

IBM® Enterprise Information Portal
for Multiplatforms



Managing Enterprise Information Portal

Version 7.1

IBM® Enterprise Information Portal for
Multiplatforms



Managing Enterprise Information Portal

Version 7.1

Note

Before using this information and the product it supports, read the information in "Notices" on page 81.

First Edition (August 2000)

This edition applies to Version 7.1 of IBM Enterprise Information Portal and IBM Enterprise Information Portal Client Kit (product numbers 5697-G29, 5697-G31) and to all subsequent releases and modifications until otherwise indicated in new editions. This edition replaces GC26-9833-00.

Portions of this product are: Copyright © 1990-2000 ActionPoint, Inc. and/or its licensors, 1299 Parkmoor Drive, San Jose, CA 95126 U.S.A. All rights reserved.

Outside In® Viewer Technology © 1992-2000 Inso Corporation. All rights reserved.

© **Copyright International Business Machines Corporation 1999, 2000. All rights reserved.**

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

About this guide	v
Who should use this guide	v
Skills that are required for Enterprise Information Portal administrators	vi
Skills that are required for business analysts or process modeler	vi
Where to find more information	vi
Information included in your product package.	vi
Support available on the Web	vii
How to send your comments	vii

Part 1. Tasks for Enterprise Information Portal administrators . . . 1

Chapter 1. Introducing Enterprise Information Portal	3
Searching for customer information.	3
The need	3
The solution	4
The Enterprise Information Portal solution	4
Enterprise Information Portal components	6
What's new in Version 7.1	8

Chapter 2. Getting started with Enterprise Information Portal	9
Starting the Enterprise Information Portal administration client.	9
Defining the business problem	9
Analyzing the business problem	10
Defining content servers and running the server inventory	10
Creating objects for a search template	12
Creating federated entities	12
Creating and mapping federated attributes.	12
Creating federated text indexes.	17
Creating search templates	17
Associating objects with the search template	18
Defining and mapping search criteria.	18
Managing access.	23
Creating privilege sets.	24
Creating users and user groups.	24
Creating access control lists	25
Optimizing the administration client	26

Chapter 3. Managing information mining	27
Configuring and starting Enterprise Information Portal Information Mining	27
Defining category schemes	28
Guidelines for defining categories	28
Providing the training information	29
Running the training wizard.	29
Importing documents from a webspace	30

Introducing the IBM Web Crawler.	30
The CMBWebCrawlerService JavaBean	30
Step 1: Setting environment variables.	31
Step 2: Starting the Control Service	32
Step 3: Defining a webspace.	32
Step 4: Crawling a webspace	34
Step 5: Importing the crawled documents	37

Chapter 4. Introducing Enterprise Information Portal workflow 39

Understanding workflow.	39
How to use workflow	40
Planning a workflow	40
Information to be processed	41
How information is handled.	41
Actions to be taken.	41
How information flows through the process	42
How everything fits together	42
Using Enterprise Information Portal workflow components	42
Workflow builder	42
Enterprise Information Portal workflow services	43
Defining worklists	43
Defining action lists	44
Predefined actions	44
Creating a workflow	45
Starting Enterprise Information Portal workflow builder	45
Starting the MQSeries Workflow server	45

Part 2. Tasks for Enterprise Information Portal users 47

Chapter 5. Getting started with Enterprise Information Portal clients . . . 49

Chapter 6. Defining document types to Enterprise Information Portal	51
Changing the server MIME type file (cmbcc2mime)	51
Changing the client MIME type file (cmbmime2app)	52

Part 3. Content Manager features 53

Chapter 7. Adding text search to your Content Manager system 55

Planning for text search	55
Planning for text search capacity	56
Planning for text search performance.	57
Prerequisites for text search	57
Installing text search	57
Installing text search on Windows.	58
Installing text search on AIX.	59

Manually configuring additional text search components	60
Setting up to administer a text search server	60
Updating the shortcut for the Content Manager system administration client	61

Chapter 8. Adding image search to your Content Manager system 63

Installing image search	63
Installing image search on Windows	63
Installing image search on AIX	65
Setting up image search	66
Setting up the environment	67
Configuring the image search server	68
Configuring the image search client	68

Chapter 9. Loading and indexing sample data 71

Before you load the sample data	71
Creating a text search index	71
Creating the image search database, catalog, and features.	72
Running the loader program	73
Indexing the sample text data	74

Part 4. Appendixes. 75

Appendix. Troubleshooting 77

Notices	81
Trademarks	83

Glossary 85

Index 91

About this guide

This guide describes how to manage, use, and maintain IBM Enterprise Information Portal. It also provides scenarios to help you understand and apply Enterprise Information Portal concepts and tasks. This guide includes three parts:

Part 1. Tasks for Enterprise Information Portal administrators

This section describes what an administrator needs to do to manage an Enterprise Information Portal system in your enterprise.

Part 2. Tasks for Enterprise Information Portal users

This section describes tasks for your users.

Part 3. Content Manager features

Text search and image search are optional Content Manager features that can be configured and run with the Content Manager servers only. This section describes what you need to do to configure these Content Manager features.

Who should use this guide

If you are an administrator responsible for document processing, this book helps you understand how to manage the document processing within your enterprise. Using Enterprise Information Portal, you set up and administer the searches and information mining that your users can perform across disparate content servers, including the Web. You also can set up, define, and manage the workflow processes of your business.

If you are a business analyst, this book helps you understand how to use the Enterprise Information Portal workflow feature to define and model your business or workflow processes.

Enterprise Information Portal administrator

Enterprise Information Portal administrators must be familiar with the content servers and business goals of their enterprise. By using the Enterprise Information Portal administration client, the administrator creates and manages the search templates used to search the content servers within the enterprise. The administrator also uses the Enterprise Information Portal administration client and the workflow feature to control the workflow processes of your enterprise. For Information Mining, the administrator uses a categorization tool to define taxonomies of documents. Users then use these taxonomies to perform searches on documents located on content servers. The administrator can also set up the IBM Web Crawler to search and import documents from an intranet or the Web.

Business analyst or process modeler

A business analyst or process modeler is responsible for defining, building, and testing the work processes of your enterprise. The business analyst or process modeler must understand the staff requirements, programs, and data structures used in the processes that they are modeling. A business analyst uses the Enterprise Information Portal workflow builder to model workflow processes and test the assumptions behind the processes.

Skills that are required for Enterprise Information Portal administrators

To manage and administer Enterprise Information Portal, you must be able to:

- Make basic administration decisions such as security assignments.
- Know how to administer Windows and AIX platforms.
- Perform database administration tasks.
- Perform network administration tasks.
- Group content server characteristics such as documents and folders.
- Apply a working knowledge of content and search criteria when creating search templates.
- Understand the schema of content servers in order to define useful search templates.
- Understand what kind of content your users want to find.
- Apply an understanding of information mining capabilities; these capabilities include extracting key information from documents and grouping documents into categories.
- Apply an understanding of your business processes when administering workflow decisions.

Skills that are required for business analysts or process modeler

To use Enterprise Information Portal workflow builder, you must be able to:

- Understand staff requirements, programs, and data structures used in the business processes of your enterprise.
- Make decisions regarding the business or workflow processes of your enterprise.

Where to find more information

IBM Enterprise Information Portal includes a complete set of information to help you plan for, install, administer, and use your system. Product documentation and support are also available on the Web.

Information included in your product package

The IBM Enterprise Information Portal publication library is available online in HTML (Hypertext Markup Language) format. On Windows operating systems, you can view the publications from the taskbar by clicking **Start** → **Programs** → **IBM Enterprise Information Portal for Multiplatforms 7.1** → **Information** and then the title of the publication that you want to view.

The *IBM Enterprise Information Portal* CD-ROM contains each publication in the Adobe Acrobat portable document format (PDF).

You can view the PDF files online using the provided Adobe Acrobat Reader for the following operating systems: Windows® or AIX®. (Adobe provides Acrobat Readers for additional operating systems on their Web site, <http://www.adobe.com>.)

Table 1 on page vii shows the Enterprise Information Portal publications included on the *IBM Enterprise Information Portal* CD-ROM.

Table 1. Enterprise Information Portal publications

File name	Title	Publication number
apgoo	<i>Application Programming Guide</i>	SC27-0877-00
eipinst	<i>Planning and Installing Enterprise Information Portal¹</i>	GC27-0873-00
eipmc	<i>Messages and Codes</i>	SC27-0874-00
eipmng	<i>Managing Enterprise Information Portal</i>	SC27-0875-00
txtse	<i>Text Search Engine Application Programming Reference</i>	SC27-0876-00

Notes:

1. You receive a printed copy of the *Planning and Installing Enterprise Information Portal* with the product.

Copying the PDF files: To copy the PDF files from the CD-ROM to your hard drive:

1. Change to the directory for the language that you are using (for example, ENU for English).
2. Copy *.PDF files to your designated hard drive directory

Installing the PDF reader: The Adobe Acrobat Reader is available in the ACROBAT directory of the CD-ROM and from <http://www.adobe.com>. To install Acrobat Reader, decompress the program files for your platform and follow the instructions in the Acrobat installation program or the installation text file.

On AIX, untar the .tar file and read INSTGUID.TXT.

On Windows, run the executable file.

Support available on the Web

Product support is available on the Web. Click **Support** from the product Web site at:

<http://www.ibm.com/software/data/eip/>

The documentation is included in softcopy on the product CD-ROM. To access product documentation on the Web, click **Library** on the product Web site.

An HTML-based documentation interface, called Enterprise Documentation Online (EDO), is also available from the Web. It currently contains the reference information about the object-oriented and Internet APIs. Go to the Enterprise Information Portal Library Web page for information about accessing EDO.

How to send your comments

Your feedback helps IBM to provide quality information. Please send any comments that you have about this publication or other Content Manager and Enterprise Information Portal documentation. You can use either of the following methods to provide comments:

- Send your comments from the Web. Visit the IBM Data Management Online Reader's Comment Form (RCF) page at:
<http://www.ibm.com/software/data/rcf>

You can use the page to enter and send comments.

- Send your comments by e-mail to comments@vnet.ibm.com. Be sure to include the name of the product, the version number of the product, and the name and part number of the book (if applicable). If you are commenting on specific text, please include the location of the text (for example, a chapter and section title, a table number, a page number, or a help topic title).

Part 1. Tasks for Enterprise Information Portal administrators

Chapter 1. Introducing Enterprise Information

Portal	3
Searching for customer information.	3
The need	3
The solution	4
The Enterprise Information Portal solution	4
Enterprise Information Portal components	6
Administration database	6
Administration client	6
Connectors	6
Enterprise Information Portal sample client application	7
Enterprise Information Portal thin client samples	7
Connector toolkits and samples	7
Information mining	7
Workflow	7
Content Manager text search server and client	8
Content Manager image search server and client	8
What's new in Version 7.1	8

Chapter 2. Getting started with Enterprise Information Portal

Starting the Enterprise Information Portal administration client.	9
Defining the business problem	9
Analyzing the business problem	10
Defining content servers and running the server inventory	10
Creating objects for a search template	12
Creating federated entities	12
Creating and mapping federated attributes	12
Creating federated text indexes.	17
Creating search templates	17
Associating objects with the search template	18
Defining and mapping search criteria.	18
Managing access.	23
Creating privilege sets.	24
Creating users and user groups.	24
Creating access control lists	25
Optimizing the administration client	26

Chapter 3. Managing information mining.

Configuring and starting Enterprise Information Portal Information Mining	27
Defining category schemes	28
Guidelines for defining categories	28
Providing the training information	29
Running the training wizard.	29
Importing documents from a webspace	30
Introducing the IBM Web Crawler.	30
The CMBWebCrawlerService JavaBean	30
Step 1: Setting environment variables.	31
Step 2: Starting the Control Service	32
Step 3: Defining a webspace.	32

Defining a webspace using the imyclean command	33
Defining a webspace using the Webspace Manager in Windows	33
Step 4: Crawling a webspace	34
What happens during a crawl	34
Starting a crawl	34
Interacting with the Web Crawler	35
Monitoring progress	36
Excluding Web Crawler from a server	36
Step 5: Importing the crawled documents	37

Chapter 4. Introducing Enterprise Information Portal workflow

Understanding workflow.	39
How to use workflow	40
Planning a workflow	40
Information to be processed	41
How information is handled.	41
Actions to be taken.	41
How information flows through the process	42
How everything fits together	42
Using Enterprise Information Portal workflow components	42
Workflow builder	42
Enterprise Information Portal workflow services	43
Defining worklists	43
Defining action lists	44
Predefined actions	44
Creating a workflow	45
Starting Enterprise Information Portal workflow builder	45
Starting the MQSeries Workflow server	45

Chapter 1. Introducing Enterprise Information Portal

Many paper-intensive enterprises, such as insurance companies and financial institutions, administer large volumes of business-related content. The need for an enterprise solution for managing and accessing business information spans many industries.

A *content server* is a software system that stores multimedia, business forms, documents, and related data, along with metadata that allows employees to process and work with the content. When there is no way to effectively connect disparate content servers, a business can waste time and money by duplicating information or training employees to perform multiple searches.

Enterprise Information Portal provides leading-edge technology to bring all of your enterprise resources to your workstation desktop. Enterprise Information Portal can help you maximize the value of your information and multimedia assets by connecting disparate content servers through a single client. With an Enterprise Information Portal client users can quickly and concurrently access all connected content servers. Users can also do information mining or “intelligent” searches across content servers, including the Web or an intranet, and they can perform workflow tasks within your business processes.

With Enterprise Information Portal, you can customize applications for your enterprise. Using the Enterprise Information Portal toolkits, application programmers can write both desktop and Web-based applications.

This chapter provides an overview of Enterprise Information Portal. A scenario about a fictitious insurance company, AA Insurance, demonstrates Enterprise Information Portal’s features and functionality.

Searching for customer information

AA Insurance (AA), a large property and casualty insurance company, has an extensive collection of photographs, claims, policies, adjuster’s notes, reports from experts, and other business documents.

AA keeps all memos that are sent to policy holders, along with medical and appraisal electronic forms in Lotus® Domino™.Doc file cabinets. AA archives all policy declarations, notices, and invoices in a Content Manager OnDemand server for long-term storage and quick access. AA stores all claim forms, photographs, and letters received from policy holders in a Content Manager for AS/400® system folder. AA keeps reports from experts in a DB2 Universal Database™ (DB2 UDB) Data Warehouse Center Information Catalog Manager. AA also stores corporate media assets such as high-resolution graphics in a Content Manager system for the advertising, public relations, and new business departments to share. In addition, AA keeps information, such as company procedures, on its company intranet.

The need

Claims, customer calls, and general policy holder servicing cannot be handled with the content from one server because employees need to access all customer information. To provide customer service, employees require simultaneous access to a variety of content servers. AA Insurance needs a solution that connects their

content servers and their company intranet for searching and retrieving information. They also want to expand their use of workflow processing.

Many different employees need to access documents, from clerks to claim adjusters to agents. AA must restrict access to certain items, while providing unlimited access to others. AA also wants an easy-to-use interface to reduce the need for training.

The solution

AA Insurance deploys Enterprise Information Portal because the comprehensive search technologies allow them to connect and search all of their content servers for the retrieval of data. Now, when an AA Call Center representative receives a call, a single federated search retrieves all of the necessary policy holder information.

AA Insurance also uses Enterprise Information Portal Information Mining feature to search for and retrieve information from the company's intranet. They also want to expand their use of workflow processes.

The Enterprise Information Portal solution

Enterprise Information Portal is a comprehensive product; its components work together to provide a solution uniquely suited to your enterprise. Centered on a multiple-tier architecture, Enterprise Information Portal provides an administration client for managing searches, clients (as samples) for running searches, and connectors for connecting to disparate content servers such as Content Manager, Content Manager ImagePlus[®] for OS/390[®], Content Manager OnDemand, Lotus Domino.Doc, DB2 UDB, DB2[®] DataJoiner[®], and DB2 Data Warehouse Center Information Catalog Manager. You can write additional connectors for additional content servers.

Figure 1 on page 5 shows the concept of the multiple-tier architecture of Enterprise Information Portal.

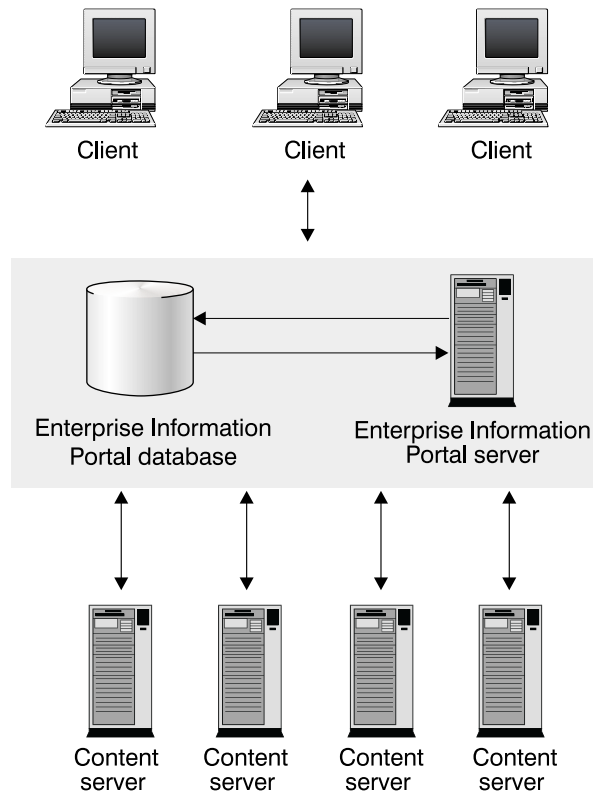


Figure 1. Multiple-tier architecture

The Enterprise Information Portal architecture allows your client applications to run single searches on one or more content servers. To perform searches, a client uses search templates that were defined by the Enterprise Information Portal administrator.

