

IBM DB2 Information Integrator  
OmniFind Edition



# Notas sobre o Release

*Versão 8.2.2*



IBM DB2 Information Integrator  
OmniFind Edition



# Notas sobre o Release

*Versão 8.2.2*

Antes de utilizar estas informações e o produto suportado por elas, leia as informações gerais na seção "Avisos".

Este documento contém informações de propriedade da IBM. Ele é fornecido sob um acordo de licença, e é protegido por leis de direitos autorais. As informações contidas nesta publicação não incluem garantias de produto, e nenhuma declaração feita neste manual deve ser interpretada como tal.

Você pode solicitar as publicações IBM on-line ou através de seu representante IBM local:

- Para solicitar publicações on-line, consulte o IBM Publications Center em [www.ibm.com/shop/publications/order](http://www.ibm.com/shop/publications/order).
- Para localizar o representante IBM local, consulte o IBM Directory of Worldwide Contacts em [www.ibm.com/planetwide](http://www.ibm.com/planetwide).

Quando o Cliente envia seus comentários, concede direitos não-exclusivos à IBM para utilizá-los ou distribuí-los da maneira que achar conveniente, sem que isto implique em qualquer compromisso ou obrigação para com o Cliente.

© Copyright International Business Machines Corporation 2004, 2005. Todos os direitos reservados.

---

# Índice

<b>Sobre as Notas sobre o Release . . . . .</b>	<b>1</b>	<b>Atualizações de Documentação . . . . .</b>	<b>13</b>
Informações de Suporte Técnico na Web . . . . .	1	Alterações em Mensagens de Erro . . . . .	13
<b>Novidades Neste Release . . . . .</b>	<b>3</b>	Atualizações da Sintaxe de Consulta . . . . .	13
Pacote do Anotador Básico de Procura Corporativa . . . . .	4	Reduzindo URIs. . . . .	13
<b>Limitações, Problemas e Soluções</b>		Ativando Acesso Direto de Documentos para o	
<b>Alternativas Conhecidos . . . . .</b>	<b>7</b>	FileNet Content Services . . . . .	14
Limites de Tamanho do Documento . . . . .	7	Plug-in de Crawler Genérico . . . . .	14
Instalação do Fix Pack Resulta em um Novo Crawl		Ajuda para a Página Meu Perfil Não Corresponde à	
Completo . . . . .	7	Interface com o Usuário . . . . .	15
Limites de Tamanho de Heap para Anotador do		Páginas de Ajuda para o Web Crawler Incluem	
Analisador Java . . . . .	7	Valores Incorretos . . . . .	15
Tamanho Máximo dos Arquivos Processing Engine		Descrição de indexBuildItem . . . . .	16
Archive . . . . .	8	Mapeando uma Estrutura de Análise Comum para o	
Crawler do Notes no AIX Requer Módulo I/O		Índice . . . . .	16
Completion Port . . . . .	9	Migrando Taxonomia Baseada em Modelo do	
A Consulta de Coletas Seguras do Notes Pode		WebSphere Portal . . . . .	17
Resultar na Contagem Incorreta de Documentos		<b>Avisos . . . . .</b>	<b>19</b>
Resultantes . . . . .	10	Marcas Registradas . . . . .	21
Nomes de Usuários Podem Conter Ponto Final . . . . .	10	<b>Entrando em Contato com a IBM . . . . .</b>	<b>23</b>
Caracteres Curinga em uma Procura de Campo . . . . .	11	Obtendo Informações sobre o Produto . . . . .	23
Local do Índice no Servidor . . . . .	11	Fornecendo Comentários sobre a Documentação . . . . .	23
Processo Orphan URLFetcher para Linux e AIX . . . . .	11		
Segmentos de Memória Compartilhada no AIX . . . . .	12		



---

## Sobre as Notas sobre o Release

O documento de notas sobre o release descreve o que há de novo neste release, limitações, problemas, soluções alternativas e atualizações de documentação conhecidos para o IBM DB2 Information Integrator OmniFind Edition, Versão 8.2.2 (versão 8.2 mais Fix Pack 2).

O DB2 II OmniFind Edition, Versão 8.2.2 é somente um upgrade. Você deve ter a versão 8.2 ou a versão 8.2.1 antes que possa instalar a versão 8.2.2.

Antes de instalar no SuSE Linux Enterprise Server 9, verifique o Web site de Suporte do DB2 II OmniFind Edition, Versão 8.2.2 para obter soluções alternativas e atualizações.

---

### 1 Informações de Suporte Técnico na Web

1 O produto DB2 II OmniFind Edition foi renomeado recentemente como WebSphere  
1 Information Integrator OmniFind Edition. Você pode consultar as referências ao  
1 WebSphere Information Integrator OmniFind Edition nas páginas da Web  
1 relacionadas ao produto, mas as interfaces, documentação e informações de  
1 resolução de problemas do produtos (technotes) ainda refletem a marca DB2.

1 Você pode localizar informações de suporte nos seguintes Web sites:

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 Universal Database para produtos Linux, UNIX e Windows Versão 8**

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



---

## Novidades Neste Release

O IBM DB2 Information Integrator OmniFind Edition, Versão 8.2.2 fornece os seguintes recursos:

### Sistemas operacionais

A versão 8.2.2 suporta os seguintes sistemas operacionais adicionais:

- AIX 5.3 (requer Nível de Manutenção 1 e a PTF de agosto de 2004 do C++ Runtime para AIX)
- SuSE Linux Enterprise Server 9 com Service Pack 1 (UnitedLinux SP1) (requer libstdc++-3.3.3-43.28)

Antes de instalar no SuSE Linux Enterprise Server 9, verifique o Web site de Suporte do DB2 II OmniFind Edition, Versão 8.2.2 para obter soluções alternativas e atualizações.

