

IBM eDiscovery Analyzer



用户指南

V 2.1

IBM eDiscovery Analyzer



用户指南

V 2.1

注意

在使用本资料及其支持的产品之前，请阅读第 19 页的『声明』中的信息。

本版本适用于 IBM eDiscovery Analyzer V2R1（产品号 5724-V36）以及所有后续发行版和修订版，直到新版本中另有指示为止。

© Copyright International Business Machines Corporation 2008.

目录

ibm.com 和相关资源	v	使用类别来缩小搜索结果的范围	13
如何发送意见	v	保存搜索	14
与 IBM 联系	vi	第 3 章 使用标志来标识文档相关性	17
第 1 章 IBM eDiscovery Analyzer 入门	1	声明	19
搜索策略	1	词汇表	23
第 2 章 搜索案件	3	索引	27
搜索技巧	4		
解释搜索结果	12		

ibm.com 和相关资源

ibm.com 提供了产品支持和文档。

支持与辅助

产品支持通过 Web 提供。请单击产品 Web 站点上的“支持”，网址如下：

IBM eDiscovery Analyzer

<http://www.ibm.com/software/data/content-management/support/ediscovery-analyzer/>

IBM eDiscovery Manager

<http://www.ibm.com/software/data/content-management/support/ediscovery-manager/>

信息中心

您可以在基于 Eclipse 的信息中心中查看产品文档（可以在安装产品时安装此信息中心）。缺省情况下，信息中心以 Web 服务器方式运行，并可供其他 Web 浏览器访问。您还可以在工作站上以本地方式运行信息中心。请参阅信息中心，其网址为 <http://publib.boulder.ibm.com/infocenter/email/v1r0m0/topic/com.ibm.email.doc/welcome.htm>。

PDF 出版物

您可以使用适合于操作系统的 Adobe Acrobat Reader 来查看在线 PDF 文件。如果尚未安装 Acrobat Reader，那么可以从 Adobe Web 站点（<http://www.adobe.com>）下载此产品。

请参阅下列 PDF 出版物 Web 站点：

产品	Web 站点
IBM eDiscovery Analyzer	http://www.ibm.com/support/docview.wss?rs=3533&context=SSJKLP&uid=swg27013410
IBM eDiscovery Manager	http://www-01.ibm.com/support/docview.wss?rs=3506&context=SS8JHU&uid=swg27012725

如何发送意见

您的反馈对于帮助我们提供最准确最高品质的信息而言至关重要。

请使用在线读者意见表（https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf）来发送您的意见。

与 IBM 联系

要与美国或加拿大的 IBM 客户服务联系，请致电 1-800-IBM-SERV (1-800-426-7378)。

要了解可用的服务选项，请拨打下列其中一个电话：

- 在美国：1-888-426-4343
- 在加拿大：1-800-465-9600

有关如何与 IBM 联系的更多信息，请访问“与 IBM 联系”Web 站点 (<http://www.ibm.com/contact/us/>)。

第 1 章 IBM eDiscovery Analyzer 入门

从这里开始，了解 IBM® eDiscovery Analyzer 如何使您能够对已归档的电子邮件文档进行搜索、分析和复审。

IBM eDiscovery Analyzer 帮助您收集可能与发现请求相关的电子邮件和附带文档，以便资深律师了解哪些文档可能需要进一步复审。eDiscovery Analyzer 可以帮助您减少通常发送到独立法律复审服务机构的文档量，从而降低此类复审的开销。

eDiscovery Analyzer 搜索引擎将搜索电子邮件的不同部分：发件人和收件人字段（收件人、发件人、抄送和密送）、主题、正文和附件。如果您指定搜索引擎只搜索电子邮件的某个部分，例如主题行，那么搜索引擎将不会搜索电子邮件的任何其他部分。

您在案件中执行搜索以标识与发现请求相关的文档，然后在执行法律复审过程前进一步缩小搜索范围。您的目标可能是：

- 更好地了解案件中电子邮件文档的内容
- 确定电子邮件文档的复审优先顺序
- 标识不需要进一步复审的电子邮件文档
- 标识并标记可能对发现指令有反应的电子邮件文档

通常，通过将文档标记为**有反应**或**第二级复审**等等，确定为了进一步复审而收集的文档集合。对相关文档作标记后，管理员可以使用 eDiscovery Manager 将文档打包，以便进一步复审。

搜索策略

您可以使用一个或多个搜索策略来标识可能与发现请求相关或者需要进一步复审的电子邮件文档。

修改搜索查询

完整的查询将显示在可编辑的搜索查询框中。您可以保存搜索以便重复使用，可以在各个查询之间前后移动以查看不同查询的结果，也可以编辑查询并运行新查询。

如果您熟悉搜索查询语法，那么可以直接在文本框中输入查询。

减少干扰

任何一组搜索结果都可能包含一些无关的文档，例如垃圾邮件、自动生成的公告以及公司声明等等。通过搜索文档中频繁出现的那些所要排除的短语或发件人，可以减少需要复审的文档数。例如，如果您知道自动生成的公告来自特定电子邮件地址，那么请搜索该地址并将结果标记为“无反应”。然后，可以在后续搜索中排除那些被标记为“无反应”的文档。

频繁出现的短语和发件人域可能是干扰电子邮件的表现。类别按频率列示结果中的词条，因此可用于标识这些短语。

复审查询集以减少搜索词条

除了减少干扰以外，资深律师可以与对方律师进行协商，以便创建所生成结果的数目最少并可接受的查询。通过导入案件双方都同意的查询，您可以获取一组所要处理的初始搜索结果。您可以对每个搜索进行复审，以标识结果数以及所返回内容的范围。使用所显示的类别对每个特定查询进行精简，使其返回满足发现请求并且数目最少的结果。您可以保存并导出这些新查询，以便与外部各方进行进一步协商。

查找相关短语

如果您正在查找案件中出现最为频繁的短语、电子邮件地址或名称等内容，那么请使用所显示的类别在搜索查询中包括或排除那些词条。通过搜索多组不同的短语，然后将搜索结果的范围缩小到相关的文档，您可以了解更多有关案件内容的信息。

限制所搜索的日期

通过使用时间线，可以限制所搜索的日期范围。通过检查发送电子邮件较多或较少的日期，可以帮助您洞察案件的内容。通过对不同的日期范围进行迭代，可以更深入地进行洞察。

减少无关的单词和短语

通过指定应该在搜索中忽略的文本（忽略文本），可以对无关的和重复的短语在搜索结果中的出现进行限制。例如，将公共法律免责声明添加到忽略文本。搜索引擎不使用公共的停用词，因此您必须对每个案件指定任何此类单词或短语。

