

IBM DB2 Information Integrator  
OmniFind Edition



# 发行说明

版本 8.2.2



IBM DB2 Information Integrator  
OmniFind Edition



# 发行说明

版本 8.2.2

在使用本资料及其支持的产品之前，请务必阅读『声明』中的一般信息。

本文档包含 IBM 的专利信息。它是根据许可协议提供的，并受版权法保护。本出版物包含的信息不包括任何产品保证，且本手册提供的任何声明不应作如此解释。

可以在线方式或通过您当地的 IBM 代表订购 IBM 出版物：

- 要以在线方式订购出版物，可访问“IBM 出版物中心”（IBM Publications Center），网址为 [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order)。
- 要查找您当地的 IBM 代表，可访问“IBM 全球联系人目录”（IBM Directory of Worldwide Contacts），网址为 [www.ibm.com/planetwide](http://www.ibm.com/planetwide)。

当您发送信息给 IBM 后，即授予 IBM 非专有权，IBM 可以它认为合适的任何方式使用或分发此信息，而无须对您承担任何责任。

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

---

# 目录

关于发行说明 . . . . .	1	文档更新 . . . . .	13
1 Web 上的技术支持信息 . . . . .	1	对错误消息的更改 . . . . .	13
此发行版中的新增内容 . . . . .	3	对查询语法的更新 . . . . .	13
企业搜索基本注释器程序包 . . . . .	4	折叠 URI . . . . .	13
已知局限性、问题和变通方法 . . . . .	7	为 FileNet Content Services 启用直接文档访问 . . . . .	14
文档大小限制 . . . . .	7	普通搜寻器插件 . . . . .	14
安装修订包将导致完全重新搜寻 . . . . .	7	我的概要文件页的帮助与用户界面不一致 . . . . .	15
Java 解析器注释器的堆大小限制 . . . . .	7	Web 搜寻器的帮助页中有不正确的值 . . . . .	15
处理引擎归档文件的最大大小 . . . . .	8	indexBuildItem 描述 . . . . .	15
AIX 上的 Notes 搜寻器需要 I/O 完成端口模块 . . . . .	9	映射公共分析结构到索引 . . . . .	16
查询安全 Notes 集合可能会导致不准确的结果文档计 数 . . . . .	10	从 WebSphere Portal 迁移基于模型的分 类法 . . . . .	16
用户名可包含句点 . . . . .	10	声明 . . . . .	17
字段搜索中的通配符 . . . . .	10	商标 . . . . .	18
服务器上的索引位置 . . . . .	11	与 IBM 联系 . . . . .	21
Linux 和 AIX 的孤立 URLFetcher 进程 . . . . .	11	获取产品信息 . . . . .	21
AIX 中的共享内存段 . . . . .	12	提供对文档的意见 . . . . .	21



---

## 关于发行说明

本发行说明文档描述了此发行版中的新内容，以及 IBM® DB2® Information Integrator OmniFind™ Edition V8.2.2（版本 8.2 加修订包 2）的已知局限性、问题、变通方法和文档更新。

DB2 II OmniFind Edition V8.2.2 只是一个升级。在安装 V8.2.2 之前，您必须已安装了 V8.2 或 V8.2.1。

在 SuSE Linux® Enterprise Server 9 上进行安装之前，应访问 DB2 II OmniFind Edition V8.2.2 支持 Web 站点获取变通方法和更新。

---

### 1 Web 上的技术支持信息

1 DB2 II OmniFind Edition 产品最近已重新命名为 WebSphere® Information Integrator  
1 OmniFind Edition。在与产品相关的 Web 页面上，您可能会看到提及 WebSphere  
1 Information Integrator OmniFind Edition，但产品界面、文档和故障诊断信息（技术说明）  
1 仍然使用 DB2 品牌。

1 可在以下 Web 站点中找到技术支持信息：

#### 1 **IBM DB2 Information Integrator OmniFind Edition**

1 [www.ibm.com/software/data/integration/db2ii/supportomni.html](http://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 通用数据库™ Linux 版、UNIX® 版和 Windows® 版的版本 8 产品**

1 [www.ibm.com/software/data/db2/udb/winos2unix/support](http://www.ibm.com/software/data/db2/udb/winos2unix/support)





---

## 此发行版中的新增内容

IBM DB2 Information Integrator OmniFind Edition V8.2.2 提供以下功能:

