



# Gerir o Information Integrator para o Content

*Versão 8 Edição 2*





# Gerir o Information Integrator para o Content

*Versão 8 Edição 2*

**Nota**

Antes de utilizar estas informações e o produto a que se referem, leia as informações incluídas em “Informações” na página 137.

**Segunda Edição (Março 2003)**

Esta edição é aplicável à Versão 8 Edição 2 do IBM Enterprise Information Portal for Multiplatforms (número de produto 5724-B43) e a todas as edições e modificações posteriores até indicação em contrário em novas edições.

Partes deste produto são: Copyright © 1990-2000 ActionPoint, Inc. e/ou detentores de licenças, 1299 Parkmoor Drive, San Jose, CA 95126 U.S.A. Todos os direitos reservados.

Tecnologia de Visualização Outside In<sup>®</sup>, ©1992-2000 Inso Corporation. Todos os Direitos Reservados.

© Copyright International Business Machines Corporation 1999, 2003. Todos os direitos reservados.

# Índice

|  |            |
|--|------------|
| <b>Acerca deste manual</b> . . . . .   | <b>vii</b> |
| Quem deve utilizar este manual . . . . .   | vii        |
| Competências necessárias para administradores . . . . .                                    | viii       |
| Competências necessárias para analistas empresariais ou modeladores de processos . . . . . | viii       |
| Onde encontrar mais informações . . . . .  | viii       |
| Informações incluídas na embalagem do produto . . . . .                                    | viii       |
| Suporte disponível na Web . . . . .  | ix         |
| Como enviar comentários . . . . .  | x          |
| Novidades no EIP Versão 8.2 . . . . .  | x          |

## Capítulo 1. Apresentação do Enterprise Information Portal . . . . .

|  |          |
|--|----------|
| <b>Capítulo 1. Apresentação do Enterprise Information Portal</b> . . . . . | <b>1</b> |
| Pesquisar informações sobre clientes . . . . .                             | 1        |
| A necessidade . . . . .  | 2        |
| A solução . . . . .  | 2        |
| Descrição Geral . . . . .  | 2        |
| Apresentação dos componentes do Enterprise Information Portal . . . . .    | 3        |

## Capítulo 2. Introdução ao cliente de administração . . . . .

|  |          |
|--|----------|
| <b>Capítulo 2. Introdução ao cliente de administração</b> . . . . .                                  | <b>7</b> |
| Utilizar os Primeiros Passos do EIP para compreender o cliente de administração de sistema . . . . . | 7        |
| Administrar o EIP . . . . .  | 7        |
| Gerir utilizadores e grupos . . . . .  | 7        |
| Utilizar as ferramentas do cliente de administração . . . . .  | 7        |
| Introduzir privilégios . . . . .   | 9        |
| Privilégios . . . . .  | 9        |
| Alternar entre vistas de produto e bases de dados . . . . .  | 10       |
| Melhoramentos e aperfeiçoamentos no cliente de administração . . . . .                               | 10       |
| Ligar um cliente de administração a uma base de dados de administração local . . . . .               | 11       |
| Ligar o cliente de administração a uma base de dados de administração remota . . . . .               | 11       |
| Passo 1 - catalogue a base de dados remota utilizando o Assistente de Configuração do DB2 . . . . .  | 12       |
| Passo 2 - utilize o Utilitário de Configuração do Servidor . . . . .                                 | 13       |
| Passo 3 - testar a ligação da base de dados remota . . . . .   | 14       |
| Definir tipos de documentos . . . . .  | 15       |
| Alterar o ficheiro de tipo MIME do servidor (cmbcc2mime.ini) . . . . .                               | 15       |

## Capítulo 3. Utilizar as características do cliente de administração de EIP . . . . .

|   |           |
|---|-----------|
| <b>Capítulo 3. Utilizar as características do cliente de administração de EIP</b> . . . . . | <b>17</b> |
| Criar uma pesquisa associada . . . . .  | 17        |
| Definir servidores . . . . .  | 18        |
| Directrizes para definir servidores . . . . .   | 19        |
| Trabalhar com o conector OnDemand: Ajuste e sockets TCP/IP . . . . .                        | 23        |

|   |    |
|---|----|
| Trabalhar com o conector de Extended Search . . . . .           | 24 |
| Criar entidades federadas . . . . .                             | 24 |
| Compreender entidades federadas . . . . .                       | 24 |
| Utilizar o assistente de Criação de Entidade Federada . . . . . | 25 |
| Criar índices de texto federados . . . . .                      | 26 |
| Criar modelos de pesquisa . . . . .                             | 26 |
| Definir o modelo de pesquisa . . . . .                          | 27 |
| Definir critérios de pesquisa . . . . .                         | 27 |
| Determinar as definições de pesquisa . . . . .                  | 27 |
| Atribuir privilégios . . . . .                                  | 28 |

## Capítulo 4. Gerir acesso de utilizadores 29

|  |    |
|--|----|
| <b>Capítulo 4. Gerir acesso de utilizadores</b> 29   |    |
| Criar IDs de utilizador e palavras-passe . . . . .   | 29 |
| Compreender a autoridade de administração de DB2 . . . . .   | 30 |
| Estabelecer ligação com o DB2 utilizando os ficheiros INI . . . . .  | 30 |
| Alterar as palavras-passe do servidor de bibliotecas e do administrador do sistema para o gestor de recursos . . . . . | 31 |
| Alterar as palavras-passe de acesso à base de dados . . . . .  | 31 |
| Importar utilizadores de LDAP . . . . .  | 32 |
| Introduzir privilégios . . . . .   | 33 |
| Criar conjuntos de privilégios . . . . .   | 33 |
| Criar grupos de privilégios . . . . .  | 34 |
| Atribuir um conjunto de privilégios a um utilizador . . . . .  | 34 |
| Atribuir a um ID de utilizador um conjunto de privilégios de concessão . . . . .                                       | 34 |
| Atribuir utilizadores a gestor de recursos . . . . .   | 34 |
| Atribuir utilizadores a recolhas . . . . .   | 35 |
| Criar grupos de utilizadores . . . . .   | 35 |
| Criar listas de controlo de acessos . . . . .  | 35 |
| Atribuir um conjunto de privilégios a uma lista de controlo de acessos . . . . .                                       | 35 |
| Criar domínios . . . . .   | 36 |
| Administrar domínios . . . . .   | 37 |
| Aceder a domínios . . . . .  | 37 |
| Atribuir um utilizador a um domínio . . . . .  | 37 |
| Atribuir um grupo de utilizadores a um domínio . . . . .   | 37 |
| Atribuir um conjunto de privilégios a um domínio . . . . .   | 38 |
| Atribuir um gestor de recursos a um domínio . . . . .  | 38 |
| Atribuir uma recolha a um domínio . . . . .  | 38 |
| Mover um utilizador de um domínio para outro . . . . .   | 38 |
| Mover um grupo de utilizadores de um domínio para outro . . . . .  | 39 |
| Mover um gestor de recursos de um domínio para outro . . . . .   | 39 |
| Mover uma recolha de um domínio para outro . . . . .   | 39 |
| Mover um conjunto de privilégios de um domínio para outro . . . . .  | 40 |
| Mover uma lista de controlo de acesso de um domínio para outro . . . . .   | 40 |

## **Capítulo 5. Gerir o information mining 41**

|  |    |
|--|----|
| O que é o Information Mining? . . . . .                                      | 41 |
| Os serviços do Information Mining do Enterprise Information Portal . . . . . | 41 |
| Componentes dos serviços do Information Mining . . . . .                     | 42 |
| Utilizar o Information Mining num ambiente empresarial . . . . .             | 44 |
| Um exemplo da utilização do Information Mining . . . . .                     | 46 |
| Idiomas e Formatos Suportados . . . . .                                      | 49 |
| Conceitos . . . . .  | 50 |
| Arquitectura do sistema . . . . .  | 50 |
| Os conceitos do Information Mining . . . . .                                 | 51 |
| As ferramentas do Information Mining . . . . .                               | 53 |
| Interfaces de Programação . . . . .  | 60 |
| Primeiros Passos. . . . .  | 62 |
| Construir uma taxonomia . . . . .  | 62 |
| Instalar a Information Structuring Tool . . . . .                            | 63 |
| Como começar . . . . .   | 63 |
| Direitos de acesso . . . . .   | 63 |
| Definir uma taxonomia . . . . .  | 64 |
| Seleccionar documentos de formação . . . . .                                 | 65 |
| Carregar documentos de formação . . . . .                                    | 66 |
| Avaliar um modelo de categorização . . . . .                                 | 68 |
| Efectuar a formação a um catálogo . . . . .                                  | 72 |
| Regulação do rendimento. . . . .   | 73 |
| Utilizar o IBM Web Crawler . . . . .   | 73 |
| Capacidades do IBM Web Crawler . . . . .                                     | 74 |
| Configurar e executar o IBM Web Crawler para a Web . . . . .                 | 74 |
| O ficheiro de configuração do IBM Web Crawler . . . . .                      | 78 |
| Registar-se no IBM Web Crawler . . . . .                                     | 87 |
| Detecção e Correção de Problemas . . . . .                                   | 88 |
| Escolher Resumidores . . . . .   | 88 |
| IBM Web Crawler for Notes . . . . .  | 90 |
| Excluir o IBM Web Crawler de um servidor . . . . .                           | 94 |

## **Capítulo 6. Introdução ao fluxo de trabalho . . . . . 97**

|  |     |
|--|-----|
| Compreender o fluxo de trabalho . . . . .  | 97  |
| Como utilizar o fluxo de trabalho . . . . .  | 97  |
| Sincronizar IDs de utilizador e grupos . . . . .   | 98  |
| Reinstalar o servidor do EIP com o Workflow activado . . . . .                                     | 99  |
| Actualizar IDs de utilizador e grupos entre o MQSeries Workflow e a base de dados do EIP . . . . . | 100 |
| Planear um fluxo de trabalho . . . . .   | 101 |
| Informações a processar . . . . .  | 101 |
| Tratamento de informações . . . . .  | 101 |
| Acções a efectuar . . . . .  | 102 |
| Como as informações fluem através do processo . . . . .  | 102 |
| Como tudo se encaixa . . . . .   | 102 |
| Utilizar componentes de fluxo de trabalho do Enterprise Information Portal . . . . .               | 102 |
| Utilizar o construtor de fluxo de trabalho . . . . .   | 103 |
| Utilizar os serviços de fluxo de trabalho . . . . .  | 103 |
| Definir listas de trabalhos . . . . .  | 104 |
| Definir listas de acções . . . . .   | 104 |
| Criar um fluxo de trabalho . . . . .   | 105 |

|  |     |
|--|-----|
| Activar construtor de fluxo de trabalho. . . . . | 105 |
| Iniciar o servidor MQSeries Workflow . . . . .   | 105 |

## **Capítulo 7. Ficheiros exemplo do IBM Web Crawler. . . . . 107**

|   |     |
|---|-----|
| exemplo de config-sample2.xml . . . . .                             | 107 |
| IBM Web Crawler exemplo de ficheiro de análise de registo . . . . . | 109 |

## **Capítulo 8. Utilizar a pesquisa de texto e QBIC . . . . . 113**

|   |     |
|---|-----|
| Pesquisar documentos utilizando o motor de pesquisa de texto . . . . .          | 113 |
| Activar um servidor de pesquisa de texto . . . . .                              | 113 |
| Pesquisar imagens utilizando a Pesquisa por Conteúdo de Imagem (QBIC) . . . . . | 114 |
| Introdução à pesquisa de imagem . . . . .                                       | 114 |
| Configurar a pesquisa de imagens . . . . .                                      | 114 |
| Carregar e indexar dados exemplo . . . . .                                      | 117 |
| Antes de carregar os dados exemplo. . . . .                                     | 117 |
| Criar um índice remissivo de pesquisa de texto . . . . .                        | 117 |
| Criar a base de dados, o catálogo e as funções da pesquisa de imagens . . . . . | 119 |
| Executar o programa de carregamento . . . . .                                   | 119 |
| Indexar os dados de texto exemplo . . . . .                                     | 121 |

## **Capítulo 9. Formatos de Documentos 123**

|  |     |
|--|-----|
| Formatos de documentos do Information mining . . . . . | 123 |
| Processamento de texto: Genérico . . . . .             | 123 |
| Processamento de Texto: DOS . . . . .                  | 123 |
| Processamento de Texto: International . . . . .        | 124 |
| Processamento de Texto: Windows . . . . .              | 124 |
| Processamento de Texto: Macintosh . . . . .            | 125 |
| Formatos de folhas de cálculo . . . . .                | 125 |
| Formatos de Bases de Dados . . . . .                   | 125 |
| Formatos padrão de gráficos . . . . .                  | 126 |
| Formatos de gráficos com grande capacidade . . . . .   | 128 |
| Formatos de Apresentação . . . . .                     | 128 |
| Formatos compactados e codificados . . . . .           | 128 |
| Outra . . . . .  | 129 |

## **Capítulo 10. Gestão de direitos. . . . . 131**

|  |     |
|--|-----|
| Proteger a sua propriedade intelectual . . . . . | 131 |
| Utilizar técnicas de marcação . . . . .          | 132 |
| Marcação visível . . . . .                       | 133 |
| Marcação invisível. . . . .                      | 133 |

## **Capítulo 11. Características de Acessibilidade . . . . . 135**

|  |     |
|--|-----|
| Entrada do teclado e navegação . . . . .                 | 135 |
| Características para ecrã acessível . . . . .            | 135 |
| Compatibilidade com tecnologias de assistência . . . . . | 136 |
| Documentação Acessível . . . . .                         | 136 |

## **Informações . . . . . 137**

|                             |     |
|-----------------------------|-----|
| Marcas Comerciais . . . . . | 139 |
|-----------------------------|-----|

## **Glossário . . . . . 141**

|                            |     |
|----------------------------|-----|
| Índice Remissivo . . . . . | 149 |
|----------------------------|-----|





---

## Acerca deste manual

Este guia fornece uma introdução a todos os conceitos básicos que tem de compreender para gerir o sistema Enterprise Information Portal (EIP). Uma vez que o EIP oferece vários componentes que podem ser geridos a partir do cliente de administração, e porque pode aceder a funcionalidades de outros produtos utilizando o EIP, este não é um guia típico de administração de sistema. Este documento centra-se nos seguintes tópicos e explica de que forma:

- Utilizar o EIP para corresponder às necessidades comerciais
- Aceder e utilizar o cliente de administração
- Gerir acesso de utilizadores
- Utilizar o EIP para pesquisar conteúdos em vários servidores de conteúdos, incluindo dados estruturados armazenados em bases de dados relacionais, conteúdos não estruturados ou multimédia, ou documentos de texto
- Concepção, implementação e gestão de fluxos de trabalho

---

## Quem deve utilizar este manual

Este guia apoia os administradores do EIP na realização das seguintes tarefas:

### **Administração do Sistema**

Incluindo administração da base de dados, do servidor e da rede

### **Gestão de utilizadores**

Definir e conceder acesso a indivíduos e grupos, manter listas de controlo de acesso

### **Pesquisas federadas**

Definir e utilizar modelos de pesquisas federadas para obter conteúdos a partir do sistema Content Management

### **Information mining**

Extraír informações a partir de documentos, categorizar documentos e resultados de pesquisas

### **Web Crawling**

Utilizar o IBM Web Crawler para pesquisar e importar conteúdos da web

### **Pesquisa de texto**

Utilizar o IBM DB2 TIE ou o IBM Text Search Engine (apenas Content Manager Versão 7.1 e anterior) para pesquisar e indexar documentos

### **Pesquisa de imagens**

Utilizar o Content Manager Versão 7.1 (e anterior) para efectuar pesquisas de imagens

### **Gestão de fluxo de trabalho**

Utilizar as ferramentas de fluxo de trabalho do EIP para gerir os fluxos de trabalho de informações de uma empresa

---

## Competências necessárias para administradores

Dependendo das tarefas que efectuar, tem de ter uma boa compreensão de:

- Protocolos de segurança para acesso de utilizador
- Sistemas operativos Windows NT, Windows XP, Windows 2000, AIX ou Solaris
- Administração de rede
- Modelos de dados dos servidores de conteúdos no seu sistema de gestão de conteúdos
- Administração de base de dados
- Como aplicar um conhecimento prático do conteúdo e critérios de pesquisa ao criar modelos de pesquisa.
- Técnicas e ferramentas do Information Mining
- Os princípios da concepção de fluxo de trabalho
- Os processos de trabalho que pretende suportar com fluxos de trabalho do EIP

---

## Competências necessárias para analistas empresariais ou modeladores de processos

Os analistas e modeladores de processos empresariais encontrarão neste guia informações conceptuais relativas ao modo de definição e moldagem dos fluxos de trabalho do EIP para as suas empresas.

Para utilizar o construtor de fluxos de trabalho do Enterprise Information Portal, tem de:

- Compreender as necessidades dos funcionários, programas e estruturas de dados utilizados nos processos empresariais da empresa.
- Tomar decisões relacionadas com os negócios ou com os processos de fluxo de trabalho da empresa.

---

## Onde encontrar mais informações

O pacote do produto inclui um conjunto completo de informações para o ajudar a planear, instalar, administrar e utilizar o sistema. Também estão disponíveis na Web documentação e suporte do produto.

### Informações incluídas na embalagem do produto

O pacote do produto contém um centro de informações e todas as publicações em portable document format (.PDF).

#### O centro de informações

O pacote do produto contém um centro de informações que pode ser instalado aquando da instalação do produto. Para obter informações sobre o centro de informações, consulte o manual *Planeamento e Instalação do Sistema Content Management*.

O centro de informações contém a documentação para Content Manager, Enterprise Information Portal e IBM Content Manager VideoCharger for Multiplatforms. As informações referentes aos tópicos estão organizadas por produto e por tarefa (por exemplo, Administração). Para além dos índices e dos mecanismos de navegação, está disponível uma função de pesquisa que facilita o acesso à informação.

## Publicações em PDF

Pode visualizar os ficheiros PDF online utilizando o Adobe Acrobat Reader para o seu sistema operativo. Se não tiver o Acrobat Reader instalado, pode descarregá-lo a partir do sítio da Web da Adobe em [www.adobe.com](http://www.adobe.com).

A Tabela 1 mostra as publicações do Content Manager incluídas no IBM Content Manager for Multiplatforms.

*Tabela 1. Publicações do Content Manager*

| Nome do ficheiro | Título  | Número de publicação |
|------------------|---|----------------------|
| install          | <i>Planeamento e Instalação do Sistema Content Management<sup>1</sup></i> | GC17-5411-01         |
| migrate          | <i>Migração para o Content Manager Versão 8</i>                           | SC17-5425-01         |
| sysadmin         | <i>Manual de Administração do Sistema</i>                                 | SC17-5428-01         |

Quando encomenda o IBM Content Manager for Multiplatforms, recebe também o IBM Enterprise Information Portal for Multiplatforms. Ou pode encomendar o IBM Enterprise Information Portal for Multiplatforms separadamente. A Tabela 2 apresenta as publicações do Enterprise Information Portal incluídas com o produto.

*Tabela 2. Publicações do Enterprise Information Portal*

| Nome do ficheiro | Título  | Número de publicação |
|------------------|---|----------------------|
| apgwork          | <i>Application Programming Guide for Windows<sup>1</sup></i>              | SC17-5414-01         |
| ecliinst         | <i>Instalação, Configuração e Gestão do eClient</i>                       | SC17-5416-02         |
| eipinst          | <i>Planeamento e Instalação do IBM Information Integrator for Content</i> | GC17-5406-01         |
| eipmanag         | <i>Gerir o Information Integrator para o Content</i>                      | SC17-5413-01         |
| messcode         | <i>Mensagens e Códigos<sup>2</sup></i>                                    | SC17-5415-01         |

### Notas:

1. O *Application Programming Guide for Windows* contém informações sobre programação de aplicações, tanto para o Content Manager como para o Enterprise Information Portal.
2. O *Mensagens e Códigos* contém as mensagens e códigos para Content Manager e Enterprise Information Portal.

## Suporte disponível na Web

O suporte do produto está disponível na Web. Clique em **Support** nos sítios da Web do produto em:

[www.ibm.com/software/data/cm/](http://www.ibm.com/software/data/cm/)

[www.ibm.com/software/data/eip/](http://www.ibm.com/software/data/eip/)

A documentação está incluída em formato electrónico juntamente com o produto. Para aceder à documentação sobre o produto na Web, clique em **Library** no sítio da Web do produto.

Uma interface de documentação baseada em HTML, denominada Enterprise Documentation Online (EDO), está também disponível a partir da Web.

Actualmente, contém as informações de referência da API. Consulte a página da Web Library do Enterprise Information Portal para obter informações sobre como aceder ao EDO.

## Como enviar comentários

Os comentários do utilizador ajudam a IBM a disponibilizar informações de qualidade. Queira enviar comentários que tenha acerca desta publicação ou de qualquer outra documentação relativa ao Content Manager ou ao Enterprise Information Portal. Pode utilizar um dos métodos que se seguem para enviar comentários:

- Através da Web. Visite a página de RCF (Reader's Comment Form) da IBM Data Management Online em:  
[www.ibm.com/software/data/rcf](http://www.ibm.com/software/data/rcf)  
Pode utilizar a página para introduzir e enviar comentários.
- Envie comentários por e-mail para [comments@vnet.ibm.com](mailto:comments@vnet.ibm.com). Certifique-se de que inclui o nome e o número da versão do produto e o nome e o part number do manual (se aplicável). Se pretender tecer comentários acerca de um texto específico, inclua a localização do texto (por exemplo, o título de um capítulo ou de uma secção, o número de uma tabela, de uma página ou o título de um tópico da ajuda).

---

## Novidades no EIP Versão 8.2

Foram feitas as seguintes alterações ao produto:

### **Suporte para Sun Solaris**

Pode instalar conectores, características e bases de dados em sistemas Solaris.

### **Administração de sistema comum**

Uma só aplicação de cliente fornece acesso separado ao Content Manager e à administração do Enterprise Information Portal .

### **Novos conectores**

- O conector ICM para Content Manager Versão 8 Edição 1 permite tirar partido das poderosas características de armazenamento de documentos do Content Manager Versão 8'.
- O novo conector C++ Extended Search Versão 3.7 é executado em AIX.

### **Conectores Aperfeiçoados**

- As pesquisas de texto paramétricas são suportadas a partir da camada federada e através de uma ligação directa do Extended Search.
- Os aperfeiçoamentos funcionais e melhoramentos de rendimento efectuados ao conector OnDemand, incluem:
  - Modificações à estrutura de um OnDemand DDO.
  - A pesquisa assíncrona é agora suportada

### **Novos serviços do information mining**

- Extração de características
- Agrupamento em Unidades
- Identificação de idioma

### **IBM Web Crawler**

O IBM Web Crawler é uma função que permite aos utilizadores procurar e resumir informações na Web e em bases de dados Lotus Notes.

### **Aperfeiçoamentos de Fluxo de Trabalho**

O fluxo de trabalho é agora completamente suportado em AIX e Solaris. O construtor de fluxo de trabalho ,as APIs e os JavaBeans fornecem funções e utilização melhoradas do fluxo de trabalho.

### **Centro de informações**

O centro de informações com base no motor de pesquisa inclui a documentação para Content Manager, Enterprise Information Portal e IBM Content Manager VideoCharger for Multiplatforms. As informações referentes aos tópicos estão organizadas por produto e por tarefa (por exemplo, Administração). Para além dos índices e dos mecanismos de navegação fornecidos, uma função de pesquisa também facilita a recuperação de informação.

### **Acessibilidade**

As características de acessibilidade auxiliam o utilizador que sofre de algum tipo de incapacidade física, tal como mobilidade reduzida ou visão limitada, a utilizar produtos de software de forma bem sucedida. As principais características de acessibilidade para este produto incluem:

- A capacidade para realizar todas as funções através do teclado em vez do rato.
- Suporte de propriedades de apresentação ampliada.
- Opções para sinais de alerta vídeo e áudio
- Compatibilidade com tecnologias de assistência
- Compatibilidade com as características de acessibilidade ao sistema operativo
- Formatos acessíveis de documentação



---

## Capítulo 1. Apresentação do Enterprise Information Portal

Muitas empresas com um grande movimento de documentos, como, por exemplo, companhias de seguros e instituições financeiras, administram grandes volumes de conteúdos relacionados com a área empresarial. A necessidade de encontrar uma solução em termos empresariais para gerir e aceder a informações comerciais é partilhada por muitas indústrias.

Um *servidor de conteúdos* é um sistema de software que guarda ficheiros de multimédia, formulários de negócios, documentos e dados relacionados, com metadados que permitem aos funcionários processar e trabalhar com os conteúdos. Quando não existe um modo de ligar eficazmente servidores de conteúdos diferentes, a empresa pode perder tempo e dinheiro com a duplicação de informações ou a formação de funcionários para a execução de múltiplas pesquisas.

O Enterprise Information Portal (EIP) providencia tecnologia de ponta para reunir todos os recursos da empresa no seu computador. O EIP pode ajudar a maximizar o valor das informações e dos recursos multimédia ligando servidores de conteúdos diferentes através de um único cliente. Com um cliente do EIP, os utilizadores podem aceder rápida e simultaneamente a todos os servidores de conteúdos que se encontrem ligados. Os utilizadores podem também efectuar a extracção de informações ou pesquisas avançadas em servidores de conteúdos, incluindo a Web ou uma intranet. Podem executar tarefas de fluxo de trabalho nos processos de negócio que definir.

O EIP permite-lhe personalizar aplicações para a sua empresa. Ao utilizar os exemplos do EIP, os programadores de aplicações podem criar aplicações de ambiente de trabalho e aplicações baseadas na Web.

Nesta secção faz-se uma descrição geral do EIP. É apresentada uma situação acerca de uma companhia de seguros fictícia, a Seguradora XYZ, que demonstra as funções e capacidades do EIP.

---

### Pesquisar informações sobre clientes

A Seguradora XYZ (XYZ), uma grande companhia de seguros multiriscos, tem um extenso acervo de fotografias, participações de sinistros, apólices, notas de avaliadores, relatórios de peritos e outros documentos empresariais.

A XYZ mantém todos os memorandos enviados aos titulares de apólices, junto com formulários médicos e de avaliação electrónicos em arquivos de ficheiros Lotus Domino.Doc. A XYZ arquiva todas as declarações de apólice, notificações e facturas num servidor Content Manager OnDemand para armazenamento a longo prazo e de rápido acesso. A XYZ guarda todos as participações de sinistros, fotografias e correspondência recebidas dos titulares de apólices numa pasta de sistema do Content Manager para iSeries. A XYZ mantém relatórios de peritos em DB2 Universal Database (DB2 UDB) Data Warehouse Center Information Catalog Manager. A XYZ armazena também conteúdos multimédia, tais como gráficos de alta resolução num sistema Content Manager por forma a serem partilhados pelos departamentos de publicidade, relações públicas e novos negócios. Além disso, a XYZ mantém informações, como por exemplo os procedimentos empresariais, na intranet da empresa.

## A necessidade

As participações, as chamadas de clientes e a assistência geral a titulares de apólices não podem ser processadas com o conteúdo de um único servidor porque os funcionários têm necessidade de aceder a todas as informações de cliente. Para fornecer assistência a clientes, os funcionários necessitam de aceder simultaneamente a diversos servidores de conteúdos. A Seguradora XYZ precisa de uma solução que ligue todos os seus servidores de conteúdos e a intranet da empresa para pesquisa e obtenção de informação. Também pretendem expandir a utilização do processamento de fluxos de trabalho.

Existem vários funcionários que necessitam de aceder a documentos, desde o atendimento ao público a peritos e mediadores. A XYZ tem de limitar o acesso a certos artigos e facultar acesso ilimitado a outros. A XYZ também precisa de uma interface de fácil utilização para reduzir a necessidade de formação.

## A solução

A Seguradora XYZ optou por implementar o EIP porque as tecnologias de pesquisa exhaustiva permitem ligar-se e pesquisar todos os servidores de conteúdos para obtenção de dados. Actualmente, quando um funcionário do Centro de Atendimento da XYZ recebe uma chamada, basta uma única pesquisa associada para obter todas as informações necessárias acerca do titular da apólice.

A Seguradora XYZ também recorre à função information mining do EIP para pesquisar e obter informações da intranet da empresa. Pretendem também expandir a utilização dos processos de fluxo de trabalho.

---

## Descrição Geral

O EIP é um produto global: os seus componentes trabalham em conjunto para fornecer uma solução exclusiva adequada à sua empresa. Centrado numa arquitectura de várias camadas, EIP fornece um cliente de administração para gerir pesquisa, clientes para executar pesquisas e conectores para ligar a servidores de conteúdos tais como IBM Content Manager, Content Manager ImagePlus para OS/390, Content Manager OnDemand, Lotus Domino.Doc, DB2 Universal Database, DB2 DataJoiner e DB2 Data Warehouse Center Information Catalog Manager. Pode criar conectores adicionais para servidores de conteúdos adicionais, através do toolkit e dos exemplos do conector de EIP.

A arquitectura do EIP permite às aplicações cliente efectuar pesquisas únicas num ou mais servidores de conteúdos. Para efectuar pesquisas, o cliente utiliza modelos de pesquisa definidos pelo administrador do EIP.

Utilizando os modelos de pesquisa, o cliente executa uma *pesquisa associada*, uma pesquisa executada em simultâneo em servidores de conteúdos cujas correspondências dos atributos nativos tenham sido definidas com os atributos associados utilizados no modelo de pesquisa. Os modelos de pesquisa do EIP contêm critérios de pesquisa que fazem referência aos atributos federados que correspondem aos atributos nativos em cada um dos servidores de conteúdos. O administrador do EIP cria os modelos de pesquisa. O EIP fornece conectores para aceder e pesquisar dados guardados em vários servidores de conteúdos. Os servidores de conteúdos devolvem então objectos de dados ao cliente.

A arquitectura do EIP fornece as seguintes vantagens:

- Acesso por consulta única a vários servidores de conteúdos que suportam transacções de e-business e aplicações de serviços a clientes.



- A capacidade do Information Mining em vários servidores de conteúdos, incluindo a Web.
- Acesso de processos de fluxo de trabalho a dados em múltiplos servidores de conteúdos.
- Suporte para o desenvolvimento de aplicações cliente independentes da localização dos dados em qualquer servidor de conteúdos, devido à separação de aplicações cliente, índices remissivos e dados.

## Apresentação dos componentes do Enterprise Information Portal

Esta secção explica cada componente do EIP e descreve as opções de instalação.

A Tabela 3 na página 3 lista os componentes e os sistemas operativos compatíveis.

*Tabela 3. compatibilidade de sistema operativo de componente EIP*

| Componente                      | Windows | AIX | Solaris | Notas  |
|---------------------------------|---------|-----|---------|--|
| Base de Dados de Administração  | sim     | sim | sim     | Base de dados inclui funcionalidade de construtor de fluxo de trabalho   |
| Cliente de administração        | sim     | não | não     | Cliente pode ligar a base de dados instalada nos sistemas operativos Windows, AIX ou Solaris.  |
| Conectores                      | sim     | sim | sim     |  |
| Information mining              | sim     | sim | sim     |  |
| IBM Web Crawler                 | sim     | sim | sim     |  |
| Cliente de Pesquisa de Texto    | sim     | sim | sim     |  |
| Cliente de pesquisa de imagens  | sim     | sim | sim     |  |
| Toolkits de conector e exemplos | sim     | sim | sim     | <ul style="list-style-type: none"> <li>• A versão Windows inclui código de fonte para compilar o cliente exemplo. Não existe nenhum código de cliente exemplo instalado em AIX.</li> <li>• Os exemplos de fluxo de trabalho e APIs são instalados com o exemplo de conector federado.</li> </ul> |

Tabela 3. compatibilidade de sistema operativo de componente EIP (continuação)

| Componente            | Windows | AIX | Solaris | Notas                                    |
|-----------------------|---------|-----|---------|--|
| Visualizador          | sim     | não | não     | Instala cliente e visualizador OnDemand. |
| Centro de informações | sim     | sim | sim     |  |

## Administração

O componente de administração fornece a base de dados de administração e subcomponentes do cliente de administração. Quando instala a base de dados de administração, também instala a função de fluxo de trabalho.

**Base de dados de administração:** A base de dados de administração é uma base de dados de DB2 que gere informações acerca dos utilizadores e grupos do EIP, níveis de privilégio, palavras-passe, IDs de utilizador e outras. A base de dados também fornece o fluxo de trabalho e, opcionalmente, a funcionalidade Information Mining. Pode instalar várias bases de dados. Cada base de dados fornece a funcionalidade do fluxo de trabalho EIP. Se possuir um sistema Content Manager Versão 8, pode partilhar a base de dados de administração do EIP com a base de dados da Biblioteca de Serviços do Content Manager Versão 8. Pode partilhar a base de dados uma vez que a base de dados da Biblioteca de Serviços contém toda a informação requerida pelo EIP.

**Cliente de administração:** O cliente de administração apenas pode ser instalado nas estações de trabalho Windows. Poderá instalar vários clientes. Se tiver um sistema Content Manager Versão 8, pode administrar EIP e Content Manager Versão 8 a partir do mesmo cliente.

O cliente fornece a interface que permite ao administrador:

- Definir cada servidor de conteúdos para pesquisas federadas.
- Identificar entidades nativas e atributos em servidores de conteúdos e correspondê-los com entidades federadas.
- Manter um inventário dos atributos e entidades de base para todos os servidores de conteúdos definidos pelo utilizador.
- Criar modelos de pesquisa.
- Identificar e gerir utilizadores e grupos.
- Atribuir privilégios e conjuntos de privilégios aos utilizadores.
- Definir o acesso a modelos de pesquisa e definir as condições para as acções que os utilizadores podem efectuar com as informações obtidas numa pesquisa.
- Conceber e administrar os processos de fluxo de trabalho da empresa.

## Conectores

Os conectores fornecem a interface de comunicações entre os clientes EIP, os servidores de conteúdos e a base de dados de administração. Os conectores de servidores de conteúdos, tais como o conector Content Manager Versão 7.1, fornecem a funcionalidade que permite ao EIP iniciar sessão no servidor, pesquisar informações e devolver as informações à administração ou aos clientes de utilizadores finais. O conector federado liga o cliente de administração à base de dados da administração.

O EIP fornece os seguintes conectores:

- Conector federado que liga o cliente EIP à base de dados da administração.

- Conector da base de dados relacional para DB2 Universal Database 7.1, controlador JDBC 1.3 (apenas Java), ODBC 3.0 (apenas C++), DataJoiner 2.1.1.
- Conector de Content Manager para servidores de Content Manager Versão 7.1.
- Conector de Content Manager para servidores Content Manager Versão 8.2.
- Conector de Content Manager OnDemand para Content Manager OnDemand Versão 7.1.
- Content Manager for VisualInfo for 400 Versão 4.3 e Versão 5.1.
- Conector Content Manager ImagePlus for OS/390 for ImagePlus/390 Folder Application Facility Versão 3.1, Image Plus/390 ODM Versão 3.1.
- Conector Lotus Domino.Doc para Domino.Doc Versão 3.0a, Desktop Enabler Versão 3.0a.
- Conector Extended Search para Versão 3.7.
- Conector Information Catalog Manager para DB2. Universal Database Visual Warehouse Versão 5.2, DB2 Universal Database Versão 7.2.

## **Funções**

EIP tem quatro funções opcionais.

### **Information mining**

O Information Mining fornece serviços linguísticos para localizar informações ocultas de documentos de texto em servidores de conteúdos. Durante o processamento de documentos de texto, são criados metadados que podem ser resumidos, divididos em categorias e pesquisados. O WebSphere Application Server 4.0 (edição normal ou avançada) é um pré-requisito do information mining. Pode também agrupar documentos semelhantes, extrair características de documentos, como por exemplo nomes de pessoas ou empresas, e determinar o idioma de um documento.

### **Cliente de pesquisa de imagens**

Fornece a interface necessária para aceder e administrar funcionalidade de Pesquisa de Imagens num servidor de conteúdos do Content Manager Versão 7.

### **Cliente de Pesquisa de Texto**

Fornece a interface necessária para aceder e administrar a funcionalidade de Pesquisa de Texto num servidor de Pesquisa de Texto.

### **IBM Web Crawler**

Web Crawler é um crawler e um extractor de conteúdos baseado em Java. O Web Crawler pode efectuar o crawl a conteúdos numa web de intranet, extranet ou Internet, em bases de dados de Lotus Notes ou utilizando o Domino, sistemas de ficheiros locais e recolhas de FTP.

O Web Crawler consegue explorar metadados e texto a partir de muitos tipos de conteúdos. Por exemplo, o conteúdo HTML pode ser pesquisado para procurar por URL, título, corpo, hora da última alteração e meta-identificadores, tais como, autor, palavras-chave, descrição e outros. Os utilizadores fazem a sua selecção a partir de uma lista de extractores predefinidos para um determinado tipo de conteúdos. O conteúdo e/ou metadados explorados são guardados no disco local.

## **Visualizador de conteúdo**

Ao instalar o visualizador OnDemand são também instalados o cliente OnDemand e outros ficheiros necessários para ver documentos obtidos a partir do servidor OnDemand.

## Toolkits de conector e exemplos

O EIP fornece um toolkit de conector que inclui programas exemplo que podem ser utilizados para experimentar e testar diversas funções do EIP, tais como:

- estabelecer e terminar a ligação com servidores de conteúdos
- executar SQL e outras consultas exemplificativas em servidores de conteúdos
- determinar tipos MIME de servidores de conteúdos e outros

**Toolkit de conector de Windows:** Para instalar o toolkit de conector e exemplos em servidores Windows tem de seleccionar o tipo de máquina Estação de Trabalho de Desenvolvimento. Em seguida, seleccione o componente Toolkit de Conector e Exemplos. Pode instalar os programas exemplo para todos os conectores ou seleccionar exemplos individuais de acordo com os conectores instalados.

Nos servidores Windows, os programas exemplo do toolkit de conector estão organizados da seguinte forma:

```
c:\CMBROOT\SAMPLES\activex\xx
c:\CMBROOT\SAMPLES\cpp\xx
c:\CMBROOT\SAMPLES\java\xx
c:\CMBROOT\SAMPLES\jsp\xx
c:\CMBROOT\SAMPLES\server\xx
```

em que *xx* é o nome do directório que contém os programas exemplo para cada conector específico, por exemplo, db2, od, dl.

Nos servidores AIX, os programas exemplo estão organizados da seguinte forma:

```
/usr/lpp/cmb/samples/cpp/icm
/usr/lpp/cmb/samples/java/xx
/usr/lpp/cmb/samples/jsp/xx
/usr/lpp/cmb/samples/server/exit
```

em que *xx* é o nome do subdirectório, por exemplo, beans, servlets, ou outros.

Nos servidores Solaris, os programas exemplo estão organizados da seguinte forma:

```
/opt/IBMcmb/samples/java/xx
/opt/IBMcmb/samples/jsp/xx
/opt/IBMcmb/samples/server/exit
```

em que *xx* é o nome do subdirectório, por exemplo, beans, servlets, ou outros.

Os programas exemplo incluem documentação que descreve os programas e que fornece as definições de servidor (definições de ambiente, memória e outras) necessárias para trabalhar com o código exemplo.

## Centro de informações

O componente do centro de informações contém o Centro de Informações do Enterprise Information Portal. O Centro de Informações é uma versão da biblioteca do Enterprise Information Portal baseada na Web e passível de ser pesquisada.

---

## Capítulo 2. Introdução ao cliente de administração

O cliente de administração fornece a interface entre a base de dados de administração de EIP e o administrador EIP. Esta secção descreve várias características e funções oferecidas pelo cliente para gestão do sistema EIP.

Acede a algumas características e funções, tais como definições de servidor e gestão de utilizadores, a partir de ícones localizados na área de janela da esquerda do cliente. Pode aceder a outras funções pela barra de acções Ferramentas.

---

### Utilizar os Primeiros Passos do EIP para compreender o cliente de administração de sistema

*Primeiros Passos* é um módulo que vem com qualquer instalação do EIP. Os *Primeiros Passos* fornecem-lhe dados exemplificativos e preenchem objectos para que não necessite de utilizar dados reais. Utilize o módulo *Primeiros Passos* para explorar definições de servidores, utilizadores e grupose outras características que o ajudem a compreender o aspecto, o comportamento e a estrutura de base do cliente de administração.

---

### Administrar o EIP

Como administrador do sistema, pode concluir uma ou mais das tarefas seguintes através do cliente de administração:

- Definir servidores de conteúdos
- Gerir utilizadores e grupos
- Gerir privilégios e níveis de acesso
- Criar modelos de pesquisa federados
- Criar entidades federadas
- Criar subdomínios, se os domínios administrativos estiverem activados.
- Trabalhar com fluxo de trabalho se o fluxo de trabalho estiver activado.
- Criar uma entidade de texto federado em Content Manager Versão 7.

---

### Gerir utilizadores e grupos

É dado acesso aos utilizadores para que pesquisem e trabalhem com documentos em vários servidores de conteúdos criando IDs de utilizador e privilégios. Restringe o acesso aos dados armazenados no sistema definindo e atribuindo privilégios apropriados aos utilizadores.

---

### Utilizar as ferramentas do cliente de administração

Esta secção descreve os componentes fornecidos pelo cliente de administração.

#### Configuração de LDAP

Ao clicar nesta opção, o EIP inicia uma janela que contém quatro separadores:

- Separador LDAP - pode activar a importação de fontes de dados de um servidor LDAP, activar a importação e a autenticação de utilizador de LDAP ou seleccionar ambos.

- Separador Servidor - contém campos para definir especificações de servidor LDAP, incluindo nome do sistema central, o nome de utilizador, o tipo de referência e assim por diante.
- Separador Autenticação - contém campos para definir informações de Secure Sockets Layer.
- Separador Avançadas - configura definições sobre registos máximos e tempo de espera do servidor.

#### **Opção de correlação de utilizador**

Esta opção permite-lhe desactivar a predefinição da Correlação Activada de Utilizador.

#### **Editor de correlação de utilizador fed**

O editor de utilizador federado apresenta uma lista de utilizadores e dá-lhe a opção para definir utilizadores para servidores de conteúdos específicos.

#### **Visualizador de modelos de pesquisa**

O visualizador de modelos de pesquisa fornece informações detalhadas sobre todos os modelos de pesquisa. O visualizador fornece três opções para visualizar detalhes do modelo de pesquisa:

- Correlações Associadas (predefinição) - fornece detalhes sobre as entidades federadas e outros pormenores sobre o modelo de pesquisa
- Modelo de Pesquisa - fornece detalhes sobre o operador predefinido, valores predefinidos, etc.
- Apresentar Resultados - fornece detalhes sobre o nome da visualização, a respectiva largura e ordem de critérios, etc.

#### **Visualizador do inventário do servidor**

Apresenta os inventários do servidor ou servidores seleccionados.

#### **Visualizador de Registo**

Utilize o visualizador de registo para visualizar o registo gerado depois de actualizar o inventário do servidor. O registo mostra uma lista de mensagens quando são encontradas diferenças entre os novos inventários e os anteriores.

#### **Serviços**

Selecione Serviços para activar fluxo de trabalho e/ou information mining.

#### **Domínios administrativos**

Selecione domínios administrativos para activar os mesmos. Os domínios administrativos não podem ser desactivados depois de serem activados.

#### **Editor de tipo MIME**

O editor de tipo MIME faz a listagem das informações que se seguem para cada servidor de conteúdos:

- Classe de Conteúdos
- Extensão de Ficheiros
- Coluna da Base de Dados Relacional (RDB)
- Tipo MIME

Os nomes dos servidores de conteúdos listados no editor de tipo MIME estão abreviados e correspondem à lista de nomes de servidores de conteúdos que surge quando define um novo servidor de conteúdos.

**Sugestão:** DL é a abreviatura para o servidor de conteúdos do Content Manager Versão 7.1. V4 é a abreviatura para o servidor de conteúdos para o Content Manager para AS/400.

Pode adicionar, remover e editar as informações predefinidas no editor de tipo MIME.

#### **MIME para editor de aplicações**

Utilize o MIME para o editor de aplicações para adicionar a, eliminar ou editar os cinco MIME predefinidos para associações de aplicações. Os valores e as definições existentes no MIME para o editor de aplicações exercem influência sobre o visualizador utilizado pelos clientes terminais.

#### **Definição do Tipo de Servidor**

Utilize esta Ferramenta para definir quaisquer servidores personalizados desenvolvidos pelos seus programadores do sistema.

#### **Alterar palavra-passe/ID de DB2**

Selecione esta opção para modificar o ID e palavra-passe de utilizador apenas para ligação do DB2. Este é um ID de utilizador completamente separado do ID de utilizador do administrador.

---

## **Introduzir privilégios**

Esta secção descreve os privilégios do Enterprise Information Portal. Expanda o ícone Autorização para aceder aos quatro privilégios. **Sugestão:** porque pode administrar o Content Manager Versão 8 e o EIP Versão 8 a partir do mesmo cliente, o cliente apresenta todos os privilégios para ambos privilégios relativos a cada lado do cliente.

## **Privilégios**

O cliente de administração fornece privilégios, grupos de privilégios e conjuntos de privilégios predefinidos. Os privilégios concedem aos utilizadores do sistema, tais como utilizadores finais de clientes e administradores, o direito ou direitos de agir de determinada forma sobre um determinado objecto.

#### **Privilégios**

O EIP fornece vários privilégios. Um privilégio é o direito de agir sobre um determinado objecto de uma forma específica. Por exemplo, se tiver utilizadores finais de clientes, pode conceder-lhes os privilégios ItemAdd e ItemDelete que lhes dão o direito de adicionar ou eliminar artigos de um servidor de conteúdos. Para ver os privilégios, expanda o ícone Autenticação e faça duplo clique sobre Privilégios. Para criar privilégios, clique com o botão direito do rato sobre Privilégios e, em seguida, clique sobre Novo.

#### **Grupos de privilégios**

O EIP fornece grupos de privilégios predefinidos. Um grupo de privilégios é um conjunto de privilégios relacionados. Por exemplo, o grupo de privilégios denominado Administer EIP contém cinco privilégios normalmente associados com a administração de um sistema EIP:

- EIPAdminServer
- EIPAdminEntity
- EIPAdminTextEntity
- EIPAdminTemplate
- EIPAdminInfoMining

Para ver ou modificar os privilégios predefinidos para um grupo de privilégios, expanda o ícone Autenticação e faça duplo clique sobre o



nome do grupo de privilégios. Para criar grupos de privilégios, faça clique com o botão direito do rato sobre Grupos de Privilégios e, em seguida, clique sobre Novo.

### **Conjuntos de privilégios**

O EIP fornece vários conjuntos de privilégios predefinidos. Os conjuntos de privilégios são um conjunto de privilégios que definem os papéis do utilizador. Por exemplo, o conjunto de privilégios denominado ClientUserCreateAndDelete contém 17 privilégios associados com papéis de utilizador final de cliente, tais como Delete (eliminar um artigo), ItemAdd (adicionar um artigo) e assim sucessivamente. Se criar um ID de utilizador para um utilizador de cliente e lhe atribuir o conjunto de privilégios ClientUserCreateAndDelete, o utilizador pode iniciar sessão num servidor de conteúdos e desempenhar quaisquer dos 17 papéis de utilizador contidos no conjunto de privilégios. Para ver ou modificar conjuntos de privilégios, clique em Conjuntos de Privilégios e faça duplo-clique sobre um nome de conjunto de privilégios. Para criar conjuntos de privilégios, faça clique com o botão direito do rato sobre Conjuntos de Privilégios e, em seguida, clique sobre Novo.

Ao administrar utilizadores e grupos, associa um grupo de privilégios a um utilizador e/ou a um grupo de utilizadores. Quando atribui conjuntos de privilégios a um grupo de utilizadores, todos os utilizadores do grupo podem desempenhar todos os papéis contidos nos conjuntos de privilégios associados.

---

## **Alternar entre vistas de produto e bases de dados**

Se dispuser do Content Manager e do Enterprise Information Portal como parte da sua solução empresarial, pode aceder a ambos os clientes de administração de sistema a partir de uma interface de utilizador. Anteriormente, se tivesse os dois produtos instalados, teria de abrir dois clientes diferentes. O facto de poder alternar entre uma e outra vista de cliente fornece uma forma cómoda para modificar informações que se apliquem a ambos os clientes e um acesso mais rápido a qualquer um dos produtos.

Para alternar de EIP de administração para Content Manager de administração sem terminar a sessão, active a janela principal de administração do sistema e utilize o menu pendente por cima da área esquerda da janela e seleccione Content Manager.

Para comutar entre bases de dados federadas, vá para a área esquerda da janela de cliente e faça duplo-clique sobre o ícone de uma base de dados federada.

Também pode administrar bases de dados diferentes sem sair do cliente e iniciar sessão na nova base de dados. O cliente de administração apresenta um ícone para todas as bases de dados de administração listadas no ficheiro cmbds.ini. Para alternar para bases de dados diferentes, faça clique no ícone. Se a nova base de dados tem um ID de utilizador diferente daquele que introduziu quando iniciou sessão no cliente, ser-lhe-á solicitado que introduza um ID de utilizador diferente.

---

## **Melhoramentos e aperfeiçoamentos no cliente de administração**

O EIP Versão 8.2 apresenta melhoramentos significativos do cliente de administração do EIP, incluindo:

### **Assistentes e diálogos melhorados**

Os novos diálogos tornam a gestão de utilizadores mais fácil. Os novos



assistentes tornam a definição e modificação de entidades federadas e modelos de pesquisa mais simples. Utilizadores podem ainda escolher os diálogos suportados pelo EIP Versão 7.1.

#### **Cliente administrativo partilhado**

Ao instalar o EIP Versão 8.2 e o Content Manager Versão 8.2 no mesmo sistema, os dois produtos partilham um cliente administrativo. Se for administrador de ambos os produtos, vai iniciar sessão no cliente uma vez e alternar entre as duas aplicações no cliente. Pode também alternar entre bases de dados de administração sem ter de sair e entrar.

#### **Administradores de domínio**

É possível criar administradores de domínio que apenas tenham privilégios de administração para um domínio definido.

#### **Início de Sessão único e suporte de LDAP**

Agora o EIP utiliza o Windows Active Directory e LDAP para permitir aos utilizadores um acesso com um único início de sessão a vários servidores de conteúdos.