### Documentos grandes

Coletas podem incluir documentos de até 32 MB de tamanho

### Segurança de nível de documento para documentos Domino

- Validação em tempo real (durante processamento de consulta) das credenciais de um usuário com controles de acesso gerenciados pelo repositório nativo
- Validação do usuário através de perfis de usuário e um registro do usuário do LDAP. Ao armazenar credenciais em um formato criptografado, os usuários não precisam especificar credenciais cada vez que consultam uma origem Domino.

### Crawlers

- Notes crawler: a capacidade de efetuar crawl em diretórios, além de bancos de dados, visualizações e pastas
- Web crawler:
  - Suporte de interface com o usuário para efetuar crawl em Web sites protegidos por senha (para sites que utilizam autenticação básica HTTP ou autenticação baseada em formulário)
  - A capacidade de efetuar crawl em Web sites servidos por servidores proxy
  - A capacidade de retornar mensagens customizadas para páginas de erros simbólicos

### Qualidade da procura

- Procura federada (a capacidade de procurar mais de uma coleta por vez)
- Suporte para caracteres curinga finais nas consultas
- Suporte para customizar o comprimento da exibição dos resumos nos resultados da procura
- Suporte para interface com o usuário para redução dos resultados da procura a partir do mesmo URI

### Análise lingüística

- Suporte para UIMA (Unstructured Information Management Architecture), que oferece a capacidade de fazer análise de texto customizada nos documentos em uma coleta
- Suporte para procura XML nativa

- Suporte para dicionários de sinônimos customizados

#### API e aplicativo de procura

- Suporte para interface com o usuário para configuração e monitoramento dos aplicativos listener dos dados de cliente customizados
- Um plug-in de crawler genérico, permitindo a especificação de regras customizadas de segurança e de negócios
- Muitos aprimoramentos no aplicativo de procura de amostra para demonstrar funções de consulta, opções de resultado de procura, procura federada e segurança de nível do documento

---

## Pacote do Anotador Básico de Procura Corporativa

Você pode utilizar o pacote do anotador básico de procura corporativa para desenvolver novos anotadores com base na saída dos anotadores de procura corporativa e para testar os anotadores de procura corporativa e seus anotadores customizados dentro do UIMA SDK (Software Development Kit).

O pacote do anotador básico de procura corporativa é um arquivo ZIP que contém as etapas de análise de texto que são utilizadas na procura corporativa. Esses anotadores serão sempre executados antes de qualquer análise customizada durante a análise de documentos.

O UIMA SDK (Software Development Kit) está disponível no Web site do IBM developerWorks: <http://www.ibm.com/developerworks/db2/zones/db2ii/>.

Para instalar o pacote do anotador:

1. Localize o pacote do anotador `OF_base_annotators.zip` na instalação da procura corporativa (DB2 II OmniFind Edition) no diretório `ES_INSTALL_ROOT/packages/uima`.
2. Copie o arquivo ZIP no diretório raiz da sua instalação UIMA SDK.
3. Extraia o arquivo ZIP para incluir os arquivos do anotador básico da procura corporativa na estrutura de diretório especificada da instalação do UIMA SDK.

O pacote do anotador básico de procura corporativa contém somente os dicionários que são requeridos para processar documentos em inglês. Caso queira processar outros idiomas em seu ambiente de desenvolvimento, siga estas etapas:

1. Localize os dicionários de procura corporativa na instalação da procura corporativa em `ES_INSTALL_ROOT/configurations/parserservice/jediidata/frost4`.
2. Copie o conteúdo do diretório em sua instalação UIMA SDK local em `UIMA_SDK_INSTALL/data/frost4`.

Para verificar se o UIMA SDK foi instalado com êxito:

1. Abra o CVD (CAS Visual Debugger) no seguinte diretório:  
`UIMA_SDK_INSTALL/bin/cvd[.bat/.sh]`.
2. Clique em **Executar** → **carregar TAE**.
3. Localize o arquivo especificador TAE `of_tokenization.xml` no diretório `UIMA_SDK_INSTALL/docs/examples/descriptors/analysis_engine`.
4. Carregue um documento de amostra e execute o TAE. Você verá anotações do tipo `uima.tt.TokenAnnotation` no CVD.

| Para utilizar os anotadores de procura corporativa para seu processamento:

- | 1. Inclua uma referência no arquivo of\_typesystem.xml na seção typeSystem do  
| seu especificador anotador customizado se seu anotador customizado utilizar  
| tipos que são definidos pelos anotadores de procura corporativa. O arquivo  
| of\_typesystem.xml está no diretório  
| UIMA\_SDK\_INSTALL/docs/examples/descriptors/analysis\_engine. Consulte o  
| arquivo jtok.xml no diretório analysis\_engine para obter uma amostra de como  
| incluir referências nos arquivos.
- | 2. Crie um especificador TAE agregado que inclua o arquivo of\_tokenization.xml  
| e seu especificador anotador customizado.
- | 3. Utilize o especificador TAE agregado ao testar o UIMA SDK.

| Ao criar um arquivo PEAR para transferir seu anotador por upload para a procura  
| corporativa, certifique-se de que este arquivo PEAR inclua somente seu  
| especificador anotador customizado. Ele pode não incluir nenhum dos arquivos a  
| seguir: of\_tokenization.xml, of\_typesystem.xml, o conteúdo do  
| of\_base\_annotators.zip ou especificador agregado criado na etapa 2.

| Para obter informações adicionais sobre os recursos de análise na procura  
| corporativa, consulte os tópicos "Custom Text Analysis Integration" e "Text  
| Analysis included in enterprise search" no centro de informações ou na publicação  
| *Programming Guide and API Reference for Enterprise Search*.