### 操作系统

V8.2.2 新增了对以下操作系统的支持:

- AIX® 5.3 (需要维护级别 1 和 2004 年 8 月的 C++ Runtime for AIX PTF)
- 带有 Service Pack 1 (UnitedLinux SP1) 的 SuSE Linux Enterprise Server 9 (需要 libstdc++-3.3.3-43.28)

在 SuSE Linux Enterprise Server 9 上进行安装之前, 应访问 DB2 II OmniFind Edition V8.2.2 支持 Web 站点获取变通方法和更新。

**大文档** 集合可包括大小最多为 32 MB 的文档

### Domino® 文档的文档级安全性

- 使用由本机存储库管理的访问控制, 在查询处理期间, 实时验证用户凭证
- 通过用户概要文件和 LDAP 用户注册表验证用户。通过以加密格式存储凭证, 用户不需要在每次查询 Domino 数据源时都指定凭证。

### 搜寻器

- Notes® 搜寻器: 除可搜寻数据库、视图和文件夹外, 还可搜寻目录
- Web 搜寻器:
  - 用户界面支持搜寻受密码保护的 Web 站点 (使用 HTTP 基本认证或基于表单的认证的站点)
  - 能够搜寻由代理服务器提供服务的 Web 站点
  - 能够返回软错误页的定制消息

### 搜索质量

- 联合搜索 (能够同时搜索多个集合)
- 支持在查询中使用通配符结尾
- 支持对搜索结果中摘要的显示长度进行定制
- 用户界面支持对来自同一 URI 的搜索结果进行折叠

### 语言分析

- 支持非结构化信息管理体系结构 (UIMA), 它提供了对集合中的文档进行定制文本分析的能力
- 支持本机 XML 搜索
- 支持定制同义词字典

### API 和搜索应用程序

- 用户界面支持配置和监视定制客户机数据侦听器应用程序
- 普通搜寻器插件使您能指定定制安全性和业务规则
- 对样本搜索应用程序提供了许多增强功能, 可演示查询函数、搜索结果选项、联合搜索和文档级安全性

---

## 企业搜索基本注释器程序包

您可以使用企业搜索基本注释器程序包根据企业搜索注释器的输出来开发新的注释器，并且在 UIMA 软件开发包（SDK）中测试企业搜索注释器和您定制的注释器。

企业搜索基本注释器程序包是一个 ZIP 文件，其中包含了在企业搜索中使用的文本分析步骤。解析文档时，这些注释器总是在任何定制分析之前运行。

UIMA 软件开发包（SDK）可从 IBM developerWorks Web 站点获得：  
<http://www.ibm.com/developerworks/db2/zones/db2ii/>。

要安装注释器程序包：

1. 在安装企业搜索（DB2 II OmniFind Edition）的 *ES\_INSTALL\_ROOT/packages/uima* 目录中找到注释器程序包 *OF\_base\_annotators.zip*。
2. 将该 ZIP 文件复制到安装 UIMA SDK 的根目录中。
3. 解压缩该 ZIP 文件，将企业搜索基本注释器文件添加到 UIMA SDK 安装的指定目录结构中。

企业搜索基本注释器程序包只包含了处理英文文档所需的字典。如果您想要在开发环境中处理其它语言，可遵循以下步骤：

1. 在安装企业搜索的 *ES\_INSTALL\_ROOT/configurations/parserservice/jediidata/frost4* 目录中找到企业搜索字典。
2. 将该目录的内容复制到安装本地 UIMA SDK 的 *UIMA\_SDK\_INSTALL/data/frost4* 目录中。

要验证 UIMA SDK 是否已成功安装：

1. 打开以下目录中的 `CAS` 可视调试器（CVD）：  
*UIMA\_SDK\_INSTALL/bin/cvd[.bat/.sh]*。
2. 单击运行 → 装入 TAE。
3. 在 *UIMA\_SDK\_INSTALL/docs/examples/descriptors/analysis\_engine* 目录中找到 TAE 说明符文件 *of\_tokenization.xml*。
4. 装入样本文档并运行 TAE。您将在 CVD 中看见类型为 *uima.tt.TokenAnnotation* 的注释。