---

## **Ligar um cliente de administração a uma base de dados de administração local**

Se instalar uma base de dados de administração no mesmo servidor onde instala o cliente de administração, a informação necessária para ligar o servidor e o cliente local já se encontra armazenada em `cmbds.ini`, um ficheiro que armazena informações de ligação da base de dados. Não será necessário executar qualquer configuração após a instalação e pode estabelecer ligação de imediato, seguindo os passos indicados nesta secção. **Requisito:** Se criar bases de dados locais adicionais utilizando o utilitário EIP Database Install, tem de modificar manualmente o ficheiro `cmbds.ini` com a informação necessária para estabelecer a ligação com a nova base de dados.

1. Clique em **Iniciar-->Programas-->Enterprise Information Portal for Multiplatforms 8.2-->Administração**
2. Selecciona a base de dados local na lista pendente no campo Servidor.
3. Escreva o ID de utilizador e a palavra-passe de administrador e prima OK.
4. Abre-se o cliente de administração do sistema. **Sugestão:** Se utilizou os Primeiros Passos do EIP, as bases de dados de exemplificação são apresentadas na área esquerda da janela do cliente.

---

## **Ligar o cliente de administração a uma base de dados de administração remota**

Existem duas formas de ligar um cliente de administração de EIP a uma base de dados AIX, Windows ou Solaris remota:

- Estabelecer ligação através de um servidor RMI (consultar *Planeamento e Instalação do Information Integrator for Content* para obter mais informações).
- Definir uma ligação por catalogação da base de dados utilizando o Assistente de Configuração do DB2 seguida de definição dos parâmetros de ligação utilizando o Utilitário de Configuração do Servidor EIP. O utilitário copia informações, como, por exemplo, o sistema operativo, o nome alternativo e o nome de esquema da base de dados, para um ficheiro de nome `cmbds.ini`. Quando inicia o cliente de administração do sistema, a lista de servidores com os quais pode iniciar sessão é compilada a partir dos servidores definidos em `cmbds.ini`.

**Requisito:** É necessário catalogar cada base de dados remota separadamente. As bases de dados remotas têm de estar listadas no ficheiro cmbds.ini para que possa ser estabelecida a sua ligação ao cliente de administração.

**Sugestão:** Um utilizador mais experiente pode ignorar os passos do Utilitário de Configuração do Servidor e modificar o ficheiro cmbds.ini num editor de texto. O caminho predefinido para cmbds.ini é C:\Program Files\IBM\CMgmt.

**Importante:** Se a pessoa que instalou o produto já tiver configurado os valores de catálogo de base de dados para a base de dados remota com a qual pretende estabelecer ligação, não será necessário executar os passos do DB2 CCA para essa base de dados. No caso de o instalador não ter introduzido os valores de catálogo da base de dados ou se pretender estabelecer ligação com uma base de dados remota adicional, tem de utilizar o DB2 CCA e modificar o ficheiro cmbds.ini com os parâmetros de ligação para a(s) base(s) de dados adicional(ais).

## **Passo 1 - catalogue a base de dados remota utilizando o Assistente de Configuração do DB2**

O Assistente de Configuração (CCA) do DB2 cataloga a base de dados remota de EIP em DB2. Para catalogar a base de dados remota utilizando o DB2 CCA é necessário conhecer o nome de sistema central do servidor remoto, o nome da base de dados e o número de porta de instância da base de dados e definir um nome alternativo para a base de dados remota.

Os passos 1a - 1f explicam como localizar o número da porta de ligação, o nome de esquema e o nome da base de dados. É necessário conhecer os nomes e o número da porta de ligação para configurar uma ligação entre o cliente de administração e uma base de dados remota.

1. Localizar as informações de ligação da base de dados remota:
  - a. Inicie sessão no servidor AIX, Windows ou Solaris remoto com um ID de utilizador que tenha autoridade administrativa do DB2.
  - b. Escreva `db2 list db directory`
  - c. Seleccione o nome da base de dados de administração com a qual pretende estabelecer ligação. Tome nota da instância de db2 na qual a base de dados está instalada, uma vez que instâncias distintas podem ter números de porta de ligação distintos.
  - d. Escreva `db2 connect to <base de dados> user <ID de utilizador> using <palavra-passe>`
  - e. Escreva `db2 list tables` e tome nota do nome de esquema da base de dados (requerido pelo utilitário de configuração do servidor).
  - f. Localizar o número da porta de ligação associado com a base de dados de administração remota:

Em Windows:

- 1) Abrir o DB2 Control Center no servidor Windows remoto.
- 2) Faça clique com o botão direito do rato sobre uma das instâncias disponíveis para a máquina local.
- 3) Seleccione "Setup Communications...."
- 4) Seleccione o botão "Properties" localizado à direita da selecção de TCP/IP. O número de porta surge listado na janela.

Em AIX ou Solaris

- 1) Escreva `cd /usr/etc`

- 2) Escreva cat services
  - 3) Percorra a lista de serviços até encontra o número da porta de ligação da instância de base de dados da base de dados remota. Por exemplo, se a base de dados estiver instalada em db2inst1, a porta de ligação poderá ser 50000.
  - 4)
2. Utilize o Assistente de Configuração do DB2 para catalogar a base de dados remota. Consulte os ficheiros de ajuda do DB2 CCA para obter mais informações.
    - a. Inicie sessão no servidor Windows onde o cliente de administração se encontra instalado. É necessário iniciar sessão com um ID de utilizador com privilégios DB2ADM totais.
    - b. Direcione-se para o Assistente de Configuração do DB2 a partir do menu Iniciar-->Programas.
    - c. Para catalogar e testar a ligação à base de dados remota, siga os pedidos de informação do Assistente de Configuração do DB2.
    - d. Se o teste da ligação efectuado pelo DB2 CCA for bem sucedido, siga os passos indicados em “Passo 2 - utilize o Utilitário de Configuração do Servidor” ou modifique o ficheiro cmbds.ini directamente de modo a definir os parâmetros de ligação da base de dados remota armazenados em cmbds.ini

## Passo 2 - utilize o Utilitário de Configuração do Servidor

O utilitário de configuração do servidor solicita informações de conectividade (número da porta, nome do sistema central, etc.) referentes à base de dados remota e armazena estes dados em cmbds.ini.

1. Clique em **Iniciar-->Programas-->IBM Enterprise Information Portal for Multiplatforms-->Utilitário de Configuração do Servidor**.
2. Introduza as informações nos campos (consulte a Tabela 4).

Tabela 4. Utilitário de configuração do servidor

| Campo            | Informação  | Notas  |
|------------------|---|--|
| Servidor         | Selecione o tipo de base de dados, Content Manager ou EIP.  | Servidor refere-se ao tipo de base de dados e não ao nome do servidor onde a base de dados se encontra instalada. <b>Sugestão:</b> Só pode utilizar o cliente de administração para gerir ambos os tipos de base de dados se o seu sistema incluir clientes de administração Content Manager e EIP na mesma máquina. |
| Nome do servidor | Introduza o nome alternativo da base de dados com a qual está a estabelecer ligação. Requisito: Tem de utilizar o mesmo nome alternativo definido no DB2 CCA. | Um nome alternativo constitui um nome exclusivo de identificação da base de dados remota na sua estação de trabalho. Os nomes alternativos têm um limite de oito caracteres. Por exemplo, se o nome da base de dados remota for ICMNLSDB, um nome alternativo possível seria REMOTE1.                                |

Tabela 4. Utilitário de configuração do servidor (continuação)

| Campo                          | Informação  | Notas   |
|--------------------------------|---|---|
| Nome de esquema                | Introduza o nome de esquema atribuído quando a base de dados remota foi criada.                             | ICMADMIN é o nome de esquema predefinido para as bases de dados de EIP e Content Manager.   |
| Nome do sistema central        | Introduza o nome do computador onde a base de dados remota foi instalada.                                   | Introduza o nome completo do sistema central ou um endereço IP do computador onde a base de dados remota se encontra instalada.   |
| Sistema operativo              | Selecione um sistema operativo na lista pendente.   | Selecione AIX, Sun Solaris ou Windows. A opção OS/390 não se encontra funcional no EIP 8.2.   |
| Número da porta                | Introduza o número de porta atribuído à base de dados remota.   | 50000 é o número de porta de ligação predefinido para as bases de dados de EIP e Content Manager instaladas em Windows, AIX e Solaris.  |
| Nome da base de dados remota   | Introduza o nome da base de dados remota. Utilize letras maiúsculas.  | ICMNLSDDB é o nome predefinido para as bases de dados de EIP e Content Manager.   |
| Nome do nó                     | Introduza o nome do nó da base de dados de EIP ou Content Manager remota.                                   | O nome do nó é um nome exclusivo atribuído à base de dados remota, semelhante ao nome alternativo que é criado para a base de dados remota. Para localizar o nome do nó de uma base de dados instalada num servidor Windows, AIX ou Solaris: <ul style="list-style-type: none"> <li>a. Abra uma sessão de linha de comandos de db2.</li> <li>b. No pedido de informação db2=&gt;, escreva LIST NODE DIRECTORY</li> <li>c. O DB2 apresenta os nomes dos nós e outros dados referentes a todas as bases de dados instaladas ou definidas no servidor remoto.</li> </ul> |
| Activar início de sessão único | Clique nesta opção se o início de sessão único foi activado durante a instalação da base de dados.          | A predefinição é não assinalado (desactivado).  |
| Opções de segurança            | Clique em autenticação de cliente se esta opção tiver sido seleccionada durante a criação da base de dados. | A predefinição é Servidor.  |

3. Faça clique sobre OK.

### Passo 3 - testar a ligação da base de dados remota

1. Inicie sessão no servidor Windows onde o cliente de administração se encontra instalado.

2. Clique em **Iniciar-->Programas-->Enterprise Information Portal for Multiplatforms 8.2-->Administração**.
3. Seleccione o nome alternativo da base de dados remota na lista pendente no campo Servidor. O nome corresponde ao nome alternativo definido no Utilitário de Configuração do Servidor e no Assistente de Configuração do DB2.
4. Introduza o ID de utilizador e a palavra-passe associados com a base de dados remota.
5. Faça clique sobre OK. O cliente de administração é aberto.

---

## Definir tipos de documentos

O EIP fornece suporte de visualizador para alguns tipos de documentos. Se definir um tipo de documento para o servidor, pode iniciar documentos nas respectivas aplicações nativas. Por exemplo, se estiver a armazenar documentos Lotus Word Pro no servidor OnDemand do Content Manager, pode definir o EIP para iniciar documentos com a extensão .lwp no Lotus Word Pro em vez do visualizador de documentos do cliente.

Para definir um tipo de documento, modifique o ficheiro `cmbcc2mime.ini` no directório CMBROOT, normalmente `x:\Program Files\IBM\CMgmt`. O ficheiro contém instruções para o desenvolvimento de definições MIME personalizadas. O ficheiro converte classes de conteúdos numa sequência de tipo MIME, para que o cliente possa ler os conteúdos a partir dos servidores de conteúdos.

**Importante:** quando inicia uma aplicação com base no tipo MIME, só será apresentado o objecto base. Qualquer markup que tenha sido utilizada no documento é apresentada. Se o documento tiver várias partes, apenas é mostrada a primeira parte. O tipo MIME nos dois ficheiros deve ser correspondente.

## Alterar o ficheiro de tipo MIME do servidor (cmbcc2mime.ini)

Quando adicionar tipos MIME de servidor, verifique se o tipo de documento que está a adicionar é um tipo MIME criado para esse ficheiro. Para obter mais informações, consulte o site na Web: <ftp://ftp.isi.edu/in-notes/iana/assignments/media-types>.

Para adicionar valores ao ficheiro `cmbcc2mime.ini`, conclua os seguintes passos:

1. Abrir `cmbcc2mime.ini` num editor de texto.
2. Utilize um seguinte formato para valores definidos pelo utilizador:
  - A classe de conteúdo inicia em 4096
  - O sinal de igual (=) segue o valor de classe do conteúdo
  - O tipo MIME deve seguir o sinal de igual. Se não se tratar de um tipo MIME standard para essa classe de conteúdo, proceda do seguinte modo:
    - a. Um tipo MIME é composto de um tipo e de um subtipo. Os tipos válidos são aplicação, texto, imagem, modelo, mensagem, áudio e vídeo.
    - b. Uma barra (/) segue o tipo
    - c. Para criar o subtipo, o sinal (x-) deve preceder o sinal utilizado para esse documento, por exemplo  
`x-minhaclasseedocumento(4096=aplicação/x-minhaclasseedocumento)`
  - Repetir 2b e 2c conforme necessário para todo o novo tipo de MIME.

**Sugestão:** Os servidores de conteúdos OnDemand correlacionam extensões de ficheiros em vez de valores numéricos de classes de conteúdos para sequências de tipo MIME.

---

## Capítulo 3. Utilizar as características do cliente de administração de EIP

Esta secção explica algumas das tarefas comuns realizadas pelos administradores de EIP.

---

### Criar uma pesquisa associada

Uma pesquisa federada é uma consulta lançada a partir de uma aplicação cliente que pesquisa mais que um servidor de conteúdo em simultâneo. O EIP fornece as ferramentas que criam modelos de pesquisa para pesquisas federadas. Uma vez que cada servidor de conteúdo armazena e organiza as informações de modo diferente, o modelo de pesquisa deve ter em conta estas diferenças para cada servidor. O modelo de pesquisa efectua o correlacionamento de entidades federadas e dos seus atributos federados a atributos nativos de modo efectuar pesquisas nos servidores de conteúdos.

A criação de pesquisas federadas envolve:

- Definir ligações a servidores de conteúdos utilizando conectores EIP
- Criar entidades federadas
  - Definir entidades federadas
  - Criar atributos federados
  - Correlacionar atributos federados a atributos nativos
  - Atribuir parâmetros
- Criar modelos de pesquisa
  - Definir o modelo de pesquisa
  - Definir o critério de pesquisa
  - Definir as definições do modelo
  - Atribuir acesso a utilizadores cliente

Os dois assistentes, fornecidos com o EIP Versão 8.2, tornam a criação de entidades federadas e modelos de pesquisa mais fácil. O assistente de entidade federada inclui um inventário de servidor que pode ser filtrado para facilitar a localização de atributos nativos. Gera também parâmetros predefinidos válidos para atributos federados, reduzindo a possibilidade de errar na sua configuração. O assistente de modelo de pesquisa ajuda a criar critérios de pesquisa. Também ajuda a definir como é que os critérios de pesquisa e apresentação de resultados serão e agirão. Até fornece uma pré-visualização que como o modelo de pesquisa será na aplicação de cliente. Adicionalmente, os diálogos para criar entidades federadas e modelos de pesquisa para Versão 7.1 EIP estão também disponíveis para aqueles que os preferirem.

Todos os assistentes, diálogos e campos estão documentados na ajuda on-line do EIP.

---

## Definir servidores

Tem de definir o servidor antes de efectuar a ligação a um servidor e executar um inventário do servidor. Ao clicar com o botão direito do rato no ícone do Servidor e, de seguida, clicar em Novo, o cliente apresenta todos os conectores suportados pelo EIP. Antes de definir um servidor, deve saber algumas informações básicas sobre os conectores:

- Quais os conectores seleccionados pela pessoa que efectuou a instalação? Os conectores instalados estão listados no ficheiro de configuração `cmbcs.ini`. Num servidor Windows, o caminho predefinido é `x:\Program Files\IBM\CMgmt`. Pergunte ao seu administrador AIX ou Solaris para a localização de ficheiros `cmbcs.ini`.
- A pessoa que efectuou a instalação seleccionou uma opção de conector local ou remoto? O ficheiro `cmbcs.ini` contém os tipos de conector Local e Remoto.
- Se o seu sistema estiver configurado para RMI, o servidor RMI está iniciado? Para iniciar RMI no servidor RMI local, utilize **Iniciar→Programas→IBM Enterprise Information Portal for Multiplatforms 8.2→Iniciar servidores RMI**. Se o seu sistema utiliza RMI remoto, procure em `cmbsvclint.ini` para encontrar o servidor remoto onde os conectores RMI estão instalados. Peça ao administrador do servidor RMI mais informações.
- Se a pessoa que instalou EIP incluiu o CM para conector AS/400, que informações foram incluídas na tabela de rede designada `frnolint.tbl`? O AS/400 `frnolint.tbl` está em `%CMBROOT%`.
- Se estiver a definir servidores de conteúdos remotos que contenham bases de dados relacionais, como Content Manager Version 8 e DB2, DataJoiner e Catálogo de Informações, tem de catalogar ou adicionar a base de dados a partir da estação de trabalho onde está a utilizar o cliente.

A lista que se segue fornece passos gerais que deve seguir ao definir um servidor:

1. Faça clique com o botão direito do rato sobre Servidores e seleccione **Novo**.
2. Selecciona um servidor da lista. Surge a janela Novo Servidor.
3. Insira o nome e a descrição do servidor no campo Nome do Servidor no separador Geral. Para alguns servidores, necessita de inserir apenas o nome da base de dados. Para outros servidores, tem de inserir o nome completo do servidor onde a base de dados foi instalada.
4. Indique parâmetros de inicialização, se necessário. Alguns servidores necessitam de parâmetros de inicialização, tais como cadeias de ligação e de configuração. Outros servidores necessitam apenas do nome da base de dados.
5. Clique em Testar Ligação ao Servidor. O EIP inicia sessão nalguns servidores utilizando o ID de utilizador e a palavra-passe que introduziu para iniciar o cliente de administração. Se o servidor necessitar de um ID de utilizador ou palavra-passe diferentes, o EIP solicita-lhe que introduza um ID de utilizador e uma palavra-passe válidos, específicos para o servidor de conteúdos que está a definir.

**Sugestão:** Pode também definir um tipo de servidor de conteúdos que não é um dos tipos de servidor predefinidos, mas tem de fornecer as classes de conector de Java ou C++ e a classe de definição de servidor para o novo tipo de servidor. Também necessita do conector Java para executar o inventário do servidor. Para obter instruções sobre a adição de servidores de conteúdos, consulte o *Manual de Programação de Aplicações de Estações de Trabalho* e o online API reference.



Se a configuração para o servidor de conteúdo não tiver sido efectuada com sucesso, consulte o manual *Mensagens e Códigos* para obter mais informações sobre a detecção e resolução de problemas para esta situação ou sobre a mensagem de erro recebida.

Pode também consultar o administrador do servidor ao qual pretende ligar para obter mais assistência.

## Directrizes para definir servidores

Esta secção fornece as directrizes para o ajudar a executar a definição inicial do servidor.

### Efectuar a ligação às bases de dados de DB2 (relacional)

Esta secção aplica-se a servidores DB2, DataJoiner, JDBC, ODBC, Catálogo de Informações e Content Manager Versão 7 e Versão 8.

- **Importante:** Deve catalogar cada base de dados de DB2 antes de definir o servidor. Pode utilizar o DB2 CCA para catalogar as bases de dados ou pode utilizar uma Linha de Comandos de DB2. Contacte o administrador de DB2 para mais informações.
- No campo do Nome do servidor no separador Geral, deve inserir o nome da base de dados à qual pretende estabelecer ligação. Use letras maiúsculas ao escrever o nome do servidor.
- Ao definir DB2, DataJoiner, JDBC, ODBC e Catálogo de Informações, clique no separador Parâmetros de Inicialização e insira o nome do Esquema associado às tabelas de bases de dados às quais está a estabelecer ligação, por exemplo SCHEMA=ICMADMIN.
- Quando define um servidor Content Manager Versão 7.1 ou Versão 8.2, apenas lhe é solicitado que insira o nome da base de dados. Não altere as predefinições no separador de Parâmetros de Inicialização.
- Quando define um servidor Content Manager Versão 7.1, tem de ter uma tabela de rede denominada `fronlnt.tbl` na unidade local em `x:\CMBROOT`. A tabela de rede contém o nome do sistema central, o número de porta e informações sobre o tipo de servidor que o EIP necessita para localizar e iniciar sessão no servidor de Bibliotecas do Content Manager Versão 7.1. Se definir servidores múltiplos do Content Manager Versão 7.1, cada servidor deve ter uma entrada em separado no ficheiro `frnolnt.tbl` antes de definir o servidor.
- Para estabelecer ligação à DB2 DataJoiner, certifique-se que o método de autenticação para Enterprise Information Portal está definido como servidor para a instância de base de dados definida em DB2 Universal Database.
- Para efectuar a ligação ao DataJoiner 2.1, tem de descarregar um programa de associações do site da Web do DataJoiner e associar a base de dados DataJoiner antes de poder definir um servidor DataJoiner.

### Estabelecer a Ligação a um servidor de Pesquisa de Texto

Para definir um Servidor de Pesquisa de Texto, primeiramente tem de definir o servidor do Content Manager Versão 7.1 associado ao Servidor de Pesquisa de Texto.

Vai inserir o nome do servidor de Pesquisa de Texto na opção "Seleccionar o nome do servidor associado do Content Manager Versão 7.1" numa lista pendente. Esta lista está no separador Servidor Associado.

O servidor do Content Manager Versão 7.1 e o servidor de Pesquisa de Texto devem estar activos e em execução antes da ligação do EIP àqueles.

## Estabelecer ligação a vários servidores Content Manager for AS/400

Se utiliza mais do que um servidor de AS/400, terá de definir os servidores adicionais na tabela de rede. A tabela de rede (`frnolint.tbl`) está localizada em `x:\<cmbrroot>`. Para o novo servidor, introduza o nome do servidor, o tipo de ligação (por exemplo, TCP/IP), o nome do sistema central, a porta e o tipo de servidor. Para o primeiro servidor, os tipos de instaladores em Servidor, Nome de Sistema Central e valores de Porta durante a instalação para criar `frnolint.tbl`.

O que se segue é um exemplo típico de informações armazenadas em `frnolint.tbl` :

```
/* VI/400 Network Table */
SERVER: VI400 REMOTE TCPIP
      HOSTNAME = vi400
      PORT     = 29000
      SERVER_TYPE = FRNLS400
```

## Configurar o conector do Extended Search

As informações que inserir para definir um servidor de Extended Search dependem de dois factores:

- O tipo de servidor Web no qual o Servidor Extensivo foi instalado - Domino Web Server, WebSphere, IIS.
- O número de porta definido para o Servidor Web onde foi instalado o Extended Search.

Ao definir um conector de Extended Search, siga os passos seguintes:

1. No campo de Nome do Servidor no separador Geral, insira o nome totalmente qualificado do sistema central do servidor Web onde foi instalado o Extended Search.
2. No separador Parâmetros de Inicialização, insira 80 no campo do Número de Porta caso a pessoa que efectuou a instalação tenha seleccionado as predefinições para o número de porta do Servidor Web, ao instalar o Extended Search.
3. No campo do ID de Aplicações, escreva Demo. Escreva o nome tal como é mostrado.
4. No campo da Palavra-passe, escreva Demo.
5. No campo de Parâmetros Adicionais:
  - a. Não altere os dois ponto e vírgula se souber que o Extended Search foi instalado num Servidor de Web Domino e que a pessoa que efectuou a instalação utilizou as definições do número de porta para o Servidor da Web e para os números de portas de Extended Search predefinidas durante a instalação.
  - b. Consulte a secção abaixo para obter informações sobre como modificar o campo de Parâmetros Adicionais para servidores do Extended Search que foram instalados utilizando definições personalizadas.

Se tiver configurado o conector de Extended Search com o servidor de aplicações WebSphere ou se a porta do Extended Search NÃO for 6001 ou se o nome do servidor de Extended Search NÃO for igual ao nome do servidor de web, é necessário configurar o conector para localizar os caminhos relativos correctos do servlet do ES, do número de porta do ES e do nome do servidor do ES.

Se tiver configurado o conector de Extended Search com WebSphere ou se a porta do Extended Search NÃO for 6001 ou se o nome do servidor de Extended Search

NÃO for igual ao nome do servidor de web, pode criar um ficheiro de configuração, por exemplo `desclient.cfg`, seguindo os passos indicados abaixo:

Defina como directório aquele onde estão localizadas as aplicações ou exemplos. Crie um ficheiro de configuração; por exemplo, `desclient.cfg`. Este ficheiro não é fornecido com o Enterprise Information Portal.

No ficheiro `desclient.cfg`, adicione `DESHOSTNAME=(nome de sistema central de ES)`

Se tiver configurado o ES com o servidor de aplicações Domino, defina `DESREQURI=/servlet/ESAdmin`. Se tiver configurado o ES com o servidor de aplicações WebSphere, defina `DESREQURI=/lotuskms/ESAdmin` onde *servlet* é o caminho do directório no servidor HTTP que suporta o conector de Extended Search.

Se o servidor de aplicações for WebSphere, o `DESREQURI` deve ser `/lotuskms/ESAdmin` em vez de `/servlet/ESAdmin`

Caso pretenda pesquisar fontes de PE a partir de clientes com Poucos/Muitos Recursos, defina um parâmetro adicional denominado

`"DESCFGPATH=<caminho absoluto de desclient.cfg>"`

nas caixas de diálogo de definição do servidor DES do cliente de administração.

Caso pretenda executar exemplos de PE, passe o caminho absoluto de `desclient.cfg` nos argumentos da linha de comandos.

Exemplo 1:

```
TConnectDES es.stl.ibm.com utilizador palavra-passe  
PORT=80;DESAPPID=Demo;DESAPPPW=palavra-passe;DESCFGPATH  
=<caminho absoluto de desclient.cfg>;
```

Exemplo 2:

```
java TConnectDES es.stl.ibm.com utilizador palavra-passe  
PORT=80;DESAPPID=Demo;DESAPPPW=palavra-passe;DESCFGPATH=<caminho absoluto de  
desclient.cfg>;
```

## Definir o servidor do Catálogo de Informações

Deve catalogar o servidor do Catálogo de Informações antes de definir o servidor. Digite o nome do servidor, por exemplo, `SAMPLE1`, no campo Nome de Servidor. No separador de parâmetros de Inicialização, digite `SCHEMA=<Nome do esquema associado com EXEMPL01>`.

## Definir o servidor OnDemand

O servidor OnDemand e o daemon do Servidor de Bibliotecas devem estar em execução antes de poder definir um servidor OnDemand. Pode efectuar um ping ao servidor OnDemand antes de definir o servidor no EIP para verificar se o servidor e o daemon estão em execução.

No separador Geral, insira o nome totalmente qualificado do sistema central do servidor onde foi instalado o OnDemand.

No separador Parâmetros de Inicialização, introduza o número de porta que foi atribuído quando o servidor OnDemand foi instalado. Se a pessoa que instalou OnDemand seleccionou o valor de porta predefinido 0 durante a instalação do OnDemand, digite 0 no campo número de porta. Se a pessoa que efectuou a

instalação seleccionou um número de porta diferente, insira esse número de porta precedido de um #. Por exemplo, # 5000 pode ser um número de porta alternativo escolhido para OnDemand num servidor Windows.

Se estiver a definir um servidor OnDemand que foi instalado num servidor AS/400 que execute software da Versão 4, deve inserir a seguinte informação no campo de Parâmetros Adicionais: STATECONNECT=#1.

Se estiver a definir um servidor OnDemand que foi instalado num servidor OS/390 que execute software da Versão 2.1, insira o número de porta personalizado designado na altura em que o OnDemand foi instalado no servidor OS/390 Versão 2.1.

O OnDemand necessita que seja mantido um socket em actividade durante a ligação.

### **Definir o servidor da pesquisa de texto**

Para definir um Servidor de Pesquisa de Texto, deve definir primeiro o servidor Content Manager Versão 7 associado ao Servidor de Pesquisa de Texto.

Insira o nome do servidor da Pesquisa de Texto no campo do Nome do Servidor no separador Geral. Seccione o servidor associado do Content Manager Versão 7.1 a partir de uma lista pendente no separador Servidor Associado.

O servidor do Content Manager Versão 7.1 e o servidor de Pesquisa de Texto devem estar activos e em execução antes da ligação do EIP àqueles.

### **Definir o servidor Domino.Doc**

No campo do nome do servidor, insira o caminho para o nome do servidor e nome da biblioteca do servidor Domino.Doc. Por exemplo: oakley/DominoDoc1/Lib.nsf.

Se estiver a utilizar conectores locais, tem de instalar o Domino Doc Desktop Enabler na estação de trabalho que tem o cliente EIP. Se estiver a utilizar RMI, tem de instalar o Domino Doc Desktop Enabler no servidor RMI. O Domino Doc Desktop Enabler tem de ter a mesma versão do servidor Domino Doc.

Não altere os dois pontos e vírgula no separador dos Parâmetros de Inicialização.

### **Definir o servidor ImagePlus para OS/390**

Ao definir um servidor ImagePlus para OS/390, tem de obter os parâmetros seguintes para efectuar a ligação ao servidor. A lista abaixo contém valores exemplificativos:

- Número de Porta FAF: 3061
- ID de Aplicação FAF: 01
- Protocolo FAF: 4000
- Endereço de IP FAF: 9.67.43.83
- CICS do Gestor de Distribuição de Objectos: 4000
- Endereço de IP do Gestor de Distribuição de Objectos: 9.67.43.83
- Número de Porta do Gestor de Distribuição de Objectos: 3082
- ID do Terminal Gestor de Distribuição de Objectos: *deixe este campo em branco*
- Parâmetros Adicionais: FAFSITE=CS61;;

### **Utilizar rastreio com o Content Manager ImagePlus for OS/390**

O rastreio pode ajudar a resolver problemas se não conseguir estabelecer ligação ao servidor Content Manager ImagePlus para OS/390. Se instalar o conector para

Content Manager ImagePlus for OS/390, pode activar o rastreio de ImagePlus for OS/390 modificando o ficheiro eypapi.ini localizado em cmbroot.

O ficheiro eypapi.ini contém as seguintes linhas:

```
; Caminho onde os ficheiros IPFAF estão armazenados
; (NÃO PODE ter um '\\' final)
; -- a predefinição é <ROOT Directory>\
;
IPFAFPath=d:\cmbroot
; Sinalizar para Registo (ficheiros EYPLmdd.LOG)
; -- a predefinição é Registo OFF (0)
; -- 0 Todos os Registos OFF
; -- 1 Ficheiros de registo criaram apenas condições de erro registadas
; -- 2 Ficheiros de registo criaram todas as condições registadas
;
Registo = 0

;-----
;
; Sinalizar para Registrar os Tipos de Parâmetros FAF criados por APIs
; -- a predefinição é Registo OFF (0)
; -- 0 Tipos de parâmetros Não registados
; -- 1 Registrar Tipos de Parâmetros Faf
;
FafTypeLogs = 0
```

#### **IPFAFPath**

Especifica o directório onde os ficheiros são escritos. Os ficheiros de registo têm os seguintes nomes:

EYPmdd.LOG

em que *mdd* é o mês e o dia em que o ficheiro de registo foi criado.

#### **Registo**

Especifica quando um ficheiro de registo é criado.

- 0** Não regista. A predefinição é 0.
- 1** Os ficheiros de registo criados apenas contêm condições de erro.
- 2** Os ficheiros de registo criados contêm todas as condições.

#### **FafTypeLogs**

Especifica o registo dos tipos de parâmetros FAF criados pelas APIs.

- 0** Não regista tipos de parâmetros; a predefinição é 0/.
- 1** Regista tipos de parâmetro FAF.

## **Trabalhar com o conector OnDemand: Ajuste e sockets TCP/IP**

Um problema conhecido de Windows pode afectar o rendimento ao efectuar a ligação a um servidor OnDemand. Durante repetidas pesquisas e obtenções num servidor OnDemand, muitos dos sockets Windows serão abertos e fechados. Duas predefinições de Windows poderão ter impacto no tráfego pesado entre o EIP e um servidor OnDemand:

- Quando uma aplicação fecha um socket do Windows, o Windows coloca a porta dos sockets no estado TIME\_WAIT durante 240 segundos; durante este tempo a porta não pode ser reutilizada.
- O Windows limita o número de portas que uma aplicação pode utilizar a 5000.

Para evitar os problemas que possam ocorrer, altere os valores do tempo de espera excedido e do número de portas que utilizam o editor de registos do Windows registry editor.

- Altere o valor do tempo de espera excedido de 240 segundos para um número mais baixo (o intervalo válido é entre 30-300 segundos). O nome das chaves é HKEY\_Local\_Machine\System\CurrentControlSet\services\Tcpip\Parameters\TcpTimedWaitDelay.
- Aumente o número máximo de portas da sua predefinição de 5000 para um número mais alto (o intervalo válido é entre 5000-65534). O nome das chaves é HKEY\_Local\_Machine\System\CurrentControlSet\services\Tcpip\Parameters\MaxUserPort

Para obter mais informações sobre o TcpTimedWaitDelay e MaxUserPort, consulte a documentação do Windows.

## Trabalhar com o conector de Extended Search

Esta secção descreve uma alteração ao conector do Extended Search no EIP Versão 8.2.

O locale especificado pelo utilizador é suportado passando o valor do locale na chave DESLOCALE. Pode passar este valor par na linha de comandos se invocar o conector ES directamente. Pode definir este valor nos argumentos **Adicionais** para as propriedades ES.

**Sugestão:** O software do servidor Extended Search está incluído na caixa de produto EIP Versão 8.2.

---

## Criar entidades federadas

Após ter definido as ligações a servidores de conteúdos, o próximo passo na criação de pesquisas federadas é criar entidades federadas, que se tornam nos blocos de construção para modelos de pesquisa. Esta secção explica o que são entidades federadas e como utilizar o assistente de Criação de Entidade Federada.

## Compreender entidades federadas

A maioria das vezes, os utilizadores de aplicações cliente não querem pesquisar informações numa base de servidor-por-servidor. Em vez disso, estes querem conduzir uma pesquisa federada única. Os modelos de pesquisa permitem aos utilizadores de aplicações cliente agrupar as suas pesquisas numa única consulta. Como administrador do EIP, pode criar estes modelos de pesquisa para utilização em aplicações cliente. Antes de criar o modelo de pesquisa, deve criar primeiro entidades federadas que correlacionam os seus atributos federados a atributos nativos em servidores de conteúdos.

Por exemplo, o DB2 armazena informações em tabelas cujas colunas representam atributos das informações armazenadas numa tabela. Uma tabela com o nome Customer\_Demographics pode conter colunas como Nome, Número\_Pol, Endereço, Telefone e Ocupação.

Por outro lado, o Content Manager, utiliza artigos, tipos de artigos e atributos em vez de tabelas e colunas. As informações armazenadas no DB2 podiam também ser armazenadas num nome de entidade CustInfo. Os atributos poderiam ser NomeDespesa, Conta, EndereçoCasa, TelefoneCasa e Emprego. Em ambos os casos, são armazenadas informações idênticas e identificadas de modo diferente.

O EIP resolve o problema de ter de se considerar todas as diferentes formas como os servidores de conteúdos armazenam informações idênticas. As entidades federadas mantêm um registo destas informações. Na realidade, uma entidade federada não armazena dados; armazena metadados sobre como cada servidor de conteúdo armazena dados. Quando cria uma entidade, são correlacionados todos os seus atributos a atributos nativos correspondentes nos servidores de conteúdos que pretende consultar.

Para o exemplo dado acima, podia criar uma entidade federada com o nome Info Apólice, com os atributos federados Nome\_Apólice, Num\_Apólice, End\_Casa e Cargo. Pode então correlacionar os atributos federados a cada um dos atributos nativos correspondentes.

O EIP pode gerar um inventário de servidor que contém estas informações. Um inventário de servidor contém estas informações e o assistente Criar Entidades Federadas permite obter um inventário de servidor que pode ser filtrado nos servidores de conteúdos. A filtragem está disponível apenas se utilizar o assistente Criar Entidades Federadas. A opção de filtragem não está disponível se utilizar o método manual (sem assistente) para criar uma entidade federada. Uma vez gerado o inventário do servidor, pode então começar a correlacionar os atributos federados a atributos nativos.

Não é suficiente correlacionar apenas atributos federados a atributos nativos. Cada atributo nativo também pode ter propriedades diferentes. Os atributos podem ser do tipo (1) anulável, (2) pesquisável, (3) actualizável e (4) pesquisável por texto. Dependendo do tipo de dados que seleccionar, pode também ter opções respeitantes ao comprimento, precisão, escala e valores mínimos e máximos dos dados.

Quando define estas propriedades, não pode torná-las mais restritivas do que as propriedades já definidas pelos atributos nativos mapeados ao atributo federado. O assistente fornece propriedades predefinidas que respeitam este critério. Se, após ter personalizado as propriedades predefinidas para o atributo federado predefinido, pretender reverter para as propriedades sugeridas pelo assistente, pode ainda escolher as definições assumidas.

Para resumir, os atributos federados estão correlacionados a atributos nativos correspondentes em vários servidores de conteúdos. As propriedades de cada atributo federado têm correspondência em todas as propriedades para os atributos nativos. Uma vez criada uma entidade federada, tem agora o caminho para informações armazenadas em diferentes servidores de conteúdos. Pode então utilizar as entidades federadas para criar modelos de pesquisa para consultas em particular.

## Utilizar o assistente de Criação de Entidade Federada

O assistente de Criação de Entidade Federada é novo no EIP Versão 8.2. Embora possa utilizar as mesmas caixas de diálogo para criação de entidades federadas que no EIP Versão 7.1 e anteriores, o assistente facilita este processo.

Para criar a entidade federada utilizando o assistente, siga os passos indicados abaixo:

1. **Definir uma Entidade Federada** Denomina e descreve uma entidade federada. Pode também determinar se pretende que a mesma seja pesquisável por texto.
2. **Definir Atributos Federados** Denomina e modifica atributos federados.



3. **Correlacionar Atributos Federados** Correlaciona atributos federados a atributos nativos. São fornecidas ferramentas para obter um inventário de servidor, para seleccionar os atributos nativos pode pretender correlacionar e modificar as suas correlações posteriormente.
4. **Definir Propriedades** Define as propriedades para cada atributo federado. É possível personalizar as propriedades ou alterar as definições assumidas.
5. **Confirmar Entidade Federada** Revê as definições que seleccionou para a sua entidade federada. Pode voltar aos painéis anteriores para modificar as suas definições. Clique sobre **Terminar** quando tiver terminado.

Estes passos correspondem à utilização do assistente. Consulte a ajuda on-line do EIP para obter mais informações sobre a forma de utilizar o assistente.

---

## Criar índices de texto federados

O Motor de Pesquisa de Texto pode ser integrado com os servidores de conteúdos do Content Manager Versão 7.1 e anteriores, para que possa automaticamente indexar, pesquisar e obter informações de texto guardadas no Content Manager. Os utilizadores podem localizar documentos pesquisando palavras ou expressões. O servidor de pesquisa de texto suporta conjuntos de caracteres de byte único e de duplo byte.

Se estiver a utilizar servidores Content Manager Versão 7.1 e anterior com o Motor de Pesquisa de Texto, pode criar um índice de texto federado. Poderá então definir a correspondência do índice remissivo de texto federado para o índice remissivo de pesquisa de texto do Content Manager nos servidores de pesquisa de texto do Content Manager.

Quando cria um índice de pesquisa de texto federada, pode activá-lo para pesquisa associada, isto é, para procurar tanto índices de texto nativo como atributos nativos. Quando activa um índice de texto federado para pesquisas combinadas, é também efectuada a correlação desse índice para uma entidade federada. Efectua depois a correlação dos atributos nativos cuja correlação foi efectuada pela mesma entidade federada e os respectivos atributos federados para os índices de pesquisa de texto nativos nos servidores de pesquisa de texto.

---

## Criar modelos de pesquisa

Após a criação de uma entidade federada, pode criar um modelo de pesquisa. Tenha em conta que um modelo de pesquisa utiliza a entidade federada como uma correlação para onde todo o conteúdo está armazenado. Quando cria o modelo de pesquisa, tem ainda que definir o que pretende procurar, o que pretende fazer com os resultados da pesquisa e quem tem permissão para utilizar o modelo. Enquanto apenas pode utilizar uma entidade federada para cada modelo, pode utilizar uma entidade federada para vários modelos. Pode também fazer pesquisas em qualquer combinação dos atributos da entidade federada como critério de pesquisa. Para criar um modelo de pesquisa, complete estes passos utilizando o assistente de Modelo de Pesquisa:

1. Definir o modelo de pesquisa
2. Definir o critério de pesquisa
3. Determinar as definições de pesquisa
4. Atribuir privilégios de acesso



Estes passos correspondem aos passos do assistente do Modelo de Pesquisa. Consulte a ajuda on-line do EIP para obter mais detalhes para completar o processo de criação do modelo de pesquisa.

## Definir o modelo de pesquisa

Após ter iniciado o assistente, ser-lhe-á pedido para definir o modelo de pesquisa. Esteja preparado para:

- Fornecer um nome e uma descrição para o modelo de pesquisa
- Seleccionar uma entidade federada para o modelo de pesquisa **Restrição:** Pode utilizar apenas uma entidade federada por modelo de pesquisa.
- Selecione um índice de texto federado, se aplicável

**Sugestão:** A caixa de selecção para índice de texto federado apenas é aplicável se estiver a utilizar Motor de Pesquisa de Texto para o Content Manager Versão 7.1 e anterior. Se estiver a utilizar o DB2 TIE para pesquisa de texto, é uma pesquisa paramétrica e pode ser configurada como tal no modelo de pesquisa.

## Definir critérios de pesquisa

Após ter definido o modelo de pesquisa, o assistente pede-lhe para que

1. Escolha um tipo de pesquisa, atributo ou documento. A opção Documento só está disponível se um índice de texto federado tiver sido seleccionado no passo anterior.
2. Nomeie o critério de pesquisa
3. Selecione um atributo federado
4. Selecione operadores disponíveis
5. Forneça uma cadeia de pesquisa predefinida (apenas pesquisa de documentos)

O assistente fornece um menu pendente com todos os atributos federados associados com a entidade federada seleccionada. Estes atributos passam a ser os critérios para o modelo de pesquisa. O assistente fornece também uma lista de operadores disponíveis.

**Sugestão:** Pode criar mais que um critério de pesquisa por modelo ou eliminar critérios existentes no modelo.

## Determinar as definições de pesquisa

Este painel permite-lhe atribuir as predefinições de pesquisa, definições de critérios e as definições de valor de apresentação. Cada uma destas definições tem um valor predefinido o qual pode ser alterado. Para modificar estas definições clique sobre o botão correspondente a cada definição.

A janela das predefinições permite-lhe:

- Controlar o que acontece se um servidor não estiver disponível quando um utilizador de uma aplicação cliente pretender utilizar um modelo de pesquisa
- Definir um carácter global para pesquisas paramétricas
- Especificar o nome da pasta para guardar os resultados da pesquisa
- Selecione se a pesquisa tem de utilizar todos (E) ou quaisquer dos critérios (OU).

A janela das Definições de Critérios permite-lhe controlar a ordem dos critérios de pesquisa, a ordem das colunas de apresentação de resultados, os cabeçalhos das colunas e larguras de coluna.

A janela das Definições de Valores de Apresentação fornece uma maneira de definir o valor de apresentação dos resultados da pesquisa. Por exemplo, se o valor para Dia da semana é Segunda num servidor, mas Seg. noutro, pode especificar que Segunda deve ser utilizado como um valor de apresentação de resultados da pesquisa para ambos os servidores.

## **Atribuir privilégios**

Para além de definir onde procurar (com entidades federadas), o que procurar (critério de pesquisa) e como apresentar resultados (definições), deve também definir quem tem acesso a modelos de pesquisa.

A janela de Atribuição de Privilégios do assistente do Modelo de Pesquisa fornece ferramentas para atribuir acesso ao modelo a utilizadores ou grupos de utilizadores existentes.

Ao atribuir os privilégios de acesso de um utilizador a um modelo de pesquisa não atribui esse acesso de utilizador aos servidores de conteúdos correlacionados para o modelo. Os utilizadores têm que preencher os requisitos de segurança para cada um dos servidores de conteúdos. Tem de utilizar listas de controlo de acesso e gestão de utilizadores para assegurar que os utilizadores têm os privilégios adequados antes de lhes atribuir acesso a um modelo de pesquisa.

Quando utiliza o assistente para pesquisar utilizadores ou grupos de utilizadores, o EIP devolve apenas utilizadores que tenham acesso adequado aos servidores de conteúdos solicitados.

---

## Capítulo 4. Gerir acesso de utilizadores

Um utilizador não pode ter acesso ao sistema do EIP sem um ID de utilizador, uma palavra-passe ou um conjunto de privilégios. Antes de criar utilizadores e de lhes atribuir privilégios, tem de decidir quem irá ter acesso ao sistema e o que necessitam para as respectivas funções. Não se vai permitir que os utilizadores tenham a facilidade de eliminar um objecto quando não compreendem as consequências desse acto. Por outro lado, não se pretende impedir que os utilizadores façam o seu trabalho não lhes concedendo os privilégios correctos. Assim, antes de atribuir privilégios a utilizadores, terá de determinar os tipos de tarefas que cada função requer.

Quando os utilizadores criam um objecto no sistema EIP, têm de definir o acesso que terão outros utilizadores a esse objecto. Os utilizadores que criem um objecto têm de definir quem pode aceder ao objecto e quais as operações que podem ser efectuadas ao mesmo. Esta definição é conhecida no sistema de EIP como uma lista de controlo de acessos ou uma ACL.

---

### Criar IDs de utilizador e palavras-passe

Se pretende que um ID de utilizador definido no cliente de administração do sistema também possa ser utilizado para autenticação de DB2, esse ID de utilizador tem de seguir as normas de nomenclatura do DB2. As normas de nomenclatura do DB2 aplicam-se a IDs de utilizador que pretenda utilizar para super-administradores ou IDs de utilizador de ligação. Não pode utilizar as seguintes palavras:

- USERS
- ADMINS
- GUESTS
- PUBLIC
- LOCAL
- Qualquer palavra reservada de SQL listada no Manual de SQL.

Um ID de utilizador não pode começar com os seguintes caracteres:

- SQL
- SYS
- IBM

Pode utilizar os seguintes caracteres:

- A a Z **Restrição:** alguns sistemas operativos permitem IDs de utilizador e palavras-passe que têm em conta maiúsculas e minúsculas. Consulte a documentação do seu sistema operativo para verificar se este tem em conta maiúsculas e minúsculas.
- 0 a 9
- #
- \$

**Restrição:** Os IDs de utilizador não podem ter mais que 30 caracteres.

---

## Compreender a autoridade de administração de DB2

Ao iniciar sessão no cliente de administração do sistema, tem dois níveis de autenticação: um ao nível da base de dados e outro ao nível do produto. Quando activa a função de domínios de administração, os administradores possuem duas classificações: super-administradores e sub-administradores. De modo geral, apenas os super-administradores possuem acesso ao cliente de administração do sistema.

Os super-administradores devem possuir privilégios de DB2: são necessários privilégios db2admin, ou seja, privilégios administrativos totais do DB2. Este ID de utilizador tem de ser definido no sistema operativo com o privilégio db2admin. A palavra-passe para este ID de sistema operativo é utilizada para estabelecer ligação com o DB2 e para iniciar sessão no servidor de bibliotecas. A palavra-passe definida para o servidor de bibliotecas não é utilizada. Privilégios do Content Manager: Este ID de utilizador é definido no servidor de bibliotecas com privilégios administrativos de Content Manager totais ("AllPrivs") para desempenhar todas as actividades administrativas.

Os sub-administradores não necessitam de privilégios de DB2. Os sub-administradores gerem apenas certas secções do servidor de bibliotecas. Assim, os sub-administradores iniciam sessão no cliente de administração do sistema de uma das seguintes formas:

- Se o ID de utilizador for um ID de utilizador de sistema operativo, a palavra-passe do sistema operativo é utilizada para estabelecer ligação com o DB2 e para iniciar sessão no servidor de bibliotecas.
- Se o ID de utilizador não for um ID de utilizador de sistema operativo, o par codificado de ID de utilizador e palavra-passe em cmbfedenv.ini (para Enterprise Information Portal) ou cmbicmenv.ini (para Content Manager) é utilizado para estabelecer ligação com o DB2 e o ID de utilizador e a palavra-passe fornecidos na janela Iniciar Sessão são utilizados para iniciar sessão no servidor de bibliotecas.

Para obter mais informações sobre o início de sessão no servidor de bibliotecas, consulte a secção seguinte.

Os sub-administradores também necessitam de privilégios de EIP. Necessitam do privilégio Administrativo de Domínio para todas as actividades de administração de sub-domínios.

## Estabelecer ligação com o DB2 utilizando os ficheiros INI

Cada entrada do ficheiro INI contém o nome de um servidor de bibliotecas e um par codificado de ID de utilizador e palavra-passe para estabelecer ligação com o DB2. Este ID de utilizador codificado (denominado ID de utilizador de ligação) e esta palavra-passe codificada são definidos quando o produto é instalado. O ID de utilizador de ligação tem de ser diferente do ID de utilizador do administrador do sistema. O Enterprise Information Portal utiliza cmbfedenv.ini para estabelecer ligação com o DB2 e o Content Manager utiliza cmbicmenv.ini. O ID de utilizador de ligação predefinido é ICMCONCT. Durante a instalação, as palavras-passe para o servidor de bibliotecas e para o gestor de recursos estão contidas em três locais. O ficheiro cmbicmenv.ini contém o ID de utilizador e a palavra-passe para aceder ao servidor de bibliotecas. O sistema operativo define o acesso à base de dados onde residem o servidor de bibliotecas e o gestor de recursos. O ficheiro ICMRM.properties contém o ID de utilizador e a palavra-passe para o gestor de recursos.

Se for utilizado o ficheiro INI, isto é, se o ID de utilizador não for um ID de utilizador de sistema operativo, então tanto o ID de utilizador como o ID de utilizador de ligação do ficheiro INI têm de existir no servidor de bibliotecas.

O ID de utilizador de ligação tem de ser definido no servidor de bibliotecas e no sistema operativo. Neste caso, é necessário o privilégio UserDB2Connect. Para alterar o ID de utilizador de ligação e a palavra-passe no ficheiro INI, seleccione **Ferramentas --> Alterar ID/palavra-passe de Base de Dados** na janela de cliente de administração.