---

## Limitações, Problemas e Soluções Alternativas Conhecidos

Estas informações descrevem limitações, problemas e soluções alternativas conhecidos para o DB2 Information Integrator OmniFind Edition, Versão 8.2.2.

Quaisquer limitações e restrições podem ou não se aplicar a outros releases do produto.

---

### Limites de Tamanho do Documento

É possível efetuar crawl e indexar documentos de até 32 MB. Você pode efetuar crawl e indexar as tabelas do DB2 Universal Database com dados LOB somente se os dados aderirem ao limite geral de tamanho do documento de 32 MB.

Ignore as instruções nos arquivos de ajuda de propriedade do crawler que especificam que o tamanho máximo da página pode ser maior que 32 MB.

---

### Instalação do Fix Pack Resulta em um Novo Crawl Completo

Após instalar o DB2 II OmniFind Edition, Versão 8.2.2, será efetuado crawl novamente em todos os documentos, com exceção daqueles nos quais foi efetuado crawl pelos crawlers da Web e NNTP.

A primeira vez que você iniciar os crawlers não-Web e não-NNTP após instalar a versão 8.2.2, será efetuado crawl em todos os documentos em todas as origens de dados e eles serão indexados novamente e não apenas os documentos novos ou alterados desde a última vez em que foi efetuado crawl em uma origem de dados.

Após efetuar crawl em todos os documentos em cada tabela, diretório ou pasta na coleta, os crawlers são alternados para o processamento normal de atualização e começam a efetuar crawl somente em documentos novos ou alterados.

Normalmente, um recrawl completo é necessário somente quando as alterações na configuração do crawler exigem isso. Por exemplo, se os nomes de exibição para os campos ou colunas em uma origem de dados forem alterados, o crawler deverá efetuar recrawl em todos os documentos para que as novas informações possam ser indexadas com os documentos.

O recrawl completo que ocorre automaticamente após você instalar a versão 8.2.2 envolve tempo, mas não causa impacto na qualidade do índice de procura corporativa.

---

### Limites de Tamanho de Heap para Anotador do Analisador Java

Se as mensagens OutOfMemory forem exibidas no registro de erros de coleta para a sessão *collection\_ID.casprocessor* em execução no anotador do analisador Java, aumente o tamanho de heap para 200 MB.

Os anotadores de análise de texto customizados que são transferidos por upload para procura corporativa são executados em uma sessão Java. Por exemplo, na procura corporativa, estas sessões são sessões *collection\_ID.casprocessor*, em que *collection\_ID* é o ID de coleta ao qual esse anotador pertence. Estas sessões têm o tamanho de heap limitado a 200 MB.

Para aumentar o tamanho de heap:

1. Pare o sistema de procura corporativa. Em todos os sistemas operacionais, digite o seguinte comando: `esadmin stop`
2. Pare o CCL:
  - No Linux e AIX, digite o seguinte comando: `stopccl.sh`
  - No Windows, pare o serviço DB2 Information Integrator OmniFind Edition  
Windows: **Iniciar** → **Painel de Controle** → **Serviços**.
3. Localize o ID de coleta ao qual seu anotador pertence. Para localizar o ID de coleta, execute o comando `esadmin check` ou vá para o diretório `ES_NODE_ROOT/master_config` no qual você localizará um arquivo `collection_ID_config.ini` para cada coleta criada.
4. Aumente o valor `max_heap` para a sessão `collection_ID.casprocessor` no arquivo `collection_ID_config.ini`.
5. Classifique o arquivo `collection_ID_config.ini` e procure as entradas do processador CAS. Por exemplo, no Linux, para 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. Edite a linha `session_number.max_heap=200` para incluir o novo tamanho de heap máximo (200). Por exemplo, para col7: `session3.max_heap=200`
7. Reinicie o CCL:
  - No Linux e no AIX, digite o seguinte comando: `startccl.sh -bg`
  - No Windows, inicie o serviço DB2 Information Integrator OmniFind Edition  
Windows: **Iniciar** → **Painel de Controle** → **Serviços**.
8. Reinicie a procura corporativa. Em todos os sistemas operacionais, digite o seguinte comando: `esadmin start`
9. Inicie o analisador para esta coleta a partir do console de administração da procura corporativa. O anotador será executado em um JVM com o novo tamanho de heap máximo.

---

## Tamanho Máximo dos Arquivos Processing Engine Archive

Para a análise de texto customizada, o tamanho máximo dos arquivos PEAR (Processing Engine Archive Files) é de 8 MB.

Somente arquivos PEAR menores que 8 MB podem ser transferidos por upload utilizando o console de administração da procura corporativa. Os arquivos PEAR maiores que 8 MB geralmente contêm recursos, tais como dicionários. Para fazer upload de um arquivo PEAR:

1. Altere seu código do anotador para que ele carregue o recurso através do gerenciador de recursos UIMA (utilizando o método `getResourceObject` do `AnnotatorContext`) ou para que ele utilize o valor de um parâmetro de configuração do anotador que aponte para o recurso.