要使用企业搜索注释器进行处理：

1. 如果您的定制注释器使用由企业搜索注释器定义的类型，则将对 *of\_typesystem.xml* 文件的引用包括在定制注释器说明符的 *typeSystem* 节中。*of\_typesystem.xml* 文件在 *UIMA\_SDK\_INSTALL/docs/examples/descriptors/analysis\_engine* 目录中。请参阅 *analysis\_engine* 目录中的 *jtok.xml* 文件获取示例，了解如何包括对文件的引用。
2. 创建一个包括文件 *of\_tokenization.xml* 和定制注释器说明符的聚集 TAE 说明符。
3. 在 UIMA SDK 中测试时，使用该聚集 TAE 说明符。

在创建 PEAR 文件以将您的注释器上载到企业搜索中时，应确保此 PEAR 文件只包括您的定制注释器说明符。它可能不包括以下任何文件：

*of\_tokenization.xml*、*of\_typesystem.xml*、*of\_base\_annotators.zip* 的内容，或在第 2 步中创建的聚集说明符。

| 要获取关于企业搜索中分析能力的更多信息，请参阅信息中心或 *Programming Guide and*  
| *API Reference for Enterprise Search* 中的“Custom Text Analysis Integration”和“Text  
| Analysis included in enterprise search”主题。



---

## 已知局限性、问题和变通方法

此信息描述 DB2 Information Integrator OmniFind Edition V8.2.2 的已知局限性、问题和变通方法。

任何局限性和限制都可能适用于也可能不适用于产品的其它发行版。

---

### 文档大小限制

可对最多 32 MB 的文档进行搜寻并建立索引。可对带有 LOB 数据的 DB2 通用数据库表进行搜寻并建立索引，但数据必须符合总文档大小为 32 MB 的限制。

请忽略搜寻器属性帮助文件中关于最大页大小可大于 32 MB 的说明。

---

### 安装修订包将导致完全重新搜寻

在安装了 DB2 II OmniFind Edition V8.2.2 之后，将重新搜寻所有文档（那些已由 Web 和 NNTP 搜寻器搜寻过的文档除外）。

在安装了 V8.2.2 后首次启动非 Web 和非 NNTP 搜寻器时，将搜寻所有数据源中的所有文档并再次对它们建立索引，而不只是针对自上次搜寻数据源以来的新文档或更改过的文档。

在搜索了集合中的每个表、目录或文件夹中的所有文档后，搜寻器将转换为正常的更新处理并开始只搜寻新文档或更改过的文档。

通常，只有在对搜寻器配置的更改需要完全重新搜寻时，才有必要进行完全重新搜寻。例如，如果数据源中字段或列的显示名发生了更改，则搜寻器必须重新搜寻所有文档，才能将新信息用于对文档建立索引。

在安装 V8.2.2 后自动进行完全重新搜寻将耗费一定的时间，但不会影响企业搜索索引的质量。

---

### Java 解析器注释器的堆大小限制

对于运行 Java 解析器注释器的 *collection\_ID.casprocessor* 会话，如果在集合错误日志中显示了“内存不足”的消息，应将堆大小增加到 200 MB。

上载到企业搜索的定制文本分析注释器是在 Java 会话中运行的。例如，在企业搜索中，这些会话是 *collection\_ID.casprocessor* 会话，其中 *collection\_ID* 是此注释器所属的集合标识。这些会话的堆大小限制为 200 MB。

要增加堆大小：

1. 停止企业搜索系统。在所有操作系统上，输入以下命令：`esadmin stop`
2. 停止 CCL：
  - 在 Linux 和 AIX 上，输入以下命令：`stopccl.sh`