## **Alterar as palavras-passe do servidor de bibliotecas e do administrador do sistema para o gestor de recursos**

Se necessitar de alterar a palavra-passe para o gestor de recursos, tem de alterar a palavra-passe para o início de sessão do servidor de bibliotecas no gestor de recursos e a palavra-passe do administrador do sistema para o gestor de recursos. **Importante:** Para alterar as palavras-passe para o início de sessão do servidor de bibliotecas e do administrador do sistema no gestor de recursos, execute os seguintes passos pela ordem indicada:

1. Inicie sessão no cliente de administração do sistema.
2. Expanda a árvore do Gestor de Recursos.
3. Clique sobre o gestor de recursos que pretende modificar e expanda a sua árvore.
4. Clique sobre Definições do Servidor e seleccione Propriedades. É aberta a janela Painel do Servidor.
5. Altere a palavra-passe no campo Palavra-passe.
6. Faça clique sobre OK.
7. Faça clique com o botão direito do rato sobre o gestor de recursos expandido (no passo 2) e seleccione Propriedades. Surge a janela Propriedades do Gestor de Recursos.
8. Altere a palavra-passe no campo Palavra-passe e clique em OK.

## **Alterar as palavras-passe de acesso à base de dados**

Se necessitar de alterar a palavra-passe de acesso à base de dados, tem de alterar a palavra-passe do sistema operativo para a ligação à base de dados e o ficheiro ICMRM.properties de forma a que o gestor de recursos possa identificar a nova palavra-passe.

Para alterar a palavra-passe do sistema operativo para a ligação à base de dados, execute os seguintes passos:

1. Consoante o seu sistema operativo, navegue para o utilitário de Utilizadores e Palavras-passe.
2. Faça clique sobre ICMRM.
3. Seleccione Definir Palavra-passe.
4. Introduza a nova palavra-passe.

Para alterar o ficheiro ICMRM.properties, execute os seguintes passos:

1. Abra o ficheiro ICMRM.properties. A localização predefinida é `X:\WebSphere\AppServer\installedApps\icrmr.ear\icrmr.war\WEB-INF\classes\com\ibm\mm\icrmr\ICMRM.properties` em que X é a localização da unidade na qual foi instalado o Content Manager.
2. Altere DBPassword conforme a palavra-passe do sistema operativo.

3. Guarde o ficheiro ICMRM.properties.

Quando tiver alterado a palavra-passe da base de dados, é necessário reiniciar a base de dados ou deixá-la emitir dois ou três erros até que se auto-reinicie.

Para obter instruções mais detalhadas sobre a forma de alterar as palavras-passe e outros campos para um gestor de recursos no cliente de administração do sistema, consulte a ajuda de administração do sistema online.

---

## Importar utilizadores de LDAP

O LDAP suporta a gestão do ID e da palavra-passe de um utilizador ao nível da empresa e não numa base de sistema-a-sistema. EIP utiliza três tecnologias LDAP: IBM Directory (denominado IBM SecureWay Directory em versões anteriores), Windows 2000 Active Directory e Lotus Domino Directory Notes Address Book (NAB). A palavra-passe do utilizador reside no servidor LDAP. Quando um utilizador inicia sessão no ou no Enterprise Information Portal, o ID de utilizador e a palavra-passe são autenticados e os privilégios específicos do ID de utilizador são verificados pelo perfil de utilizador na base dados do EIP. O LDAP pode ter sido activado durante a instalação do EIP. Se o LDAP não tiver sido activado durante a instalação, pode activá-lo em qualquer altura.

Para activar o LDAP, seleccione **Iniciar → Programas → EIP for Multplatforms → Programador de Importação de ID de Utilizador de LDAP** e, em seguida, inicie o cliente de administração do sistema. Active a janela Configuração de LDAP (Ferramentas --> Configuração de LDAP). Seleccione o quadrado de confirmação Activar importação e autenticação de utilizador de LDAP e introduza as informações referentes ao servidor LDAP na página Servidor.

Depois de activar o LDAP, pode importar utilizadores fazendo clique no botão LDAP na janela Novo Utilizador. Isto permite importar selectivamente os utilizadores do servidor LDAP para o EIP. Como alternativa, pode importar os utilizadores em grupos utilizando o utilitário Programador de Importação de ID de Utilizador de LDAP. Durante o início de sessão, o servidor de bibliotecas estabelece automaticamente ligação com o servidor LDAP para autenticar o utilizador. Se por algum motivo o servidor LDAP não puder autenticar a palavra-passe do utilizador, a autenticação falha.

Pode modificar a configuração do servidor LDAP na janela principal do cliente de administração do sistema, fazendo clique sobre **Ferramentas -> Configuração de LDAP**. Também pode alterar o seu servidor LDAP actual no Utilitário de Importação de Registo de Utilizador de LDAP. Este pode ser acedido a partir do menu Iniciar no EIP. Para obter mais informações sobre o planeamento do LDAP, consulte *Planeamento e Instalação do Sistema Content Management*. Para obter mais informações sobre a configuração das informações do servidor LDAP nas janelas de administração do sistema, consulte a ajuda do cliente de administração do sistema online.

Para informações sobre planear para LDAP, consulte o manual *Planeamento e Instalação do Sistema Content Management*. Para mais informações sobre como implementar o LDAP, consulte a ajuda on-line do cliente de administração de sistema.

---

## Introduzir privilégios

O cliente de administração fornece grupos de privilégios, conjuntos de privilégios e privilégios individuais. Se administrar um sistema combinado Content Manager/EIP, os privilégios são comuns a ambas as partes do cliente. Os privilégios incorporados no cliente podem ajudar a evitar obstáculos no

### Grupo de privilégios

Um grupo de privilégios é uma recolha de tarefas de utilizador destinadas a facilitar aos administradores a criação de novos conjuntos de privilégios ou a utilização de papéis na caixa de diálogo Conjunto de privilégios.

### Grupos de privilégios

Os conjuntos de privilégios são recolhas de papéis de utilizador.

### Privilégio

Um privilégio representa uma acção de utilizador. Por exemplo,

**Exemplo 1 - privilégios:** Pretende atribuir os privilégios ClientScan e ClientImport a um grupo de utilizadores que geralmente utilizam um cliente apenas para pesquisar e importar documentos para o Content Manager. Se possuir vários utilizadores que geralmente realizam essa tarefa, pode criar um ID de utilizador (por exemplo, utilizador1). Em seguida, associa os privilégios ClientScan e ClientImport ao ID de utilizador utilizador1. Atribui então Utilizador1 a um grupo denominado Grupo1. Assim, quando um utilizador final introduzir utilizador1 para iniciar sessão no seu cliente e aceder ao Content Manager, esse utilizador vai poder apenas pesquisar e importar documentos.

**Exemplo 2 - grupos de privilégios:** Tem um grupo de utilizadores finais experientes que requerem privilégios de acesso a todas as tarefas típicas de cliente. Pode criar um ID de utilizador (por exemplo, utilizador2). Em seguida, atribua utilizador2 a um grupo (por exemplo, grupo2). Associe então o grupo de privilégios denominado ClientTaskAll a utilizador2. Assim, quando um utilizador final introduzir utilizador2 para iniciar sessão no seu cliente e aceder ao Content Manager, esse utilizador vai realizar todas as tarefas incluídas no grupo de privilégios denominado ClientTaskAll.

**Exemplo 3 - conjuntos de privilégios:** Tem um grupo de utilizadores que requerem acesso só de leitura. Crie um ID de utilizador (por exemplo, utilizador3). Em seguida, atribua utilizador3 a um grupo (por exemplo, grupo3). Associe então o conjunto de privilégios denominado ClientUserReadOnly a utilizador3. Assim, quando um utilizador final introduzir utilizador3 para iniciar sessão no seu cliente e aceder ao Content Manager, esse utilizador vai poder realizar apenas as tarefas incluídas no conjunto de privilégios denominado ClientUserReadOnly.

---

## Criar conjuntos de privilégios

Ao planificar a configuração do seu sistema EIP, tem também de decidir quem vai ter acesso ao sistema e até que ponto esses utilizadores vão poder aceder aos objectos no seu sistema. O sistema EIP define o acesso através de privilégios.

Um privilégio concede o direito a aceder a um objecto específico de um modo específico. Os privilégios incluem direitos como, criar, eliminar e seleccionar objectos armazenados num sistema. Um grupo de privilégios atribuídos a um utilizador é um conjunto de privilégios.



A primeira tarefa do utilizador na gestão de acesso é criar conjuntos de privilégios para utilizadores. Um *conjunto de privilégios* identifica as tarefas ou acções que o utilizador pode realizar. Os conjuntos de privilégios combinam privilégios e estão adaptados a certos tipos de utilizadores. Por exemplo, pode querer que um conjunto de administradores faça a gestão do servidor de encaminhamento de documentos e que outro faça a gestão de um domínio. Quando um administrador inicia sessão, o EIP verifica o conjunto de privilégios do administrador.

O cliente de administração de sistema tem vários privilégios predefinidos que pode agrupar num conjunto de privilégios. Atribui então os conjuntos de privilégios que criar a utilizadores individuais. Não pode atribuir um conjunto de privilégios a um grupo de utilizadores.

## **Criar grupos de privilégios**

Os grupos de privilégios são como grupos de utilizadores para utilizadores. Ao criar um grupo de privilégios, coloca juntos os privilégios que se assemelham para que possa encontrar facilmente os privilégios que pretende incluir num conjunto de privilégios. Por exemplo, se tiver dois privilégios que atribua a quase todos os utilizadores no seu sistema, em vez de pesquisar nos diversos privilégios que tem sempre que criar um conjunto de privilégios, agrupa estes dois privilégios básicos num grupo de privilégios denominado BasicPrivs.

## **Atribuir um conjunto de privilégios a um utilizador**

O cliente de administração de sistema tem vários privilégios predefinidos que pode agrupar num conjunto de privilégios. Atribui então os conjuntos de privilégios que criar a utilizadores individuais. Não pode atribuir um conjunto de privilégios a um grupo de utilizadores.

Pode criar nomes de privilégios, mas não pode criar o próprio privilégio. Tem de trabalhar em conjunto com o programador do sistema para criar quaisquer privilégios que ainda não estejam definidos no cliente de administração de sistema.

Pode utilizar os conjuntos de privilégios que vêm com o EIP, ou pode criar os seus.

## **Atribuir a um ID de utilizador um conjunto de privilégios de concessão**

Para impedir que os utilizadores criem um ID de utilizador com mais privilégios do que aqueles que têm, o EIP implementou a utilização de um conjunto de privilégios de concessão. Ao atribuir um ID de utilizador com um conjunto de privilégios de concessão, está a conceder autoridade para criar IDs de utilizador dentro dos limites dos privilégios concedidos. Por exemplo, pode dar a um ID de utilizador um conjunto de privilégios de administração do sistema para gerir um domínio. Pode, no entanto, querer certificar-se que o ID de utilizador não possui o privilégio para criar utilizadores. Quando criar este ID de utilizador, seleccione "Noprivs" no campo do conjunto de privilégios de concessão. Com efeito, o ID de utilizador pode gerir o domínio mas não pode criar utilizadores para esse domínio.

## **Atribuir utilizadores a gestor de recursos**

Para permitir que os utilizadores acessem a um gestor de recursos específico, tem de atribuir um gestor de recursos a um domínio ao qual os utilizadores tenham acesso. Para mais informações sobre a atribuição de gestor de recursos a domínios, consulte "Atribuir um gestor de recursos a um domínio" na página 38.



## Atribuir utilizadores a recolhas

Para permitir que os utilizadores acedam a recolhas, tem de atribuir uma recolha num gestor de recursos a um domínio ao qual os utilizadores tenham acesso. Para mais informações sobre a atribuição de recolhas a domínios, consulte “Atribuir uma recolha a um domínio” na página 38.

---

## Criar grupos de utilizadores

Muitas vezes, os utilizadores com a mesma descrição de trabalho têm as mesmas tarefas e, por conseguinte, o mesmo nível de acesso a objectos existentes no sistema. Poderá agrupar os utilizadores com necessidades de acesso comuns num grupo de utilizadores. Não pode encaixar grupos de utilizadores.

Um grupo de utilizadores é apenas um agrupamento de conveniência de utilizadores individuais com tarefas semelhantes. Não irá atribuir a um grupo de utilizadores um conjunto de privilégios. Cada utilizador num grupo de utilizadores tem o seu próprio conjunto de privilégios. Um grupo de utilizadores torna mais fácil criar listas de controlo de acesso para objectos existentes no seu sistema.

Se tiver domínios activados, antes de atribuir um ID de utilizador a um grupo, verifique se esse grupo de utilizadores se encontra num domínio específico ou no domínio PÚBLICO (consulte “Administrar domínios” na página 37 para mais informações sobre domínios). Certifique-se que o grupo de utilizadores está no domínio no qual pretende que esteja o seu ID de utilizador. Se pretende criar um ID de utilizador especificamente para um domínio, pode fazer clique sobre **Novo Utilizador** na janela de Grupo de Utilizadores. Pode, então, adicionar o utilizador que criou ao grupo de utilizadores e certificar-se que o utilizador está no mesmo domínio.

---

## Criar listas de controlo de acessos

Providencia aos utilizadores os privilégios que necessitam para realizarem as suas tarefas. Os objectos, numa base individual, têm certos problemas de controlo de acesso.

Uma lista de controlo de acessos (ACL) é uma lista que consiste num ou mais IDs de utilizador individuais ou grupos de utilizadores e os respectivos privilégios associados. As ACLs são utilizadas para controlar o acesso dos utilizadores a objectos existentes no sistema do EIP. Os objectos que podem ser associados a listas de acessos são: os objectos de dados armazenados por utilizadores, tipos de artigos e sub-conjuntos de tipos de artigos, listas de trabalho e processos.

Os conjuntos de privilégios definem uma capacidade máxima dos utilizadores individuais, uma ACL restringe o acesso do utilizador individual a um objecto. Uma ACL que tenha um privilégio que não esteja definido pelo conjunto de privilégios do utilizador não irá conceder ao utilizador esse privilégio. Apenas os utilizadores que têm esse privilégio podem utilizá-lo num objecto. Uma ACL limita o acesso do utilizador, não lhe concede mais acesso. As listas de controlo de acessos fornecem outro nível de segurança ao gerir um sistema.

## Atribuir um conjunto de privilégios a uma lista de controlo de acessos

Cada ID de utilizador que adicione a uma lista de controlo de acessos (ACL) necessita de um conjunto de privilégios associado àquele. O ID de utilizador e o

conjunto de privilégios definem quais os utilizadores que têm acesso a um objecto e qual o tipo de acesso que têm a esse objecto.

Os utilizadores não podem aceder a quaisquer objectos, a não ser que se encontrem na ACL. Para adicionar um utilizador ou um grupo de utilizadores a uma ACL, necessita de seleccionar um ID de utilizador e um conjunto de privilégios para a ACL e clicar em **Adicionar**. Para cada ACL definida, vai encontrar os IDs de utilizador e grupos listados na janela da Lista de Controlo de Acessos. Pode modificar esta tabela adicionando e removendo IDs de utilizador e grupos. Para mais informações sobre criar e modificar uma ACL, consulte a ajuda on-line do cliente de administração de sistema.

---

## Criar domínios

Um domínio é uma secção de um de uma base de dados de administração que um ou mais administradores gerem. Os domínios consistem em IDs de utilizadores, grupos de utilizadores, listas de controlo de acesso, conjuntos de privilégios, listas de controlo de acesso, gestor de recursos e conjuntos de SMS. Não são visíveis aos utilizadores, de forma que a denominação que der aos domínios apenas farão sentido para si e para os administradores do sistema que os gerem. Os utilizadores não têm conhecimento de que os limitou a uma parte da base de dados de administração, o que significa que apenas têm conhecimento sobre os artigos nesse domínio.

Os domínios limitam acesso administrativo e de utilizador a uma subsecção da base de dados de administração. Um administrador com plenos privilégios para o base de dados de administração pode delegar privilégios administrativos para outro administrador. O administrador com plenos privilégios, um super-administrador, tem acesso a todas as secções de um base de dados de administração enquanto que um administrador com privilégios limitados, um sub-administrador, tem acesso a apenas uma secção do base de dados de administração.

Os domínios restringem o acesso que o sub-administrador tem a listas de controlo de acessos (ACLs). Apenas os super-administradores podem criar ACLs que os sub-administradores podem utilizar para adicionar ou eliminar IDs de utilizador e grupos de utilizadores. Os sub-administradores não podem criar, actualizar ou eliminar ACLs.

Um sub-administrador pode partilhar combinações diferentes das responsabilidades do super-administrador, mas apenas para o respectivo domínio. Ao criar domínios e designar administradores para gerir os mesmos, os super-administradores delegam subtarefas para que se possam concentrar em todo o sistema e geri-lo de forma eficiente como os sub-administradores gerem utilizadores e tarefas específicas ao seu domínio.

Antes de activar domínios, deve ter em consideração as condições que se seguem:

- Não pode desactivar domínios
- Gestor de recursos, recolhas, IDs de utilizador e grupos de utilizadores podem existir apenas num domínio de cada vez.
- Conjuntos de privilégios e listas de controlo de acesso podem existir em mais do que um domínio de cada vez.
- À excepção do domínio PÚBLICO (partilhado), os domínios não se sobrepõem
- Qualquer objecto criado num super domínio administrativo não pode ser movido, seja ele gerado pelo sistema ou criado pelo utilizador.

Para activar domínios, vá ao menu do ficheiro, seleccione **Ferramentas** → **Domínios Administrativos** e depois seleccione **Activar Domínios Administrativos**. Tem de reiniciar o cliente de administração do sistema para os domínios entrarem em efeito. Para instruções específicas sobre como configurar a base de dados de administração para domínios, consulte a ajuda online do cliente de administração de sistema.

## Administrar domínios

Dependendo do conjunto de privilégios, pode administrar toda a base de dados de administração ou um domínio específico. Um administrador que tem pleno acesso à base de dados de administração é um super-administrador. Um sub-administrador tem pleno acesso a objectos num domínio específico.

Cada tipo de administrador tem a capacidade de criar, obter, actualizar e eliminar os objectos nos respectivos domínios, incluindo utilizadores e recolhas. Os sub-administradores podem ver e obter objectos apenas no seu domínio e listar ou recuperar no domínio PÚBLICO, ou partilhado.

## Aceder a domínios

Os sub-administradores não podem alterar o domínio de um objecto. Podem contudo aceder aos conteúdos do seu domínio e listar ou recuperar qualquer objecto no domínio PÚBLICO ou partilhado.

Os super-administradores têm acesso a todos os domínios na base de dados de administração. Podem criar um objecto e atribuí-lo a um domínio. Alguns objectos, tais como conjuntos de privilégios e ACLs, apenas podem ser criados por aqueles para que os sub-administradores os utilizem.

Os sub-administradores só podem criar, recuperar, actualizar e eliminar (CRUD) para quaisquer objectos no domínio.

## Atribuir um utilizador a um domínio

Ao criar um ID de utilizador, tem a opção de atribuí-lo a um domínio ou deixá-lo no domínio predefinido. Pode alterar o domínio do ID de utilizador posteriormente através das propriedades do utilizador.

Um ID de utilizador pode ter acesso apenas a um domínio de cada vez. Não pode adicionar um utilizador ao domínio PÚBLICO ou partilhado.

Apenas os super-administradores têm autoridade para criar domínios e atribuir utilizadores a esses domínios. Um domínio pode ter mais do que um sub-administrador, mas apenas o super-administrador pode definir quem são esses administradores, concedendo-lhes privilégios de administração do sistema dentro de um grupo de privilégios. O campo **Conjunto de Privilégios de Concessão** no Novo Utilizador ou na janela Propriedades de Utilizador indicará quais os privilégios administrativos que um sub-administrador tem num domínio.

## Atribuir um grupo de utilizadores a um domínio

Atribuir um grupo de utilizadores a um domínio altera o domínio designado para cada ID de utilizador nesse grupo de utilizadores. Um ID de utilizador pode ter acesso apenas a um domínio de cada vez. Então, qualquer ID de utilizador incluído num grupo por si atribuído é também movido para o novo domínio.

Um nome de grupo de utilizador não pode estar em apenas um domínio de cada vez. Pode atribuir o grupo de utilizadores ao domínio PÚBLICO ou partilhado.

## Atribuir um conjunto de privilégios a um domínio

Qualquer ID de utilizador que adiciona a um domínio tem de ter também um conjunto de privilégios associado. Se não incluir os conjuntos associados de privilégios, os utilizadores não vão conseguir desempenhar as respectivas tarefas. O melhor local para armazenar conjuntos de privilégios para torná-los disponíveis para qualquer utilizador é o domínio PÚBLICO, ou partilhado.

## Atribuir um gestor de recursos a um domínio

Pode restringir o acesso de utilizadores a determinados gestor de recursos atribuindo-lhes um domínio específico. Quando define um novo gestor de recursos para ser acedido por um da base de dados de administração, tem a opção que permite seleccionar um domínio.

A predefinição para todos os gestor de recursos é PÚBLICO. Se não pretende que todos tenham acesso ao gestor de recursos, tem de atribuí-lo a um domínio. Caso não veja um domínio ao qual possa atribuir o gestor de recursos, pode ainda definir este e, de seguida, criar o domínio que necessita. Após ter o domínio adequado definido, abra as propriedades do gestor de recursos e seleccione o domínio.

## Atribuir uma recolha a um domínio

Pode restringir o acesso de utilizadores a uma determinada recolha num gestor de recursos atribuindo-o a um domínio específico. Se o gestor de recursos estiver no domínio PÚBLICO, pode atribuir uma recolha a qualquer outro domínio definido. Se, no entanto, o gestor de recursos já estiver definido para um domínio específico, então não pode atribuir a recolha a outro domínio, mesmo que queira atribuí-la ao domínio PÚBLICO.

Um utilizador necessita de acesso ao gestor de recursos para aceder a recolhas neste, por isso não pode restringir ao acesso ao gestor de recursos sem impor as mesmas restrições às respectivas recolhas.

## Mover um utilizador de um domínio para outro

Pode deparar-se com circunstâncias em que tenha de remover certos utilizadores de um domínio e adicioná-los a outro. Considere utilizar o campo **Descrição** na janela definição de Utilizador como forma de lembrar em que grupos de utilizadores é que um utilizador está agrupado. Pode tornar esta tarefa um pouco mais fácil.

**Importante:** Esta tarefa é muito morosa e pode resultar em problemas no acesso ao sistema se não o fizer de forma correcta. Tem de ser um super-administrador para alterar o domínio do utilizador.

Siga estes passos com atenção:

1. Encontre todos os grupos ao qual pertence o utilizador.
2. Para todos os grupos ao qual pertence o utilizador, mova estes grupos para o domínio PÚBLICO ou remova o utilizador de todos os grupos.
3. Mova quaisquer gestor de recursos associados a este utilizador para o domínio PÚBLICO, seguido de todas as recolhas para cada gestor de recursos que vai mover para o domínio de destino.

4. Crie, *não mova*, todos os conjuntos de privilégios associados ao utilizador no domínio de destino, se aí não estiverem.
5. Crie, *não mova*, todas as listas de controlo de acesso associadas ao utilizador, se não tiverem no domínio de destino.
6. Mova o utilizador para o domínio de destino abrindo as Propriedades do utilizador e alterando o domínio do utilizador.
7. **Opcional:** Pode mover os grupos e gestor de recursos que moveu nos passos 1 na página 38, 2 e 3 na página 38 do domínio PÚBLICO para o domínio de destino, mas só o pode fazer se não houver mais utilizadores no domínio de fonte que são associados aos grupos e gestor de recursos que moveu. Senão, os grupos e gestor de recursos necessitam de ficar no domínio PÚBLICO para permitir a partilha a utilizadores em domínios diferentes.

**A Lembrar:** Um utilizador nunca pode estar num domínio PÚBLICO. Os utilizadores não podem ser partilhados.

## Mover um grupo de utilizadores de um domínio para outro

**Importante:** Esta tarefa é muito morosa e pode resultar em problemas no acesso ao sistema se não o fizer de forma correcta. Tem de ser um super-administrador para alterar o domínio de um grupo de utilizadores.

Siga estes passos para mover um grupo de utilizadores para um domínio diferente:

- Se o grupo de utilizadores estiver vazio, elimine o grupo do seu actual domínio depois recrie o grupo e atribua-o ao domínio de destino.
- Se o grupo de utilizadores não estiver vazio, siga estes passos:
  1. Encontre todos os utilizadores que pertencem a este grupo.
  2. Apague o grupo do seu actual domínio, que apagará todos os utilizadores.
  3. Recrie o grupo e atribua-o ao domínio de destino.
  4. Adicione todos os utilizadores ao grupo recentemente criado.

## Mover um gestor de recursos de um domínio para outro

Tem de ser um super-administrador para alterar o domínio de um gestor de recursos.

Para mover um gestor de recursos para outro domínio, siga os seguintes passos:

- Se o gestor de recursos não contiver recolhas, mova o gestor de recursos para o domínio de destino abrindo as suas propriedades e alterando o domínio para o domínio de destino.
- Se o gestor de recursos contiver recolhas, siga estes passos:
  1. Mova o gestor de recursos para o domínio PÚBLICO.
  2. Mova as recolhas para o domínio de destino abrindo Propriedades e seleccionando o domínio de destino.
  3. Mova o gestor de recursos para o domínio de destino abrindo Propriedades e seleccionando o domínio de destino.

## Mover uma recolha de um domínio para outro

Tem de ser um super-administrador para alterar o domínio de uma recolha.

Siga estes passos para mover uma recolha de um domínio para outro:

1. Descubra o gestor de recursos ao qual pertence a recolha.

2. Mova o gestor de recursos associado para o domínio PÚBLICO.
3. Mova a recolha para o domínio de destino abrindo Propriedades e seleccionando o domínio de destino.
4. Mova o gestor de recursos para o domínio de destino abrindo Propriedades e seleccionando o domínio de destino.

### **Mover um conjunto de privilégios de um domínio para outro**

Como os conjuntos de privilégios podem residir em vários domínios, pode adicioná-los ao domínio de destino sem ter de movê-los.

### **Mover uma lista de controlo de acesso de um domínio para outro**

Como as listas de controlo de acesso podem residir em vários domínios, pode adicioná-los ao domínio de destino sem ter de movê-los.

---

## Capítulo 5. Gerir o information mining

Esta secção começa com uma descrição do Information Mining e de como o utilizar num ambiente empresarial. Em seguida são apresentadas secções acerca dos Primeiro Passos do Information Mining, conceitos do Information Mining, Information Structuring Tool e, por fim, comentários acerca de ajustamentos do rendimento.

---

### O que é o Information Mining?

O Information Mining do Enterprise Information Portal é uma tecnologia que permite às empresas, através da automatização de vários aspectos da extracção e da análise de informações, facilitarem aos utilizadores o acesso a informações relevantes a custos reduzidos.

O primeiro desafio do Information Mining consiste em aceder prontamente a informações sob forma de texto não estruturado para utilização por computadores. Uma interpretação completa de conhecimentos factuais expressos em linguagem natural não restrita ainda se encontra fora do alcance da tecnologia actual. Todavia, as ferramentas que aplicam técnicas de reconhecimento de padrões e heurística são capazes de extrair informações valiosas a partir de texto simples e arbitrário. As informações extraídas podem incluir a identificação das chamadas palavras importantes, como nomes, instituições ou lugares mencionados num documento, assim como resumos inteiros de um documento.

Contudo, o Information Mining proporciona muito mais do que simples extracção de partes de informação de documentos únicos. O information mining torna-se realmente importante quando tem de trabalhar com vastas recolhas de documentos. A metáfora "exploração" é utilizada quer para o processo de descoberta de conhecimento, isto é, identificar e extrair informações codificadas de documentos únicos e armazenar estas informações como metadados, quer para o processo analítico de distribuição destas características ao longo de recolhas de documentos, detectando fenómenos interessantes, padrões ou tendências.

### Os serviços do Information Mining do Enterprise Information Portal

Os serviços do Information Mining do Enterprise Information Portal fornecem uma infra-estrutura para a criação e manutenção de informações relacionadas com documentos individuais ou com recolhas de documentos. Estas informações denominam-se metadados. Os exemplos de informações que caracterizam o conteúdo do documento e que são armazenadas como metadados incluem:

- Título
- Sinopse ou resumo
- Nomes, termos ou expressões
- Categorias às quais pertence um documento

O que distingue o Information Mining de um arquivo de metadados tradicional que associe documentos com metadados é o facto de em disponibilizar capacidades de criação automática de metadados, mesmo que estes não estejam explicitamente disponíveis. Os algoritmos de extracção e obtenção são capazes de aceder a informações relevantes em enormes recolhas de documentos, quer



mediante metadados que orientem o processo de obtenção, quer mediante a execução de modelos estatísticos com base nos metadados, de modo a encontrar relações interessantes entre documentos que podem não ser evidentes ao contemplar os documentos individuais de um grupo.

Dado que as operações de extracção e obtenção operam sobre conjuntos bem definidos de metadados em vez de considerarem o conteúdo do documento original, a velocidade destes processos pode ser significativamente aumentada se os metadados forem guardados numa memória dedicada, o chamado *arquivo de dados do Information Mining*. Acelera-se assim o acesso aos metadados, visto que as aplicações os podem obter a partir de um repositório único e não precisam de voltar a um servidor de conteúdos arbitrariamente remoto. Como os metadados são cruciais para obtenção e navegação e se utilizam frequentemente para apurar um resultado de pesquisa, podem-se executar muitos passos sem ter de voltar a um servidor de conteúdos.

Outra vantagem na utilização do arquivo de metadados consiste em manter os dados associados a um documento, separados do documento real. Naturalmente que no caso de documentos disponíveis para acesso só de leitura, como por exemplo documentos remotos na Web, a manutenção do conteúdo e dos metadados dentro do mesmo repositório não constitui opção alguma.

Os serviços do Information Mining do Enterprise Information Portal fornecem os seguintes mecanismos para a criação automática de metadados:

- **Categorização**, que atribui uma ou mais categorias a um documento, com base numa taxonomia definida pelo utilizador. O componente de categorização contém uma aplicação que faculta uma interface gráfica de utilizador para criação e manutenção de taxonomias, denominada *Information Structuring Tool*.
- **Resumo**, que extrai as frases mais importantes de um documento e assim pode ajudar o utilizador a decidir se pretende ler o documento todo, ou não. O utilizador pode especificar o comprimento do resumo pretendido e desta forma influenciar directamente o equilíbrio entre a complexidade dos metadados extraídos e o volume de informação no documento.
- **Identificação de Idioma**, que determina o idioma no qual um documento é escrito. Constitui um passo útil de pré-processamento, antes de aplicar outros serviços do Information Mining.
- **Extracção de informações**, que reconhece de forma automática significativos artigos de vocabulário, como nomes, termos e expressões, em documentos de texto.
- **Agrupamento em Unidades**, que divide um conjunto de documentos em grupos semelhantes ou conjuntos de unidades. Os conjuntos de unidades derivam automaticamente da recolha de documentos.

## Componentes dos serviços do Information Mining

Uma aplicação do Information Mining engloba, normalmente, as seguintes tarefas:

1. Organizar dados de forma a torná-los pesquisáveis e navegáveis
2. Aceder a fontes de dados díspares
3. Utilizar a operação de pesquisa avançada para filtrar os dados necessários para uma previsão ou tendência

A Figura 1 na página 43 ilustra estas tarefas do Information Mining.



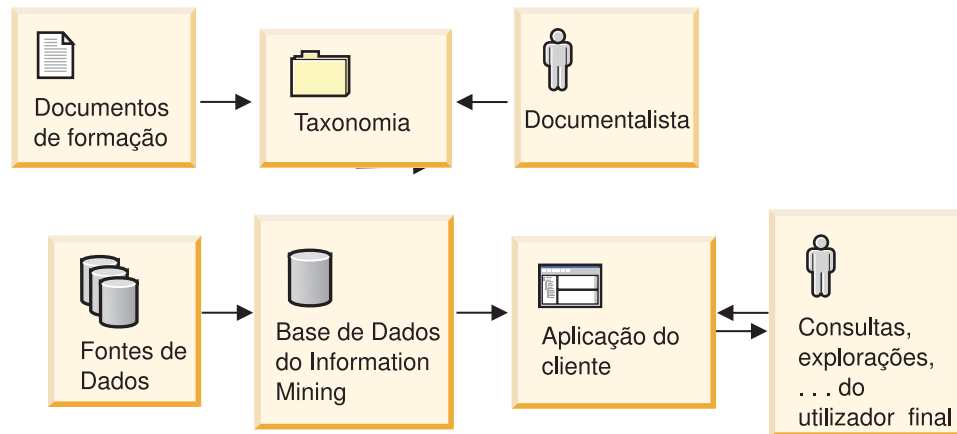


Figura 1. As tarefas do Information Mining

Para tirar partido da funcionalidade do Information Mining, os documentos têm de ser organizados de forma a puderem ser pesquisados e navegados. Esta tarefa é, normalmente, desempenhada por um documentalista ou por um engenheiro do conhecimento. O documentalista utiliza a Information Structuring Tool (IST) para definir uma taxonomia, uma caracterização temática hierárquica dos dados nos documentos para que sejam posteriormente explorados. O IST é uma aplicação com uma interface gráfica de utilizador que permite criar e manter taxonomias. As categorias são sujeitas a testes e, uma vez que a taxonomia estabilize, podem ser acedidas diferentes fontes de dados.

Os documentos podem ser importados de qualquer servidor de conteúdos ou a partir da Web utilizando o serviço Web Crawler (apenas disponível como JavaBeans) e distribuídos por categorias.

As funções de análise de texto e de criação de metadados que são subsequentemente aplicadas aos conteúdos do documento, estão disponíveis como interface de programações ao nível dos JavaBeans não-visuais e de uma API de Serviços Java.

Os JavaBeans do Information Mining são componentes de software para desenvolvimento rápido de aplicações e estão em conformidade com os JavaBeans. A API de Serviços Java contém a funcionalidade completa do Information Mining como blocos individuais de construção para criar aplicações. Os exemplos baseados em JavaBeans e os exemplos de JSPs são fornecidas para suporte à criação de aplicações.

A identificação do conteúdo de um documento é um pré-requisito de todas as operações da Information Mining que processem documentos. Os subpassos destas tarefas incluem:

1. Identificação da página de códigos de um documento
2. Identificação das secções de texto para processamento, ignorando informações de markup ou dados binários, como, por exemplo, imagens.

Dado que os documentos em servidores de conteúdos podem estar estruturados arbitrariamente, os serviços do Information Mining facultam um meio de escrever módulos específicos que identificam e extraem partes de texto relevantes de formatos de documentos. Também está disponível nos serviços Information Mining um módulo predefinido para abranger um vasto leque de formatos documentais frequentemente usados. Consulte Capítulo 9, “Formatos de Documentos”, na página 123

página 123 para obter uma lista dos formatos suportados e o *Manual de Programação da Aplicações* para obter informações mais detalhadas sobre a utilização do módulo predefinido de exemplificação.

A criação de metadados para cada um dos documentos seleccionados envolve o processamento dos conteúdos dos documentos e da aplicação de métodos estatísticos ou heurísticos com base em recursos de conhecimento, por exemplo, dicionários ou perfis de frequência.

As APIs do Information Mining suportam as seguintes operações:

- Resumo
- Categorização
- Identificação de idioma
- Extração de Informações
- Agrupamento em Unidades

Os metadados criados para todos os documentos encontram-se armazenados no arquivo de dados do Information Mining.

Uma vez preenchido o arquivo de dados, fica disponível mais uma opção para efectuar a selecção de documentos, nomeadamente a selecção de documentos com base na informação deste arquivo de dados. A operação de pesquisa avançada combina uma consulta textual com uma categoria, restringindo desta forma a pesquisa a documentos que pertençam a uma determinada categoria.

## Utilizar o Information Mining num ambiente empresarial

A infra-estrutura de uma organização que suporte a implementação de qualquer tecnologia de Information Mining geralmente envolve pelo menos as seguintes funções:

- Um administrador de sistemas para a generalidade das TI, não necessariamente restrito ao Information Mining
- Um programador de aplicações
- Um documentalista ou engenheiro especializado
- Pessoal que trabalhe com uma aplicação Information Mining (utilizadores finais)

Poderá encontrar também, consoante a natureza da organização, as seguintes funções mais específicas, além das mencionadas acima:

- Um Web designer
- Um arquitecto ou consultor

A Figura 2 na página 45 ilustra estas funções e acções do Information Mining.

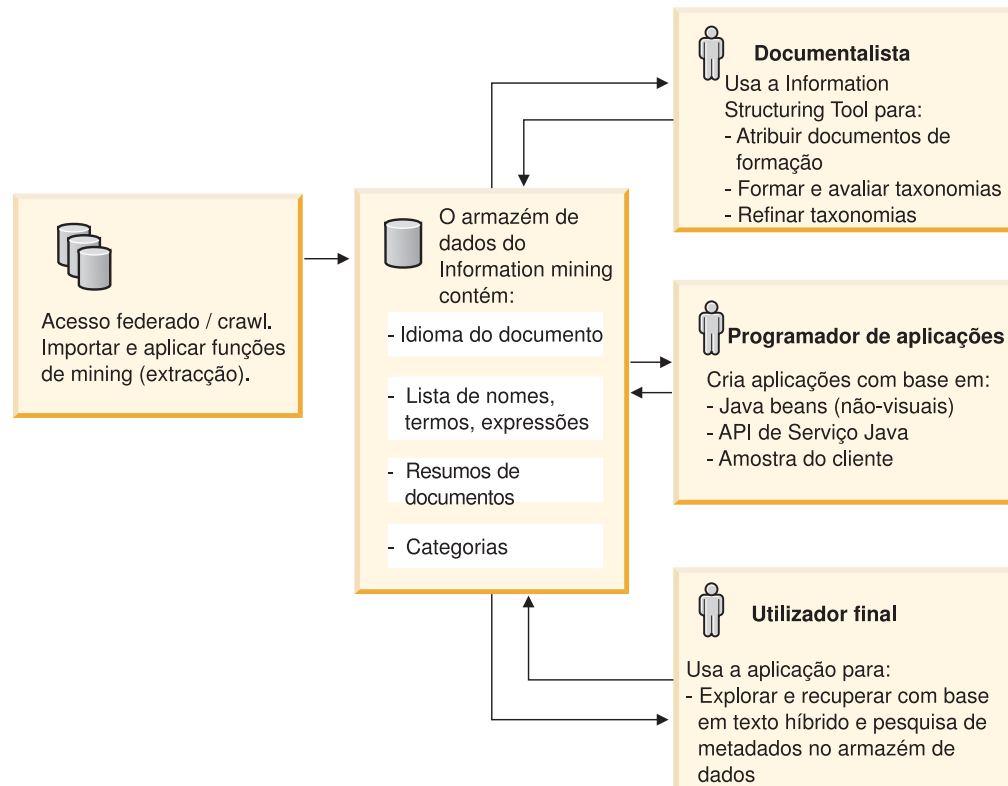


Figura 2. As funções e as acções do Information Mining

O *administrador do sistema* configura o ambiente de hardware e software e mantém os recursos necessários, por exemplo, o espaço do sistema de ficheiros e os direitos de acesso. O administrador do sistema instala os componentes necessários do Enterprise Information Portal e configura as origens de conteúdos e a aplicação de administração do Enterprise Information Portal de forma a possibilitar o acesso aos vários repositórios. O administrador do sistema também administra a memória de metadados do Information Mining a um nível de base de dados.

O *programador de aplicações* cria uma aplicação utilizando JavaBeans ou a API de Serviços. Consulte o *Application Programming Guide for Windows* para obter exemplos de JavaBeans ou o *Manual de Programação de Aplicações* para obter uma descrição da API de Serviços. O programador de aplicações pode fazer isto com a ajuda de um *Web designer*.

O *documentalista* ou o *engenheiro do conhecimento* é responsável pela configuração e manutenção dos recursos e das recolhas de documentos a utilizar para exploração (mining) e obtenção. O documentalista utiliza a aplicação de administração do Enterprise Information Portal para criar correlações e modelos de pesquisa e o Information Structuring Tool (ver “Construir uma taxonomia” na página 62) para definir catálogos e a taxonomia. O preenchimento do arquivo de metadados do Information Mining com documentos de fontes de dados ou da Web, utilizando aplicações construídas pelo programador de aplicações, é, normalmente, também da responsabilidade do documentalista.

Os *utilizadores finais* trabalham com a aplicação criada pelo programador de aplicações para executar tarefas de extração e obtenção de informações com base nos recursos que o documentalista ou o engenheiro especializado criaram e mantêm. Consoante seja a distribuição de trabalho entre utilizadores finais e

documentalista, os primeiros também poderão ficar envolvidos na selecção de documentos a partir de servidores de conteúdos e no preenchimento do arquivo de metadados do Information Mining.

## Um exemplo da utilização do Information Mining

A Electro Corp. é uma empresa que fabrica dispositivos electrónicos para o mercado ----. O catálogo da empresa abrange pelo menos cinco produtos diferentes com uma vasta gama de configurações individuais.

O departamento comercial dispõe da informação sobre as preferências dos clientes relativamente a certas áreas de aplicação dos dispositivos. Estes perfis de utilização explicam a forma como os clientes usam os produtos e as diversas opções de configuração. Cada perfil remete para uma estratégia de marketing, ----- e gestão de relacionamento específica.

O departamento de serviços dispõe de informações sobre as peças que constituem os dispositivos, o respectivo modo de montagem, quem são os fornecedores dessas peças e sobre a capacidade de manutenção e fiabilidade das mesmas a nível individual.

O departamento de administração de contratos mantém informações sobre revendedores e subcontratados. Do mesmo modo, dispõe de acesso a documentos jurídicos relativos aos termos e condições válidas para os tipos de contratos específicos.

Ultimamente registou-se uma queda acentuada nas vendas a determinado tipo de clientes. A concorrência entre métodos de aplicação dos diversos dispositivos electrónicos alterou-se, assim como as expectativas do cliente, de modo a fazer face aos recentes progressos tecnológicos.

No intuito de acompanhar estas mudanças e de as explorar em benefício próprio, a Electro Corp. constituiu uma equipa para desenvolver uma estratégia que ponha a empresa de volta ao mercado.

A primeira etapa consiste em conceber uma infra-estrutura de TI para aceder a estas informações relevantes de modo a que os responsáveis do planeamento possam tomar decisões rápidas e bem fundamentadas.

Os exemplos deste tipo de informações incluem:

- Dados sobre produtos concorrentes, características, preços e aceitação do cliente
- Conhecimento das forças e fraquezas da linha de produtos da Electro Corp. na perspectiva do cliente
- Tendências e perspectivas nas áreas em que os produtos são mais utilizados

Esta informação reside em variadas fontes de dados, com plataformas de hardware e software e níveis de organização diferentes, por exemplo, ficheiros hierarquizados, indexados ou planos, e tipos de documentos, por exemplo, registos de base de dados ou HTML.

A Figura 3 na página 47 ilustra utilizando o Information Mining.

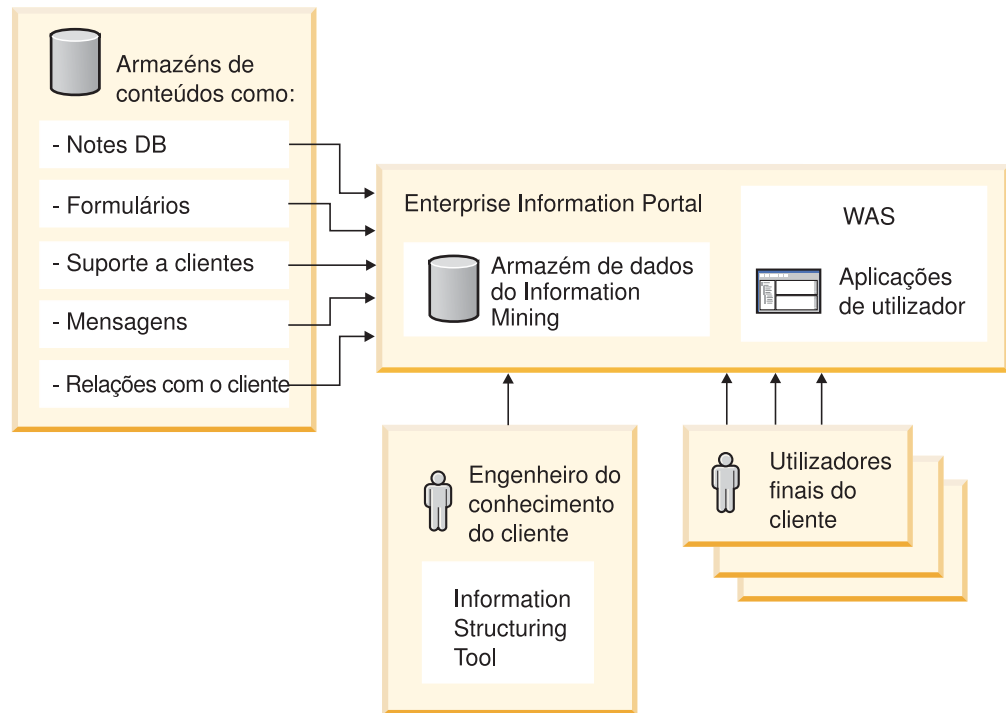


Figura 3. Um exemplo de Information Mining

A equipa decide criar um portal para aceder a todas estas fontes variadas a partir de um cliente ligeiro (um browser) que reside na máquina cliente do responsável do planeamento. Decidem utilizar o Enterprise Information Portal uma vez que este fornece todos os componentes para a criação do portal e permite mesmo a criação de conectores personalizados a fontes de dados mais antigas, criadas quando eram poucos os standards na área das aplicações empresariais.

As etapas de criação do portal são:

1. Configurar a infra-estrutura de hardware e software
2. Definir métodos de acesso a fontes de dados, configurando as ligações necessárias e criando correlações com os dados relevantes.
3. Organizar os dados de modo a que os responsáveis do planeamento possam orientar-se e navegar neles
4. Criar uma aplicação de utilizador final

O Information Mining entra na etapa 3. Uma vez que a infra-estrutura se encontre disponível e em execução, as fontes de dados identificadas, as ligações correspondentes estabelecidas e as correlações relevantes criadas, é possível aceder aos dados necessários a partir de um único ponto e pode-se definir subconjuntos destes dados usando uma pesquisa associada. A questão que se segue consiste em filtrar os dados necessários para uma certa previsão ou tendência, e saber como organizar estes dados de modo a ajustá-los ao processo de planeamento da estratégia.

A equipa define um *engenheiro cognitivo* que seja responsável pela manutenção, organização e actualização das informações de planeamento estratégico. Para extrair as informações relevantes do enorme conjunto de documentos que reside em diversas fontes de dados, o engenheiro cognitivo entrevista o pessoal envolvido

no planeamento estratégico anterior para saber como é a interacção dos processos e quais as boas práticas, além de pesquisar as relações com a clientela e as bases de dados de suporte.

Utilizando as capacidades de pesquisa do Enterprise Information Portal, é possível aceder facilmente aos documentos destas bases de dados através do nome ou morada do cliente ou através das propriedades do dispositivo. No entanto, as informações necessárias para determinar os perfis de utilização estão ocultas no texto e a única maneira de obter estas informações é analisando o conteúdo do documentos de uma forma inteligente utilizando os serviços do Information Mining do Enterprise Information Portal.

Um dos tipos de informação útil facultado por estes serviços consiste na caracterização temática do conteúdo documental denominada categoria, por exemplo, *este documento é sobre PDAs*. O serviço de categorização da Information Mining atribui documentos às categorias mediante análise do respectivo conteúdo. As categorias são estruturadas numa hierarquia de tópicos a que se chama taxonomia. Ambos os metadados explicitamente disponíveis e aqueles criados automaticamente residem em repositórios mantidos pelos chamados catálogos do serviço de Information Mining para ajudar a acelerar o acesso e a obtenção.

Ao utilizar a Information Structuring Tool, o engenheiro cognitivo define um catálogo que ilustra a maneira como os dispositivos são usados pelos clientes.

A Figura 4 na página 48 mostra um catálogo.

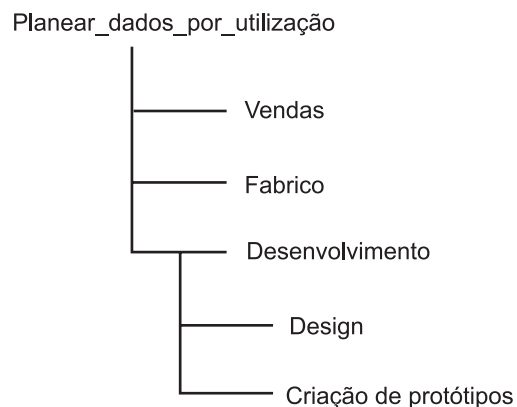


Figura 4. Um exemplo de catálogo

Ao efectuar uma pesquisa associada em todos os clientes e na forma como estes utilizam os dispositivos nas bases de dados de vendas e de suporte, o engenheiro do conhecimento é capaz de identificar conjuntos de documentos de formação representativos de cada uma das categorias na taxonomia. Este facto pode conduzir à reorganização da taxonomia, à medida que novas categorias surgem quando se analisa os dados relevantes com mais rigor.

Uma vez que a taxonomia estabilize e cada categoria tenha um número suficiente de documentos de formação, o engenheiro do conhecimento efectua a formação a uma taxonomia, utilizando a Information Structuring Tool. A formação cria um modelo de categorização que pode ser utilizado para atribuir documentos a categorias utilizando o serviço de Categorização.

Entretanto, os programadores do departamento de informática criam uma aplicação de utilizador final para os planificadores estratégicos utilizando o cliente de poucos recursos do Enterprise Information Portal e JavaBeans ou a API de Serviços Java.

Esta aplicação consiste num número de modelos de pesquisa personalizados para utilização no planeamento estratégico. Utilizando estes modelos, os planificadores podem preencher o catálogo com documentos obtidos a partir de diversos programas de fundo. Ao preencher o catálogo, o serviço do Information Mining automaticamente atribui categorias aos documentos. Se for identificado um novo perfil de utilização, o engenheiro cognitivo reorganiza a taxonomia em conformidade, usando os novos documentos identificados pelos responsáveis do planeamento como sendo material de formação. O catálogo é sujeito novamente a formação e os novos resultados são transmitidos aos responsáveis do planeamento.