2. Copie os arquivos de recurso em um diretório dedicado para o nó Indexador do seu sistema de procura corporativa.
3. Altere a configuração do anotador (tanto a especificação <externalResource> quanto <namevaluePair> do parâmetro de configuração que você utiliza) para incluir o caminho absoluto no recurso utilizando o diretório criado na etapa 2.
4. Recrie seu arquivo PEAR com a configuração atualizada. Certifique-se de que ele não contenha os recursos e faça seu upload para a procura corporativa.

---

## Crawler do Notes no AIX Requer Módulo I/O Completion Port

Para utilizar o crawler do Notes e os processos de descoberta nas plataformas AIX IBM, você deve instalar o módulo IOCP (I/O Completion Port) (bos.iocp.rte) do CD-ROM do produto AIX e configurar seu status como **Disponível**.

Após instalar o módulo bos.iocp.rte, é preciso aplicar nele a correção apropriada. Consulte as informações no link a seguir para obter instruções:

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

Sem o módulo IOCP, o processo de descoberta do Notes falhará durante a criação de um novo crawler e a seguinte mensagem de erro será exibida:

```
FFQM0105E Recebido um erro do servidor -
Mensagem: FFQG0024E Foi capturada uma exceção inesperada: descoberta
```

Além disso, a seguinte mensagem, que inclui o erro ENOEXEC, será gravada no arquivo *ES\_NODE\_ROOT/logs/system\_yyyymmdd.log*. (Alguns textos de mensagens são divididos em várias linhas para melhorar a leitura.)

```
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 Foi capturada uma exceção com o detalhe 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
carregar ENOEXEC na(s) biblioteca(s)
compartilhada(s)/opt/lotus/notes/latest/ibmpow/libnotes_r.a'
e um rastreamento de pilha de 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
carregar ENOEXEC na(s) biblioteca(s) compartilhada(s)
/opt/lotus/notes/latest/ibmpow/libnotes_r.a
em java.lang.ClassLoader$NativeLibrary.load(Native Method)
em java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)
em java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)
em java.lang.Runtime.loadLibrary0(Runtime.java:824)
em java.lang.System.loadLibrary(System.java:908)
em lotus.domino.NotesThread.load(NotesThread.java:306)
em lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)
em lotus.domino.NotesThread.sinitThread(NotesThread.java:181)
em com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover
(Origem Desconhecida)
em com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Origem Desconhecida)
em com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover
(ComponentDiscoveryW.java:72)
em sun.reflect.NativeMethodAccessorImpl.invoke0(Método Nativo)
em sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
em sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
em sun.reflect.DelegatingMethodAccessorImpl.invoke
(DelegatingMethodAccessorImpl.java:60)
em java.lang.reflect.Method.invoke(Method.java:391)
em com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)
```

Certifique-se de que o módulo IOCP esteja instalado corretamente antes de criar um crawler do Notes:

1. Digite o seguinte comando `lslpp` para garantir que o módulo IOCP seja instalado no servidor de crawler:

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

A saída do comando `lslpp` é semelhante ao seguinte exemplo:

Conjunto de Arquivos	Nível	Estado	Descrição
-----			
Caminho: /usr/lib/objrepos bos.iocp.rte	5.2.0.10	COMMITTED	I/O Completion Ports API
Caminho: /etc/objrepos bos.iocp.rte	5.2.0.10	COMMITTED	I/O Completion Ports API

2. Digite o seguinte comando `lsdev` para garantir que a porta IOCP esteja **Disponível**:

```
$ lsdev -Cc iocp
```

A saída do comando `lsdev` deve corresponder ao seguinte exemplo:

```
iocp0 Disponível I/O Completion Ports
```

3. Se o status da porta IOCP estiver **Definido**, siga as seguintes etapas para alterar o status para **Disponível**:
  - a. Efetue login no servidor de crawler como usuário root e emita o seguinte comando:

```
# smit iocp
```
  - b. Selecione **Alterar / Mostrar Características do I/O Completion Ports** e altere o **ESTADO** a ser configurado na reinicialização do sistema de **Definido** para **Disponível**.
  - c. Reinicie o servidor de crawler.
  - d. Digite o comando `lsdev` novamente e confirme se o status da porta IOCP foi alterado para **Disponível**.

---

## A Consulta de Coletas Seguras do Notes Pode Resultar na Contagem Incorreta de Documentos Resultantes

O número de documentos resultantes em um conjunto de resultados pode não ser exato para consultas que procuram coletas seguras nas quais é selecionada a opção para validar credenciais atuais, durante o processamento de consulta.

Quando você procura uma coleta de procura corporativa, o número total de documentos resultantes é mostrado nos resultados da procura. Este número pode ser maior que o número real de documentos resultantes se todas as condições a seguir forem verdadeiras:

- A segurança foi ativada para a coleta quando foi criada.
- A coleta inclui um crawler do Notes.
- A opção para validar as credenciais atuais durante o processamento de consulta foi especificada quando as opções de segurança em nível de documento foram configuradas para o crawler do Notes.

Este problema afeta o valor `SI-API ResultSet.getAvailableNumberOfResults` e o aplicativo de procura de amostra `ESSearchApplication`.

---

## Nomes de Usuários Podem Conter Ponto Final

As instruções on-line para inclusão de nomes de administrador de procura corporativa estão incorretas.

Ao utilizar o console de administração de procura corporativa para incluir um ID de administrador de procura corporativa no sistema, as instruções na página especificam que o nome do usuário não pode conter um ponto final. Essa instrução está incorreta. Nomes de usuários com ponto final (tais como `j_user@this.company.com`) são permitidos.