Using search templates, the client runs a *federated search*, a search that runs simultaneously across content servers whose native attributes have been mapped with the federated attributes used in the search template. The Enterprise Information Portal search templates contain search criteria, which reference federated attributes that are mapped to native attributes on each of the content servers. The Enterprise Information Portal administrator creates the search templates. Enterprise Information Portal provides connectors to make compatible the heterogeneous interfaces of content servers. The content servers then return data objects to the client.

The Enterprise Information Portal architecture provides the following advantages:

- Access using a single query to multiple and varying content servers that support e-business™ transactions and customer service applications.
- Information mining capability across multiple content servers, including the Web.
- Workflow process access to data across multiple, heterogeneous content servers.
- Support for the development of client applications that are independent of data's location on any content server, because of the separation of client applications, indexes, and data.

Enterprise Information Portal components

This section describes each Enterprise Information Portal component. These components are shipped as part of the Enterprise Information Portal product.

Administration database

The Enterprise Information Portal database is a DB2 UDB database that stores the information managed by the administration client. This data is used to perform federated searching, information mining, and workflow processes.

Migrating your Enterprise Information Portal Version 6.1 database: You must migrate your data from Enterprise Information Portal Version 6.1 *before* using your Enterprise Information Portal Version 7.1 administration database.

Administration client

The system administrator uses the Enterprise Information Portal administration client to:

- Define each content server for federated searching.
- Identify native entities and attributes on content servers and map them to federated entities.
- Maintain an inventory of the search criteria for all content servers.
- Create search templates.
- Identify and manage users who can access search templates, the information mining feature, and workflow processes.
- Define business workflow processes.
- Administer the information mining server: stopping, starting, and testing the information mining server connection and setting up trace and log levels.

This information is stored in the Enterprise Information Portal database.

It is recommended that you install the administration client on the same workstation, or server, as the Enterprise Information Portal database.

You can also have as many administration clients as you want on other workstations. To do this configuration, you need to do one of the following:

- Have DB2 Client Application Enabler installed and configured on each workstation on which the administration client is installed.
- Use a Remote Method Invocation (RMI) configuration by starting the RMI server on which the Enterprise Information Portal database is installed. You need to make sure that your `\CMBROOT\cmbclient.ini` file points to this server.

Connectors

Connector classes permit client applications to access content servers and the Enterprise Information Portal database. Enterprise Information Portal ships the following connectors for content servers:

- Relational database connectors
- Federated connector (to the Enterprise Information Portal database)
- Content Manager connector
- Content Manager OnDemand connector
- Content Manager ImagePlus for OS/390 connector
- Content Manager for AS/400 connector
- Lotus Domino.Doc connector
- Extended search connector

- DB2 Data Warehouse Manager Information Catalog Manager connector

The federated connector contains the connector class for the Enterprise Information Portal database. Each content server connector contains the appropriate connector class.

Connectors are either local or remote. Local connectors are a set of connector classes you use for directly connecting to various content servers. Local connectors can reside on an Enterprise Information Portal desktop client or on an Enterprise Information Portal RMI server. Remote connectors, which connect to a content server through an RMI server or an RMI server pool member, are always installed.

Enterprise Information Portal sample client application

The sample client application, built by IBM from the JavaBeans™ in the federated access toolkit, provides an interface for performing a single search to retrieve and display documents from multiple content servers.

When a user runs the Enterprise Information Portal sample client application, the search templates available to that user are retrieved from the Enterprise Information Portal database. The user then selects a template and enters values for the search criteria. A list of documents satisfying the search criteria is returned to the user. Documents can then be selected from the search results list, viewed, and updated.

Enterprise Information Portal thin client samples

The Enterprise Information Portal thin client samples run only in a Web browser. Written in HTML and JavaServer Pages™ (JSP), the thin client samples are easily changed and deployed through a Web browser interface. The thin client samples are a set of files that you install on your Web server.

Users use the thin client samples for performing a single search to retrieve and display documents and folders from multiple content servers. Users can also use the thin client samples for performing tasks in a workflow process.

Connector toolkits and samples

Use the connector toolkits and samples to build your own Internet or desktop client applications that access data and content in content servers. The toolkits provide:

- Java®, C++, and ActiveX classes
- Content server-specific samples

Information mining

Information mining provides linguistic services to find hidden information in text documents on content servers. During processing of the text documents, metadata (information about data) is created that can be summarized, categorized, and searched. Enterprise Information Portal provides samples that show how to use information mining capabilities in a thin client. You can build your own desktop client or thin client to work with information mining.

Workflow

With Enterprise Information Portal you can control the flow of work in your business. By using Enterprise Information Portal workflow feature, you can define and run the workflow process of a work group, department, or enterprise. Using a graphical builder, you can construct a comprehensive, easy-to-understand graphical representation of a workflow process in the Enterprise Information Portal

workflow builder. Your users can then use the defined workflow process to perform their tasks, using a client that you develop or the Enterprise Information Portal thin client samples.

Content Manager text search server and client

You can use this feature to automatically index, search, and retrieve documents stored in Content Manager. Users can locate documents by searching for words or phrases

Restriction: The Content Manager text search server and client is an optional Content Manager feature that you can configure and run with Content Manager servers only. If you do not use Content Manager servers, do not install this feature.

Content Manager image search server and client

This feature uses IBM QBIC[®] (query by image content) technology with which you can search for objects by certain visual properties, such as color and texture.

Restriction: The Content Manager image search server and client is an optional Content Manager feature that you can configure and run with Content Manager servers only. If you do not use Content Manager servers, do not install this feature.

What's new in Version 7.1

Enterprise Information Portal Version 7.1 provides unprecedented access to disparate content servers. The new features and components include:

- Improved installation procedures
- Additional connectors for relational databases
Enterprise Information Portal provides relational database connectors for DB2 UDB, DB2 DataJoiner, DB2 Data Warehouse Manager Information Catalog Manager, and other databases through JDBC or ODBC drivers.
- Advanced information mining and search capabilities
Information Mining offers advanced text searching using a flexible query that you can restrict to documents of certain categories.
- Workflow capabilities
By using Enterprise Information Portal workflow feature, you can define and run the workflow process of a work group, department, or enterprise.
- Federated level access control
You can control access to Enterprise Information Portal information mining and workflow processes through the use of privilege sets and access control lists. Additional access control to data can be managed by the access control features of each content server.
- Additional support for Content Manager:
 - List, add, retrieve, update, and delete of content class
 - Asynchronous retrieval of object content

Chapter 2. Getting started with Enterprise Information Portal

IBM Enterprise Information Portal provides an administration client that you use to manage and administer federated searches. This chapter describes concepts and procedures for using the administration client for federated searches, information mining, and workflow processing. Your users can find documents in one content server, images in another, and forms in a third using a single search with one set of criteria. Your users can then use this content for workflow processes.

This chapter uses a scenario to provide an overview of federated searching. For more detailed instructions, see the Enterprise Information Portal online help system. For a scenario and overview of workflow processing, see “Chapter 4. Introducing Enterprise Information Portal workflow” on page 39.

If you used Enterprise Information Portal Version 6.1: If you want to migrate your data from the Enterprise Information Portal Version 6.1 to Enterprise Information Portal Version 7.1, see *Planning and Installing Enterprise Information Portal*.

Starting the Enterprise Information Portal administration client

Requirement for starting the administration client on Windows 95 or Windows 98: You must increase your environment space for the shortcut to 4096 bytes by completing the following steps:

1. Right-click **Start** from the Windows taskbar.
2. Click **Open**.
3. Double-click **Programs**.
4. Double-click **IBM Enterprise Information Portal**.
5. Right-click the **Enterprise Information Portal Administration** icon and click **Properties**.
6. Click the **Memory** tab.
7. Enter 4096 for the number of bytes for environment.

To start and log on to the Enterprise Information Portal administration client:

1. From the Windows taskbar, click **Start** → **Programs** → **Enterprise Information Portal** → **Administration**.
2. In the Enterprise Information Portal Administration Log On window, enter a valid Enterprise Information Portal user ID and password. The default user ID is cmbadmin and the default password is password. You should change the default password at this step.
3. Select the Enterprise Information Portal database that you want to access.
4. Click **OK**.

Defining the business problem

Usually, the decision maker who purchased Enterprise Information Portal has already defined the business problem that you need to use the Enterprise Information Portal to solve. The Enterprise Information Portal administrator must develop a clear definition of the problem and create search templates for solving the business problem.

For example, Pat Kirkland Simmons is an AA Insurance service manager. The vice president chooses her to be the Enterprise Information Portal administrator. To clearly define the business problem, Pat asks the call center representatives about what kind of information they need to find to service customer calls. Most calls received are about claims that have not yet been paid. Call center representatives cannot immediately answer questions about the claim, because they have to search several content servers to find the information they need.

Call center representatives want to enter the policy holder's name, policy number, and the date the claim was filed and receive all the claim-related information.

Analyzing the business problem

When analyzing the business problem, you need to answer three questions:

- What values are users going to enter for the search?
- What objects need to be returned?
- Where are the returned objects stored?

For example, Pat knows the call representatives want to search for objects using the policy holder's name, the policy number, and the date the claim was filed.

After talking to the call representatives, Pat creates the following list of items that call center representatives need to know to effectively process a claim call:

- Policy declarations
- Claims
- Appraisals
- Letters to policy holders
- Photographs
- Reports from experts

Pat then determines on which content servers the items are stored, as shown in Table 2.

Table 2. Pat's list of where the business content is stored

Content server type	Server name	Items
IBM Content Manager	LS1	Photographs
	LS2	Letters to policy holders
IBM Content Manager OnDemand	OD1	Policy declarations
Lotus Domino.Doc	DD1	Appraisals
IBM DB2 Warehouse Manager Information Catalog Manager	IC1	Reports from experts

Defining content servers and running the server inventory

After you have a list of where the needed items are stored, you can use the Enterprise Information Portal administration client to define the content servers. To define a content server to Enterprise Information Portal, you select the type of server and then define the details for that specific server and connection.

When you define the content servers, you can provide additional parameters for the initialization, connection strings, and configuration strings. See *Planning and Installing Enterprise Information Portal* and the Enterprise Information Portal online help for a list of supported strings.

Tip: You can also define a content server type that is not one of the predefined server types, but you must provide the Java or C++ connector classes and the server definition class for the new server type. You also need the Java connector to run the server inventory. For instructions about adding content servers, see the *Application Programming Guide* and the online API reference.

After you define the content servers, you run a server inventory to collect and list the native entities and their characteristics, called native attributes. These native entities and attributes define the objects that are stored on that type of content server. For example, on a Content Manager content server the index classes are the native entities and key fields are the native attributes. These entities and attributes are stored in the Enterprise Information Portal database.

Tip: After you run the server inventory, you should test the connection to each content server.

In our scenario, Pat defines the following content servers:

- LS1
- LS2
- OD1
- DD2
- IC1

Pat then runs the server inventory to collect and list the native entities and attributes of each content server. She also tests the connection to each content server. Pat knows she wants to query using a policy holder's name, policy number, and the date the claim was filed. She lists the native attributes that are related to her search items as shown in Table 3.

Table 3. Pat's list of related native attributes

Content server name	Policy holder name	Policy number	Filing date
LS1	last_name	acct_number	date
LS2	l_name	a_number	DATE
OD1	primary_recipient	pol_number	date_time
DD1	client	number	date
IC1	NAME	—	—

As an Enterprise Information Portal administrator, you must understand the native entities and native attributes of all necessary content servers. Your task is to map the appropriate native attributes to federated attributes, which are used for searching. Table 4 on page 12 lists the supported content servers and their native entities and attributes.

Table 4. Content server native entities and native attributes

Content server type	Native entities	Native attributes
IBM Content Manager	Index class	<ul style="list-style-type: none"> • Attribute • Key attribute
IBM Content Manager for AS/400	Index class	<ul style="list-style-type: none"> • Attribute • Key attribute
IBM DB2 UDB	Table	Column
IBM DB2 DataJoiner	Table	Column
JDBC	Table	Column
ODBC	Table	Column
IBM Content Manager OnDemand	<ul style="list-style-type: none"> • Application group • Search template 	<ul style="list-style-type: none"> • Field • Criteria
Lotus Domino.Doc	<ul style="list-style-type: none"> • Library • Cabinet • Binder 	Field
Domino Extended Search	Database	Field
IBM Content Manager ImagePlus for OS/390	Entity Examples: <ul style="list-style-type: none"> • Document • Folder 	Attribute Examples: <ul style="list-style-type: none"> • FolderId • DocId
DB2 Warehouse Manager Information Catalog Manager	Object	Property

Creating objects for a search template

After you have the server information, you can start building the objects necessary to create the search template. Enterprise Information Portal associates reusable federated entities to the search templates, so the specific query information contained in the search templates can be reused. Federated entities contain federated attributes that are mapped to one or more content server native attributes. It is the mapping of native attributes that makes the disparate content servers available for one query.

Using the Enterprise Information Portal sample client application or thin client samples, users select search templates and provide search criteria to query content servers.

Creating federated entities

You must create a federated entity to associate with the search template. A *federated entity* is an Enterprise Information Portal object that contains federated attributes. You should create a federated entity that describes the information that the federated attributes contain. For example, Pat creates a federated entity named `claim`.

Creating and mapping federated attributes

After you create a federated entity that is used for your federated search, your next task is to map the federated attribute to the appropriate native attribute. A *federated*

attribute is an Enterprise Information Portal metadata category that is mapped to native attributes in one or more content servers. When the search runs, Enterprise Information Portal uses the associated federated attribute to search the native attributes mapped to it from each content server. You should map a federated attribute only to a native attribute with the same properties, such as the ability to be queried or updated. You can also map a federated attribute that is less restrictive than the native attribute.

Requirement: A federated attribute can only be deleted after all mappings using that federated attribute are deleted.

For example, Pat defines the following federated attributes in `claim`:

- `policy_number`
- `policy_holder_name`
- `date`

Pat then maps the federated attributes as listed in Table 5.

Table 5. How Pat maps the federated attributes

Federated attribute	Maps to:
<code>policy_holder_name</code>	<ul style="list-style-type: none"> • <code>last_name</code> • <code>l_name</code> • <code>primary_recipient</code> • <code>client</code> • <code>NAME</code>
<code>policy_number</code>	<ul style="list-style-type: none"> • <code>acct_number</code> • <code>a_number</code> • <code>pol_number</code> • <code>number</code>
<code>date</code>	<ul style="list-style-type: none"> • <code>date</code> • <code>DATE</code> • <code>date_time</code> • <code>date</code>

Table 6 lists all the supported data types by content server. It also includes instructions for mapping.

Table 6. Supported data types by content server

Content server	Supported data types	To define federated attributes, you can use:
Content Manager	<ul style="list-style-type: none"> • Character • Varchar • Date • Time • Timestamp • Decimal • Double • Long • Short 	All of these data types, because they map one-to-one between the Enterprise Information Portal database and the Content Manager server.

Table 6. Supported data types by content server (continued)

Content server	Supported data types	To define federated attributes, you can use:
Content Manager for AS/400	<ul style="list-style-type: none"> • String • Long 	<p>Varchar Map to Content Manager for AS/400 string to represent Strings</p> <p>Date Map to Time, Timestamp, Decimal, and Double</p> <p>Long Map to Content Manager for AS/400 Long to represent Short and Long integers</p>
DB2 UDB (Built-in)	<ul style="list-style-type: none"> • Char • Time • Timestamp • Date • Varchar • Long varchar • Blob • Graphic • Vargraphic • Long vargraphic • Clob • Dblob • Real • Double • Smallint • Integer • Bigint • Decimal 	All of these data types, because they map one-to-one between federated and DB2 UDB datastores.

Table 6. Supported data types by content server (continued)

Content server	Supported data types	To define federated attributes, you can use:
JDBC	<ul style="list-style-type: none"> • Bit • Tinyint • Smallint • Integer • Bigint • Float • Real • Double • Numeric • Decimal • Character • Varchar • Longvarchar • Date • Time • Timestamp • Binary • Varbinary • Longvarbinary • Null • Other 	All of these data types, because they map one-to-one between federated and datastores supported by JDBC.
ODBC	<ul style="list-style-type: none"> • Character • Decimal • Numeric • Bit • Tinyint • Smallint • Integer • Bigint • Real • Float • Double • Binary • Date • Time • Timestamp 	All of these data types, because they map one-to-one between federated and datastores supported by ODBC.

Table 6. Supported data types by content server (continued)

Content server	Supported data types	To define federated attributes, you can use:
DB2 DataJoiner	<ul style="list-style-type: none"> • Character • Varchar • Date • Time • Timestamp • Integer • Smallint • Blob • Clob • Dbclob • Float • Decimal • Double • Graphic • Vargraphic • Longvarchar • Longvargraphic 	All of these data types, because they map one-to-one between federated and DB2 DataJoiner datastores.
Content Manager OnDemand	<ul style="list-style-type: none"> • Character • Varchar • Decimal • Long • Short 	<p>Varchar Map to OnDemand Varchar</p> <p>Decimal/Double Map to OnDemand Decimal</p> <p>Long and short Map to OnDemand Long and Short</p> <p>Date Map to OnDemand Long</p>
Domino.Doc	Varchar	Varchar to map to the Domino.Doc Varchar
Extended search	<ul style="list-style-type: none"> • String • Integer • Double • Date 	<p>Varchar Map to Domino Extended Search String</p> <p>Date/Time/Timestamp Map to Domino Extended Search Date</p> <p>Decimal and Double Map to Domino Extended Search Double</p> <p>Long/Short Map to Domino Extended Search Integer</p>

Table 6. Supported data types by content server (continued)

Content server	Supported data types	To define federated attributes, you can use:
Content Manager ImagePlus for OS/390	<ul style="list-style-type: none"> • Character • Varchar • Date • Time • Timestamp • Decimal • Double • Long • Short 	All of these data types, because they map one-to-one between federated and Content Manager datastores.
DB2 Warehouse Manager Information Catalog Manager	<ul style="list-style-type: none"> • Character • Varchar • Date • Time • Timestamp • Decimal • Double • Smallint • Integer 	All of these data types because they map one-to-one between federated and to the DB2 Warehouse Manager Information Catalog Manager ¹

Notes:

1. For more information about how to define data types of the DB2 Warehouse Manager Information Catalog Manager, see *DB2 Warehouse Manager Information Catalog Manager System Administration Guide*.