O exemplo anterior demonstra como, combinando as funcionalidades do Information Mining de forma a reflectir as expectativas e necessidades do cliente, uma empresa como a Electro Corp. pode fazer face à evolução do mercado e manter-se competitiva.

## Idiomas e Formatos Suportados

Os serviços do Information Mining do Enterprise Information Portal suportam os seguintes idiomas (ver Tabela 5):

*Tabela 5. Idiomas suportados*

| Idioma                              | Identificação do idioma | Exploração de Informação | Resumo | Categorização | Agrupamento em Unidades |
|-------------------------------------|-------------------------|--------------------------|--------|---------------|-------------------------|
| Inglês                              | x                       | x                        | x      | x             | x                       |
| Alemão                              | x                       |                          | x      | x             | x                       |
| Francês                             | x                       |                          | x      | x             | x                       |
| Dinamarquês                         | x                       |                          |        |               |                         |
| Finlandês                           | x                       |                          |        |               |                         |
| Italiano                            | x                       |                          | x      | x             | x                       |
| Norueguês                           | x                       |                          |        |               |                         |
| Português                           | x                       |                          | x      | x             | x                       |
| Espanhol                            | x                       |                          | x      | x             | x                       |
| Sueco                               | x                       |                          |        |               |                         |
| Coreano                             | x                       |                          | x      | x             | x                       |
| Japonês                             | x                       | x                        | x      | x             | x                       |
| Chinês (Tradicional e Simplificado) | x                       |                          | x      | x             | x                       |

Para uma lista dos formatos de documentos suportados, consulte Capítulo 9, “Formatos de Documentos”, na página 123.

## Conceitos

A quantidade de informação a que se acede está em constante crescimento. A maioria das organizações dispõem de um número cada vez maior e sempre em crescimento de documentos online que contêm informações com um grande potencial de valor, por exemplo, dados de retorno dos clientes, informações estratégicas vitais num mercado cada vez mais competitivo ou informações que facultam conhecimentos para novas e mutáveis oportunidades de negócio. Os serviços do Information Mining são concebidos para serem utilizados como aplicações que lidam com grandes quantidades de documentos online.

## Arquitectura do sistema

A Figura 5 ilustra a arquitectura do sistema Information Mining.

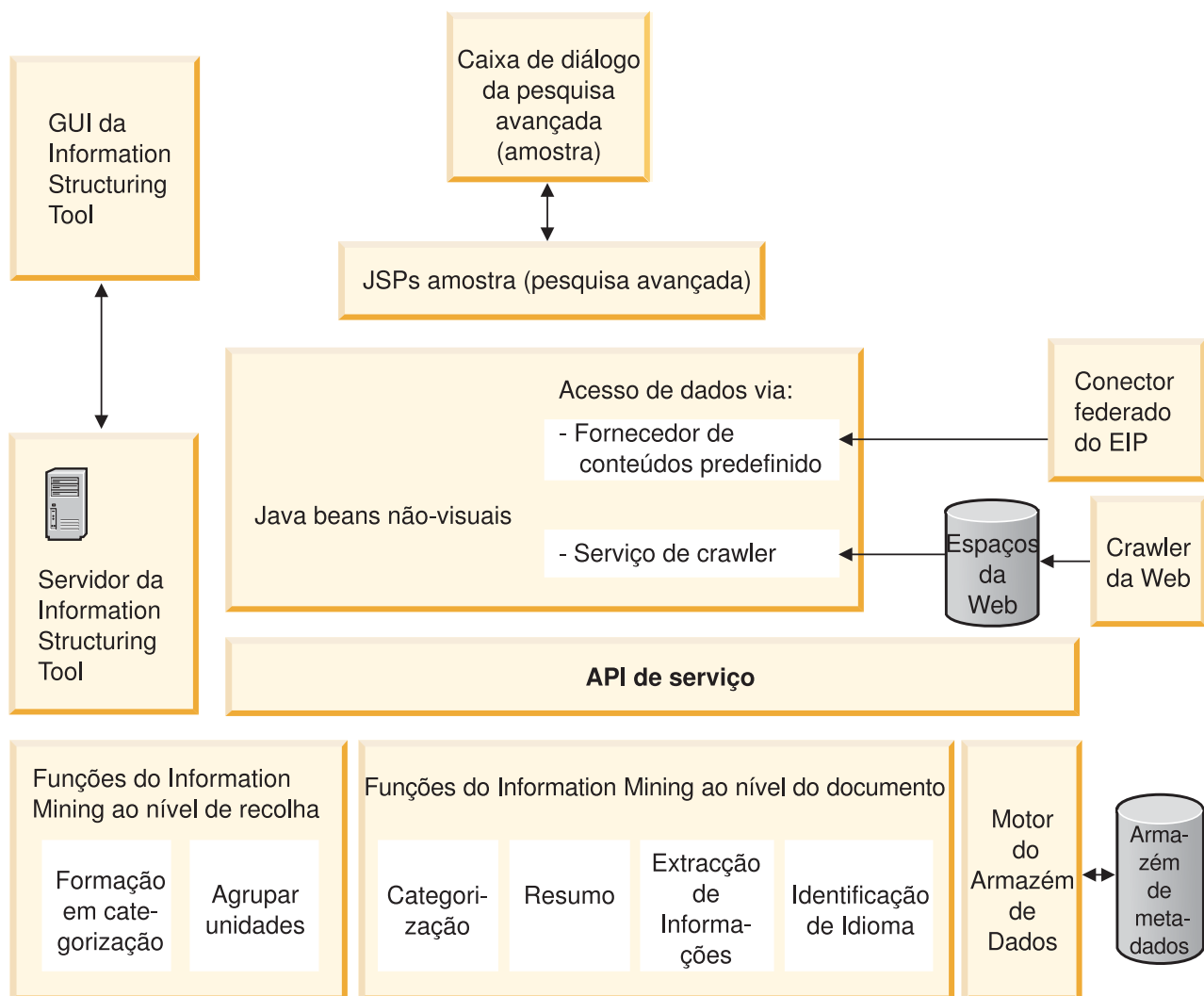


Figura 5. Arquitectura do sistema do Information Mining

As caixas do lado direito do diagrama referem-se aos componentes utilizados pelos serviços do Information Mining mas não são considerados como fazendo parte dos serviços, nomeadamente:

- conector federado (parte da API OO)
- Web Crawler



Existem dois níveis diferentes de funcionalidade do Information Mining, designadamente:

1. API de Serviços Java

Este nível revela a funcionalidade e a persistência dos metadados do Information Mining como uma API de Java coerente.

2. JavaBeans não-visuais

Este nível utiliza componentes prontos a utilizar baseados nas especificações de JavaBeans que aplicam tipos de eventos e convenções provenientes de beans padrão.

3. Páginas de Servidor de Java Exemplificativas

Este nível é constituído por código exemplificativo que utiliza JavaBeans não-visuais e demonstra uma aplicação para pesquisa avançada, isto é, pesquisa textual com restrição por categorias.

4. Information Structuring Tool

Uma aplicação com uma interface gráfica de utilizador para criação e manutenção de taxonomias.

## Os conceitos do Information Mining

Para compreender na totalidade e conseguir utilizar eficientemente a funcionalidade do Information Mining, esta secção trata dos conceitos principais.

Os serviços do Information Mining fornecem uma infra-estrutura para a criação e manutenção de informações relacionadas com documentos individuais ou com recolhas de documentos. Estas informações sobre um documento são denominadas de **metadados**.

A **biblioteca** é uma vista conceptual dos conteúdos da base de dados do Information Mining. A biblioteca contém um conjunto de catálogos.

Um **catálogo** é o arquivo dos metadados para os documentos de texto e contém:

- Um **esquema de catálogo** que define quais os atributos que são armazenados para cada documento.
- Uma **taxonomia** que é uma estrutura hierárquica em árvore de **categorias**.
- Um **modelo de categorização** baseado nos resultados da formação do documento, que pode ser utilizado para atribuir categorias automaticamente a documentos. Este modelo é gerado pela uso da **Information Structuring Tool** onde uma taxonomia pode ser criada e sujeita a formação. O modelo serve como dados de entrada ao serviço de categorização.

O esquema especifica os nomes e tipos de atributos que podem ser gerados ou armazenados para um documento no catálogo. O esquema está predefinido e contém os seguintes atributos:

- IKF\_CONTENT da cadeia de tipo
- IKF\_TITLE da cadeia de tipo
- IKF\_AUTHOR da cadeia de tipo
- IKF\_CATEGORIES da cadeia de tipo
- IKF\_SUMMARY da cadeia de tipo
- IKF\_LANGUAGE da cadeia de tipo
- IKF\_FEATURES da cadeia de tipo
- IKF\_COMMENTS da cadeia de tipo

- IKF\_DATE da marca de hora de tipo
- IKF\_IDNUMBER do número inteiro de tipo

O catálogo cria um **registo** de acordo com o esquema de catálogo para armazenar as informações extraídas ou criadas a partir de um documento importado. Um registo tem um identificador único e um conjunto de pares de valores de nomes. O identificador único, conhecido como um ID de Objecto Persistente, ou PID, estabelece a ligação dos registos criados de volta à fonte do documento original.

A Figura 6 na página 52 ilustra um registo exemplificativo.

| Registo        |   |
|----------------|---|
| IKF_TITLE      | "Aves"                                      |
| IKF_AUTHOR     | "J. Smith"                                  |
| IKF_SUMMARY    | "Este é um resumo do livro denominado Aves" |
| IKF_CATEGORIES | Aves/Insectívoros                           |
| IKF_DATE       | 07/01/2001                                  |

Figura 6. Um registo exemplificativo

Se definir os valores de um registo de Information Mining utilizando JavaBeans ou a API de Serviços, certifique-se de que os valores que vai utilizar se encontram dentro dos limites de tamanho definidos. Caso contrário, receberá a mensagem `DKIKFSizeOutOfBoundsException`. Os limites definidos são:

Tabela 6.

| Chave          | Tamanho máximo em bytes |
|----------------|-------------------------|
| IKF_CONTENT    | 209715                  |
| IKF_TITLE      | 2048                    |
| IKF_AUTHOR     | 2048                    |
| IKF_CATEGORIES | 8192                    |
| IKF_SUMMARY    | 8192                    |
| IKF_LANGUAGE   | 8                       |
| IKF_FEATURES   | 524288                  |
| IKF_COMMENTS   | 8192                    |

Uma vez criado um registo, é armazenado no catálogo, atribuindo-o a uma categoria. Normalmente, os resultados da categorização podem ser utilizados para seleccionar a categoria adequada, embora a categoria também possa ser escolhida dependendo de outro valor armazenado no registo. Os registos tem de ser atribuídos a uma categoria pois esta opção inclui também a indexação do conteúdo do documento para permitir a pesquisa de texto. Cada catálogo tem um índice único de texto o que significa que todos os resultados de pesquisa encontram-se sempre automaticamente no âmbito de pesquisa do catálogo.

O **motor do arquivo de dados** é o componente que mantém o acesso ao arquivo de dados persistente.

## As ferramentas do Information Mining

Os serviços do Information Mining fornecem funções para lidar com documentos online. Incluem:

- A Information Structuring Tool que cria e mantém catálogos.
- O serviço de Identificação de Idiomas que detecta automaticamente o idioma em que um documento está escrito.
- O serviço de Categorização que atribui automaticamente documentos a categorias que definiu previamente através da Information Structuring Tool.
- O serviço de Resumo que analisa palavras e frases num documento para produzir um resumo do documento.
- O serviço de Information Extraction (Extracção de Informações) que reconhece automaticamente artigos significativos no texto, sem solicitar ao utilizador que defina um vocabulário dependente de domínio.
- O serviço de Agrupamento em Unidades que divide um conjunto de documentos em grupos ou conjuntos de unidades. Os documentos em cada conjunto de unidades partilham características comuns. Os conjuntos de unidades não são predefinidos; derivam automaticamente.
- A Pesquisa Avançada que pesquisa texto em documentos armazenados no catálogo, limitado a categorias específicas.

### O Information Structuring Tool

O Information Structuring Tool é uma aplicação baseada na Web que permite criar e manter um conjunto de catálogos denominado uma biblioteca. Usa-se um catálogo para armazenar metadados extraídos de um documento e está associado a uma taxonomia que organiza documentos respeitando uma organização predefinida. Uma taxonomia consiste numa estrutura hierárquica de categorias que classificam os documentos segundo o respectivo conteúdo temático.

Por exemplo, utilizando o Information Structuring Tool, um documentalista pode definir um catálogo que ilustra hábitos alimentares de aves.

A Figura 7 na página 54 mostra um catálogo exemplificativo.

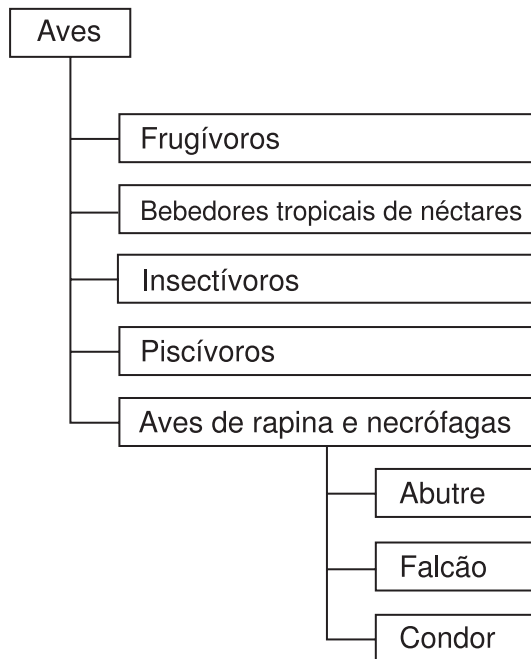


Figura 7. Um catálogo exemplificativo

Um sistema de categorias bem estruturado auxilia de forma significativa a encontrar as informações relevantes num grande volume de dados. As categorias são escolhidas para que correspondam à utilização pretendida de um grupo de documentos e têm de ser antecipadamente sujeitas a testes, utilizando para isso documentos exemplificativos. O modelo de categorização criado utilizando a Information Structuring Tool pode ser utilizado subsequentemente pelo serviços de categorização para atribuir automaticamente categorias a documentos.

As funções da Information Structuring Tool incluem:

- Criar, renomear e eliminar catálogos.
- Editar a descrição de um catálogo.
- Criar, renomear e eliminar categorias.
- Adicionar e remover documentos de formação de e para uma categoria.
- Visualizar o conteúdo de um documento de formação.
- Iniciar e parar o processo de formação do catálogo.
- Obter retorno de informações sobre a qualidade dos dados para formação num catálogo.

Para informações mais detalhadas sobre como instalar e utilizar a Information Structuring Tool, consulte “Construir uma taxonomia” na página 62.

### Identificação de Idioma

O serviço de Identificação de Idioma selecciona, num dado conjunto de idiomas, o idioma mais provável no qual está escrito um documento de texto.

O serviço de Identificação de Idioma devolve uma lista de idiomas e de valores de fiabilidade ordenados para cada documento. A chave do idioma está especificada num código de duas letras de acordo com o ISO Standard 639. O valor de fiabilidade é uma medida que indica até que ponto o documento se encaixa no idioma e é representado por um número flutuante que vai de 0 (mau) a 1 (bom). O algoritmo de identificação de idioma é concebido para determinar o idioma em

documentos mono-linguísticos. Por isso, para documentos com diversos idiomas, a classificação por valor de fiabilidade não é garantia de uma correcta identificação do idioma do documento.

Podem ser detectados os seguintes idiomas:

- Inglês EN
- Alemão DE
- Francês FR
- Dinamarquês DA
- Finlandês FI
- Italiano IT
- Norueguês/Bokmal NB
- Norueguês/Nynorsk NO
- Português PT
- Espanhol ES
- Sueco SV
- Coreano KO
- Japonês JA
- Chinês simplificado e tradicional ZH

Pode ser definida a seguinte propriedade:

- **maxResults** (apenas quando se utiliza a API de Serviços Java):  
O número máximo de idiomas a determinar e devolver para cada documento. É um valor inteiro maior ou igual a 0. O valor predefinido é 1, o que significa que é devolvido o resultado com a melhor classificação. Se o valor for definido para 0, todos os idiomas reconhecidos são devolvidos numa ordem de classificação, excepto se o valor de fiabilidade para o idioma for abaixo de 0,01.

Pode utilizar a identificação de idioma como uma etapa de pré-processamento para os outros serviços do Information Mining. Por exemplo, pode utilizar a identificação de idioma para localizar todos os documentos em Inglês ou em Japonês antes de proceder a extracção de informações.

## Categorização

A categorização é um meio de atribuição de categorias a documentos e de organização de documentos, obedecendo a um modelo organizacional predefinido, criado pela Information Structuring Tool.

Por isso, antes de poder utilizar o serviço de categorização, tem de definir e uma taxonomia para criar tal modelo, utilizando para isso a Information Structuring Tool.

O resultado da categorização contém a categoria e um valor de fiabilidade que indica até que ponto o documento se encaixa na categoria. É devolvido um conjunto de tais resultados para cada documento. A lista é classificada de acordo com os valores de fiabilidade devolvidos.

Podem definir-se as seguintes propriedades:

- **maxResults:**  
O número máximo de categorias a devolver para cada documento. O valor predefinido é -1, o que significa que são devolvidas todas as categorias. A lista de resultados é ordenada.
- **minConfidence:**  
O valor de fiabilidade atribuído a um documento indica de que forma o documento se adequa à categoria. O parâmetro minConfidence especifica o valor

mínimo desde 0 (não muito adequado) e 1 (mais adequado). O valor predefinido é 0, o que significa que são devolvidas todas as categorias atribuídas a um documento. A lista de resultados é ordenada desde o mais adequado até ao pouco adequado.

- **catalogName:**

Especifica o catálogo que vai ser utilizado na elaboração de categorias. Um catálogo pode ser criado e adaptado para a elaboração de categorias através de Information Structuring Tool.

Consulte “Idiomas e Formatos Suportados” na página 49 para obter uma lista dos idiomas suportados.

## Resumo

O resumo de um documento consiste num grupo de frases extraídas de um documento, reunindo características do conteúdo do documento. A ferramenta de Resumo pode, por exemplo, ajudar a decidir se um documento é relevante e se deve ser lido na totalidade ou, quando é devolvido como parte do resultado de uma consulta, ajudar a decidir se se deve seguir a ligação do documento.

O resultado do resumo contém o resumo como uma cadeia única e como uma estrutura de dados (uma matriz) que pode ser utilizada por uma aplicação para seleccionar frases individuais e determinar se ocorrem próximo umas com as outras.

O serviço de Resumo pode ser utilizado em diferentes modos. Os modos determinam como os valores `maxLength` e `proporção` são utilizados para determinar o comprimento do resumo.

- **maxLength:**

O número máximo de frases no resumo. O resumo criado não será maior do que `maxLength`. O valor predefinido é 3.

- **proporção:**

O número de frases relativas ao comprimento total do documento. O comprimento do resumo criado é determinado pelo comprimento total do documento. O valor predefinido é 0.1.

- **modo:**

Determina a relação entre `maxLength` e `proporção` necessárias para a definição do comprimento do resumo. Os modos diferentes são:

- **MODE\_LESS\_THAN\_MAXLENGTH:**

O resumo tem no máximo um número de frases de `maxLength`. Este é o modo predefinido.

- **MODE\_EQUALS\_RATIO:**

O resumo tem exactamente o número proporcional de frases. Este é determinado através da multiplicação da `proporção` pelo número total de frases no documento.

- **MODE\_EQUALS\_RATIO\_BUT\_AT\_MOST\_MAXLENGTH:**

O resumo tem pelo menos o número proporcional de frases (`proporção` vezes o número total de frases no documento), mas não mais do que o número de frases de `maxLength`.

Consulte “Idiomas e Formatos Suportados” na página 49 para obter uma lista dos idiomas suportados.

## Extracção de Informações

Uma tarefa importante para a análise de documentos é a extracção de artigos que fornecem informações sobre o conteúdo do documento. Estes elementos chave podem ser utilizados:

- Para indicar informações importantes que auxiliem na avaliação do interesse de um documento
- Para encontrar e armazenar conceitos chave para utilizar num aperfeiçoamento da consulta
- Como critérios para recolher documentos relacionados

Como exemplos de elementos chave temos artigos de vocabulário, tal como palavras, nomes ou termos com várias palavras.

Para Inglês, o serviço de Extracção de Informações normaliza os elementos chave que encontra e agrupa ocorrências de tais elementos chave num texto caso se refiram à mesma entidade ou expressem o mesmo conceito. Por exemplo, se José J. Silva, Sr. Silva, José e Silva ocorrerem num documento, são todos marcados como referindo-se à mesma pessoa, correlacionando-os para o mesmo formato normalizado. As formas inflectidas das palavras também são correlacionadas para o respectivo formato normalizado, por exemplo, crianças para criança.

Para o Japonês, no entanto, todos os elementos chave são extraídos à medida que ocorrem no documento. Não é efectuada qualquer normalização excepto para a data, hora e expressões de moeda, em que a normalização está em conformidade com o ISO8601 e o ISO4217.

O serviço de Extracção de Informações permite-lhe analisar documentos em busca de:

- Artigos de vocabulário de palavras únicas ou múltiplas, por exemplo, anúncio, valor, ciclo de produto
- Nomes de lugares, pessoas e organizações, por exemplo, Lisboa, Sampaio, Instituto Nacional de Estatística
- Abreviaturas, por exemplo, MB (megabytes)
- Termos para datas, dinheiro e números, por exemplo, 11 Jan. 1958, 01/11/58, ?30, trinta cêntimos, 4.5, 5000

Pode especificar três tipos de informações a extrair:

- Nome
- Termo
- Expressão

Através da API de Serviços Java, pode também identificar os subtipos listados abaixo e um valor de fiabilidade que indique de que forma o subtipo se encaixa no artigo de vocabulário extraído. O valor de fiabilidade vai de 0 (mau) a 1 (bom). Os subtipos para o nome do tipo, termo e expressão incluem:

- **Nome**
  - Local, por exemplo, Montreal ou Londres
  - Pessoa, por exemplo, Tiago Ribeiro
  - Organização, por exemplo, Silva e Filho
  - Desconhecido, por exemplo, Smashing Pumpkins, Silicon Valley, CCTV (abreviatura sem um formato completo)

- Outros, por exemplo, AIS Plan, ISO Conference, Internet, Privacy Act Officer, JCAHO Performance Report
- **Termo**
  - Termo não especificado, por exemplo, entretenimento, aglomerado, mundo da arte, variável de classe, código fonte, definição de dados, iniciativa de melhoramento de processo
- **Expressão**
  - Cardinal, por exemplo, quatro, cinquenta, 70
  - Ordinal, por exemplo, quarto, quinquagésimo
  - Percentagem, por exemplo, 12%, sessenta por cento
  - Data, por exemplo, 28/07/98
  - Hora, por exemplo, 18 hrs, 4 horas
  - Dinheiro, por exemplo, DM90, trinta libras
  - Abreviatura, por exemplo, NY

A extracção de informações funciona apenas em documentos em língua inglesa e japonesa. Pode usar o serviço de Language Identification (Identificação de Idioma) como uma etapa de pré-processamento para identificar documentos no seu grupo de documentos que não se encontrem em Inglês ou em Japonês. Pode combinar o serviço Information Extraction com outra funcionalidade da exploração (mining), por exemplo, utilize-o como uma etapa de pré-processamento para o serviço de Resumo para resumir apenas os documentos sobre Jorge Sampaio como Presidente da República e não como Presidente da Câmara de Lisboa.

### **Agrupamento em Unidades**

A ferramenta de Agrupamento Unidades ordena uma recolha de documentos de forma a que os documentos similares fiquem agrupados conjuntamente e que os documentos em grupos diferentes (conjuntos de unidades) sejam distintos uns dos outros no que diz respeito ao respectivo conteúdo. Desta forma, o agrupamento em unidades pode ser utilizado como um meio para fornecer uma vista geral de uma recolha de documentos de grandes dimensões e identificar documentos relacionados. Pode, da mesma forma, ser utilizado para suportar a construção de uma taxonomia através da Information Structuring Tool, agrupando em unidades os documentos de formação numa área de aplicações. O agrupamento em unidades também é útil para encontrar documentos similares num grupo que possam direccioná-lo para novas tendências ou novas tecnologias e duplicados, ou documentos muito parecidos, que possam ter interesse do ponto de vista da análise competitiva.

O agrupamento de unidades é um processo iterativo que organiza documentos em conjuntos de unidades de forma a que os documentos em cada conjunto sejam tão semelhantes quanto possível, no que diz respeito ao conteúdo, e os conjuntos de unidades sejam tão diferentes quanto possível uns dos outros. O agrupamento opera sobre a recolha de documentos na sua totalidade, ao contrário dos serviços do Information Mining indicados acima, como a categorização ou o resumo, que operam ao nível do documento. O agrupamento de unidades funciona comparando as características representativas de cada documento com os outros e agrupa os documentos de acordo com as similaridades das características.

Durante uma fase de agrupamento de unidades, não podem ser adicionados novos documentos ao conjunto de documentos.

Podem definir-se as seguintes propriedades:



- **maxClusterCount**  
O número máximo de conjuntos de unidades a devolver.
- **minClusterCount**  
O número mínimo de conjuntos de unidades a devolver.
- **clusterFeatureCount**  
O número de etiquetas (palavras-chave) devolvidos por conjunto de unidades.

Estes valores, no entanto, não são vinculativos para quem efectua o agrupamento, mas têm o propósito de servir como meros limites de directrizes. Os dados resultantes do serviço de Agrupamento em Unidades apresentam-se como uma lista de resultados.

O Agrupamento de Unidades funciona apenas em documentos de língua inglesa. Pode usar o serviço de Language Identification (Identificação de Idioma) como uma etapa de pré-processamento para identificar documentos no seu grupo de documentos que não se encontrem em Inglês.

## Pesquisa Avançada

Ao contrário do que acontece com uma pesquisa de Enterprise Information Portal padrão, que é executada na totalidade do Servidor de Conteúdos do Enterprise Information Portal, a pesquisa avançada pesquisa apenas os documentos cujos IDs se encontrem guardados num catálogo criado pelo Information Structuring Tool. Para limitar ainda mais a pesquisa, a consulta de pesquisa avançada não pesquisa apenas texto, mas também pode restringir esta pesquisa a documentos em categorias específicas.

Podem ser definidos os seguintes parâmetros:

- **catalogName:**  
Esta opção especifica o catálogo a ser utilizado para a pesquisa. Um catálogo pode ser criado e sujeito a formação utilizando a Information Structuring Tool.
- **maxResults:**  
O número máximo de resultados da pesquisa devolvido para cada consulta. O valor predefinido é 0 e significa que são devolvidos todos os resultados.

Podem ser submetidos os tipos seguintes de consultas:

1. Consulta apenas de texto. Esta pesquisa devolve todos os documentos que correspondam à consulta de texto. A lista de resultados é ordenada por relevância.
2. Pesquisa apenas de categoria. Esta pesquisa devolve todos os documentos atribuídos à categoria. Os resultados são apresentados numa ordem arbitrária.
3. Pesquisa combinada de texto e categoria. Esta pesquisa devolve todos os documentos que correspondam a consulta de texto e que estejam atribuídos à categoria. A lista de resultados é ordenada por relevância.

As consultas de pesquisa avançada submetidas ao sistema estão sempre associadas a um catálogo específico. Denomina-se âmbito de pesquisa de catálogo. Não é possível qualquer pesquisa cruzada de catálogos pois os catálogos representam uma vista dos documentos importados que tem de ser respeitada.

O BNF (sintaxe da consulta) para uma cadeia de consultas é o seguinte:

```
query_string    ::= term
term            ::= ( term )
               ::= single_term
               ::= compound_term
```

```

single_term          ::= category_term
                    ::= text_search_term
                    ::= string_term
                    ::= number_term
compound_term        ::= term binary_bool_operator term
                    ::= unary_bool_operator single_term
category_term         ::= ( DKIKFCategory category_operator category_path_value)
text_search_term      ::= ( "attribute name" CONTAINS text_search_value )
string_term           ::= ( "attribute name" string_operator string_value )
number_term           ::= ( "attribute name" basic_operator number_value )
binary_bool_operator  ::= AND | OR
unary_bool_operator   ::= NOT
category_operator     ::= >= | =
string_operator        ::= LIKE | basic_operator
basic_operator         ::= > | < | <= | >= | != | =
category_path_value   ::= "category path"
text_search_value     ::= "'string'"
string_value           ::= "string"
number_value          ::= "integer" | "decimal number"

```

- Os terminais, cadeia e número, representam termos comuns.
- O category\_operator '=' limita o âmbito da pesquisa a apenas uma categoria.
- O category\_operator '>=' expande a pesquisa a esta categoria e a todas as respectivas subcategorias na árvore de categorias.
- A cadeia de pesquisa numa cláusula CONTAINS pode incluir os caracteres globais ('\_') para um único carácter e ('%') para uma quantidade arbitrária de caracteres. Por exemplo, \_LOB pode corresponder a BLOB e CLOB, enquanto que %nome poderia corresponder a nomeficheiro. Apenas os atributos de esquema que tenham sido sinalizados como pesquisáveis, por exemplo IKF\_CONTENT, podem ser consultados com o operador de cadeias CONTAINS.
- A cadeia de pesquisa numa cláusula LIKE podem também incluir caracteres globais tal como são utilizados em SQL.
- Para uma lista completa de nomes de atributos actualmente suportados, consulte “Os conceitos do Information Mining” na página 51.

Exemplos de consultas:

- Consultas apenas de texto:
 

```
("IKF_CONTENT" CONTAINS "'África do Sul'") AND NOT
("IKF_CONTENT" CONTAINS "'Cabo'")
```
- Pesquisa apenas de categoria:
 

```
("DKIKFCATEGORY" >= "aves/Frutívoros")
```
- Pesquisa combinada de texto e categoria:
 

```
("IKF_CONTENT" CONTAINS "'África do Sul'") AND
("DKIKFCATEGORY" >= "aves/Aves de rapina/Falcão")
```
- Consulta de atributo:
 

```
("IKF_SUMMARY" LIKE "colibris dos tópicos")
```

ou

```
("IKF_FEATURES" LIKE "Goethe") AND ("IKF_TITLE" = "Fausto")
```

## Interfaces de Programação

A funcionalidade do Information Mining está disponível para construir aplicações como:

- API de Serviços Java
- Information Mining JavaBeans

A **API de Serviços Java** engloba todas as funcionalidades do Information Mining, excepto a manutenção de catálogos, que está integrada no Information Structuring Tool, como um serviço de Enterprise Information Portal. Fornece comunicações cliente/servidor baseadas em RMI de Java.

Uma aplicação que utilize a API de Serviços Java pode:

- Determinar o idioma no qual o documento está escrito
- Criar resumos de documentos de texto
- Atribuir categorias a documentos
- Extrair informações, por exemplo nomes, termos ou expressões, de um documento de texto
- Agrupar documentos similares
- Armazenar e procurar metadados para documentos num catálogo
- Executar uma pesquisa de texto em documentos restringida a certas categorias e em atributos, por exemplo um resumo

A API de Serviços Java pode ser executada em modo local utilizando chamadas de método directo ou em modo remoto utilizando Remote Method Invocation (RMI) de Java. A execução em modo remoto permite configurar um servidor como o servidor de aplicações que executa as aplicações Web e outro servidor como o servidor do Information Mining que executa análise de texto, indexação e pesquisa. O mecanismo de tarefas do servidor é o meio através do qual podem ser enviadas tarefas completas para o servidor do Information Mining (computador remoto) e todo o processamento é levado a cabo nesse computador.

Se utilizar o Web Crawler, os mecanismos de acesso correspondentes têm de ser implementados utilizando JavaBeans. O acesso de Web Crawler não está disponível no nível de API de Serviços Java.

Para obter uma descrição detalhada da API de Serviços Java do Information Mining, consulte o *Application Programming Guide for Windows*.

A figura Figura 8 na página 61 ilustra a configuração remota do Information Mining.

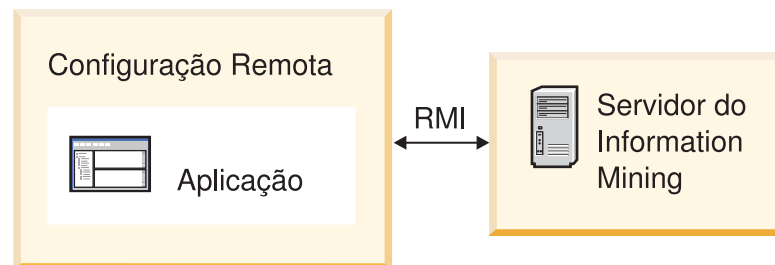


Figura 8. Configuração remota do Information Mining

Os beans do **Information Mining** são uma API de Java de alto nível para desenvolvimento rápido de aplicações e são construídos segundo as especificações dos JavaBeans. Os beans não suportam processamento de tarefas do servidor, por isso, por razões de rendimento, todos os desenvolvimentos de aplicações que utilizem os beans devem ser realizados no mesmo computador.

Cada um dos beans pode ser utilizado como sendo direccionado pelos eventos e alguns fornecem métodos que podem ser chamados directamente. A integração

com os beans de Enterprise Information Portal já existentes é conseguida pelo suporte dos eventos de resultados utilizados pelos beans de Enterprise Information Portal. Isto significa que eventos de resultados de Web Crawler e pesquisas federadas são compatíveis com os serviços do Information Mining e os resultados do mesmo podem ser processados pelo EIP. Para obter uma descrição detalhada dos beans do Information Mining, consulte o *Manual de Programação de Aplicações para Windows*.

---

## Primeiros Passos

Primeiros Passos do Information Mining do Enterprise Information Portal é um guia de iniciação em formato de storyboard que demonstra a engenheiros de conhecimento, administradores e programadores de aplicações como a tecnologia Information Mining da IBM pode ser aplicada num ambiente empresarial, usando como exemplo uma situação realista. O guia de iniciação está estruturado da seguinte forma:

- Breve Introdução
- O que é necessário fazer antes de iniciar o módulo Primeiros Passos
- Organizar os dados exemplificativos
- Aceder a estes dados
- Utilizar o cliente exemplificativo
- Remover os dados exemplificativos
- Mais referências para leitura

Para aceder aos Primeiros Passos do Information Mining, execute  
<CMBROOT>\ikf\firststeps\first\_steps.html.

Para executar os Primeiros Passos do Information Mining em Windows, clique no botão Start e seleccione **Enterprise Information Portal for Multiplatforms 8.1** → **<Information Mining>** → **Primeiros Passos**.

Os Primeiros Passos do Information Mining do Enterprise Information Portal também funcionam como uma verificação da instalação.

---

## Construir uma taxonomia

A Information Structuring Tool é uma aplicação baseada na Web que faculta um meio de criar e manter um conjunto de catálogos, denominado biblioteca. Usa-se um catálogo para armazenar metadados e o mesmo está associado a uma taxonomia que organiza as respectivas informações. A taxonomia consiste numa estrutura hierárquica de categorias que classificam os documentos segundo o respectivo conteúdo temático. Por exemplo, o nível superior de uma taxonomia pode conter categorias como *negócios*, *cultura*, *desporto*, enquanto *desporto* se subdivide em *desportos de equipa* e *atletismo* no nível imediatamente abaixo. Um nível mais abaixo e *desporto de equipa* subdivide-se em *futebol*, *basquetebol*, *ténis*.

Ao atribuir documentos de formação a categorias e ao efectuar a formação de um catálogo, a Information Structuring Tool cria um modelo de categorização que pode ser usado subsequentemente pelo bean do serviço de categorização para atribuir categorias a um documento.

## Instalar a Information Structuring Tool

A Information Structuring Tool deve ser implementada como uma aplicação de Web num contentor de servlet, como, por exemplo, o motor de servlet WebSphere Application Server (WAS) da IBM.

Porém, tenha presente que não é permitido que duas aplicações de Web de Information Structuring Tool (por exemplo, uma denominada IST1 e a outra, IST2) trabalhem sobre a mesma instância do Information Mining.

Antes de implementar a Information Structuring Tool, certifique-se de que foi instalada e configurada a versão correcta do WAS. Consulte *Planeamento e Instalação do Enterprise Information Portal* para obter informações mais detalhadas.

Os direitos de acesso do utilizador necessários para a implementação da Information Structuring Tool são:

- Para Windows: Autoridade de administrador
- Para AIX: Privilégios de utilizador root
- Para Sun: Privilégios de utilizador root

## Como começar

A aplicação faculta uma interface baseada na Web onde se podem definir, manter e efectuar formação a taxonomias. Existem duas molduras. A da esquerda chama-se vista de *catálogo* e usa-se para criar e manter as taxonomias. A da direita chama-se *bloco de notas* e dá informações sobre a utilização da aplicação. O bloco de notas apresenta uma série de separadores para formação e avaliação de documentos para o catálogo. Para poder usar a aplicação, é preciso criar primeiro catálogos e categorias na moldura da vista de catálogo do lado esquerdo.

## Direitos de acesso

Na Information Structuring Tool, as informações de nome de utilizador e de palavra-passe são mantidas pelo Enterprise Information Portal. Só pode introduzir um nome de utilizador e uma palavra-passe conhecidos por Enterprise Information Portal.

Quando se inicia o Information Structuring Tool, é apresentada uma mensagem de aviso de segurança porque a Applet Java utilizado para carregar documentos de treino requer acesso de leitura ao sistema de ficheiros. Se recusar, não lhe será permitido descarregar documentos de formação e, por isso, não poderá efectuar formação de taxonomias.

A Information Structuring Tool pode ser executada num ambiente de vários utilizadores. Para permitir que vários utilizadores visualizem a mesma taxonomia, a Information Structuring Tool faculta um mecanismo de bloqueio para controlar o acesso a um catálogo e respectivas categorias.

Um utilizador pode optar por bloquear explicitamente um catálogo, seleccionando-o e bloqueando-o antes de começar a trabalhar com ele, ou pode começar a trabalhar com o catálogo, por exemplo, adicionando documentos de treino, e este é automaticamente bloqueado para evitar conflitos de acesso. Os outros utilizadores podem ver este catálogo mas não podem alterar o que quer que seja no mesmo até que seja novamente desbloqueado pelo utilizador que o bloqueou.

Tenha em atenção que se o servidor de aplicações em que estiver distribuída a Information Structuring Tool for encerrado, serão eliminados todos os bloqueios.

## Definir uma taxonomia

O catálogo é um ponto de ancoragem de uma taxonomia que é uma estrutura tipo árvore que consiste em categorias.

Os passos para definir um novo catálogo e seleccionar categorias apropriadas são:

1. Decida quais as categorias que pretende definir e crie uma nova taxonomia.  
Na vista catálogo, seleccione **Biblioteca** e clique com o botão direito do rato. Surge um menu. Seleccione **novo catálogo** e será criado um novo ícone de catálogo. Mude o nome deste ícone introduzindo o nome do catálogo e clique no botão **Introduzir**. É criada uma pasta com o mesmo nome. Esta é a categoria raiz. O conteúdo do bloco de notas é alterado.

Um outro método alternativo para adicionar um novo catálogo à Biblioteca é importar uma taxonomia já existente, criada fora da Information Structuring Tool, por exemplo, no sistema de ficheiros. Consulte “Carregar documentos de formação” na página 66 para obter mais informações.

Para trabalhar activamente numa taxonomia, o catálogo tem que estar bloqueado para si. Ao criar um catálogo, este procedimento é automático. Para bloquear um catálogo existente, seleccione o catálogo, e clique com o botão direito do rato. É apresentado um menu. Seleccione **bloquear catálogo** e o estado do ícone do catálogo altera-se. Os ícones são utilizados para representar os diferentes tipos de estado do catálogo:



A árvore da taxonomia está resumida e não está bloqueada por um utilizador.



A árvore de taxonomia está expandida e não bloqueada por um utilizador.



A árvore de taxonomia está resumida e está bloqueada pelo utilizador actual.



A árvore de taxonomia está expandida e está bloqueada pelo utilizador actual.



A árvore de taxonomia está resumida e está bloqueada por outro utilizador.



A árvore de taxonomia está expandida e está bloqueada por outro utilizador.

A Figura 9 é um exemplo de dois catálogos.

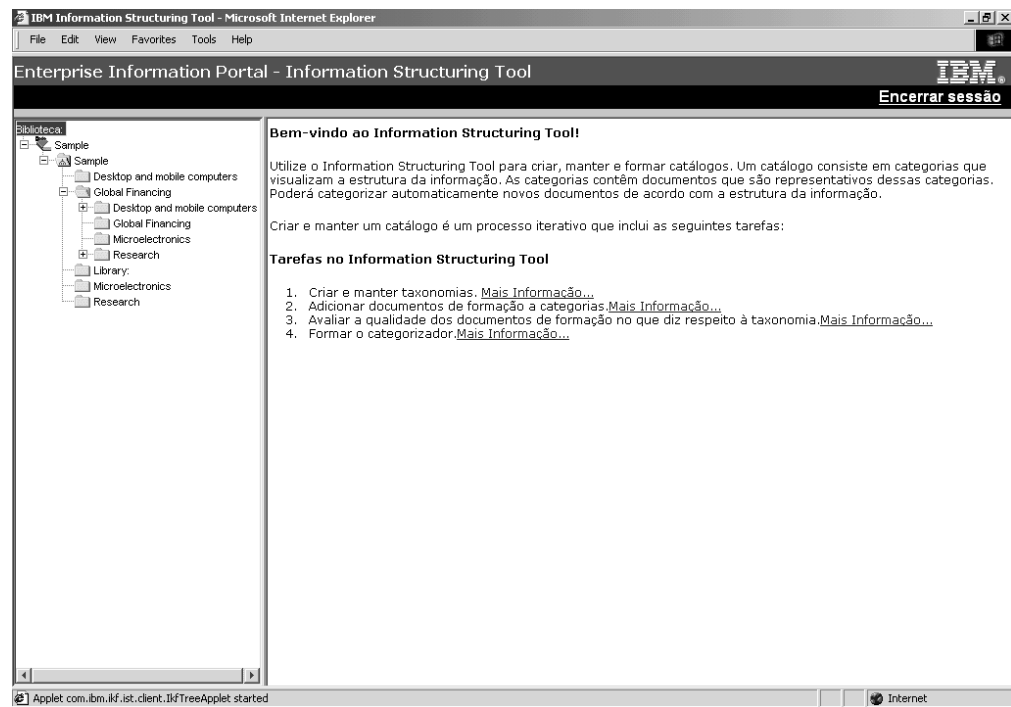


Figura 9. Um exemplo de um catálogo

2. Uma vez criado e bloqueado o catálogo, pode adicionar, mudar o nome ou eliminar as categorias. Quando cria um novo catálogo, é também criada uma categoria raiz com o mesmo nome que o catálogo. Pode mudar o nome a esta categoria. Para fazer isto, seleccione a categoria e clique com o botão direito do rato. Seleccione **mudar o nome**, escreva o novo nome e prima **Enter**.

Para adicionar novas categorias, destaque a categoria à qual pretende adicionar uma nova subcategoria e clique com o botão direito do rato. Seleccione **nova categoria**, escreva o novo nome e prima **Enter**. Os nomes de categoria do mesmo nível da estrutura da árvore devem ser únicos.

A categoria raiz não pode ser eliminada. Apenas pode ser eliminada quando o catálogo for eliminado. A eliminação de uma categoria remove também todas as subcategorias (as categorias inferiores na estrutura em árvore), todos os documentos de treino e todos os registos da categoria.

3. No separador **Propriedades** são apresentadas mais informações descritivas relacionadas com o catálogo seleccionado. Na primeira vez, o campo de descrição está vazio. Para adicionar ou editar uma descrição de catálogo, clique no botão **Editar Descrição**. É apresentada uma janela, na qual pode ser introduzida uma descrição.

## Seleccionar documentos de formação

A qualidade de um modelo de categorização é fortemente dependente da qualidade dos documentos de formação atribuídos a cada categoria.

Os documentos de treino devem estar num dos formatos de documentos suportados pelo Enterprise Information Portal. Consulte Capítulo 9, "Formatos de Documentos", na página 123 para obter uma lista dos formatos suportados.



A selecção de um conjunto apropriado de documentos de formação é essencial. Os documentos devem:

- Ser representativos da categoria
- Conter uma quantidade significativa de texto descritivo, sem demasiadas markups ou listas de palavras
- Todos têm que estar escritos no mesmo estilo de escrita, por exemplo, evitar escrita em prosa se os outros documentos apresentarem um estilo de relatório
- Todos têm de ter um comprimento semelhante, e de preferência não demasiado extensos; os documentos de formação devem ser também de tamanho aproximado ao dos documentos que pretende categorizar utilizando o serviço de Categorização

Recomenda-se uma recolha de cerca de 40 documentos de formação; todavia, se a categoria escolhida for mais generalista serão necessários mais documentos. As categorias têm de ser escolhidas com cuidado e têm de fazer sentido; as categorias e os documentos de formação que pareçam vagos e indistintos a um indexador humano irão, com toda a certeza, colocar problemas durante o processamento automático.

## Carregar documentos de formação

Para adicionar documentos de formação, seleccione a respectiva categoria e será apresentada a janela **Lista de Documentos de formação**. Para adicionar documentos a esta lista, clique no botão **Adicionar Documento...**

Surge a janela **Adicionar Documentos de Formação**. Não necessita fechar esta janela após carregar ficheiros; pode ser utilizada para carregamentos subsequentes de documentos para outra categoria ou outro catálogo.

Para adicionar documentos de formação, clique no botão **Procurar...** e seleccione os ficheiros relevantes ou um directório na janela **Abrir**.

Pode seleccionar um ou mais ficheiros do mesmo directório ou mesmo um directório completo para carregar. Se o directório que seleccionar estiver vazio, será informado do facto.

Este facto permite que importe e trabalhe com taxonomias existentes que foram criadas fora da Information Structuring Tool, por exemplo, no sistema de ficheiros.

Se o directório seleccionado no seu sistema de ficheiros, por exemplo, Desenvolvimento, contiver um subdirectório Design, e se a categoria Desenvolvimento na sua árvore de taxonomia também contiver uma subcategoria Design, os ficheiros no subdirectório são adicionados à subcategoria. Se a subcategoria não existir, é criada, e os ficheiros são adicionados a esta subcategoria agora criada.

Selecione o formato e o idioma dos documentos para todos os ficheiros seleccionados. Com a excepção de ficheiros de texto simples, utilize sempre o formato "detecção automática".

Para adicionar os ficheiros à lista de documentos de formação, clique no botão **Submeter**.

A Figura 10 na página 67 mostra a adição de documentos de formação.



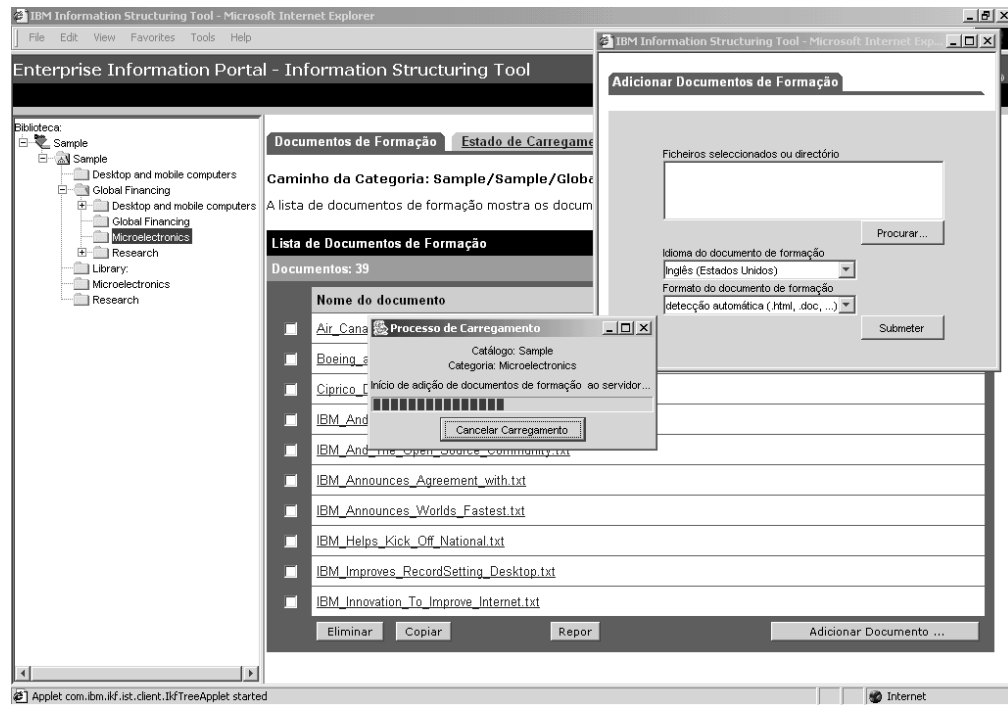


Figura 10. Adicionar documentos de formação

Surge a janela **Processo de Carregamento**. Pode cancelar o processo de carregamento do documento nesta janela. Se o processo de carregamento for cancelado, não são adicionados mais ficheiros. Os ficheiros que já tiverem sido adicionados como documentos de formação são removidos. No entanto, se tiverem sido adicionadas algumas subcategorias durante o carregamento, não serão eliminadas do catálogo.

Se surgir alguma falha ao carregar os ficheiros, é automaticamente mostrada uma janela com o **Estado do Carregamento** com informações sobre a falha no carregamento dos ficheiros. Por exemplo:

- Já existe um ficheiro com o mesmo nome.
- O ficheiro está vazio.
- O ficheiro não pôde ser carregado no servidor.

Na janela de Estado do Carregamento, prima o separador **Documentos de Formação** para voltar à lista de Documentos de Formação. Todos os documentos de formação que foram carregados com êxito na categoria são aqui apresentados. Use os botões **Anterior** e **Seguinte** para visualizar todos os documentos na lista.

Se quiser adicionar o mesmo documento como documento de formação a mais do que uma categoria, carregue o ficheiro na primeira categoria e, de seguida, copie-o para as outras. Não volte a carregar este ficheiro. Para copiar, seleccione um ou mais documentos na janela Lista de Documentos de Formação e prima **Copiar**. Na janela que surge, prima **Procurar...** para seleccionar uma ou mais categorias para as quais pretende copiar os documentos e prima **Submeter**.