---

## Caracteres Curinga em uma Procura de Campo

Geralmente é permitida a utilização de `field:pref*` como um termo de consulta. O DB2 II OmniFind Edition recuperará documentos que incluem palavras que começam com `pref` marcadas como membros do campo denominado `field`. Há uma exceção para esta regra: para alguns campos especiais para os quais consultas curingas são menos relevantes, um termo de consulta `field:pref*` é tratado como se fosse `field:pref`. Esses campos especiais são:

- `site`
- `link`
- `docid`
- `samegroupas`
- `url`
- grupo (novos grupos, que são semelhantes a `site`, mas são utilizados para NNTP)

Por exemplo, a consulta `url:porta*` pode não retornar nenhum resultado, enquanto que `url:portal` retornará resultados.

---

## Local do Índice no Servidor

Se você criar uma coleta e especificar um local de índice diferente do local padrão no diretório raiz de dados do servidor, o uso de espaço no disco para este local de índice não padrão não será monitorado.

Em uma instalação com vários servidores, quando você cria uma coleta, se quiser especificar um local para o índice diferente do local padrão, o local do índice deve primeiro existir em todos os quatro servidores.

---

## Processo Orphan URLFetcher para Linux e AIX

**Problema:** O crawler da Web utiliza dois processos. Quando um processo sai, ele sinaliza um segundo processo para sair. Um processo pode, algumas vezes, sair inesperadamente e deixar o outro processo sendo executado indefinidamente.