- 在 Windows 上，停止 DB2 Information Integrator OmniFind Edition Windows 服务：开始 → 控制面板 → 服务。
3. 查找您的注释器所属的集合标识。要查找该集合标识，可运行命令 `esadmin check` 或转至 `ES_NODE_ROOT/master_config` 目录，在该目录中对于已创建的每个集合都有一个 `collection_ID_config.ini` 文件。
  4. 在 `collection_ID_config.ini` 文件中，增加 `collection_ID.casprocessor` 会话的 `max_heap` 值。
  5. 对 `collection_ID_config.ini` 文件排序，并查找 CAS 处理器条目。例如，在 Linux 上，对于 col7:
 

```
% cd $ES_NODE_ROOT
% sort col7_config.ini
...

session3.collectionid=col7
session3.configDir=col7.casprocessor
session3.description=Fenced CAS Processor Session
session3.displayname=Fenced CAS Processor Server
session3.domain=.
session3.flags=0
session3.id=col7.casprocessor
session3.max_heap=200

...
```
  6. 编辑 `session_number.max_heap=200` 这一行，以包括新的最大堆大小（200）。例如，对于 col7: `session3.max_heap=200`
  7. 重新启动 CCL:
    - 在 Linux 和 AIX 上，输入以下命令: `startccl.sh -bg`
    - 在 Windows 上，启动 DB2 Information Integrator OmniFind Edition Windows 服务：开始 → 控制面板 → 服务。
  8. 重新启动企业搜索。在所有操作系统上，输入以下命令: `esadmin start`
  9. 从企业搜索管理控制台启动此集合的解析器。注释器将使用新的最大堆大小在 JVM 上运行。

---

## 处理引擎归档文件的最大大小

对于定制文本分析，处理引擎归档文件（PEAR 文件）的最大大小是 8 MB。

只有小于 8 MB 的 PEAR 文件才能使用企业搜索管理控制台上载。大于 8 MB 的 PEAR 文件通常包含了如字典这类资源。要上载这类 PEAR 文件：

1. 更改注释器代码，使它通过 UIMA 资源管理器（使用 `AnnotatorContext` 的 `getResourceObject` 方法）装入资源，或使它使用指向资源的注释器配置参数的值。
2. 将资源文件复制到企业搜索系统的“索引器”节点的专用目录中。
3. 更改注释器配置（您使用的配置参数的 `<externalResource>` 指定或 `<namevaluePair>`），通过使用在第 2 步中创建的目录将指向资源的绝对路径包括在该配置中。
4. 使用更新的配置重新创建 PEAR 文件。确保它不包含此资源，并将它上载到企业搜索。

## AIX 上的 Notes 搜寻器需要 I/O 完成端口模块

要在 IBM AIX 平台上使用 Notes 搜寻器和 discovery 进程，必须从 AIX 产品 CD-ROM 中安装 I/O 完成端口 (IOCP) 模块 (bos.iocp.rte)，并将其状态设置为可用。

在安装了 bos.iocp.rte 模块之后，必须对它应用适当的修订。请访问以下链接获取指示信息：

<http://www-1.ibm.com/support/docview.wss?uid=swg21086556>

如果没有 IOCP 模块，Notes discovery 进程将在创建新搜寻器期间失败，并显示以下错误消息：

```
FFQM0105E 从服务器接收到错误 -  
消息: FFQG0024E 捕获到意外的异常: discovery
```

此外，以下包括了 ENOEXEC 错误的消息也将写入 ES\_NODE\_ROOT/logs/system\_yyyymmdd.log 文件中。（某些消息正文分为几行显示，以便阅读。）

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]  
dominoserver.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:  
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86  
FFQ00277E An exception was caught with the detail 'java.lang.UnsatisfiedLinkError:  
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:  
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a'  
and a stack trace of 'java.lang.UnsatisfiedLinkError:  
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:  
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a  
at java.lang.ClassLoader$NativeLibrary.load(Native Method)  
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)  
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)  
at java.lang.Runtime.loadLibrary0(Runtime.java:824)  
at java.lang.System.loadLibrary(System.java:908)  
at lotus.domino.NotesThread.load(NotesThread.java:306)  
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)  
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)  
at com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover  
(Unknown Source)  
at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)  
at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover  
(ComponentDiscoveryW.java:72)  
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)  
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)  
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)  
at sun.reflect.DelegatingMethodAccessorImpl.invoke  
(DelegatingMethodAccessorImpl.java:60)  
at java.lang.reflect.Method.invoke(Method.java:391)  
at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)
```

在创建 Notes 搜寻器之前应确保 IOCP 模块已正确安装：

1. 输入以下 ls1pp 命令以确保 IOCP 模块安装在搜寻器服务器上：

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

ls1pp 命令的输出与以下示例类似：

文件集	级别	状态	描述
-----			
路径: /usr/lib/objrepos			
bos.iocp.rte	5.2.0.10	已落实	I/O 完成端口 API

路径: /etc/objrepos  
bos.iocp.rte 5.2.0.10 已落实 I/O 完成端口 API

2. 输入以下 `lsdev` 命令以确保 IOCP 端口是**可用**的:

```
$ lsdev -Cc iocp
```

`lsdev` 命令的输出应符合以下示例:

```
iocp0 可用 I/O 完成端口
```

3. 如果 IOCP 端口的状态是**已定义**, 请执行以下步骤将状态更改为**可用**:

a. 作为 `root` 用户登录到搜寻器服务器, 并发出以下命令:

```
# smit iocp
```

b. 选择**更改 / 显示 I/O 完成端口的特征**, 并将**要在系统重新启动时配置的状态**从**已定义**更改为**可用**。

c. 重新启动搜寻器服务器。

d. 再次输入 `lsdev` 命令, 并确认 IOCP 端口的状态已更改为**可用**。

---

## 查询安全 Notes 集合可能会导致不准确的结果文档计数

结果集中结果文档的数目对于搜索安全集合的查询可能不准确, 在这些集合中已选择了在查询处理期间验证当前凭证的选项。

在搜索企业搜索集合时, 结果文档的总数显示在搜索结果中。如果满足以下所有条件, 此数目可能会大于结果文档的实际数目:

- 在创建集合时, 已对该集合启用了安全性。
- 集合包括了 Notes 搜寻器。
- 在对 Notes 搜寻器配置文档级安全性选项时, 指定了在查询处理期间验证当前凭证的选项。

此问题将影响 `SI-API ResultSet.getAvailableNumberOfResults` 值和 `ESSearchApplication` 样本搜索应用程序。

---

## 用户名可包含句点

关于添加企业搜索管理员名称的联机指示信息不正确。

在使用企业搜索管理控制台将企业搜索管理员标识添加到系统中时, 该页中的指示信息指示用户名不能包含句点。该指示信息不正确。使用带有句点的用户名 (例如, `j_user@this.company.com`) 是允许的。

---

## 字段搜索中的通配符

使用 `field:pref*` 作为查询项通常是允许的。DB2 II OmniFind Edition 将检索的文档中包括了以 `pref` 开头且标记为 `field` 字段成员的单词。此规则有个例外: 对于少数通配符查询相关程度较低的特殊字段, 查询项 `field:pref*` 将作为 `field:pref` 来处理。那些特殊字段有:

- `site`



- link
- docid
- samegroupas
- url
- group (newsgroup, 与 *site* 类似, 但用于 NNTP 中)

例如, 查询 *url:porta\** 可能不会返回任何结果, 而 *url:portal* 将会返回结果。

---

## 服务器上的索引位置

如果您创建一个集合, 并且将索引位置指定为其它位置, 而不是服务器数据根目录中的缺省位置, 则此非缺省索引位置的磁盘空间的使用将不会受到监视。

对于多服务器安装, 当创建集合时, 如果想要指定不是缺省位置的索引位置, 该索引位置必须首先在四个服务器上已经存在。

---

## Linux 和 AIX 的孤立 URLFetcher 进程

**问题:** Web 搜寻器使用两个进程。当一个进程退出时, 它会发信号让另一个进程也退出。有时, 一个进程可能会意外退出, 使另一个进程无限期地运行下去。

如果从 UNIX `ps` 命令的输出中看到类似以下进程的孤立进程, 就可以确定发生了这种情况:

```
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:...
```

注意:

- PPID (父进程标识) 值为 1 表示该进程是孤立的。
- 类路径中存在 `URLFetcher.jar` 表示这是 Web 搜寻器进程。

### 解决方案

找到该孤进程的进程标识 (PID), 并杀死它。在以上示例中, 该孤进程是 26222。在 Linux 和 AIX 上, 使用不带自变量的 `kill` 命令。您可能还会在进程表中看到“子代”条目。子代条目表示由孤立进程创建的线程。当您杀死这些子代条目的祖代进程时, 这些子代条目应改被删除。

当您看到孤立 URLFetcher 进程时, 应该杀死该进程, 原因是如果它位于进程表中, 操作系统就会为它保留资源。您应该定期检查是否存在这些孤立进程, 尤其是当您怀疑 Web 搜寻器异常终止时。如果存在这些进程, 则杀死它们。

**重要:** 在 DB2 II OmniFind Edition V8.2.1 中, 孤进程很少出现, 因为现在 URL 访问器注意到 C++ 组件消失时会自行终止。如果真的出现此问题, 请注意不要由于疏忽而杀死另一用户的 `httpauth.ini` 编辑会话。HTTPAuthEdit 实用程序是一个从 URLFetcher JAR 文件运行的 JVM, 在进程表中可能看起来很类似。

---

## AIX 中的共享内存段

如果在 AIX 上看到消息 SQL1224N，则表示您可能已超过了 AIX 允许的共享内存段数目。

要解决此问题，请参阅 IBM 支持 Web 站点上的主题：[SQL1224N Shared Memory Segments](#)。

---

## 文档更新

文档更新包括对 DB2 Information Integrator OmniFind Edition V8.2.2 文档的内容更正或少量添加。

---

### 对错误消息的更改

在文档送交翻译之后，又更新了错误消息。

#### **FFQD1034E**

错误：未定义环境变量 CMBROOT / IBMCMROOT。请验证是否正确安装了 DB2 Information Integrator for Content。

解释：未定义环境变量 CMBROOT / IBMCMROOT。可能未安装 DB2 Information Integrator for Content。

#### 用户响应：

验证是否正确安装了 DB2 Information Integrator for Content。

#### **FFQD1035E**

错误：未找到 cmbicmsrvs.ini 文件。请验证是否正确安装了 DB2 Information Integrator for Content。

解释：在指定的路径中未找到 cmbicmsrvs.ini 文件。可能未安装 DB2 Information Integrator for Content。

#### 用户响应：

验证是否正确安装了 DB2 Information Integrator for Content。

---

### 对查询语法的更新

只包含了一个 - (减号) 查询项的查询是无效的。减号查询项只能用作后过滤器，而不能产生结果。

*Programming Guide and API Reference* (iiysp.pdf) 和信息中心中的主题“Query syntax”描述了可应用于搜索查询的语法类型。

---

### 折叠 URI

关于如何在搜索结果中折叠 URI 的主题包含了有误导性及不正确的信息。

- 该文档暗示所有数据源类型的 URI 都可在搜索结果中折叠。但实际上，只有那些带有 URL 的文档（即，由 Web 或 Nntp 搜寻器搜寻的文档）才能折叠 URI。
- 该文档中显示的 samegroupas 查询语法的用法也不正确。要在查询中使用此关键字，必须指定标准 URL；不能指定已创建的组的名称来将多个 URI 折叠在一起。

例如，要查询属于同一个组（例如，名称为 News 的组）的文档，可指定该组中其中一个文档的 URI，而不是该组的名称：

正确：samegroupas:http://mycompany.server3.com/news/

不正确：samegroupas:news

---

## 为 FileNet Content Services 启用直接文档访问

当您使用 VeniceBridge 搜寻器搜寻由 DB2 Information Integrator Content Edition (以前称为 VeniceBridge) 服务器管理的 FileNet Content Services 存储库时, 目标 FileNet Content Services 库和 FileNet Web Services 将配合使用 FileNet Content Services 连接器, 并且必须配置为使用直接文档访问。请参阅 *FileNet Web Services Administrator's Guide* 和 FileNet Content Services 管理文档以建立直接文档访问支持。

---

### 普通搜寻器插件

当配置搜寻器属性时, 可以指定想要使用的 Java™ 类的名称, 以用于强制执行文档级别访问控制, 并将元数据与企业搜索索引中的文档相关联。不能将普通搜寻器插件用于 Web 或 NNTP 搜寻器。

此搜寻器插件是在产品中实施的, 但在企业搜索文档中未对它进行描述。

普通搜寻器插件包括了 DB2 II OmniFind Edition V8.2 中提供的安全性令牌插件的所有功能。虽然安全性令牌插件仍受支持, 但我们建议您转为使用普通搜寻器插件。

