone, telle que Auteur, et de trouver des documents ayant des noms de
1 zones internes différents dans leurs sources de données respectives (telles que
1 Auteur, Créateur, Producteur, etc).

1 Configuration et surveillance du programme d'écoute de données

1 Vous pouvez utiliser la console d'administration de recherche d'entreprise pour
1 configurer le support de vos applications d'écoute de données personnalisées.

1 Dans DB2 II OmniFind Edition, version 8.2, il était nécessaire d'éditer
1 manuellement plusieurs fichiers avant que vos applications puissent mettre à jour
1 les collections de recherche d'entreprise. Les informations suivantes remplacent les
1 instructions figurant dans le chapitre *Configuration du support des moteurs de balayage*
1 *externes* dans *Administering Enterprise Search*.

1 Une application d'écoute de données peut ajouter des données à une collection,
1 supprimer des URI d'une collection ou commander à un moteur de balayage sur le
1 Web de visiter ou revisiter des adresses URL. Lorsque l'utilisateur (client) se
1 connecte au programme d'écoute de données, il a besoin de transférer un ID client,
1 un mot de passe et l'ID de la collection à mettre à jour. Le programme d'écoute de
1 données authentifie l'ID client et le mot de passe et, si ce dernier est valide,
1 autorise l'application à mettre à jour la collection spécifiée.

1 Pour configurer le support de vos applications d'écoute de données, sélectionnez
1 l'option **Système** sur la barre d'outils de la console d'administration. Les fenêtres
1 suivantes permettent d'indiquer des informations relatives à vos applications
1 d'écoute de données et de contrôler leur activité :

- 1 • «Vue Système : page Programme d'écoute de données»
- 1 • «Vue Système du moniteur : page Programme d'écoute de données», à la page
1 18
- 1 • «Page Détails du programme d'écoute de données», à la page 19

1 Vue Système : page Programme d'écoute de données

1 La page Programme d'écoute de données de la vue Système permet de spécifier
1 des informations permettant à vos applications personnalisées d'écoute de données
1 de mettre à jour les collections.

1 Une application d'écoute de données vous permet d'explorer les types de sources
1 de données qui ne peuvent pas être explorées par les moteurs de balayage par
1 défaut fournis avec DB2 Information Integrator OmniFind Edition. Votre
1 application peut ajouter des données à une collection, supprimer les données d'une
1 collection ou or charger un moteur de balayage du Web de visiter ou de revisiter
1 des adresses URL.

1 **Zones et commandes**

1 **Nombre d'unités d'exécution**

1 Entrez le nombre maximal d'unités d'exécution que le composant
1 programme d'écoute de données peut créer pour traiter des requêtes issues
1 des applications d'écoute de données. La valeur par défaut est 6.

1 **Port du programme d'écoute de données**

1 Tapez le numéro de port correspondant au composant programme d'écoute
1 de données. La valeur par défaut est 6668.

1 **Important :** Si vous changez de numéro de port, vous devez relancer le
1 composant Programme d'écoute de données pour que les
1 modifications entrent en vigueur.

1 **Ajouter un ID de client d'écoute de données**

1 Ajoutez une ligne vierge à la liste d'identificateurs de client afin de
1 pouvoir spécifier des informations sur une nouvelle application d'écoute
1 de données.

1 **ID de client d'écoute de données**

1 Tapez l'ID du client pour l'application d'écoute de données que vous
1 souhaitez autoriser à accéder aux collections de recherche d'entreprise.
1 Chaque ID doit être unique dans le système de recherche d'entreprise.

1 **Mot de passe**

1 Entrez un mot de passe pour l'application d'écoute de données. Pour
1 accéder à une collection de recherche d'entreprise, l'application d'écoute de
1 données doit utiliser les données d'identification (ID et mot de passe) que
1 vous indiquez.

1 **Nom de collection**

1 Choisissez la collection que cette application d'écoute de données mette à
1 jour. Chaque application peut mettre à jour le contenu d'une seule
1 collection.



1 **Supprimer**

1 Supprimez l'application d'écoute de données sélectionnée du système de
1 recherche d'entreprise. Cette application ne sera pas utilisée pour ajouter
1 des données dans des collections de recherche d'entreprise, ni pour en
1 supprimer.

1 **Vue Système du moniteur : page Programme d'écoute de 1 données**

1 La page Programme d'écoute de données permet, lorsque vous contrôlez le
1 système, de visualiser des informations relatives au composant programme
1 d'écoute de données et à l'activité de l'application d'écoute de données.

Zones et commandes

Régénérer

Régénère les informations contenues sur cette page. Un message de statut indique quand ces informations ont été régénérées pour la dernière fois.

Nom Affiche le nom du composant (**Programme d'écoute de données**).

Détails

Affiche des informations détaillées concernant l'activité du programme d'écoute de données. (Les détails ne sont pas disponibles si le programme d'écoute de données est arrêté.)

Statut Les icônes suivantes indiquent le statut actuel du programme d'écoute de données :

Exécuter

Indique que le composant programme d'écoute de données est actif. Ce composant s'exécute automatiquement lorsque le système de recherche d'entreprise est lancé. En général, il est inutile de l'arrêter ou de le redémarrer à moins que vous ne changiez le numéro de port que vous avez configuré pour lui.

Arrêt terminé

Indique que le composant programme d'écoute de données est arrêté.

Page Détails du programme d'écoute de données

La page Détails du programme d'écoute de données permet de contrôler les informations relatives à l'activité en cours du programme d'écoute de données, telles que par exemple l'état de l'unité d'exécution et les requêtes sur l'application d'écoute de données.

Zones et commandes

Régénérer

Régénère les informations contenues sur cette page. Un message de statut indique quand ces informations ont été régénérées pour la dernière fois.

Statut Les icônes suivantes indiquent le statut actuel du programme d'écoute de données :

Exécuter

Indique que le composant programme d'écoute de données est actif. Ce composant s'exécute automatiquement lorsque le système de recherche d'entreprise est lancé. En général, vous n'avez pas besoin de l'arrêter ni de le relancer.

Arrêt terminé

Indique que le composant programme d'écoute de données est arrêté.

Nombre de requêtes en attente de traitement

Affiche le nombre de requêtes sur l'application d'écoute de données qui sont en attente de traitement.

Nombre total d'unités d'exécution

Affiche le nombre total d'unités d'exécution que le composant programme d'écoute de données peut créer pour traiter les requêtes issues des applications d'écoute de données.

Etat de l'unité d'exécution

Affiche les états possibles de l'unité d'exécution :

- Unités d'exécution initialisées
- Les unités d'exécution lisent une requête du client
- Les unités d'exécution traitent une requête du client
- Les unités d'exécution envoient une réponse au client
- Les unités d'exécution complétées envoyant une réponse au client
- Les unités d'exécution en attente de la prochaine requête du client
- Les unités d'exécution sont en pause
- Les unités d'exécution sont dans un processus d'arrêt

Nombre d'unités d'exécution

Affiche le nombre d'unités d'exécution actives pour un état donné (par exemple, 5 unités extraient activement des données).

Arrêt et démarrage du programme d'écoute de données

Si vous changez le numéro de port qui est configuré pour le programme d'écoute de données, vous devez arrêter le composant, puis le relancer.

Procédure

Pour arrêter et redémarrer le programme d'écoute de données, procédez comme suit :

1. Sur le serveur d'index, connectez-vous en tant qu'administrateur de recherche d'entreprise. Cet ID utilisateur a été indiqué lors de l'installation de DB2 II OmniFind Edition.
2. Placez-vous dans le répertoire ES_NODE_ROOT/master_config/datalistener.
3. Pour arrêter et redémarrer le programme d'écoute de données, entrez les commandes suivantes :

```
esadmin datalistener stop  
esadmin datalistener start
```

Estimation du nombre de documents dans une collection

Lors de la création d'une collection, vous pouvez fournir une estimation du nombre de documents que vous attendez. Le gestionnaire de ressources de DB2 II OmniFind Edition utilise ce nombre pour estimer la mémoire et les ressources disque nécessaires à la collection.

Le gestionnaire de ressources vous envoie des avertissements lorsque votre estimation de ressources de mémoire et de disque dépasse la capacité de votre système. Les avertissements vous permettent de prévenir les éventuels problèmes de manque de ressources.