Você pode determinar se isso ocorreu procurando um processo órfão como o processo a seguir na saída do comando `ps` do 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:...
```

Note que:

- Um PPID (Parent Process ID) igual a 1 indica que o processo é órfão.
- A presença do `URLFetcher.jar` no caminho de classe indica que este é um processo do crawler da Web.

**Solução**

Localize o PID (ID do Processo) do processo órfão e o elimine. No exemplo anterior, o processo órfão é 26222. No Linux e AIX, utilize o comando kill sem argumentos. Você pode também ver entradas “filhas” na tabela de processos. As entradas filhas representam encadeamentos que são criados pelo processo órfão. Essas entradas filhas devem ser excluídas quando você eliminar seus processos ascendentes.

Quando você vê um processo URLFetcher órfão, deverá eliminar o processo porque se ele estiver na tabela de processos, o sistema operacional reservará recursos para ele. Você deve verificar regularmente se esses processos órfãos existem, especialmente se suspeitar que o crawler da Web foi concluído de maneira anormal. Se os processos existirem, elimine-os.

**Importante:** No DB2 II OmniFind Edition Versão 8.2.1, os processos órfãos raramente ocorrem porque o localizador de URL é encerrado automaticamente ao observar que o componente C++ foi eliminado. Se o problema ocorrer, tome cuidado para não eliminar a sessão de edição httpauth.ini de outro usuário acidentalmente. O utilitário HTTPAuthEdit é um JVM executado a partir do arquivo JAR URLFetcher e pode parecer semelhante na tabela de processos.

---

## Segmentos de Memória Compartilhada no AIX

Caso receba a mensagem SQL1224N no AIX, você pode ter excedido o número de segmentos de memória compartilhada permitido pelo AIX.

Para solucionar este problema, consulte o tópico Segmentos de Memória Compartilhada SQL1224N no Web site de Suporte da IBM.

---

## Atualizações de Documentação

As atualizações de documentação incluem correções ou adições menores na documentação da versão 8.2.2 para o DB2 Information Integrator OmniFind Edition.

---

## Alterações em Mensagens de Erro

Mensagens de erro foram atualizadas após a documentação ser enviada para a tradução.

### FFQD1034E

ERRO: A variável de ambiente CMBROOT / IBMCMROOT não está definida. Verifique se o DB2 Information Integrator para Content foi instalado corretamente.

#### Explicação:

A variável de ambiente CMBROOT / IBMCMROOT não foi definida. O DB2 Information Integrator para Content pode não ter sido instalado.

#### Resposta do Usuário:

Verifique se o DB2 Information Integrator para Content foi instalado corretamente.

### FFQD1035E

ERRO: O arquivo cmbicmsrvs.ini não foi localizado. Verifique se o DB2 Information Integrator para Content foi instalado corretamente.

#### Explicação:

O arquivo cmbicmsrvs.ini não foi localizado no caminho especificado. O DB2 Information Integrator para Content pode não ter sido instalado.

#### Resposta do Usuário:

Verifique se o DB2 Information Integrator para Content foi instalado corretamente.

---

## Atualizações da Sintaxe de Consulta

Uma consulta que consiste em um único termo – (menos) é inválida. Um termo menos pode ser utilizado somente como pós-filtro e não para gerar resultados.

O tópico “Query syntax”, que descreve os tipos de sintaxe que você pode aplicar nas consultas de procura, está no *Programming Guide and API Reference* (iiysp.pdf) e no centro de informações.

---

## Reduzindo URIs

Tópicos sobre como reduzir URIs em resultados de procura contêm informações enganosas e incorretas.

- A documentação implica que URIs para todos os tipos de origem de dados podem ser reduzidos nos resultados da procura. No entanto, a capacidade de reduzir URIs é limitada a documentos que têm URLs (ou seja, documentos nos quais foi efetuado crawl por um crawler da Web ou NNTP).

- A documentação também mostra o uso incorreto da sintaxe de consulta `samegroupas`. Para utilizar esta palavra-chave em uma consulta, você deve especificar uma URL completa; você não deve especificar o nome de um grupo que foi criado para reduzir vários URIs juntos.

Por exemplo, para consultar documentos que pertencem ao mesmo grupo (tais como um grupo denominado Notícias), especifique o URI de um dos documentos no grupo e não o nome do grupo:

Correto: `samegroupas:http://mycompany.server3.com/news/`

Incorreto: `samegroupas:news`

---

## Ativando Acesso Direto de Documentos para o FileNet Content Services

Ao utilizar o crawler VeniceBridge para efetuar `crawl` nos repositórios do FileNet Content Services gerenciados pelo servidor DB2 Information Integrator Content Edition (anteriormente chamado VeniceBridge), a biblioteca de destino do FileNet Content Services e o FileNet Web Services trabalham com o conector do FileNet Content Services e devem ser configurados para utilizar o acesso direto de documentos. Consulte o *Guia do Administrador* do FileNet Web Services e a documentação de administração do FileNet Content Services para configurar o suporte ao acesso direto de documentos.

---

## Plug-in de Crawler Genérico

Quando configura as propriedades do crawler, você pode especificar o nome de uma classe Java que deseja utilizar para reforçar o controle de acesso em nível de documento e para associar metadados com documentos em um índice de procura corporativa. Você não pode utilizar um plug-in de crawler genérico com crawlers da Web ou NNTP.

O plug-in de crawler é implementado no produto, mas não é descrito na documentação da procura corporativa.

O plug-in de crawler genérico cobre todas as funções do plug-in do token de segurança que estava disponível na versão 8.2 do DB2 II OmniFind Edition. Embora o plug-in do token de segurança ainda seja suportado, é recomendado que você utilize o plug-in de crawler genérico em seu lugar.

Para utilizar o plug-in, você deve criar uma nova coleta. Você não pode utilizar um novo plug-in com coletas criadas anteriormente.

O plug-in consiste em uma classe Java que é chamada para cada documento no qual o crawler efetua `crawl`. A classe Java é transmitida para o identificador de documentos (URI), tokens de segurança designados e metadados. A classe retorna um conjunto novo ou modificado de tokens de segurança e metadados ou a classe pode indicar que o documento será ignorado ou desconsiderado.

Após ter sido efetuado `crawl` em todos os documentos no `crawl space`, o plug-in é chamado somente para documentos novos ou alterados. Se quiser alterar os tokens de segurança e metadados para documentos que não foram atualizados, efetue `crawl` novamente em todos os documentos no `crawl space` e reorganize o índice.

### Utilizando o Plug-in para Reforçar a Segurança

A segurança em nível de documento é realizada através da associação de um ou mais tokens de segurança (uma cadeia delimitada por vírgula) com

cada documento em que o crawler efetua crawl. Identificadores de grupo são comumente utilizados como os tokens de segurança.

Por padrão, cada documento tem um token público designado que disponibiliza um documento para todos. O token público pode ser substituído por um valor fornecido pelo administrador ou por um valor extraído de um campo do documento no qual foi efetuado crawl.