Não são permitidas as seguintes acções durante o carregamento de ficheiros:

- Desbloquear o catálogo
- Fechar sessão na Information Structuring Tool

- Iniciar formação ou avaliação do catálogo
- Renomear o catálogo
- Eliminar a informação do estado do carregamento para o catálogo

No entanto, são permitidas as seguintes acções:

- Iniciar outro processo de carregamento de ficheiro, quer para a mesma categoria quer para outra diferente
- Trabalhar noutra catálogo

A Figura 11 mostra a lista de documentos de formação.

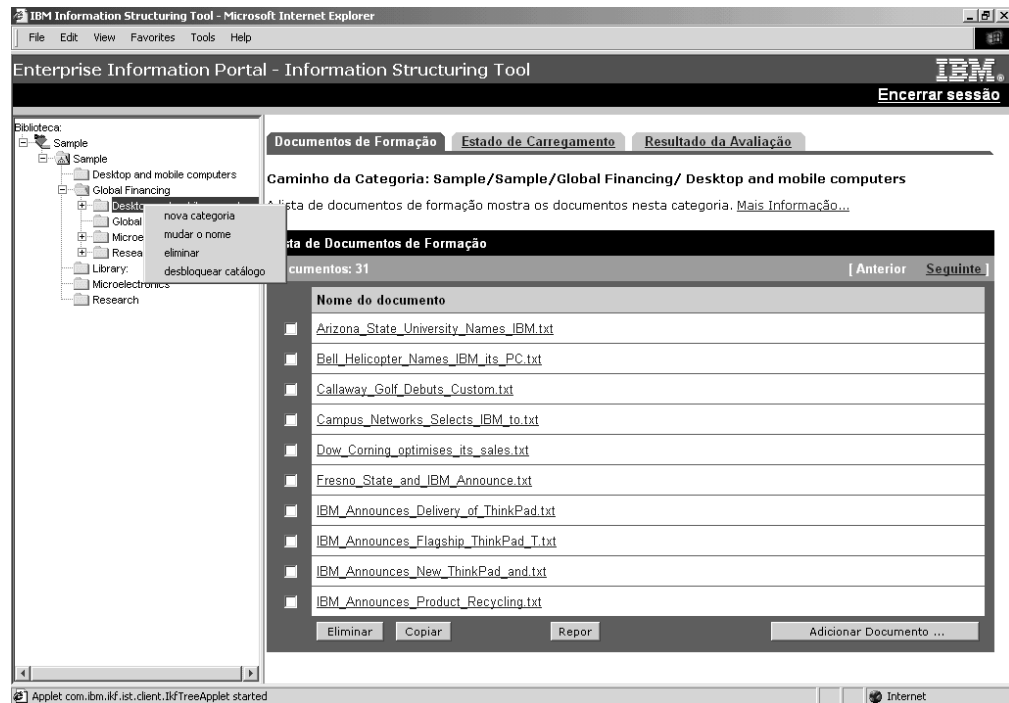


Figura 11. Lista de documentos de formação

Se o browser estiver encerrado durante o carregamento de ficheiros, os ficheiros serão adicionados como documentos de formação caso já tenham sido transferidos para o servidor. Se tal não tiver acontecido, não são adicionados quaisquer ficheiros.

Após ter carregado todos os documentos de formação, seleccione o catálogo para iniciar a avaliação.

## Avaliar um modelo de categorização

Uma vez que a categoria tenha sido definida e que os documentos de formação tenham sido atribuídos a cada categoria, a taxonomia deve ser sujeita a avaliação. A avaliação da taxonomia ajuda a determinar a qualidade dos documentos de formação no que diz respeito à sua taxonomia predefinida. Este é um processo iterativo que consiste nos passos seguintes:

1. Iniciar a avaliação
2. Avaliar os resultados da avaliação
3. Efectuar alterações à taxonomia ou aos documentos de formação

#### 4. Voltar a executar a avaliação

Durante cada iteração da avaliação, o processo de avaliação:

- Divide os documentos de formação num conjunto de formação (aproximadamente 80% dos documentos) e um conjunto de testes (aproximadamente 20% dos documentos). Efectua a formação do catálogo utilizando o conjunto de formação e utiliza o serviço de Categorização no conjunto de testes.
- Verifica se os documentos estão atribuídos às categorias correctas com um valor de fiabilidade suficientemente alto. O intervalo situa-se entre 0 e 1, em que 1 significa que o documento é ideal. Pode definir o valor de fiabilidade. A predefinição é 0.5.

Pode seleccionar entre três e cinco iterações. Três iterações é a predefinição e fornece uma perspectiva com significado dos pontos fortes e fracos da taxonomia. Ao seleccionar cinco, todos os documentos terão estado quer no conjunto de formação quer no de testes.

Prima **Iniciar Avaliação** para iniciar o processo de avaliação.

Para cada categoria, o que se segue é calculado durante cada iteração da avaliação:

- *Corrigir* documentos. O número de documentos de formação na categoria c atribuídos a c durante a avaliação.
- Documentos *de saída*. O número de documentos de formação na categoria c atribuídos a uma categoria diferente durante a avaliação.
- Documentos *de entrada*. O número de documentos de formação atribuídos à categoria c durante a avaliação mas que têm origem numa categoria diferente.
- Documentos *por atribuir*. O número de documentos de formação na categoria c que não foram atribuídos a qualquer categoria durante a avaliação. Este podem incluir documentos que foram atribuídos a uma categoria mas que ficaram abaixo do nível de fiabilidade indicado.

É apresentada uma descrição geral da progressão do processo de avaliação a nível do catálogo, que mostra:

- O estado da avaliação que indica se a avaliação ainda está em execução ou se já parou
- Última avaliação que faz a listagem da data da última avaliação
- Número de iterações de avaliação concluídas
- Número total de iterações
- Média da precisão geral, indicando a percentagem de documentos correctos, ou seja, aqueles originalmente atribuídos à categoria e os documentos de entrada
- Média de recuperação geral, indicando a percentagem de documentos da categoria que estão correctos
- Documentos correctos, ou seja, o número de documentos correctamente atribuídos
- Documentos mal colocados, ou seja, o número de documentos de entrada ou de saída
- Documentos por atribuir

A precisão e a recuperação têm uma relação próxima com o valor de fiabilidade atribuído. Se o valor de fiabilidade for baixo, a precisão cai e a recuperação aumenta, ou vice-versa. Um valor elevado de precisão significa que muitos

documentos de formação foram atribuídos correctamente; um valor elevado de recuperação, por outro lado, significa que a maior parte dos documentos de formação foram atribuídos a uma categoria; por outras palavras, não existem ou existem muito poucos documentos que não estejam atribuídos.

## Resultados da avaliação

Os resultados detalhados da avaliação são obtidos premindo o separador **Resultado da Avaliação**. Se seleccionar o catálogo e premir o separador, são apresentados os resultados para todo o catálogo; se seleccionar uma categoria, são apresentados os resultados para a categoria.

A Figura 12 apresenta o resultado da avaliação ao nível do catálogo:

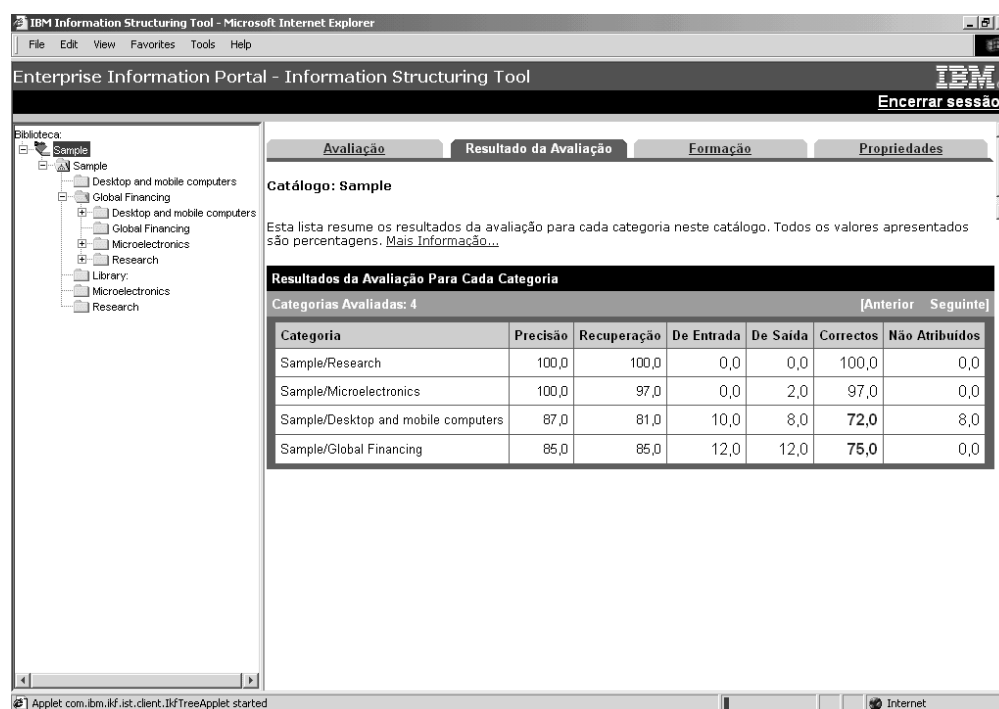


Figura 12. Resultados da avaliação a nível do catálogo

Iniciar pelos resultados globais do catálogo. Os valores a vermelho (crítico) e azul (menos crítico) indicam que algo está mal, ou com a categoria ou com os documentos de formação. A qualidade de uma categoria está definida pelos respectivos documentos de formação, por isso recomenda-se o estudo do movimento dos documentos para a (entrada) e da (saída) categoria.

As alterações efectuadas a uma categoria ou a documentos de formação estão fortemente dependentes da importância colocada quer na precisão e na recuperação da mesma forma, quer apenas na precisão. Quanto mais elevado for o valor de precisão, mais distinta é a categoria no que diz respeito a outras na taxonomia. Por outro lado, quanto mais elevado for o valor de recuperação, menos são os documentos de formação por atribuir.

Os resultados da avaliação são apresentados em dois níveis:

1. Ao nível do **catálogo**:

Para cada categoria no catálogo:

- As percentagens de precisão e de recuperação

- A percentagem de documentos de entrada
  - A percentagem de documentos de saída
  - A percentagem de documentos correctos
  - A percentagem de documentos por atribuir
2. Ao nível da **categoria**:
- Para tipos de documentos de entrada e de saída:
- Os documentos de formação e as categorias de origem ou de destino
- Para tipos de documentos correctos e por atribuir:
- Os documentos de formação

## Interpretar os resultados da avaliação

A secção seguinte sugere uma forma de interpretar os resultados da avaliação, mas tenha sempre presente que a taxonomia funciona como um todo, o que significa que as alterações efectuadas a uma secção da mesma podem ter um efeito adverso nos resultados produzidos noutra localização da taxonomia.

A Figura 13 mostra os resultados da avaliação ao nível da categoria:

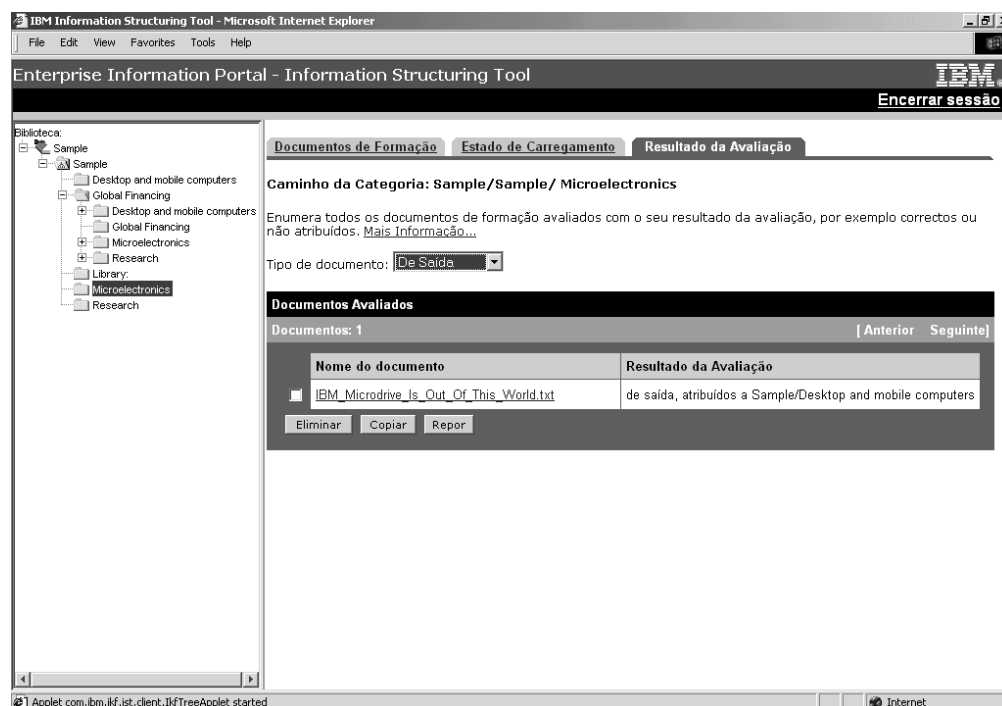


Figura 13. Resultados da avaliação a nível da categoria

Começando pelos resultados da avaliação ao nível do catálogo, selecione as categorias com valores de precisão e de recuperação potencialmente baixos e com valores de documentos de formação que tenham sido assinalados quer a vermelho (mais crítico) quer a azul (menos crítico). Para cada categoria, à vez, (possíveis de seleccionar na moldura à esquerda), verifique o seguinte:

- Muitos documentos de **entrada**:
  - A categoria ganha documentos de muitas outras categorias:
  - A categoria não é suficientemente distintiva. Copie estes documentos para esta categoria.

- Torne esta categoria mais distintiva seleccionando documentos de formação mais adequados ou divida a categoria em subcategorias.
- A categoria ganha documentos de uma ou duas categorias:
  - Copie estes documentos para esta categoria.
  - Verifique a categoria de origem. Faz sentido manter esta categoria ou devem ser intercaladas? Para intercalar, copie todos os documentos de treino para a categoria que pretende manter e elimine a outra categoria.
- Muitos documentos de **saída**:
  - A categoria perdeu documentos para muitas outras categorias:
    - A categoria não é suficientemente distintiva. Seleccionar documentos de formação mais adequados ou considere a opção de eliminar a categoria.
  - A categoria perdeu documentos para uma ou duas categorias:
    - Copie estes documentos para as outras categorias.
    - Verifique esta categoria. Faz sentido manter a categoria ou devem ser intercaladas?
- Muitos documentos **por atribuir**:
  - Compare os documentos por atribuir com os correctos em relação a:
    - Tamanho
 

Se os documentos por atribuir forem mais pequenos, por exemplo, considere a opção de juntar três para criar um novo documento mais extenso, para que corresponda ao tamanho dos documentos correctos.
    - Estilo
 

Se o estilo for diferente, elimine o documento.
    - Tópico
 

Se o tópico for ligeiramente diferente, mas de alguma forma ainda relacionado com o tópico dos documentos correctos, considere a opção de pesquisar documentos adicionais de formação que cubram este tópico e carregá-los na mesma categoria ou numa categoria criada de novo.

Se os valores de recuperação e precisão forem bons, e se os mesmos documentos continuarem por atribuir durante repetidas iterações de avaliação, elimine-os.

Só terá necessidade de iniciar novamente o processo de avaliação caso tenha efectuado alterações à taxonomia, por exemplo, se tiver eliminado uma categoria ou se tiver intercalado duas, ou se moveu muitos documentos de um lado para o outro. Podem ser efectuada pequenas alterações ao nível da categoria, como adicionar um novo documento de formação ou copiar documentos, para várias categorias, sem ter de voltar a iniciar o processo de avaliação. Tenha em atenção de que não existe um processo de desfazer a acção, caso pretenda inverter as suas alterações.

Pare a avaliação quando já não existirem valores vermelhos ou azuis que sejam do seu interesse ou logo que atinja um nível de precisão e de recuperação acima de 90%.

## Efectuar a formação a um catálogo

Após ter feito a avaliação da taxonomia e se estiver satisfeito com os resultados, é necessária uma formação da mesma utilizando para isso todos os documentos de formação para produzir um tipo especial de metadados, nomeadamente um

modelo de categorização, que pode ser utilizado, subsequentemente, para categorizar novos documentos através do serviço de Categorização do Information Mining.

Para iniciar a fase de formação, seleccione o catálogo que necessita de formação, prima o separador **Formação** e, de seguida, o botão **Iniciar Formação**.

O nome do catálogo seleccionado aparece no topo da moldura do lado direito. Pode ver também o estado de formação do catálogo. Os tipos de estados de formação são:

- Os documentos não podem ser categorizados. Efectue a formação do catálogo para que actualize o categorizador. (Esta é também a predefinição para um catálogo criado de novo. Neste caso, a data da **Última Formação** está vazia.) Ou o catálogo é novo e nunca foi sujeito a formação ou os últimos resultados da formação não são válidos, porque, por exemplo, foi alterado o nome ou foi eliminada uma categoria desde a última vez em que o catálogo foi sujeito a formação. Qualquer tentativa de usar o serviço de categorização para categorizar novos documentos resultará em erro.
- O conjunto de documentos de formação do catálogo foi alterado. Os documentos ainda são categorizados de acordo com os resultados da última formação. Efectue a formação do catálogo para que actualize a categorização.
- A formação do catálogo está em execução.
- O catálogo está actualizado, não é necessária qualquer formação.

Para parar o processo de formação, faça clique sobre o botão **Parar Formação**. Tenha em conta que, se a formação for interrompida, o processo de formação do catálogo tem que ser efectuado de novo.

Não pode carregar novos documentos de formação para o catálogo durante a formação do mesmo.

---

## Regulação do rendimento

Ao criar registos num catálogo, os índices de texto para os atributos pesquisáveis por texto, por exemplo IKF\_CONTENT, ficam muito grandes e o rendimento da pesquisa diminui de forma considerável. Para otimizar o armazenamento e aumentar o rendimento, precisa reorganizar os índices de texto periodicamente, em especial após grandes actualizações realizadas aos índices.

É preferível iniciar a reorganização de índices de texto em períodos de carga mínima a nível informático, por exemplo, à noite. Para executar uma organização de índices, mude para:

- Para Windows: ...\\ikf\\IkfReorg.cmd
- Para AIX: .../ikf/bin/IkfReorg.sh
- Para Solaris: .../ikf/bin/IkfReorg

Os parâmetros incluem: IkfReorg <IDUtilizador><Palavra-passe><NomeBD>

---

## Utilizar o IBM Web Crawler

Esta secção descreve e explica como configurar a função IBM Web Crawler. A função é instalada pelo programa de instalação do EIP caso seleccione o quadrado de verificação Funções.

O EIP Versão 8.2 contém um crawler da Web, um crawler do Lotus Notes, resumidores para extracção de dados de ficheiros sujeitos ao crawl, documentação baseada em HTML, exemplos de configuração e utilitários de suporte. O IBM Web Crawler (também referido como GCS) requer Java Versão 1.3 ou superior.

O IBM Web Crawler é um crawler e um miner (explorador) de conteúdos baseado em Java. Quando direccionado para o conteúdo, adquire e explora esse conteúdo.

O IBM Web Crawler pode efectua o crawl a conteúdos numa web de intranet, extranet ou Internet, em bases de dados do Lotus Notes de forma nativa ou através do Domino, em sistema de ficheiros locais. O IBM Web Crawler é construído para facilitar a adição de novos protocolos. O conteúdo pode ser de qualquer tipo, por exemplo, HTML, anexos de Notes e multimédia.

O IBM Web Crawler consegue explorar metadados e texto a partir de muitos tipos de conteúdos. Por exemplo, o conteúdo de HTML pode ser explorado para:

- URL
- Título
- Corpo
- Hora da última modificação
- Meta-identificadores, tais como, autor, palavras-chave, descrição e outros

Selecciona a partir de uma lista de extractores predefinidos para um determinado tipo de conteúdos. Os conteúdos e/ou metadados explorados são guardados no disco local. O IBM Web Crawler pode utilizar a tecnologia Network Solutions Outside In para extrair texto de mais de 200 tipos de conteúdo, uma parceria ideal para utilizar em aplicações de pesquisa. O IBM Web Crawler é concebido também para facilitar a adição de novos miners (exploradores).

O IBM Web Crawler está disponível para os sistemas operativos Windows NT 4.0 e Windows 2000. Pode instalar, configurar e utilizar o IBM Web Crawler em cerca de meia hora. Adquire e explora conteúdos numa média de cerca de dez ficheiros por segundo num PC a 500 MHz. É testado para efectuar o escalonamento de um 1 milhão de objectos (200000 Notas). Suporta vários utilizadores e diversas configurações de crawl/mine por utilizador e está habilitada para suportar o Idioma Nacional preferido pelo utilizador.

## Capacidades do IBM Web Crawler

O programa de instalação instala dois ficheiros:

**x:./<directório instalação>/run**

IBM Web Crawler para os ficheiros de comandos da Web e configurações exemplificativas.

**x:./<directório instalação>/notes-run**

Web Crawler para os ficheiros de comandos do Notes e configurações exemplificativas.

**x:./<directório instalação>/lib**

Ficheiros .jar, .zip e de filtragem do IBM Web Crawler.

## Configurar e executar o IBM Web Crawler para a Web

Esta secção descreve como configurar e executar o IBM Web Crawler para a Web. O IBM Web Crawler para a Web acede a servidores HTTP, FTP, de notícias ou de



ficheiros e cria resumos de documentos HTML e de outros objectos. Os resumos são ficheiros, um para cada documento ou objecto, que contém metadados e texto integral.

### Uma configuração básica

Esta secção contém instruções que explicam de que forma pode editar um ficheiro de configuração do IBM Web Crawler em formato XML. São fornecidas duas configurações exemplificativas para o ajudar a iniciar-se:

- Um ficheiro config-db2.xml para utilizar o IBM Web Crawler com o DB2 UDB.
- Um ficheiro config-sample.xml para utilizar o IBM Web Crawler sem o DB2 UDB.

1. Abra uma linha de comandos.
2. Mude o directório para o subdirectório de execução onde instalou o IBM Web Crawler. Por exemplo, se instalou o IBM Web Crawler num servidor Windows, escreva `cd x:<cmbrroot>\gcs\run`. Se instalou o IBM Web Crawler em AIX, escreva `cd /usr/lpp/cmb/gcs`.

**Sugestão:** É muito importante manter uma cópia do ficheiro original. Um erro no ficheiro pode interromper o IBM Web Crawler. Seja prudente ao editar.

3. Para executar o IBM Web Crawler com uma base de dados do DB2 UDB (mais escalável, mais lenta), edite o ficheiro config-db2.xml. Por exemplo, escreva `edit config-db2.xml` na linha de comandos.
4. Para executar o IBM Web Crawler sem uma base de dados do DB2 UDB (menos escalável, mais rápida), edite o ficheiro config-sample.xml. Por exemplo, escreva `config-sample.xml` na linha de comandos.

Para executar crawls de  $n$  URLs sem uma base de dados, necessita de aproximadamente  $n/1000$  MB de RAM no computador para conter os metadados dos URL sujeitos ao crawl. Por exemplo, para efectuar o crawl a 500.000 URLs, necessita de 512 MB de RAM. Para aproveitar da melhor maneira esta memória, edite o ficheiro crawlweb.bat e aumente o valor de JVMXmx.

### Configurar o Web Crawler para information mining

Para utilizar o Web Crawler com o Information Mining, ou seja, para poder aplicar funções de extracção de informações a documentos em que tenha sido efectuado crawl, são necessárias as seguintes definições de configuração específicas, que se distinguem das definições dos ficheiros de configuração exemplificativos já mencionados, o config-sample.xml e o config-db2.xml:

```
<globals ...
  max-urls="5000"
  temp-filepool-class="FullPathFilePool"
  summaries-dir="webspaces/ikf/disks/1/"
  summaries-filepool-class="DistributedEIPFilePool"
  ... >
...
</globals>
<group-list>
  <group ...>
    ...
    <summarizer-config>
      ...
      <resource-handler content-type="*"
        summarizable="EipHtmlSummarizable"
        summary-maker="EipHtmlRawSummaryMaker" />
      ...
    </summarizer-config>
  </group>
</group-list>
```

```

        </summarizer-config>
    ...
</group>
</group-list>

```

Estas definições fazem parte de um ficheiro de configuração exemplificativo para o componente do information mining, `im-crawler-config-sample.xml`, localizado no seguinte directório:

Em Windows:

```
<CMBROOT>\samples\java\beans\infomining\webcrawler\
```

Em UNIX (AIX e Solaris):

```
<CMBROOT>/samples/java/beans/infomining/webcrawler/
```

## Configurar a opção DB2 do IBM Web Crawler

Para configurar a opção DB2 deve criar uma nova base de dados. Isto requer autoridade de administrador de DB2. Pode ser necessário mudar para a conta do administrador do DB2. Pode atribuir um nome à base de dados conforme o DB2 o permitir, mas se o seu nome de base de dados for diferente de `gcs`, será necessário actualizar o `nomebd` no ficheiro de configuração do Web Crawler.

Se possuir autoridade de administrador de base de dados, pode executar este comando numa linha de comandos de DB2 para criar a base de dados:

```
db -createdb <utilizador><palavra-passe>[nome_basededados]
```

Se não especificar um nome de base de dados, é utilizado `gcs`. Uma vez criada a base de dados, adicione tabelas de IBM Web Crawler emitindo o comando:

```
db -createtables<utilizador><palavra-passe>[nome_basededados]
```

Para que seja possível utilizar o DB2 com o IBM Web Crawler deve concluir-se a criação de bases de dados e de tabelas do IBM Web Crawler.

As seguintes definições do ficheiro de configuração (na secção `urlpool-config`) são necessárias para poder utilizar a sua nova base de dados: `nomebd`:

- O nome da base de dados (como foi criado acima): Exemplo `gcs`.
- Nome de utilizador: O nome do utilizador, por exemplo: `db2admin`
- Palavra-passe: A palavra-passe do utilizador, por exemplo: `db2admin`.

Defina a base de dados, o nome de utilizador e a palavra-passe com os valores apropriados. Não modifique o tamanho da memória cache nem o controlador. Continue a editar o ficheiro de forma a definir o âmbito da pesquisa para o seu sistema.

## Definir o âmbito da pesquisa

Estas definições do ficheiro de configuração são necessárias para estabelecer o âmbito de pesquisa independentemente de estar a utilizar o DB2 ou não.

Verifique as seguintes definições na secção `crawler-config` e defina estas entradas de acordo com as suas necessidades.

### seed list

Um ou mais URLs absolutos de partida. O URL tem de estar disponível. Verifique utilizando o seu browser, por exemplo:  
<http://www.<omeusite>.com/>

**content-type-pattern-list**

Pesquisar o URL encontrado nas páginas apenas se alguma extensão de ficheiro corresponder a estes padrões, por exemplo: htm\*

**include-pattern-list**

Pesquisar o URL encontrado nas páginas apenas se estas corresponderem a estes padrões, por exemplo: <mysite>.com

Pode também definir estas entradas:

**recursion-depth**

A distância máxima em ligações pesquisadas a partir de qualquer ponto de partida. Utilize -1 para uma profundidade de pesquisa ilimitada.

**exclude-pattern-list**

Pesquisar o URL encontrado nas páginas apenas se estas não corresponderem a estes padrões, por exemplo: \*cgi-bin\*

**system properties (propriedades do sistema)**

Para efectuar buscas através de uma firewall a partir de um computador sem socks vai ser necessário definir os valores socksProxy neste ficheiro.

**Iniciar o IBM Web Crawler**

Se editar o ficheiro de configuração .xml, guarde-o.

Para iniciar o IBM Web Crawler, utilize o ficheiro de comandos crawlweb e um ficheiro de configuração. Abra uma linha de comandos e escreva:

- Para Windows: crawlweb.bat<CONFIGFILE>
- Para AIX: crawlweb.sh<CONFIGFILE>

Para executar com o DB2 UDB, escreva: crawlweb config-db2.xml e prima Enter. Para executar sem o DB2 UDB, escreva: crawlweb config-sample.xml e prima Enter.

**Sugestão:** Planifique a produção regular de relatórios de progresso de crawl/resumir. Logo que o crawl tiver sido concluído, são gravados resumos na localização configurada em summaries-dir. Os resumidores predefinidos gravam o objecto original e um prólogo de metadados como ficheiros .html numa árvore. Durante ou após o crawl, pode examinar o ficheiro-registo para obter informações adicionais.

**Configuração Avançada**

Neste momento, pode estudar as opções de configuração. Consulte o ficheiro config-sample2.xml em Capítulo 7, “Ficheiros exemplo do IBM Web Crawler”, na página 107 para obter uma exemplo de configuração. O exemplo ilustra a configuração de:

- Crawler e módulos de resumidor
- O monitor de gráficos
- Opções de registo
- SOCKS
- Crawling de Lotus Domino
- Tipos múltiplos de conteúdo
- Mais exclusões
- A utilização de InsoSummarizable para obter resumos de objectos tal como ficheiros .pdf

Consulte o ficheiro `config.dtd` para obter uma definição formal dos parâmetros disponíveis no ficheiro de configuração. **Recomendação:** Não edite este ficheiro. Faça uma cópia do ficheiro e atribua-lhe um outro nome.

## O ficheiro de configuração do IBM Web Crawler

O ficheiro de configuração é um ficheiro XML que indica ao IBM Web Crawler quais os recursos baseados na web a reunir e de que forma os resumir. Esta secção descreve cada elemento e atributo que pode definir em `config.xml`. Para informações sobre como utilizar o IBM Web Crawler for Notes, consulte “IBM Web Crawler for Notes” na página 90.

O IBM Web Crawler confere se os conteúdos do ficheiro de configuração estão em conformidade com `gcs-config.dtd`. Se existirem erros significativos, por exemplo, inexistência de URLs para efectuar o crawl, o IBM Web Crawler sai e imprime uma mensagem de erro. Para problemas menores (um atributo ou valor desconhecido), o programa irá registar um aviso no ficheiro de registos e continuar.

**Recomendação:** Faça uma cópia de segurança do seu ficheiro de configuração antes de editar. Um erro no ficheiro pode interromper o IBM Web Crawler.

São enviados ficheiros de configuração exemplificativos com o IBM Web Crawler.

### <gcs-config>

O ficheiro `gcs-config` contém duas secções: **globals** e **group-list**. Consulte o Capítulo 7, “Ficheiros exemplo do IBM Web Crawler”, na página 107 para obter um exemplo de um ficheiro `gcs-config`.

#### globals

O elemento `globals` (globais) captura definições para o IBM Web Crawler tais como sistema de ficheiros, rendimento e informações da rede.

#### group-list

O elemento `group-list` configura o crawl e o resumo de grupos, sendo que um grupo é um conjunto de recursos tal como um domínio de negócios ou de rede.

### <globals>

O elemento `globals` representa as definições globais do IBM Web Crawler. As definições estão codificadas como atributos globais e elementos dependentes.

A lista seguinte define os atributos globais. Para definições de elementos dependentes globais, consulte “<logger-config>” na página 80.

#### max-urls

O número máximo de URLs para efectuar o crawl. Deve ser um número inteiro positivo e a predefinição é de 100000.

#### summaries-dir

O directório no qual são gravados os resumos de recurso. Por predefinição é utilizado o directório `summaries/`.

#### summaries-filepool-class

O tipo de conjunto de ficheiros utilizado para resumos de recursos. Este facto determina de que forma os ficheiros de resumo são denominados e qual a estrutura de subdirectório (caso exista) utilizada. Por predefinição, é utilizado o `FullPathFilePool`, que cria um directório para o sistema central e utiliza a mesma estrutura de subdirectório e o mesmo nome de ficheiro como URL.

**num-crawlers**

O número de módulos de crawlers a utilizar. Deve ser um número inteiro positivo e a predefinição é de 20.

**num-summarizers**

O número de módulos de resumidor a utilizar. Deve ser um número inteiro positivo e a predefinição é de 5. Siga estes passos para configurar num-crawlers e num-resumidores:

1. Defina crawlers para a velocidade do seu computador em MHz/20. Por exemplo, num sistema a 600 MHz, utilize 30.
2. Defina o número de resumidores para 1/4 do número no passo 1, por exemplo, 8.
3. Faça uma execução-ensaio tendo em atenção o painel Task Manager Performance do Windows. Se a CPU *alguma vez* trabalhar mais do que um segundo a 100%, volte ao passo 1 e defina um número menor, por exemplo, 3/4 da última vez, até que a CPU nunca trabalhe a 100%.

Se, durante o ensaio, verificar que o texto-ecrã comunica consistentemente números de Summarizer: ToDo bastante abaixo do número de resumidores configurados, pode diminuir o número de resumidores (quantos menos, melhor) e aumentar o número de crawlers (quantos mais, melhor), desde que não esteja a ir contra o passo 3. Para um melhor desempenho, utilize a rede mais rápida que possa e distribua resumos, BD, espaço temp e registos em vários discos diferentes, sempre que possível.

**text-monitor**

Quando definido para ON, o ecrã-texto imprime o estado do IBM Web Crawler a cada cinco segundos para saída standard. Se lhe é indicado um valor decimal, o ecrã-texto define o tempo entre actualizações (em segundos) da saída de texto. A definição estabelecida é off (desligado).

**graph-monitor**

Quando definido para on, o ecrã-gráfico indica ao IBM Web Crawler para apresentar o estado utilizando uma GUI gráfica. Quando lhe é fornecido um valor decimal, o monitor-gráfico define o tempo entre actualizações (em segundos) do ecrã GUI. A definição estabelecida é off (desligado).

**log-file**

Especifica o ficheiro principal de registos a utilizar. A predefinição é log/log.txt.

**Sugestão:** Pode especificar informações adicionais de registador no elemento de config-registador.

**log-priority**

Define a predefinição da prioridade de registo. Introduza um valor de informações, aviso ou erro. O valor predefinido é aviso.

**Sugestão:** pode especificar informações adicionais de registador no elemento de config-registador.

**temp-dir**

O directório no qual se gravam os ficheiros temporários. **Sugestão:** Todos os ficheiros neste directório podem ser eliminados pelo IBM Web Crawler. Não deve ser necessário alterar a predefinição de x:/temp/gcs.

**temp-filepool-class**

O tipo de conjunto de ficheiros a utilizar para ficheiros temporários.

**Recomendação:** Não altere este facto na predefinição de TempFilePool.

**content-dir**

O directório no qual o IBM Web Crawler grava o ficheiro de conteúdos. Normalmente, content-dir é o mesmo que temp-dir.

**content-filepool-class**

O tipo de conjunto de ficheiros para utilizar para ficheiros de conteúdos. Normalmente este é o mesmo que temp-filepool-class.

**how-often-to-gc**

O número de URLs a efectuar o crawl entre pedidos de recolha de lixo.

**Recomendação:** Defina um número inteiro  $\geq 50$ . A definição predefinida é 100.

**max-resource-pool-size**

O tamanho máximo de fila de recursos aguardando resumo.

**Recomendação:** Defina um número inteiro  $\geq 10$ . A predefinição estabelecida permite 10 recursos a aguardar por resumidor.

**connect-timeout**

Define os milissegundos a aguardar antes de esgotar o tempo de uma ligação na rede. O valor predefinido é 4000. O âmbito válido é 1000-60000.

**read-timeout**

Define os milissegundos a aguardar antes de esgotar o tempo de uma leitura na rede. O valor predefinido é 6000. O âmbito válido é 1000-60000.

**cookies**

Define se verifica cookies no cabeçalho de HTTP e se os armazena numa base de dados. A definição estabelecida é off (desligado). Pode activar os cookies definindo o valor para on (ligado).

**locale** Define o idioma a usar para resumos e registos. O valor predefinido é en\_US.

Os elementos dependentes globais incluem logger-configs, uma urlpool-config e system-properties.

**<logger-config>**

O ficheiro logger-config fornece controlo avançado sobre o que é registado, de forma é formatado e onde é gravado o ficheiro de registo. Os default log-file e log-priority predefinidos são especificados como atributos globais (globals). Para obter mais informações sobre registos, consulte “Registar-se no IBM Web Crawler” na página 87.

**category**

A categoria do registador que está a ser configurado, por exemplo gcs.crawler. Caso não esteja especificado, é então configurado o registador predefinido. Tenha em conta que as definições de uma categoria em particular irão afectar todas as categorias dependentes.

**priority**

A prioridade mínima que uma mensagem deve possuir para ser registada. Se não for especificado, então este registador vai buscar a prioridade à sua categoria ascendente (e, em último caso, à log-priority predefinida de globais).

**log-file**

Define em que local gravar o ficheiro de registo. Se começar por um sinal '+', este ficheiro de registo será utilizado em adição a quaisquer outros

ficheiros de registo (ascendentes). Se não for especificado, será utilizado o ficheiro de registo ascendente (e, em último caso, o log-file predefinido de globais).

**Sugestão:** Certifique-se de que não especifica o mesmo log-file para registadores múltiplos, pois neste caso irão sobrepor-se.

#### **log-layout**

Define o esquema utilizado para cada mensagem impressa no ficheiro de registos.

#### **<urlpool-config>**

O ficheiro `urlpool-config` configura o componente do IBM Web Crawler onde estão armazenados os URLs. Existem várias opções para o conjunto de URLs. Pode armazenar o conjunto na memória, pode utilizar o DB2 ou pode ainda utilizar uma versão especial de uma memória reduzida que não armazena tanta informação sobre cada URL. Se não especificar um elemento de `config-conj-url`, o conjunto de URLs é armazenado na memória. O `urlpool-config` pode ter elementos de `urlpool-param` dependentes, para especificar coisas tal como informações da base de dados.

#### **urlcontainer-class**

O tipo de contentor de URLs a utilizar. Especifique:

- DB2URLContainer ao qual efectuar o crawl com o DB2 UDB
- MemoryURLContainer ao qual efectuar o crawl sem o DB2 UDB (predefinição).
- BigMemoryURLContainer ao qual efectuar o crawl sem o DB2 UDB e utilizar memória adicional (armazena alguns URLs de referência e outras informações).

#### **urlcollection-class**

O tipo de recolha de URLs a utilizar. Especifique:

- DB2URLCollection ao qual efectuar o crawl com o DB2 UDB
- MemoryURLCollection ao qual efectuar o crawl sem o DB2 UDB (predefinição).
- BigMemoryURLCollection ao qual efectuar o crawl sem o DB2 UDB e utilizar memória adicional (armazena alguns URLs de referência e outras informações).

#### **<urlpool-param>**

Utilizado para transmitir parâmetros para a `urlcollection-class`. Para um exemplo, consulte a informação da ligação da base de dados na configuração exemplificativa utilizando o DB2 UDB em Capítulo 7, “Ficheiros exemplo do IBM Web Crawler”, na página 107.

**name** Define o nome do parâmetro.

**value** Define o valor do parâmetro.

**Sugestão:** Seja prudente na utilização destes parâmetros, pois não existe uma verificação de erros para os mesmos.

#### **<system-properties>**

As propriedades do sistema representam uma lista de definições de propriedades do sistema.



### <property>

Para um exemplo, consulte a configuração para a utilização de uma porta de ligação SOCKS no exemplo de configuração avançada.

**name** O nome do parâmetro.

**value** O valor do parâmetro.

Alternativamente, pode configurar o acesso de IBM Web Crawler a servidores externos através de uma porta de ligação PROXY utilizando:

```
<system-properties>
  <nome propriedade="proxySet" valor="true"/>
  <nome propriedade="proxyHost" valor="proxy.nome sistema central"/>
  <nome porta proxy="proxyPort" valor="80"/>
</system-properties>
```

**Sugestão:** Não existe verificação de erros nestes parâmetros, por isso seja prudente ao utilizá-los.

### <group-list>

A group-list é uma lista de um ou mais elementos de grupo.

### <group>

O elemento de grupo representa um único grupo de recursos que irá ser sujeito a um crawl e resumido de forma idêntica. Cada grupo deve ter um atributo de nome único e pelo menos um elemento dependente de config-crawler que lhe indique sobre o que deve efectuar crawl. O Grupo pode ter um elemento de summarize-config dependente caso não pretenda utilizar os resumidores predefinidos. **Sugestão:** Os grupos que se sobrepõem (o mesmo URL está em dois ou mais grupos) podem levar a resultados não esperados. Os URLs em múltiplos grupos são associados apenas ao primeiro grupo onde são encontrados.

**Indique** um nome único para este grupo (necessário).

### <crawler-config>

Utilize estas regras para definir o âmbito do crawling. O crawler recupera cada URL na seed-list, analisa o URL a partir do conteúdo e adiciona à lista para-ser-efectuado-crawl aqueles URLs que correspondem:

- a pelo menos uma regra na content-type-pattern-list
- e a pelo menos uma regra na include-pattern-list
- e a nenhuma das regras na exclude-pattern-list.

O config-crawler também necessita de um único atributo: **profundidade-recorrência**. A profundidade-recorrência define a quantidade de ramificações de ligações de cada originador pelos quais o crawler pode viajar. A predefinição é -1, que representa profundidade infinita.

### <seed-list>

Esta é uma lista de originadores de URL, possivelmente com informações de autenticação.

### <seed>

Originador representa um originador de URL que iniciar o crawler, com um atributo de URL e, possivelmente, informações de autenticação. Cada originador deve ser um URL absoluto, por exemplo `http://<oseu.servidor>.com/`. É de evitar originadores que são redireccionados, que não se encontrem disponíveis ou que apontem para páginas que não sejam de texto. É útil que indiquem uma



página que editou por forma a conter os seus próprios originadores. Tal página torna-se fácil de actualizar, rever e testar com um browser.

**URL** Um URL originador por onde iniciar o crawl.

### **<authentication>**

Autenticação opcional enviada a um URL originador protegido por *Basic Authentication* tal como está definido em rfc2617.

#### **username**

O nome do utilizador utilizado para autenticação.

#### **password**

A palavra-passe utilizada para autenticação.

Por exemplo:

```
<url originador="http://oseu.servidor.com/"><nome do utilizador de autenticação="eu"
palavra-passe="minha"/></originador>
```

### **<content-type-pattern list>**

Esta é uma lista de padrões para incluir tipos de conteúdos aos quais será efectuado um crawl, tal como identificado pela extensão de ficheiro. Qualquer extensão de ficheiro de URL (.html, .gif, .doc, etc.) que corresponda a qualquer padrão-nome-url nesta lista passa este teste. Os URLs que não possuam uma extensão passam o teste por predefinição. Se a *content-type-pattern-list* (lista-padrão-tipo-conteúdo) não for especificada ou se se encontrar vazia, apenas os URLs sem uma extensão de ficheiro serão aceites.

### **<include-pattern list>**

Esta é uma lista de padrões para incluir URLs aos quais será efectuado um crawl, por exemplo, por servidor ou nome de domínio. Qualquer URL que corresponda a qualquer *url-obj-pattern*, *url-regex-pattern*, *url-name-pattern* ou *url-predicate-pattern* nesta lista, passa o teste. Se a *include-pattern-list* não for especificada ou se se encontrar vazia, todos os URLs serão aceites.

### **<exclude-pattern list>**

Esta é uma lista de padrões para excluir URLs da acção de crawl. Qualquer URL que corresponda a qualquer *url-obj-pattern*, *url-regex-pattern*, *url-name-pattern* ou *url-predicate-pattern* nesta lista não será sujeito a crawl. Se a *exclude-pattern-list* não for especificada ou se se encontrar vazia, não serão rejeitados quaisquer URLs.

### **<url-obj-pattern>**

Este é um padrão para fazer corresponder as diferentes partes de um URL (protocolo, sistema central, etc.) com caracteres globais. Pode ser utilizado na *exclude-pattern-list* (lista-padrão-excluir) e na *include-pattern-list* (lista-padrão-incluir). O padrão para cada parte pode ter um carácter global '\*' no início e/ou no fim, o que fará corresponder a tudo. No entanto, não podem existir caracteres globais no meio de qualquer padrão. A correspondência não é dependente de maiúsculas e minúsculas. É de imediato feita a correspondência a qualquer padrão de parte de URL que seja omitido.

A lista seguinte contém um exemplo de como Java e IBM Web Crawler interrompem o URL <http://www.ibm.com/products/index.html?query#ref>:

- O protocolo é: http
- O sistema central é: www.ibm.com
- A porta é: -1 (não especificada)

- O ficheiro é: /products/index.html?query
- O caminho é: /products/index.html
- O dir é: /products/
- O nome do ficheiro é: index.html
- A extensão é: .html
- A consulta é: query
- A ref é: ref

A lista seguinte fornece mais pormenores sobre cada elemento do url-obj-pattern:

#### **protocolo**

O padrão do carácter global ao qual deve corresponder um protocolo de URL, por exemplo, http.

#### **sistema central**

O padrão do carácter global que deve corresponder a um sistema central de URL, por exemplo, \*.ibm.com

**porta** O padrão do carácter global que deve corresponder a uma porta de URL, por exemplo, 80.

#### **ficheiro**

O padrão do carácter global que deve corresponder a um ficheiro de URL, por exemplo, \*.htm\*. A parte do ficheiro de um URL inicia-se com a primeira barra após o sistema central e pode incluir uma consulta mas não uma ref. A parte do *ficheiro* de  
http://www.ibm.com/products/index.html?query#ref é  
/products/index.html?query.

#### **caminho**

O padrão do carácter global ao qual deve corresponder um caminho de URL, por exemplo, \*.html. A parte do caminho de um URL inicia-se com a primeira barras após o sistema central e não inclui nem uma consulta nem uma ref. No exemplo, a parte do *caminho* de  
http://www.ibm.com/products/index.html?query#ref é  
/products/index.html

**dir** O padrão de carácter global que deve corresponder a um directório num URL, por exemplo, /products/. O directório é a parte do caminho que é iniciada com a primeira barra e que termina com a última barra. No exemplo, a parte do *dir* de  
http://www.ibm.com/products/index.html?query#ref é /products/. Não inclui uma consulta ou uma ref. Tenha em atenção que os URLs incorrectos que omitam a última barra, por exemplo, http://www.ibm.com/products, não irão corresponder correctamente ao dir. No exemplo do URL incorrecto, o dir do URL é /.

#### **nome do ficheiro**

O padrão do carácter global que deve corresponder a um nome de ficheiro num URL, no exemplo, index.html. O nome do ficheiro é a parte do caminho a seguir à última barra. No exemplo, o *nomeficheiro* de  
http://www.ibm.com/products/index.html?query#ref é index.html. Não inclui uma consulta ou uma ref.

#### **extensão**

O padrão do carácter global que deve corresponder a uma extensão de ficheiro URL, por exemplo, htm\*. É preferível utilizar a content-type-pattern-list onde for possível.

**consulta**

O padrão do carácter global ao qual deve corresponder uma consulta de URL.

**ref**

O padrão do carácter global ao qual deve corresponder uma ref de URL (não utilizado para HTTP). Por exemplo, `<url-obj-pattern host="*.ibm.com"/>` irá corresponder páginas de HTML em qualquer site da IBM.

**<url-regex-pattern>**

O `url-regex-pattern` é um padrão que faz correspondências de URLs utilizando uma expressão global. Pode ser utilizado na `exclude-pattern-list` ou na `include-pattern-list`. Utiliza o pacote `com.ibm.regex (regex4j)` e possui a maior parte da funcionalidade das expressões globais de Perl 5. Pode ter duas expressões globais, uma que o URL *tem* de corresponder e outra que o URL *não tem* de corresponder. Podem ser especificadas outras opções, tais como `i` para que não seja dependente do tipo de letra. Consulte `Regex4j Expressões Globais` para obter mais detalhes.

**corresponder**

Uma expressão global de estilo Perl 5 que um URL deve corresponder.

**no-match**

Uma expressão regular de estilo Perl 5 à qual um URL não deve corresponder.

**opções**

Modificadores opcionais, tal como `i` para não dependente do tipo de letra.

Por exemplo, `<url-regex-pattern`

`match="^http://www\.ibm\.com/.*\.html?$"/>` irá corresponder a páginas HTML no site principal da IBM.

**<url-name-pattern>**

Este é um padrão simples com caracteres globais para correspondência a um URL total ou a uma extensão de ficheiro de URL. Pode ser utilizado na `content-type-pattern-list` e na `include-pattern-list` e na `exclude-pattern-list`. Pode ter um `'*'` no início e/ou no fim da cadeia padrão, que corresponde a tudo. No entanto, não podem existir caracteres globais no meio do nome. A correspondência não é dependente de maiúsculas e minúsculas.

Por exemplo, `<url-name-pattern name="*.ibm.com/*"/>` iria corresponder a todos os ficheiros no site da IBM, mas `<url-name-pattern name="*.ibm.com/*.html"/>` não é válido pois tem um carácter global no meio.

**nome** Padrão de carácter global que uma cadeia de URL deve corresponder, com caracteres globais `'*'` opcionais no início e/ou final.

**<url-predicate-pattern>**

Este padrão carrega uma classe `UnaryPredicate` de Java para corresponder a um URL. Pode ser utilizado na `lista-padrão-excluir` ou na `lista-padrão-incluir`. A classe deve ter um método de execução booleana pública (`URL url`) que devolve verdadeiro se o URL corresponde ao predicado.

**classe** O nome completo de classe `UnaryPredicate`.

### <summarizer-config>

Esta é a configuração para um resumidor, com uma lista de operadores de recursos dependentes. Neste momento, pode apenas existir um summarizer-config por grupo.

### <resource-handler>

Determina qual o tipo de resumos efectuados para um recurso (tal como uma página da web ou um artigo de grupo de discussão), baseado no seu tipo de conteúdo, por exemplo (text/html) ou extensão de nome de ficheiro (htm). Quando um recurso se encontra pronto para ser resumido, o IBM Web Crawler verifica os operadores-recursos, por ordem, e utiliza o primeiro que corresponda ao tipo de conteúdo ou à extensão de ficheiro. Caso não exista nenhuma correspondência, serão utilizados os `Copy2RdfSummarizable` e `Copy2RdfSummaryMaker` por predefinição. Pode sobrepor esta definição adicionando um operador-recurso sem tipo de conteúdo ou extensão de ficheiro no final da lista.

Um operador-recursos também pode ter summarizer-param dependentes que transmitam parâmetros especiais à sua classe `SummaryMaker`.