Le moniteur utilise le nombre de documents estimé conjointement avec un pourcentage de seuil que vous configurez pour la réception d'alertes, afin de déterminer la date d'envoi des notifications concernant le risque de dépassement du nombre maximal de documents dans une collection.

Pour fournir une estimation de taille potentielle d'une collection, tapez un nombre dans la zone relative au **nombre de documents estimé** lorsque vous créez la collection. La valeur par défaut est 1 million de documents (1 000 000).

Après la création d'une collection, vous pouvez changer cette estimation en modifiant la collection et en spécifiant une nouvelle valeur dans la page Général.

Lorsque la collection atteint la taille estimée, le système n'interrompt pas l'ajout de documents à l'index. Cependant, si vous configurez des alertes pour la collection et activez l'option de notification lorsque le nombre de documents dans l'index dépasse une limite, vous recevrez des notifications lorsque la taille de la collection atteint le pourcentage de notifications spécifié.

Remarque : Lorsque vous configurez cette alerte dans la version 8.2, vous deviez spécifier une limite pour le nombre de documents dans l'index en tapant le nombre de documents que l'index pouvait contenir. Avec la version 8.2.1, le nombre spécifié dans la zone relative au **nombre de documents estimé** s'affiche automatiquement dans la zone **Limite**, et vous configurez le seuil de notification en spécifiant que vous souhaitez être averti lorsque le nombre de documents de l'index atteint un certain pourcentage de cette limite.

Le contrôle des ressources système est automatique avec l'assistant Collection

Si vous utilisez l'assistant Collection, le système estime automatiquement quelles sont les ressources système adéquates pour l'ajout d'une nouvelle collection ou d'un moteur de balayage à votre système de recherche d'entreprise. Dans DB2 II OmniFind Edition, version 8.2, vous ne pouvez vérifier les ressources système qu'en cliquant sur le bouton **Vérifiez les ressources système**.

Si les ressources système sont vérifiées après que vous créez une collection ou un moteur de balayage, mais avant de commencer à explorer des documents, vous pouvez déterminer si les ressources système disponibles sont adéquates pour exécuter DB2 II OmniFind Edition à pleine capacité en fonction des paramètres de la configuration courante.

Pour plus d'informations sur cette fonction, consultez la note technique dont le numéro de référence est 1193761 sur le site Web de DB2 II OmniFind Edition : Checking DB2 Information Integrator OmniFind Edition system resources (Vérification des ressources système DB2 Information Integrator OmniFind Edition).

Contrôle de l'activité de l'analyseur syntaxique

Pendant que vous contrôlez une collection, vous pouvez afficher les informations détaillées concernant l'activité de l'analyseur syntaxique.

Dans la version 8.2 de DB2 II OmniFind Edition, vous pouviez lancer ou interrompre l'analyseur syntaxique, mais vous ne pouviez pas afficher les informations de statut détaillées, ni obtenir des informations concernant le nombre de documents analysés.

Pour contrôler l'analyseur syntaxique d'une collection et afficher les informations de statut détaillées :

- 1 1. Dans la console d'administration de recherche d'entreprise, cliquez sur l'icône
- 1 de **surveillance** pour la collection que vous souhaitez contrôler.
- 1 2. Cliquez sur la page Analyseur syntaxique.
- 1 3. Si l'état de l'analyseur syntaxique est **en cours d'exécution**, cliquez sur l'icône
- 1 **Détails**.
- 1 4. Utilisez la fenêtre suivante pour afficher les détails concernant l'activité de
- 1 l'analyseur syntaxique :
- 1
 - «Page Détails de l'analyseur syntaxique»

1 Page Détails de l'analyseur syntaxique

1 Utilisez la page d'informations sur l'analyseur syntaxique pour afficher les

1 informations de statut détaillées concernant les documents analysés pour une

1 collection. Certaines options vous permettent d'analyser des rapports statistiques et

1 de gérer l'activité de l'analyseur syntaxique.

1 Zones et commandes

1 Régénérer

1 Régénère les informations contenues sur cette page. Un message de statut

1 indique quand ces informations ont été régénérées pour la dernière fois.

1 Parser status summary (Résumé du statut de l'analyseur syntaxique)

1 Les options disponibles dans cette zone vous indiquent le statut de

1 l'analyseur syntaxique et fournissent des informations statistiques sur les

1 documents qui sont analysés mais pas encore indexés.

1 **Statut** La première icône de la ligne **Statut** indique le statut actuel du

1 moteur de balayage :

1 Exécuter

1 Indique que l'analyseur syntaxique est en cours

1 d'exécution.

1 Selon l'état réel de l'analyseur, un message contenant des

1 informations supplémentaires sur l'activité d'analyse en

1 cours risque de s'afficher. Par exemple, l'analyseur peut

1 analyser des documents de manière active, être en veille

1 (jusqu'au prochain document prêt pour l'analyse, en

1 attente d'être relancé suite à une condition d'erreur, ou en

1 attente du redémarrage du service), ou en pause (en

1 attente, par exemple, qu'une réorganisation de l'index soit

1 effectuée).

1 Arrêt terminé

1 Indique que l'analyseur syntaxique est arrêté.

1 Selon le statut actuel de l'analyseur syntaxique et votre rôle

1 administratif, les icônes suivantes vous permettent de gérer

1 l'activité de l'analyseur :

1 Démarrer

1 Démarre l'analyseur syntaxique.

1 Une fois que l'analyseur syntaxique est lancé, il analyse

1 généralement les documents explorés jusqu'au dernier.

1 L'analyseur syntaxique entre alors dans un mode

d'économie d'énergie et ne s'exécute pas avant d'avoir vérifié la disponibilité des nouveaux documents explorés à analyser.



Arrêter

Arrête l'analyseur syntaxique.

Vous pouvez arrêter et redémarrer l'analyseur syntaxique, par exemple, si vous changez le type de catégorisation ou que vous ajoutez des règles de catégorie.

Number of documents in the data store that need to be indexed (Nombre de documents du magasin de données à indexer)

Affiche le nombre de documents que l'analyseur syntaxique a analysé et enregistré dans le magasin de données. Ces documents n'ont pas encore été ajoutés à l'index de recherche d'entreprise.

Les statistiques suivantes sont disponibles pour chaque moteur de balayage que contient la collection :

Snapshot statistics (Statistiques instantanées)

Les options disponibles dans cette zone vous indiquent le statut courant des activités de l'analyseur syntaxique et fournissent des informations statistiques sur les documents qui sont collectés par les moteurs de balayage individuels. Pour obtenir les meilleurs résultats possibles, ne générez pas l'index avant que l'analyseur ait analysé tous les documents explorés par les divers moteurs de la collection.

Snapshot start (Démarrer l'instantané)

Indique la date et l'heure auxquelles les statistiques de la vue en cours du statut de l'analyseur ont été collectées.

Parse rate (Débit d'analyse)

Affiche le nombre de pages que l'analyseur syntaxique analyse par seconde.

Les statistiques suivantes sont disponibles pour chaque moteur de balayage que contient la collection :

Nom du moteur de balayage

Affiche le nom de chaque moteur de balayage dans la collection.

Type de moteur de balayage

Pour chaque moteur de balayage, affiche le type de données que le moteur de balayage explore.

Documents parsed (Documents analysés)

Pour chaque moteur de balayage, affiche le nombre de documents qui sont prêts pour l'analyse syntaxique et le nombre de documents qui ont été analysés jusqu'à présent.

Statut Pour chaque moteur de balayage, affiche le statut des activités d'analyse de documents :



Exécuter

Indique que l'analyseur syntaxique analyse activement les documents explorés par ce moteur de balayage.



Arrêt terminé

Indique que l'analyseur syntaxique n'analyse pas de

documents pour ce moteur de balayage (peut-être parce qu'aucun document n'est prêt pour l'analyse).



En pause

Indique que l'analyse syntaxique des documents explorés par ce moteur de balayage est en suspens.



Erreurs

Indique que des erreurs se sont produites lors de l'analyse syntaxique des documents explorés par ce moteur de balayage. Pour analyser les erreurs, ouvrez la page Journal lors du contrôle de la collection et examinez les fichiers journaux relatifs à l'analyseur syntaxique.