查看电子邮件论题

从搜索结果中，您可以查看电子邮件论题（对话）以便更好地了解所复审的电子邮件文档中出现的内容和对话。请使用结果列表中的**查看论题**和**按论题排序**选项。

相关概念

第 12 页的『解释搜索结果』

相关任务

第 3 页的第 2 章，『搜索案件』

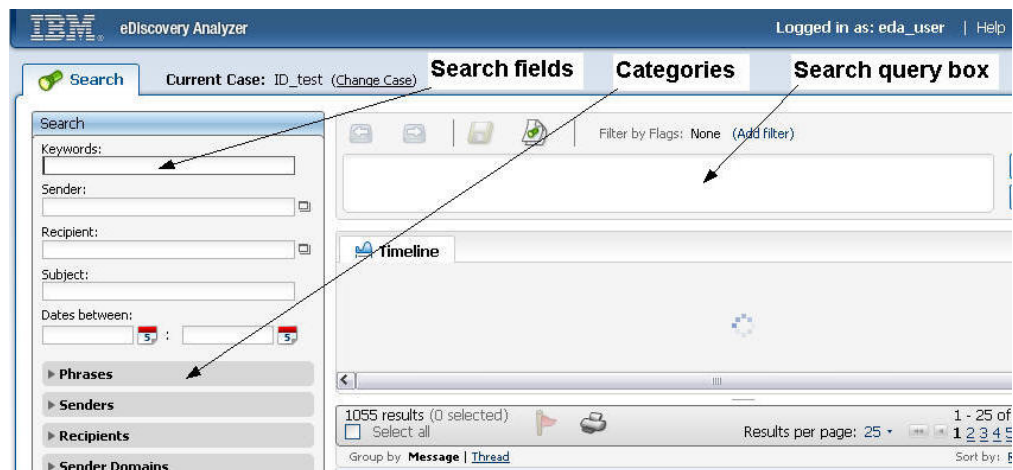
相关参考

第 4 页的『搜索技巧』

第 2 章 搜索案件

搜索案件，以便标识相关的内容、减小案件大小以及标识需要进一步复审的文档。

下图说明搜索字段、类别和搜索查询框在搜索界面中的位置。



您第一次打开案件时，将自动运行缺省的搜索查询。除非您定义另一个缺省查询，否则该查询将返回该案件中的所有电子邮件。

要搜索案件，请执行下列操作：

1. 登录到 eDiscovery Analyzer 并选择案件。
2. 在搜索字段中输入关键字、发件人、收件人、内容或日期的任意组合。您必须至少输入一个搜索词条。用于返回案件中所有文档的缺省查询是 *:*，即所有字段和所有词条。
3. 单击**搜索**。查询将显示在搜索查询框中，结果将显示在搜索页面的“结果”和“时间线”部分中。
4. 可选：要缩小搜索范围，请在搜索字段中输入其他搜索词条，选择**添加到搜索**，然后单击**搜索**。新的搜索词条将被追加到搜索查询文本框中。
5. 可选：通过在搜索查询框中编辑查询、从类别中的词条中进行选择、使用时间线或**日期范围**字段更改日期范围或者在“搜索”部分中添加或删除词条，尝试其他搜索词条。
6. 可选：从**按标志过滤**菜单中选择一个或多个标志，以使将来执行的每个查询都包括或排除带有所选标志的文档。“按标志过滤”设置随用户不同而有所变化。

相关概念

第 1 页的『搜索策略』

相关任务

第 17 页的第 3 章，『使用标志来标识文档相关性』

搜索技巧

您可以通过使用搜索查询运算符来优化搜索。

下列规则适用于所有搜索:

- 标点通常被视为单词分隔符。查询 `case-sensitive` 与精确匹配查询 `"case sensitive"` 返回相同的文档。但是, 某些带有连字符的词条被视为单个的词, 例如 `self-inflicted`。
- 搜索词条不区分大小写。查询 `IBM` 与包含 `ibm` 的文档匹配。
- 搜索词条与同一个词的其他变形匹配 (基本形式匹配)。查询 `stock option` 与 `stock options` 匹配。查询 `mouse` 与 `mice` 匹配。例如, 如果输入 `king`, 那么搜索引擎将返回所有包含单词 `king` 或 `kings` 的文档。如果输入查询 `king lear`, 那么搜索引擎将返回包含词条 `king` 或 `kings` 以及 `lear` 的文档。

要执行简单关键字搜索, 可以在**关键字**字段中输入一个或多个查询词条 (关键字)。搜索引擎将返回主题、正文或附件包含所有那些关键字或其变体的文档。缺省布尔运算符是 `AND`。

要显示更精确的结果, 请使用更具体的关键字。例如, 使用 `tax avoidance strategies` 来代替 `taxes`。或者, 使用 `California energy usage` 来代替 `energy`。

如果简单关键字搜索所返回的很多文档不是您要查找的文档, 那么您可以使用运算符或其他搜索词条字段来优化搜索。

排除词条

使用减号 (-) 或布尔 `NOT` 运算符来排除词条。例如, 如果要查找包含词条 `lear` 的文档, 并且不想查看包含 `edward` 的文档, 那么请输入查询 `lear -edward` 或 `lear NOT edward`。

减号 (-) 或 `NOT` 也适用于词条及其变体。例如, 查询 `-edward` 将排除包含单词 `edward's` 的文档。

精确匹配词条

如果要确保词条在结果中完全按您输入的顺序显示, 那么可使用双引号 (")。例如, 如果要查看包含短语 `mouse trap` 的文档, 并且不想与 `trap a mouse` 或 `mice escaped the trap` 之类的相关短语匹配, 那么请将该查询括在双引号中。对于精确匹配搜索而言, 不会将词条变体 (例如 `mouse` 和 `mice`) 视为匹配项。查询 `"mouse trap"` 与 `mouse traps` 和 `mice trap` 都不匹配。并且, 在短语中的任何两个词条之间都不能出现其他词条, 因此 `mouse in trap` 也不匹配。注意, 搜索仍然不区分大小写, 因此将与 `Mouse Trap` 匹配。但是, 标点将被忽略, 因此 `mouse, trap` 和 `See the mouse? Trap it` 都将匹配。

下列规则适用于精确匹配搜索:

- 词条必须完全按照查询中的顺序出现。例如, 查询 `"life sentence"` 与 `sentence to life` 不匹配。
- 词条必须以相同的形式出现; 词条变体将不匹配。查询 `"stock option"` 与 `stock options` 不匹配。

- 字段型查询如果包含多个由空格分隔的词条，那么必须通过将各个词条括在圆括号中对其进行组合。

表 1. 精确匹配搜索示例

要查找的内容	搜索查询框中的样本查询
没有相关短语的短语 mouse trap。	"mouse trap"

根据特定邻近性执行搜索

要将结果限制为只包括出现在同一个句子、同一个段落或特定单词范围中的词条，请指定那些词条，并在查询中指定 `WITHIN context`。要规定指定的词条必须按指定的顺序出现，请添加 `INORDER`。例如，查询 `WITHIN filed patents in December INORDER` 将找到与 `filed patents in December` 匹配的文本，但与 `patents filed in February` 和 `patents filed in December` 不匹配。

- `WITHIN SENTENCE` 与同一个句子中包含所指定词条的文档匹配。
- `WITHIN PARAGRAPH` 与同一个段落中包含所指定词条的文档匹配。
- `WITHIN n` 与所指定数目的单词范围中包含所指定词条的文档匹配。 n 的值必须是正数，并且必须至少与指定的词条数一样大。

邻近性搜索中指定的词条必须在任何邻近性搜索查询进行文档匹配的文档部分中出现。例如，邻近性查询不会返回主题和正文字段或者正文和附件部分都包含所指定词条的文档。

表 2. 邻近性搜索示例

要查找的内容	搜索查询框中的样本查询
在同一个段落中以任意顺序出现的 3 个短语 <code>article but</code> 、 <code>cancer dancer</code> 和 <code>elegant funeral</code> 。	("article but" "cancer dancer" "elegant funeral") WITHIN PARAGRAPH
在 5 个单词的范围中以任意顺序出现的 3 个单词 <code>and</code> 、 <code>begin</code> 和 <code>court</code> ，并且包括这些单词的变体。	(and begin court) WITHIN 5
在 5 个单词的范围中按以下顺序出现的 3 个单词 <code>and</code> 、 <code>begin</code> 和 <code>court</code> ，并且不包括这些单词的变体。	("and" "begin" "court") WITHIN 5 INORDER