O plug-in permite que você aplique suas próprias regras de negócios para determinar o valor dos tokens de segurança para documentos em que tenha sido efetuado crawl. Os tokens de segurança que são associados com cada documento são armazenados no índice. Eles são utilizados para filtrar documentos que correspondem aos tokens de segurança e garantem que somente os documentos que um usuário tem permissão para visualizar são retornados nos resultados da procura.

#### **Utilizando o Plug-in para Incluir Metadados**

Metadados de documentos, tais como dados da última data de modificação ou os dados do campo dos documentos do Lotus Notes, são criados para todos os documentos em que foi efetuado crawl. O plug-in do crawler permite que você aplique suas próprias regras de negócios para determinar o valor dos metadados que devem ser indexados para cada documento.

Os metadados são definidos como um par nome-valor. Você pode procurar os metadados com uma consulta sem texto ou com uma consulta que especifica o nome do campo.

Consulte o Web site Suporte para DB2 II OmniFind Edition para obter informações adicionais sobre este tópico e sobre o código do plug-in de amostra que você pode modificar com sua própria lógica.

---

## **Ajuda para a Página Meu Perfil Não Corresponde à Interface com o Usuário**

A interface com o usuário para o aplicativo de procura de amostra foi atualizada após a documentação ser traduzida.

Na página Meu Perfil no aplicativo de procura ESSearchApplication, as etiquetas **Credenciais da Origem de Dados** e **Nome da Origem de Dados** foram alteradas para **Credenciais do Domínio** e **Nome do Domínio**, respectivamente. O arquivo de ajuda para esta página ainda reflete o texto da etiqueta antiga.

---

## **Páginas de Ajuda para o Web Crawler Incluem Valores Incorretos**

Valores padrão incorretos são especificados na ajuda on-line para a configuração de um Web crawler.

Consulte as seguintes correções ao configurar um Web crawler.

#### **Página Propriedades do Web Crawler**

- Em **Número de Encadeamentos Simultâneos para Recuperação de Documentos**, a ajuda on-line afirma que o valor padrão e o valor máximo são 200. O valor padrão correto é 50. O valor máximo é 200.
- Em **Tamanho Máximo da Página**, a ajuda on-line afirma que valores aceitáveis variam de 250 KB a 10000 KB. O valor máximo é 32 MB.

#### **Página Propriedades do Web Crawler Avançadas**

- Em **Número Máximo de URLs**, a ajuda on-line afirma que o valor padrão é 2 000 000 URLs. O valor padrão correto é 3 000 000.
- Em **Número Máximo de Hosts Ativos**, a ajuda on-line afirma que o valor padrão é 200. O valor padrão correto é 500.
- Em **Número Máximo de Documentos Exclusivos**, a ajuda on-line afirma que o valor padrão é 7 000 000. O valor padrão correto é 1 000 000.
- Em **Porcentagem de Novos Documentos nos quais Efetuar Crawl**, a ajuda on-line afirma que o valor padrão é 99 por cento. O valor padrão correto é 75 por cento.
- Em **Número Máximo de Documentos nos quais Foi Efetuado Crawl no Armazenamento Temporário**, a ajuda on-line afirma que o valor padrão é 5 000 000 documentos. O valor padrão correto é 500 000.

#### Página Web Crawl Space

- Em **Profundidade do Caminho da URL**, a ajuda on-line afirma que o valor padrão é 25. O valor padrão correto é 16.

---

## Descrição de `indexBuildItem`

No *Programming Guide and API Reference*, há um tópico chamado "Writing the index build configuration file." A descrição da subseção chamada "O The `<indexBuildItem>` element" está incorreta.

O texto deve ser o seguinte:

Se o tipo A for um subtipo do tipo B (por exemplo, `?Person?` como um subtipo de `?Entity?`) e houver elementos `<indexBuildItem>` Ia e Ib definidos para ambos os tipos, o processamento será como a seguir:

- Cada regra de índice que é definida em Ib será aplicada nas estruturas de recurso do tipo B e estruturas de recurso do tipo A.
- Cada regra de índice que é definida em Ia será aplicada nas estruturas de recurso do tipo A.

Com este processamento, pode ser definida uma regra que armazena todas as anotações de "Entity" em uma extensão "entity," que também se aplica para anotações de "Person" e outra regra para anotações de "Person" que cria uma extensão adicional denominada "Person." Cada regra é aplicada independentemente de outras regras, pois elas podem conter elementos `<attributemapping>` diferentes para os mesmos recursos.

---

## Mapeando uma Estrutura de Análise Comum para o Índice

O arquivo XML para o mapeamento de uma estrutura de análise comum para o índice deve utilizar `xmlns`, e não `xmlns:namespace`.

O arquivo XML deve iniciar com as seguintes instruções:

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

Não anexe `:namespace` ao atributo `xmlns`. Caso contrário, a mensagem de erro FFQT0138E aparecerá quando o analisador for iniciado.

Utilize `<name>` em vez de `<typeName>` no elemento `<indexBuildItem>`. A configuração de amostra contém uma tag incorreta. Ela deveria ser:

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

em vez de

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

---

## Migrando Taxonomia Baseada em Modelo do WebSphere Portal

Em *Administrando Procura Corporativa*, o tópico denominado “Migrando uma Taxonomia Baseada em Modelo do WebSphere Portal” requer informações adicionais.

Quando o utilitário de migração é executado pela primeira vez, ele precisa do caminho para a instalação do WebSphere Portal. Em seguida, ele atualizará os arquivos de configuração da procura corporativa. Para que essas atualizações de configuração funcionem, o sistema de procura corporativa deve ser parado. Caso contrário, a opção para migrar arquivos de taxonomia baseados em modelo ficará esmaecida. Para execuções subsequentes do utilitário de migração, o sistema de procura corporativa pode ficar ativo.



---

## Avisos

Estas informações foram desenvolvidas para produtos e serviços oferecidos nos Estados Unidos. É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos nesta publicação em outros países. Consulte um representante IBM local para obter informações sobre os produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM, poderá ser utilizado em substituição a este produto, programa ou serviço. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não-IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes pendentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não garante ao Cliente nenhum direito sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para: Gerência de Relações Comerciais e Industriais da IBM Brasil, Av. Pasteur, 138-146 Botafogo, Rio de Janeiro, RJ, CEP 22290-240

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para: IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

O parágrafo a seguir não se aplica a nenhum país/região em que tais disposições não estejam de acordo com a legislação local: A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO “NO ESTADO EM QUE SE ENCONTRA” SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS IMPLÍCITAS DE NÃO-INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em certas transações; portanto, essa disposição pode não se aplicar ao Cliente.

