

**IBM Content
Manager for Multiplatforms**



Information Integrator for Content の管理

バージョン 8 リリース 2

**IBM Content
Manager for Multiplatforms**



Information Integrator for Content の管理

バージョン 8 リリース 2

お願い

本書、および本書で記述されている製品をご使用になる前に、147 ページの『特記事項』をお読みください。

本書は、IBM Enterprise Information Portal for Multiplatforms (製品番号 5724-B43) のバージョン 8 リリース 2 に適用されます。また、改訂版で特に断りのない限り、これ以降のすべてのリリースおよびモディフィケーションにも適用されます。

本マニュアルに関するご意見やご感想は、次の URL からお送りください。今後の参考にさせていただきます。

<http://www.ibm.com/jp/manuals/main/mail.html>

なお、日本 IBM 発行のマニュアルはインターネット経由でもご購入いただけます。詳しくは

<http://www.ibm.com/jp/manuals/> の「ご注文について」をご覧ください。

(URL は、変更になる場合があります)

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： SC27-1346-01
IBM Content
Manager for Multiplatforms
Managing Information Integrator for
Content
Version 8 Release 2

発行： 日本アイ・ビー・エム株式会社

担当： ナショナル・ランゲージ・サポート

第1刷 2003.2

この文書では、平成明朝体™W3、平成明朝体™W9、平成角ゴシック体™W3、平成角ゴシック体™W5、および平成角ゴシック体™W7を使用しています。この(書体*)は、(財)日本規格協会と使用契約を締結し使用しているものです。フォントとして無断複製することは禁止されています。

注* 平成明朝体™W3、平成明朝体™W9、平成角ゴシック体™W3、
平成角ゴシック体™W5、平成角ゴシック体™W7

© Copyright International Business Machines Corporation 1999, 2003. All rights reserved.

© Copyright IBM Japan 2003

目次

本書について	v
本書の対象読者	v
管理者に必要なスキル	vi
業務アナリストまたはプロセス・モデラーに必要なスキル	vi
本書の追加情報の入手先	vi
製品パッケージに含まれている情報	vi
Web 上で入手可能なサポート	vii
I EIP バージョン 8.2 の新機能	viii

第 1 章 Enterprise Information Portal

の紹介	1
カスタマー情報の検索	1
ニーズ	2
ソリューション	2
概説	2
Enterprise Information Portal コンポーネントの紹介	3

第 2 章 管理クライアントの紹介

EIP First Steps の使用によるシステム管理クライアントの理解	9
EIP の管理	9
ユーザーおよびグループの管理	9
管理クライアント・ツールの使用	9
特権の紹介	11
特権	11
製品ビューおよびデータベースの切り替え	12
管理クライアントの改善と機能強化	13
ローカル管理データベースへの管理クライアントの接続	13
リモート管理データベースへの管理クライアントの接続	14
I ステップ 1 - DB2 Configuration Assistant を使用してリモート・データベースをカタログする	14
I ステップ 2 - Server Configuration ユーティリティを使用する	16
ステップ 3 - リモート・データベース接続をテストする	17
文書タイプを定義する	18
サーバーの MIME タイプ・ファイル (cmbcc2mime.ini) の変更	18

第 3 章 EIP 管理クライアント・フィー

チャーの使用	19
Federated Search の作成	19
サーバーの定義	20
サーバー定義のガイドライン	21
OnDemand コネクタの使用: TCP/IP チューニングとソケット	26
Extended Search コネクタの使用	27

統合エンティティの作成	27
統合エンティティの理解	27
統合エンティティの作成ウィザードの使用	28
統合テキスト索引の作成	29
検索テンプレートの作成	29
検索テンプレートの定義 (Define search template)	30
検索基準の定義 (Define search criteria)	30
検索設定の決定 (Determine search settings)	30
特権の割り当て (Assign privileges)	31

第 4 章 ユーザー・アクセスの管理

ユーザー ID とパスワードの作成	33
I DB2 管理者権限について	34
I INI ファイルを使用した DB2 への接続	34
I リソース・マネージャーへのライブラリー・サーパーとシステム管理者のパスワードの変更	35
I データベース・アクセス・パスワードの変更	35
I LDAP からのユーザーのインポート	36
I 特権の紹介	37
特権セットの作成	38
特権グループの作成	38
ユーザーへの特権セットの割り当て	38
ユーザー ID への認可特権セットの割り当て	39
リソース・マネージャーへのユーザーの割り当て	39
コレクションへのユーザーの割り当て	39
ユーザー・グループの作成	39
アクセス制御リストの作成	40
アクセス制御リストへの特権セットの割り当て	40
ドメインの作成	40
ドメインの管理	41
ドメインへのアクセス	42
ドメインへのユーザーの割り当て	42
ドメインへのユーザー・グループの割り当て	42
ドメインへの特権セットの割り当て	42
ドメインへのリソース・マネージャーの割り当て	43
ドメインへのコレクションの割り当て	43
ドメイン間のユーザーの移動	43
ドメイン間のユーザー・グループの移動	44
ドメイン間のリソース・マネージャーの移動	44
ドメイン間のコレクションの移動	45
ドメイン間の特権セットの移動	45
ドメイン間のアクセス制御リストの移動	45