Terminé

Indique que l'analyseur syntaxique a terminé d'analyser les documents explorés par ce moteur de balayage.

Mappage des métadonnées HTML pour la recherche de zones

L'analyseur syntaxique peut mapper les éléments HTML sur les noms de zones. Lorsque vous configurez l'analyseur syntaxique pour une collection, vous pouvez définir quels éléments HTML doivent être indexés comme zones de recherche. Les utilisateurs peuvent ensuite rechercher des parties spécifiques des documents HTML en indiquant ces noms de zones dans les requêtes.

Lorsque vous mappez des métadonnées HTML sur les zones, vous spécifiez si vous souhaitez que les utilisateurs puissent rechercher la zone, rechercher la zone avec une requête paramétrique, ou afficher la zone dans les résultats de recherche.

Pour mapper les éléments de métadonnées HTML pour rechercher des zones, veuillez effectuer les opérations suivantes :

1. Dans la console d'administration, cliquez sur l'icône **Edition** pour la collection que vous souhaitez modifier.
2. Sélectionnez la page Analyseur syntaxique.
3. Cliquez sur **Mapper les métadonnées HTML avec des zones**.
4. Utilisez la fenêtre suivante pour spécifier des informations concernant les zones de métadonnées que vous souhaitez rechercher :
 - «Page de mappages de zone de métadonnées HTML»

Page de mappages de zone de métadonnées HTML

Cette page permet de mapper des éléments de métadonnées HTML sur des noms de zones. En définissant les noms des zones mappées dans les requêtes, les utilisateurs peuvent rechercher des parties spécifiques des documents HTML et obtenir des résultats de recherche plus précis.

Lorsque vous créez un mappage de zone HTML, le changement devient effectif lorsque vous arrêtez, puis relancez l'analyseur syntaxique la fois suivante. Le nouveau mappage n'a pas d'effet sur les données qui sont déjà analysées et indexées.

Zones et commandes

Ajouter une zone

Ajoute une ligne vierge dans laquelle vous pouvez mapper un autre élément de métadonnées HTML sur un nom de zone de recherche.

Pour chaque élément de métadonnées HTML que vous souhaitez mapper sur un nom de zone de recherche, fournissez les informations suivantes :

Nom de zone

Entrez un nom que vous souhaitez associer à l'élément de métadonnées HTML. Les utilisateurs peuvent indiquer ce nom de zone lorsqu'ils lancent une requête sur les documents HTML de cette collection. Par exemple, pour mapper un code de métadonnées HTML appelé `publisher`, vous pouvez entrer `publisher` comme nom de zone de recherche.

Cette valeur doit être un mot unique qui contient seulement des caractères alphanumériques.

Zone de métadonnées HTML

Entrez le nom de l'élément de métadonnées HTML qui mappe sur le nom de zone que vous avez indiqué dans la zone **Nom de zone**. Pour utiliser le même exemple, tapez `publisher` dans cette zone car vous mappez l'élément de métadonnées HTML `publisher` sur une zone de recherche appelée `publisher`.

Cette valeur doit être un mot unique qui contient seulement des caractères alphanumériques.

Recherche par nom de zone

Cochez cette case si vous souhaitez que les utilisateurs puissent rechercher des documents HTML en définissant le nom de zone que vous avez spécifié dans la zone **Nom de zone**. Si cette case à cocher est désactivée, les utilisateurs ne peuvent pas définir le nom de la zone lorsqu'ils recherchent des documents HTML dans cette collection.

Recherche paramétrique

Activez cette case à cocher si vous souhaitez que les utilisateurs puissent formuler des requêtes paramétriques lorsqu'ils recherchent cette zone. Les requêtes paramétriques permettent aux utilisateurs d'évaluer les valeurs numériques et de date. Par exemple, un utilisateur peut déterminer si la valeur de cette zone est inférieure ou supérieure à une autre valeur.

Si vous cochez cette case, assurez-vous que le type de données de l'élément de métadonnées HTML est `DECIMAL`, `DOUBLE`, `INTEGER`, `SHORT`, `TIME` ou `TIMESTAMP`.

Afficher dans les résultats de la recherche

Activez cette case à cocher si vous souhaitez afficher cette zone dans les résultats de la recherche. Si cette case à cocher est désactivée, cette zone n'est pas incluse dans les résultats de la recherche.



Supprimer

Supprime le mappage de la zone de métadonnées HTML sélectionnée.

Suppression de nouveaux caractères de changement de ligne provenant d'espaces dans des documents japonais ou chinois

Dans les langues pour lesquelles les espaces ne sont pas utilisés pour délimiter les limites des mots, comme le japonais ou le chinois, vous pouvez configurer l'analyseur syntaxique de manière à ce qu'il supprime certains caractères d'espaces entraînant des sauts de lignes.

Pour activer cette fonctionnalité pour une collection :

1. Connectez-vous en tant qu'administrateur de recherche d'entreprise (ces ID utilisateur et mot de passe ont été configurés pendant l'installation de DB2 II OmniFind Edition).
2. Utilisez un éditeur de texte pour modifier le fichier `collection_ID.parserdriver/collection.properties`, où `collection_ID` représente l'ID que le système a affecté à la collection lors de sa création.
3. Modifiez la valeur de la propriété `removeCjNewlineChars` de `false` en `true`.

Si cette fonctionnalité est activée et qu'un document de la collection spécifiée est en japonais ou en chinois, l'analyseur syntaxique supprime toute suite de nouveaux caractères de lignes ou de tabulation séparant deux caractères de lettres du champ de caractères non ASCII. Les caractères suivants sont supprimés : tabulation (0x09), saut de ligne (0x0A) et retour chariot 0x0D.

Configuration de types de documents pour un analyseur syntaxique de collection et une session Stellent

Dans DB2 II OmniFind Edition version 8.2, les types de fichiers et de contenu sont pré-associés à un analyseur syntaxique de collection particulier et à une session de filtrage de document Stellent. Vous pouvez maintenant créer des fichiers de configuration pour spécifier les types de documents devant être analysés.

Dans la version 8.2, certains types de documents ne sont pas analysés (comme les documents PostScript), certains sont figés dans le code pour être envoyés à la session Stellent pour le traitement (Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT et Ichitaro). Tous les autres formats de documents sont traités par les analyseurs HTML ou XML intégrés. Les métadonnées peuvent être incorrectes, les documents en texte brut et HTML peuvent donc être envoyés à la session Stellent par erreur, puis renvoyés à l'un des analyseurs syntaxiques (une opération qui peut affecter la performance).

Pour vous assurer que les documents de votre espace d'exploration sont correctement et efficacement analysés, vous pouvez créer des fichiers de configuration pour spécifier les types de documents à analyser par l'analyseur syntaxique de collection et ceux à analyser par la session Stellent.

L'association des types de documents à l'analyseur syntaxique de collection et à la session Stellent nécessite les tâches suivantes :

1. La configuration de types de documents pour l'analyseur syntaxique de collection. Cette étape entraîne la création d'un fichier de configuration qui mappe les types de documents sur l'analyseur syntaxique utilisé par une collection. Vous pouvez créer l'un de ces fichiers de configuration par collection.
2. La configuration de types de documents pour la session Stellent. Cette étape entraîne la création d'un fichier de configuration qui mappe les types de

documents sur les filtres de document Stellent utilisés par une collection. Vous pouvez créer l'un de ces fichiers de configuration par collection.

3. Pour que les modifications prennent effet, utilisez la console d'administration de recherche d'entreprise pour contrôler la collection pour laquelle vous avez configuré des types de documents, puis arrêtez et redémarrez l'analyseur syntaxique.

Configuration de types de documents pour l'analyseur syntaxique de collection

Pour associer des types de documents particuliers à l'analyseur syntaxique de collection, créez un fichier de configuration. Cette fonctionnalité n'est pas prise en charge dans la console d'administration de recherche d'entreprise.

Le fichier de configuration spécifie :

- Les documents que vous souhaitez envoyer à la session Stellent, selon l'extension de fichier ou le type de contenu.
- La façon d'analyser les documents dont le type est inconnu à cause de métadonnées incomplètes.

Si le fichier de configuration n'existe pas, l'analyseur syntaxique de collection utilise les paramètres de configuration par défaut, qui correspondent exactement au comportement de DB2 II OmniFind Edition, version 8.2. Pour obtenir une description de ces paramètres par défaut, voir «Règles de service par défaut de l'analyseur syntaxique», à la page 28.