Creating federated text indexes

Text Search Engine can be integrated with your IBM Content Manager servers, so that you can automatically index, search, and retrieve text information stored in Content Manager. Users can locate documents by searching for words or phrases. The text search server supports both single-byte and double-byte character sets.

If you are using Content Manager servers with the Text Search Engine, you can create a federated text index. You then map the federated text index to the Content Manager text search index on the Content Manager text search servers.

When you create the federated text search index, you can enable it for combined searching, that is, for searching both native text indexes and native attributes. When enabling a federated text index for combined searches, you also map that index to a federated entity. You then map the native attributes mapped by the federated entity, and its federated attributes, to the native text search indexes on the text search servers.

Creating search templates

After you have created the necessary federated entities and attributes, you can now create a search template. Creating a search template involves associating the appropriate Enterprise Information Portal federated entities and attributes and defining search criteria. Each search criterion is mapped to a federated attribute.

Associating objects with the search template

When you create a search template, you must:

- Associate the template with one federated entity
- Associate the template with one federated text index, if applicable
- Add one or more users or user groups to the access list (see “Managing access” on page 23)

Because federated entities, user IDs, and user groups are separate objects that are associated with the template, you can use the same objects in a different search template. The reusable objects reduce the administrative overhead.

For example, Pat creates a search template called `claim calls` and associates it with the federated entity `claims`. She adds the user groups `agents` and `ccrs` to the search template access list.

Defining and mapping search criteria

Search criteria are specific field values defined in a search template. As an administrator, you create the search criteria for each search template and map each criterion to a federated attribute. You can also assign each field a default operator and value. Assigning default operators restrict the operators that a user can use to perform searches. Default values are the initial search criteria values that a user sees when using a given search template.

A default operator such as `like` or `equals`, should be assigned when it helps refine the search. If you assign a default operator to a search criterion, the federated attribute associated with the search criterion must support the operator. A federated attribute supports the intersection of the native entities that are mapped to it. For example, Table 7 lists the valid operators for Content Manager ImagePlus for OS/390 content servers.

Table 7. Valid operators for Content Manager ImagePlus for OS/390

Native entity	Attribute	Valid operators
Document	FolderId ¹	equals
	ReceiveDate	equals, greater than, less than, greater than or equal, less than or equal, between
	FileTab	equals
	UserStatus	equals
Folder	FolderId ¹	equals
	or FolderToken	
	IndexId	equals
Folders	IndexId ¹	equals
	and IndexValue	
	FolderType	equals
	CreateDate	equals, greater than, less than, greater than or equal, less than or equal, between

Notes:

1. You must include this attribute in every search template that queries a Content Manager ImagePlus for OS/390 search template.

For example, you can map a federated attribute such as date to CreateDate and ReceiveDate. When you map the search criterion, Claim date to the federated attribute date, you can then set the following default operators:

- equals
- greater than
- less than
- greater than or equal
- less than or equal
- between

Both of the native entities, CreateDate and ReceiveDate, support all six operators for Content Manager ImagePlus for OS/390.

Another example is to map the federated attribute record to FileTab and ReceiveDate. When you map the search criterion, Claim record to the federated attribute record, you can set the default operator only to equal, because FileTab does not support the rest of the operators.

The default value you assign to a search criterion should also be consistent with the federated attribute's data type.

For example, Pat knows she wants to query using a policy holder's name, policy number, and the date the claim was filed. She creates three search criteria, which she maps to federated attributes as shown in Table 8.

Table 8. How Pat maps the search criteria

Search criterion	Federated attribute
Name	policy_holder_name
Policy Number	policy_number
Filing Date	date

Pat then defines the default operators and values as listed in Table 9.

Table 9. How Pat defines default operators and values

Search criterion	Default operator	Default value	Explanation
Name	like		Pat uses like because names are rarely spelled accurately by the call center representatives. They often need to use a wildcard character. There is no default value because the names are never the same from call to call.
Policy Number	equals		Policy numbers match exactly, so Pat uses equals as the default operator. There is no default value because the policy number changes from call to call.
Filing Date	greater than	1/1/2000	The majority of calls are for the current year, so Pat sets a default value that returns items filed during 2000.

Table 10 lists all of the supported data types and operators by content server. Some operators are supported by only specific data types. Exceptions are noted by data type in the table.

Table 10. Supported operators by content server

Content server type	Supported operators
Content Manager	<ul style="list-style-type: none"> • equals • not equal • greater than • less than • like • not like • between • not between • in • not in • greater than or equal • less than or equal • and • or
Content Manager for AS/400	<ul style="list-style-type: none"> • equals • not equal • greater than • less than • like • not like • between • not between • in • not in • greater than or equal • less than or equal • and • or
JDBC	<ul style="list-style-type: none"> • = • <> • > • < • > or = • < or = • between • and • or • exists • in • like

Table 10. Supported operators by content server (continued)

Content server type	Supported operators
ODBC	<ul style="list-style-type: none"> • = • <> • > • < • > or = • < or = • between • and • or • exists • in • like
DB2 UDB	<ul style="list-style-type: none"> • = • <> • > • < • > or = • < or = • between • and • or • exists • in • like <p>For Blob, Clob, Dbclob</p> <ul style="list-style-type: none"> • = NULL • <> NULL
DB2 DataJoiner	<ul style="list-style-type: none"> • = • <> • > • < • > or = • < or = • between • and • or • exists • in • like <p>For Blob, Clob, Dbclob</p> <ul style="list-style-type: none"> • = NULL • <> NULL

Table 10. Supported operators by content server (continued)

Content server type	Supported operators
Content Manager OnDemand	<ul style="list-style-type: none"> • equals • not equal • greater than • less than • like • not like • greater than or equal • less than or equal • and • or
Lotus Domino.Doc	<ul style="list-style-type: none"> • equals • not equal • like • not like
Domino Extended Search	<ul style="list-style-type: none"> • equals • greater than or equal • less than or equal • greater than • less than • between • and • or
Content Manager ImagePlus/390	<p data-bbox="886 1129 1365 1213">Character, Varchar, Date, Time, Timestamp, Decimal, Double, Long, and Short equals</p> <p data-bbox="886 1234 984 1262">For Date</p> <ul style="list-style-type: none"> • equals • greater than or equal • less than or equal • greater than • less than • between • and • or

Table 10. Supported operators by content server (continued)

Content server type	Supported operators
DB2 Warehouse Manager Information Catalog Manager	<ul style="list-style-type: none"> • = • != • <> • > • < • > or = • < or = • like • between • in • and • or • not

Restriction: When you use an apostrophe (') in queries on a Content Manager for AS/400 content server, you are restricted to the `in` and `not in` operators. Table 11 shows how to revise queries that use an apostrophe (').

Table 11. How to revise queries for Content Manager for AS/400 servers that use the apostrophe (')

Operator	Example that fails	Alternative query	Explanation
equals	user ID equals "O'Neal"	user ID in "('O' 'Neal ')"	Switch operator
not equal	user ID not equal "O'Neal"	user ID not in "('O' 'Neal ')"	Switch operator
like	user ID like "O'Neal"	user ID like "O_Neal%"	Without wildcard character (%), no document is returned.
not like	user ID not like "O'Neal"	user ID not like "O_Neal%"	Without wildcard character (%), no document is returned.
<ul style="list-style-type: none"> • greater than • less than • greater than or equal • less than or equal 	user ID less than "O'Neal"	user ID less than "O_Neal"	Wildcard character (%) is not necessary.

Managing access

After you decide what a search template searches for, you must decide who needs to perform the searches. As an Enterprise Information Portal administrator, you create user IDs and user groups to give users access to search templates.

In our AA Insurance example, Pat knows the call center representatives and insurance agents need to search for claim-related information. Pat creates a user ID for each of the call representatives and insurance agents. Pat knows insurance agents have more access privileges than call center representatives, but all agents

have similar access needs. She creates a user group named agents, which includes all of the insurance agent user IDs. Pat also creates a user group named representatives that includes the call center representatives user IDs. Because she is defining these user groups for federated searches only, Pat assigns a default privilege set, UserPrivileges, to both user groups.

As an Enterprise Information Portal administrator, you are responsible for managing access to Enterprise Information Portal. You must decide who needs to perform administrative tasks and which users are to perform the tasks or actions in a workflow process. Your task is to create privilege sets, user IDs, user groups, and access control lists. If you have a number of users with similar access, making a user group saves time because you can enter one name to grant access to multiple users. You use privilege sets and access control lists to restrict access further.

Creating privilege sets

Your first task in managing access is to create privilege sets for users. A *privilege set* identifies the tasks or actions that a user can perform. Privilege sets combine privileges together and are tailored for certain types of users. For example, you might want one set of administrators to manage your workflow servers and another to manage your Web servers for information mining. For example, when you want to manage one of these servers, Enterprise Information Portal checks the administrator's privilege set to ensure that you have the authority to perform that task.

Enterprise Information Portal provides three predefined privilege sets:

AdminPrivileges

Includes all privileges, including administrative authority. You can:

- Define users, user groups, privilege sets, and access control lists
- Access content servers, workflow features, and Information Mining features
- Define federated entities, federated text indexes, and search templates

CoordinatorPrivileges

Includes privileges for workflow processes. You can define a privilege set for a user or user group who can coordinate the actions in a workflow process, such as transfer a work item from one user ID to another. For more information about workflow processes, see "Chapter 4. Introducing Enterprise Information Portal workflow" on page 39.

UserPrivileges

Includes privileges for workflow processes, such as viewing, creating, and updating work packets.

For more information about these privilege sets and how to define privilege sets, see the Enterprise Information Portal online help.

Creating users and user groups

After you create privilege sets, you must create user IDs and user groups to give users access to search templates, information mining, and workflow processes. You give users access to Enterprise Information Portal by assigning them a user ID, password, and associated privilege set. If you have a number of users with similar access, making a user group saves time because you can enter one name to grant access to multiple users.

For more information about assigning user IDs, passwords, and privilege sets, see the Enterprise Information Portal online help.

Creating access control lists

To further restrict access to Enterprise Information Portal, you can create access control lists. An *access control list* contains one or more individual user IDs or user groups, along with the privilege set associated with each user ID or user group. When a user attempts to perform a task, Enterprise Information Portal first checks the user's privilege set to ensure that the user is allowed to perform the requested task. If the user's user ID is part of a user group, then Enterprise Information Portal checks the user group's privilege set to ensure that the user is allowed to perform the requested task. You use access control lists to limit a user's authority to perform tasks within Enterprise Information Portal.

For example, at AA Insurance Pat wants to restrict who can perform certain actions within a workflow process, such as approving an insurance claim. Table 12 shows the user IDs and user groups with associated predefined privilege sets that Pat defines.

Table 12. User IDs and user groups at AA Insurance

User ID or user group	Privilege set
cmbadmin	AdminPrivileges, with complete administrative authority
admin1	AdminPrivileges, with complete administrative authority
claimgrp (includes user ID cmbadmin)	AdminPrivileges, with authority for workflow administrative actions
mrgsrp (includes user ID admin1)	UserPrivileges, with no administrative authority

Then, to limit access to the claims approval process, Pat completes the following steps:

1. Creates an access control list named `claimacl`, which includes the user groups, `claimgrp` and `mrgsrp`.
2. Defines a workflow process for approving claims.
3. In the workflow process, defines a worklist, `approve`, which includes the action, **Access Worklist**.
4. Associates the access control list, `claimacl`, with the worklist, `approve`.

In this scenario, the user ID `admin1` is limited to the `mrgsrp` privileges, even though the personal privilege set of `admin1` has administrative privileges. The access control list limits `admin1`'s access to the worklist `approve`. Only the user ID `cmbadmin` is allowed to access this worklist, because of `cmbadmin`'s privilege set (AdminPrivileges, with complete administrative authority).

Important: Additional access control might be needed to restrict access to the content servers supported by Enterprise Information Portal. Each content server has its own security requirements and capabilities, which are neither surfaced, nor overridden by the Enterprise Information Portal. To understand and implement user access control for these content servers, see the appropriate system administration books for each content server.

Optimizing the administration client

You can increase the performance of the Enterprise Information Portal administration client by setting a maximum heap size in `cmbfestart71.bat`. To edit `cmbfestart71.bat`:

1. On the workstation where the Enterprise Information Portal administration client is installed, open `cmbfestart71.bat` in a text editor. The `cmbfestart71.bat` file is in the Enterprise Information Portal installation directory. The Enterprise Information Portal default installation directory is `cmbroot`.

2. Add the option `-Xmx50M` after the `-Xms16M` option in the `java` command. For example:

```
java -Xms16M -Xmx50M fedsa.client.FeCCMain
```

3. It is recommended that your starting heap size and maximum heap size be the same. Update the option `-Xms16M` to read:

```
java -Xms50M -Xmx50M fedsa.client.FeCCMain
```

Chapter 3. Managing information mining

Most organizations have large numbers of online documents such as electronic mail, Web documents, technical reports, and news wires. With Enterprise Information Portal Information Mining, your users can search for relevant documents using powerful queries and organize documents by subject.

Enterprise Information Portal Information Mining offers the capability to create intelligent searches through two components: a categorization tool and the IBM Web Crawler. The categorization tool assigns documents to predefined categories, topics, or themes. The IBM Web Crawler keeps a copy of every object from a defined webspace, and your users can view these objects with a Web browser or index them using a search engine.

Configuring and starting Enterprise Information Portal Information Mining

If you choose to install the Information Mining feature after the initial installation of Enterprise Information Portal, you need to configure the Enterprise Information Portal system for the Information Mining feature. You also need to change the configuration if you install the Information Mining feature on a different workstation than the workstation on which the administration client program is installed.

1. From the Administration window, click the file member **Tools**.
2. Select **Services** from the menu.
3. Select the **Information Mining** check box.
4. After the configuration is complete, log off of the Enterprise Information Portal administration client and then log on again to initialize the Information Mining feature. After you log on to the Enterprise Information Portal administration client, the **Information Mining** icon appears in the left pane.

Tip: Administrators do not see the **Information Mining** icon unless they have authority to administer the Information Mining feature. If you want to restrict access to the Information Mining feature, see the appropriate system administration books for each content server. See the Enterprise Information Portal online help for more information about authorizing administrators to manage the Information Mining feature.

To start Enterprise Information Portal Information Mining:

1. In the left pane of the Enterprise Information Portal Administration window, double-click **Information Mining**.
2. In the right pane, right-click **Information Mining**, and select **Start/Stop Server**.
3. Leave the command window open to continue running the server.

To stop the Information Mining server:

1. In the left pane of the Enterprise Information Portal Administration main window, double-click **Information Mining**.
2. In the right pane, right-click **Information Mining**, and select **Start/Stop Server**.

Defining category schemes

As the Enterprise Information Portal administrator, you define the categories that your users use to intelligently search documents. A category scheme contains information about the expected data and its distribution in documents that belong to a certain category. In Enterprise Information Portal Information Mining you define categories using sample documents. This “training” phase produces a special index, which is used to categorize new documents.

To define the categories, you:

1. Decide which categories to define and create an appropriate directory structure.
2. Provide a set of sample documents for each category and store them in the directory structure.
3. Run the Categorization Training Tool to create a taxonomy and categorization scheme to be used by the categorization JavaBean.

Guidelines for defining categories

To get high quality results for the categorization, selecting the appropriate sets of documents for training is essential. Use the following guidelines for selecting the appropriate sets of documents as candidates. Candidate documents should:

- Contain a significant amount of text (avoid too much markup and avoid a list of words).
- Be typical for category they represent (most of the content should relate to this category).
- Be as distinct as possible from the sets of documents that represent other categories.

Beginning with 10 to 20 documents is a good start if the documents comply with these criteria.

To refine the categorization results, complete the following steps:

1. Create a categorization scheme using sets of documents to represent categories.
2. Select documents from different training sets and run them through an application that uses the categorization JavaBean.
3. Compare the results with the category to which the document belongs.

Tip: Sometimes a document selected from a training set does not end up in the same category for which it has been selected. Some reasons relate to the statistical nature of the category assignment process. A success rate above 80% is usually a good result. If too many mismatches occur in a category assignment, consider doing the following:

- a. Reorganize the taxonomy; for example, split the category into two categories.
 - b. Rearrange the training sets; for example, replace a document with more typical documents.
 - c. Make sure the type of documents you selected meet the guidelines listed above. If they do not, replace the documents with more appropriate ones.
 - d. Repeat step 3b to receive a representative number of documents arbitrarily selected from the training sets.
4. Categorize some documents that are not part of the training data and check whether the result is as expected.
 5. Refine the taxonomy and training sets as necessary.

Providing the training information

In the following scenario, an Enterprise Information Portal administrator wants to organize information into these categories:

- business
- culture
- politics
 - people
 - parties
- science
- sport

To train Enterprise Information Portal Information Mining to recognize information that belongs in these categories, the administrator must first provide files of information that are typical for these categories, using the following directory structure for the files:

```
.../trainingDirectory/home/business
.../trainingDirectory/home/culture
.../trainingDirectory/home/politics/people
.../trainingDirectory/home/politics/parties
.../trainingDirectory/home/science
.../trainingDirectory/home/sport
```

Some rules to observe when providing the training information are:

- The *trainingDirectory* must contain only one subdirectory (the root category), which, in the above example, is the home directory.
- You can put training documents in any directory that represents a leaf of the taxonomy, that is, in the lowest directories of the tree structure. Documents in other directories are ignored.
- The names of the directories in the taxonomy must be unique.
- The training information must be in one of the document formats supported by Enterprise Information Portal, including common word processing, spreadsheet, graphics, database, and presentation formats. For a list of supported formats, see:

```
ftp://ftp.software.ibm.com/ps/products/IMiner/fortext/fixes/NT/GeneralNT/Documentation/README012000-NT.html#formats
```

You must type this address exactly as shown.