#### content-type

Um padrão de carácter global a que o tipo-conteúdo de um recurso deve corresponder, por exemplo: `*htm*`

#### file-extension

Um padrão de carácter global a que a extensão de ficheiro de um recurso deve corresponder, por exemplo: `htm*`

#### summarizable

O nome de classe de recurso de Resumível, por exemplo:  
`HtmlRawSummarizable`

#### summary-maker

O nome de classe do `SummaryMake` (CriadorResumos) de recursos, por exemplo: `HtmlRawSummaryMaker`

Os padrões de tipo-conteúdo e de extensão-ficheiro permitem caracteres globais. Um padrão pode ter um `*` no início e/ou final da cadeia-padrão, que corresponde a tudo. No entanto, não podem existir caracteres globais no meio. A correspondência não é sensível a maiúsculas e minúsculas.

Um operador-recursos irá corresponder se o tipo-conteúdo e a extensão-ficheiro corresponderem, e um padrão não especificado irá sempre corresponder, por isso

```
<resource-handler content-type="*htm*"
summarizable="*HtmlRawSummarizable" summary-maker=
"HtmlRawSummaryMaker"/>
```

corresponderá a todos os ficheiros com o tipo de conteúdo `text/html`, independentemente da extensão de ficheiro.

Para `summarizable` e `summary-maker`, não necessita especificar o caminho completo das classes se as classes se encontrarem no pacote `com.ibm.ibm Web Crawler.summarizer.resource`.

### <summarizer-param>

Estes são parâmetros especiais que são transmitidos a uma classe `SummaryMaker`. A utilização é específica para aquela classe.

**name** O nome do parâmetro.

**value** O valor do parâmetro.

**Sugestão:** Não pode existir verificação de erros nestes parâmetros, por isso seja prudente ao utilizá-los.

## Registrar-se no IBM Web Crawler

Esta é uma introdução às características de registo do IBM Web Crawler.

O IBM Web Crawler fornece grande capacidade de controlo sobre o que fica registado, onde fica registado e de que forma é formatado. Por exemplo, pode optar por gravar o código de resposta de cada página em que foi efectuado o crawl para um ficheiro, o estado do IBM Web Crawler (quantos URLs foram sujeitos a crawl, quantos os módulos a trabalhar, etc.) para outro ficheiro, os URLs resumidos para um terceiro ficheiro, todos os avisos do IBM Web Crawler para um quarto ficheiro e todas as mensagens de registo no pacote de utilitários da rede para outro ficheiro para depuração.

Consulte “IBM Web Crawler exemplo de ficheiro de análise de registo” na página 109 para um exemplo de um ficheiro de análise de registos.

### Utilidade dos registos

O registo é útil para a contabilização de rede/web/crawl/resumo, a comunicação com outros componentes da aplicação e para a depuração de IBM Web Crawler.

A contabilização de crawl e de exploração pode dar a conhecer um leque alargado de funções interessantes, por exemplo servidores erradamente configurados, páginas em falta e número de objectos por tipo-conteúdo. O script de Perl `loganalysis.pl` fornece um exemplo de contabilização de resumos de registos. As aplicações podem necessitar de informações do IBM Web Crawler, tal como quando foi removido o conteúdo.

### Configurar os registadores

Pode especificar a configuração de um ou mais registadores no ficheiro de config do IBM Web Crawler. Os atributos de prioridade-registo e ficheiro-registo dos elementos globais estabelecem a política predefinida de registos.

Para expandir a política de registos, há que criar instruções de config-registador como dependentes do elemento de globais. Cada instrução selecciona um subconjunto de mensagens de registo de IBM Web Crawler, encaminha-os para um ficheiro específico e grava-os utilizando um formato específico. O subconjunto de mensagens registadas é seleccionado utilizando atributos de prioridade e de categoria. Os valores legais de prioridade são: rastreio, depuração, info e aviso (não dependente de maiúsculas e minúsculas).

- Definir o valor de prioridade determina como o registador é verboso, sendo que rastreio é o mais verboso.
- rastreio e depuração são níveis de manutenção - as mensagens são em código estático em Inglês.
- info e aviso são níveis de utilizador, com suporte para idiomas nacionais.
- info produz muitas mensagens. Reduza a saída de mensagens especificando um prioridade de aviso.

### Exemplos de configuração de registo

Registrar de/para hiperligações, sem informações de data/hora/módulo, para o ficheiro `log/fromto.txt`

```
<logger-config category="gcs.url.fromto" priority="info"
log-layout="%m\n" log-file="log/fromto.txt"/>
```

#### **Registar objectos resumidos para o ficheiro log/resources.txt**

```
logger-config category="gcs.summaries.list.resource"
priority="info" log-file="log/resources.txt"/
```

#### **Registar o URL ignorado e a razão para este facto**

```
<logger-config category="gcs.url.skipped"
priority="info" log-file="log/urls_skipped.txt"/>
```

#### **Registar códigos de resposta processados de HTTP**

```
<logger-config category="gcs.http.302"
priority="info" log-file="log/urls_redirected.txt"/>
<logger-config category="gcs.http.404"
priority="info" log-file="log/urls_not_found.txt"/>
```

#### **Registar todas as mensagens na categoria de resumidor, incluindo a sua prioridade**

```
<logger-config category="gcs.summarizer"
priority="TRACE" log-file="summarizer_trace.txt"
log-layout="%d: %t: %c: %p: %m\n"/>
```

## **Detecção e Correção de Problemas**

Caso se depare com problemas, tem que verificar primeiro:

#### **As páginas da lista originadora são alcançáveis?**

As páginas devem existir (evite redireccionar originadores) e devem ser alcançáveis a partir do seu sistema e através de SOCKS, caso esteja a ser utilizado.

#### **As páginas na lista originadora são HTML normais?**

Estruturas, Flash, javascripts e outros artigos idênticos não representam uma boa escolha de originadores. Escolha páginas normais de HTML.

#### **Se estiver a utilizar o DB2 UDB, já efectuou o crawl?**

O DB2 UDB mantém o registo do que foi sujeito a crawl. Se todas as páginas foram processadas, o DB2 UDB fica em silêncio. Utilize o comando `db -emptytables` para iniciar um novo crawl.

#### **Se estiver a utilizar o DB2 UDB, as entradas de acesso da base de dados no ficheiro de configuração estão correctas?**

Se a ligação da base de dados falhar, o crawl também falha.

#### **Verificou cuidadosamente as edições ao seu ficheiro de configuração?**

Os erros podem interromper o IBM Web Crawler. Não se apercebeu de um `max-urls` restritivo ou um valor de profundidade-recorrência?

#### **Mantém-se o problema?**

Edite o ficheiro de configuração e altere a prioridade do registo para "depuração". Volte a iniciar o Web Crawler e, logo que pare, examine o ficheiro de registos.

## **Escolher Resumidores**

O objectivo de um resumidor é tomar um recurso (tal como uma página da web) ou um sistema central (tal como um servidor de web) e produzir um ficheiro que contém a informação que pretende num formato fácil de utilizar.

O IBM Web Crawler inclui uma variedade de resumidores que funcionam com diferentes tipos de conteúdos, extraem tipos diferentes de dados a partir do

recurso e fazem a saída em diferentes formatos de ficheiros. Esta secção descreve as funções e os requisitos dos resumidores disponíveis. Se nenhum destes tem as funções que pretende, pode também gravar o seu próprio resumidor.

Existem dois tipos de resumidores no IBM Web Crawler. Um resumidor de recursos produz um resumo de um único recurso, tal como uma página da web e um resumidor do sistema central produz um resumo de um sistema central, tal como um servidor web. Por agora, apenas são configuráveis os resumidores de recursos.

### **Escolher um resumidor de recursos**

O que há a considerar ao escolher ou gravar um resumidor de recursos:

- Qual é o formato da saída? Página da web HTML, PDF, documento em WordPro, ficheiro XML,
- Quais os metadados que pretendem extrair? Cabeçalho de HTTP, título, ligações anotadas, corpo do texto
- Qual o formato de saída que pretende? XML, HTML, RDF

Especifica-se qual o resumidor a utilizar para um tipo específico de recurso utilizando o elemento do operador de recurso no ficheiro config do IBM Web Crawler. Primeiro especifica-se o tipo de conteúdo e/ou extensão de nome de ficheiro para o qual o resumidor é utilizado. De seguida, especificam-se as classes Java Summarizable e SummaryMaker que fazem estas acções. A classe resumível representa o recurso a ser resumido e a classe de summary-maker representa o tipo de resumo que será efectuado.

### **Resumidor predefinido (Cópia + Resumidor RDF)**

A cópia + resumidor RDF é o resumidor que é aplicado a qualquer objecto com um tipo de conteúdo não configurado explicitamente para processamento por outro resumidor. Este resumidor pode ser utilizado com qualquer tipo de recurso, e grava dois ficheiros. O primeiro ficheiro é uma cópia exacta do recurso original e o segundo ficheiro é um resumo RDF que contém o URL original, o nome do ficheiro e o ficheiro armazenado e informações de cabeçalho de HTTP. Pode também ser configurado explicitamente utilizando o DefaultSummarizable e Copy2RdfSummaryMaker.

### **Resumidores para páginas HTML (resumidor de caracteres de HTML)**

Para recursos de HTML, o Resumidor de caracteres de HTML produz simplesmente uma cópia do ficheiro, com o URL e as informações de cabeçalho de HTTP num comentário no cimo. É configurado utilizando o HtmlSummarizable e HtmlRawSummaryMaker.

```
<resource-handler content-type="*htm*"
    summarizable="HtmlSummarizable"
    summary-maker="HtmlRawSummaryMaker" />
```

### **Resumidores para páginas HTML (resumidor de HTML do EIP)**

Para recursos de HTML, o Resumidor de caracteres de HTML produz simplesmente uma cópia do ficheiro, com o URL e as informações de cabeçalho de HTTP num comentário no cimo. É configurado utilizando o EIPHtmlSummarizable e o EIPHtmlRawSummaryMaker.

```
<resource-handler content-type="*htm*"
    summarizable="EIPHtmlSummarizable"
    summary-maker="EIPHtmlRawSummaryMaker" />
```



### Resumidores para páginas HTML (resumidor HTML sem escrita)

Este resumidor vai efectuar o crawl de HTML e ligações de seguimento, mas não vai gravar qualquer resumo do ficheiro. Este facto pode ser útil se, por exemplo, pretender efectuar crawl a todos os ficheiros PDF num site (utilizando o INSO para XML Summarizer) mas não pretender armazenar os ficheiros HTML. É configurado utilizando o HtmlSummarizable e o NoWriteSummaryMaker.

```
<resource-handler content-type="*htm*"
                    summarizable="InsoSummarizable"
                    summary-maker="InsoSummaryMaker" />
```

### Resumidores para outros tipos de conteúdo (INSO para resumidor de XML)

Este resumidor vai criar um resumo XML para mais de 200 tipo de recursos, tais como documentos em Microsoft Word, ficheiros PDF, apresentações em PowerPoint e outros. Contém alguma meta-informação e o corpo do texto é extraído pelos filtros da Network Solutions INSO (é necessária uma licença da INSO). É configurado utilizando o InsoSummarizable e o InsoSummaryMaker.

```
<resource-handler content-type="pdf"
                    summarizable="InsoSummarizable"
                    summary-maker="InsoSummaryMaker" />
```

### Outros resumidores

Se necessitar de resumir outros tipos de recursos, explorar outros dados ou efectuar a saída de dados noutros formatos, contacte a IBM ou crie um resumidor personalizado.

## IBM Web Crawler for Notes

Esta secção descreve como configurar e executar o IBM Web Crawler for Notes. O IBM Web Crawler for Notes acede a bases de dados de Notes e cria resumos de documentos e anexos de Notes. Os resumos são ficheiros de formato XML, um por documento ou anexo, contendo objecto e texto completo.

### Pré-requisito

São necessários os seguintes pré-requisitos antes de executar o IBM Web Crawler for Notes:

- Lotus Notes Versão 5.0.5 ou posterior.
- PKZIP Versão 2.50, caso queira lidar com ficheiros anexados que são ficheiros zip que se auto-extraem.

### Executar um teste de crawl

Selecione **Iniciar**→**Programas**→**Linha de Comandos**. Na nova janela, modifique para o directório no qual instalou o IBM Web Crawler for Notes e, de seguida, mude para o subdirectório notes-run. Por exemplo:

```
cd c:\directório de instalação>\gcs\notes-run
```

Um crawl de Notes é controlado por dois ficheiros:

- Uma lista de fontes que edita para identificar bases de dados de Notes nas quais pode efectuar o crawl. Inclui nomes de servidores de Notes, endereços de IP, nomes de ficheiros de .nsf, etc. Por exemplo, uma lista de fontes pode indicar 34 bases de dados de Notes; é estabelecido numa configuração descrita abaixo quais os que são sujeitos ao crawl. A lista de fontes pode ser um ficheiro .xml ou uma base de dados de Notes (um ficheiro .nsf).
- Um ficheiro de configuração que especifica uma lista de fontes, a quais das fontes é efectuado o crawl, quais os tipos de anexos a processar, formatos de saída, etc. O ficheiro de configuração é sempre um ficheiro .xml.



Para verificar se o IBM Web Crawler está instalado correctamente, efectue o crawl à base de dados de teste. Utilizando um editor, certifique-se de que a lista de fontes `testSources.xml` tem o caminho e o nome de ficheiro correctos para a base de dados `test.nsf`, que se encontra no subdirectório `notes-run` onde instalou o IBM Web Crawler. Faça uma cópia de salvaguarda dos ficheiros originais.

**Recomendação:** *Edite com prudência* : um erro no ficheiro provoca uma falha no IBM Web Crawler. Guarde quaisquer alterações.

Teste a instalação efectuando o crawl à base de dados `test.nsf` incluída.

Introduza: `crawlNotes crawlTestXml`

O ficheiro `crawlNotes.bat` inicia o IBM Web Crawler com o `crawlTestXml` como ficheiro de configuração; `.xml` é anexado automaticamente ao nome do ficheiro de configuração. O IBM Web Crawler deve participar o crawl e resumir dois documentos, cada um com um anexo.

Quando o IBM Web Crawler terminar, pode ver os resumos no directório de resumos e os ficheiros de registo do crawl no directório de resumos especificado no ficheiro de configuração.

### Configurar um crawl de Notes personalizado

Após um crawl de teste bem sucedido, vai querer efectuar o crawl a outras bases de dados.

1. Crie a lista de fontes de base de dados. Adicione bases de dados de Notes nas quais vai ser efectuado o crawl a um ficheiro de fontes.

Para identificar bases de dados de Notes às quais vai ser efectuado o crawl num ficheiro XML, comece por editar o ficheiro `testSources.xml`. Para identificar bases de dados de Notes às quais vai ser efectuado um crawl numa base de dados de Notes, utilize Notes para abrir e actualizar a base de dados de Notes `testSources.nsf`. Os parâmetros que pode definir nos ficheiros fonte estão explicados nas Listas de Editar Fonte.

2. Definir a configuração do crawler. Necessita editar um ficheiro de configuração em formato XML.
  - Se as suas fontes estão listadas num ficheiro XML, comece por editar `crawlTestXml.xml` e defina `sourcesInXmlFile` para direccionar para o seu ficheiro de fontes.
  - Se as suas fontes se encontram numa base de dados de Notes, comece por editar `crawlTestNsf.xml` e defina `sourcesInNotesDB` para direccionar para a sua base de dados de fontes. Os parâmetros que pode definir nos ficheiros de configuração encontram-se explicados em “O ficheiro de configuração do IBM Web Crawler” na página 78.

Uma vez que tenha completado a sua lista de fontes e de configuração, invoque o IBM Web Crawler: `crawlNotes your_config`

Ou, se a sua lista de fontes está numa base de dados de Notes, inicie o IBM Web Crawler tal como no seguinte exemplo de trabalho: `crawlNotes crawlTestNsf`

Quando o Notes de IBM Web Crawler terminar, pode ver os resumos no directório de resumos e os ficheiros de registo do crawl no directório de resumos especificado no ficheiro de configuração.

## Parâmetros da lista de fontes

Uma lista de fontes contém descrições de bases de dados de Notes que podem ser sujeitas a um crawl. As listas de fontes no formato `file.xml` contém um elemento de `notesDataSources` com um ou mais elementos de `oneDBInfo`. Cada elemento de `oneDBInfo` contém:

**id** Um id numérico para esta base de dados. É referido pelo parâmetro de âmbito no ficheiro de configuração.

### **Nomeservidor**

O nome do servidor que serve a base de dados. Utilize a cadeia nula "" para as bases de dados locais.

### **pathAndFileName**

O caminho completo e nome de ficheiro da base de dados no servidor. Termine o caminho e o nome de ficheiro com `.nsf`.

### **viewName**

O nome da visualização de Notes da base de dados sobre a qual vai ser efectuado o crawl.

### **ipAddress**

Opcional. O endereço de IP do servidor; caso seja fornecido, o DNS não é utilizado. Se o DNS não consegue processar o nome do servidor de destino, pode especificar o seu endereço de IP neste local. Em Windows, o endereço de IP pode ser determinado utilizando o comando `nslookup server_name`.

### **dateLastCrawled**

Opcional. A data da última vez em que foi efectuado um crawl à base de dados. Será automaticamente modificado a não ser que defina actualizar Data do Último Crawl para não no ficheiro de configuração.

### **tentativas**

Opcional. O número de vezes que pretende tentar efectuar de novo o crawl à base de dados, caso o crawl não seja bem sucedido (tempo de espera esgotado).

### **fieldSubstitutions**

Direccionamentos que especificam de que forma os nomes dos campos da base de dados de Notes são substituídos no documento XML de saída. Contém um ou mais elementos de substituição, cada um com dois atributos:

- **Original:** o nome do campo que será substituído no documentoXML de saída, se existir
- **Substituir:** Pelo novo nome do campo que irá substituir o nome do campo original no documento XML de saída.

Listas de fontes numa base de dados de Notes podem ser examinadas e actualizadas utilizando o cliente Notes. Inicie Notes e seleccione **File →Database →Open**. Clique sobre **Browse** para localizar e abrir a base de dados `testSources.nsf` no directório `x:\<install directory>/gcs/notes-run`. O teste da base de dados enviado no pacote pode também ser examinado e actualizado utilizando o cliente Notes. Inicie Notes e seleccione **File →Database →Open**. Clique em **Browse** para localizar e abrir a base de dados `test.nsf` no directório `x:\<install directory>/gcs/notes-run`.

## Parâmetros de configuração do ficheiro

Os parâmetros que podem ser definidos num ficheiro de configuração estão descritos abaixo. Pode omitir parâmetros com as predefinições listadas.

Um ficheiro `sourcesInXml` ou um elemento `sourcesInNotesDB` identifica a lista de fontes. Fontes é um ficheiro de formato XML ou uma base de dados de Notes, respectivamente, que contém informações que identificam a(s) base(s) de dados à(s) quais irá ser efectuado o crawl.

Um elemento de `runInfo` contendo parâmetros que controlam uma única execução. Isto é, estes parâmetros aplicam-se a todas as bases de dados que foram sujeitas a crawl na utilização dada do crawler:

#### **rangeSpecify**

Os ids das bases de dados a efectuar o crawl. Os ids são os números fornecidos no campo do id da sua lista de fontes. Pode ser especificado como uma lista separada por vírgulas de ids individuais e/ou âmbitos indicados por hífenes, como em 1-4, 15, 25-31.

#### **SummaryDirectory**

Especifica o directório-raiz para resumos de saída. Os resumos são escritos em subdirectórios deste directório.

#### **MaxThreads**

Especifica o número de módulos paralelos de crawl. Cada base de dados de Notes é sujeita ao crawl por um único módulo. Bases de dados múltiplas serão sujeitas ao crawl em paralelo.

#### **doIncrementalCrawl**

A predefinição é não. Se for sim, apenas processa documentos de Notes novos/modificados desde '`summarizeThisDateAndLater`'. Se '`summarizeThisDateAndLater`' não for especificado, o crawler vai utilizar cada campo de `DateLastCrawled` da base de dados especificado na lista de fontes da base de dados. Se `doIncrementalCrawl=no`, todos os documentos vão ser processados, independentemente da data.

#### **summarizeThisDateAndLater**

O formato para este campo é: MM/dd/aaaa hh:mm a tz, por exemplo 01/01/2000 01:11 PM PDT. Caso não seja fornecida nenhuma data ou hora, resumem-se todos os documentos desde o último crawl gravado na lista de fontes (se `doIncrementalCrawl` estiver definido para sim), ou desde sempre (se `doIncrementalCrawl` estiver definido para não.)

#### **detachAttachments**

A predefinição é sim. Se for este o caso, separa e resume anexos. Os tipos de ficheiros de anexos para processamento estão listados no ficheiro de configuração. Se não, os anexos são ignorados.

#### **attachmentFilenameFormat**

A predefinição é l (longo). Pode também ser s (short-curto). O nome de ficheiro longo codifica o tipo, o servidor, o nome da base de dados e o id de Notes. O nome de ficheiro curto codifica o tipo e o id de Notes.

#### **processAttachmentsAfterwards**

A predefinição é não. Se sim, os ficheiros anexados não são resumidos durante o resumo da base de dados de Notes. Em vez disso, é escrito um registo no `notesCrawl-attachments.bat` para cada anexo, especificando comandos para o resumo dos ficheiros. O utilizador, de seguida, escreve e executa um ficheiro de comandos que resuma e depois elimine os anexos. O processamento adiado dos anexos requer normalmente espaço de armazenamento substancial de disco.

#### **saveAttachmentFiles**

A predefinição é não. Se sim, os originais dos ficheiros anexados não são

apagados após o processamento. Esta opção é válida apenas se processAttachmentsAfterwards estiver definido como não. Se processAttachmentsAfterwards estiver definido como sim, é necessário espaço em disco para armazenar os anexos gravados.

**MaximumNumberOfDetachingErrors**

A predefinição é 10. O número máximo de erros ao processar o anexo, por exemplo, ficar sem espaço em disco ao gravar ficheiros anexado, tolerados pelo crawler antes de abortar o crawl.

**saveURLsToFile**

Predefinição é não. Se especificado, os URLs encontrados nos artigos de documento Notas são escritos para num ficheiro do nome na forma: nomebasedados(sem caminho e .nsf) + ".html".

**updateDateLastCrawled**

A predefinição é sim. Se for definido não, não actualiza a dataÚltimoCrawl no ficheiro de fontes.

**tempDirectory**

A predefinição é c:\temp . Este directório é utilizado para escrever quaisquer ficheiros temporários.

**logSummaryDirectory**

A predefinição é registar. O directório no qual se gravam os ficheiros de registo.

**loggerPriority**

A predefinição é informação. Estas definições definem a prioridade do registador. As definições podem ser, da mais elevada à mais baixa, erro, aviso ou informação. Por exemplo, se a prioridade do registador for definida como aviso, apenas as mensagens de registo com a prioridade de aviso e erro serão registadas.

Um elemento de anexação contendo elementos de inclusão que identificam extensões de ficheiros anexados a serem processados, por exemplo .prz.

## Excluir o IBM Web Crawler de um servidor

Por razões de segurança e rendimento, um administrador do EIP pode optar por excluir determinados servidores ou páginas da sequência de hiperligações. O utilizador pode ter a necessidade de limitar as actividades dos crawlers nos respectivos servidores e páginas.

Pode dar instruções ao IBM Web Crawler para evitar servidores ou páginas utilizando um ficheiro de *política de acesso*. O ficheiro é construído de acordo com as directivas publicadas em *A Standard for Robot Exclusion* (consulte <http://info.webcrawler.com/mak/projects/robots/norobots.html>).

- O IBM Web Crawler necessita do ficheiro de política de acesso `http://oseuservidor/robots.txt` antes de efectuar o crawl a um servidor e periodicamente a partir dessa data.
- O ficheiro é constituído por linhas no formato  
`campo:<espaçoopcional>valor<espaçoopcional>`

Se *campo* for Utilizador-Agente e o *valor* for IBM-WebCrawler ou \*, as seguintes linhas Desactivar (através da seguinte linha Utilizador-Agente) especificam endereços parciais a evitar. Este pode ser um caminho completo ou um caminho parcial, qualquer endereço que se inicie com este valor não é obtido.

Por exemplo:

Desactivar: /ajuda

desactiva /ajuda.html e /ajuda/índice.html.

Desactivar: /ajuda/

desactiva /ajuda/índice.html mas activa /ajuda.html.

Um valor em branco permite a obtenção de todos os endereços.

- Estas linhas podem ser separadas por linhas em branco.
- Pode incluir comentários escrevendo um carácter #. O resto da linha é considerado um comentário.

Aqui estão alguns exemplos:

- Este /robots.txt especifica que todos os robots devem evitar este servidor.

# não permitir qualquer pessoa

Agente-utilizador: \*

Desactivar: /

- Este /robots.txt especifica que apenas o IBM Web Crawler pode efectuar um crawl a este servidor e que o crawler não tem limites.

# permitir apenas IBM

Agente-utilizador: \*

Desactivar: /

Agente-utilizador: IBM-WebCrawler

Desactivar: # não permitir nada

- Este /robots.txt especifica que todos os robots devem evitar os endereços nas árvores de documentos html modelo, desenvolvimento e teste, que um agente utilizador denominado IBM-WebCrawler tem uma excepção para as árvores de desenvolvimento e de teste (é permitido aqui) e que os robots xyz e wxyz devem ser totalmente excluídos.

# um exemplo mais realista

Agente-utilizador: \*

Desactivar: /dochtml/modelo

Desactivar: /dochtml/desenvolvimento

Desactivar: /dochtml/teste

Agente-utilizador: xyz

Agente-utilizador: wxyz

Desactivar: /

Agente-utilizador: IBM-WebCrawler

Desactivar: /dochtml/modelo



---

## Capítulo 6. Introdução ao fluxo de trabalho

É possível utilizar o fluxo de trabalho do EIP para controlar o fluxo e o rendimento do trabalho na empresa. Quando os utilizadores trabalham com os resultados de pesquisas federadas, devem frequentemente tomar decisões relativamente às acções a executar. Pode utilizar o fluxo de trabalho do EIP para determinar antecipadamente como pretende que os utilizadores executem o trabalho.

Pode automatizar os fluxos de trabalho definindo perfis e regras que controlam o modo como os componentes do fluxo de trabalho trabalham em conjunto. Pode também optar por escolher o grau de restrição do seu sistema através do controlo do acesso e da autoridade de utilizadores através de conjuntos de privilégios e listas de controlo de acesso.

---

### Compreender o fluxo de trabalho

A maioria das operações empresariais podem ser caracterizadas como um conjunto de processos inter-relacionados. O trabalho passa de um funcionário para outro e de um departamento para outro. Alguns processos simples poderão necessitar de poucos passos, enquanto processos mais complexos envolvem um certo número de funcionários em departamentos diferentes.

O fluxo de trabalho permite mover trabalhos através de um processo e tomar decisões relacionadas com trabalho durante todo o processo. Por exemplo, a XYZ Seguros recebe grandes volumes de formulários de participações por correio. Durante o processo de verificação, os avaliadores de pedidos de indemnização devem reunir documentos, como, por exemplo, fotografias e avaliações. Os funcionários gastam várias horas por dia para abrir, ordenar, arquivar e supervisionar informações, bem como a recolher documentos pertinentes para aprovação final.

Estas informações são trocadas entre funcionários à medida que as informações são recebidas e verificadas. À medida que a participação vai seguindo o seu trajecto, é tratada por funcionários de vários departamentos.

---

### Como utilizar o fluxo de trabalho

Como no exemplo da XYZ Seguros, a maioria das empresas que lida com documentos efectua parte ou a totalidade das tarefas seguintes:

- Arquivar documentos para uma consulta posterior.
- Recolher documentos, formulários, relatórios e informações junto de diversas fontes e, em seguida, entregar estes documentos no local adequado para serem processados.
- Fazer corresponder correio recebido com documentos que estão a ser processados actualmente.

Um *fluxo de trabalho* representa o processo do trabalho. Descreve as acções que podem ser executadas num grupo de um ou mais documentos ou conteúdos e o caminho deste grupo de documentos através do processo de fluxo de trabalho. Um fluxo de trabalho reflecte o modo como é executado, com um âmbito e limites bem definidos. Define a sequência de actividades e tarefas e as ligações e relações entre estas actividades e tarefas. Um fluxo de trabalho determina os critérios utilizados

para tomar decisões sobre o fluxo de trabalho. Para informações acerca do processo de criação de fluxo de trabalho, consulte o *Manual de Programação de Aplicações de Estações de Trabalho*. Para informações acerca de como utilizar um cliente com fluxo de trabalho, consulte *Instalação, Configuração e Gestão do eClient*

---

## Sincronizar IDs de utilizador e grupos

Esta secção explica como sincronizar IDs de utilizador e grupos entre o EIP, o Content Manager e o MQ Series Workflow.

Quando gere utilizadores no Content Manager ou no Enterprise Information Portal e o MQSeries Workflow é parte do seu sistema, também gere utilizadores no MQSeries Workflow. Assim, se criar, modificar ou eliminar um ID de utilizador no Content Manager ou no Enterprise Information Portal, deve também fazê-lo no servidor MQSeries Workflow.

Uma vez que o Content Manager e o Enterprise Information Portal partilham IDs de utilizador e grupos, se criar um ID de utilizador ou um grupo no cliente de administração do sistema e o serviço de fluxo de trabalho estiver activado mas o servidor MQSeries Workflow não estiver activo, recebe uma mensagem de erro indicando que esse utilizador ou grupo não foi criado no servidor MQSeries Workflow. O ID de utilizador continua a existir como utilizador no Content Manager ou no Enterprise Information Portal mas não existe no servidor MQSeries Workflow.

Para sincronizar os IDs de utilizador e os grupos de gestão de conteúdos com os do MQSeries Workflow, é necessário executar o utilitário de sincronização de utilizadores do Workflow. Se instalou o servidor MQSeries Workflow manualmente, certifique-se de que o servidor MQSeries Workflow foi iniciado. Abra Serviços e verifique o estado do MQSeries Workflow X.X - FMC (sendo X.X a versão de fluxo de trabalho que tem instalada. Se o MQSeries Workflow não estiver activo, inicie o serviço ou, caso tenha efectuado a instalação em modo não assistido, vá para o directório WFInstall e execute o ficheiro de comandos CMBWFStart.bat. Siga os passos indicados abaixo para executar o utilitário de sincronização:

1. Altere o directório onde foi instalado o Enterprise Information Portal. O directório predefinido é C:\CMBROOT.
2. Escreva EIPUser2WF.bat.

**Importante:** Se possuir um servidor de fluxo de trabalho remoto, tem de iniciar os ficheiros de comandos ou de interface (cmbsvregist81.bat ou cmbsvregist81.sh) do servidor RMI antes de escrever EIPUser2WF.bat. Para além disso, EIPUser2WF.bat utiliza informações de dois ficheiros INI, cmbsvcs.ini e cmbsvclient.ini, para localizar o servidor local ou remoto. Certifique-se de que cmbsvcs.ini indica LOCAL se o seu servidor for local ou REMOTE se o seu servidor for remoto. O ficheiro cmbsvclient.ini contém a localização do servidor.

3. Introduza as informações necessárias para o nome da base de dados do EIP, o ID de utilizador, a palavra-passe e o esquema.

Depois de introduzidas as informações necessárias, o utilitário de sincronização copia os utilizadores do servidor de Content Manager or Enterprise Information Portal para o servidor do MQSeries Workflow. Quando o utilitário tiver terminado, não deverá receber mensagens de erro indicando que um dado utilizador ou grupo não é reconhecido pelo servidor do MQSeries Workflow.



Se eliminar um ID de utilizador ou um grupo no Content Manager ou no Enterprise Information Portal, deve também fazê-lo no servidor do MQSeries Workflow. Se um ID de utilizador ou um grupo não existir no servidor do MQSeries Workflow, não poderá eliminá-lo no Content Manager ou no Enterprise Information Portal. Por exemplo, cria um ID de utilizador ou um grupo no Content Manager ou no Enterprise Information Portal com a opção de servidor de fluxo de trabalho do EIP desactivada. Em seguida, activa a opção de servidor de fluxo de trabalho quando tenta eliminar o ID de utilizador ou um grupo. Uma vez que o ID de utilizador ou o grupo não existe no fluxo de trabalho do MQSeries, vai receber uma mensagem de erro indicando que o ID de utilizador não existe no servidor do MQSeries Workflow. Para corrigir este erro tem de executar o utilitário EIPUser2WF para sincronizar os IDs de utilizador e os grupos e, em seguida, eliminar o ID de utilizador e o grupo.

## Reinstalar o servidor do EIP com o Workflow activado

Se tiver o EIP Workflow activado no servidor do EIP e pretende reinstalar um novo servidor de EIP, tem de remover todos os dados de fluxo de trabalho do EIP do servidor do MQSeries Workflow.

Antes de remover o servidor do EIP, execute os seguintes passos de forma a remover todos os dados de fluxo de trabalho do EIP do servidor do MQSeries Workflow. **Importante:** execute os passos na ordem indicada abaixo.

1. Termine todas as instâncias de fluxo de trabalho utilizando o eClient ou as APIs de Workflow. Ao terminar as instâncias de fluxo de trabalho são eliminados todos os artigos de trabalho.
2. Elimine as listas de trabalho e os modelos de fluxo de trabalho utilizando o cliente de administração do sistema do EIP ou as APIs de administração do sistema.
3. Elimine todos os utilizadores e grupos de utilizadores do EIP quando o fluxo de trabalho estiver activado, de forma a que todos os IDs de utilizadores e grupos de utilizadores sejam também eliminados do servidor do MQSeries Workflow.
4. Reconfigure o servidor do MQSeries Workflow para EIP:
  - a. Insira o CD-ROM de instalação do EIP. Utilize uma linha de comandos para passar para o directório WfInstall.
  - b. Na linha de comandos, escreva

```
fmcibie -iCMBWAdmin.fdl -uadmin -ppassword -o -f
```

**Importante:** Se remover a base de dados do EIP antes de eliminar os IDs de utilizador e grupos do EIP e do MQSeries Workflow, e em seguida tentar criar os mesmos IDs de utilizador e grupos que existiam na base de dados do EIP que eliminou, vai receber uma mensagem de erro indicando que os utilizadores e os grupos não podem ser adicionados ao EIP. A mensagem de erro é: DGL2616A: Não foi possível adicionar o utilizador: XXX -DGL2485A: Este utilizador de fluxo de trabalho já existe. Para resolver o problema execute os seguintes passos:

1. Desactive o serviço de fluxo de trabalho do EIP no cliente de administração do sistema do EIP.
2. Termine a sessão do cliente de administração do sistema do EIP e volte a iniciar sessão. Quando iniciar sessão depois de desactivar o serviço de fluxo de trabalho, pode criar os mesmos IDs de utilizador e grupos que se encontravam no EIP e ainda existem no servidor MQSeries Workflow.

3. Active o serviço de fluxo de trabalho do EIP depois de criar os IDs de utilizador e grupos que já existem no servidor MQSeries Workflow.

## Actualizar IDs de utilizador e grupos entre o MQSeries Workflow e a base de dados do EIP

Para cada ID de utilizador ou grupo que exista no servidor MQSeries Workflow e não exista na base de dados do EIP, pode executar os seguintes passos para sincronizar os IDs de utilizador e os grupos:

1. Crie um ficheiro e introduza o seguinte texto:

- CODEPAGE 1252
- FM\_RELEASE V3R3 2
- DELETE PERSON 'Utilizador1'
- DELETE PERSON 'Utilizador2'
- DELETE PERSON 'Utilizador3'

em que Utilizador1, Utilizador2 e Utilizador3 são os utilizadores que pretende eliminar. Pode listar quantos utilizadores necessitar.

2. Guarde o ficheiro e execute o seguinte comando numa linha de comandos:

```
fmcibie -u admin -ppassword -i DeletePersons.fdl -f -o
```

em que DeletePersons.fdl é o nome do ficheiro que foi criado no passo anterior.

Para cada grupo que exista no servidor MQSeries Workflow e que já não exista na base de dados do EIP, pode executar os seguintes passos:

1. Crie um ficheiro de texto e introduza as seguintes informações:

- CODEPAGE 1252
- FM\_RELEASE V3R3 2
- DELETE ROLE 'Grupo1'
- DELETE ROLE 'Grupo2'
- DELETE ROLE 'Grupo3'

em que Grupo1, Grupo2 e Grupo são os grupos que pretende eliminar. Pode listar quantos grupos sejam necessários.

2. Guarde o ficheiro e introduza o seguinte comando numa linha de comandos:

```
fmcibie -u admin -p password -i DeleteGroups.fdl -f -o
```

onde DeleteGroups.fdl é o nome do ficheiro que foi criado no passo anterior.

Quando elimina um ID de utilizador do EIP e o suporte de fluxo de trabalho se encontra activado, pode receber a seguinte mensagem de erro:

Não foi possível eliminar o utilizador [RC=12]

Se a operação de eliminação do utilizador falhar, o sistema gera temp.log, que fornece informações detalhadas sobre a falha. O ficheiro temp.log é gravado em x:\CMBROOT. **Sugestão:** Se tentar eliminar o ID de utilizador que usou para iniciar sessão no cliente de fluxo de trabalho do MQSeries, a operação de eliminação do utilizador falhará.

---

## Planear um fluxo de trabalho

Antes de começar a definir um fluxo de trabalho, deve determinar o trabalho que a sua empresa executa, além do local e do modo de execução e quem o vai executar. Um administrador ou um analista empresarial efectua este passo do planeamento.

Qual é o produto final? O produto final poderá ser o resultado de todo o trabalho realizado pela sua empresa, por um departamento na sua empresa ou por determinados funcionários de diferentes departamentos. Por exemplo, o produto final do processo de compensação de indemnizações da XYZ Seguros é a carta enviada ao detentor da apólice aprovando ou rejeitando a participação.

Analise as informações que devem ser processadas para produzir o produto final, determine as acções que devem ser executadas e onde devem ser executadas e decida o modo como pretende que as informações passem pelo fluxo de trabalho.

### Informações a processar

Considere quais as informações que devem ser tratadas pelos utilizadores na sua empresa. Que tipos de entrada de dados suportam o produto final? Que documentos específicos devem ser processados?

Um *artigo de trabalho* não é um documento real. Um *artigo de trabalho* contém uma referência para um documento e outras informações sobre o documento, tais como, o estado do documento, a data da sua criação e outras. Um *artigo de trabalho* pode ser qualquer conteúdo (documentos ou objectos) de um servidor de conteúdos. Por exemplo, a XYZ Seguros recebe inicialmente formulários de apólices e mais tarde recebe documentos de seguimento, tais como fotografias, apreciações e relatórios de peritos.

### Tratamento de informações

Quem está mais habilitado para tratar cada passo do processo? Por exemplo, um assistente administrativo pode verificar se um formulário de participação está completo, em seguida, arquivar o mesmo até que o titular da apólice envie um determinado documento. Quando o documento é recebido, o avaliador de pedidos poderá ser responsável por fazer corresponder o documento com o pedido de indemnização, bem como pela aprovação do documento.

Uma *lista de trabalhos* pode ser considerada como uma fila de trabalhos criados para utilização por um ou mais funcionários. É possível agrupar os pedidos de indemnização numa *lista de trabalhos* acessível a um certo número de avaliadores de pedidos. A lista de trabalhos constitui uma vista filtrada dos artigos de trabalho. Os funcionários só vêem os artigos numa lista de trabalhos para a qual tenham autorização.

As listas de trabalho podem ser definidas para filtrar artigos de trabalho de uma forma que trata cada parte do processo de participação, tal como recolher fotografias, avaliações e relatórios. Uma lista de trabalho pode também consistir em trabalhos de diferentes fluxos de trabalhos. Por exemplo, lista de trabalhos para um avalista de pedidos pode conter avaliações para um pedido, fotografias relacionadas com um segundo pedido e um relatório de perito para um terceiro pedido. As acções do avalista para cada artigo nesta lista de trabalhos pode ser diferente. O avalista pode rever a avaliação e aprovar o primeiro pedido. Pode preferir esperar por mais informações sobre o segundo pedido antes de executar qualquer acção relacionada com as fotografias. Para o terceiro pedido, pode enviar o relatório do perito para que outro funcionário execute uma acção.

## Acções a efectuar

Tenha em conta quais as acções a efectuar no conteúdo de um artigo de trabalho durante o fluxo de trabalho. Por exemplo, o avaliador de pedidos pode aceitar um pedido de indemnização ou rejeitá-lo se estiver incompleto. Uma *lista de acções* define as acções que um utilizador pode executar no trabalho.

Por exemplo, dependendo do facto de um pedido satisfazer ou não os requisitos iniciais, um avaliador de pedidos pode seleccionar uma opção para que o pedido continue no fluxo de trabalho ou outra opção para rejeitar o pedido.

## Como as informações fluem através do processo

Pense no modo como pretende que as informações e actividades fluam. Por exemplo, quando será revisto o pedido de indemnização inicial? Que documentos de apoio são necessários para avançar para o passo seguinte no processo? Que critérios determinam se um pedido é aceite ou rejeitado? Este fluxo de informações é a base do fluxo de trabalho.

Um fluxo de trabalho consiste nos caminhos que conduzem o trabalho ao longo do processo. Qual a origem da entrada de dados? O fluxo de trabalho deve começar num determinado ponto. Na XYZ Seguros, o pedido de indemnização enviado pelo titular da apólice é o documento que inicia o fluxo de trabalho

Quando todos os documentos tiverem sido recebidos, o artigo de trabalho pode continuar no caminho até chegar a uma acção final—por exemplo, a aprovação do pedido.

## Como tudo se encaixa

Após analisar as informações que pretende processar, determinar as acções que pretende executar e a decidir como pretende que seja efectuado o fluxo das informações através do fluxo de trabalho, está apto a criar um diagrama do fluxo de trabalho, que é a representação gráfica do fluxo de trabalho. Utilize a função construtor de fluxo de trabalho do EIP para criar o diagrama.

Um diagrama de fluxo de trabalho mostra como o trabalho se move nas diferentes actividades do processo, anotando as tarefas que a actividade envolve. Descreve o fluxo, os elementos principais e os pontos-chave de um fluxo de trabalho.

Cada símbolo no diagrama do fluxo de trabalho representa um ponto em que o trabalho é efectuado. Uma participação de sinistro tem de ser revista, tem de ser reunida documentação de suporte e a participação tem de ser aprovada ou rejeitada, dependendo de determinados critérios. Consulte “Criar um fluxo de trabalho” na página 105 para obter mais informações sobre os símbolos do processo utilizados no construtor de fluxo de trabalho.

---

## Utilizar componentes de fluxo de trabalho do Enterprise Information Portal

Esta secção descreve os componentes do fluxo de trabalho. Accede a todos os componentes através do cliente de administração. **Dica:** O fluxo de trabalho EIP Versão 8 inclui várias alterações, incluindo mudanças ao contentor da Versão 7.1 para acomodar a nova arquitectura Content Manager Versão 8.




## Utilizar o construtor de fluxo de trabalho

Pode utilizar o construtor de fluxo de trabalho para definir e construir graficamente o fluxo de trabalho de um grupo de trabalho, departamento ou empresa. **Restrição:** O processo de migração EIP transfere utilizadores de bases de dados Versão 7.1. O EIP Versão 8.2 não fornece qualquer transferência automática de dados de fluxo de trabalho. Deve voltar a desenhar os diagramas do fluxo de trabalho da Versão 7.1 utilizando o construtor de fluxo de trabalho do EIP Versão 8.2 e voltar a implementar os processos de fluxo de trabalho do EIP Versão 7.1.

Antes de utilizar o construtor de fluxo de trabalho para criar um modelo dos fluxos de trabalho, tem de definir os conjuntos de privilégios, listas de controlos de acesso, utilizadores, grupos de utilizadores, acções, listas de acções e listas de trabalhos. Ao definir um fluxo de trabalho no cliente de administração, pode definir também uma lista de acções predefinida para todo o fluxo de trabalho. Pode também atribuir uma lista de acções diferente em cada nó no fluxo de trabalho. Para obter mais informações sobre estas tarefas, consulte “Definir listas de acções” na página 104, “Definir listas de trabalhos” na página 104 e a ajuda on-line.

Embora utilize o construtor de fluxo de trabalho para construir fluxos de trabalho, não pode utilizar o construtor de fluxo de trabalho para executar o fluxo de trabalho. Utilizando um cliente, os utilizadores vêem e executam acções em listas de trabalhos e artigos de trabalhos. Para obter mais informações sobre como programar um cliente para trabalhar com fluxo de trabalho de Enterprise Information Portal, consulte o *Manual de Programação de Aplicações de Estações de Trabalho* e a online API reference. A Tabela 7 descreve três dos ícones de fluxo de trabalho que são comuns a todos os fluxos de trabalho. Existem outros ícones de fluxo de trabalho, como sub-fluxo de trabalho, evento, recolha e saída de utilizador. A ajuda on-line do EIP descreve os ícones da barra de ferramentas em detalhe.

Tabela 7. Ícones básicos de fluxo de trabalho

| Ícone   | Descrição  |
|---|--|
|  | O nó de início começa o processo de fluxo de trabalho. O diagrama do processo de fluxo de trabalho tem de ter um e apenas um nó de início.   |
|  | O nó de término termina o processo de fluxo de trabalho. Cada novo diagrama de processo de fluxo de trabalho contém um nó de término. Quando cria um processo, será gerado um nó de término. Pode mover o nó de término para qualquer posição na superfície do desenho. O diagrama de processo de fluxo de trabalho tem de ter um e apenas um nó de término.                                 |
|  | Um nó de trabalho associa uma lista de trabalho e uma lista de acções a um ponto específico do processo de fluxo de trabalho. Um nó de trabalho representa um ponto do processo de fluxo de trabalho onde é realizado trabalho. Todos os nós, incluindo o nó inicial e o nó final, necessitam de um conjunto de acções e alguém a quem seja atribuída a realização desse conjunto de acções. |

## Utilizar os serviços de fluxo de trabalho

O Enterprise Information Portal fornece serviços de fluxo de trabalho que mantêm informações sobre fluxos de trabalho. As definições de fluxo de trabalho e de lista

de acções que o utilizador cria com o construtor de fluxo de trabalho são mantidas na base de dados de administração do Enterprise Information Portal e na base de dados do fluxo de trabalho do IBM MQSeries.

Quando o administrador do sistema cria uma lista de trabalho, as informações associadas à mesma são armazenadas permanentemente na base de dados da administração. O administrador do sistema pode actualizar, eliminar e adicionar listas de trabalho utilizando EIP. Quando um administrador de sistema dá saída de um fluxo de trabalho, o fluxo de trabalho é fechado na base de dados do Enterprise Information Portal e é marcado na base de dados como estando a ser utilizado pelo utilizador, evitando que qualquer outra pessoa actualize até o utilizador tenha terminado.

---

## Definir listas de trabalhos

Uma lista de trabalho pode ser vista como um filtro de trabalho disponível. Uma lista de trabalhos é uma lista filtrada de artigos atribuídos a utilizadores ou grupos de utilizadores específicos. Quando os utilizadores iniciam sessão no Enterprise Information Portal podem ver listas filtradas de artigos de trabalho que lhes são atribuídas. Utilize o cliente de administração do Enterprise Information Portal para definir as listas de trabalhos.

A definição de uma lista de trabalhos inclui as regras que governam a apresentação, o estado e a segurança do respectivo artigo de trabalho. O utilizador especifica as regras para cada lista de trabalhos ao mesmo tempo que cria a lista de trabalhos. Para gerir o acesso a uma lista de trabalhos, o utilizador cria uma lista de controlos de acesso para a lista de trabalhos. Para obter uma descrição completa de como definir listas de trabalhos, consulte a ajuda on-line. A definição da lista de trabalhos inclui:

### **Lista de controlo de acesso**

Uma lista de controlos de acesso consiste num ou mais IDs de utilizadores individuais ou grupos de utilizadores e no conjunto de privilégios associados. O conjunto de privilégios é utilizado para definir uma autorização do utilizador para aceder ou realizar determinadas tarefas no trabalho. As listas de controlo de acesso são utilizadas para limitar o acesso do utilizador a artigos numa lista de trabalhos.

### **Filtrar e ordenar listas de trabalhos**

Critérios de visualização de uma lista de trabalhos filtrada e ordenada.

### **Número máximo de artigos numa lista de trabalho.**

Número máximo de artigos que pretende que uma lista de trabalho contenha.

---

## Definir listas de acções

Uma lista de acções é uma lista abrangente de todas as acções que um utilizador pode executar no trabalho num fluxo de trabalho.

A ajuda on-line do cliente de administração fornece instruções passo-a-passo que explicam como definir acções e listas de acções.



---

## Criar um fluxo de trabalho

Depois de definir acções, listas de acções e listas de trabalhos, utilize o construtor de fluxo de trabalho para criar o modelo do fluxo de trabalho. A ajuda online do cliente de administração fornece instruções passo-a-passo que explicam como definir acções e listas de acções. O construtor de fluxo de trabalho fornece sugestões visuais para criar um fluxo de trabalho.

---

## Activar construtor de fluxo de trabalho

Neste passo, está a iniciar um fluxo de trabalho numa base de dados de administração. **Restrição:** A base de dados que selecciona para fluxo de trabalho deve estar no mesmo servidor onde instalou o MQ Series, e serviços MQSeries devem ser iniciados.

Para activar um fluxo de trabalho EIP e criar uma definição de fluxo de trabalho:

1. Inicie sessão no cliente de administração.
2. Se tiver várias bases de dados de administração, faça clique no ícone para a base de dados onde pretende activar o fluxo de trabalho.
3. Clique em Ferramentas->Serviços. Clique em Activar Fluxo de Trabalho.
4. Termine a sessão do cliente e inicie novamente. Se tiver várias bases de dados, seleccione o ícone para a base de dados onde activou o fluxo de trabalho EIP. O ícone da pasta Fluxos de Trabalho será apresentado.
5. Na área esquerda da janela principal da administração do Enterprise Information Portal, faça duplo clique na pasta Fluxos de Trabalho **Fluxos de Trabalho**.
6. Faça clique com o botão direito do rato no ícone Definições de Fluxo de Trabalho e seleccione **Novo** para criar uma definição de fluxo de trabalho.