要使用此插件, 必须创建新集合。不能将新插件用于先前创建的集合。

该插件由对搜寻器搜寻的每个文档调用的 Java 类组成。该 Java 类将传递给文档标识 (URI)、指定的安全性令牌和元数据。该类返回一组新的或已修改的安全性令牌和元数据, 或者该类可指示将忽略或跳过该文档。

在搜寻了搜寻空间中的所有文档后, 将只对新的或更改过的文档调用该插件。如果想要更改尚未更新的文档的安全性令牌和元数据, 可重新搜寻搜寻空间中的所有文档并重组索引。

#### 使用该插件以强制执行安全性

文档级安全性是通过将一个或多个安全性令牌 (用逗号分隔的字符串) 与搜寻器搜寻的每个文档相关联来实现的。组标识常被用作安全性令牌。

缺省情况下, 将对每个文档都指定一个公共令牌, 该令牌使每个人都可以使用该文档。公共令牌可以被替换为由管理员提供的值或从已搜寻文档中的字段抽取的值。

该插件允许您应用自己的业务规则来确定已搜寻的文档的安全性令牌值。与每个文档相关联的安全性令牌都存储在索引中。它们用于过滤与安全性令牌匹配的文档, 并确保只有那些用户有权查看的文档才会返回到搜索结果中。

#### 使用该插件来添加元数据

文档元数据, 例如, 上次修改日期数据或 Lotus Notes® 文档的字段数据, 是对所有搜寻过的文档创建的。搜寻器插件使您可以应用自己的业务规则, 以确定应对每个文档建立索引的元数据值。

元数据定义为名称 - 值对。可以使用自由文本查询或指定了字段名的查询来搜索元数据。

请访问 DB2 II OmniFind Edition 支持 Web 站点, 获取关于此主题的更多信息, 以及您可以用自己的逻辑进行修改的样本插件代码的信息。

---

## 我的概要文件页的帮助与用户界面不一致

样本搜索应用程序的用户界面在文档翻译之后又进行了更新。

在 ESSearchApplication 搜索应用程序的“我的概要文件”页中，**数据源凭证**和**数据源名称**标注已分别更改为**域凭证**和**域名称**。此页的帮助文件仍然使用旧的标注文本。

---

## Web 搜寻器的帮助页中有不正确的值

在配置 Web 搜寻器的联机帮助中指定了不正确的缺省值。

在配置 Web 搜寻器时，请参考以下更正过的缺省值。

### Web 搜寻器属性页

- 在**用于检索文档的并发线程的数目**中，联机帮助指出缺省值和最大值是 200。正确的缺省值是 50。最大值是 200。
- 在**最大页大小**中，联机帮助指出可以接受的值的范围是从 250 KB 到 10000 KB。最大值是 32 MB。

### 高级 Web 搜寻器属性页

- 在**最大 URL 数**中，联机帮助指出缺省值是 2 000 000 000 个 URL。正确的缺省值是 3 000 000。
- 在**最大活动主机数**中，联机帮助指出缺省值是 200。正确的缺省值是 500。
- 在**最大唯一文档数**中，联机帮助指出缺省值是 7 000 000。正确的缺省值是 1 000 000。
- 在**要搜寻的新文档的百分比**中，联机帮助指出缺省值是 99%。正确的缺省值是 75%。
- 在**临时存储器中的最大已搜寻文档数**中，联机帮助指出缺省值是 5 000 000 个文档。正确的缺省值是 500 000。

### Web 搜寻空间页

- 在 **URL 路径深度**中，联机帮助指出缺省值是 25。正确的缺省值是 16。

---

## indexBuildItem 描述

在 *Programming Guide and API Reference* 中，有一个主题是“Writing the index build configuration file”。其中，子章节“The <indexBuildItem> element”的描述不正确。

这段文字应该是：

If type A is a subtype of type B (for example, ?Person? as a subtype to ?Entity?), and there are <indexBuildItem> elements Ia and Ib defined for both types, the processing is as follows:

- Each index rule that is defined in Ib will be applied to feature structures of type B and feature structures of type A.
- Each index rule that is defined in Ia will be applied to feature structures of type A.

With this processing, one rule can be defined that stores all "Entity" annotations under a span "entity," which applies also for "Person" annotations, and another rule for "Person"

annotations that creates an additional span named "Person." Each rule is applied independently of other rules, so they can contain different <attributemapping> elements for the same features.

---

## 映射公共分析结构到索引

将公共分析结构映射到索引的 XML 文件必须使用 xmlns，而不是 xmlns:namespace。

该 XML 文件必须由以下语句开始:

```
<?xmlversion="1.0" encoding="UTF-8"?>
<indexBuildSpecification xmlns="http://www.ibm.com/of/822/consumer/index/xml">
```

不要附加 :namespace 到 xmlns 属性。否则，在解析器启动时将出现错误消息 FFQT0138E。

在 <indexBuildItem> 元素中使用 <name> 替代 <typeName>。配置样本中包含了错误的标记。它应该是:

```
<indexBuildItem><name>com.ibm.uima.tt.PersonAnnotation</name>
```

而不是:

```
<indexBuildItem><typeName>com.ibm.uima.tt.PersonAnnotation</typeName>
```

---

## 从 WebSphere Portal 迁移基于模型的分法

《管理企业搜索》中的章节『从 WebSphere Portal 迁移基于模型的分法』需要增加一些信息。

迁移实用程序首次运行时，它需要指向 WebSphere Portal 安装的路径。然后，它将更新企业搜索配置文件。为使这些配置更新生效，企业搜索系统必须停止。否则，用于迁移基于模型的分法文件的选项将变灰。对于此后运行的迁移实用程序，企业搜索系统会是活动的。

---

## 声明

本信息是为在美国提供的产品和服务编写的。IBM 可能未在所有国家或地区提供本文档中讨论的产品、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。

本信息包含日常商业运作所使用的数据和报表的示例。为了尽可能全面地说明这些数据和报表，这些示例包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，如与实际商业企业所使用的名称和地址有任何雷同，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。用户如果是为了按照 IBM 应用程序编程接口开发、使用、经销或分发应用程序，则可以任何形式复制、修改和分发这些样本程序，而无须向 IBM 付费。

凡这些样本程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明:

Outside In (®) Viewer Technology, ©1992-2005 Stellent, Chicago, IL., Inc. All Rights Reserved.

IBM XSLT Processor Licensed Materials - Property of IBM ©Copyright IBM Corp., 1999-2005. All Rights Reserved.

---

## 商标

本主题列示 IBM 商标和某些非 IBM 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:

IBM  
AIX  
AIX 5L  
DB2  
DB2 Universal Database  
Domino  
Domino.doc  
Hummingbird



Informix  
Lotus  
Lotus Notes  
Notes  
OmniFind  
POWER4  
POWER5  
RISC System/6000  
Tivoli  
WebSphere  
Workplace  
xSeries  
z/OS

下列各项是其他公司的商标或注册商标:

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国和 / 或其他国家或地区的商标或注册商标。

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

Intel、Intel Inside (徽标)、MMX 和 Pentium 是 Intel Corporation 在美国和 / 或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Linux 是 Linus Torvalds 在美国和 / 或其他国家或地区的商标。

其他公司、产品或服务名称可能是其他公司的商标或服务标记。



---

## 与 IBM 联系

在中国，请致电下列其中一个号码以与 IBM 联系：

- 800-810-1818 或 (010) 84981188 分机 5151，可获得售前客户服务；
- 800-810-1818 或 (010) 84981188 分机 5200，可获得售后客户服务；
- 800-810-1818 或 (010) 84981188 分机 5017，可获得市场营销与销售的信息；

要查找您所在国家或地区的 IBM 营业处，可在网上查看 IBM 全球联系人目录 (Directory of Worldwide Contacts)，网址为：[www.ibm.com/planetwide](http://www.ibm.com/planetwide)。

---

## 获取产品信息

关于 DB2 Information Integrator 的信息可通过万维网获取，网址为：<http://www-900.ibm.com/cn/software/db2/>。

此站点包含有关 DB2 产品家族、DB2 解决方案、技术前沿与趋势、DB2 服务、成功案例、市场活动、培训与认证、DB2 开发者园地、合作伙伴、下载中心、资料库、第三方分析报告、殊荣与奖项、DB2 新闻以及如何购买 DB2 的最新信息。

---

## 提供对文档的意见

请发送您对本书或其它 DB2 Information Integrator 文档的任何意见。

您的反馈有助于 IBM 提供高质量的信息。请发送您对本书或其它 DB2 Information Integrator 文档的任何意见。可以使用下列任何一种方法提出意见：

1. 使用 [www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf) 上的在线读者意见表发送您的意见。
2. 通过电子邮件将您的意见发送至 [ctscrcf@cn.ibm.com](mailto:ctscrcf@cn.ibm.com)。确保包括产品的名称、产品的版本号和书籍的名称及部件号（如果适用的话）。如果您对特定文本有意见，请包括此文本的位置（例如，标题、表号或页码）。



中国印刷



**Java**<sup>™</sup>  
**COMPATIBLE**