Running the training wizard

After you have the training information in place, you can begin the training tool:

1. Click **Start** → **Programs** → **Enterprise Information Portal** → **Categorization Training Tool**.
2. Enter the training directory and a taxonomy name.

When the training process is completed, the training results are stored in two files that have the same name as the taxonomy:

```
.../trainingDirectory/taxonomyname.tcd
.../trainingDirectory/taxonomyname.taxonomy
```

These two files are used by the categorization JavaBean and by the advanced search query builder bean for defining a query that searches in documents of a particular category.

Importing documents from a webspace

You can automate the process of importing documents from a predefined area of the Web or an intranet (a *webspace*) on the Enterprise Information Portal Information Mining server. Follow these steps:

1. **Set up environment variables:** Set up the environment variables TZ and PATH for IBM Web Crawler. See “Step 1: Setting environment variables” on page 31.
2. **Start the Control Service:** Start the Web Crawler Control Service program that performs start and stop requests. See “Step 2: Starting the Control Service” on page 32.
3. **Define a webspace:** Specify the Internet or intranet area where you want Web Crawler to obtain information. See “Step 3: Defining a webspace” on page 32.
4. **Crawl the Web:** Run Web Crawler to monitor the Web and gather documents. See “Step 4: Crawling a webspace” on page 34.
5. **Import the crawled documents:** Use the CMBWebCrawlerService JavaBean to import the documents into Enterprise Information Portal Information Mining. See “Step 5: Importing the crawled documents” on page 37.

Tip: If you want to begin importing documents from a webspace now, see “Step 1: Setting environment variables” on page 31.

Introducing the IBM Web Crawler

IBM Web Crawler monitors a particular area on a Web, specified by a user-defined set of addresses. The Web Crawler keeps a current copy of every object in that area, ready to be viewed or indexed using IBM Text Search Engine.

You can run several web crawlers to concurrently monitor different areas on a Web at the same time.

IBM Web Crawler is a ready-to-run web crawler. The Web Crawler understands the Hypertext Transfer Protocol (HTTP), knows how to follow addresses within Web-based objects, and can easily be configured to track objects of your choice. It tracks the type and status of all servers and addresses, and logs unusual events. It interacts on request, informs external subsystems of the creation of new objects and of changes to existing objects, and reviews rapidly changing objects more frequently.

The CMBWebCrawlerService JavaBean

After the Web Crawler places a copy of the Web documents that it has been monitoring into a region on disk, you use the CMBWebCrawlerService JavaBean to import the documents that have been found into Enterprise Information Portal Information Mining.

The WebCrawler runs as a persistent process to monitor and imports documents as the documents change. The CMBWebCrawlerService JavaBean must be run within an application with a CMBCategorizationService JavaBean to import crawled objects into the catalog. After the CMBWebCrawlerService JavaBean starts, it continues to run, importing any documents that the Web Crawler writes into a disk region. New documents are imported and made searchable, and previously imported documents are updated.

A WebCrawler sample, `WebCrawler.java`, is in `\CMBROOT\samples\java\jsp\infomining`. The sample demonstrates monitoring and loading of crawled objects.

Modifying the import operation: You can change some aspects of the import operation on the `CMBWebCrawlerService` JavaBean. By default, the program looks for crawled documents every 30 minutes, but you can change this value. It also notifies (using a `CMBResultEvent`) all listeners after every 100 imported documents. You can change this number, or cause a notification occur when crawled files are imported to the Information Mining set.

The `CMBWebCrawlerService` JavaBean has the following properties:

pollCycles

Number of times to look for documents to crawl.

pollMinutes

Number of minutes to wait before the next poll. The default is 30.

rootDir

The root directory of the webspace. This directory is completed by the Web Crawler. For example, if the webspace is `ibm` and the Web Crawler keeps the file system image of the webspace in `..\TextTools\Spidey\webspaces`, then the root directory and the data for webspace `ibm` is kept in `..\TextTools\Spidey\webspaces\ibm`.

archive

Keep files in `rootDir/webspace/archives`. The default is false.

ikfCatalog

Alias identifying DB2 Information Mining instances as defined on `ikfEngine` using portal administration. The default is `IKF` and should not be changed.

pageSize

Number of items to import before notifying listeners. The default is 100.

webSpace

The webspace that is monitored by the Web Crawler.

Step 1: Setting environment variables

To see or set one of the environment variables used by the Web Crawler, set the appropriate values in the **System Variables** field on the Environment page in the Control Panel.

TZ Some Web Crawler functions use the `TZ` environment variable to derive the correct coordinated universal time (UTC) (formerly known as Greenwich Mean Time or GMT) from the local time. To avoid possible conflict with other applications, the Web Crawler does not set this required variable. Ensure that you set it correctly.

If `TZ` is not present or is incorrect, type `TZ` in the **Variable** field and in the **Value** field, type a value of the format `xxxnyyy` (for example, `EST5EDT`).

xxx The three-character label for your time zone.

n The number of hours your time zone differs from UTC. If you are east of UTC, prefix the number with a minus (-) sign. If you are west of UTC, you may optionally prefix the number with a plus (+) sign. The *n* value ranges from -12 to +12.

yyy The three-character label for your time zone when in daylight savings time. If you do not observe daylight savings time, omit this value.

PATH If you want to use the Web Crawler Toolkit to build and use a custom crawler, you might want to change the PATH environment variable. Append the name of the Web Crawler executable files BIN directory to the current PATH variable. For example, if you installed the Web Crawler in c:\CMBROOT\TextTools, your PATH must include c:\CMBROOT\TextTools\bin.

Step 2: Starting the Control Service

The Control Service is a Windows service program that executes start and stop requests from the Web Crawler Control Monitor or the `imystart` command line program. To use either of these methods, the Control Service must be running.

To create the service program, use the `imycrtsv` command:

1. Enter `imycrtsv -c .\username`
2. You are prompted for a valid Windows user ID and password.
3. From the taskbar, click **Start** → **Settings** → **Control Panel**.
4. Double-click **Services**.
5. If the user has not been granted the Log On As a Service authority, click **Startup** and confirm the Startup Dialog by clicking **OK**.
6. Click **Start** to start the Web Crawler Control Service. After the next shutdown the service is started automatically.

Step 3: Defining a workspace

A workspace is the area that you want to crawl on the Web. You define the workspace in a set of files, either by using the command `imyclean`, or in Windows by using the Workspace Manager.

The files that define a workspace are:

include.re

Use this file to identify the servers to include. The service (server name and optional port specification) portion of each address must match a pattern in this file for the address to be of interest to the Web Crawler. The file can contain multiple lines and so any specific set of servers can be included.

Note that specifying such lines does not guarantee that these sites are crawled, but rather that any address found pointing to those sites could be of interest. For example, you might have two Web servers: Web Server A and Web Server B. Each Web server has multiple documents, but the documents on Web Server A do not contain links to the documents on Web Server B, and vice versa. Both servers are considered disconnected. To crawl both of these Web servers, you must specify both addresses.

exclude.re

Use this file to define patterns found in addresses that are to be excluded from the set passed by the `include.re` file. If any portion of an address matches any pattern in this file, that address is not included within your workspace. Typical entries for exclusion are backup files that might appear in a workspace.

imy.ini

Use this file to customize your webspace size, detect rate of change, set connection time-outs and retries, set disk storage, and so on. Entries in this file are for advanced needs and typically do not need to be changed. Instructions for making changes are in the file.

types.re

Use this file to identify the file formats to crawl. If the path portion of any address ends with .type, then that extension must match a pattern in types.re, or the address is skipped. For example, if types.re contains a line htm, then all addresses that end in an htm extension are crawled.

Defining a webspace using the imyclean command

This section describes an example where you crawl documents with the extension htm at the address `http://any.url.com`. You exclude documents at the address `http://any.url.com/trash` in a webspace called `myspace`.

1. Use `imyclean` to create and define a webspace having the name `myspace`

```
imyclean -C myspace
```

The `imyclean` command copies a default set of the webspace definition files and guides you to modify them and prepare a webspace that meets your needs.

2. To instruct the Web Crawler to include document metadata in each local copy, make this change to the webspace `imy.ini` configuration file:

```
[store] SAVE_HEADLINES = 1
```

3. Edit `include.re` to specify the addresses to be crawled. For example, to include `http://any.url.com`, add:

```
"any.url.com"
```

Replace the file contents with the domain and server names that you want to be within the crawled webspace. Be sure to include a backslash (\) before any period (.) in any word.

Tip: The servers listed in `include.re` are not crawled unless webspace validated addresses point to objects on those servers. To ensure that potentially disconnected objects are crawled, you can enter their addresses (or an address of any page pointing to them) on the Web Crawler command line. For a webspace with many disconnected pages, consider creating an HTML file containing links to those pages and specifying that file as your starting point. An added benefit is that additions to this file are recognized automatically by the Web Crawler according to the configured visitation schedule.

4. Edit `exclude.re` to specify the addresses not to crawl. For example, to exclude: `http://any.url.com/trash`, add:

```
"any.url.com/trash"
```

For efficiency, delete any lines that are outside the scope of your included webspace. Remember to insert a backslash (\) in front of characters that have a special meaning in regular expressions (metacharacters) as shown in the example.

5. Edit `types.re` to specify required type extensions. If there is a type extension at the end of the address, such as `htm` in `/path/file.htm`, a document is copied by Web Crawler only if this extension is specified in `types.re`.

The type can be any 1- to 12-character extension.

Defining a webspace using the Webspace Manager in Windows

In Windows, you can create a webspace by using the Webspace Manager.

1. Start the Webspaces Manager; double-click the **Webspaces Manager** icon in the IBM Web Crawler folder.
2. To create a webspaces:

Click **Create** and enter a 2- to 8-character (alphabetic) webspaces name. This is the name that is used in all future references to the webspaces; it must be unique across the system.

The base identifies the location on disk to place the webspaces. Use the default location or a location such as e:\. Ensure that enough storage is available at that location. Hard disk requirements depend on the size of the webspaces to be crawled. In general, plan 10 MB for 1000 Web objects.

Click **Create** to create a DB2 database for address and service metadata, and a directory/file structure at %BASE%.
3. Edit the files described in “Step 3: Defining a webspaces” on page 32 that define the scope of the crawl:
 - a. Click **Include**.
 - b. Click **Exclude**.

Your webspaces is now defined. You can define multiple webspaces. Note also that the Webspaces Manager supports webspaces cleanup, restart, and removal.

Step 4: Crawling a webspaces

This section describes how to crawl the webspaces to get the documents you want to import.

If you want to crawl a webspaces now, try this example, which crawls the webspaces you have just defined:

```
imycrawl -i http://any.url.com/adiirectory myspace
```

where `-i` starts the Web Crawler in interactive mode, allowing you to observe progress. The Web Crawler acquires and saves any new or changed objects that it finds, and follows links to objects within the defined webspaces.

You can allow the Web Crawler to continue running.

What happens during a crawl

When the Web Crawler begins crawling, it requests addresses and stores the returned objects into the file system that `imyclean` prepared at the location specified by the `BASE` attribute of the `[STORE]` section in the `imy.ini` file.

It also examines the content of the object, scanning for any HTML tags containing an address, such as `` or `<frame>`. Addresses not conforming to any of the rules that you defined using the `imyclean` command are ignored; otherwise, they are scheduled to be visited.

Starting a crawl

Start a crawl by entering the command:

```
imycrawl [-i] [addresses] webspaces
```

-i Starts the crawler in interactive mode, allowing you to observe progress and to query the values of some Web information. You should use the `-i` option the first time you use this command.

addresses

One or more starting addresses in the form `http://<server>[:port][/<path>]`. If you do not specify an address, the

Web Crawler looks in the Scheduler for addresses. To generate multiple crawlers on the same Web, specify several addresses. Alternatively, you can use the above procedure repeatedly using a different starting address. Be aware that running multiple Web Crawlers can affect response time and this might affect other users who share the same network resources.

Requirement: At least one crawler must have a starting address.

webservice

Name of a previously defined webservice.

Interacting with the Web Crawler

The `imycrawl` command offers an interactive mode that lets you observe progress and query the values of some Web information. To start the interactive mode, run the command `imycrawl` using the `-i` option, or interrupt a crawl, by pressing `Ctrl+C`.

In the interactive mode, the Web Crawler displays a command line:

```
[c]ont, o)bject, s)ervice, rR)esched, i)my, a)cct, #)verbose, g)o, q)uit ?
```

Enter one of the following command letters:

c (default)

Continue one step, typically an attempt to visit an address.

o View metadata about the most recently visited object.

O View metadata about a specific object. The crawler requests that you enter the address or object ID (OID) of a particular object, which it then attempts to retrieve immediately.

s View metadata about the most recently visited service.

S View metadata about a specific service. The crawler requests that you enter the name or service ID (SID) of a particular service and then displays metadata for that service.

r Print a segment of the object crawl schedule. The crawler requests a beginning object ID (OID) and a count; it then displays visit pace (the number of ticks between crawls) and appointment information (at which tick the object will next be crawled) for each object.

R Reschedule an object for visitation during the current tick. (You can display the current tick by entering `'i'`.) The crawler requests an OID; that OID is rescheduled for visitation during the current tick. If you enter no OID, the crawler reschedules all OIDs for visitation during the current tick.

i View crawler general metadata including configuration input.

a Display the current accounting report.

Enter a number to switch verbosity on in a single subsystem:

0 Verbosity off

1 General verbosity (this is the default in interactive mode)

2 Network connection details

3 HTTP details

4 Object storage details

5 HTML or address parsing details

6 Scheduling details

- 7 Service tracking and reservation details
- 8 Database access details
- 9 Web object details
- 10 Workflow details
- 11 Interaction details
- 12 DNS details
- 13 Request details
- 14 Response details
- g** Exit interactive mode, and reenter continuous operation. The last selected verbosity option remains in effect.
- q** Print the accounting report and exit.

Note that a sleeping crawler does not interact until the sleep completes.

Monitoring progress

To watch the Web Crawler's progress, execute the command:

```
imystat webspac
```

`imystat` maintains a status listing of `imycrawl` processes working on the named `webspac`.

You can use commands to query your database tables, such as:

```
db2 connect to webspac
db2 select '*' from imy.document
db2 select '*' from imy.service
db2 quit
```

For detailed progress monitoring and operational debugging, use the `imycrawl -i` interactive mode. **In Windows**, you can also use the Control Monitor window to view the state and progress of each crawler.

Excluding Web Crawler from a server

For security and performance reasons, an Enterprise Information Portal administrator might want to exclude certain servers or pages from being crawled. You might need to be able to limit the crawlers activities on your servers and pages.

You can instruct the Web Crawler to avoid servers or pages using an *access policy* file. The file is built according to guidelines published in *A Standard for Robot Exclusion* (see <http://info.webcrawler.com/mak/projects/robots/norobots.html>).

- The Web Crawler requests the access policy file `http://yourserver/robots.txt` before crawling a server and periodically thereafter.

- The file consists of lines in the form
`field:<optionalspace>value<optionalspace>`

If *field* is `User-Agent`, and *value* is `IBM-Spidey` or `*`, the following `Disallow` lines (through the next `User-Agent` line) specify partial addresses to avoid. This can be a full path or a partial path; any address that starts with this value is not retrieved.

For example:

```
Disallow: /help
```

disallows both /help.html and /help/index.html.

```
Disallow: /help/
```

disallows /help/index.html but allows /help.html.

An empty value allows all addresses to be retrieved.

- These lines can be separated by blank lines.
- You can include comments by typing a # character. The rest of the line is considered a comment.

Here are some examples:

- This /robots.txt specifies that all robots should avoid this server.

```
# disallow everybody
```

```
User-agent: *
```

```
Disallow: /
```

- This /robots.txt specifies that only Web Crawler can crawl this server, and that the crawler has no limits.

```
# allow only IBM
```

```
User-agent: *
```

```
Disallow: /
```

```
User-agent: IBM-Spidey
```

```
Disallow: # disallow nothing
```

- This /robots.txt specifies that all robots should avoid addresses within the temp, development, and testing htmldocs trees, that a user agent called IBM-Spidey has an exception for the development and testing trees (it is allowed there) and that the xyz and wxyz robots should stay away entirely.

```
# a more realistic example
```

```
User-agent: *
```

```
Disallow: /htmldocs/temp
```

```
Disallow: /htmldocs/development
```

```
Disallow: /htmldocs/testing
```

```
User-agent: xyz
```

```
User-agent: wxyz
```

```
Disallow: /
```

```
User-agent: IBM-Spidey
```

```
Disallow: /htmldocs/temp
```

Step 5: Importing the crawled documents

To import documents found during a crawl:

1. Start Enterprise Information Portal Information Mining:
 - a. In the left pane of the Enterprise Information Portal Administration main window, double-click **Information Mining**.
 - b. In the right pane, right-click **Information Mining**, and select **Start/Stop Server**.
 - c. Leave the command window open to continue running the server.
2. Run an application that uses the CMBWebCrawlerService such as the example provided in \CMBROOT\samples\java\beans\infomining.

Chapter 4. Introducing Enterprise Information Portal workflow

You can use Enterprise Information Portal to control the flow and performance of work in your business. When users work with the results of federated searches, they often must make decisions on what actions to perform. You can use workflows to determine in advance how you want users to perform the work.

You can automate the workflows of Enterprise Information Portal. You do this by setting up profiles and rules that control the way Enterprise Information Portal workflow components work together. You also choose how restrictive to make your system by controlling user access and authority through privilege sets and access control lists.

If you choose to install the workflow feature after the initial installation of Enterprise Information Portal, you must configure the Enterprise Information Portal system for the workflow feature. You also need to change the configuration if you install the workflow feature on a different workstation than the workstation on which the administration client is installed.

1. From the Administration window, click on the file member **Tools**.
2. Click **Services** from the menu.
3. Select the **Workflow** check box.
4. After the configuration is complete, log off of the Enterprise Information Portal administration client and log on again to initialize the workflow feature. After you log on to the Enterprise Information Portal administration client, the **Workflow Definitions** icon appears in the left pane.