Pour associer des types de documents à l'analyseur syntaxique de collection :

1. Sur le serveur d'index, connectez-vous en tant qu'administrateur de recherche d'entreprise. Cet ID utilisateur a été spécifié pendant l'installation de DB2 II OmniFind Edition.

2. Créez le fichier de configuration comme suit, où *collection_ID* identifie la collection que vous souhaitez configurer :

```
ES_NODE_ROOT/master_config/collection_ID.parserdriver/parserTypes.cfg
```

Ce fichier présente les extensions de fichier et types de contenu à envoyer à la session Stellent. Le format du fichier est une suite de lignes, où chaque ligne représente l'une des règles suivantes :

EXTENSION *analyseur d'extension*

Tous les documents dont l'adresse URL se termine par l'extension spécifiée seront traités par l'analyseur syntaxique spécifié. N'incluez aucun point dans l'extension. La comparaison ne respecte pas les majuscules/minuscules.

CONTENTTYPE *analyseur de type/sous-type*

Tous les documents dont le type de contenu correspond au type/sous-type spécifié seront traités par l'analyseur syntaxique spécifié. Selon le type de contenu t/s d'un document, une correspondance se produit si t est équivalent au type, et s est équivalent au sous-type ou bien le sous-titre est un caractère générique (astérisque *).

UNKNOWN *analyseur*

Tous les documents dont l'extension et le type de contenu sont inconnus (c'est-à-dire non rendus disponibles par le moteur de balayage) seront traités par l'analyseur syntaxique spécifié.

1 **DEFAULT** *analyseur*
1 Tous les documents qui ne sont pas concernés par d'autres règles seront
1 traités par l'analyseur syntaxique spécifié.

1 Dans tous les cas, *l'analyseur syntaxique* doit être `html`, `xml`, `stellent` ou `none`, où
1 `none` signifie que le document ne doit pas être analysé.

1 **Priorité de règle** : Si plusieurs règles correspondent à un document, c'est la
1 règle la plus spécifique qui prévaut, quel que soit l'ordre dans lequel elles
1 apparaissent. Dans la liste ci-dessus, les règles apparaissent de la plus
1 spécifique à la moins spécifique.

- 1 • Une règle `EXTENSION` est plus spécifique qu'une règle `CONTENTTYPE`.
- 1 • Une règle `CONTENTTYPE` incluant un sous-type est plus spécifique qu'une
1 règle incluant un caractère générique. Par exemple, une règle pour un type
1 de contenu `application/postscript` est prioritaire par rapport à une règle
1 `application/*`.
- 1 • Il ne doit pas exister deux règles pour la même extension ou le même type
1 de contenu. Si c'est le cas, c'est la mise en oeuvre qui fixera la règle
1 prioritaire.

1 **Exemple**

1 Dans cet exemple, l'analyseur syntaxique intégré HTML traite tous les documents
1 avec une extension `txt`, `htm` ou `html`, avec un type de contenu commençant par
1 `text/`, ou avec une extension ou un type de contenu inconnu(e). L'analyseur
1 syntaxique intégré XML traite tous les documents avec une extension `xml` ou un
1 type de contenu `text/xml`. Tous les autres documents, notamment ceux dont le
1 type de contenu commence par `application/`, sont envoyés à la session `Stellent`.

```
1                   EXTENSION doc stellent  
1                   EXTENSION txt html  
1                   EXTENSION htm html  
1                   EXTENSION html html  
1                   EXTENSION xml xml  
1                   EXTENSION ps none  
1                   CONTENTTYPE text/xml xml  
1                   CONTENTTYPE text/* html  
1                   CONTENTTYPE application/* stellent  
1                   UNKNOWN html  
1                   DEFAULT stellent
```

1 **Règles de service par défaut de l'analyseur syntaxique**

1 Si vous ne créez pas de fichier de configuration pour mapper les types de fichiers
1 et les types de contenus sur l'analyseur syntaxique pour une collection, les règles
1 par défaut sont utilisées pour analyser des documents.

1 Les règles par défaut utilisées par l'interpréteur de la collection sont les suivantes :

```
1                   EXTENSION pdf stellent  
1                   EXTENSION ppt stellent  
1                   EXTENSION prz stellent  
1                   EXTENSION lwp stellent  
1                   EXTENSION doc stellent  
1                   EXTENSION rtf stellent  
1                   EXTENSION xls stellent  
1                   EXTENSION 123 stellent  
1                   EXTENSION vsd stellent  
1                   EXTENSION vdx stellent  
1                   EXTENSION jxw stellent  
1                   EXTENSION jsw stellent  
1                   EXTENSION jtw stellent  
1                   EXTENSION jaw stellent
```

```

1      EXTENSION jvw stellent
1      EXTENSION jbw stellent
1      EXTENSION jvw stellent
1      EXTENSION jfw stellent
1      EXTENSION jtt stellent
1      EXTENSION jtd stellent
1      EXTENSION jttc stellent
1      EXTENSION jtdc stellent
1      EXTENSION jtdx stellent
1      EXTENSION ps none
1      EXTENSION xml xml
1      EXTENSION txt text
1      EXTENSION htm html
1      EXTENSION html html
1      EXTENSION shtml html
1      EXTENSION xhtml html
1      EXTENSION asp html
1
1      CONTENTTYPE application/postscript none
1      CONTENTTYPE application/* stellent
1      CONTENTTYPE text/rtf stellent
1      CONTENTTYPE text/richtext stellent
1      CONTENTTYPE text/xml xml
1      CONTENTTYPE text/html html
1      CONTENTTYPE text/plain text
1
1      UNKNOWN html
1      DEFAULT html

```

Configuration de types de documents pour la session Stellent

Pour spécifier les types de documents à analyser par les filtres de documents Stellent, créez un fichier de configuration. Cette fonctionnalité n'est pas prise en charge dans la console d'administration de recherche d'entreprise.

Le fichier de configuration spécifie :

- Les types de fichiers à analyser par la session Stellent. Un type de fichier correspond à un des types de fichiers reconnus par la bibliothèque Stellent.
- Les types de fichiers à renvoyer à l'analyseur syntaxique de collection pour traitement par l'un des analyseurs intégrés. (Cette opération est nécessaire car l'analyseur syntaxique de collection peut potentiellement envoyer un document à la session Stellent par erreur, à cause de métadonnées incorrectes.)
- Les types de fichiers à rejeter car ils ne sont pas pris en charge par DB2 II OmniFind Edition.

Si le fichier de configuration a été spécifié mais n'existe pas, l'analyseur syntaxique ne pourra pas démarrer. Si aucun fichier de configuration n'a été spécifié pour la propriété `OutsideInSupportedTypes` du fichier `stellent.properties`, ce sont les paramètres de configuration par défaut qui sont utilisés. Pour obtenir une description de ces paramètres par défaut, correspondant au comportement de DB2 II OmniFind Edition version 8.2, voir «Règles d'analyse syntaxique par défaut pour les sessions Stellent», à la page 31.

Pour configurer les types de documents pour la session Stellent :

1. Sur le serveur d'index, connectez-vous en tant qu'administrateur de recherche d'entreprise. Cet ID utilisateur a été spécifié pendant l'installation de DB2 II OmniFind Edition.
2. Modifiez le fichier `ES_NODE_ROOT/master_config/collection_ID.stellent/stellent.properties`, où `collection_ID` identifie la collection que vous souhaitez configurer.