使用布尔运算符执行搜索

布尔运算符 `OR` 指定返回的文档必须至少包含查询的其中一个词条。例如，查询 (`othello OR otello`) 将返回包含词条 `othello` 和/或 `otello` 的文档。布尔运算符不区分大小写。

您还可以借助圆括号对布尔运算符 `AND`、`OR` 或 `NOT` 进行组合使用。例如，查询 `cougar OR (jaguar AND NOT car)` 将返回任何包含 `cougar` 的文档以及不包含 `cougar` 但包含单词 `jaguar` 并且不包含单词 `car` 的文档。

使用圆括号进行组合。要搜索同时也是布尔运算符的单词，必须将该单词括在双引号中或者在前面加上转义字符 (`\`)。

例如:

- 查询 `subject:(hedge OR fund)` 与主题包含 hedge 或 fund 或者 hedge 或 fund 的其他形式的文档匹配。
- 要搜索 AND、OR 或 NOT 之类的保留字（含义与搜索中的运算符相同的单词），如果未将这些单词括在双引号中，那么必须对其进行转义。例如，您可以创建 `car AND \not`、`car AND "not"` 或 `"car not"` 之类的查询，但查询 `car AND NOT` 无效。

表 3. 使用布尔运算符进行的搜索的示例

要查找的内容	搜索查询框中的样本查询
主题中的单词 hedge 或 fund 或者 hedging 或 funds。	<code>subject:(hedge OR fund)</code>
单词 car 和 not	<code>car AND \not</code> 、 <code>car AND "not"</code> 或者 <code>"car not"</code>
othello 和/或 otello 的其他拼写。	<code>(othello OR otello)</code>

搜索未知或可变的字符或词条

如果您不知道完整的拼写，或者要查找词条的变体，那么可以使用通配符来帮助您查找文档。例如，查询 `czech*` 将返回包含词条 `czech`、`czechoslovakia`、`czechoslovakian` 以及其他以字符 `czech` 开头的单词（包括没有任何后续字符的单词 `czech`）的文档。

在查询开头添加通配符（例如 `*zech`）可能导致搜索引擎返回结果的时间延长。

还可以在短语搜索中使用通配符。通配符（*）与零个或多个字符匹配，但只与一个词条中的字符匹配。例如，查询 `"John * Kennedy"` 将返回包含词条 `John Fitzgerald Kennedy` 和 `John F Kennedy` 但不包含 `John Kennedy` 的文档。但是，通配符（*）不会跨单词分隔符进行匹配，因此 `"Jo*ennedy"` 不会与 `"John Kennedy"` 或者任何其他包含空格的姓名变体匹配。通配符（*）与单一词条中任意数目的字符（包括零个字符）匹配，但星号不会跨单词分隔符（两个或多个词条）进行匹配，并且必须要有至少包含一个字符的词条才能成功匹配。例如，`"Jo*dy"` 与 `"John Kennedy"` 不匹配（两个词条），`"John * Kennedy"` 与 `"John Kennedy"` 不匹配（没有中间词条）。此外，查询词条 `fin*int` 与 `fingerprint` 匹配，但与 `fine print` 不匹配。

使用问号（?）作为代表单一字符的通配符。搜索词条 `ra?or` 与 `razor` 或 `rasor` 匹配，但与 `raptor` 不匹配。查询 `"John ? Kennedy"` 将返回包含词条 `John F Kennedy` 或 `John F. Kennedy` 的文档，但不会返回包含 `John Fitzgerald Kennedy` 的文档。

表 4. 通配符搜索示例

要查找的内容	搜索查询框中的样本查询
姓名 John Fitzgerald Kennedy 或 John F Kennedy	<code>person:"John * Kennedy"</code>
姓名 John F Kennedy 或 John F. Kennedy	<code>person:"John ? Kennedy"</code>
姓名 Mikhail Gorbachev	<code>person:Mi*1 Gorbachev</code>

搜索特殊字符

通过在特殊字符前面加上转义字符（\），可以将特殊字符（例如 `" \ ()`）或布尔运算符（AND、OR 和 NOT）作为文字词条包括在搜索查询中。如果转义字符（\）后面的字符不是特殊字符，那么转义字符 \ 将被忽略。

需要转义字符的单一字符包括 Unicode 字符，例如 \u3000、\t、\n、\r、+、-、(、)、:、^、@、[、]、\、{、}、~、*、?、<、>、= 和 !。但是，在查询中指定特殊字符并将其括在双引号中时，只有下列特殊字符需要转义字符：*、?、" 和 \。

如果要搜索在搜索查询中同时也是运算符的单词或字符，例如布尔运算符、邻近性运算符或算术运算符（例如 <=、>= 和 != ），那么请在该字符或单词前插入反斜杠（\）。由多个符号组成的特殊字符序列（例如 <= 和 >=）必须对每个符号进行转义，例如 \<=。由字母字符组成的特殊字符序列（例如 SENTENCE）只需要一个转义字符，例如 \SENTENCE。

提示：对于 Lotus Notes® 电子邮件地址而言，特殊字符 / 是保留字符，在没有转义字符的情况下也可以被识别。例如，包含 "Raul Martinez San Jose" 的查询将返回 Notes® 地址 Raul Martinez/San Jose/Example。

表 5. 包含特殊字符或单词的搜索示例

要查找的内容	搜索查询框中的样本查询
小于或等于号 (<=)	\<=
单词 car 和 not	car AND \not
括在圆括号中的短语 (but not here)	\(but not here\)

按字段或类别执行搜索

可以通过在搜索字段中输入文本或者在搜索查询框中输入有效查询来执行搜索。要按字段或类别执行搜索，请输入字段名或类别名，接着输入冒号 (:)，然后输入词条。例如，要查找主题行包含短语“tax evasion”的文档，请在**主题**字段中输入 "tax evasion"，或者在搜索查询框中输入 subject:"tax evasion"。要查找 joe.morales@example.org 所发送的电子邮件，请在**发件人**字段中输入 joe.morales@example.org，或者在搜索查询框中输入 from:joe.morales@example.org。

通过在查询词条前面加上字段名或类别名前缀，可以将查询限制为只在特定字段或类别中进行匹配。如果该查询包含不在短语中的空格，那么请将查询括在圆括号中，以便对该短语的适当部分进行组合。例如，subject:(IBM Software) 仅返回电子邮件的主题包含单词 IBM 和 software 的文档。查询 subject:"IBM Software" 将返回按顺序包含 IBM Software 的文档（不包括这些单词的变体）。

查询词条 subject:IBM Software 将返回主题包含 IBM 并且主题、正文或附件包含 software 的文档。搜索查询 senderdomains:example.org 将返回发件人域名为 example.org 的所有文档，而查询 -senderdomains:example.com 将返回域名不是 example.com 的所有文档。

表 6. 类别搜索示例

要查找的内容	搜索查询框中的样本查询
短语 case sensitive	phrase:"case sensitive"
从 john.smith@example.com 发送的电子邮件	senders:"john.smith@example.com"
发送到 jill.jones@example.com 的电子邮件	recipients:"jill.jones@example.com"
从域 example.com 发送的电子邮件	senderdomains:"example.com"
发送到域 example.org 的电子邮件	recipientdomains:"domain.example.com"
对名为 John Smith 的任何提及	person:"John Smith"

表 6. 类别搜索示例 (续)