Tip: Administrators do not see the **Workflow Definitions** icon unless they have authority to administer the workflow feature. If you want to restrict access to the workflow feature, see the appropriate system administration books for each content server. See the Enterprise Information Portal online help for more information about authorizing administrators to manage the workflow feature.

Understanding workflow

Most business operations can be characterized as a set of interrelated processes. Work flows from one employee to another, and from one department to another. Some simple processes might require only a few steps, while more complex processes involve a number of employees in different departments.

Enterprise Information Portal's workflow feature lets you move work through a process and make decisions about work throughout the process. For example, AA Insurance receives large volumes of claims forms in the mail. During the verification process, insurance claims adjusters need to gather documents such as photographs, appraisals, and expert reports. Employees spend several hours each day opening, sorting, filing, and monitoring information, as well as collecting pertinent documents for final approval.

This work packet of information moves from one employee to another as the information is received and checked. As the claims work packet is completed, it might be handled by employees in more than one department. Figure 2 on page 40 shows a high-level view of a claims workflow.

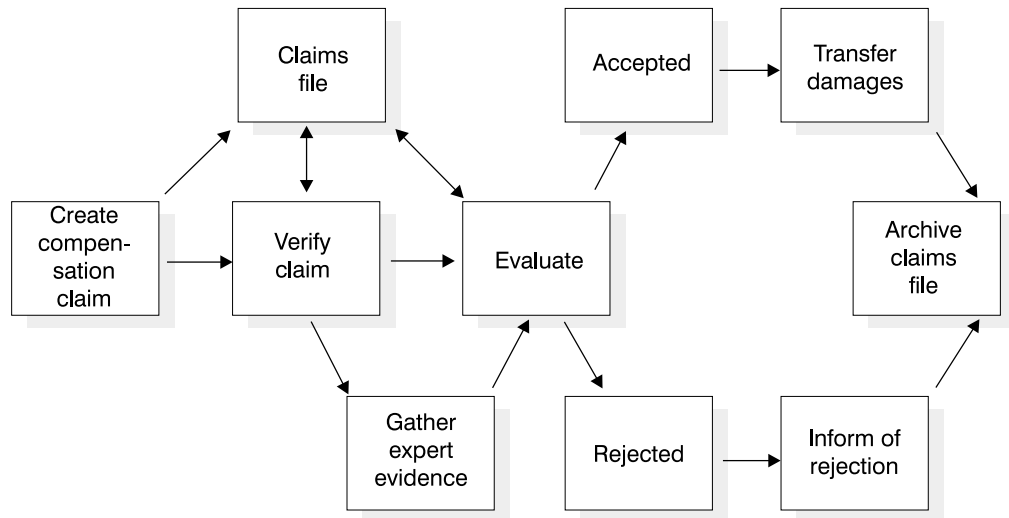


Figure 2. High-level view of a work process

How to use workflow

As in our AA Insurance example, most enterprises that handle documents perform some or all of the following tasks:

- File documents for later retrieval.
- Collect documents, forms, reports, and information from different sources, then deliver these documents to somewhere to be processed.
- Match incoming mail with documents currently being processed.

A *workflow* represents the flow of work. It describes the actions that can be performed on a group of one or more documents or content and the path this group of documents takes throughout the workflow. A workflow reflects work the way that it is performed, with a clearly defined scope and boundaries. It defines the sequence of activities and tasks, and the connections and relationships among those activities and tasks. A workflow determines the criteria that are used to make decisions about the flow of work.

Planning a workflow

Before you begin to define a workflow, you must analyze the work that your business performs, where and how it is performed, and by whom. An Enterprise Information Portal administrator or business analyst does this planning step.

What is the final product? The final product might be the result of all the work accomplished by your business, by one department in your business, or by certain employees from different departments. For example, the final product of the claims compensation process of AA Insurance is the letter sent to the policy holder approving or rejecting the claims form.

Analyze the information that must be processed to produce the final product, determine the actions that must be performed and where they are performed, and decide how you want the information to flow through the workflow.

Information to be processed

Consider the information that must be handled by users in your enterprise. What types of input support the final product? What are the specific documents that must be processed?

A *work packet* consists of zero or more work items, or activities, that are moved together from one place to another in a workflow. A *work item* can be any content (documents or objects) from a content server. For example, AA Insurance initially receives claims forms and later receives follow-up documents, such as photographs, appraisals, and expert reports. This information forms the work packet.

How information is handled

Who can best handle each step of the process? For example, an administrative assistant might verify that a claims form is complete, then file the application until a certain document is received from the policy holder. When the document arrives, the claims adjuster might be responsible for matching the document with the claims form, and for approving that document.

The claims forms can be grouped into a *worklist* that is accessible to a number of claims adjusters. A *worklist* can be thought of as a queue of work that you create for one or more employees to use. This queue of work consists of work packets. The worklist is filtered view of the workflow. Employees only see work packets in a worklist that they are allowed to see. Work packets assigned to a single worklist can be worked on by as many employees as you have assigned to that worklist.

Worklists can be defined within the workflow to handle each part of the claims process, such as gathering photographs, appraisals, and reports. A worklist can also consist of work packets from different workflows. For example, the worklist for one claims adjuster can contain appraisals for one claim, photographs for a second claim, and an expert report for a third claim. The actions the adjuster performs for each work packet in this worklist can be different. The adjuster might review the appraisal and approve the first claim. She might need to wait for more information about the second claim before taking action on the photographs. For the third claim, she could send the expert report to another employee for action.

Actions to be taken

Consider what actions should be taken on the contents of a work packet during the workflow. For example, a claims adjuster can accept a claims form or reject it as incomplete. An *action list* defines the actions that a user can perform on the work packet. An action list can describe the following:

- Selectable options that are available to the user
- Customer-defined options

For example, depending on whether a claim meets the initial requirements, an adjuster can select one option to continue the claims form through the workflow, or another option to reject the form.

When a user works with a worklist in Enterprise Information Portal, the user cannot change the actions defined for the work packet at the worklist.

How information flows through the process

Consider how you want information and activities to flow. For example, when is the initial claims form reviewed? What supporting documents are needed to move on to the next step in the process? What criteria determine whether a claims is accepted or rejected? This flow of information is the basis of your workflow.

A workflow consists of the paths that guide the work throughout processing. Where does the input originate? Your workflow must begin at some point. For AA Insurance, the claims form submitted by the policy holder is the document that starts the workflow.

When all of the documents are received, the work packet can continue along the path to a final action—for example, approval of the claim.

How everything fits together

After you analyze the information that you want to process, determine the actions that you want to perform, and decide how you want the information to flow, you are ready to create a workflow diagram, which is the graphical representation of your workflow.

A workflow diagram shows how work moves through the various activities in the process, noting what tasks the activity involves. It describes the flow, the main elements, and the key points of a workflow.

The workflow diagram in Figure 3 shows how information might flow through an insurance company such as AA Insurance.

Each symbol in the workflow diagram represents a point at which work is done. An insurance claim must be reviewed, supporting documentation must be collected, and the claim must be approved or rejected, depending on certain criteria. See “Creating a workflow” on page 45 for more information about the process symbols used in the workflow builder.

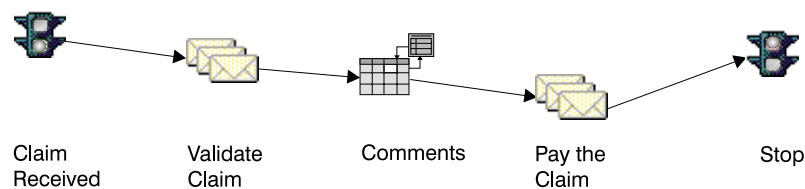


Figure 3. Sample workflow diagram

Using Enterprise Information Portal workflow components

This section describes the components of the Enterprise Information Portal workflow feature.

Workflow builder

You use the workflow builder to graphically define and build the workflow of a workgroup, department, or enterprise. The workflow builder is available from the Enterprise Information Portal administration client.

Before you use the workflow builder to create a model of your workflows, you need to define your privilege sets, access control lists, users, user groups, actions, action lists, and worklists using the Enterprise Information Portal administration

client. When you define a workflow in the administration client, you can set a default action list for the entire workflow. You can also assign a different action list at each node in the workflow. For more information about these tasks, see “Managing access” on page 23, “Defining action lists” on page 44, “Defining worklists”, and the Enterprise Information Portal online help.

Although you use the workflow builder to build workflows, you cannot use workflow builder to run the workflow. You must run the workflow from either a desktop client or thin client that you develop. You can also use the Enterprise Information Portal thin client samples for examples on how to code a sample thin client for workflow. Using a client, your users view and perform actions on worklists and work packets. For more information about how to create your own client to work with Enterprise Information Portal workflows, see the *Application Programming Guide* and online API reference.

Enterprise Information Portal workflow services

Enterprise Information Portal provides workflow services that maintain workflow information. The workflows and action list definitions that you create using the workflow builder are maintained in the Enterprise Information Portal database and the MQSeries Workflow database. Enterprise Information Portal also provides workflow services that route and keep track of activities and their state within a workflow.

Using a client you’ve developed to work with Enterprise Information Portal, users can log into Enterprise Information Portal to perform actions in a workflow. After a user performs a federated search and receives a worklist, the information associated with the worklist is permanently stored in the Enterprise Information Portal database. When a user suspends a workflow, the workflow is locked in the Enterprise Information Portal database, and is marked within the database as being checked out to the user, preventing anyone else from updating it until the user is finished.

Defining worklists

A worklist can be thought of as a queue of work you create for one or more employees to use. A worklist is a filtered list of work items assigned to specific users or user groups. When users log on to Enterprise Information Portal they can see filtered lists of work items that are assigned them. You use the Enterprise Information Portal administration client to define your worklists.

A worklist definition includes the rules that govern the presentation, status, and security of its work packets. You specify the rules for each worklist at the same time as you create the worklist. To manage the access to a worklist, you create an access control list for the worklist. For a complete description of how to define worklists, see the Enterprise Information Portal online help. The worklist definition includes the following:

Access control list

An access control list consists of one or more individual user IDs or user groups and the privilege set associated with each. The privilege set is used to define a user’s authorization to access or perform certain tasks on the work packet. You use access control lists to limit user access to items in a worklist.

Filtering and sorting worklists

Criteria by which a user can view a filtered and sorted worklist.

Maximum number of items in a worklist

Maximum number of items that you want a worklist to contain.

Defining action lists

An action list is a list of the actions that a user can perform on work items or work packets. The action list must be a comprehensive list of all actions performed on work items or work packets. If an action is not listed on the action list for a work item or work packet, the user cannot perform that action.

Use the Enterprise Information Portal administration client to define your action lists. See the Enterprise Information Portal online help for details about how to define actions and action lists.

Predefined actions

The following alphabetically ordered list describes the Enterprise Information Portal predefined actions that you can use in action lists.

Change workflow

Allows users to change a selected work packet within a workflow, and change the priority of the work packet.

Continue workflow

Releases the work packet to continue to the next step of the workflow.

Delete notification

Removes the selected process notification from the Notifications worklist.

Delete work packet

Deletes a work packet from a workflow.

Explore

Opens a folder and updates the tree view to indicate which folder is being viewed.

Open Opens a document, folder, worklist, or work packet in viewer.

Resume workflow

Changes the status of a work packet to Running, from its current state of Suspended.

Suspend workflow

Holds a work packet in a worklist until a defined time period has been satisfied or until the work packet is resumed.

Terminate workflow

Stops a workflow.

Update work packet attributes

Allows users to edit the attributes of a work packet. Attributes include: priority, name of workflow, item status, workflow status, start time, due time, and owner.




Restriction: You cannot override the Enterprise Information Portal predefined actions, but you can define your own actions. If you define your own actions, they must match the Enterprise Information Portal API specifications for workflow actions. For more information, see the *Application Programming Guide* and online API reference.

Creating a workflow

After you define actions, action lists, and worklists, you use Enterprise Information Portal workflow builder to create the model of your workflow. The workflow builder window contains icons or symbols that represent groups of worklists and workflows.

The workflow builder provides visual cues for creating a workflow. Table 13 shows an overview of the process icons that you use to create a workflow in the workflow builder.

Table 13. Overview of workflow builder process icons

	Represents the beginning of a workflow. When you create a workflow, workflow builder automatically generates this icon. You can move this icon to any position on the drawing surface. Every workflow must have a beginning.
	Represents the end of a workflow. When you create a workflow, workflow builder automatically generates this icon. You can move this icon to any position on the drawing surface. Every workflow must have an end.
	A <i>work node</i> associates a worklist and an action list for a specific point in a workflow. A work node represents a point in the workflow where work is performed.

Starting Enterprise Information Portal workflow builder

To start and log on to Enterprise Information Portal workflow builder:

1. In the left pane of the Enterprise Information Portal Administration main window, double-click **Workflows**.
2. Right-click **Workflow Definitions**, and select **New** to create a workflow.

Requirement: You must create at least one access control list, one action, and one action list, before defining a workflow.

Starting the MQSeries Workflow server

Start the MQSeries Workflow server by entering: `cmbwfstart` at a command prompt. Two windows open for the MQSeries Workflow server. Leave these command windows open to continue running the server.

Part 2. Tasks for Enterprise Information Portal users

Chapter 5. Getting started with Enterprise Information Portal clients 49

Chapter 6. Defining document types to Enterprise Information Portal. 51
Changing the server MIME type file (cmbcc2mime) 51
Changing the client MIME type file (cmbmime2app) 52

Chapter 5. Getting started with Enterprise Information Portal clients

Requirement for starting the sample client application on Windows 95 or Windows 98: You must increase your environment space for the shortcut to 4096 bytes by completing the following steps:

1. Right-click **Start** from the Windows taskbar.
2. Click **Open**.
3. Double-click **Programs**.
4. Double-click **IBM Enterprise Information Portal**.
5. Right-click the **Enterprise Information Portal Client** icon and click **Properties**.
6. Click the **Memory** tab.
7. Enter 4096 for the number of bytes for environment.

To start and log into the Enterprise Information Portal sample client application:

1. From the Windows taskbar, click **Start** → **Programs** → **Enterprise Information Portal** → **Enterprise Information Portal Client**. The Login window opens.
2. Type a valid Enterprise Information Portal user ID and password.
3. In the **Server** field, select the Enterprise Information Portal database you want to access.
4. Click **Login**.

To start and log in to the Enterprise Information Portal thin client samples:

1. Open a Web browser.
2. Enter your Web server address. For example:
`http://server/eip/com/ibm/mm/servlet/login.jsp`

The Login window opens.

3. Enter a valid Enterprise Information Portal user ID and password.
4. In the **Server** field, select the Enterprise Information Portal database you want to access.
5. Click **Login**.

Chapter 6. Defining document types to Enterprise Information Portal

Enterprise Information Portal provides viewer support for some document types. If you define a document type to the server, you can launch documents within their native applications. For example, if you are storing Lotus Word Pro documents in your Content Manager OnDemand server, you can set Enterprise Information Portal to launch documents with an .lwp extension in Lotus Word Pro rather than the client document viewer.

You can define a document type by changing the following files:

cmbcc2mime.ini

Translates content classes to a MIME type stream, so content from content servers can be read by a client. The `cmbcc2mime.ini` file is in `installation_dir\cmb` where `installation_dir` is the Enterprise Information Portal installation directory. The default installation directory is CMBROOT.

cmbmime2app.ini

Translate the MIME type stream to the client application that you use to view the documents. The `cmbmime2app.ini` file is in `installation_dir\cmb` where `installation_dir` is the Enterprise Information Portal installation directory. The default installation directory is CMBROOT.

Important: When launching an application based on MIME type, only the base object is displayed. Any markup that was made to the document is not displayed. If the document has multiple parts, only the first part is displayed. The MIME type in both files must match.

For more information about CC2MIME, see *Planning and Installing Enterprise Information Portal*.

Changing the server MIME type file (cmbcc2mime)

When adding server MIME types, verify that the document type you are adding is a MIME type created for that file. For more information, refer to the Web site:

<ftp://ftp.isi.edu/in-notes/iana/assignments/media-types>

To add values to the `cmbcc2mime.ini` file, complete the following steps:

1. Open `cmbcc2mime.ini` in a text editor.
2. Use the following format for user-defined values:
 - Content class starts at 4096
 - The equal sign (=) follows the content class value
 - MIME type should follow the equal sign. If this is not a standard MIME type for that content class, follow these steps:
 - a. A MIME type is composed of a type and a subtype. Valid types are application, text, image, model, message, audio, and video.
 - b. A slash (/) follows the type

- c. To create the subtype, the token (x-) must precede the token used for that document; for example

`x-mydocumentclass (4096=application/x-mydocumentclass)`

- Repeat 2b on page 51 and 2c as necessary for every new MIME type.

If you are using Content Manager OnDemand content servers: OnDemand maps file extensions rather than content class numeric values to MIME type streams.

Changing the client MIME type file (cmbmime2app)

The `cmbmime2app` file is a file that is installed with every client and is used to add a client MIME type or change the application values of existing MIME types.

To add a client MIME type, or change the application values of existing MIME types, complete the following steps:

1. Open `cmbmime2app.ini` in a text editor.
2. Add or change the MIME type to the application association:
 - a. Verify that the MIME type is defined in the server file `cmbcc2mime.ini`.
 - b. Initially, when the MIME type file is opened, all existing MIME type definitions are commented out with the number sign (#). For MIME type definitions that the administrator chooses, the number sign (#) must be removed.
 - c. Specify the MIME type followed by the equal sign (=).

Important: Specify only the name of the executable program. Its fully qualified path name must be defined in the PATH environment variable. If the fully qualified path name is specified, you can use double backslashes (\\) or forward slashes (/) as directory separators. You can specify only valid DOS directory and file names (8.3).

- d. Specify a space or blank, then `x=`.
- e. Specify the most common file extension used for the document type; include a period (.) before the file extension. For example:

```
text/plain=a=notepad x=.txt
text/richtext=a=c:\\progra~1\\window~1\\access~1\\wordpad x=.rtf
image/tiff\\c:/progra~1/window~1/access~1/imagevue/wangimg x=.tiff
```

- f. Save the `cmbmime2app.ini` file.