- 1 3. Pour la propriété `OutsideInSupportedTypes`, spécifiez le chemin absolu du
1 fichier de configuration que vous créez.
1 Par exemple, vous pouvez créer le fichier de configuration suivant pour une
1 collection unique, puis le stocker avec d'autres fichiers spécifiques aux
1 collections :
- ```
1 ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```
- 1 Pour utiliser les mêmes paramètres pour toutes les collections, vous pouvez  
1 créer le fichier de configuration suivant, puis le stocker avec les autres fichiers  
1 au niveau du système. (Si vous optez pour cette solution, assurez-vous de  
1 spécifier le chemin dans le fichier `stellent.properties` pour chaque collection,  
1 tel que spécifié à l'étape 2, à la page 29.)
- ```
1 ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```
- 1 4. Créez le fichier de configuration. Ce fichier présente des types de documents,
1 ainsi que la manière dont ils sont traités. Le format du fichier est une suite de
1 lignes, où chaque ligne représente l'une des règles suivantes :
- ```
1 accept DEFAULT
1 accept ALL doctype
1 accept type doctype
1 native DEFAULT
1 native type doctype
1 reject type
```
- 1 Où :
- 1 *doctype*
- 1 Représente la valeur à utiliser pour le jeton de requête de type de  
1 document (*doctype*). Le jeton de type de document est interrogeable en  
1 indiquant, par exemple, `doctype:pdf` dans une requête.
- 1 *type* Représente l'une des valeurs `FI_` de la bibliothèque Stellent, et *doctype*  
1 représente la valeur à utiliser pour le jeton de type de document en cas  
1 d'application d'une règle.
- 1 **DEFAULT**
- 1 Signifie que la liste des types acceptés ou natifs, selon le type de règle,  
1 inclut la liste par défaut. Cette option vous permet de développer la  
1 configuration par défaut au lieu de la remplacer.
- 1 **ALL** Signifie que tous les types qui ne sont pas explicitement présentés sont  
1 acceptés par le jeton de type de document spécifié.
- 1 **Traitement de règle :**
- 1 • S'il existe une règle `accept` pour *type*, incluant la liste par défaut si `accept`  
1 `DEFAULT` a été spécifié, la valeur est acceptée.
  - 1 • Ou, s'il existe une règle `reject` pour *type*, la valeur n'est pas acceptée.
  - 1 • Ou, si `accept ALL` a été spécifié, la valeur est acceptée.
  - 1 • Dans les autres cas, la valeur n'est pas acceptée.
- 1 Si le type de document est accepté, c'est la valeur *doctype* ayant été spécifiée  
1 dans la règle appliquée qui est utilisée. Cette valeur est renvoyée à l'analyseur  
1 syntaxique de collection avec le contenu analysé. Si le type de document n'est  
1 pas accepté, le comportement suivant se produit :
- 1 • S'il existe une règle `native` pour *type*, incluant la liste par défaut si `native`  
1 `DEFAULT` a été spécifié, le document est renvoyé à l'analyseur intégré,  
1 indiquant la valeur pour le jeton *doctype* spécifié par cette règle. La valeur  
1 de *doctype* doit être `txt`, `htm` ou `xml`, indiquant respectivement le mode de  
1 texte brut, HTML ou XML.
  - 1 • Sinon, le document est rejeté et ne sera pas analysé.

## 1 Exemple

1 Selon le fichier de configuration suivant, la session Stellent accepte, en plus de la  
1 liste par défaut, le format Microsoft Visio.

```
1 accept DEFAULT
1 accept FI_VISI03 visio
1 accept FI_VISI04 visio
1 accept FI_VISI05 visio
1 accept FI_VISI06 visio
```

1 Selon le fichier de configuration suivant, les documents Postscript seront pris en  
1 charge et interrogeables avec un type de document ps, le format X pixmap (XPM)  
1 sera renvoyé à l'analyseur syntaxique de texte intégré, le format d'image PNG sera  
1 rejeté et tous les autres types de fichiers seront acceptés et interrogeables avec un  
1 type de document other.

```
1 accept DEFAULT
1 accept FI_POSTSCRIPT ps
1 native FI_XPIXMAP txt
1 accept ALL other
1 reject FI_PNG
```

## 1 Règles d'analyse syntaxique par défaut pour les sessions 1 Stellent

1 Si vous ne créez pas de fichier de configuration pour mapper les types de fichiers  
1 sur les filtres de document d'une session Stellent, les règles par défaut sont  
1 utilisées pour analyser des documents.

1 Les règles par défaut utilisées par la session Stellent sont les suivantes :

```
1 ACCEPT FI_WORD4 doc
1 ACCEPT FI_WORD5 doc
1 ACCEPT FI_RTF rtf
1 ACCEPT FI_WINWORD1 doc
1 ACCEPT FI_WINWORD1COMPLEX doc
1 ACCEPT FI_WINWORD2 doc
1 ACCEPT FI_WORD6 doc
1 ACCEPT FI_WINWORD6 doc
1 ACCEPT FI_ICHITAR03 jxw
1 ACCEPT FI_ICHITAR04 jsw
1 ACCEPT FI_WINWORD1J doc
1 ACCEPT FI_WINWORD5J doc
1 ACCEPT FI_RTfJ rtf
1 ACCEPT FI_WINWORD7 doc
1 ACCEPT FI_WORDPRO lwp
1 ACCEPT FI_WINWORD97 doc
1 ACCEPT FI_ICHITAR08 jtd
1 ACCEPT FI_WORDPRO97 lwp
1 ACCEPT FI_WINWORD2000 doc
1 ACCEPT FI_WINWORD2002 doc
1 ACCEPT FI_WINWORD2003 doc
1 ACCEPT FI_123R1 123
1 ACCEPT FI_123R2 123
1 ACCEPT FI_123R3 123
1 ACCEPT FI_EXCEL xls
1 ACCEPT FI_EXCEL3 xls
1 ACCEPT FI_EXCEL4 xls
1 ACCEPT FI_123R4 123
1 ACCEPT FI_EXCEL5 xls
1 ACCEPT FI_123R6 123
1 ACCEPT FI_EXCEL97 xls
1 ACCEPT FI_123R9 123
1 ACCEPT FI_EXCEL2000 xls
1 ACCEPT FI_EXCEL2002 xls
1 ACCEPT FI_EXCEL2003 xls
```