要查找的内容	搜索查询框中的样本查询
对位置 Silicon Valley 的任何提及	location:"silicon valley"
对公司 IBM 的任何提及	company:IBM
所有被标记为机密的电子邮件	flags:"机密" 标志搜索区分大小写。

搜索电子邮件发件人和收件人

要搜索电子邮件发件人或收件人，请在**发件人**或**收件人**搜索字段中输入姓名或电子邮件地址，或者从**发件人**或**收件人**类别中进行选择。这些字段中任何包含 @ 符的字符串都被认为是电子邮件地址。不完整的电子邮件地址不会与已建立索引的电子邮件中的完整地址匹配。包含通配符的地址不一定不完整。例如，`John*@example` 与 `John.Smith@example.com` 不匹配，但 `John*@example.com` 将匹配。此外，`Smith@example.com` 将不会与 `example.com` 的所有 `Smith` 匹配。也可以同时在地址的这两个部分（身份和域）中使用通配符：`John*@example*` 将找到 `John.Smith@example.com`。通过单击**发件人**或**收件人**字段旁边的图标，可以通过一个较大的文本框输入大批电子邮件地址。

也可以通过搜索查询框中的电子邮件地址字段 `from:`、`to:`、`cc:`、`bcc:`、`to:cc:bcc:` 和 `sentrepresenting:` 来输入姓名或电子邮件地址。这些字段分别只在发件人头、收件人列表、抄送列表、密送列表、收件人抄送密送列表或主体（代表身份）头中执行搜索。

允许使用通配符。通配符可以与电子邮件地址中用作单词分隔符的标点字符匹配，但通配符不会跨单词分隔符进行匹配。

搜索电子邮件地址时，将考虑其中可以出现的标点和特殊字符。电子邮件地址不区分大小写。单词可以与电子邮件地址或者发件人/收件人的姓名匹配。在电子邮件地址搜索中，不支持词条的备用形式。

例如，`from:"John Doe"` 与来自 `john.doe@example.com` 的电子邮件不匹配，但与来自 `John Doe <jdoe@example.com>` 的电子邮件匹配。关键字或短语搜索只在单个姓名或地址中进行匹配。例如，查询词条 `to:(craig smith)` 将与发送给 `Craig Miller` 和 `John Smith` 的电子邮件匹配。如果查询包含 @ 符，那么该查询将被解释为要求对电子邮件地址进行精确匹配。例如，查询 `from:smith@example.com` 只与电子邮件地址 `smith@example.com` 匹配，而与来自 `john.smith@example.com` 的电子邮件不匹配。

表 7. 电子邮件地址搜索示例

要查找的内容	搜索查询框中的样本查询
从 John Doe 发送的电子邮件	John Doe <jdoe@example.com>
发送给 Craig Miller 和 John Smith 的电子邮件	to:(craig smith)
从 smith@example.com 发送的电子邮件	from:smith@example.com
John Smith 发送的电子邮件	from:"john smith",from:jsm?th@example.com,
发送给 John Smith 的电子邮件	to:john.sm*th@example.com
密送给 Jill Jones 的电子邮件	bcc:jill.jo*es@example.net

表 7. 电子邮件地址搜索示例 (续)

要查找的内容	搜索查询框中的样本查询
抄送给 John Smith 或 Jill Jones 的电子邮件	cc:"john smith" OR cc:"jill jones"
Jill Jones 代表另一发件人发送的电子邮件	sentrepresenting:"Jill Jones"

搜索日期或日期间隔

在搜索查询中，可以使用保留字 docdate 将搜索限制到某个日期范围内发送的文档。搜索查询 docdate < "2007-12-18" 只返回 2007 年 12 月 18 日之前发送的文档。搜索查询 docdate > "2007-12-18" 只返回 2007 年 12 月 18 日之后发送的文档。搜索查询 docdate >= "2006-05-16" AND docdate <= "2007-12-18" 只返回从 2006 年 5 月 16 日到 2007 年 12 月 18 日发送的文档。

您还可以使用保留字 postedtime (电子邮件的发布日期) 或 deliveredtime (电子邮件的递送日期和时间) 来代替 docdate (电子邮件的发送日期)。例如，postedtime="2002-01-02 15:57" 将搜索 2002 年 1 月 2 日下午 3:57 发布的电子邮件，deliveredtime>="2002-01-02 15:56" AND deliveredtime<"2002-01-02 15:58" 将搜索 2002 年 1 月 2 日下午 3:56 到 3:58 之间递送的电子邮件。

在字段型搜索中，请以 ISO 8601 标准来指定日期: YYYY-MM-DD 或 YYYY-MM-DD HH:MM。

可以使用 >=、<=、=、> 或 < 运算符。在 docdate、postedtime 或 deliveredtime 后面，冒号 (:) 不是必需的。这些是可以在查询表达式中使用的保留字，而不是字段标签。

表 8. 日期和时间搜索示例

要查找的内容	搜索查询框中的样本查询
在 2007 年 12 月 18 日之前发送的文档	docdate < "2007-12-18"
在 2007 年 12 月 18 日之后发送的文档	docdate > "2007-12-18"
从 2006 年 5 月 16 日到 2007 年 12 月 18 日发送的文档	docdate >= "2006-05-16" AND docdate <= "2007-12-18"
在 2002 年 1 月 2 日下午 3:56 到 3:58 之间递送的电子邮件	deliveredtime>="2002-01-02 15:56" AND deliveredtime<"2002-01-02 15:58"

搜索查询框使用技巧

您可以在搜索查询框中输入或编辑使用特定字段名和保留字的搜索查询。

搜索字段

表 9. 关键字和主题语法

搜索字段	搜索查询框语法	所搜索的电子邮件部分	在搜索查询框中输入的内容示例
关键字	无	主题、正文和附件	stock option
主题	subject:	仅主题	subject:401k

- 如果未将搜索词条括在双引号中，那么支持备用形式，除非备用形式不受支持。词条的顺序没有意义。区分大小写。如果将特殊字符和保留字用作查询词条，那么必须对其进行转义。字段标签型查询如果包含多个由空格分隔的词条，那么必须通过将各个词条括在圆括号中对其进行组合。
- 如果将搜索词条括在双引号中（“短语”语法），那么将禁用备用形式。词条的顺序有意义，并且查询中的连续词条必须在文档中的连续位置出现才能匹配。不区分大小写。只有几个特殊字符必须进行转义才能用作查询词条（请参阅详细的查询语法页面）。在字段标签型查询中，多个由空格分隔的词条通过双引号进行组合。
- 在**主题**字段中，允许使用通配符。通配符不会与用作单词分隔符的标点字符匹配，也不会跨单词分隔符进行匹配。请参阅有关通配符的详细语法规则。

表 10. 发件人和收件人语法

搜索框字段	搜索查询框语法	所搜索的电子邮 件部分	在搜索查询框中输入的内容示例
发件人	from:	发件人和主体	from:"john smith" from:jsm?th@example.com from:jjones@example.com
收件人	toccbcc:	收件人列表、抄 送列表和密送列 表	toccbcc:"jill jones" toccbcc:jjones@example.com

- 允许使用通配符。通配符不会与用作单词分隔符的标点字符匹配，也不会跨单词分隔符进行匹配。请参阅有关通配符的详细语法规则。
- 不支持备用形式（不进行词形还原）。
- 这些字段中任何包含 @ 符的字符串都被认为是电子邮件地址。不完整的电子邮件地址不会与已建立索引的电子邮件中的完整地址匹配。包含通配符的地址不一定不完整。例如，John*@example 与 John.Smith@example.com 不匹配，但 John*@example.com 将匹配。此外，Smith@example.com 将不会与 example.com 的所有 Smith 匹配。也可以同时在地址的这两个部分（身份和域）中使用通配符：John*@example* 将找到 John Smith@example.com。
- 这些注意事项也适用于您在单击搜索字段右边的图标时显示的对话框中输入的词条。