If you are using Content Manager OnDemand servers: The `cmbmime2app` file is included on the Enterprise Information Portal CD with the value `application/pdf=a=acrord32 x=.pdf` defined to allow OnDemand PDF documents to be displayed by the Adobe Acrobat Reader. The Adobe Acrobat Reader must be installed on the Enterprise Information Portal client, and the directory of `acrord32.exe` must be included in your PATH.

Part 3. Content Manager features

Chapter 7. Adding text search to your Content Manager system	55	Installing image search on AIX	65
Planning for text search	55	Setting up image search	66
Planning for text search capacity	56	Setting up the environment	67
Planning for text search performance	57	AIX example	67
Prerequisites for text search	57	Windows example	67
Installing text search	57	Configuring the image search server	68
Installing text search on Windows	58	Configuring the image search client	68
Starting the installation	58	Assigning an alias	68
Installing text search on AIX	59	Verifying the connection	68
Manually configuring additional text search components	60	Updating the Windows shortcut for the Content Manager system administration client.	69
Setting up to administer a text search server	60	Chapter 9. Loading and indexing sample data	71
Updating the shortcut for the Content Manager system administration client.	61	Before you load the sample data	71
Chapter 8. Adding image search to your Content Manager system	63	Creating a text search index	71
Installing image search	63	Creating the image search database, catalog, and features.	72
Installing image search on Windows	63	Running the loader program	73
		Indexing the sample text data	74

Restriction: Text search and image search are optional Content Manager features that can be configured and run with the Content Manager servers only. If you do not use Content Manager servers, do not install these features.

Chapter 7. Adding text search to your Content Manager system

Text search can be integrated with your Content Manager servers, so that you can automatically index, search, and retrieve documents stored in Content Manager. Users can locate documents by searching for words or phrases. The text search server supports both single-byte and double-byte character sets and runs on both AIX and Windows®.

Text search includes structured document support for XML, HTML, and tagged ASCII documents, which allows you to search for terms within specified sections of a document. You can search for data in nested sections. You can search by full XML context, for example, you can search for IBM within titles, and search for IBM within a title that is in a specific section. When you specify a DTD path, text search can dynamically use the appropriate DTD for each document, assuming that a reference to the DTD is stored as metadata for that document.

Planning for text search

In a typical LAN environment, a text search installation comprises one or more servers and several clients. You can install the Text Search Engine (server and client) on a machine with other Content Manager components. You can install additional client components on client workstations.

In addition to the server and client components, Text Search Engine uses dictionaries to support the linguistic processing of documents in different languages during indexing and retrieval. Dictionaries installed at the client workstation are used when applications retrieve the matches of a previous search for later highlighting. Dictionaries installed at the server workstation are used by the search service while indexing and searching.

You can install the following components:

Text search client

The client component that works with the text search server and the Content Manager client to access the Content Manager documents.

Text search server

The server component includes the text search library services, which access the Content Manager documents. The text search server requires the Content Manager toolkit shipped with Enterprise Information Portal. The text search server can reside on the same workstation.

Dictionaries

Monolingual dictionaries and stop-word lists are available for the following languages:

Arabic	Hebrew
Brazilian Portuguese	Icelandic
Catalan	Italian
Chinese, simplified	Norwegian
Chinese, traditional	Portuguese
Danish	Russian
Dutch	Spanish
Finnish	Swedish
French	Swiss German
French Canadian	UK English
German	US English

The dictionaries are installed automatically. You do not have to select them.

If you require additional dictionaries, contact your IBM representative about availability.

Content Manager Toolkit

The Content Manager toolkit includes the text search APIs and header files and is a prerequisite for all of the text search components

Planning for text search capacity

In addition to meeting the Content Manager prerequisites, you need to reserve resources for indexing. You need permanent disk space for document indexes, plus temporary disk space for merging index parts and for indexing (the merging space and indexing space are not additive, because merging and indexing do not run concurrently). Carefully planning your capacity requirements helps you avoid future capacity problems.

The amount of permanent storage that you need for the document index depends on the number and size of the documents and several other factors:

Number of graphics

The more graphics your documents contain, the less space you need for the index.

Document types

Word processing documents require less space than text files, because most of the word processing file content consists of control characters rather than text.

Number of unique words or numbers

Documents containing lists of sequential numbers, for example, take up relatively more space, because each number is unique and is indexed separately.

Type of index

- Precise indexes contain words as they appear in the documents. Because there is little linguistic processing, precise indexes require the most storage space.
- Linguistic indexes store words without variations, for example, verbs are reduced to their stem, plural words are changed to singular. Because there is some linguistic processing, linguistic indexes require less room than precise indexes.

- N-gram (or GTR) indexes eventually become as large as the indexed documents they store. Initially, the N-gram indexes are larger than the documents they store.

Temporary disk space for merging and indexing

- When you estimate your disk storage, plan for an additional 70% above the size of the real documents. The additional 70% storage is used to store the data associated with the document index.
- For merging, reserve about 2.5 times the size of the primary part and the secondary part of the document index.
- As with permanent disk space, the amount of temporary disk space that you need for indexing varies. It depends on the size of your secondary index, and on the size of the documents you index.

Estimate your temporary capacity requirements with the following steps:

1. Take the current size of the secondary part of the document index.
2. Add about 70% of the size of the documents that you index.
3. Again, add about 70% of the documents to be indexed. If this last calculation comes to more than 10 MB, then use the figure 10 MB.

Planning for text search performance

To reduce the performance impact, run the text search indexing function when processor demand is low. If you index documents while many other applications are using the processor, then you might experience reduced performance.

Tip for AIX: To make backing up your data easier, create a file system for the document index files. Index work directories do not need to be backed up. You can complete the following tasks:

- Create a file system for the document index files.
- Create a file system for work files.

All of the directories in these file systems must have read/write/execute permission (mode 775) for users in the text search administration group.

Prerequisites for text search

The text search server requires the Content Manager access toolkit provided with Enterprise Information Portal.

Ensure that you have read the sections about planning for text search and have calculated enough hardware resource to install the Text Search Engine.

Installing text search

You must install the text search client on the Content Manager system administration gateway workstation to enable the Content Manager system administration program to administer text searches. You must install the text search client on the Enterprise Information Portal administration program workstation to use the text search feature for federated searches.

Both the text search client and text search server run on Windows and AIX. Though the configuration tasks on Windows and AIX are similar, the installation tasks differ. You can install the server and client on different operating systems.

Installing text search on Windows

Start the installation program and follow the instructions in the installation windows. After you enter the information requested on a window, click **Next** to proceed to the next window. If you want to review or change any information that you entered, click **Back** to return to the previous window.

Starting the installation

To install text search:

1. Insert your Enterprise Information Portal CD in the CD drive.
2. The setup wizard starts automatically. If the wizard does not start, use the Windows Explorer to open the CD-ROM folder and double-click the **Setup** icon in the WIN\ENU directory where ENU is the three-character abbreviation for your workstation's locale.

The Enterprise Information Portal logo displays.

3. Click **Nextt**.

The Copyright and License window opens.

4. After reading the license terms in the Copyright and License window, indicate that you are willing to accept them by clicking **Accept**.

The Component Selection window opens and displays the components that are available for installation.

5. From the list on the left, select **Features**. From the list on the right, select **Text search server** and **Text search client**.
6. You can accept the default file path for your installation, or you can select your own. To select a different file path, click **Browse**. A window opens from which you can select a location.

If any Content Manager components are already installed, this area of the window is disabled.

If there is insufficient space on the selected drive, the installation program prompts you to select a different drive. If no drive on the target workstation has sufficient space, installation ends.

7. A window prompts you for system configuration information for the Text search components.

For the server configuration:

Instance name

Enter the text search server instance name. (Typically, a text search server has a single instance).

Host name

Enter the TCP/IP host name of the text search server.

Port number

Enter the TCP/IP port number for the text search server.

User ID

Enter the Content Manager user ID for logging on to the text search server.

Password

Enter the password for the Content Manager user ID.

For the client installation:

Requirement: The host name and port number of your client configuration must match the host name and port number of your server configuration.

User ID

Use the default value.

Server name

Enter the name of the alias for the text search server instance.

Recommendation: Use the TCP/IP host name for your server name. It creates a consistent host name for loading documents for various text search clients.

Server host name

Enter the TCP/IP host name of the workstation where the text search is installed.

Server port number

Enter a unique number that identifies the TCP/IP port for the text search server. This number must differ from the library server and the object server port numbers.

When you have entered your configuration information, click **Next**. The National Language Dictionary window opens.

8. Select the national language dictionary to use for your text search components, and then click **Next**.

After you install text search you must create a text search index, see the *Content Manager System Administration Guide* for instructions about starting the text search server and creating a document index.

If you are using feature indexes from Digital Library Version 2.4: You must migrate your feature indexes to a different index type. See the *Content Manager System Administration Guide* for instructions about migrating text search indexes.

Installing text search on AIX

Important: When you install AIX components, the install program and the administration IDs of the selected components must be running under the same locale. Otherwise, language-dependent files might not be available.

This section describes how to install text search components using the installation wizard on AIX. Do not use SMIT to install text search.

To install text search:

1. Log in as root.
2. Mount the Enterprise Information Portal CD.
3. Start the installation wizard by entering `./frnxsetup.sh` from the `/usr/lpp/cmb` directory.
4. Follow the instructions in the installation windows.

After you install text search you must create a text search index, see the *Content Manager System Administration Guide* for instructions about starting the text search server and creating a document index.

If you are using feature indexes from Digital Library Version 2.4: You must migrate your feature indexes to a different index type. See the *Content Manager System Administration Guide* for instructions about migrating text search indexes.

Manually configuring additional text search components

When you install Enterprise Information Portal the installation wizard configures your text search components. If you want to add text search components, you can configure them manually using Table 14.

Table 14. Manual configuration tasks for text search components on AIX

Steps	Text search client	Any text search server
Log in as root.	X	X
Enter: <code>su tsadmin</code>	X	X
where <i>tsadmin</i> is the user ID for the text search component		
Enter: <code>./usr/lpp/cmb/bin/frnconfig.clt</code>	X	X
If there is an <code>frnsetup.clt</code> statement in the <code>tsadmin.profile</code> change it to: <code>../frnsetup.clt</code>	X	X
Add the following statement to your <code>tsadmin.profile</code> . Line breaks are insignificant in each export statement: <code>export NLSPATH=\$NLSPATH:./usr/lpp/cmb/msg/En_US/%N:/usr/lpp/cmb/bin/msg/En_US/%N</code> <code>export LIBPATH=\$LIBPATH:./usr/lpp/cmb/imlsetup.ksh</code>	X	X
Where <code>En_US</code> is the 4-character code for the language used by your Enterprise Information Portal system.		
Add the following line to your <code>tsadmin.profile</code> : <code>export IMLCONFIGSRV=/home/tsadmin/ts</code>		X
Add the following line to your <code>tsadmin.profile</code> : <code>export IMLCONFIGCL=/home/tsadmin/ts/client</code>	X	

Run the following command on the text search server after installation:

```
imlcfgsv -i instance name -c tcpip -n port number -m
hostname -k nn -t xx
```

where *nn* is the initial number of connections and *xx* is the maximum number of connections. Due to a limitation to be corrected later, you must enter the same value for both *nn* and *xx*.

Setting up to administer a text search server

If you want to use a text search server, you must enable administration for it before starting the IBM Content Manager system administration program. To enable administration:

1. Start your IBM Content Manager library server.
Allow the library server to finish building the index classes.
2. Start the text search server on the workstation where it is installed by entering:
`imlss -start dinst`

where *dlist* is the name of your text search server instance chosen at installation time or when using the *imlcfgsv* command utility.

Updating the shortcut for the Content Manager system administration client

On Windows:

1. Right-click the **Start** button on the Windows taskbar to display the pop-up menu.
2. Select **Open**.
3. Double-click **Programs**.
4. Double-click **IBM Content Manager**.
5. Right-click the Content Manager **System Admin Client** icon to open the pop-up menu.
6. Select **Properties** to open the Properties notebook.
7. Click the **Shortcut** tab.
8. Add a **-t** to the **Target** field:

The **Target** field should read:

```
%cmbroot%\frnnacl.exe sa.client.SajcAdmin  
hostname -p 6021 -t -l en US
```

where `%cmbroot%\frnnacl.exe sa.client.SajcAdmin` specifies the application and its location on the system, `hostname` is the host name of the Content Manager system administration gateway, `-p 6021` specifies the port number, `-t` specifies that the text search component is installed, and `-l en US` specifies the locale.

9. Click **OK**.

Chapter 8. Adding image search to your Content Manager system

The image search server uses IBM's QBIC (query by image content) technology to help you search for objects by certain visual properties, such as color and texture. The image search server analyzes images and stores the image information in a database. Then users can run image queries, which use the visual properties of images, to match colors, textures, and their positions without describing them in words. You can combine content-based queries with text and keyword searches for more powerful retrieval of image and multimedia data.

Each image search server has a data directory containing one or more image search databases, which hold the image search catalogs. An image search catalog stores the data about the visual features of a collection of images. The actual image objects are stored on object servers in the IBM Content Manager system. The image search server runs on AIX and Windows.

Installing image search

Both the image search client and image search server run on Windows and AIX. Though the configuration tasks on Windows and AIX are similar, the installation tasks differ. You can install the server and client on different operating systems.

On Windows, you must complete the configuration using the image search command interpreter.

On AIX, the installation wizard completes the configuration for you.

Requirements:

- The image search server and the library server must be installed on the same operating system. They can reside on different machines, but these machines must run the same operating system. For example, if your library server is installed on a Windows machine, you can install the image search server on a different Windows machine.
- You must install the image search client on the Content Manager system administration gateway workstation to enable the Content Manager system administration program to administer image searches.

Installing image search on Windows

Start the installation program and follow the instructions in the installation windows. After you enter the information requested on a window, click **Next** to proceed to the next window. If you want to review or change any information that you entered, click **Back** to return to the previous window.

To install image search:

1. Insert your Enterprise Information Portal CD in the CD drive.
2. The setup wizard starts automatically. If the wizard does not start, use the Windows Explorer to open the CD-ROM folder and double-click the **Setup** icon.

The Enterprise Information Portal logo displays.

3. Click **Next**.
The Copyright and License window opens.
4. After reading the license terms in the Copyright and License window, indicate that you are willing to accept them by clicking **Accept**.
The Component Selection window opens and displays the components that are available for installation.
5. From the list on the left, select **Features**. From the list on the right, select **Image search server** and **Image search client**.
6. You can accept the default file path for your installation, or you can select your own. To select a different file path, click **Browse**. A window opens from which you can select a location.
If any Content Manager components are already installed, this area of the window is disabled.
If there is insufficient space on the selected drive, the installation program prompts you to select a different drive. If no drive on the target workstation has sufficient space, installation ends.
7. Both the server and client components require a configuration file path. The default value for this setting is `\CMBROOT\lib`. If you are installing the image search server, gather the information shown in Table 15. If you are installing the image search client, gather the information shown in Table 16.
After you provide the necessary values, click **Next**. Then confirm the settings, and click **Next** again. The installation program copies the program files, creates shortcuts and menu items, and creates the following three initialization files:

```
qbic.ini
qbic_alias.ini
qbic_usr.ini
```

Table 15. Installation inputs for image search server

Installation field	Input descriptions	Default value
Port Number	A unique number that identifies a TCP/IP port for the image search server.	6012
Library Server name	The library server to use with the image search server.	LIBSRVRN
Administration user ID	The Content Manager user ID that manages image search.	frnadmin
Content Manager password	Password for the Content Manager user ID.	password
Image path	The path to the control information for the image search server.	c:\frniss

Table 16. Installation inputs for the image search client

Installation field	Input descriptions	Default value
Image search server name	The image search server instance name.	QBICSRV
Image search server host	The host name of the image search server machine.	

Table 16. Installation inputs for the image search client (continued)

Installation field	Input descriptions	Default value
Image server port number	The TCP/IP port number used by the image search server. (This value must match the server setting, which also defaults to 6012.)	6012
Library server name	The name of the library server to use to administer the image search server. (This field is active only if the Content Manager system administration client is selected.)	LIBSRVRN

See *Planning and Installing Content Manager* for more information about installing Content Manager, including information about:

- Removing Content Manager.
- Installing in unattended mode.
- Reinstalling components.
- Generating network tables.

Installing image search on AIX

Important: When you install AIX components, the install program and the administration IDs of the selected components must be running under the same locale. Otherwise, language-dependent files might not be available.

This section explains how to install the image search component using the installation wizard.

See *Planning and Installing Content Manager* for more information about installing Content Manager.

To install image search:

1. Gather the information shown in Table 17 and Table 18 on page 66. Where the **Your value** column shows “default,” you can use the default value. Otherwise, you must enter a value specific to your Content Manager system.

Table 17. Image search server installation inputs

Installation field	Input descriptions	Your value
Image search server administration user ID	The AIX user ID that manages the image search server. This ID can be shared with the client toolkit user ID.	default ¹
Image search server name	The name of the image search server	QBICSRV
Port number	An integer between 5001 and 65535	default
Control data path	The path to the initialization files (*.ini)	default

Table 17. Image search server installation inputs (continued)

Installation field	Input descriptions	Your value
Library server database name	This value can be up to eight characters.	default
Content Manager user ID	The product user ID	default
Content Manager password	The product password	default

Notes:

- a. This value defaults to the client toolkit user ID. If you used the default value when you installed the client toolkit, it will be `cltadmin`. If you changed the user ID, the value defaults to your user ID.