**Requisito:** é necessário criar pelo menos uma lista de controlos de acesso, uma acção e uma lista de acções antes de definir um fluxo de trabalho.

---

## Iniciar o servidor MQSeries Workflow

Iniciar servidor de Fluxo de Trabalho MQSeries introduzindo cmbwfstart numa linha de comandos. São abertas duas janelas para o servidor do MQSeries Workflow. Deixe estas janelas de comandos abertas para continuar a executar o servidor.

Se instalar o fluxo de trabalho depois da instalação inicial do Enterprise Information Portal, deverá configurar o sistema do Enterprise Information Portal para a função do fluxo de trabalho. Também é necessário alterar a configuração se instalar a função fluxo de trabalho numa estação de trabalho diferente da estação de trabalho em que o programa do cliente de administração foi instalado.

1. A partir da janela de administração, clique sobre as **Ferramentas** do membro do ficheiro.
2. Faça clique sobre **Serviços** no menu.
3. Seleccione a caixa de verificação **Fluxo de Trabalho**.
4. Depois de concluída a configuração, termine sessão do cliente de administração do Enterprise Information Portal e, em seguida, inicie sessão de novo para inicializar a função fluxo de trabalho. Após ter iniciado sessão no cliente de administração Enterprise Information Portal, o ícone das **Definições de Fluxo de Trabalho** aparece no painel da esquerda.

**Sugestão:** Os administradores não visualizam o ícone **Definições do Fluxo de Trabalho** a não ser que tenham autoridade para administrar a função fluxo de trabalho. Se pretende restringir o acesso à função fluxo de trabalho, consulte os manuais de administração de sistema apropriados para cada servidor de conteúdo. Consulte a ajuda on-line para obter mais informações sobre como autorizar administradores a gerir a função fluxo de trabalho.

O cliente pode ser criado a partir de uma aplicação personalizada utilizando o EIP Connector Toolkit e exemplos ou pode utilizar o cliente exemplo do EIP.



---

## Capítulo 7. Ficheiros exemplo do IBM Web Crawler

Esta secção fornece dois exemplos de código. Os ficheiros exemplificativos `config-sample2.xml` ilustram os parâmetros de configuração `<gcs-config>`. O exemplo de análise do registo ilustra um exemplo de um relatório que contém informações de um crawl concluído.

---

### exemplo de config-sample2.xml

O exemplo de código nesta secção ilustra o ficheiro `gcs-config`.

```
<!DOCTYPE gcs-config SYSTEM "config.dtd">
<gcs-config>
  <!-- Definições globais: -->
  <globals max-urls="1000000"
    num-crawlers="30"
    num-summarizers="8"
    summaries-dir="summaries"
    log-file="log/LOG.txt"
    temp-dir="temp"
    log-priority="warn"
    text-monitor="60"
    graph-monitor="2"
    connect-timeout="120"
    read-timeout="100">

    <!-- definições do registador exemplificativas -->
    <logger-config category="gcs.summaries.list.resource"
priority="info" log-file="log/resources.txt"/>
    <logger-config category="gcs.summaries.list.host" priority="info"
log-file="log/hosts.txt"/>
    <logger-config category="gcs.url.skipped" priority="info"
log-file="log/skipped_urls.txt"/>
    <logger-config category="gcs.url.fromto" priority="info"
log-layout="%m\n" log-file="log/fromto.txt"/>
    <logger-config category="gcs.http" priority="info"
log-file="log/http.txt"/>
    <logger-config category="gcs.http.connect" priority="info"
log-file="log/connecterrs.txt"/>

    <!--utilize este para especificar uma base de dados
    <urlpool-config urlcontainer-class="DB2URLContainer"
urlcollection- class="DB2URLCollection">
      <urlpool-param name="dbname" value="gcs"/>
      <nome parâm-conjurl="user" valor="xxxxxx"/>
      <urlpool-param name="password" value="xxxxxx"/>
      <urlpool-param name="cachesize" value="1000"/>
      <urlpool-param name="driver"
value="COM.ibm.db2.jdbc.app.DB2Driver"/>
    </urlpool-config> -->

    <!--utilize este para especificar um proxy SOCKS
    <system-properties>
      <property name="socksProxySet" value="true"/>
      <property name="socksProxyHost" value="socks2.server.ibm.com"/>
      <property name="socksProxyPort" value="1080"/>
    </system-properties> -->

  </globals>

  <group-list>
    <group name="ibm">
```

```

<crawler-config recursion-depth="-1">
  <seed-list>
    <!-- URLs to start crawling at: -->
    <seed url="http://gcs.stl.ibm.com/gcs/testurl.html"/>
    <seed url="http://gcs.stl.ibm.com/gcs/stl.html"/>
    <seed url="http://gcs.stl.ibm.com/gcs/ibm.html"/>
  </seed-list>

  <content-type-pattern-list>
    <!-- extensões de ficheiros URL que não correspondam a estes padrões
    não serão sujeitos a crawl: -->
    <url-name-pattern name="htm*"/>
    <url-name-pattern name="pdf"/>
    <url-name-pattern name="gif"/>
    <url-name-pattern name="zip"/>
    <url-name-pattern name="txt"/>
  </content-type-pattern-list>

  <include-pattern-list>
    <!-- URLs que não correspondam a estes padrões não serão sujeitos a crawl: -->
    <url-obj-pattern host="*.ibm.com"/>
    <!-- sbo - url-obj-pattern query="*OpenDocument*" / -->
    <!-- sbo - url-obj-pattern query="*OpenView*" / -->
    <!-- url-obj-pattern query="*OpenDocument =>
OpenDocument&ExpandAll*" / -->
    <!-- url-obj-pattern query="*OpenView =>
OpenView&ExpandAll&Count=999999*" / -->
  </include-pattern-list>

  <exclude-pattern-list>
    <!-- URLs que correspondam a estes padrões não serão sujeitos
    a crawl: -->
    <!-- ignorar estes padrões comuns na nossa intranet -->
    <url-obj-pattern file="*news*"/>
    <url-obj-pattern file="*search*"/>
    <url-obj-pattern file="*/afs*/>
    <url-obj-pattern file="*/.../*"/>
    <url-obj-pattern file="*bluepages*"/>
    <!-- ignorar páginas pessoais -->
    <url-obj-pattern file="*/~*"/>
    <!-- ignorar SOCKS: nenhum URL deve especificar isto directamente -->
    <url-regex-pattern match=".*:1080/*"/>
    <!-- ignorar portas de ligação: recomendado para meros mortais -->
    <url-regex-pattern match=".*[\\?\\=\\+\\;\\%&quot;&amp;].*"/>
    <!-- senão efectua o crawl a portas de ligação tal como se encontra
    aqui configurado... -->
    <!-- skip Domino -->
    <url-obj-pattern file="*.nsf*"/>
    <!-- senão efectua crawl a Domino: permite apenas OpenDocument -->
    <url-obj-pattern query="*OpenServer*"/>
    <url-obj-pattern query="*OpenDatabase*"/>
    <url-obj-pattern query="*OpenElement*"/>
    <url-obj-pattern query="*OpenView*"/>
    <url-obj-pattern query="*OpenAbout*"/>
    <url-obj-pattern query="*OpenHelp*"/>
    <url-obj-pattern query="*OpenIcon*"/>
    <url-obj-pattern query="*OpenForm*"/>
    <url-obj-pattern query="*OpenNavigator*"/>
    <url-obj-pattern query="*OpenAgent*"/>
    <url-obj-pattern query="*CreateDocument*"/>
    <url-obj-pattern query="*DeleteDocument*"/>
    <url-obj-pattern query="*EditDocument*"/>
    <url-obj-pattern query="*SaveDocument*"/>
    <url-obj-pattern query="*SearchSite*"/>
    <url-obj-pattern query="*SearchView*"/>
    <url-obj-pattern query="*&login*"/>

```

```

        <url-obj-pattern    query="*Command*"/>
        <!-- crawl Domino: avoid OpenDocument permutations -->
        <url-obj-pattern    query="*ExpandSection*"/>
        <url-obj-pattern    query="*Navigate*"/>
        <url-obj-pattern    query="*Start*"/>
        <!-- -->

    </exclude-pattern-list>
</crawler-config>

<summarizer-config>
<!-- Copy2Rdf e o resumidor predefinido. Para estes tipos, utilize: -->

    <resource-handler content-type="*htm*"
                        summarizable="EipHtmlSummarizable"
                        summary-maker="EipHtmlRawSummaryMaker" />
    <resource-handler content-type="*pdf"
                        summarizable="InsoSummarizable"
                        summary-maker="InsoSummaryMaker" />

</summarizer-config>
v    </group>
    </group-list>
</gcs-config>

```

---

## IBM Web Crawler exemplo de ficheiro de análise de registo

```
D:\gcs\run\log>perl loganalysis.pl log.txt
```

Elapsed time in Log.txt for 7710 lines was 1.84 minutes.

```
GCS was configured for 20 crawlers
999 total crawls attempted
137 - total crawl failures:
      21 GCSHttpConnection.ABANDONING
      12 GCSHttpConnection.CONNECT_ERROR
      16 GCSHttpConnection.UNKNOWN_HOST
       4 HTTP 403
      29 HTTP 404
       2 HTTP 500
       8 HTTP 599
       1 Read timed out
      39 Robots not allowed
       4 over max redirects
       1 unknown protocol

```

```
-----
862 = successfully crawled
  0 - unchanged since earlier crawl

```

```
-----
862 = new or changed
468 crawled per minute

```

```
GCS was configured for 5 summarizers
855 total summaries attempted
  0 - total summary failures:
-----
855 = successfully summarized
144 gcs.summaries.list.host
855 gcs.summaries.list.resource
465 summarized per minute

```

```
GCS successfully crawled 134 servers to obtain 862 URL:
afqa0854.mop.ibm.com: 15
als1f1.yamato.ibm.com: 1
apache.btv.ibm.com: 1

```

apc.endicott.ibm.com: 2  
 as400service.ibm.com: 1  
 atlas.bocaraton.ibm.com: 1  
 autopproxy.ibm.com: 1  
 cer.si.ibm.com: 1  
 commerce.www.ibm.com: 1  
 crmweb.boulder.ibm.com: 3  
 d02ntcl01.ibm.com: 1  
 dacs.endicott.ibm.com: 1  
 duke.toraix.can.ibm.com: 1  
 ebcweb.austin.ibm.com: 1  
 ecspubs.ibmus2.ibm.com: 5  
 edaw3.fishkill.ibm.com: 1  
 endwww.endicott.ibm.com: 1  
 gcs.stl.ibm.com: 1  
 gustwick.austin.ibm.com: 1  
 ibmfnsys.somers.hqregion.ibm.com: 1  
 ibmpnyil.somers.hqregion.ibm.com: 2  
 ifw-www.mul.ie.ibm.com: 1  
 iplswww.nas.ibm.com: 2  
 itirc.ibm.com: 1  
 logosite.services.ibm.com: 1  
 lt.lahulpe.ibm.com: 17  
 messaging.ibm.com: 1  
 mrsmrn04.leeds.uk.ibm.com: 1  
 online.lahulpe.ibm.com: 1  
 page.sg.ibm.com: 1  
 procure.sbyl.ibm.com: 1  
 reso.somers.hqregion.ibm.com: 1  
 ristal.leipzig.de.ibm.com: 1  
 rrhhar.argentina.ibm.com: 1  
 seashore.stl.ibm.com: 1  
 secureway.raleigh.ibm.com: 15  
 service.software.ibm.com: 1  
 software.ibmus2.ibm.com: 1  
 techcenter.austin.ibm.com: 1  
 tr2.fishkill.ibm.com:8080: 1  
 ucd.torolab.ibm.com: 1  
 usmweb.boulder.ibm.com: 1  
 w3-1.ibm.com: 32  
 w3-2.ibm.com: 3  
 w3-3.ibm.com: 108  
 w3-5.ibm.com: 4  
 w3.a-nz.au.ibm.com: 1  
 w3.academy.ibm.com: 1  
 w3.almaden.ibm.com: 2  
 w3.alphaworks.ibm.com: 1  
 w3.ap.ibm.com: 1  
 w3.asca.ibm.com: 7  
 w3.austin.ibm.com: 3  
 w3.boulder.ibm.com: 1  
 w3.br.ibm.com: 1  
 w3.btv.ibm.com: 1  
 w3.can.ibm.com: 40  
 w3.chq.ibm.com: 4  
 w3.coc.ibm.com: 1  
 w3.corporatetechnology.ibm.com: 1  
 w3.cupertino.ibm.com: 1  
 w3.dds.dfw.ibm.com: 17  
 w3.demopkg.ibm.com: 4  
 w3.design.ibm.com: 1  
 w3.developer.ibm.com: 3  
 w3.education.ibm.com: 1  
 w3.emea.ibm.com: 14  
 w3.enterlib.ibm.com: 7  
 w3.finsys.ibm.com: 1  
 w3.gcg.ibm.com: 1

w3.globalfinancing.de.ibm.com: 1  
w3.hakozaki.ibm.com: 1  
w3.houston.ibm.com: 1  
w3.hursley.ibm.com: 5  
w3.iabc.ibm.com: 1  
w3.ibm.com: 180  
w3.ibmfax.ibm.com: 1  
w3.ibmlla.ibm.com: 14  
w3.isicc.de.ibm.com: 1  
w3.itso.ibm.com: 1  
w3.japan.ibm.com: 1  
w3.knowledge.raleigh.ibm.com: 1  
w3.linux.ibm.com: 1  
w3.marketiq.ibm.com: 1  
w3.micro.ibm.com: 2  
w3.mtlisc.can.ibm.com: 1  
w3.munich.ibm.com: 1  
w3.ode.raleigh.ibm.com: 1  
w3.paylink.au.ibm.com: 1  
w3.pisc.uk.ibm.com: 1  
w3.pl.ibm.com: 1  
w3.printers.ibm.com: 1  
w3.pssc.mop.ibm.com: 1  
w3.pssed.au.ibm.com: 1  
w3.raleigh.ibm.com: 3  
w3.rchland.ibm.com: 1  
w3.research.ibm.com: 3  
w3.reserve.ibm.com: 1  
w3.rs6000.ibm.com: 1  
w3.security.ibm.com: 1  
w3.software.ibm.com: 6  
w3.ssd.ibm.com: 1  
w3.stl.ibm.com: 1  
w3.techline.ibm.com: 1  
w3.techsupp.yamato.ibm.com: 1  
w3.torolab.ibm.com: 2  
w3.usergroup.ibm.com: 1  
w3.vendor.pok.ibm.com: 1  
w3.viewblue.ibm.com: 1  
w3.watson.ibm.com: 2  
w3.wdg.uk.ibm.com: 1  
w3.ytal.yasu.ibm.com: 1  
w3.zurich.ibm.com: 1  
w3chq.disbursements.ibm.com: 1  
w3is.lagaude.ibm.com: 1  
w3md.btv.ibm.com: 1  
w3ssd.mainz.de.ibm.com: 1  
w3vm.demopkg.ibm.com: 1  
widweb.raleigh.ibm.com: 1  
wtscpok.itso.ibm.com: 1  
wwas.raleigh.ibm.com: 1  
www-1.ibm.com: 63  
www-3.ibm.com: 4  
www-4.ibm.com: 86  
www.almaden.ibm.com: 1  
www.as400.ibm.com: 1  
www.chips.ibm.com: 1  
www.ibm.com: 52  
www.ieg.ibm.com: 1  
www.patents.ibm.com: 1  
www.pc.ibm.com: 2  
www.rs6000.ibm.com: 23  
www.software.ibm.com: 9  
www.storage.ibm.com: 1  
www.watson.ibm.com: 1

GCS timed out 1 times:

```

w3-3.ibm.com: 1

GCS ignored 42 URL prohibited by robots.txt:
reso.somers.hqregion.ibm.com: 1
w3.education.ibm.com: 1
w3.rchland.ibm.com: 34
w3.zurich.ibm.com: 1
www.ibm.com: 5

GCS skipped 3846 URL (requires gcs.url logging)
59 specified an unsupported protocol:
  protocol not supported gopher: 6
  protocol not supported mailto: 53
1206 had content-types (lower or UPPER case, > 10) that were not included
  .2: 12
  .faq: 13
  .1: 14
  .asp: 16
  .cgi: 21
  .shtml: 90
  .pl: 92
  .gif: 157
  .nsf: 160
  .jpg: 214
  .css: 240
516 URL were on servers and/or paths that were not included
2065 were excluded for these reasons:
  URL longer than 254: 1
  excluded by rule 1: 1210
  excluded by rule 2: 854

```

---

## Capítulo 8. Utilizar a pesquisa de texto e QBIC

A primeira secção deste apêndice explica como configurar e utilizar a pesquisa de texto e a Query by Image Content (QBIC), duas funções disponíveis apenas quando selecciona o conector Content Manager Versão 7.1 durante a instalação do EIP. A segunda secção deste apêndice fornece informações sobre o carregamento do texto exemplo e dados de imagem utilizados com aplicações de demonstração.

---

### Pesquisar documentos utilizando o motor de pesquisa de texto

A pesquisa de texto pode ser integrada com o servidor Content Manager Versão 7.1, para que possa indexar, pesquisar e obter documentos armazenados no Content Manager automaticamente. Os utilizadores podem localizar documentos pesquisando palavras ou frases. O servidor de pesquisa de texto suporta tanto conjuntos de caracteres de único-byte e de duplo-byte e pode ser executado em AIX e Windows.

A pesquisa de texto inclui suporte de documentos estruturado para XML, HTML e documentos com código ASCII, que permite ao utilizador pesquisar termos em secções especificadas de um documento. O utilizador pode pesquisar dados em secções imbricadas. Pode pesquisar por contexto completo XML, por exemplo, pode pesquisar por IBM em títulos e pesquisar por IBM dentro de um título que está numa secção específica. Quando o utilizador especifica um caminho de DTD, a pesquisa de texto pode utilizar de modo dinâmico o DTD adequado a cada documento, assumindo que uma referência ao DTD está guardada como metadados para o documento.

Consulte o manual *Planeamento e Instalação do Enterprise Information Portal* para obter informações sobre planeamento e instalação de um sistema EIP utilizando pesquisa de texto.

### Activar um servidor de pesquisa de texto

Para utilizar um servidor de pesquisa de texto, deve activar a administração para o servidor antes de iniciar o cliente de administração IBM Content Manager for Multiplatforms. Para activar a administração:

1. Inicie o IBM Content Manager for Multiplatforms servidor de bibliotecas.  
Permita que o servidor de bibliotecas termine a construção das classes de índices remissivos.
2. Inicie o servidor de pesquisa de texto na estação de trabalho em que está instalado introduzindo:  

```
imlss -start dlinst
```

onde *dlinst* é o nome da instância do servidor de pesquisa de texto escolhido na altura da instalação ou aquando da utilização do utilitário de comando *imlcfgsv*.

---

## Pesquisar imagens utilizando a Pesquisa por Conteúdo de Imagem (QBIC)

Esta secção apresenta a Pesquisa por Conteúdo de Imagem (QBIC) e explica como configurar e utilizar a QBIC. A função QBIC apenas está disponível se instalar o conector Content Manager Versão 7.1. O QBIC é compatível com os sistemas operativos Windows e AIX.

### Introdução à pesquisa de imagem

O servidor de pesquisa de imagem utiliza a tecnologia IBM QBIC (pesquisa por conteúdo de imagem) para o ajudar a pesquisar objectos por certas propriedades visuais, tal como cor e textura. O servidor de pesquisa de imagens analisa imagens e guarda as informações de imagens numa base de dados. Em seguida, os utilizadores podem executar consultas de imagens, que utilizam as propriedades visuais das imagens, para combinar cores, texturas e as respectivas posições sem as descrever com palavras. Pode combinar consultas baseadas no conteúdo com pesquisas de texto e de palavra-chave para conseguir métodos de obtenção com maior capacidade para bases de dados de imagens e multimédia.

Cada servidor de pesquisa de imagens possui um directório de dados que contém um ou mais bases de dados de pesquisa de imagens, que mantêm os catálogos de pesquisa de imagens. Um catálogo de pesquisa de imagens guarda os dados relativos às funções visuais de uma recolha de imagens. Os objectos de imagem actuais são armazenados em servidores de objectos no sistema IBM Content Manager for Multiplatforms. O servidor de pesquisa de imagens é executado no AIX e no Windows.

Consulte o manual *Planeamento e Instalação do Enterprise Information Portal* para obter informações sobre a instalação da pesquisa de imagens.

### Configurar a pesquisa de imagens

Estas instruções aplicam-se após a instalação da Pesquisa de Imagens, que é automaticamente instalada se seleccionar o conector Content Manager Versão 7.1. A configuração da pesquisa de imagens consiste em:

1. Configurar o ambiente
2. Configurar o servidor de pesquisa de imagens
3. Configurar o cliente de pesquisa de imagens
4. Carregar dados exemplificativos de imagens

**Se utilizar o assistente de instalação no AIX:** Não é necessário executar os scripts de configuração e de definições ou emitir o comando de configuração do servidor. O assistente conclui estas tarefas pelo utilizador.

**Se estiver a instalar no Windows:** deve concluir estas tarefas.

#### Configurar o ambiente

Conclua as tarefas de configuração do ambiente nesta secção nas máquinas de servidor e de cliente. O servidor de pesquisa de imagens requer as seguintes variáveis de ambiente.

##### QBICTOP

Resolve nomes de ficheiros durante a configuração da pesquisa de imagens

##### QbicImagePath

Resolve nomes de ficheiros num ficheiro de imagens do servidor



### **QbicMaskPath**

Resolve nomes de ficheiros num ficheiro de máscaras do servidor

### **QbicSketchPath**

Resolve nomes de ficheiros num ficheiro de desenhos do servidor

### **QbicTextPath**

Resolve nomes de ficheiros num ficheiro de texto do servidor

O cliente de pesquisa de imagens requer apenas a variável de ambiente QBICTOP.

**Exemplo AIX:** No AIX, execute o script de configuração, que gera o script da configuração e, em seguida execute o script de configuração para configurar o ambiente.

1. Execute o seguinte script de configuração:

```
/usr/lpp/cmb/bin/frnconfig.iss QBICTOP
```

onde QBICTOP é o caminho do directório de instalação para os ficheiros de controlo (\*.ini). Defina QBICTOP para /utilizador1/cmb/qbic, em que /utilizador1 é o directório inicial do ID de utilizador de administração da pesquisa de imagem. O ID de utilizador da pesquisa de imagens tem de ter acesso de leitura/escrita a este directório.

Este script gera o script de configuração: frnsetup.iss.

2. No directório inicial do ID de utilizador da pesquisa de imagens, execute:

```
./frnsetup.iss
```

Este script preenche as variáveis de ambiente para servidores e clientes de pesquisa de imagens.

**exemplo Windows:** Para definir as variáveis de ambiente:

1. Selecione **Start**→ **Settings**→ **Control Panel**.
2. Faça duplo clique sobre **System**.
3. Faça clique sobre o separador Ambiente.
4. Defina as variáveis e os valores apresentados na Tabela 8 introduzindo-os nos campos adequados e fazendo clique sobre **Definir**.

**Requisito:** O cliente de pesquisa de imagens requer apenas a variável QBICTOP. Para ambientes de cliente, defina apenas a variável QBICTOP.

*Tabela 8. Variáveis de ambiente de pesquisa de imagens*

| Variável   | Valor          |
|--|----------------|
| QBICTOP  | d:\cmbroot\iss |
| QbicImagePath  | d:\cmbroot\iss |
| QbicMaskPath   | d:\cmbroot\iss |
| QbicSketchPath   | d:\cmbroot\iss |
| QbicTextPath   | d:\cmbroot\iss |
| Em que d: é a unidade em que a pesquisa de imagens está instalada. |                |

## **Configurar o servidor de pesquisa de imagens**

Antes de iniciar o servidor de pesquisa de imagens, deve configurá-lo. A configuração do servidor consiste em concluir a configuração inicial e a verificação de ligação.

Para configurar o servidor:

1. Inicie o intérprete de comandos introduzindo: `qbicadm`
2. Introduza o comando **config server**. Por exemplo,

```
config
server LIBSRVRN FRNADMIN PALAVRAPASSE 9999
```

em que LIBSRVRN é o nome de servidor da biblioteca, FRNADMIN é o ID de utilizador do Content Manager, PALAVRAPASSE é a palavra-passe do Content Manager e 9999 é o número da porta do servidor de pesquisa de imagens.

Consulte “Verificar a ligação” para obter mais informações.

### Configurar o cliente de pesquisa de imagens

Antes de iniciar um cliente de pesquisa de imagens, incluindo o programa de administração do sistema de pesquisa de imagens, é necessário configurá-lo. A administração do sistema Content Manager requer a atribuição de um nome alternativo. Teste a configuração verificando a ligação.

**Atribuir um nome alternativo:** Antes de poder utilizar a administração do sistema Content Manager da pesquisa de imagens, que funcionam como um cliente de pesquisa de imagens, deve atribuir pelo menos um nome alternativo de servidor.

Para atribuir um nome alternativo:

1. Inicie o intérprete de comandos introduzindo: `qbicadm`
2. Introduza o comando **add alias**. Por exemplo,

```
add alias QBICSRV NOMESISCENTRAL 9999
```

em que QBICSRV é o nome alternativo, NOMESISCENTRAL é o nome de sistema central do servidor de pesquisa de imagem e 9999 é o número da porta do servidor de pesquisa de imagem.

### Verificar a ligação:

#### Importante:

1. O servidor de bibliotecas deve estar em execução para estabelecer ligação com o servidor de pesquisa de imagens.
2. A administração do sistema de pesquisa de imagens requer um ID de utilizador do Content Manager existente. Para ligar com êxito à biblioteca de servidores, o ID do utilizador da pesquisa de imagens e o ID do utilizador do servidor de bibliotecas devem ser iguais. O valor predefinido para este ID de utilizador é `frnadmin`. Se alterar este valor, certifique-se que os IDs são correspondentes.

Para verificar a ligação:

1. Após a configuração do servidor de pesquisa de imagens da introdução do nome alternativo, inicie o servidor introduzindo `commsrv` a partir da linha de comandos do servidor.
2. Para iniciar o interpretador de comandos, introduza: `qbicadm`.
3. No intérprete de comandos, introduza o comando `connect`.

```
connect QBICSRV FRNADMIN PALAVRAPASSE
```

em que QBICSRV é o nome alternativo, FRNADMIN é o ID de utilizador do Content Manager, PALAVRAPASSE é a palavra-passe do Content Manager.

Após a ligação ter sido bem sucedida, é apresentada a mensagem 0 servidor da Biblioteca é LIBSRVRN.

4. Para desligar do servidor, introduza: disconnect.
5. Para sair do interpretador de comandos, introduza: quit.

---

## Carregar e indexar dados exemplo

Esta secção explica como carregar e indexar dados de texto e de imagem exemplo, que utiliza com as aplicações exemplo. Esta secção apenas se aplica se tiver instalado o conector Content Manager Versão 7.1 e seleccionado a opção de pesquisa de texto.

Existem vários programas de carregamento exemplo fornecidos com o CD do Enterprise Information Portal. Esta secção descreve como carregar dados de imagem e de texto utilizando o programa de carregamento exemplo LoadSampleTSQBICDL. Pode carregar dados de pesquisa de texto e imagem em separado para se certificar do funcionamento correcto de ambas as funções.

### Antes de carregar os dados exemplo

Antes de executar o programa de carregamento, o utilizador deve:

1. Inicie sessão no cliente de administração do EIP. Clique em Iniciar → Programas → **Enterprise Information Portal** for Multiplatforms 8.2 → Administração.
2. Seleccionar uma base de dados e inicie sessão utilizando o ID de utilizador e palavra-passe correctos. Se seleccionar a base de dados predefinida icmn1sdb, insira **icmadmin** como o ID de utilizador e insira palavra-passe no campo da palavra-passe. Se utilizar outra base de dados, introduza o ID de utilizador aplicável.
3. Crie uma configuração de servidor de bibliotecas utilizando o programa de administração do sistema do Content Manager. Consulte a ajuda online do programa de administração do sistema para obter ajuda para esta tarefa.
4. Altere as propriedades de **Acesso** da configuração do servidor de bibliotecas executando os seguintes passos:
  - a. Clique com o botão direito sobre a nova configuração e clique em **Propriedades** para abrir o bloco de notas Propriedades.
  - b. Faça clique sobre o separador Acesso.
  - c. Faça clique sobre o botão **Sessões ilimitadas a partir de qualquer estação de trabalho**.

### Criar um índice remissivo de pesquisa de texto

Antes de carregar dados, o utilizador deve criar um índice remissivo de pesquisa de texto vazio, que pode utilizar para indexar os exemplos de texto. **Dica:** Só pode criar um índice de pesquisa de texto num servidor Content Manager Versão 6.1 ou Versão 7.1.

Para criar um índice remissivo de pesquisa de texto:

1. Inicie o servidor de pesquisa de texto na estação de trabalho em que está instalado utilizando o seguinte comando:

```
imlss -start dlinst
```

Onde `dlinst` é o nome da instância do servidor de pesquisa de texto escolhido na altura da instalação ou aquando da utilização do utilitário de comando `imlcfgsv`.

2. Inicie sessão no programa de administração do sistema do Content Manager.
3. Seleccione **Pesquisa de Texto** na lista no painel superior esquerdo.
4. Faça duplo clique sobre **Servidores de Pesquisa** no painel da esquerda.
5. Faça duplo clique sobre **TM**. TM é o nome alternativo do servidor de pesquisa para o servidor de pesquisa de texto.
6. Faça duplo clique sobre a pasta **Índices Remissivos** no painel da esquerda. Se a mensagem `RC_EMPTY_LIST` mostrar, então, a partir da barra de menus, clique em **Seleccionado → Novo** para criar um índice remissivo.
7. Na janela Novo Índice Remissivo, defina índice remissivo. Clique sobre **Ajuda** para obter uma descrição detalhada de cada campo.

Por exemplo:

**Para o Windows:**

**Nome** TMINDEX

**Tipo** Precise

**Ficheiros de índice remissivo**

`x:\cmbroot\ts\index\tmlindex` em que `x` é a unidade de instalação; se o caminho não existir será criado.

**Ficheiros de trabalho no índice remissivo**

`x:\cmbroot\ts\work\tmlindex` em que `x` é a unidade de instalação; se o caminho não existir, será criado.

**Introdução de informações**

Nome do servidor de bibliotecas do Content Manager.

Não altere os nomes das DLL assumidas do cliente e do servidor DLL.

**Para o AIX:**

**Nome** TMINDEX

**Tipo** Precise

**Ficheiros de índice remissivo**

`/home/cltadmin/tsindex/index/tmlindex`; se o caminho não existir, será criado.

**Ficheiros de trabalho no índice remissivo**

`/home/cltadmin/tsindex/work/tmlindex`; se o caminho não existir, será criado. Certifique-se de que o utilizador tem autorização para escrever no directório.

**Introdução de informações**

Nome do servidor de bibliotecas do Content Manager.

8. Faça clique sobre **OK**.
9. Faça duplo clique sobre **TMINDEX** para abrir o TMINDEX Administration Notebook.

## Criar a base de dados, o catálogo e as funções da pesquisa de imagens

Depois de ter criado um índice de pesquisa para os dados da pesquisa de texto de exemplo, tem de criar uma base de dados de pesquisa de imagens e catálogo para os dados de imagem de exemplo.

Para criar a base de dados, o catálogo e as funções da pesquisa de imagens:

1. Inicie o servidor de pesquisa de imagens na estação de trabalho em que está instalado introduzindo o seguinte comando:  
`commsrv`
2. Inicie sessão no programa de administração do sistema do Content Manager.
3. Seccione **Pesquisa de Imagens** na lista no painel superior esquerdo.
4. Faça clique sobre **Servidores de Pesquisa de Imagens** no painel da esquerda.
5. Faça clique sobre **QBICSRV**.  
onde QBICSRV é o nome do servidor de pesquisa de imagens que especificou durante a instalação.
6. Faça clique com o botão direito do rato sobre **Bases de dados** no painel da esquerda e seccione **Nova Base de Dados**.
7. Na janela Nova Base de Dados, escreva SAMPLEDB no campo **Nome** e clique sobre **OK**.
8. No painel da esquerda, faça clique sobre **Bases de Dados** para apresentar o ícone **SAMPLEDB** na janela da esquerda.
9. Faça clique sobre **SAMPLEDB**.
10. No painel da esquerda, faça clique com o botão direito do rato sobre **Catálogos** e faça clique sobre **Novo Catálogo**.
11. Na janela Novo Catálogo, escreva SAMPLECAT no campo **Nome** e clique sobre **OK**.
12. No painel da esquerda, faça clique sobre **Catálogos** para apresentar o ícone **SAMPLECAT**.
13. Faça clique sobre **SAMPLECAT**.
14. No painel da esquerda, faça clique com o botão direito do rato sobre **Funções** e faça clique sobre **Novas Funções**.
15. Na janela Novas Funções, seccione cada função no campo **Nome** e faça clique sobre **Aplicar**. Quando as quatro funções estiverem seleccionadas é apresentada a seguinte mensagem:  
Foram adicionadas ao catálogo todas as funções possíveis.
16. Faça clique sobre **OK**.
17. Faça clique sobre **Cancelar**.

## Executar o programa de carregamento

Pode carregar dados exemplo para testar a pesquisa de texto e imagem.

Os dados de imagem exemplo encontram-se nos seguintes ficheiros:

**No Windows:**

`x:\cmbroot\samples\java\dl\samples.jar`

**No AIX:**

`/usr/lpp/cmb/samples/java/dl/samples.jar`

O programa de carregamento exemplo carrega os dados no Content Manager e indexa-os. Leia o prólogo do programa origem para obter instruções sobre a sintaxe utilizada para executar o programa. Os programas de carregamento exemplo são:

**No Windows:**

`x:\cmbroot\samples\java\d1\LoadSampleTSQBICDL.jar`

**No AIX:**

`/usr/lpp/cmb/samples/java/d1/LoadSampleTSQBICDL.jar`

Para executar o programa de carregamento de dados exemplo:

1. Descomprima os ficheiros .jar introduzindo:

`jar -xvf samples.jar`

Os ficheiros são descomprimidos para os directórios correctos.

2. Defina as variáveis de ambiente da estação de trabalho para compilar o programa de carregamento exemplo concluindo as seguintes tarefas:

**No Windows:**

- a. Abra `x:\cmbroot\cmbenv71.bat` num editor de texto e altere as primeiras três linhas para definir as variáveis de ambiente para a estação de trabalho:

```
set CMBROOT = e:\cmbroot
set DB2HOME = e:\sql1ib
set JAVAHOME = d:\jdk117
```

- b. Guarde `cmbenv71.bat` e defina as variáveis de ambiente introduzindo:

`cmbenv71`

**No AIX:**

- a. Vá para `/usr/lpp/cmb/bin/` e execute a configuração introduzindo  
`. /cmbenv71.sh`
- b. Certifique-se de que os subdirectórios em  
`/usr/lpp/cmb/samples/java/d1/` e os ficheiros exemplo podem ser escritos por todos os utilizadores.

3. Compile o programa de carregamento introduzindo o comando sensível a maiúsculas e minúsculas:

`javac LoadSampleTSQBICDL.java`

4. **Requisito:** Os seguintes servidores devem estar em execução antes de iniciar o programa de carregamento:

- Servidor de bibliotecas
- Servidor de objectos
- Servidor de pesquisa de texto
- Servidor de pesquisa de imagens

Se estiver a ser executado numa Versão de Idioma Nacional do Content Manager, defina a variável `FRNDEFLANG` para `ENU` antes de executar o programa de carregamento. O comando do AIX para definir a variável de ambiente é:  
`export FRNDEFLANG=ENU`

5. Carregue os dados exemplo com o programa de carregamento introduzindo:

`java LoadSampleTSQBICDL sampleQBIC.dat load.log frnadmin palavra-passe LIBSRVRN`

onde o seu ID de utilizador é frnadmin, a sua palavra-passe é password e o servidor da biblioteca é LIBSRVRN.

6. Verifique o load.log para se certificar de que os dados de demonstração são carregados com sucesso.

Depois de terminar o carregamento dos dados exemplo, utilize a ferramenta de administração do sistema do Content Manager ou as ferramentas da linha de comandos pesquisa de texto para indexar os dados de texto exemplo.

## Indexar os dados de texto exemplo

Para indexar os dados exemplo:

1. Inicie sessão no programa de administração do sistema do Content Manager.
2. Seccione **Pesquisa de Texto** na lista no painel superior esquerdo.
3. Faça duplo clique sobre **Servidores de Pesquisa**.
4. Faça duplo clique sobre **TM**. TM é o nome alternativo do servidor de pesquisa para o servidor de pesquisa de texto.
5. Faça duplo clique sobre **novo índice de texto** e clique sobre **Propriedades**.
6. Na página Explícito do bloco de notas Propriedades, faça clique sobre **Actualizar**.
7. O campo **Total do índice remissivo** deve apresentar o número de documentos carregados com o programa de carregamento.
8. Faça clique sobre **Índice remissivo** para indexar os ficheiros.
9. Após alguns instantes, faça clique sobre **Actualizar** para ver o número de documentos indexados com êxito no campo **Índice de documentos principal**.

Após a indexação dos dados, pode utilizar a aplicação Java de demonstração para pesquisar a recolha ou executar uma pesquisa simples utilizando a ferramenta de linha de comandos imlsrch.





---

## Capítulo 9. Formatos de Documentos

---

### Formatos de documentos do Information mining

Esta apêndice descreve os formatos de documentos suportados pelo information mining.

#### Processamento de texto: Genérico

|   |                                      |
|---|--------------------------------------|
| Texto ANSI (7 & 8 bits)                         | Todas as versões                     |
| Texto ASCII (versões disponíveis de 7 & 8 bits) | Todas as versões                     |
| HTML  | Versões até 3.0 (algumas limitações) |
| IBM FFT   | Todas as versões                     |
| IBM Revisable Form Text                         | Todas as versões                     |
| Microsoft Rich Text Format (RTF)                | Todas as versões                     |
| Texto em Unicode                                | Todas as versões                     |

#### Processamento de Texto: DOS

|                          |                          |
|--------------------------|--------------------------|
| DEC WPS Plus (DX)        | Versões até 4.0          |
| DEC WPS Plus (WPL)       | Versões até 4.1          |
| DisplayWrite 2 & 3 (TXT) | Todas as versões         |
| DisplayWrite 4 & 5       | Versões até à edição 2.0 |
| Enable                   | Versões 3.0, 4.0 e 4.5   |
| First Choice             | Versões até 3.0          |
| Framework                | Versão 3.0               |
| IBM Writing Assistant    | Versão 1.01              |
| Lotus Manuscript         | Versões até 2.0          |
| MASS11                   | Versões até 8.0          |
| Microsoft Word           | Versões até 6.0          |
| Microsoft Works          | Versões até 2.0          |
| MultiMate                | Versões até 4.0          |
| Navy DIF                 | Todas as versões         |
| Nota Bene                | Versão 3.0               |
| Office Writer            | Versão 4.0 até 6.0       |
| PC-File Letter           | Versões até 5.0          |
| PC-File+ Letter          | Versões até 3.0          |
| PFS:Write                | Versões A, B e C         |

|                              |                           |
|------------------------------|---------------------------|
| <b>Professional Write</b>    | Versões até 2.1           |
| <b>Q&amp;A</b>               | Versão 2.0                |
| <b>Samna Word</b>            | Versões até 4.0           |
| <b>SmartWare II</b>          | Versão 1.02               |
| <b>Sprint</b>                | Versões 1.0               |
| <b>Total Word</b>            | Versão 1.2                |
| <b>Volkswriter 3 &amp; 4</b> | Versões até 1.0           |
| <b>Wang PC (IWP)</b>         | Versões até 2.6           |
| <b>WordMARC</b>              | Versões até Composer Plus |
| <b>WordPerfect</b>           | Versões até 6.1           |
| <b>WordStar</b>              | Versões até 7.0           |
| <b>WordStar 2000</b>         | Versão até 3.0            |
| <b>XyWrite</b>               | Versões até III Plus      |

## **Processamento de Texto: International**

|                             |                                   |
|-----------------------------|-----------------------------------|
| <b>JustSystems Ichitaro</b> | Versões 5.0, 6.0, 8.0, 9.0 e 10.0 |
|-----------------------------|-----------------------------------|

## **Processamento de Texto: Windows**

|  |                                |
|--|--------------------------------|
| <b>AMI/AMI Professional</b>                            | Versões até 3.1                |
| <b>Corel WordPerfect for Windows</b>                   | Versões até 9.0                |
| <b>JustWrite</b>                                       | Versões até 3.0                |
| <b>Legacy</b>  | Versões até 1.1                |
| <b>Lotus WordPro (Win32 / Intel plataformas)</b>       | SmartSuite 96, 97 e Millennium |
| <b>Lotus WordPro (plataformas Unix - apenas texto)</b> | SmartSuite 97 e Millennium     |
| <b>Microsoft Windows Works</b>                         | Versões até 4.0                |
| <b>Microsoft Windows Write</b>                         | Versões até 3.0                |
| <b>Microsoft Word 97</b>                               | Word 97                        |
| <b>Microsoft Word 2000</b>                             | Word 2000                      |
| <b>Microsoft Word for Windows</b>                      | Versões até 7.0                |
| <b>Microsoft WordPad</b>                               | Todas as versões               |
| <b>Novell Perfect Works</b>                            | Versão 2.0                     |
| <b>Novell WordPerfect for Windows</b>                  | Versões até 7.0                |
| <b>Professional Write Plus</b>                         | Versão 1.0                     |
| <b>Q&amp;A Write for Windows</b>                       | Versões 3.0                    |
| <b>WordStar for Windows</b>                            | Versão 1.0                     |

## Processamento de Texto: Macintosh

|                   |                      |
|-------------------|----------------------|
| Microsoft Word    | Versões 4.0 até 6.0  |
| Microsoft Word 98 | Word 98              |
| WordPerfect       | Versões 1.02 até 3.0 |
| Microsoft Works   | Versões até 2.0      |
| MacWrite II       | Versão 1.1           |

## Formatos de folhas de cálculo

|                                    |                            |
|------------------------------------|----------------------------|
| VP Planner 3D                      | Versão 1.0                 |
| Enable                             | Versões 3.0, 4.0 e 4.5     |
| First Choice                       | Versões até 3.0            |
| Framework                          | Versão 3.0                 |
| Lotus 1-2-3 (DOS & Windows)        | Versões até 5.0            |
| Lotus 1-2-3 for SmartSuite         | SmartSuite 97 e Millennium |
| Lotus 1-2-3 Charts (DOS & Windows) | Versões até 5.0            |
| Lotus 1-2-3 (OS/2)                 | Versões até 2.0            |
| Lotus 1-2-3 Charts (OS/2)          | Versão até 2.0             |
| Lotus Symphony                     | Versões 1.0, 1.1 e 2.0     |
| Microsoft Excel 97                 | Excel 97                   |
| Microsoft Excel 2000               | Excel 2000                 |
| Microsoft Excel Macintosh          | Versão 3.0 até 4.0, 98     |
| Microsoft Excel Windows            | Versões 2.2 até 7.0        |
| Microsoft Excel Charts             | Versões 2.x até 7.0        |
| Microsoft Multiplan                | Versão 4.0                 |
| Microsoft Windows Works            | Versões até 4.0            |
| Microsoft Works (DOS)              | Versões até 2.0            |
| Microsoft Works (Mac)              | Versões até 2.0            |
| Mosaic Twin                        | Versão 2.5                 |
| Novell Perfect Works               | Versão 2.0                 |
| QuattroPro for DOS                 | Versões até 5.0            |
| QuattroPro for Windows             | Versões até 9.0            |
| PFS:Professional Plan              | Versão 1.0                 |
| SuperCalc 5                        | Versão 4.0                 |
| SmartWare II                       | Versão 1.02                |

## Formatos de Bases de Dados

|              |             |
|--------------|-------------|
| SmartWare II | Versão 1.02 |
|--------------|-------------|

|                                |                        |
|--------------------------------|------------------------|
| <b>Access</b>                  | Versões até 2.0        |
| <b>dBase</b>                   | Versões até 5.0        |
| <b>DataEase</b>                | Versão 4.x             |
| <b>dBXL</b>                    | Versão 1.3             |
| <b>Enable</b>                  | Versões 3.0, 4.0 e 4.5 |
| <b>First Choice</b>            | Versões até 3.0        |
| <b>FoxBase</b>                 | Versão 2.1             |
| <b>Framework</b>               | Versão 3.0             |
| <b>Microsoft Windows Works</b> | Versões até 4.0        |
| <b>Microsoft Works (DOS)</b>   | Versões até 2.0        |
| <b>Microsoft Works (Mac)</b>   | Versões até 2.0        |
| <b>Paradox (DOS)</b>           | Versões até 4.0        |
| <b>Paradox (Windows)</b>       | Versões até 1.0        |
| <b>Personal R:BASE</b>         | Versão 1.0             |
| <b>R:BASE 5000</b>             | Versões até 3.1        |
| <b>R:BASE System V</b>         | Versão 1.0             |
| <b>Q &amp; A</b>               | Versões até 2.0        |
| <b>Reflex</b>                  | Versão 2.0             |

## Formatos padrão de gráficos

|  |  |
|--|--|
| <b>PNG - Portable Network Graphics Internet Format</b>           | Versão 1.0                                       |
| <b>Binary Group 3 Fax</b>  | Todas as versões                                 |
| <b>BMP(incluindo RLE, ICO, CUR &amp; os/2 DIB)</b>               | Windows  |
| <b>CDR (se a imagem TIFF estiver aí incorporada)</b>             | Coral Draw versões 2.0 – 9.0                     |
| <b>CGM - Computer Graphics Metafile</b>                          | ANSI, CALS, NIST, Versão 3.0                     |
| <b>CMX - Corel Clip Art Format</b>                               | Versões 5 até 6                                  |
| <b>DCX (PCX de múltiplas páginas)</b>                            | Microsoft Fax                                    |
| <b>DRW - Micrografx Designer</b>                                 | Versão 3.1                                       |
| <b>DRW - Micrografx Draw</b>                                     | Versões até 4.0                                  |
| <b>DXF (Binary and ASCII) AutoCAD Drawing Interchange Format</b> | Versões até 14                                   |
| <b>EMF</b>   | Windows Enhanced Metafile                        |
| <b>EPS Encapsulated PostScript</b>                               | Se a imagem TIFF estiver aí incorporada          |
| <b>FMV - FrameMaker graphics</b>                                 | Formato vector e por quadrícula até à Versão 5.0 |
| <b>FPX - Kodak Flash Pix</b>                                     | Sem formato específico                           |

|   |                       |
|---|-----------------------|
| <b>GDF - IBM Graphics Data Format</b>   | Versão 1.0            |
| <b>GEM - Graphics Environment Manager Metafile</b>                              | Bitmap e Vector       |
| <b>GIF - Graphics Interchange Format</b>  | Compuserve            |
| <b>GP4 - Formato Group 4 CALS</b>   | Tipo I e Tipo II      |
| <b>HPGL - Hewlett Packard Graphics Language</b>                                 | Versão 2.0            |
| <b>IMG - GEM Paint</b>  | Sem versão específica |
| <b>JFIF (JPEG sem ser em formato TIFF)</b>                                      | Todas as versões      |
| <b>JPEG - Formato Joint Photographic Experts Group</b>                          | Todas as versões      |
| <b>MET - OS/2 PM Metafile</b>   | Versão 3.0            |
| <b>PBM - Portable Bitmap</b>  | Sem versão específica |
| <b>Kodak Photo CD</b>   | Versão 1.0            |
| <b>PCD - PCX Bitmap</b>   | PC Paintbrush         |
| <b>Perfect Works (Draw)</b>   | Novell versão 2.0     |
| <b>PGM - Portable Graymap</b>   | Sem versão específica |
| <b>PIC - Lotus 1-2-3 Picture File Format</b>                                    | Sem versão específica |
| <b>PICT1 &amp; PICT2 (Raster)</b>   | Macintosh Standard    |
| <b>PIF - IBM Picture Interchange Format</b>                                     | Versão 1.0            |
| <b>PNTG</b>   | MacPaint              |
| <b>PPM - Portable Pixmap</b>  | Sem versão específica |
| <b>Progressive JPEG</b>   | Sem versão específica |
| <b>PSP - Paintshop Pro (apenas Win32)</b>                                       | Versões 5.0, 5.0.1    |
| <b>RND - AutoShade Rendering File Format</b>                                    | Versão 2.0            |
| <b>SDW Ami Draw Snapshot (Lotus)</b>  | Todas as versões      |
| <b>SRS - Sun Raster File Format</b>   | Sem versão específica |
| <b>Targa</b>  | Truevision            |
| <b>TIFF</b>   | Versões até 6         |
| <b>TIFF CCITT Group 3 &amp; 4</b>   | Sistemas de Fax       |
| <b>VISO (Modo de Visualização Prévia apenas para Versão 4) Visio 4, 5, 2000</b> | Visio 4, 5, 2000      |
| <b>WMF</b>  | Windows Metafile      |

|  |                    |
|--|--------------------|
| <b>WordPerfect Graphics [WPG e WPG2]</b> | Versões até 2.0    |
| <b>XBM - X-Windows Bitmap</b>            | compatível com x10 |
| <b>XPM - X-Windows Pixmap</b>            | compatível com x10 |
| <b>XWD - X-Windows Dump</b>              | compatível com x10 |

## **Formatos de gráficos com grande capacidade**

|   |   |
|---|---|
| <b>PSD - Adobe Photoshop File Format</b>              | Versão 4.0  |
| <b>AI - Adobe Illustrator File Format</b>             | Versões até 7.0                                     |
| <b>CDR - Corel Draw</b>                               | Versões até 8.0                                     |
| <b>DSF - Micrografx Designer</b>                      | Windows 95, versão 6.0                              |
| <b>DWG - AutoCAD Native Drawing Format</b>            | Versões 12 até 14                                   |
| <b>IGES - Initial Graphics Exchange Specification</b> | Versão 5.1  |
| <b>PDF - Portable Document Format</b>                 | Acrobat versão 2.1, 3.0, 4.0, incluindo PDF Japonês |
| <b>PS - Postscript</b>                                | Nível 2   |

## **Formatos de Apresentação**

|  |                            |
|--|----------------------------|
| <b>Microsoft PowerPoint para Macintosh</b> | Versão 4.0, 98             |
| <b>Apresentações em Corel</b>              | Versão 8.0 e 9.0           |
| <b>Apresentações em Novell</b>             | Versões 3.0 e 7.0          |
| <b>Harvard Graphics for DOS</b>            | Versões 2.x e 3.x          |
| <b>Harvard Graphics</b>                    | Versões Windows            |
| <b>Freelance 96</b>                        | Freelance 96               |
| <b>Freelance for Windows 95</b>            | SmartSuite 97 e Millennium |
| <b>Freelance for Windows</b>               | Versão 1.0 e 2.0           |
| <b>Freelance for OS/2</b>                  | Versões até 2.0            |
| <b>Microsoft PowerPoint for Windows</b>    | Versões até 7.0            |
| <b>Microsoft PowerPoint 97</b>             | PowerPoint 97              |
| <b>Microsoft PowerPoint 2000</b>           | PowerPoint 2000            |

## **Formatos compactados e codificados**

|                                     |                       |
|-------------------------------------|-----------------------|
| <b>ZIP PKWARE</b>                   | Versões até 2.0g      |
| <b>GZIP</b>                         | Sem versão específica |
| <b>LZA Self Extracting Compress</b> | Sem versão específica |

|                                |                       |
|--------------------------------|-----------------------|
| <b>LZH Compress</b>            | Sem versão específica |
| <b>Microsoft Binder</b>        | Versão 7.0, Binder 97 |
| <b>MIME (correio de texto)</b> | Sem versão específica |
| <b>UUEncode</b>                | Sem versão específica |
| <b>UNIX Compress</b>           | Sem versão específica |
| <b>UNIX TAR</b>                | Sem versão específica |

## **Outra**

|   |                                      |
|---|--------------------------------------|
| <b>vCard Electronic Business Card</b>   | Versão 2.1                           |
| <b>Executable (EXE, DLL)</b>            | Sem versão específica                |
| <b>Executable for Windows NT</b>        | Sem versão específica                |
| <b>MSG (apenas texto)</b>               | Formato de correio Microsoft Outlook |
| <b>Microsoft Project (apenas texto)</b> | Project 98                           |





---

## Capítulo 10. Gestão de direitos

Este capítulo introduz a característica gestão de direitos EIP e descreve conceitos de gestão de direitos. Explica técnicas de marcação que pode utilizar para proteger a sua propriedade.