表 11. 日期字段语法

搜索字段	搜索查询框语法	所搜索的电子邮 件部分	在搜索查询框中输入的内容示例
日期范围	docdate	发送日期	(docdate>="2008-08-12" AND docdate<="2008-08-13")

在 docdate 后面，冒号 (:) 不是必需的。docdate 是在查询表达式中使用的保留字，不是字段标签。不支持备用形式。

类别

表 12. 类别

搜索字段	搜索查询框语法	所搜索的电子邮件部分	在搜索查询框中输入的内容示例
短语	phrase:	主题、正文和附件	phrase:"case sensitive"
发件人	senders:	发件人和主体	senders:"john.smith@example.com"
收件人	recipients:	收件人列表、抄送列表和密送列表	recipients:"jill.jones@example.com"
发件人域	senderdomains:	发件人和主体	senderdomains:"example.com"
收件人域	recipientdomains:	收件人列表、抄送列表和密送列表	recipientdomains:"domain.example.com"
人员	person:	主题、正文和附件	person:"John Smith"
公司	company:	主题、正文和附件	company:IBM
位置	location:	主题、正文和附件	location:"silicon valley"

不支持词条的备用形式。

表 13. 标志

搜索字段	搜索查询框语法	所搜索的电子邮件部分	在搜索查询框中输入的内容示例
标志	flags:	标志表	flags:"机密"

标志名区分大小写。不支持备用形式。

搜索查询框

表 14. 发件人和收件人

搜索字段	搜索查询框语法	所搜索的电子邮件部分	在搜索查询框中输入的内容示例
无	to:	仅收件人列表	to:john.sm*th@example.com
无	bcc:	仅密送列表	bcc:jill.jo*es@example.net
无	cc:	仅抄送列表	cc:"john smith" OR cc:"jill jones"
无	sentrepresenting:	仅主体	sentrepresenting:"Jill Jones"

- 如果未将搜索词条括在双引号中，那么支持备用形式，除非“注 5”适用。词条的顺序没有意义。不区分大小写。如果将特殊字符和保留字用作查询词条，那么必须对其进行转义。字段标签型查询如果包含多个由空格分隔的词条，那么必须通过将各个词条括在圆括号中对其进行组合。
- 如果将搜索词条括在双引号中（“短语”语法），那么将禁用备用形式。词条的顺序有意义，并且查询中的连续词条必须在文档中的连续位置出现才能匹配。不区分大小写。只有几个特殊字符必须进行转义才能用作查询词条（请参阅详细的查询语法页面）。在字段标签型查询中，多个由空格分隔的词条通过双引号进行组合。
- 允许使用通配符。通配符不会与用作单词分隔符的标点字符匹配，也不会跨单词分隔符进行匹配。请参阅有关通配符的详细语法规则。
- 不支持备用形式。

表 15. 日期和时间保留字

搜索字段	保留字	所搜索的电子邮件部分	在搜索查询框中输入的内容示例
无	postedtime	电子邮件的发布日期和时间	postedtime="2002-01-02 15:57"
无	deliveredtime	电子邮件的递送日期和时间	deliveredtime>="2002-01-02 15:56" AND deliveredtime<"2002-01-02 15:58"

在 `postedtime` 或 `deliveredtime` 后面，冒号 (:) 不是必需的。这些是查询表达式中使用的保留字，而不是字段标签。

解释搜索结果

使用搜索结果来了解更多有关案件内容的信息，然后缩小结果的范围。

搜索页面显示了下列内容：

- 搜索查询
- 搜索结果的时间线
- 搜索结果列表
- 在搜索结果中找到的类别的内容

搜索查询框

完整的查询将显示在可编辑的搜索查询框中。您可以保存搜索以便重复使用，可以在连续的查询之间前后移动，也可以编辑查询并运行新查询。

如果您熟悉搜索查询语法，那么可以直接在文本框中输入查询。

时间线

时间线以图形方式显示搜索结果中的文档数并按文档的发送日期、月份或年份进行排序。您可以通过移动光标悬停在时间线上时显示的滑块来调整日期范围。单击**更新结果**可以将搜索结果中的日期范围更改为新范围。

使用滑块来调整日期范围之后，可以在新指定的日期范围内单击以进行放大。单击时间条形图中的一个垂直条形时，将仅显示该时间段的结果。

时间线同时显示所显示的每个日期范围的实际文档数和预期文档数。预期文档数的计算假定结果均匀分布于整个案件的活动日期范围内。如果实际文档数高于预期文档数，那么可能表明此时间段内的活动量较大，因此更应引起复审者注意。例如，如果搜索短语性骚扰，并且在案件的 500 万封电子邮件中共有 5% 与该短语匹配，那么您可以预期任何所选时间段内都有 5% 的电子邮件与该短语匹配。但是，如果 2007 年 1 月发送的电子邮件有 20% 与该短语匹配，那么您可以复审该时间段的电子邮件，以便讨论性骚扰事件。

类别

搜索引擎将评估文档内容并根据文本分析规则对单词和短语进行分类。搜索结果中的单词和短语将显示在类别中。每个类别中的单词和短语都按频率排序。您可以在搜索查询中添加或排除选择的单词或短语。

在您单击**更多**时显示的每个类别的扩展列表中，小型的条形图将显示每个类别中的单词和短语与当前搜索结果中的电子邮件数之间的相关性。数值较大表明该单词或短语在许多结果中出现。数值较小表明该单词或短语在少数结果中出现。

相关概念

第 1 页的『搜索策略』

相关任务

『使用类别来缩小搜索结果的范围』

使用类别来缩小搜索结果的范围

您可以使用搜索类别来缩小搜索结果的范围。

要使用类别来缩小搜索结果的范围，请执行下列操作：

1. 在“搜索”框中，展开一个或多个类别。
2. 修改查询：

目标:	执行的操作:
将所选词条立即添加到当前搜索查询	单击词条。 将运行一个新搜索，该搜索使用布尔 AND 运算符来包括所选词条，从而缩小结果的范围。
在当前搜索查询中包括一个或多个词条	单击所选词条的绿色加号 (+)，然后单击搜索。 当您使用类别来执行搜索时，将自动选中 添加到搜索 。要在仅使用所选词条的情况下运行新搜索，请选择 新建搜索 并单击 搜索 。
从当前搜索查询中排除一个或多个词条	单击所选词条的红色减号 (-)，然后单击搜索。 当您使用类别来执行搜索时，将自动选中 添加到搜索 。要在仅使用所选词条的情况下运行新搜索，请选择 新建搜索 并单击 搜索 。

- 在同一个类别（例如**短语**）中选择的带有同一符号的词条将通过布尔 OR 运算符包括在搜索查询中。
- 在同一个类别中选择的带有红色减号（-）的词条将通过布尔 AND NOT 运算符包括在搜索查询中。
- 在不同类别（例如**短语**和**发件人**）中选择的词条将通过布尔 AND 运算符包括在搜索查询中。

例如，如果在同一个类别中选择多个绿色加号（+），那么将生成类似于 `phrase: ("legal" OR "illegal")` 的查询。如果在同一个类别中选择多个红色减号（-），那么将生成类似于 `phrase:(NOT ("acceptable" OR "legal"))` 的查询。