Esta publicação pode conter imprecisões técnicas ou erros tipográficos. Periodicamente, são feitas alterações nas informações aqui contidas; tais alterações serão incorporadas em novas edições da publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

Referências nestas informações a Web sites não-IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a esses Web sites. Os materiais contidos nesses Web sites não fazem parte dos materiais desse produto IBM e a utilização desses Web sites é de inteira responsabilidade do Cliente.

A IBM pode utilizar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre este assunto com objetivo de permitir: (i) a troca de informações entre programas criados

independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

Gerência de Relações Comerciais e Industriais da IBM Brasil  
Av. Pasteur, 138-146  
Botafogo  
Rio de Janeiro, RJ  
CEP 22290-240

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriadas, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito neste documento e todo o material licenciado disponível são fornecidos pela IBM sob os termos do Contrato com o Cliente IBM, do Contrato de Licença do Programa Internacional IBM ou de qualquer outro contrato equivalente.

Todos os dados de desempenho aqui contidos foram determinados em um ambiente controlado. Portanto, os resultados obtidos em outros ambientes operacionais podem variar significativamente. Algumas medidas podem ter sido tomadas em sistemas de nível de desenvolvimento e não há garantia de que estas medidas serão iguais em sistemas geralmente disponíveis. Além disso, algumas medidas podem ter sido estimadas por extrapolação. Os resultados reais podem variar. Os usuários deste documento devem verificar os dados aplicáveis para seu ambiente específico.

As informações sobre produtos não-IBM foram obtidas junto aos fornecedores dos respectivos produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou estes produtos e não pode confirmar a precisão de seu desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não-IBM. Dúvidas sobre os recursos de produtos não-IBM devem ser encaminhadas diretamente a seus fornecedores.

Todas as declarações relacionadas aos objetivos e intenções futuras da IBM estão sujeitas a alterações ou cancelamento sem aviso prévio e representam apenas metas e objetivos.

Estas informações contêm exemplos de dados e relatórios utilizados nas operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e endereços utilizados por uma empresa real é mera coincidência.

#### LICENÇA DE DIREITOS AUTORAIS:

Estas informações contêm programas aplicativos de amostra na linguagem fonte, ilustrando as técnicas de programação em diversas plataformas operacionais. O Cliente pode copiar, modificar e distribuir estes programas de exemplo sem a necessidade de pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos em conformidade com a interface de programação de aplicativo para a plataforma operacional para a qual os programas de exemplo são criados. Esses exemplos não foram testados completamente em todas as condições. Portanto, a IBM não pode garantir ou implicar a confiabilidade, manutenção ou função destes programas. O Cliente pode copiar, modificar e distribuir estes programas de exemplo sem a necessidade de

pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos, em conformidade com interfaces de programação de aplicativos da IBM.

Cada cópia ou parte destes programas de exemplo ou qualquer trabalho derivado deve incluir um aviso de direitos autorais com os dizeres:

Outside In (®) Viewer Technology, ©1992-2004 Stellent, Chicago, IL., Inc. Todos os direitos reservados.

Materiais Licenciados do IBM XSLT Processor - Propriedade da IBM © Direitos Autorais IBM Corp., 1999-2004. Todos os direitos reservados.

---

## Marcas Registradas

Este tópico lista marcas registradas IBM e determinadas marcas registradas não-IBM.

Os termos a seguir são marcas registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países:

IBM  
AIX  
AIX 5L  
DB2  
DB2 Universal Database  
Domino  
Domino.doc  
Hummingbird  
Informix  
Lotus  
Lotus Notes  
Notas  
OmniFind  
POWER4  
POWER5  
RISC System/6000  
Tivoli  
WebSphere  
Workplace  
xSeries  
z/OS

Os termos a seguir são marcas ou marcas registradas de outras empresas:

Java e todas as marcas registradas e logotipos baseados em Java são marcas ou marcas registradas da Sun Microsystems, Inc. nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas registradas da Microsoft Corporation nos Estados Unidos e/ou em outros países.

Intel, Intel Inside (logotipos), MMX e Pentium são marcas registradas da Intel Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada do The Open Group nos Estados Unidos e em outros países.

Linux é uma marca registrada de Linus Torvalds nos Estados Unidos e/ou em outros países.

Outros nomes de empresas, produtos ou serviços podem ser marcas registradas ou marcas de serviço de terceiros.

---

## Entrando em Contato com a IBM

Para entrar em contato com o serviço de atendimento a clientes IBM nos Estados Unidos ou Canadá, ligue para 1-800-IBM-SERV (1-800-426-7378). No Brasil, ligue para 0-800-7014-262.

Para informações sobre opções de serviços disponíveis, ligue para um dos seguintes números:

- Nos Estados Unidos: 1-888-426-4343
- No Canadá: 1-800-465-9600

Para localizar um escritório da IBM em seu país ou região, consulte o IBM Directory of Worldwide Contacts na Web no endereço [www.ibm.com/planetwide](http://www.ibm.com/planetwide).

---

## Obtendo Informações sobre o Produto

Informações sobre os produtos DB2 Information Integrator estão disponíveis através de telefone ou na Web.

Informações sobre os produtos DB2 Information Integrator estão disponíveis através de telefone ou na Web. Os números de telefone fornecidos aqui são válidos nos Estados Unidos. No Brasil, ligue para 0-800-7014-262 - para Atendimento ao Cliente ou para 0-800-7014-850 para FAX.

1. Para solicitar produtos ou obter informações gerais: 1-800-IBM-CALL (1-800-426-2255)
2. Para solicitar publicações: 1-800-879-2755
3. Visite o endereço da Web [www.ibm.com/software/data/integration/db2ii/support.html](http://www.ibm.com/software/data/integration/db2ii/support.html).

Este site contém as informações mais recentes sobre:

- A biblioteca técnica
- Solicitação de manuais
- Downloads para clientes
- Newsgroups
- Fix packs
- Notícias
- Links para recursos na Web

---

## Fornecendo Comentários sobre a Documentação

Envie seus comentários sobre este manual ou outra documentação do DB2 Information Integrator.

Sua opinião ajuda a IBM a fornecer informações de alta qualidade. Envie seus comentários sobre este manual ou outra documentação do DB2 Information Integrator. Você pode utilizar qualquer um dos métodos a seguir para enviar comentários:

1. Envie seus comentários utilizando o formulário on-line de comentários do leitor em [www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf).

2. Envie seus comentários por e-mail para [comments@us.ibm.com](mailto:comments@us.ibm.com). Inclua o nome do produto, o número da versão do produto e o nome e o número de peça do manual (se aplicável). Se você estiver fazendo comentários sobre um texto específico, por favor, inclua a localização do texto (por exemplo, um título, um número de tabela ou número de página).





Impresso em Brazil



**Java**<sup>™</sup>  
**COMPATIBLE**