|   |                             |     |
|---|-----------------------------|-----|
| 1 | ACCEPT FI_FREELANCE         | prz |
| 1 | ACCEPT FI_POWERPOINT4       | ppt |
| 1 | ACCEPT FI_POWERPOINT3       | ppt |
| 1 | ACCEPT FI_POWERPOINT7       | ppt |
| 1 | ACCEPT FI_FREELANCE3        | prz |
| 1 | ACCEPT FI_POWERPOINTMAC3    | ppt |
| 1 | ACCEPT FI_POWERPOINTMAC4    | ppt |
| 1 | ACCEPT FI_PDF               | pdf |
| 1 | ACCEPT FI_EXTPOWERPOINT4    | ppt |
| 1 | ACCEPT FI_EXTPOWERPOINTMAC4 | ppt |
| 1 | ACCEPT FI_POWERPOINTMACB3   | ppt |
| 1 | ACCEPT FI_POWERPOINTMACB4   | ppt |
| 1 | ACCEPT FI_POWERPOINT97      | ppt |
| 1 | ACCEPT FI_PDFMACBIN         | pdf |
| 1 | ACCEPT FI_POWERPOINT9597    | ppt |
| 1 | ACCEPT FI_POWERPOINT2000    | ppt |
| 1 | ACCEPT FI_POWERPOINT2       | ppt |
| 1 |                             |     |
| 1 | NATIVE FI_HTML              | htm |
| 1 | NATIVE FI_HTML_LATIN2       | htm |
| 1 | NATIVE FI_HTML_JAPANESESJIS | htm |
| 1 | NATIVE FI_HTML_JAPANESEEUC  | htm |
| 1 | NATIVE FI_HTML_CHINESEBIG5  | htm |
| 1 | NATIVE FI_HTML_CHINESEEUC   | htm |
| 1 | NATIVE FI_HTML_CHINESEGB    | htm |
| 1 | NATIVE FI_HTML_KOREANHANGUL | htm |
| 1 | NATIVE FI_HTML_CYRILLIC1251 | htm |
| 1 | NATIVE FI_HTML_CYRILLICKO18 | htm |
| 1 | NATIVE FI_CYRILLIC1251      | txt |
| 1 | NATIVE FI_CYRILLICKO18      | txt |
| 1 | NATIVE FI_W2KHTML           | htm |
| 1 | NATIVE FI_XL2KHTML          | htm |
| 1 | NATIVE FI_PP2KHTML          | htm |
| 1 | NATIVE FI_XML               | xml |
| 1 | NATIVE FI_WML               | xml |
| 1 | NATIVE FI_HTML_JAPANESEJIS  | htm |
| 1 | NATIVE FI_WML_CHINESEBIG5   | xml |
| 1 | NATIVE FI_WML_CHINESEEUC    | xml |
| 1 | NATIVE FI_WML_CHINESEGB     | xml |
| 1 | NATIVE FI_WML_CYRILLIC1251  | xml |
| 1 | NATIVE FI_WML_CYRILLICKO18  | xml |
| 1 | NATIVE FI_WML_JAPANESEJIS   | xml |
| 1 | NATIVE FI_WML_JAPANESESJIS  | xml |
| 1 | NATIVE FI_WML_JAPANESEEUC   | xml |
| 1 | NATIVE FI_WML_KOREANHANGUL  | xml |
| 1 | NATIVE FI_WML_LATIN2        | xml |
| 1 | NATIVE FI_HTMLUNICODE       | htm |
| 1 | NATIVE FI_XML_DOCTYPE_HTML  | htm |
| 1 | NATIVE FI_XHTML             | htm |
| 1 | NATIVE FI_ASCII             | txt |
| 1 | NATIVE FI_ANSI              | txt |
| 1 | NATIVE FI_UNICODE           | txt |
| 1 | NATIVE FI_ASCII8            | txt |
| 1 | NATIVE FI_ANSI8             | txt |
| 1 | NATIVE FI_MAC               | txt |
| 1 | NATIVE FI_MAC8              | txt |
| 1 | NATIVE FI_SHIFTJIS          | txt |
| 1 | NATIVE FI_CHINESEGB         | txt |
| 1 | NATIVE FI_HANGEUL           | txt |
| 1 | NATIVE FI_CHINESEBIG5       | txt |
| 1 | NATIVE FI_LATIN2            | txt |
| 1 | NATIVE FI_JAPANESE_EUC      | txt |
| 1 | NATIVE FI_HEBREW_OLDCODE    | txt |
| 1 | NATIVE FI_HEBREW_PC8        | txt |
| 1 | NATIVE FI_HEBREW_E0         | txt |
| 1 | NATIVE FI_HEBREW_WINDOWS    | txt |
| 1 | NATIVE FI_ARABIC_710        | txt |



```
1 NATIVE FI_ARABIC_720 txt
1 NATIVE FI_ARABIC_WINDOWS txt
1 NATIVE FI_7BITTEXT txt
1 NATIVE FI_JAPANESE_JIS txt
1 NATIVE FI_CENTRALEU_1250 txt
1 NATIVE FI_UTF8 txt
1 NATIVE FI_EBCDIC_37 txt
1 NATIVE FI_EBCDIC_273 txt
1 NATIVE FI_EBCDIC_277 txt
1 NATIVE FI_EBCDIC_278 txt
1 NATIVE FI_EBCDIC_280 txt
1 NATIVE FI_EBCDIC_284 txt
1 NATIVE FI_EBCDIC_285 txt
1 NATIVE FI_EBCDIC_297 txt
1 NATIVE FI_EBCDIC_500 txt
1 NATIVE FI_EBCDIC_870 txt
1 NATIVE FI_EBCDIC_871 txt
1 NATIVE FI_EBCDIC_1026 txt
```

---

## 1 Améliorations du traitement des requêtes

1 Dans une configuration à multiples serveurs, la protection par basculement est  
1 disponible au niveau de la collection, et non pas au niveau du serveur. Si une  
1 collection sur un serveur de recherche devient indisponible pour une raison ou une  
1 autre, les requêtes de cette collection sont alors automatiquement transmises à un  
1 autre serveur de recherche.

---

## 1 Réduction des résultats issus du même site Web

1 Lorsque vous éditez le fichier SiteDefs.txt afin de spécifier des options  
1 permettant de réduire les résultats de la recherche issus du même site Web, il est  
1 possible de définir une chaîne comme identificateur de site avec le préfixe de  
1 l'adresse URL.

1 Dans DB2 II OmniFind Edition, version 8.2, le fichier SiteDefs.txt contenait  
1 uniquement une liste de préfixes URL. Il est désormais possible de définir une  
1 chaîne comme identificateur pour chaque site. Comme indiqué dans l'exemple  
1 suivant, vous tapez les préfixes URL au début de la ligne et, mais c'est facultatif,  
1 un identificateur de chaîne à droite :

```
1 http://mycompany.server1.com/hr/ hr
1 http://mycompany.server2.com/hr/ hr
1 http://mycompany.server3.com/hr/ hr
1 http://mycompany.server1.com/finance/ finance
1 http://mycompany.server1.com/news news
1 http://mycompany.server1.com/USA
```

1 Les applications de recherche peuvent soit utiliser le préfixe URL, soit  
1 l'identificateur pour regrouper et réduire les documents dans les résultats de la  
1 recherche. Par exemple, l'identificateur hr permet la réduction des résultats de la  
1 recherche provenant de trois serveurs qui hébergent des documents sur les  
1 ressources humaines.

1 La syntaxe d'une requête de recherche d'entreprise a été étendue de façon à  
1 prendre en charge cette fonctionnalité. Par exemple, pour extraire des documents  
1 qui appartiennent au même groupe, utilisez la syntaxe suivante, où *groupe*  
1 correspond soit à l'identificateur de groupe, soit à l'adresse URL complète d'un  
1 document qui appartient au groupe :

```
1 samegroupas:groupe
```

1 Par exemple :  
1 samegroupas:hr  
1 samegroupas:http://mycompany.server3.com/hr/

1 Pour plus d'informations sur la mise à jour du fichier SiteDefs.txt, reportez-vous  
1 à la rubrique permettant de *réduire les résultats provenant du même site Web* dans  
1 *Administering Enterprise Search*.

---

## 1 Mise en évidence étendue dans les résultats de recherche

1 Une propriété de l'interface Requête active permet la mise en évidence des termes  
1 de requête dans plusieurs zones des détails des résultats de recherche. Dans DB2 II  
1 OmniFind Edition, version 8.2, la mise en évidence était limitée à la zone du  
1 récapitulatif.

1 Lorsque vous créez une application de recherche, vous pouvez manipuler le  
1 traitement des requêtes en utilisant les méthodes getProperty et setProperty de  
1 l'interface Requête. Une nouvelle propriété que vous pouvez utiliser avec ces  
1 méthodes, HighlightingMode, vous permet de contrôler la manière dont les termes  
1 de requêtes sont mis en évidence dans les résultats de recherche. Cette propriété a  
1 les valeurs suivantes :

### 1 DefaultHighlighting

1 Met en évidence les termes de requête dans la zone de récapitulatif  
1 uniquement. Si votre application de recherche fait abstraction de la propriété  
1 HighlightingMode, il s'agit de l'action par défaut.

### 1 ExtendedHighlighting

1 Etend la mise en évidence de termes de requête vers d'autres zones de  
1 résultats de recherche, telles que le titre et l'adresse URL par exemple.

1 Vous pouvez extraire la zone de l'adresse URL mise en évidence des propriétés de  
1 l'objet Résultats en utilisant les méthodes getProperty ou getProperties. Le nom  
1 de la propriété est HighlightedDocumentID, sa valeur est l'adresse URL mise en  
1 évidence. La méthode de résultat getDocumentID continuera à renvoyer l'adresse  
1 URL non mise en évidence.

---

## 1 Prise en charge de délai d'attente dans les applications de recherche

1 Une propriété de l'interface API à distance vous permet de spécifier une valeur de  
1 délai d'attente pour les requêtes à distance.

1 Les applications de recherche peuvent transmettre la propriété de délai d'attente  
1 avec l'objet Properties à l'aide des méthodes suivantes :

```
1 SearchFactory.getSearchService(Properties)
1 BrowseFactory.getBrowseService(Properties)
```

1 Par exemple :

```
1 // create a new Properties object.
1 Properties config = new Properties();
1 config.setProperty("hostname", hostname);
1 config.setProperty("port", portNumber);
1 config.setProperty("locale", "en_US");
1 config.setProperty("timeout", "60");
1 SearchService searchService = factory.getSearchService(config);
```

1 La valeur de délai d'attente, qui représente le nombre de secondes à attendre avant  
1 la fin de la requête distante, doit être un entier (tel que 60, et non pas 60,5 ou

1 soixante). Si vous ne spécifiez pas de valeur de délai d'attente, l'interface API  
1 distante utilise la valeur par défaut, 30 secondes.