查询 `country` 在**短语**类别中返回 Mexico、Italy、Tanzania 和 India 之类的结果，并在**公司**类别中返回 IBM、Tata 和 Toyota 之类的结果。如果单击 India 旁边的绿色加号（+）、Italy 旁边的红色减号（-）以及 Tata 旁边的绿色加号（+），那么单击**搜索**时，生成的搜索查询将是 `(country AND (phrase: (India AND NOT Italy))) AND (company: Tata)`。

相关概念

第 12 页的『解释搜索结果』

保存搜索

您可以保存已运行的搜索查询、编辑已保存的搜索、打开并运行已保存的搜索以及导入搜索。

要保存搜索，该搜索查询必须显示在搜索查询框中。如果搜索查询框未显示任何查询，那么请通过在搜索字段中输入词条来运行搜索。

已保存的搜索与用户和案件相关联。不存在系统范围的已保存搜索。要与另一个用户共享已保存的搜索，请复制该搜索查询并将其作为文本发送给该用户。

您可以选择一个已保存的搜索，以便在您每次对案件打开搜索页面时将其作为缺省搜索运行。

可以对已保存的搜索进行分组，以便将相关的搜索放在一起。

要保存搜索查询，请执行下列操作：

1. 在“搜索”页面中，输入或运行搜索查询。
2. 单击**保存**。
3. 指定或选择已保存搜索的一个组，以便将该搜索查询保存到其中。
4. 单击**保存**。

管理已保存的搜索

您可以运行、编辑或删除已保存的搜索。您还可以指定一个已保存的搜索作为缺省搜索查询。

要管理已保存的搜索，请执行下列操作：

1. 在“搜索”页面中，单击**已保存的搜索**。

2. 选择操作:

目标:	执行的操作:
运行已保存的搜索	选择已保存的搜索并单击 运行 ，或者双击已保存的搜索。
编辑已保存的搜索	选择已保存的搜索，编辑组或查询，然后单击 保存 。
删除搜索	对于要删除的已保存搜索，单击 删除 。

3. 可选: 选中**缺省搜索**以便将所选搜索查询指定为缺省搜索，您每次对案件打开搜索页面或者在搜索界面中单击**复位**时，都将自动运行该搜索。

导出已保存的搜索

您可以导出已保存的搜索，以便与其他复审者共享该搜索或者将其用于另一个案件。

可以从文本文件中导入搜索查询，并可以将已保存的搜索导出到文本文件。在用于存储已保存的搜索的文本文件中，格式为每行一个查询。

要导出已保存的搜索，请执行下列操作:

1. 在“搜索”页面中，单击**已保存的搜索**。
2. 选择要导出的已保存搜索组。
3. 单击**导出**。
4. 指定已保存搜索的文件名以及位置。

导入搜索

您可以导入由另一位复审者或者从另一个案件中导出的已保存搜索。

可以从文本文件中导入搜索查询，并可以将已保存的搜索导出到文本文件。此文本文件的格式为每行一个搜索查询，并使用受支持的查询语法。

要导入搜索，请执行下列操作:

1. 在“搜索”页面中，单击**已保存的搜索**。
2. 单击**导入已保存的搜索**。
3. 指定或者浏览到包含已保存的搜索的文本文件。
4. 指定或选择一个组，以便将该搜索查询保存到其中。如果未指定组，那么该搜索将保存在**未分组**组中。
5. 单击**导入**。

第 3 章 使用标志来标识文档相关性

设置标志，以便根据与每个案件的发现请求要求的相关性来标识所选文档。当案件管理员将案件传递回到归档时，可以按标志对文档进行分组。

标志是与一个或多个文档相关联的元数据，但不是文档的组成部分。

例如，如果您认为某个电子邮件文档对发现请求可能有反应，那么可以将其标记为“有反应”。复审案件中的所有电子邮件文档并标记部分文档之后，可以搜索所有已被标记为“有反应”的文档并对他们进行分组，以供律师复审。您也可以标记明显无关的电子邮件文档，然后在后续搜索期间排除那些文档。

您对文档设置的标志最终确定发送哪些文档以便进一步复审。可以将所标记的文档存储在归档中的新文件夹中，这一点由管理员确定。

产品附带提供了一组缺省的常用标志。管理员可以根据业务需求对缺省标志集添加或删除标志，并且可以对案件添加标志。此外，管理员还可以对现有标志进行编辑。标志可以由案件管理员定制，因此，您应该与管理员一起确定将哪些标志用于特定案件。

缺省标志包括：

- 有反应
- 无反应
- 有特权
- 热
- 从不生成
- 机密
- 第二级复审
- 外语

执行初始搜索后，对结果列表中文档设置的标志将显示在“标志”类别中，您可以通过选择这些标志来进一步缩小搜索结果的范围。您还可以选择特定的标志，以便在您运行的所有搜索中包括或排除这些标志。

要设置标志，请执行下列操作：

1. 在“结果”列表中，选择一个或多个文档。每次可以选择一个或多个文档，也可以选中**全部选中**复选框。
2. 单击**标志**。
3. 选择一个或多个标志，然后单击**保存**。如果已对所选电子邮件设置那些标志，那么您可以修改所作的选择。通过单击文档的 ***n*** 个**标志**链接，还可以查看或设置各个文档的标志。

相关任务

第 3 页的第 2 章，『搜索案件』

声明

本信息是为在美国提供的产品和服务编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

有关双字节（DBCS）信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区： International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 允许对已经交换的信息进行相互使用，请与下列地址联系：

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

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

所有关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，纯属巧合。

版权许可：

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口（API）进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。

凡这些样本程序的每份拷贝或其任何部分或任何演绎作品，都必须包括如下版权声明：©（贵公司的名称）（年）。此部分代码是根据 IBM 公司的样本程序衍生出来的。
© Copyright IBM Corp. 2004, 2006. All rights reserved.

本产品包含：

- Copyright © Outside In® Viewer Technology 1992-2005 Stellant Chicago, Inc. All rights reserved.

商标

下列各项是 International Business Machines Corporation 在美国和/或其他国家或地区的商标：<http://www.ibm.com/legal/copytrade.shtml>

Microsoft、Windows 和 Windows NT 是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

Java 和所有基于 Java 的商标是 Sun Microsystems, Inc. 在美国和/或其他国家或地区的商标。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

其他公司、产品和服务名称可能是其他公司的商标或服务标记。

词汇表

使用词汇表以帮助您了解搜索系统概念。

安全套接字层 (Secure Sockets Layer, SSL)

提供通信隐私的安全协议。通过使用 SSL, 客户机/服务器应用程序可采用能够防止窃听、篡改和消息伪造的方式进行通信。

变音符号 (Diacritic)

指示字符或字符组合的语音值更改的标记。

标记 (Token)

由企业搜索建立索引的基本文本单元。标记可以是语言中的词语或其他适合建立索引的文本单元。

标记化 (Tokenization)

将输入解析成标记的过程。

参数搜索 (Parametric Search)

一种搜索类型, 它在指定范围内查找包含数字值或属性 (例如, 日期、整型或其他数字数据类型) 的对象。

查询日志 (Query Log)

记录查询活动。可查看查询日志以了解哪些用户在进行搜索、查询所花时间及其他信息。

词法亲缘关系 (Lexical Affinity)

文档中含义接近的搜索词之间的关系。词法亲缘关系用于计算结果的相关性。

词干 (Stemming)

请参阅单词衍生形态查询 (Word Stemming)。

词条 (Lemma)