Table 18. Image search client installation inputs

Installation field	Input descriptions	Your value
Image search server administration user ID	The AIX user ID that manages the image search server. This ID can be shared with the client toolkit user ID.	default ¹
Image search server name	The name of the image search server	QBICSRV
Host name	The host name of the image search server machine	
Port number	An integer between 5001 and 65535	default
Control data path	The path to the initialization files (*.ini)	default

Notes:

- a. You must create this ID before starting the installation wizard or use an existing user ID.
2. Log in as user root.
 3. Start the installation wizard by entering `./frnxsetup.sh` from the `DLAIX` directory at the CD path.
 4. Follow the instructions in the installation wizard.

Setting up image search

Setting up image search consists of:

1. Setting up the environment
2. Configuring the image search server
3. Configuring the image search client
4. Loading the sample image data

If you use the installation wizard on AIX: You do not have to run the configuration and setup scripts or issue the server configuration command. The wizard completes these tasks for you.

If you are installing on Windows: You must complete these tasks.

Setting up the environment

Complete the environment setup tasks in this section on both the server and client machines. The image search server requires the following environment variables:

QBICTOP

Resolves file names during image search configuration

QbicImagePath

Resolves file names on a server image file

QbicMaskPath

Resolves file names on a server mask file

QbicSketchPath

Resolves file names on a server sketch file

QbicTextPath

Resolves file names on a server text file

The image search client requires only the QBICTOP environment variable.

AIX example

On AIX, run the configuration script, which generates the setup script, and then run the setup script to set up the environment.

1. Run the following configuration script:

```
/usr/lpp/cmb/bin/frnconfg.iss QBICTOP
```

where QBICTOP is the directory path to the control files (*.ini). Set QBICTOP to /user1/cmb/qbic, where /user1 is the home directory of the image search administration user ID. The image search user ID must have read/write access to this directory.

This script generates the setup script: frnsetup.iss.

2. From the home directory of the image search user ID, run:

```
./frnsetup.iss
```

This script populates the environment variables for image search servers and clients.

Windows example

To set the environment variables:

1. Select **Start** → **Settings** → **Control Panel**.
2. Double-click **System**.
3. Click the **Environment** tab.
4. Set the variables and the values shown in Table 19 by typing them in the appropriate fields and clicking **Set**.

Requirement: The image search client requires only the QBICTOP variable. For client environments, set only the QBICTOP variable.

Table 19. Image search environment variables

Variable	Value
QBICTOP	d:\cmbroot\iss
QbicImagePath	d:\cmbroot\iss
QbicMaskPath	d:\cmbroot\iss
QbicSketchPath	d:\cmbroot\iss

Table 19. Image search environment variables (continued)

Variable	Value
QbicTextPath	d:\cmbroot\iss

Where d: is the drive where image search is installed.

Configuring the image search server

Before starting the image search server, you must configure it. Configuring the server consists of completing the initial configuration and verifying the connection.

To configure the server:

1. Start the command interpreter by entering: `qbicadm`
2. Enter the **config server** command. For example,
`config server LIBSRVRN FRNADMIN PASSWORD 9999`

where LIBSRVRN is the library server name, FRNADMIN is the Content Manager user ID, PASSWORD is the Content Manager password, and 9999 is the port number of the image search server.

See “Verifying the connection” for more information.

Configuring the image search client

Before starting an image search client, including the image search system administration program, you must configure it. Content Manager system administration requires that you assign one alias. Test your configuration by verifying the connection.

Assigning an alias

Before you can use the image search Content Manager system administration, which acts as an image search client, you must assign at least one server alias.

To assign an alias:

1. Start the command interpreter by entering: `qbicadm`
2. Enter the **add alias** command. For example,
`add alias QBICSRV HOSTNAME 9999`

where QBICSRV is the alias name, HOSTNAME is the image search server’s host name, and 9999 is the port number of the image search server.

Verifying the connection

Important:

1. The library server must be running to connect to the image search server.
2. Image search system administration requires an existing Content Manager user ID. To successfully connect to the library server, the image search user ID and library server user ID must be the same. (The default value for this user ID is `frnadmin`; if you change this value, make sure that the IDs match.)

To verify the connection:

1. After configuring the image search server and adding the alias, start it by entering `commsrv` from the command line.
2. To start the command interpreter, enter: `qbicadm`.
3. In the command interpreter, enter the **connect** command.

```
connect QBICSRV FRNADMIN PASSWORD
```

where QBICSRV is the alias name, FRNADMIN is the Content Manager user ID, PASSWORD is the Content Manager password.

After successfully connecting, the message Library Server is LIBSRVRN displays.

4. To disconnect from the server, enter: disconnect.
5. To exit the command interpreter, enter: quit.

Updating the Windows shortcut for the Content Manager system administration client

1. Right-click the **Start** button on the Windows taskbar to display the pop-up menu.
2. Select **Open**.
3. Double-click **Programs**.
4. Double-click **IBM Content Manager**.
5. Right-click the Content Manager **System Admin Client** icon to open the pop-up menu.
6. Select **Properties** to open the Properties notebook.
7. Click the **Shortcut** tab.
8. **Important:** The **Target** field contains the following information, which you must not alter.

```
%cmbroot%\frnnacl.exe sa.client.SajcAdmin  
hostname -p 6021 -t -l en US
```

where %cmbroot%\frnnacl.exe sa.client.SajcAdmin specifies the application and its location on the system, hostname is the host name of the Content Manager system administration gateway, -p 6021 specifies the port number, -t specifies that the text search component is installed, and -l en US specifies the locale.

In the **Target** field, append the following text:

```
-q QBICSRV HOSTNAME 9999 LIBSRVRN
```

where QBICSRV is the alias, HOSTNAME is the image search server's host name, 9999 is the port number of the image search server, and LIBSRVRN is the library server name.

Requirement: This information must match the information in the server configuration exactly. "Configuring the image search server" on page 68 describes the information that this shortcut must match.

9. Click **OK**.

See "Chapter 9. Loading and indexing sample data" on page 71 for instructions on loading the sample data.

Chapter 9. Loading and indexing sample data

This chapter explains how to load and index sample text and image data, which you use with the sample applications.

There are several sample loaders provided on the Enterprise Information Portal CD. This section describes how to load both image and text data using the sample loader LoadSampleTSQBICDL. You can load text and image search data separately to ensure both features are working properly.

Before you load the sample data

Before running the loader program, you must:

1. Create a library server configuration using the Content Manager system administration program. See the system administration program online help for assistance with this task.
2. Change the **Access** properties of the library server configuration by completing the following steps:
 - a. Right-click your new configuration and click **Properties** to open the Properties notebook.
 - b. Click the **Access** tab.
 - c. Click the **Unlimited sessions from any workstation** radio button.

Creating a text search index

Before you load data you must create an empty text search index that you can use to index the text samples. To create a text search index:

1. Start the text search server on the workstation where it is installed by using the following command:

```
imlss -start dlinst
```

where `dlinst` is the name of your text search server instance chosen at installation time or when using the `imlcfgsv` command utility.

2. Start and log in to the Content Manager system administration program.
3. Select **Text Search** from the list at the top left pane.
4. Double-click **Search Servers** in the left pane.
5. Double-click **TM**. TM is the search server alias for the text search server.
6. Double-click the **Indexes** folder from the left pane. If the message `RC_EMPTY_LIST` displays, then, from the menu bar, click **Selected** → **New** to create an index.
7. In the New Index window, define your index. Click **Help** for a detailed description of each field.

For example:

For Windows:

Name TMINDEX

Type Precise

Index files

x:\cmbroot\ts\index\tmlindex where x is your installation drive; if the path does not exist, it is created.

Index work files

x:\cmbroot\ts\work\tmlindex where x is your installation drive; if the path does not exist, it is created.

Information entry

Name of the Content Manager library server.

Do not change the client and server default DLL names.

For AIX:

Name TMINDEX

Type Precise

Index files

/home/cltadmin/tsindex/index/tmlindex; if the path does not exist, it is created.

Index work files

/home/cltadmin/tsindex/work/tmlindex; if the path does not exist, it is created. Ensure that the user has authorization to write to the directory.

Information entry

Name of the Content Manager library server.

8. Click **OK**.
9. Double-click **TMINDEX** to open the TMINDEX Administration Notebook.

Creating the image search database, catalog, and features

After you create a text search index for the sample text search data, you must create an image search database and catalog for the sample image data.

To create the image search database, catalog, features:

1. Start the image search server on the workstation where it is installed by entering the following command:


```
commsrv
```
2. Start and log in to the Content Manager system administration program.
3. Select **Image Search** from the list at the top left pane.
4. Click **Image Search Servers** in the left pane.
5. Click **QBICSRV**.

where QBICSRV is the image search server name you specified during installation.
6. Right-click **Databases** in the left pane and select **New Database**.
7. In the New Database window, enter SAMPLEDB in the **Name** field, and click **OK**.
8. In the left pane, click **Databases** to display the **SAMPLEDB** icon in the left pane.
9. Click **SAMPLEDB**.
10. In the left pane, right-click **Catalogs** and click **New Catalog**.
11. In the New Catalog window, enter SAMPLECAT in the **Name** field, and click **OK**.

12. In the left pane, click **Catalogs** to display the **SAMPLECAT** icon.
13. Click **SAMPLECAT**.
14. In the left pane, right-click **Features** and click **New Features**.
15. In the New Features window, select each feature in the **Name** field and click **Apply**. When all four features are selected the following message displays:
All possible features have been added to catalog.
16. Click **OK**.
17. Click **Cancel**.

Running the loader program

You can load sample data to test your text and image search.

The sample image data is in the following files:

On Windows:

x:\cmbroot\samples\java\d1\samples.jar

On AIX:

/usr/lpp/cmb/samples/java/d1/samples.jar

The sample loader program loads the data into Content Manager and indexes it. Read the prolog of the source program for instructions about the syntax for running the program. The sample loader programs are:

On Windows:

x:\cmbroot\samples\java\d1\LoadSampleTSQBICDL.jar

On AIX:

/usr/lpp/cmb/samples/java/d1/LoadSampleTSQBICDL.jar

To run the sample data loader program:

1. Decompress the .jar files by entering:

```
jar -xvf samples.jar
```

The files decompress into the correct directories.

2. Set your workstation environment variables to compile the sample loader program by completing the following tasks:

On Windows:

- a. Open x:\cmbroot\cmbenv.bat in a text editor and change the first three lines to set your workstation environment variables:

```
set CMBROOT = e:\cmbroot
set DB2HOME = e:\sql1ib
set JAVAHOME = d:\jdk117
```

- b. Save cmbenv.bat and set the environment variables by entering:

```
cmbenv
```

On AIX:

- a. Go to /usr/lpp/cmb/bin/ and run setup by entering

```
cmbenv
```

- b. Ensure that the subdirectories under /usr/lpp/cmb/samples/java/d1/ and the sample files are writable by all users.

3. Compile loader program by entering the case-sensitive command:

```
javac LoadSampleTSQBICDL.java
```

4. **Requirement:** The following servers must be running before you start the loader program:

- Library server
- Object server
- Text search server
- Image search server

If you are running in a National Language Version of Content Manager, set the `FRNDEFLANG` variable to `ENU` before running the loader program. The AIX command to set the environment variable is: `export FRNDEFLANG=ENU`

5. Load the sample data with the loader program by entering:

```
java LoadSampleTSQBICDL sampleQBIC.dat load.log frnadmin password LIBSRVRN
```

where your user ID is `frnadmin`, your password is `password` and your library server is `LIBSRVRN`.

6. Check the `load.log` to ensure that the sample data loaded successfully.

After you finish loading the sample data, use the Content Manager system administration program or text search command line tools to index the sample text data.

Indexing the sample text data

To index the sample data:

1. Start and log on to the Content Manager system administration program.
2. Select **Text Search** from the list at the top left pane.
3. Double-click **Search Servers**.
4. Double-click **TM**. TM is the search server alias for the text search server.
5. Right-click **new text index** and click **Properties**.
6. On the Explicit page of the Properties notebook, click **Refresh**.
7. The **Index count** field should display the number of documents that you loaded with the loader program.
8. Click **Index** to index the files.
9. After a few moments, click **Refresh** to see the number of successfully indexed documents in the **Primary document index** field.

After you index the data, you can use the sample Java application to query the collection or run a simple query by using the `imlsrch` command line tool.

Part 4. Appendixes

Appendix. Troubleshooting

This section describes possible solutions to typical errors.

Unable to connect to server <database name>, contact your system administrator

Ensure that the RMI server port number in the `cmbclient.ini` file on your client workstation is set correctly. See *Planning and Installing Enterprise Information Portal*.

Java exception: java.net.ConnectException

Shut down the RMI server and ensure that the master RMI server is running before you start the RMI server.

Java exception: java.net.UnknownHostException

Ensure that you have the correct host name in the RMI server's `cmbregist71.bat` file. See *Planning and Installing Enterprise Information Portal*.

Enterprise Information Portal client does not start

If your client application does not start, you might need to set parameters for the workstation Java environment and sample client application in `cmbclientapp71.bat` before you start the client application. (The `cmbclientapp71.bat` file is a file that is installed with every client and is used to start the sample client application.) If you create a shortcut for this step, and you change your configuration in the future, remember to update the shortcut as well as the `cmbclientapp71.bat` file. This section describes the parameters in `cmbclientapp71.bat` and how they should be set. You can set the parameters by opening `cmbclientapp71.bat` in a text editor.

Important: Before you set the parameters, ensure that the environment variables listed in the top of the file are set correctly. The environment variables are defined in the file comments.

connectionType

Specifies the location of the content server connector.

Example:

```
connectionType=DYNAMIC
```

LOCAL

All connectors are installed on the client workstation. If `connectionType=LOCAL`, then `clientUrlString` and `csUrlString` are ignored.

REMOTE

All connectors are installed on the RMI server broker or an RMI server pool member.

If `connectionType=REMOTE`, then `clientUrlString` is required and `csUrlString` is ignored.

DYNAMIC

Some connectors are installed on the client workstation and some are installed on an RMI server. For example, the Enterprise Information Portal administration client and Enterprise Information Portal database are running on the

client workstation, and the connectors are running on an RMI server. DYNAMIC is the default.

If `connectionType=DYNAMIC`, then `clientUrlString` is required and the `csUrlString` is optional. If you do not specify the `csUrlString`, you must ensure that the `csUrlString` is in the `cmbroot` directory of the remote host defined in `clientUrlString`.

Tip: The default setting is DYNAMIC for this variable. If you have server connections that are all LOCAL or all REMOTE, changing the setting to one of those values can result in a slight performance improvement.

clientUrlString

Specifies the full path name of the `cmbclient.ini` file.

Examples:

```
clientUrlString=file:///c:/cmbroot/cmbclient.ini  
clientUrlString=http://webserver/cmbclient.ini
```

csUrlString

Specifies the full path name of the `cmbcs.ini` file.

Examples:

```
csUrlString=file:///c:/cmbroot/cmbcs.ini  
csUrlString=http://webserver/cbmcs.ini
```

mime2app

Specifies the full path name or Web address of the `cmbcmime2app.ini` file. This is an optional parameter. If `mime2app` is not set, the application looks for the `cmbmime2app.ini` in the client application directory. The application displays an error message if the file is not found.

Examples:

```
mime2app=file:///c:/cmbroot/cmbmime2app.ini  
mime2app=http://webserver/cmbmime2app.ini
```

server The name of the Enterprise Information Portal database. This is a required parameter.

DGL0368: Error while retrieving a part. API SimLibOpenObj [FRN6052: ExtRC=0]

Content Manager for AS/400 server can return only references to documents stored externally to its repository. Currently, these documents cannot be viewed using Enterprise Information Portal viewer. If you receive this error message, it is because you tried to retrieve an external reference document.

Temporary restriction: For Content Manager ImagePlus for OS/390 and Content Manager documents that reference a form overlay, only the document part is displayed by the Enterprise Information Portal viewer.

Cannot restart the Enterprise Information Portal client after abnormal termination

If the Enterprise Information Portal sample client application terminates abnormally, the file used to log its output can be inaccessible due to a lock remaining on the file. Because the client tries to reuse the locked file, you cannot restart the program. The file is usable after restarting, or when the locking process is terminated.

To restart the sample client application:

1. In the %CMBROOT%\cmbclientapp.bat file, locate the line that ends with:
">clientApp_out.log"
2. Change the file name, for example:
">clientApp_out2.log"

Ensure that no other file exists with the name you choose; the output from your sample client application session is written to the new file.

Enterprise Information Portal client doesn't start from the shortcut

Sometimes, the sample client application does not start from the shortcut because Windows 95 and Windows 98 have a limit on the length of the shortcut. As a result, the environment variable %INIURL%, which is used by the program that starts the sample client application (cmbclientapp71.bat) fails with an error indicating that the cmbmime2app.ini file wasn't found.

To correct the shortcut:

1. Open cmbclientapp71.bat in a text editor.
2. Change the environment variable %INIURL%, to:
x:\directory_path

You must use double backslash (\) characters.

Viewer cannot find file to be viewed

Close any open instances of the viewer application that you are using to view the document. For example, if Lotus 1-2-3 is running and you try to view a Lotus 1-2-3 document, shut down all instances of Lotus 1-2-3. Then, try opening the document from the viewer.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

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

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

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

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J74/G4
555 Bailey Avenue
P.O. Box 49023
San Jose, CA 95161-9023
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this information and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

The following terms are trademarks of the IBM Corporation in the United States or other countries or both:

AIX	First Failure Support Technology
AS/400	IBM
Cryptolope	ImagePlus
DataJoiner	OnDemand
DB2	OS/2
DB2 Universal Database	OS/390
e-business	QBIC
FFST	

Domino and Lotus are trademarks of Lotus Development Corporation in the United States, other countries, or both.

Tivoli is a trademark of Tivoli Systems Inc. in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product or service names may be trademarks or service marks of others.

Glossary

This glossary defines terms and abbreviations specific to the Enterprise Information Portal system, but not necessarily to this particular document. Terms shown in *italics* are defined elsewhere in this glossary.

A

abstract class. An object-oriented programming *class* containing one or more abstract *methods*, which other classes can implement and thus make concrete.

access control. The process of ensuring that certain functions and stored objects can be accessed only by authorized users in authorized ways.

access list. A list consisting of one or more individual user IDs or user groups and their associated *privileges*. You use access lists to control user access to search templates in the Enterprise Information Portal system.

action list. An approved list of the actions, defined by a coordinator, that a user can perform in *workflow*.