---

## Limitations connues, problèmes et solutions disponibles

Cette section présente les limitations connues, les problèmes et les solutions disponibles pour DB2 Information Integrator OmniFind Edition, version 8.2.

Les limitations et restrictions peuvent ou non s'appliquer à d'autres versions du produit.

---

### Décompression du fichier tar DB2 UDB pour AIX

Pour DB2 Universal Database Enterprise Server Edition, version 8.2 pour AIX, vous devez décompresser l'image du CD avant de pouvoir installer DB2 UDB.

Décompressez le CD avant de commencer l'installation de la recherche d'entreprise.

Pour plus d'informations sur la décompression du CD, rendez-vous dans le Centre d'aide et d'information DB2 version 8.2 à l'adresse suivante : <http://publib.boulder.ibm.com/infocenter/db2help/index.jsp>. Dans le Centre d'aide et d'Information DB2, cliquez sur **Installation** → **Systèmes de bases de données** → **DB2 Universal Database pour Linux, UNIX et Windows** → **Serveurs DB2** → **DB2 UDB Enterprise Server Edition (non partitionné)** → **AIX** → **Lancement de l'assistant d'installation DB2**.

Ce fichier doit être décompressé pour que le programme d'installation de recherche d'entreprise puisse installer DB2 UDB.

Pour décompresser l'image de CD, procédez comme suit :

1. Copiez produit.tar.Z, produit représentant le programme dont vous avez la licence d'installation, sur un système de fichiers temporaire.
2. Pour lancer l'assistant d'installation DB2, entrez la commande suivante :

```
zcat produit.tar.Z | tar -xf - ;./produit/db2setup
```

Par exemple, si le nom de produit pour DB2 UDB Enterprise Server Edition est ese, entrez la commande suivante :

```
zcat ese.tar.Z | tar -xf - ;./ese/db2setup
```

Lors de l'installation de la recherche d'entreprise, affectez le répertoire qui contient le fichier db2setup dans le chemin d'accès aux fichiers décompressé. Par exemple, le répertoire du fichier décompressé pourrait être .\ese.sbcsaix1.

---

### Segments de mémoire partagés dans AIX

Si vous obtenez un message SQL1224N dans AIX, il est possible que le nombre de segments de mémoire partagés autorisés par AIX soit dépassé.

Pour résoudre ce problème, voir la rubrique SQL1224N Shared Memory Segments du site Web d'IBM.

---

## Redémarrage du système après la désinstallation de Windows

Si vous désinstallez DB2 II OmniFind Edition sous Windows, vous devez redémarrer le système pour terminer la suppression de tous les fichiers du programme.

Vous devez également supprimer manuellement les répertoires d'installation de recherche d'entreprise. Si vous ne redémarrez pas le système, ni ne supprimez les fichiers du programme, avant d'installer la recherche d'entreprise à nouveau, des problèmes peuvent survenir.

---

## Processus URLFetcher orphelin pour Linux et AIX

**Problème :** Le moteur de balayage Web utilise deux processus. Lorsqu'un processus se termine, il demande au second processus de se terminer. Un processus peut quelquefois s'interrompre de manière inattendue et laisser l'autre processus s'exécuter pour un temps indéterminé.

Vous pouvez savoir que cet incident s'est produit si vous voyez un processus orphelin tel qu'il apparaît ci-dessous depuis la commande ps de UNIX :

```
F S UID PID PPID C PRI NI ADDR SZ WCHAN STIME TTY TIME CMD
0 S user1 26222 1 0 75 0 - 152641 schedu 10:34 pts/8 00:00:00
```

```
/opt/IBMJava2-141/jre/bin/java -classpath ../../URLFetcher.jar:...
```

Remarquez que :

- Un PPID (Parent Process ID, soit ID de processus parent) de 1 indique que le processus est orphelin.
- La présence de URLFetcher.jar dans le chemin de classe indique qu'il s'agit d'un processus de moteur de balayage Web.

### Solution

Trouvez le PID (ID du processus) du processus orphelin. Dans l'exemple précédent, il s'agit de 26222. Exécutez le processus kill. Sous Linux et AIX, utilisez la commande kill sans argument. Vous pouvez également voir des entrées «child» (enfant) dans la table de processus. Les entrées enfant représentent des seuils créés par le processus orphelin. Ces entrées doivent être supprimées lorsque vous exécutez le processus kill sur le processus ancêtre.

Lorsque vous voyez un processus URLFetcher orphelin, exécutez le processus kill car s'il est situé dans la table, le système d'exploitation lui réserve des ressources. Vérifiez régulièrement s'il existe des processus orphelins, surtout si vous pensez que le moteur de balayage Web s'est terminé de manière anormale. Le cas échéant, exécutez le processus kill.

**Remarque :** Dans DB2 II OmniFind Edition version 8.2.1, il existe peu de processus orphelins car l'extracteur URL s'auto-termine s'il remarque que le composant C++ a disparu. Si le problème survient, veillez à ne pas exécuter le processus kill dans la session d'édition httpauth.ini d'un autre utilisateur (l'utilitaire HTTPAuthEdit est un JVM exécuté depuis le fichier JAR URLFetcher qui peut sembler identique dans la table de processus).

---

## Emplacement d'index sur le serveur

Lors de la création d'une collection, si vous souhaitez spécifier un emplacement d'index différent de l'emplacement par défaut, l'utilisation d'espace disque pour l'emplacement que vous avez spécifié ne sera pas contrôlé s'il n'est pas situé dans le répertoire racine des données serveur.

Pour une installation à multiples serveurs, lors de la création d'une collection, si vous spécifiez un emplacement d'index différent de l'emplacement par défaut, ce dernier doit exister sur les quatre serveurs.

---

## Limites de taille de segment

L'analyseur syntaxique de document Java (également appelé service d'analyseur syntaxique) a une limite de taille de segment de 400 Mo.

La valeur de taille de segment est stockée ainsi :

```
ES_NODE_ROOT/config/services.ini:session7.max_heap=400
```

Si des messages d'insuffisance mémoire apparaissent dans le journal d'erreur du système pour la session de service d'analyseur syntaxique, arrêtez la session, augmentez la valeur `max_heap` dans le fichier `services.ini` du service et redémarrez la session d'analyseur syntaxique.

---

## Tables DB2 UDB avec des objets de grande taille

Les tables DB2 Universal Database avec des données LOB dont la taille est au moins équivalente à 128 Mo ne peuvent pas être explorées. Les documents dont la taille dépasse 8 Mo ne peuvent pas être indexés.

---

## Le statut de l'index ne s'affiche pas

Si vous installez à nouveau une nouvelle version de DB2 II OmniFind Edition sans supprimer l'installation précédente, il est possible que le statut de surveillance de l'index ne s'affiche pas correctement.

Lorsque vous démarrez la génération d'index, le statut de la génération ne s'affiche pas correctement à la page Index de la console d'administration. Cependant, l'index se génère, et lorsque l'opération est terminée, la page Index affiche le nombre exact de documents indexés.

Il est conseillé de supprimer la version précédente de DB2 II OmniFind Edition avant d'installer une nouvelle version.

---

## Autorisations ID utilisateur pour Linux et AIX

Il peut être nécessaire d'ajouter des droits en écriture pour un nouvel ID utilisateur de recherche d'entreprise.

Si vous installez à nouveau DB2 II OmniFind Edition avec un utilisateur Linux ou AIX différent de l'administrateur de recherche d'entreprise, lorsque vous ajoutez un moteur de balayage DB2 et spécifiez que vous souhaitez explorer des bases de données non cataloguées distantes (utilisant un pilote JDBC DB2 Universal Database Type 4), une erreur peut se produire. Dans ce cas, vous devez ajouter des

droits en écriture pour le nouvel utilisateur administrateur de recherche d'entreprise dans le fichier `/home/db2inst1/sqllib/tmp/0.spf`.

---

## 1 Activation de l'accès au document direct pour FileNet Content

### 1 Services

1 Lorsque vous utilisez le moteur de balayage VeniceBridge pour explorer les  
1 référentiels FileNet Content Services gérés par le serveur DB2 Information  
1 Integrator Content Edition (anciennement VeniceBridge), la bibliothèque cible  
1 FileNet Content Services et FileNet Web Services interagissent avec le connecteur  
1 FileNet Content Services et doivent être configurés pour utiliser un accès au  
1 document direct. Voir la documentation FileNet Web Services *Administrator's Guide*  
1 et FileNet Content Services pour configurer une prise en charge d'accès au  
1 document direct.