单词的基本形式。词条在字形高度变化的语言 (如捷克语) 中十分重要。

词形还原 (Lemmatization)

标识单词的根本形式和不同语法形式的过程。例如, 搜索 mouse 还会找到包含单词 mice 的文档, 而搜索 go 还会找到包含 going、gone 或 went 的文档。

单词衍生形态查询 (Word Stemming)

将单词的变体形式缩减为一般形式的语言规

范化过程。例如, 诸如 *connections*、*connective* 和 *connected* 之类的单词缩减为 *connect*。

访问控制表 (Access Control List, ACL)

在计算机安全性中, 它指的是与某个对象关联的列表, 该列表标识可以访问该对象的所有人员以及他 (她) 们的访问权。

分段 (Segmentation)

将文本划分成不同的词汇单元。非基于字典的处理包括空格和 n-gram 分段, 而基于字典的支持包括单词、语句、段分段和词形还原。

附着词 (Clitic)

具有单独语法功能但在语音上与另一个词相关的词。可以将附着词写成与它附着的词相连或分开。附着词的常见示例包括英语中缩约词的最后一部分, 如 *wouldn't* 或 *you're*。

关键字匹配排名 (Keyword Match Ranking)

对于每个结果, 搜索引擎会确定结果与查询的匹配度。文档中出现查询项的次数越多, 出现关键字的位置相互越近, 结果在结果列表中显示的位置越靠前。关键字匹配是返回搜索结果的最重要排名因子。

规范化 (Normalization)

请参阅字符规范化 (Character Normalization)。

规范化程序 (Normalizer)

一种字符规范化程序, 它扫描文本并将字符的变体形式 (如大写字母和变音符号) 缩减为一般形式。

忽略文本 (Ignore text)

搜索查询将忽略的用户定义文本, 例如样板免责声明。与停用词类似。另请参阅停用词 (stop word)。

记号化器 (Tokenizer)

一个文本分段程序, 它扫描文本并确定是否以及何时可将一系列字符识别为标记。

基于 HTML 表单的认证 (HTML Form-based Authentication)

一种直接在 HTML 页面中使用表单提供用户凭证的认证方法。

基于 Unicode 的空格分段 (Unicode-based White Space Segmentation)

一种标记化方法，它使用 Unicode 字符属性来区分标记字符和分隔字符。

解析器 (Parser)

用于解释将添加至企业搜索数据存储器的文档的程序。解析器从文档中抽取信息并准备文档以建立索引、进行搜索和检索。

密钥环 (Key Ring)

在计算机安全性中，它指的是包含公用密钥、专用密钥、可信根和证书的文件。另请参阅密钥库文件 (Keystore File)。

密钥库文件 (Keystore File)

同时包含作为签署者证书存储的公用密钥和存储在个人证书中的专用密钥的密钥环。

目录深度 (Directory Depth)

文件系统路径中子目录的长度或数目。例如，目录 C:\My Company\personnel\private\records 比 C:\My Company\personnel 要深 (包含更多子目录)。另请参阅 URL 深度 (URL Depth)。

排名 (Ranking)

为查询产生的搜索结果中的每个文档指定一个整数值。搜索结果中文档的顺序取决于与查询的相关性。排名越高，表明匹配程度越好。

凭证 (Credential)

认证过程中获取的详细信息，它们描述用户、任何组关联以及其他与安全性相关的标识属性。凭证可用于执行许多服务，如授权、审计和委派。例如，用户的登录信息 (用户标识和密码) 就是允许用户访问帐户的凭证。

起始 Web 地址 (Starting Web address)

Web 搜寻的起始位置。如果输入 www.example.org 之类的起始 Web 地址，那么搜寻器将通过访问起始页面中的 HTTP 链接来搜寻该站点上的所有 Web 页面 (文档)。

起始目录 (Starting Directory)

文件系统搜寻的起始位置。如果输入 C:\mydocuments 之类的起始目录，那么搜寻器会搜寻该目录及子目录中的所有文档。

轻量级目录访问协议 (Lightweight Directory Access Protocol, LDAP)

一种开放式协议，它使用 TCP/IP 允许访问支持 X.500 模型的目录并且不会产生更复杂的 X.500 目录访问协议 (DAP) 的资源需求。例如，可使用 LDAP 在因特网或内部网目录中查找人员、组织和其他资源。

认证 (Authentication)

验证用户或服务器身份的过程。

认证中心 (Certificate Authority)

一个可信的第三方组织或公司，它颁发用于创建数字签名和公用/专用密钥对的数字证书。认证中心保证授予了唯一证书的人员的身份。

软错误页面 (Soft Error Page)

一种类型的 Web 页面，它提供关于无法返回请求的 Web 页面的原因的信息。例如，HTTP Server 可以返回详细解释状态码的页面，而不是返回简单的状态码。

数据源 (Data Source)

可从其中检索文档的任何数据存储库，如 Web、文件系统或数据库。

搜索高速缓存 (Search Cache)

保存先前搜索请求的数据和结果的缓冲区。

搜索结果 (Search Result)

与搜索请求相匹配的文档列表。

搜索页面 (Search Page)

一个 Web 页面，用于接受用户查询和显示搜索结果列表。

搜索引擎 (Search Engine)

接受搜索请求并返回一系列文档给用户的程序。

搜寻 (Crawling)

搜寻是指从内容存储库和文件系统等数据源中检索文档的软件程序活动。然后，搜索引擎将处理检索到的文档，进行分析，对其建立索引，并且使这些文档可供用户搜索。另请参阅搜寻器 (Crawler) 和搜寻空间 (Crawl Space)。

搜寻空间 (Crawl Space)

与指定模式 (例如统一资源定位符 (URL)、数据库名称、内容存储库、文件系统路径、域名和 IP 地址) 相匹配的一组源, 搜寻器通过读取这些源来检索要建立索引的项。

搜寻器 (Crawler)

一种软件程序, 它可以从数据源或内容存储库中检索文档并收集可用于创建搜索索引的信息。

索引 (Index)

一种数据结构, 它引用数据项来启用搜索以查找包含查询项的文档。

特色链接 (Featured Link)

特色链接是伴有标题和描述的连接, 它们可配置为在每次用户输入特定查询时出现在搜索页面的顶部。特色链接分为四个部分: 查询、Web 地址、标题和可选摘要。

停用词 (Stop Word)

被搜索应用程序忽略的常用词, 例如, *the*、*an* 或 *and*。

停用词删除 (Stop Word Removal)

从查询中除去停用词以忽略常见词语并返回更多相关结果的过程。

通配符 (Wildcard Character)

用于表示搜索项的前面、中间或后面的可选字符的字符。

同义词 (Synonym)

具有相同含义的不同词语。可通过定义同义词来提高用户找到正确文档的机率, 搜索引擎会使用它们以便在查询中出现同义词之一时扩大搜索目标的范围。

统一资源标识 (Uniform Resource Identifier, URI)

标识抽象或物理资源的字符的压缩字符串。

统一资源定位符 (Uniform Resource Locator, URL)

可在网络 (如因特网) 中访问的信息资源的唯一地址。URL 包括用于访问信息资源的协议的缩写名以及协议用来查找信息资源的信息。

文本分段 (Text Segmentation)

请参阅分段 (Segmentation)。

文件系统搜寻器 (File system crawler)

一种从目录中检索文档的搜寻器。

系统日志 (System Log)

记录搜索引擎运行时的计算机活动。可查看系统日志以检查企业搜索系统上是否存在问题。

修改日期排名 (Modification Date Ranking)