API. See *application programming interface*.

application programming interface (API). A software interface that enables applications to communicate with each other. An API is the set of programming language constructs or statements that can be coded in an application program to obtain the specific functions and services provided by the underlying licensed program.

archive. Persistent storage used for long-term information retention, typically very inexpensive for each stored unit and slow to access, and often in a different geographic location to protect against equipment failures and natural disasters.

attribute. A characteristic that identifies and describes a managed object. The characteristic can be determined, and possibly changed, through operations on the managed object. For example, title and duration might be attributes of a video object. See *federated attribute* and *native attribute*.

B

binary large object (BLOB). A sequence of bytes with a size ranging from 0 bytes to 2 gigabytes. This string does not have an associated code page and character set. Image, audio, and video objects are stored in BLOBs.

BLOB. See *binary large object*.

C

cache. An area of storage used to temporarily store objects on the client workstation. See also *LAN cache*.

catalog. A temporary index used by Enterprise Information Portal applications to insert and remove objects that are not yet assigned to a *list*.

CGI. See *Common Gateway Interface*.

CGI script. A computer program that runs on a Web server and uses the *Common Gateway Interface (CGI)* to perform tasks that are not usually done by a Web server (for example, database access and form processing). A CGI script is a CGI program that is written in a scripting language such as Perl.

CIF. See *common interchange file*.

CIU. See *common interchange unit*.

class. In object-oriented design or programming, a model or template that can be instantiated to create objects with a common definition and therefore, common properties, operations, and behavior. An object is an instance of a class.

client application. An application written with the object-oriented or Internet APIs to customize a user interface.

collection. A group of objects with a similar set of management rules.

combined search. A query that combines one or more of the following types of searches: *parametric*, text, or image.

Common Gateway Interface (CGI). A standard for the exchange of information between a Web server and programs that are external to it. The external programs can be written in any programming language that is supported by the operating system on which the Web server is running. See *CGI script*.

common interchange file (CIF). A file that contains one ImagePlus Interchange Architecture (IPIA) data stream.

common interchange unit (CIU). The independent unit of transfer for a common interchange file (CIF). It is the part of the CIF that identifies the relationship to the receiving database. A CIF can contain multiple CIUs.

connector class. Object-oriented programming *class* that uses native methods to connect *content servers* to client applications.

constructor. In programming languages, a method that has the same name as a class and is used to create and initialize objects of that class.

container. A visual user-interface component that holds objects. In the *folder manager*, an object that can contain other folders or documents.

content class. The term used in the APIs for *data format*.

content server. A software system that stores multimedia and business data and the related metadata required for users to work with that data. Content Manager and Content Manager ImagePlus for OS/390 are examples of content servers.

cursor. A named control structure used by an application program to point to a specific row within some ordered set of rows. The cursor is used to retrieve rows from the set.

D

data format. A logical name assigned to a file type, valid only within Content Manager. Content Manager provides a large range of predefined data formats. You can also define your own data format. For example, you can define your own data format called MYGIF and use it for GIF files.

datastore. Generic term for a place (such as a database system, file, or directory) where data is stored.

DDO. See *dynamic data object*.

destager. A function of the Content Manager *object server* that moves objects from the *staging area*.

document. An *item* that can be stored, retrieved, and exchanged among Content Manager systems and users as a separate unit. It can be any multimedia digital object. A single document can include varied types of content, including for example, text, images, and spreadsheets.

document content architecture (DCA). An architecture that guarantees information integrity for a document being interchanged in an office system network. DCA provides the rule for specifying form and meaning of a document. It defines revisable form text (changeable) and final form text (unchangeable).

dynamic data object (DDO). In an application program, a generic representation of a stored object that is used to move that object in to, and out of, storage.

dynamic page builder. An API in the Internet application development toolkit that is used to create

applications that dynamically format the results of queries and display those results on a Web page.

E

extended data object (XDO). In an application program, a generic representation of a stored complex multimedia *object* that is used to move that object in to, and out of, storage. XDOs are most often contained in *DDOs*.

F

feature. The visual content information that is stored in the image search server. Also, the visual traits that image search applications use to determine matches. The four *QBIC* features are average color, histogram color, positional color, and texture.

federated attribute. An Enterprise Information Portal metadata category that is mapped to *native attributes* in one or more *content servers*. For example, the federated attribute, policy number, can be mapped to a *key field*, policy num, in Content Manager and to a key field, policy ID, in Content Manager ImagePlus for OS/390.

federated collection. A grouping of objects that results from a *federated search*.

federated datastore. Abstract representation of any number of specific *content servers*, such as Content Manager.

federated entity. An Enterprise Information Portal metadata object that is comprised of *federated attributes* and optionally associated with one or more *federated text indexes*.

federated search. A query issued from Enterprise Information Portal that simultaneously searches for data in one or more *content servers*, which can be heterogeneous.

federated text index. An Enterprise Information Portal metadata object that is mapped to one or more *native text indexes* in one or more *content servers*.

FFST™. See *First Failure Support Technology™*.

file system. In AIX, the method of partitioning a hard drive for storage.

First Failure Support Technology (FFST). Technology and software that provide first failure data capture for software events, and by automating event tracking and management.

folder. A container used to organize objects, which can be other folders or documents.

folder manager. The Content Manager model for managing data as online documents and folders. You

can use the folder manager APIs as the primary interface between your applications and the Content Manager content servers.

form element. In the dynamic page builder, an element of the user interface that can be created with HTML tags. The dynamic page builder supports several HTML form elements.

H

handle. A character string that represents an object, and is used to retrieve the object.

history log. A file that keeps a record of activities for a *workflow*.

HTML. See *Hypertext Markup Language*.

Hypertext Markup Language (HTML). A markup language that conforms to the SGML standard and was designed primarily to support the online display of textual and graphical information that includes hypertext links.

I

Image Object Content Architecture (IOCA). A collection of constructs used to interchange and present images.

index class. A category for storing and retrieving objects, consisting of a named set of attributes. When you create an object in the Content Manager system, your application must assign an index class and supply the *key field* values required by that class. An index class identifies the key fields, automatic processing requirements, and storage requirements for an object.

index class subset. A view of an *index class* that an application uses to store, retrieve, and display folders and objects.

index class view. The term used in the APIs for *index class subset*.

information mining. The automated process of extracting key information from text (summarization), finding predominant themes in a collection of documents (categorization), and searching for relevant documents using powerful and flexible queries.

interchange. The capability to import or export an image along with its index from one Content Manager ImagePlus for OS/390 system to another ImagePlus system using a *common interchange file* or *common interchange unit*.

IOCA. See *Image Object Content Architecture*.

item. Generic term for the smallest unit of information that Enterprise Information Portal administrators. An item

contains one or more item *parts*. Each item has an identifier. An item can be a *folder* or a *document*.

iterator. A class or construct that you use to step through a collection of objects one at a time.

J

JavaBeans. A platform-independent, software component technology for building reusable Java components called "beans." Once built, these beans can be made available for use by other software engineers or can be used in Java applications. Also, using JavaBeans, software engineers can manipulate and assemble beans in a graphical drag-and-drop development environment.

K

key field. See *attribute*.

L

LAN cache. An area of temporary storage on a local *object server* that contains a copy of objects stored on a remote object server.

library client. The component of a Content Manager system that provides a low-level programming interface for the library system. The library client includes APIs that are part of the software developer's kit.

library server. The component of a Content Manager system that contains index information for the objects stored on one or more *object servers*.

list. A data structure, or electronic queue, that your application uses to temporarily insert or remove *objects* for work. The *list manager* allocates lists in its file space for your application.

M

media archiver. A physical device that is used for storing audio and video stream data. The VideoCharger is a type of media archiver.

media server. An AIX-based component of the Content Manager system that is used for storing and accessing video files.

method. In object-oriented design or programming, the software that implements the behavior specified by an operation.

N

native attribute. A characteristic of an object that is managed on a specific *content server* and that is specific

to that content server. For example, the *key field* policy num might be a native attribute in a Content Manager content server, whereas the field policy ID might be a native attribute in an Content Manager OnDemand content server.

native entity. An *object* that is managed on a specific *content server* and that is comprised of *native attributes*. For example, Content Manager *index classes* are native entities comprised of Content Manager *key fields*.

native text index. An index of the text *parts* of a defined group of *items* that are managed on a specific *content server*. For example, a single text search index on a Content Manager content server.

network table file. A text file that contains the system-specific configuration information for each node in a Content Manager system. Each node in the system must have a network table file that identifies the node and lists the nodes that it needs to connect to.

The name of a network table is always FRNOLINT.TBL.

O

Object Linking and Embedding (OLE). A Microsoft® specification for both linking and embedding applications so that they can be activated from within other applications.

object server. The component of a Content Manager system that physically stores the *objects* or information accessed by client applications.

object server cache. The working storage area for the *object server*. Also called the *staging area*.

OLE. See *Object Linking and Embedding*.

overlay. A collection of predefined data such as lines, shading, text, boxes, or logos, that can be merged with variable data on a page during printing.

P

package. A collection of attributes, notifications, operations, or behaviors that are treated as a single module in the specification of a managed object *class*. Packages can be mandatory or conditional when referenced in a definition of a managed object class.

parametric search. A query for *objects* that is based on the *properties* of the objects.

part. A subelement of an *item* that corresponds to a discrete multimedia data object, such as an image, video, audio, or text. An item can contain parts with diverse binary formats, for example, an image, a word processing file, and a spreadsheet.

persistent identifier (PID). An identifier that uniquely identifies an *object*, regardless of where it is stored. The PID consists of both an item ID and a location.

PID. See *persistent identifier*.

privilege. The right to access a specific object in a specific way. Privileges includes rights such as creating, deleting, and selecting objects stored in the system. Privileges are assigned by the administrator.

privilege set. A collection of *privileges* for working with system components and functions. The administrator assigns privilege sets to users (user IDs) and user groups.

property. A characteristic of an *object* that describes the object. A property can be changed or modified. Type style is an example of a property.

Q

QBIC. See *query by image content*.

query by image content (QBIC). A query technology that enables searches based on visual content, called *features*, rather than plain text. Using QBIC, you can search for objects based on their visual characteristics, such as color and texture.

query string. A character string that specifies the properties and property values for a query. You can create the query string in an application and pass it to the query.

S

search criteria. In Enterprise Information Portal, specific fields that an administrator defines for a *search template* that limit or further define choices available to the users.

search template. A form, consisting of *search criteria* designed by an administrator, for a specific type of federated search. The administrator also identifies the *users* and *user groups* who can access each search template.

server definition. The characteristics of a specific content server that uniquely identify it to Enterprise Information Portal.

server inventory. The comprehensive list of native entities and native attributes from specified content servers.

server type definition. The list of characteristics, as identified by the administrator, required to uniquely identify a custom server of a certain type to Enterprise Information Portal.

staging area. The working storage area for the *object server*. Also referred to as *object server cache*.

streamed data. Multimedia data sent over a network connection at a specified rate. A stream can be video only, audio only, or a combination of both video and audio in one stream. Data rates, which are expressed in bits per second, vary for different types of streams and networks.

subclass. A *class* that is derived from another class. One or more classes might be between the class and subclass.

superclass. A *class* from which a class is derived. One or more classes might be between the class and superclass.

suspend. To remove an *object* from its *workflow* and define the suspension criteria needed to activate it. Later activating the object enables it to continue processing.

T

thin client. A client that has little or no installed software but has access to software that is managed and delivered by network servers that are attached to it. A thin client is an alternative to a full-function client such as a workstation.

Tivoli® Storage Manager (TSM). A client/server product that provides storage management and data access services in a heterogeneous environment. It supports various communication methods, provides administrative facilities to manage the backup and storage of files, and provides facilities for scheduling backup operations.

TSM. See *Tivoli Storage Manager*.

TSM management class. A logical area of storage that is managed by *Tivoli Storage Manager*.

U

user. In Enterprise Information Portal, anyone who is identified in the Enterprise Information Portal administration program as using the Enterprise Information Portal client.

user exit. (1) A point in an IBM-supplied program at which a user exit routine can be given control. (2) A programming service provided by an IBM software product that can be requested during the execution of an application program for the service of transferring control back to the application program upon the later occurrence of a user-specified event.

user exit routine. A user-written routine that receives control at predefined user exit points.

user group. A group consisting of one or more defined individual users, identified by a single group name.

user mapping. Associating Enterprise Information Portal user IDs and passwords to corresponding user IDs and passwords in one or more content servers. User mapping enables single logon to Enterprise Information Portal and multiple content servers.

V

volume. A representation of an actual physical storage device or unit on which the objects in your system are stored.

W

work item. A work activity that is active within a *workflow*.

work packet. A collection of *documents* that is routed from one location to another. Users access and work with work packets through *worklists*.

work state. The status of an individual *work item*.

workflow. A sequence of steps, and the rules governing those steps, through which a *work packet* travels while it is being processed.

For example, a process called `claims approval` would describe the steps that a *work packet* related to approving an insurance claim must follow.

workflow coordinator. User who receives notification that a *work item* in the *workflow* has not been processed in some specified time. The user is selected for a specific *user group* or upon creation of the workflow.

workflow state. The status of an entire *workflow*.

worklist. A collection of *work items* assigned to a user and retrieved from a workflow management system.

X

XDO. See *extended data object*.

Index

A

- access control lists, creating 25
- action list
 - defining 41, 44
 - predefined actions 44
- actions, defining 44
- administration client
 - creating
 - access control lists 25
 - privilege sets 24
 - search criteria 18
 - search templates 17
 - user groups 23, 24
 - users 23, 24
 - workflow 45
 - defining
 - action list 44
 - actions 44
 - worklists 43
 - description 6
 - optimizing 25
- administration database, description 6
- attributes
 - federated 12
 - native 11

C

- categories, Information Mining 28
- cmccc2mime.ini 51
- cmbmime2app 52
- CMBWebCrawlerService JavaBean 30
- configuring
 - image search 66
 - Information Mining feature 27
 - text search 59
- connectors, description 6
- connectors toolkit, description 7
- content server
 - attributes 11
 - defining 10
 - entities 11
- content servers
 - Content Manager
 - attributes 11
 - entities 11
 - features 53
 - Content Manager for AS/400
 - attributes 11
 - entities 11
 - Domino.Doc
 - attributes 11
 - entities 11
 - Domino Extended Search
 - attributes 11
 - entities 11
 - ImagePlus for OS/390
 - attributes 11
 - entities 11

- content servers (*continued*)
 - OnDemand
 - attributes 11
 - entities 11
 - relational databases
 - attributes 11
 - entities 11
- crawling the Web, starting 34
- customizing MIME types 51

D

- data types, supported 13

E

- Enterprise Information Portal
 - components 6
 - administration client 6
 - administration database 6
 - connector toolkit 7
 - connectors 6
 - image search 8
 - Information Mining feature 7
 - sample client application 7
 - text search 8
 - thin client samples 7
 - workflow builder 7
 - workflow feature 7
 - creating
 - access control lists 25
 - privilege sets 24
 - search criteria 18
 - search templates 17
 - user groups 23, 24
 - users 23, 24
 - workflow 45
 - defining
 - action list 44
 - actions 44
 - worklists 43
 - importing documents from the Web 37
- entities
 - federated 12
 - native 11

F

- federated attributes, creating and mapping 12
- federated text indexes 17

I

- IBM Web Crawler
 - description 30
 - environment variables, setting up 31
 - excluding from server 36
 - imycrawl command 35

- IBM Web Crawler (*continued*)
 - monitoring progress 36
 - workspace, defining 32
 - Windows Control Monitor 36

- image search
 - assigning an alias 68
 - configuring 66
 - description 8, 63
 - installing 63
 - setting up 66
 - updating shortcut 69
 - verifying the connection 68

- Information Mining
 - concepts 27
 - defining categories 28
 - importing documents from the Web 30
 - starting 27

- Information Mining feature
 - description 7

- installing
 - image search
 - on AIX 65
 - on Windows 63
 - text search
 - on AIX 59
 - on Windows 57

L

- loading sample data 71
- loading text and image search documents 73

M

- mapping federated attributes 12
- MIME type file
 - changing for clients 52
 - changing for servers 51

N

- Notices 81

O

- operators, supported 20
- optimizing administration client 25

P

- parameters, client
 - clientUrlString 78
 - csUrlString 78
 - mime2app 78
 - server 78
- parameters, sample client application
 - connectionType 77
- planning for
 - Enterprise Information Portal 10

- planning for *(continued)*
 - text search capacity 56
 - text search performance 57
- privilege sets, creating 24

S

- sample client application
 - configuring
 - options 77
 - description 7
 - parameters 77
 - starting 49
- sample loader program, running 73
- search criteria
 - defining and mapping 18
 - operators for 20
- search templates, creating 17
- server inventory 10
- shortcut
 - updating for image search 69
 - updating for text search 61
- starting
 - a Web crawl 34
 - administration client 9
 - Information Mining feature 27
 - sample client application 49
 - thin client samples 49
 - workflow builder 45

T

- text search
 - configuring 59
 - description 8, 55
 - installing on AIX 59
 - installing on Windows 57
 - planning 55
 - capacity 56
 - performance 57
 - setting up 60
 - updating shortcut 61
 - XML support 55
- thin client samples
 - description 7
 - starting 49

U

- user groups, creating 23, 24
- users, creating 23, 24

W

- webspaces, crawling 34
- work packet, description 40
- workflow
 - concepts 39
 - creating 45
 - planning for 40
- workflow builder
 - creating a workflow 45
 - description 42
 - starting 45
- workflow builder, description 7
- workflow feature
 - components 42

- workflow feature *(continued)*
 - configuring 39
 - description 7
- worklist
 - defining 43
 - description 41



Program Number: 5697-G29
5697-G31



Printed in the United States of America
on recycled paper containing 10%
recovered post-consumer fiber.

SC27-0875-00