---

## 1 Limitations de l'exploration de tables de pseudonymes

1 Il existe des limitations à l'exploration des tables de pseudonymes d'IBM DB2  
1 Information Integrator pour IBM DB2 Universal Database pour z/OS et Microsoft  
1 SQL Server.

1 Si la table cible est une table de pseudonymes DB2 Information Integrator pour  
1 DB2 Universal Database pour z/OS ou Microsoft SQL Server et si la zone de  
1 contenu du document est spécifiée, les identificateurs uniques ne peuvent pas  
1 inclure de colonnes des types de données suivantes :

- 1 • LONG VARCHAR
- 1 • LONG VARCHAR FOR BIT DATA
- 1 • LONG VARGRAPHIC

1 Si la table cible est une table de pseudonymes DB2 Information Integrator pour  
1 Microsoft SQL Server et si cette table est créée dans une base de données avec  
1 UTF-8, le contenu de document des colonnes avec le type de données BLOB peut  
1 être corrompu.

1 Si la table cible est une table de pseudonymes DB2 Information Integrator pour  
1 DB2 Universal Database pour z/OS ou Microsoft SQL Server et si la zone de  
1 contenu du document est spécifiée, suivez les étapes suivantes pour améliorer la  
1 performance du moteur de balayage DB2 :

- 1 1. Créez un serveur avec l'option `COLLATING_SEQUENCE` paramétrée sur `Y`.
- 1 2. Créez le pseudonyme sous ce serveur.

1 Si la table de pseudonymes a déjà été définie au moment du paramétrage de  
1 l'option sur `Y`, supprimez le pseudonyme et créez-le à nouveau.



---

## Remarques

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent document. La remise de ce document ne vous donne aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante : IBM EMEA Director of Licensing IBM Europe Middle-East Africa Tour Descartes 92066 Paris-La Défense Cedex 50 France Pour le Canada, veuillez adresser votre courrier à : IBM Director of Commercial Relations IBM Canada Ltd 3600 Steeles Avenue East Markham, Ontario L3R 9Z7 Canada

Les informations sur les licences concernant les produits utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante : IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales : LE PRESENT DOCUMENT EST LIVRE «EN L'ETAT». IBM DECLINE TOUTE RESPONSABILITE, EXPLICITE OU IMPLICITE, RELATIVE AUX INFORMATIONS QUI Y SONT CONTENUES, Y COMPRIS EN CE QUI CONCERNE LES GARANTIES DE VALEUR MARCHANDE OU D'ADAPTATION A VOS BESOINS. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Il est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut modifier sans préavis les produits et logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Corporation J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 U.S.A.

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions de l'ICA, des Conditions internationales d'utilisation des logiciels IBM ou de tout autre accord équivalent.

Les données de performance indiquées dans ce document ont été déterminées dans un environnement contrôlé. Par conséquent, les résultats peuvent varier de manière significative selon l'environnement d'exploitation utilisé. Certaines mesures évaluées sur des systèmes en cours de développement ne sont pas garanties sur tous les systèmes disponibles. En outre, elles peuvent résulter d'extrapolations. Les résultats peuvent donc varier. Il incombe aux utilisateurs de ce document de vérifier si ces données sont applicables à leur environnement d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Toute ressemblance avec des noms de personnes, de sociétés ou des données réelles serait purement fortuite.

Licence sur les droits d'auteur :

Le présent logiciel contient des exemples de programmes d'application en langage source destinés à illustrer les techniques de programmation sur différentes plateformes d'exploitation. Vous avez le droit de copier, de modifier et de distribuer ces exemples de programmes sous quelque forme que ce soit et sans paiement d'aucune redevance à IBM, à des fins de développement, d'utilisation, de vente ou de distribution de programmes d'application conformes aux interfaces de programmation des plateformes pour lesquels ils ont été écrits ou aux interfaces de programmation IBM. Ces exemples de programmes n'ont pas été rigoureusement testés dans toutes les conditions. Par conséquent, IBM ne peut garantir expressément ou implicitement la fiabilité, la maintenabilité ou le fonctionnement de ces programmes. Vous avez le droit de copier, de modifier et de distribuer ces

exemples de programmes sous quelque forme que ce soit et sans paiement d'aucune redevance à IBM, à des fins de développement, d'utilisation, de vente ou de distribution de programmes d'application conformes aux interfaces de programmation IBM.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

Outside In (®) Viewer Technology, © 1992-2004 Stellent, Chicago, IL., Inc. Tous droits réservés.

IBM XSLT Processor Eléments sous licence - Propriété d'IBM © Copyright IBM Corp., 1999-2004. Tous droits réservés.

---

## Marques

La présente section contient la liste des marques IBM et de certaines marques non-IBM.

Les termes qui suivent sont des marques d'International Business Machines Corporation aux Etats-Unis et/ou dans certains autres pays :

IBM  
AIX  
AIX 5L  
DB2  
DB2 Universal Database  
Domino  
Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
WebSphere  
xSeries  
z/OS

Les termes qui suivent sont des marques d'autres sociétés :

Java, ainsi que tous les logos et toutes les marques incluant Java, sont des marques de Sun Microsystems, Inc. aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

Intel, Intel Inside (logos), MMX et Pentium sont des marques d'Intel Corporation aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

D'autres sociétés sont propriétaires des autres marques, noms de produits ou logos qui pourraient apparaître dans ce document.



---

## Pour prendre contact avec IBM

Pour prendre contact avec le service clientèle d'IBM aux Etats-Unis ou au Canada, appelez le 1-800-IBM-SERV (1-800-426-7378).

Pour en savoir plus sur les options de service, appelez l'un des numéros suivants :

- Aux Etats-Unis : 1-888-426-4343
- Au Canada : 1-800-465-9600

Pour obtenir l'adresse de votre agence IBM locale ou nationale, consultez la page IBM Directory of Worldwide Contacts sur le Web à l'adresse suivante : [www.ibm.com/planetwide](http://www.ibm.com/planetwide).

---

## Informations sur le produit

Vous pouvez obtenir des informations sur DB2 Information Integrator par téléphone ou sur le Web.

Les informations relatives au produit DB2 Information Integrator sont disponibles par téléphone ou via Internet. Les numéros de téléphone fournis ci-dessous sont valables uniquement aux Etats-Unis.

1. Pour commander un produit ou pour obtenir des informations générales, appelez le numéro suivant : 1-800-IBM-CALL (1-800-426-2255)
2. Pour commander des publications, appelez le numéro suivant : 1-800-879-2755
3. Rendez-vous sur le Web à l'adresse suivante : [www.ibm.com/software/data/integration/db2ii/support.html](http://www.ibm.com/software/data/integration/db2ii/support.html).

Ce site contient les dernières informations disponibles dans les domaines suivants :

- La bibliothèque technique
- La commande de manuels
- Les téléchargements sur des postes clients
- Les groupes de nouvelles
- Les correctifs (Fix packs)
- Les nouvelles
- Les liens aux ressources disponibles sur le Web

---

## Commentaires sur la documentation

Veillez nous faire parvenir tout commentaire éventuel sur le présent manuel ou sur toute autre documentation relative à DB2 Information Integrator.

Vos appréciations en retour aident IBM à fournir des informations de qualité. Veillez nous faire parvenir tout commentaire éventuel sur le présent manuel ou sur toute autre documentation relative à DB2 Information Integrator. Pour ce faire, vous pouvez faire appel à la méthode suivante :

1. Send your comments using the online readers' comment form at [www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf).
2. Envoyez vos commentaires par courrier électronique à l'adresse [comments@us.ibm.com](mailto:comments@us.ibm.com). Veillez inclure le nom et le numéro de version du

produit, ainsi que le nom et le numéro de référence de la documentation (si applicable). Si vos appréciations portent sur une portion de texte spécifique, n'oubliez pas de préciser l'emplacement de ce texte (un titre, un numéro de tableau ou de page par exemple).





Imprimé en France



**Java**<sup>™</sup>  
**COMPATIBLE**