第 5 章 情報マイニングの管理

情報マイニングの概要	47
Enterprise Information Portal 情報マイニング・サービス	47
情報マイニング・サービスのコンポーネント	48
業務環境における情報マイニングの使用	50
情報マイニングの使用例	52
サポートされる言語およびフォーマット	55

概念	56
システム体系	56
情報マイニングの概念	57
情報マイニング・ツール	59
プログラミング・インターフェース	67
First Steps	68
分類法の作成	69
Information Structuring Tool のインストール	69
始めに	69
アクセス権限	70
分類法の定義	70
トレーニング文書の選択	72
トレーニング文書のアップロード	73
カテゴリ化モデルの評価	76
カタログのトレーニング	81
パフォーマンスのチューニング	81
IBM Web Crawler の使用	82
IBM Web Crawler の機能	82
Web 用の IBM Web Crawler の構成および実行	83
IBM Web Crawler 構成ファイル	86
IBM Web Crawler でのロギング	96
トラブルシューティング	97
要約機能の選択	98
Notes 用の IBM Web Crawler	100
サーバーからの IBM Web Crawler の除外	104
第 6 章 ワークフローの紹介	107
ワークフローの理解	107
ワークフローの使用法	107
ユーザー ID とグループの同期化	108
EIP サーバー (Workflow 使用可能) の再インス	
トール	109
MQSeries Workflow と EIP データベース間のユ	
ーザー ID とグループの更新	110
ワークフローの計画	111
処理する情報	111
情報を処理する方法	111
実行するアクション	112
プロセスでの情報の流れ	112
作業プロセス全体の統合	112
Enterprise Information Portal ワークフロー・コンポ	
ーネットの使用	113
ワークフロー・ビルダーの使用	113
ワークフロー・サービスの使用	114
ワーク・リストの定義	114
アクション・リストの定義	115
ワークフローの作成	115
ワークフロー・ビルダーの使用可能化	115
MQSeries Workflow サーバーの開始	116
第 7 章 IBM Web Crawler サンプル・	
ファイル	117

config-sample2.xml サンプル	117
IBM Web Crawler ログ分析ファイルの例	119

第 8 章 テキスト検索および QBIC® の使用

テキスト検索エンジンを用いた文書検索	123
Text Search Server の使用可能化	123
イメージ・コンテンツによる照会 (Query by Image Content (QBIC)) を用いたイメージ検索	124
イメージ検索の紹介	124
イメージ検索の設定	124
サンプル・データのロードおよび索引作成	127
サンプル・データをロードする前に	127
テキスト検索索引の作成	128
イメージ検索データベース、カタログ、およびフ	
ィチャーの作成	129
ローダー・プログラムの実行	130
サンプル・テキスト・データの索引作成	131

第 9 章 文書フォーマット

情報マイニングの文書フォーマット	133
ワード・プロセッシング: 一般	133
ワード・プロセッシング: DOS	133
ワード・プロセッシング: インターナショナル	134
ワード・プロセッシング: Windows	134
ワード・プロセッシング: Macintosh	135
スプレッドシート・フォーマット	135
データベース・フォーマット	136
標準グラフィック・フォーマット	136
ハイエンド・グラフィックス・フォーマット	138
プレゼンテーション・フォーマット	139
圧縮およびエンコードのフォーマット	139
その他	139

第 10 章 権限の管理

知的財産の保護	141
各種のマーク付け方法	142
可視マークの適用	143
非可視マークの適用	143

第 11 章 アクセス支援機能

キーボード入力およびナビゲーション	145
アクセス支援表示機能	145
支援テクノロジーとの互換性	146
アクセス支援のドキュメンテーション	146

特記事項

商標	149
----	-----

用語集

索引

本書について

本書では、Enterprise Information Portal (EIP) システムを管理するうえで理解しておく必要のあるすべての基本的概念を紹介しています。EIP には、管理クライアントから管理できるいくつかのコンポーネントが用意され、また、EIP を使用して他の製品の機能を利用することができるため、本書は、通常のシステム管理ガイドではありません。本書では、以下のトピックについてフォーカスし、その実行方法について説明します。

- 業務上のニーズを満たすための EIP の使用
- 管理クライアントへのアクセスおよびその使用方法
- ユーザー・アクセスの管理方法
- EIP を使用して、リレーショナル・データベースに保管された構造化データ、非構造化コンテンツまたはマルチメディア・コンテンツ、あるいはテキスト文書など、複数のコンテンツ・サーバー上のコンテンツを検索する方法
- ワークフローの設計、インプリメント、および管理の方法

本書の対象読者

本書は、EIP 管理者が以下の作業を行うのに役に立ちます。

システム管理

データベース、サーバー、およびネットワークの管理が含まれます。

ユーザー管理

個人およびグループへのアクセスの定義と認可、アクセス制御リストの保持を行います。

統合検索

統合検索テンプレートを定義して使用し、コンテンツ管理システムのコンテンツを検索します。

情報マイニング

文書から情報を抽出し、文書および検索結果をカテゴリー化します。

Web クロール

IBM® Web Crawler を使用して、Web のコンテンツを検索し、インポートします。

テキスト検索

IBM DB2® TIE または IBM Text Search Engine (Content Manager バージョン 7.1 以前のみ) を使用して、文書を検索し、索引を作成します。