日期越新的文档排名越高。

按修改日期排名对于许多类型的文档特别重要。但是, Web 站点文档的文档日期通常不太可靠, 原因是许多 Web 服务器返回的最近一次修改日期并未反映文档的真实存在时间。

掩码字符 (Masking Character)

用于表示搜索项的前面、中间和后面的可选字符的字符。掩码字符通常用于在索引中查找术语的变体。另请参阅通配符 (Wildcard Character)。

用户代理 (User Agent)

用于浏览 Web 并在所访问的站点中留下自身信息的应用程序。在企业搜索中, Web 搜寻器就是一个用户代理。

语言搜索 (Linguistic Search)

一种搜索类型, 它浏览和检索文档, 并使用缩减为其基本形式 (例如, 将 *mice* 作为 *mouse* 建立索引) 或扩展其基本形式 (类似复合词) 的词语为文档建立索引。

元数据 (Metadata)

描述特定信息块的数据, 该数据可帮助检索 (根据搜索)、浏览 (根据目录) 或过滤 (根据重点) 信息。元数据通常包含在分类法或分类方案中。

证书 (Certificate)

在计算机安全性中, 它指的是将公用密钥绑定至证书所有者标识的一个数字文档, 这样就允许对证书所有者进行认证。证书由认证中心颁发, 并由该中心进行数字签名。

注释器 (annotator)

UIMA 注释器用于分析 UIMA 框架中包含的非结构化信息。另请参阅 UIMA。

字段 (Field)

可在其中输入特定类别的数据或控制信息的区域。

字段搜索 (Fielded Search)

限于特定字段的查询。

字符规范化 (Character Normalization)

将字符的变体形式 (如大写字母和变音符) 缩减为一般形式的过程。

自由格式文本 (Free-form Text)

由单词或句子组成的非结构化文本。

自由文本搜索 (Free Text Search)

搜索项表示为自由格式文本的搜索。

GET 命令 (GET Command)

从 Web 服务器请求文件的 HTTP 命令。

HTTP 代理服务器 (HTTP Proxy Server)

充当由应用程序或 Web 服务器主管的 HTTP Web 请求的中介的服务器。代理服务器充当企业中内容服务器的代用品。

HTTP 基本认证 (HTTP Basic Authentication)

HTTP 协议中指定的标准认证方案, 旨在控制对 Web 站点的访问。HTTP 基本认证仅使用用户名和密码。

HTML 元标记 (HTML Meta Tag)

HTML 元标记可添加至 HTML 文档, 例如, 以对搜寻器给出指令。

IP 地址 (IP Address)

使用 IP 标准的网络上的设备或逻辑单元的唯一地址。

Java 数据库连接 (Java Database Connectivity, JDBC)

一种工业标准, 适用于 Java 平台与大量数据库之间与数据库无关的连接。JDBC 接口提供调用级别 API 来访问基于 SQL 的数据库。

MIME 类型 (MIME Type)

一种因特网标准, 用于标识在因特网中传送的对象类型。

no-follow 伪指令 (no-follow Directive)

Web 页面中的一个伪指令, 它指示机器人 (如 Web 搜寻器) 不要访问在该页面中找到的链接。

n-gram 分段 (n-gram Segmentation)

一种分析方法, 它将给定数目的字符的重叠序列视为单个单词, 而不是像在基于 Unicode 的空格分段中那样使用空格或标点定界单词。例如, 如果 n 为 2, 那么文本 ABEFD 会分段为序列 AB BE EF FD。

no-index 伪指令 (No-index Directive)

Web 页面中的一个伪指令, 它指示机器人 (如 Web 搜寻器) 不要将该页面的内容包括在索引中。

POST 命令 (POST Command)

将信息发送至 Web 服务器以进行处理的 HTTP 命令。POST 方法在 HTML 文件中得到广泛实现, 用于将包含输入数据的表单发送至服务器。

Robots Exclusion Protocol 协议 (Robots Exclusion Protocol)

一种协议, 它允许 Web 站点管理员指示访问机器人 (如 Web 搜寻器) 网站的哪些部分不应由机器人访问。

robots.txt 文件 (robots.txt File)

为阻止搜寻器搜寻某一服务器, 某些 Web 站点管理员在 Web 服务器上创建了一个文件以定义搜寻器的访问策略。按照 Robots Exclusion Protocol 协议, 此文件称为 robots.txt。

UIMA

即“非结构化信息管理体系结构”, 这是一种开放式源代码框架和 SDK, 用于开发从大量非结构化信息中发现相关知识的应用程序。另请参阅 Apache UIMA。

URL 深度 (URL Depth)

Web 站点地址的长度或其中的斜杠数。例如, Web 站点地址 www.example.org/personnel/private/records 比 Web 站点地址 www.example.org/personnel 更深 (斜杠更多)。

通常 URL 深度越小, 文档越重要。另请参阅目录深度 (Directory Depth)。

User-Agent 字符串 (User-Agent String)

用于标识 Web 搜寻器。其中一个用途是在 robots.txt 文件中使用, 该文件可根据特定 User-Agent 字符串拒绝访问。

Web 链接分析排名 (Web Links Analysis Ranking)

一种排名方法, 指向文档的链接越多, 该文档的排名越高。

Web 搜寻器 (Web Crawler)

一种搜寻器, 可通过检索 Web 文档并访问该文档中的链接来浏览 Web。

索引

[A]

案件 3

[B]

变量 6

标志 7

从不生成 17

第二级复审 17

机密 17

列表 17

热 17

设置 17

外语 17

无反应 17

有反应 17

有特权 17

标志类别 11

布尔运算符 5

[C]

查询 9, 12

[D]

短语 7

短语类别 11

短语搜索 4

[F]

发件人 7, 8

发件人类别 11

发件人域 7

发件人域类别 11

非字母字符 6

[G]

概述 1

公司 7

公司类别 11

构面式浏览 13

[J]

结果 12

精确匹配 4

[L]

类别 3, 7, 12

标志 13

短语 13

发件人 13

发件人域 13

公司 13

人员 13

收件人 13

收件人域 13

位置 13

邻近性 5

[R]

人员 7

人员类别 11

日期 9

[S]

设置标志 17

时间 9

时间线 12

收件人 7, 8

收件人类别 11

收件人域 7

收件人域类别 11

搜索 12

标志 17

策略 1

类别 13

已保存的搜索 14

搜索查询框 3

搜索查询语法

包括词条 4

布尔运算符 4

短语 4

排除词条 4

通配符 4

字段 4

搜索技巧 4, 9

搜索运算符

布尔运算符 4

搜索运算符 (续)

邻近性 4

通配符 4

搜索字段 3

[T]

特殊字符 6

通配符 4, 6

[W]

位置 7

位置类别 11

[Y]

已保存的搜索 14

导出 15

导入 15

管理 14

语法 9

[Z]

主题 7

转义码 6

字段 7

标志 9

短语 9

发件人 9

发件人域 9

公司 9

关键字 9

人员 9

日期 9

收件人 9

收件人域 9

位置 9

主题 9

字符, 特殊 6

A

AND 5

B

bcc: 8

C

cc: 8

D

deliveredtime 9

docdate 9

F

from: 8

I

INORDER 5

N

NOT 4, 5

O

OR 5

P

postedtime 9

T

to: 8

to: 8

W

WITHIN 5

[特别字符]

“从不生成”标志 17

“第二级复审”标志 17

“机密”标志 17

“热”标志 17

“外语”标志 17

“无反应”标志 17

“有反应”标志 17

“有特权”标志 17



程序号: 5724-V36

SC19-2707-00