---

### Proteger a sua propriedade intelectual

Os objectos multimédia armazenados em forma digital são a sua propriedade intelectual. A protecção destes objectos pode ser crítica para os seus objectivos empresariais, especialmente quando estes objectos estão no World Wide Web onde copiar é relativamente fácil. Pode também utilizar tecnologia de marcação fornecida com o Content Manager para dissuadir a utilização não-autorizada da propriedade intelectual através da marcação dos objectos digitais de multimédia para protecção ou através da impressão digital dos objectos para identificação.

Pode aplicar uma marca aos objectos valiosos para:

- Identificar a fonte, para dissuadir cópia ou reutilização não-autorizada. Isto é conhecido por marca de água e é tipicamente visível.
- Identificar o recipiente do conteúdo, para dissuadir cópia ou reutilização não-autorizada. Isto é conhecido como uma impressão digital e é tipicamente invisível.
- Fornecer um contacto para obter informações adicionais.
- Fornecer informações, como a hora e a data, para utilização numa cadeia de distribuição de valor acrescentado.

Pode marcar os objectos digitais antes de o objecto ser entregue aos clientes. Ambas as marcas de água e impressões digitais podem ser aplicadas antes da entrega. Contudo, aplicar uma impressão digital antes da entrega implica que o recipiente seja conhecido e pode necessitar que a marca seja aplicada de forma dinâmica no processo de entrega. Aplicar a marca a partir do seu próprio ambiente controlado antes de a entrega aumentar a segurança, porque existe menos risco de interferência.

Pode aplicar marcas em várias etapas no processo de gestão e entrega. A sua situação irá influenciar aquilo considerado apropriado. As marcas podem ser aplicadas nos seguintes pontos no seu processo:

- Antes de o objecto ser armazenado  
Se uma marca comum for para ser utilizada para o objecto (por exemplo, uma marca de água visível para identificar o proprietário), pode aplicar a marca antes ou enquanto armazena o objecto. Pode armazenar ambos os objectos não-marcado e marcado no sistema Content Manager. Ou pode armazenar apenas o objecto marcado e manter o objecto não-marcado num depósito separado.
- Depois de o objecto ser armazenado  
Se pretende marcar objectos que são armazenados no sistema Content Manager, pode recuperar o objecto, marcá-lo, e substituir o objecto não-marcado com a versão marcada ou armazenar o objecto marcado como um artigo novo.
- Quando o objecto é recuperado

Se a marca a ser aplicada varia com base no recipiente, pode aplicar a marca de forma dinâmica depois de o objecto ser recuperado. O objecto marcado pode então ser entregue em vez do objecto original.

Se tiver muitos objectos de legado no sistema que não estejam marcados e não pretende utilizar os recursos para voltar a marcar cada objecto, pode marcar objectos quando estes são recuperados.

---

## Utilizar técnicas de marcação

Tem à sua disposição uma variedade de técnicas de marcação de conteúdo para escolher. Cada técnica trata um problema específico e difere na resistência à remoção e modificação.

As marcações são caracterizadas por:

- Informações transmitidas

### **Marca de Água**

Identifica a fonte do conteúdo. Pode conter informações tais como o proprietário e versão do objecto.

### **Impressão digital**

Identifica o recipiente do conteúdo. Pode conter informações tais como onde e a quem o objecto é entregue.

- Visibilidade

### **Visível**

A marca é visível e pode ser identificada.

### **Invisível**

A marca é escondida numa imagem.

- Integridade

**Frágil** A marca é quebrada por qualquer modificação.

### **Robusto**

A marca resiste a modificações ao objecto marcado tais como redimensionar, compactar, rodar e cortar.

- Hora de aplicação

- Quando o objecto é capturado
- Quando o objecto é armazenado
- Quando o objecto é recuperado para distribuição para um cliente
- Quando o objecto é recebido na estação de trabalho do recipiente

- Localização

- Se se pretende que uma marca visível funcione como dissuasora da reutilização ilegal, aquela pode ser aplicada a uma grande parte de uma imagem. A marca pode ser movida para sobrepor uma área da imagem com mais textura para dificultar a remoção.
- Se for utilizada uma marca invisível, uma área com textura da imagem permite que os dados sejam incorporados com o menor efeito na imagem.
- Se se pretende que uma marca visível indique pertença, pode ser colocada sem obstruir num canto da imagem.
- Se tiverem de ser utilizadas marcas visíveis e invisíveis, a marca visível deve ser aplicada primeiro.

- Formato

### **Binário**

A marca pode ser uma sequência aleatória de bites repetida ao longo da imagem. Esta sequência aleatória é a chave que pode ser utilizada para marcar ou desmarcar a imagem.

A marca também pode ser uma imagem.

### **Dados estruturados**

A marca pode ser constituída por dados textuais incorporados.

## **Marcação visível**

Uma marca visível é uma máscara transparente colocada sobre uma imagem de tal forma que tanto a máscara e a imagem são visíveis. Uma marca visível que é difícil de remover é um dissuasor eficiente para a apropriação inadequada dos objectos.

Utilize uma marca visível para as seguintes situações:

- Quando pretende fornecer imagens para os clientes revisualizarem, mas desencorajar os clientes de reutilizar as cópias de revisualização.
- Quando pretende utilizar uma imagem numa publicidade na World Wide Web

## **Marcação invisível**

Uma marca invisível é constituída por dados escondidos numa imagem de tal forma que a imagem parece não ter sido alterada. É necessária uma aplicação para aplicar, detectar e decifrar a marca.

Utilize uma marca invisível para as seguintes situações:

- Quando pretende incorporar informações para identificar pertença e desencorajar cópias ilegais de objectos marcados (marcação de água).
- Quando pretende incorporar informações para traçar um caminho de distribuição (impressão digital)
- Quando pretende incorporar uma anotação ou uma legenda numa imagem



---

## Capítulo 11. Características de Acessibilidade

Este produto inclui um número de características que o tornam mais acessível a pessoas com deficiências. Estas características incluem:

- A capacidade para realizar todas as funções através do teclado em vez do rato.
- Suporte para as propriedades de ecrã aumentado
- Compatibilidade com tecnologias de assistência
- Compatibilidade com as características de acessibilidade ao sistema operativo
- Formatos acessíveis de documentação

---

### Entrada do teclado e navegação

Estão disponíveis as seguintes funções para entrada de teclado e navegação:

#### **Entrada de teclado**

Pode utilizar o teclado em vez de um rato para trabalhar com o produto.

Os artigos e controlos do menu fornecem teclas de acesso que lhe permitem activar um controlo ou seleccionar um artigo de menu directamente a partir do teclado. Estas teclas são auto-documentativas; as teclas de acesso estão sublinhadas no controlo ou no menu onde aparecem.

#### **Destaque do teclado**

Em sistemas baseados em Windows, a posição de destaque do teclado está realçada, indicando qual a área da janela que está activa e em que local os batimentos de teclas surtirão efeito.

#### **Ajustes ao tempo de resposta**

Em sistemas baseados em Windows, pode ajustar os tempos de resposta através do painel de controlo.

---

### Características para ecrã acessível

Os clientes dispõem de uma série de características que aperfeiçoam a interface do utilizador, melhorando a acessibilidade para utilizadores com deficiências de visão. Estes aperfeiçoamentos incluem suporte para definições de contraste elevado e propriedades de tipos de letra personalizáveis.

#### **Modo de contraste elevado**

Os clientes suportam a opção de modo de contraste elevado que é fornecida pelo sistema operativo. Esta característica suporta um maior contraste entre cores de fundo e de primeiro plano.

#### **Definições do Tipo de Letra**

Em sistemas baseados em Windows, pode especificar definições de visualização que determinam a cor, tamanho e tipo de letra para o texto nos menus e janelas de diálogo. O cliente permite-lhe seleccionar o tipo de letra para a lista de documentos.

#### **Não-dependência de cor**

Não é necessário ter a capacidade de distinguir cores para poder utilizar qualquer das funções deste produto.

---

## Compatibilidade com tecnologias de assistência

Os clientes são compatíveis com aplicações de leitura de ecrã tal como Narrator e Via Voice. Os clientes possuem propriedades requeridas para estas aplicações de acessibilidade de forma a disponibilizar informações no ecrã para utilizadores com deficiências visuais.

---

## Documentação Acessível

A documentação para este produto está disponível em formato PDF. Pode converter os ficheiros PDF em HTML ou texto utilizando as ferramentas gratuitas disponibilizadas pela Adobe em [access.adobe.com](https://access.adobe.com). Este facto permite que os utilizadores possam visualizar a documentação de acordo com as preferências de visualização definidas nos respectivos browsers. Permite também a utilização de leitores de ecrã e outras tecnologias de assistência.

---

## Informações

Estas informações foram desenvolvidas para produtos e serviços disponibilizados nos E.U.A.

Os produtos, os serviços ou as funções descritas neste documento poderão não ser disponibilizados pela IBM noutros países. Consulte o seu representante IBM para obter informações sobre os produtos e serviços actualmente disponíveis na sua área. Quaisquer referências, nesta publicação, a programas licenciados IBM ou outros produtos ou serviços IBM, não significam que apenas esses programas licenciados, produtos ou serviços IBM possam ser utilizados. Qualquer outro produto, programa ou serviço, funcionalmente equivalente, poderá ser utilizado em substituição daqueles, desde que não infrinja nenhum dos direitos de propriedade intelectual da IBM. No entanto, é da inteira responsabilidade do utilizador avaliar e verificar o funcionamento de qualquer produto, programa ou serviço não IBM.

Nesta publicação, podem ser feitas referências a patentes ou pedidos de patente pendentes. O facto de este documento lhe ser oferecido não lhe confere quaisquer direitos sobre essas patentes. Todos os pedidos de informação sobre licenças deverão ser endereçados a:

IBM Director of Licensing  
IBM Portugal, SA  
North Castle Drive  
Armonk, NY 10504-1785  
PORTUGAL

Pode endereçar os seus pedidos de informação sobre licenças relacionados com informação de duplo byte (DBCS) ao Departamento de Propriedade Intelectual IBM no seu país. Também pode enviá-los, por escrito, para:

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

**O parágrafo seguinte não se aplica ao Reino Unido ou a nenhum outro país onde tais cláusulas sejam compatíveis com a lei local:** A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO “TAL COMO ESTÁ” SEM GARANTIA DE QUALQUER ESPÉCIE, QUER EXPLÍCITA QUER IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS IMPLÍCITAS DE NÃO INFRACÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO FIM. Alguns Estados não permitem a exclusão de garantias, quer explícitas quer implícitas, em determinadas transacções; esta declaração pode, portanto, não se aplicar ao seu caso.

É possível que estas informações contenham imprecisões técnicas ou erros de tipografia. A IBM permite-se fazer alterações periódicas às informações aqui contidas; essas alterações serão incluídas nas posteriores edições desta publicação. A IBM pode introduzir melhorias e/ou alterações ao(s) produto(s) e/ou programa(s) descrito(s) nesta publicação em qualquer altura sem aviso prévio.

Quaisquer referências, nesta publicação, a sítios da Web não IBM são fornecidas apenas para conveniência e não deverão nunca servir como aprovação desses sítios da Web. Os materiais existentes nesses sítios da Web não fazem parte dos materiais destinados a este produto IBM e a utilização desses sítios da Web será da exclusiva responsabilidade do utilizador.

A IBM pode utilizar ou distribuir qualquer informação que lhe seja fornecida, de qualquer forma que julgue apropriada, sem incorrer em qualquer obrigação para com o autor dessa informação.

Os possuidores de licenças deste programa que pretendam obter informações sobre o mesmo com o objectivo de permitir: i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização recíproca das informações que tenham sido trocadas, deverão contactar a:

IBM Portugal, SA  
Praça de Alvalade N° 7  
1700-036 Lisboa  
PORTUGAL

Tais informações poderão estar disponíveis, sujeitas aos termos e às condições adequadas, incluindo, nalguns casos, o pagamento de uma taxa.

O programa licenciado descrito neste documento e todo o material licenciado disponível para o programa são fornecidos pela IBM de acordo com os termos do IBM Customer Agreement, do IBM International Program License Agreement or de qualquer outro acordo equivalente entre ambas as partes.

Quaisquer dados de rendimento aqui contidos foram determinados num ambiente controlado. Por conseguinte, os resultados obtidos em outros ambientes operativos podem variar significativamente. Algumas medições podem ter sido efectuadas em sistemas ao nível do desenvolvimento, pelo que não existem garantias de que estas medições sejam iguais nos sistemas normalmente disponíveis. Para além disso, algumas medições podem ter sido calculadas por extrapolação. Os resultados reais podem variar. Os utilizadores deste documento devem verificar os dados aplicáveis ao seu ambiente específico.

A informação relativa a produtos não IBM foi obtida a partir dos fornecedores desses produtos, dos seus comunicados ou de outras fontes de divulgação ao público. A IBM não testou esses produtos e não pode confirmar a exactidão do rendimento, da compatibilidade ou de quaisquer outras afirmações relacionadas com produtos não IBM. Todas as questões sobre as capacidades dos produtos não IBM deverão ser endereçadas aos fornecedores desses produtos.

Todas as afirmações relativas às directivas ou tendências futuras da IBM estão sujeitas a alterações ou descontinuação sem aviso prévio, representando apenas metas e objectivos.

Esta publicação contém exemplos de dados e relatórios utilizados em operações comerciais diárias. Para os ilustrar o melhor possível, os exemplo incluem nomes de indivíduos, firmas, marcas e produtos. Todos estes nomes são fictícios e qualquer semelhança com nomes e moradas reais é mera coincidência.

#### LICENÇA DE COPYRIGHT:

Estas informações contém programas de aplicações exemplo no idioma de origem, que ilustram as técnicas de programação em várias plataformas operativas. Pode



copiar, modificar e distribuir estes programas exemplo de qualquer forma, sem encargos para com a IBM, com a finalidade de desenvolver, utilizar, comercializar ou distribuir programas de aplicação conformes à interface de programação de aplicações e destinados à plataforma operativa para a qual os programas exemplo são escritos. Estes exemplos não foram testados exaustivamente sob todas as condições. Deste modo, a IBM não garante nem se responsabiliza pela fiabilidade, assistência ou funcionamento implícito destes programas. Pode copiar, modificar e distribuir estes programas exemplo de qualquer forma, sem encargos para com a IBM, com o objectivo de desenvolver, utilizar, comercializar ou distribuir programas de aplicação em conformidade com as interfaces de programação de aplicações da IBM.

---

## Marcas Comerciais

Os seguintes termos são marcas comerciais da International Business Machines Corporation nos Estados Unidos, noutros países ou em ambos:

|                                |               |                  |
|--------------------------------|---------------|------------------|
| IBM                            | DisplayWrite  | PowerPC          |
| 400                            | e-business    | PTX              |
| Funcionamento Avançado em Rede | HotMedia      | QBIC             |
| Unidade-a-Unidade              |               |                  |
| AIX                            | Hummingbird   | RS/6000          |
| AIXwindows                     | ImagePlus     | SecureWay        |
| APPN                           | IMS           | SP               |
| AS/400                         | Micro Channel | VideoCharger     |
| C Set ++                       | MQSeries      | Visual Warehouse |
| CICS                           | MVS/ESA       | VisualAge        |
| DATABASE 2                     | NetView       | VisualInfo       |
| DataJoiner                     | OS/2          | WebSphere        |
| DB2                            | OS/390        |                  |
| DB2 Universal Database         | PAL           |                  |

Approach, Domino, Lotus, Lotus 1-2-3, Lotus Notes e SmartSuite são marcas comerciais ou marcas comerciais registadas da Lotus Development Corporation nos Estados Unidos, em outros países ou ambos.

Intel e Pentium são marcas comerciais ou marcas comerciais registadas da Intel Corporation nos Estados Unidos, em outros países ou ambos.

Microsoft, Windows e Windows NT são marcas registadas da Microsoft Corporation nos Estados Unidos, outros países ou ambos.

Java e todas as marcas comerciais e todos os logotipos baseados em Java são marcas comerciais da Sun Microsystems, Inc., nos Estados Unidos, noutros países ou ambos.

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

Outros nomes de empresas, produtos e serviços podem ser marcas comerciais ou marcas de serviços dessas empresas.



---

## Glossário

Este glossário define termos e abreviaturas específicas deste sistema. Os termos em *itálico* são definidos noutra local deste glossário.

### A

**ADSM.** Consulte *Tivoli Storage Manager*.

**agente iterativo.** Uma classe ou construção utilizada para se deslocar numa recolha de objectos, um de cada vez.

**API.** Consulte *interface de programação de aplicações*.

**aplicação cliente.** Uma aplicação escrita com as APIs orientadas para objectos e de Internet para aceder aos *servidores de conteúdos* a partir do Enterprise Information Portal.

**área de transferência ascendente.** A área de armazenamento de trabalho para o *gestor de recursos*. Também denominada como memória cache do *gestor de recursos*.

**arquivador de suportes.** Um dispositivo físico utilizado para armazenar sequências de dados de áudio e vídeo. O VideoCharger é um tipo de arquivador de suportes.

**arquivo de dados.** (1) Termo genérico para designar um local (tal como um sistema de bases de dados, ficheiro ou directório) onde são armazenados dados. (2) Num programa de aplicações, uma representação visual de um *servidor de conteúdos*.

**arquivo de dados associado.** Representação virtual de qualquer número de *servidores de conteúdos* específicos, tais como Content Manager.

**artigo .** Termo genérico para a unidade mais pequena de informação que o Enterprise Information Portal administra. Cada artigo tem um identificador. Por exemplo, um artigo pode ser uma *pasta* ou um *documento*.

**artigo de trabalho.** No fluxo de trabalho do anterior Content Manager e em fluxo de trabalho avançado do Enterprise Information Portal, qualquer actividade de trabalho que esteja activa dentro de um dado *fluxo de trabalho*.

**atributo.** Uma unidade de dados que descreve uma certa característica ou propriedade (por exemplo, nome, endereço, idade e por aí adiante) de um artigo, e que pode ser utilizada para atribuir esse artigo. Um atributo tem um tipo, que indica o intervalo das informações armazenadas por esse atributo, e um valor, que está

dentro desse intervalo. Por exemplo, as informações acerca dum ficheiro num sistema de ficheiros multimédia, tais como, título, tempo de execução ou tipo de codificação (MPEG1, H.263, e por aí adiante). Para o Enterprise Information Portal, consulte também o *atributo federado* e o *atributo nativo*.

**atributo federado.** Uma categoria de metadados do Enterprise Information Portal definida por correspondência com *atributos nativos* num ou mais *servidores de conteúdos*. Por exemplo, o atributo federado, número de política, pode ser correlacionado para um *atributo*, num política, no Content Manager e para um atributo, policy ID, em Content Manager ImagePlus for OS/390.

**atributo nativo.** Uma característica de um objecto que é gerido num *servidor de conteúdos* e que é específico desse servidor de conteúdos. Por exemplo, o *campo-chave* número de política pode ser um atributo nativo num servidor de conteúdos do Content Manager enquanto o campo ID de política pode ser um atributo nativo num servidor de conteúdos do Content Manager OnDemand.

**Áudio/Vídeo Intercalado (AVI).** Uma especificação de ficheiro RIFF (*Formato de Ficheiro de Intercâmbio de Recursos*) que permite que os dados áudio e vídeo para serem intercalados num ficheiro. As pistas separadas podem ser acedidas em porções alternadas para ouvir ou gravar enquanto mantém o acesso sequencial no dispositivo de ficheiros.

**AVI.** Consulte *Áudio/Vídeo Intercalados*.

### B

**BLOB.** Consulte *objecto binário grande*

### C

**cadeia de consulta.** Uma cadeia de caracteres que especifica as propriedades e valores de propriedade de uma consulta. Pode criar a cadeia de consulta numa aplicação e transmiti-la à consulta.

**campo-chave.** Consulte *atributo*.

**carácter global.** Um carácter especial, tal como, um asterisco (\*) ou um ponto de interrogação (?) que pode ser utilizado para representar um ou mais caracteres. Qualquer carácter ou conjunto de caracteres pode substituir um carácter global.

**cardinalidade.** O número de linhas numa tabela de base de dados.

**CGI.** Consulte *Interface de porta de ligação comum*.

**CIÉ.** Consulte *ficheiro de intercâmbio comum*.

**CIU.** Consulte *unidade de intercâmbio comum*.

**classe.** Na concepção ou programação orientada para objectos, pode ser definido um modelo para criar objectos com uma definição comum e, consequentemente, propriedades, operações e comportamento comuns. Um objecto é uma ocorrência de uma classe.

**classe abstracta.** Uma *classe* de programação orientada por objecto que representa um conceito; as classes que dela derivam representam implementações do objecto. Não poderá construir um objecto de uma classe abstracta; isto não pode ser transformado em ocorrência.

**classe de conectores.** A *classe* de programação orientada para objectos que fornece acesso normal às APIs que são nativas a *servidores de conteúdos* específicos.

**classe de índice.** Consulte *tipo de artigo*

**classificação.** Um valor inteiro que representa a importância de uma determinada parte relativamente aos resultados de uma consulta. Um valor mais elevado significa uma correspondência mais exacta.

**classificação de tipo de artigo.** Uma categorização dentro de um *tipo de artigo* que fornece mais identificação aos *artigos* desse tipo de artigo. Todos os artigos do mesmo tipo de artigo têm a mesma classificação de tipo de artigo.

O Content Manager fornece as seguintes classificações de tipo de artigo: *pasta*, *documento*, *objecto*, *vídeo*, *imagem* e *texto*; os utilizadores também podem definir os suas próprias classificações de tipo de artigo.

**cliente de biblioteca.** O componente de um sistema de Content Manager que fornece uma interface de programação de baixo nível para o sistema de bibliotecas. O cliente da biblioteca inclui API que fazem parte do kit do programador de software.

**cliente estreito.** Um cliente tem pouco ou nenhum software instalado mas que tem acesso a software gerido e disponibilizado por servidores de rede a ele ligados. Um cliente estreito constitui uma alternativa a um cliente de plenas funções, como uma estação de trabalho.

**cliente/servidor.** Nas comunicações, o modelo de interacção no processamento de dados distribuído no qual um programa num local envia um pedido para um programa para outro local e aguarda uma resposta. O programa que está a fazer o pedido é chamado de cliente; o programa que pede é chamado de servidor.

**Common Gateway Interface (CGI).** Um standard para o intercâmbio de informações entre um servidor da Web e programas que lhe são externos. Os programas externos podem ser escritos em qualquer linguagem de programação suportada pelo sistema operativo no qual o servidor da Web esteja a ser executado. Consulte *script CGI*.

**componente descendente.** Nível secundário ou inferior opcional de um *tipo de artigo* hierárquico. Cada componente descendente está directamente associado ao nível que acima dele.

**componente raiz.** O primeiro ou único nível de um *tipo de artigo* hierárquico que consiste em *atributos* relacionados definidos pelo sistema e pelo utilizador.

**componente.** Termo genérico para um *componente raiz* ou um *componente derivado*

**conjunto.** Um grupo de objectos com um conjunto semelhante de regras de gestão.

**conjunto de privilégios.** Um conjunto de *privilégios* para trabalhar com componentes e funções do sistema. O administrador atribui conjuntos de privilégios a utilizadores (IDs de utilizador) e *grupos de utilizadores*.

**construtor.** Em linguagens de programação, consiste num método com o mesmo nome que a classe e que é utilizado para criar e inicializar objectos dessa classe.

**consulta por conteúdo de imagem (QBIC).** Uma tecnologia de consultas que permite pesquisas baseadas em conteúdo visual, denominadas funções, em vez de texto simples. Utilizando a QBIC, pode procurar objectos com base nas suas características visuais, como, por exemplo, cor e textura.

**controlo de acesso.** O processo que garante que determinadas funções e *objectos* podem ser acedidos apenas por utilizadores autorizados de forma autorizada.

**correlação de utilizadores.** Associar palavras-passe e IDs de utilizador do Enterprise Information Portal às palavras-passe e aos IDs de utilizador correspondentes num ou mais servidores de conteúdos. A correlação de utilizador permite o início de sessão única no Enterprise Information Portal e *servidores de conteúdos* múltiplos.

**critérios de pesquisa.** No Enterprise Information Portal, campos específicos que um administrador define para um *modelo de pesquisa* que limita e define as opções que estão disponíveis para os *utilizadores*.

**cursor.** Uma estrutura de controlo designada utilizada por um programa de aplicação para apontar para uma linha específica num conjunto ordenado de linhas. O cursor é utilizado para obter linhas a partir do conjunto.

## D

**dados de sequência.** Quaisquer dados que são enviados através de uma ligação em rede a uma velocidade especificada. Uma corrente pode ser um tipo de dados ou uma combinação de tipos. As velocidades dos dados, expressas em bits por segundo, variam conforme os diferentes tipos de sequências e redes.

**DDO.** Consulte *objecto de dados dinâmicos*.

**definição de tipo de documento (DTD).** As regras que especificam a estrutura para uma classe específica de documentos XML. A DTD define a estrutura com elementos, atributos e notações, e estabelece as restrições para como cada elemento, atributo e notação podem ser utilizados dentro de uma classe específica de documentos. Uma DTD é análoga a um esquema de base de dados, já que a DTD descrever completamente a estrutura para um idioma de marca específica.

**definição do servidor.** As características de um *servidor de conteúdos* específico que o identifica, de forma única, no Enterprise Information Portal.

**definição do tipo de servidor.** A lista de características, de acordo com o que foi definido pelo administrador, necessárias para identificar de modo exclusivo um servidor personalizado de um determinado tipo perante o Enterprise Information Portal.

**documento.** Um *artigo* que pode ser armazenado, obtido e trocado entre utilizadores e sistemas do Content Manager como uma unidade independente. Presume-se que um artigo com o *tipo semântico* documento contenha informações que formem um documento, mas isto não significa necessariamente que se trate de uma implementação do modelo de documento do Content Manager.

Um artigo criado a partir de um tipo de artigo classificado de documento (uma implementação específica do modelo de documento do Content Manager), tem de conter partes de documento. Pode utilizar tipos de artigo classificado de documento para criar artigos com o tipo semântico documento ou pasta.

As partes de documento podem incluir vários tipos de conteúdos, por exemplo, texto, imagens e folhas de cálculo.

**DTD.** Consulte *definição de tipo de documento*

## E

**entidade federada.** Um *objecto* de metadados do Enterprise Information Portal composto por *atributos federados* e opcionalmente associados a um ou mais *índices de texto federado*.

**entidade nativa.** Um *objecto* que é gerido num *servidor de conteúdos* específico e que é composto por *atributos nativos*. Por exemplo, as *classes de índice* do Content Manager são entidades nativas compostas por *campos-chave* do Content Manager.

**estado do fluxo de trabalho.** O estado de um *fluxo de trabalho*.

**estado do trabalho.** O estado de um determinado *artigo de trabalho, documento, ou pasta*.

**extended data object (XDO).** Num programa de aplicação, uma representação genérica de um *objecto* de multimédia complexo armazenado que é utilizado para mover esse *objecto* para dentro e para fora da memória. Os XDOs são, a maior parte das vezes, contidos dentro de *DDOs*.

**Extensible Markup Language (XML).** Uma metalinguagem padrão para definir idiomas de marca que são derivados de, e são um sub-conjunto de, SGML. O XML omite as partes mais complexas e menos utilizadas de SGML e torna mais fácil escrever aplicações para processar tipos de documentos, autor e gerir informações estruturadas, e transmitir e partilhar informações estruturadas através de diversos sistemas. A utilização de XML não requer as aplicações e processamento robusto que são necessários para SGML. XML está a ser desenvolvido sob as normas do World Wide Web Consortium (W3C).

## F

**ficheiro de intercâmbio comum (CIF).** Um ficheiro que contém uma sequência de dados ImagePlus Interchange Architecture (IPIA).

**ficheiro de tabelas de rede.** Um ficheiro de texto que contém as informações de configuração específicas do sistema para cada nó num sistema do Content Manager. Cada nó do sistema tem de ter um ficheiro de tabelas de rede que identifique o nó e enuncie os nós com os quais tem de estabelecer ligação.

O nome da tabela de rede é FRNOLINT.TBL.

**ficheiro README.** Um ficheiro que deve ser visualizado antes do programa a ele associada ser instalado ou executado. Um ficheiro README normalmente contém informações de último minuto sobre o produto, informações sobre a instalação ou sugestões sobre como utilizar o produto.

**fluxo de trabalho.** Em Enterprise Information Portal, uma sequência de *passos de trabalho* e das regras que os governam, ao longo dos quais passa um *pacote de trabalho* um *documento* ou uma *pasta* enquanto está a ser processado/a.

Por exemplo, a aprovação de indemnizações descreveria o processo que uma indemnização de seguros individual deve seguir para aprovação.

**formato de dados.** Consulte *tipo de MIME*

**função.** As informações de conteúdo visual armazenadas no servidor de pesquisa de imagens. Do mesmo modo, as características visuais que as aplicações de pesquisa de imagens utilizam para determinar correspondências. As quatro funções QBIC são: cor média, cor do histograma, cor posicional e textura.

## H

**HTML.** Consulte *Hypertext Markup Language*.

**Hypertext Markup Language (HTML).** Uma linguagem de marcação que está em conformidade com a norma SGML e que foi concebida principalmente para suportar a apresentação online de informações textuais e gráficas que incluem ligações de hipertexto.

## I

**identificador persistente (PID).** Um identificador que identifica de modo único um *objecto*, independentemente do local onde está armazenado. O PID é composto por um ID do artigo e por uma localização.

**Image Object Content Architecture (IOCA).** Uma recolha de construções utilizadas para intercâmbio e apresentação de imagens.

**índice.** Para adicionar ou editar os valores de atributo que identificam um *artigo* ou um *objecto* específico, de forma a este poder ser posteriormente recuperado.

**índice de texto federado.** Um *objecto* de metadados do Enterprise Information Portal que é definido para um ou mais *índices de texto nativo* num ou em mais *servidores de conteúdos*.

**índice de texto nativo.** Um índice dos *artigos* de *objectos* que são geridos num *servidor de conteúdos* específico. Por exemplo, um índice de pesquisa de texto simples num servidor de conteúdos do Content Manager.

**information mining.** O processo automático de extracção de informações-chave de texto(resumo), localizando temas predominantes num conjunto de documentos (categorização), e pesquisando documentos relevantes utilizando consultas poderosas e flexíveis.

**intercâmbio.** A capacidade de importar ou exportar uma imagem com o seu índice a partir de um sistema Content Manager ImagePlus for OS/390 para outro sistema ImagePlus utilizando um *ficheiro de intercâmbio comum* ou uma *unidade de intercâmbio comum*.

**interface de programação de aplicações (API).** Uma interface de software que permite que as aplicações comuniquem entre si. Uma API é o conjunto de

instruções em linguagem de programação que podem ser codificadas num programa de aplicação para obter as funções e os serviços específicos fornecidos pelo programa licenciado subjacente.

**inventário do servidor.** A lista global de *entidades nativas* e *atributos nativos* a partir dos *servidores de conteúdos* especificados.

**IOCA.** Consulte *Image Object Content Architecture*.

## J

**JavaBeans.** Uma tecnologia de componente de software, independente de plataformas, para construir componentes de Java reutilizáveis chamados “beans.” Depois de serem construídos, esses beans pode ficar disponíveis para serem utilizados por outros engenheiros de software ou podem ser utilizados em aplicações de Java.Utilizando o JavaBeans, os engenheiros de software podem manipular e reunir beans num ambiente gráfico de programação de arrastar-e-largar.

**Joint Photographic Experts Group (JPEG).** (1) Um grupo que trabalho para estabelecer a norma para a compressão de imagens digitalizadas de tom contínuo. (2) A norma para imagens paradas desenvolvidas por este grupo.

**JPEG.** Consulte *Joint Photographic Experts Group*.

## L

**LAN.** Consulte *rede de área local*.

**libertar.** Remover critérios de suspensão de um *artigo*. Um artigo suspenso é libertado quando os critérios tiverem sido satisfeitos ou quando um utilizador com autoridade adequada substituir os critérios e o libertar manualmente.

**ligação.** Uma relação direccional entre dois *artigos*: a origem e o destino. Poderá utilizar um conjunto de ligações para modelar demasiadas associações. Contrasta com *referência*

**lista de acções.** Uma lista aprovada das acções, definidas por um administrador de sistemas ou outro *coordenador de fluxo de trabalho*, que um utilizador possa efectuar num processo de *fluxo de trabalho* ou de encaminhamento de documentos.

**lista de controlo de acesso.** Uma lista constituída por um ou mais IDs de utilizador ou grupos de utilizador e *privilegios* associados. Pode utilizar as listas de controlo de acesso para controlar o acesso dos utilizadores para *modelos de pesquisa* no sistema do Enterprise Information Portal.



**lista de trabalhos.** Uma recolha de *artigos de trabalho*, *documentos* ou *pastas* que estão atribuídos a um utilizador.

## M

**memória cache.** Uma memória tampão com propósitos especiais, mais pequena e mais rápida do que a memória principal, utilizada para manter uma cópia dos dados que pode ser acedida frequentemente. A utilização da memória cache reduz o tempo de acesso, mas poderá aumentar os requisitos de memória.

**memória cache do gestor de recursos.** A área de armazenamento de trabalho para o *gestor de recursos*. Também denominada *área de transferência ascendente*.

**memória cache do servidor de objectos.** Consulte memória cache do *gestor de recursos*.

**método.** Em concepção ou programa de Java, o software que implementa o comportamento especificado por uma operação. Sinónimo de uma função de membro no ++.

**modelo de pesquisa.** Um formato, composto por *critérios de pesquisa* concebido por um administrador para um determinado tipo de pesquisa associada. O administrador também identifica os *utilizadores* e os *grupos de utilizador* que podem aceder a cada modelo de pesquisa.

**multimédia.** Combinar elementos de suporte de dados diferentes (texto, gráficos, áudio, imagem parada, vídeo, animação) para visualizar e controlar a partir de um computador.

**Multipurpose Internet Mail Extensions (MIME) .** Consulte *tipo de MIME*

## O

**Object Linking and Embedding (OLE).** Uma especificação da Microsoft para ligar e incorporar aplicações para que possam ser activadas a partir de outras aplicações.

**objecto.** Qualquer conteúdo digital que um utilizador pode armazenar, recuperar e manipular como uma unidade única, por exemplo, imagens de *JPEG*, de áudio *MP3*, vídeo *AVI* e um bloco de texto a partir de um livro.

**objecto binário grande (BLOB).** Uma sequência de bytes com um tamanho variável entre 0 bytes e 2 gigabytes. Esta cadeia não tem associada uma página de códigos nem um conjunto de caracteres. As imagens, os objectos de áudio e de vídeo são armazenados em BLOBs.

**objecto de dados dinâmico (DDO).** Num programa de aplicações, uma representação genérica de um

objecto armazenado utilizado para mover esse objecto para dentro e para fora do armazenamento.

**OLE.** Consulte *Object Linking and Embedding*.

## P

**pacote.** Uma recolha de *classes* relacionadas e interfaces que fornecem protecção de acesso e gestão de espaço de nome.

**pacote de trabalho.** Na Versão 7.1 do Enterprise Information Portal, uma recolha de *documentos* que é encaminhada de uma localização para outra. Os utilizadores acedem e trabalham com pacotes de trabalho através de *listas de trabalho*.

**parâmetro identificador.** Uma cadeia de caracteres que representa um objecto e é utilizada para obter o objecto.

**parte.** Consulte *objecto*.

**passo de trabalho.** Um ponto discreto do *fluxo de trabalho* ou do *processo de encaminhamento de documentos* pelo qual um *artigo de trabalho*, *documento* ou *pasta* deve passar.

**pasta.** Um *artigo* de qualquer *tipo de artigo*, independentemente da classificação, com o *tipo semântico* pasta. Qualquer artigo com o tipo semântico pasta possui funcionalidade específica de pasta fornecida por Content Manager, para além de todas as capacidades de artigo não-recurso e quaisquer funcionalidades adicionais disponíveis numa classificação de tipo de artigo, como por exemplo *documento* ou artigo de recurso. As pastas podem conter um número qualquer de artigos de qualquer tipo, incluindo documentos e sub-pastas. Uma pasta é indexada por *atributos*

**pesquisa associada.** Uma consulta emitida a partir do Enterprise Information Portal que pesquisa simultaneamente dados num ou mais *servidores de conteúdos* que podem ser heterogéneos.

**pesquisa combinada.** Uma consulta que combina um ou mais dos seguintes tipos de pesquisas: *paramétrica*, texto ou imagens.

**pesquisa paramétrica.** Uma consulta de *objectos* baseada nas *propriedades* dos objectos.

**PID.** Consulte *identificador persistente*.

**porta de ligação.** Uma unidade funcional que interliga duas redes de computadores com arquitecturas de rede diferentes. Uma porta de ligação liga redes ou sistema de diferentes arquitecturas. U ma ponte interliga redes ou sistema com a mesma arquitectura ou com arquitecturas semelhantes.

**privilégio.** O direito de aceder a um *objecto* específico de um modo específico. Os privilégios incluem direitos como criação, eliminação e selecção de *objectos* armazenados no sistema. Os privilégios são atribuídos pelo administrador.

**propriedade.** Uma característica de um *objecto* que o descreve. Uma propriedade pode ser alterada ou modificada. O estilo de escrita é um exemplo de uma propriedade.

## Q

**QBIC.** Consulte *consulta por conteúdo de imagem*.

## R

**recolha associada.** Um agrupamento de *objectos* que resulta de uma *pesquisa associada*.

**rede de área local (LAN).** Uma rede na qual um conjunto de dispositivos estão ligados uns aos outros para comunicar e que podem ser ligados a uma rede maior.

**referência.** Associação um-a-um de direcção única entre uma raiz ou um *componente descendente* e outro *componente raiz*. Contrasta com *ligação*.

**registo do histórico.** Um ficheiro que mantém um registo das actividades de um *fluxo de trabalho*.

**Remote Method Invocation (RMI).** Um conjunto de APIs que permite programação distribuída. Um *objecto* de uma Máquina Virtual Java (JVM) pode invocar métodos em *objectos* de outras JVMs.

**reproduzir.** Tomar dados que não são normalmente orientados para imagens apresentá-los como uma imagem. No Content Manager, os documentos de processamento de texto podem ser convertidos em imagens para fins de apresentação.

**Resource Interchange File Format (RIFF) .** Utilizado para armazenar som e gráficos para serem reproduzidos em diferentes tipos de equipamento informático.

**RIFF.** Consulte *Resource Interchange File Format*.

**rotina de saída de utilizador.** Uma rotina escrita pelo utilizador que recebe controlo em *saídas de utilizador* predefinidas.

## S

**saída do utilizador.** Um ponto num programa fornecido pela IBM em que é possível dar o controlo a uma rotina de saída de utilizador.

**Script CGI.** Um programa de computador executado num servidor da Web, que utiliza a *Common Gateway Interface (CGI)* para efectuar tarefas que normalmente não são efectuadas por um servidor da Web (por exemplo, acesso a bases de dados e processamento de formulários). Um script CGI é um programa CGI escrito numa linguagem de scripts como, por exemplo, Perl.

**servidor de bibliotecas.** O componente de um sistema do Content Manager que armazena, gere e processa consultas em *artigos*.

**servidor de conteúdos.** Um sistema de software que armazena dados multimédia e comerciais e os respectivos metadados necessários para os utilizadores trabalharem com esses dados. O Content Manager e o Content Manager ImagePlus for OS/390 são exemplos de servidores de conteúdos.

**servidor de objectos.** Consulte *gestor de recursos*.

**servidor de suportes.** Um componente do sistema Content Manager baseado no AIX utilizado para armazenar e aceder a ficheiros de vídeo.

**Servidor RMI.** Um servidor que implementa o modelo de *objecto* distribuído Java *Remote Method Invocation (RMI)*.

**sistema de ficheiros.** No AIX, o método de criação de partições para memória num disco rígido.

**sistema de ficheiros multimédia.** Um sistema de ficheiros que é optimizado para o armazenamento e entrega de vídeo e áudio.

**sobreposição.** Um conjunto de dados predefinidos como, por exemplo, linhas, sombreado, texto, caixas ou logotipos que podem ser intercalados com os dados da variável numa página durante a impressão.

**subclasse.** Uma *classe* derivada de outra classe. Pode existir uma ou mais classes entre a classe e a subclasse.

**subconjunto de classe de índices remissivos.** No Content Manager anterior, uma vista de uma *classe de índice* que uma aplicação utiliza para armazenar, recuperar e visualizar pastas e *objectos*.

**super-classe.** Uma *classe* a partir da qual uma classe é derivada. Uma ou mais classes podem estar entre a classe e a super-classe.

**suspender.** Remover um *objecto* do *fluxo de trabalho* e definir os critérios de suspensão necessários para o activar. A activação posterior do *objecto* permite que este possa continuar a processar.



## T

**tipo de artigo.** Um modelo para definir e mais tarde localizar como *artigos*, consiste num *componente raiz*, zero ou mais *componentes descendentes*, e uma classificação.

**Tipo de MIME.** Uma norma da Internet para identificar o tipo de objecto que está a ser transferido através da Internet. Os tipos de MIME incluem várias variantes de áudio, imagem e vídeo. Cada objecto tem um tipo de MIME.

**tipo semântico.** A utilização ou as regras para um *artigo*. Base, anotação e nota são tipos semânticos fornecidos pelo Content Manager; os utilizadores também podem definir os seus próprios tipos semânticos.

**Tivoli Storage Manager (TSM).** Um produto *cliente/servidor* que fornece serviços de gestão da memória e de acesso de dados num ambiente heterogéneo. Suporta vários métodos de comunicação, faculta funções administrativas para a gestão da cópia de segurança e armazenamento de ficheiros e faculta funções para programação de operações de cópia de segurança.

**transferência ascendente.** O processo de mover um *objecto* armazenado a partir de um dispositivo off-line ou de baixa prioridade de volta para um dispositivo online ou de prioridade mais elevada, normalmente a pedido do sistema ou a pedido de um utilizador. Quando um utilizador pede um objecto armazenado na memória permanente, é guardada uma cópia de trabalho na *área de transferência ascendente*.

**TSM.** Consulte *Tivoli Storage Manager*.

## U

**unidade de intercâmbio comum (CIU).** A unidade de transferência independente de um ficheiro de intercâmbio comum (CIF). Esta é a parte do CIF que identifica a relação para a base de dados de recepção. Um CIF pode incluir várias CIUs.

**uniform resource locator (URL).** Uma sequência de caracteres que representa os recursos de informações num computador ou numa rede, tais como a Internet. Esta sequência de caracteres inclui o nome abreviado do protocolo utilizado para aceder aos recursos de informações e às informações utilizadas pelo protocolo para localizar o recurso da informação. Por exemplo, no contexto da Internet, estes são nomes abreviados de alguns protocolos utilizados para aceder a vários recursos de informações: http, ftp, gopher, telnet e news.

**utilizador.** No Enterprise Information Portal, qualquer utilizador que é identificado no programa de administração do Enterprise Information Portal.

## V

**vista de classes de índice.** No Content Manager anterior, o termo utilizado nas APIS para *sub-conjunto de classes de índices*.

**volume.** Uma representação de um dispositivo ou unidade de memória física na qual os objectos do sistema são armazenados.

**Volume TSM.** Uma área lógica de memória gerida pelo *Tivoli Storage Manager*.

## X

**XDO.** Consulte *objecto de dados expandidos*.

**XML.** Consulte *Extensible Markup Language*.



---

# Índice Remissivo

## A

acções, definir 104  
acessibilidade 135  
armazenar metadados  
    utilizar o Information mining 41

## C

Carregar dados exemplo 117  
carregar documentos de pesquisa de  
    texto e imagem 119  
catálogo  
    adicionar 64  
    adicionar documentos de  
        formação 65  
    avaliar 68  
    eliminar 64  
    formação 72  
    mudar o nome 64  
cliente de administração  
    criar  
        critérios de pesquisa 27  
        fluxo de trabalho 105  
        modelos de pesquisa 26  
    definir  
        acções 104  
        lista de acções 104  
        listas de trabalhos 104  
cmbcc2mime.ini 15  
componentes do EIP  
    administração 4  
    centro de informações 6  
    compatibilidade de sistema  
        operativo 3, 4  
    conectores 4  
    information mining 5  
    pesquisa de imagens 5  
    pesquisa de texto 5  
    visualizador de conteúdos 5  
    Web Crawler 5  
conectores 4  
configurar pesquisa de imagens 114  
conjunto de privilégios 29, 34  
    criar 33  
    mover domínios 40  
conjunto de privilégios de concessão 34  
construtor de fluxo de trabalho  
    criar um fluxo de trabalho 105  
    descrição 102  
    iniciar 105  
critérios de pesquisa  
    definir e fazer correspondências 27

## D

domínio 39  
domínio administrativo 35  
domínios  
    compreender 36  
    criar 36

domínios (*continuação*)  
    privilégios de sub-administrador 37  
    privilégios de super-  
        administrador 37

## E

EIP  
    cliente de pesquisa de imagens 5  
    cliente de pesquisa de texto 5  
    cliente visualizador de conteúdos 5  
    componente de administração 4  
    componente de centro de  
        informações 6  
    conectores 4  
    opção information mining 5  
    opção Web Crawler 5  
    toolkit de conector 6  
Enterprise Information Portal  
    criar  
        critérios de pesquisa 27  
        fluxo de trabalho 105  
        modelos de pesquisa 26  
    definir  
        acções 104  
        lista de acções 104  
        listas de trabalhos 104

## F

ficheiro de tipo MIME  
    alterar para servidores 15  
fluxo de trabalho  
    conceitos 97  
    criar 105  
    planear 100  
função fluxo de trabalho  
    componentes 102  
    configurar 97

## G

gestor de recursos  
    atribuir a um domínio 38  
    atribuir utilizadores a 34  
gestor de recursos, mover domínios 39  
grupo de privilégios 34  
grupo de utilizadores 35  
    mover domínios 39

## I

IBM Enterprise Information Portal for  
    Multiplatforms  
        componentes 3  
ID de utilizador 29  
incapacidade 135  
Information mining  
    componentes de 42

Information mining (*continuação*)  
    construir uma taxonomia 62  
    descrição 41  
    formatos de documentos  
        suportados 49  
    grupo alvo 44  
    idiomas suportados 49  
    serviço 41  
    trabalhar num ambiente  
        empresarial 44  
    um exemplo 46  
Information Structuring Tool  
    avaliar uma taxonomia 68  
    como começar 63  
    definir uma taxonomia 64  
    descrição 62  
    efectuar a formação de uma  
        taxonomia 72  
    instalar 62  
    mecanismo de bloqueio 63  
    seleccionar documentos de  
        formação 65  
    utilizar WAS 62  
iniciar  
    construtor de fluxo de trabalho 105  
inventário do servidor 17

## L

LDAP  
    configurar 32  
    importar 32  
lista de acções  
    acções predefinidas 104  
    definir 101, 104  
lista de controlo de acesso  
    mover domínios 40  
lista de trabalhos  
    definir 104  
    descrição 101

## M

modelos de pesquisa, criar 26

## O

opção pesquisa de imagem 5  
opção visualizador de conteúdos 5

## P

pacote de trabalhos, descrição 101  
personalizar tipos MIME 15  
pesquisa de imagens  
    atribuir um nome alternativo 116  
    configurar 114  
    verificar a ligação 116

- pesquisa de texto
  - configurar 113
  - suporte de XML 113
- planejar
  - Enterprise Information Portal 2
- programa de carregamento exemplo,
  - executar 119

## R

- recolha
  - atribuir a um domínio 38
- recolhas
  - mover domínios 39

## S

- servidor de conteúdos
  - definir 17

## T

- taxonomia
  - utilizar a Information Structuring Tool 62
- teclado 135

## U

- utilizador 29
  - conjunto de privilégios 34
  - mover domínios 38

## W

- Web Crawler
  - opção EIP 5





Número do Programa: 5724-B43

SC17-5413-01