イメージ検索

Content Manager バージョン 7.1 以前を使用して、イメージ検索を実行します。

ワークフロー管理

EIP ワークフロー・ツールを使用して、企業の情報ワークフローを管理します。

管理者に必要なスキル

実行する作業によって、以下のことを理解していなければなりません。

- ユーザー・アクセスのためのセキュリティー・プロトコル
- Windows NT[®]、Windows[®] XP、Windows 2000、AIX[®]、または Solaris オペレーティング・システム
- ネットワーク管理
- コンテンツ管理システムのコンテンツ・サーバーのデータ・モデル
- データベース管理
- 検索テンプレートの作成時に、コンテンツおよび検索基準の実用的な知識を適用する方法。
- 情報マイニング技法およびツール
- ワークフロー設計の原則
- EIP ワークフローを使ってサポートする業務プロセス

業務アナリストまたはプロセス・モデラーに必要なスキル

本書には、業務アナリストおよびプロセス・モデラーのための、企業の EIP ワークフローを定義およびモデル化する方法についての概念情報も記載されています。

Enterprise Information Portal ワークフロー・ビルダーを使用するには、以下のことを行う必要があります。

- 社内の業務プロセスで使用されるスタッフ要件、プログラム、およびデータ構造について理解する。
- 企業の業務またはワークフロー・プロセスについて決定する。

本書の追加情報の入手先

製品のパッケージには、システムの計画、インストール、管理、および使用に役立つ情報の完全なセットが含まれています。製品のドキュメンテーションおよびサポートは、Web 上でも利用可能です。

製品パッケージに含まれている情報

製品パッケージには、Information Center およびそれぞれの資料が PDF (.PDF) 形式で収められています。

Information Center

製品パッケージには Information Center が含まれています。これは、製品のインストール時にインストールすることができます。Information Center のインストールについての詳細は、「*Content Management System の計画とインストール*」を参照してください。

Information Center には、Content Manager、Enterprise Information Portal、および IBM Content Manager VideoCharger for Multiplatforms 用のドキュメンテーションが

含まれています。トピック・ベースの情報は、製品別およびタスク別 (たとえば、管理) に編成されています。付属しているナビゲーション機構と索引に加えて、検索機能も検索能力を補助します。

PDF 資料

各オペレーティング・システム用の Adobe Acrobat Reader を使って、PDF ファイルをオンラインで表示することができます。Acrobat Reader がインストールされていない場合は、Adobe の Web サイト (www.adobe.com) からダウンロードできます。

表 1 には、IBM Content Manager for Multiplatforms に含まれた Content Manager 資料が示されています。

表 1. Content Manager 資料

ファイル名	タイトル	資料番号
install	Content Management System の計画とインストール ¹	GC88-9200-01
migrate	Content Manager バージョン 8.1 へのマイグレーション	SC88-9202-01
sysadmin	システム管理ガイド	SC88-9201-01

IBM Content Manager for Multiplatforms を注文すると、IBM Enterprise Information Portal for Multiplatforms も受け取ります。あるいは、別個に IBM Enterprise Information Portal for Multiplatforms を注文することもできます。表 2 には、製品に含まれた Enterprise Information Portal 資料を示します。

表 2. Enterprise Information Portal 資料

ファイル名	タイトル	資料番号
apgwork	ワークステーション・アプリケーション・プログラミング・ガイド ¹	SC88-9205-01
ecliinst	e-client のインストール、構成と管理	SC88-9207-02
eipinst	Information Integrator for Content の計画とインストール	GC88-9203-01
eipmanag	Information Integrator for Content の管理	SC88-9204-01
messcode	メッセージとコード ²	SC88-9206-01

注:

1. 「ワークステーション・アプリケーション・プログラミング・ガイド」には、Content Manager および Enterprise Information Portal の両方のアプリケーションのプログラミングについての情報が含まれています。
2. 「メッセージとコード」には、Content Manager および Enterprise Information Portal 用のメッセージとコードが含まれています。

Web 上で入手可能なサポート

製品のサポートは、Web 上でも入手可能です。以下に示す製品の Web サイトから、**Support** をクリックします。

www.ibm.com/software/data/cm/

www.ibm.com/software/data/eip/

ドキュメンテーションは、製品に付随したソフトコピーに含まれています。 Web 上で製品のドキュメンテーションにアクセスするには、製品の Web サイトで **Library** をクリックします。

Enterprise Documentation Online (EDO) と呼ばれる、HTML ベースのドキュメンテーション・インターフェースも、Web から入手できます。これには現在、API 参照情報が含まれています。EDO へのアクセスについての情報は、Enterprise Information Portal Library Web ページを参照してください。

「 EIP バージョン 8.2 の新機能

この製品の変更点は次のとおりです。

Sun Solaris のサポート

Solaris システムにコネクタ、フィーチャー、およびデータベースをインストールすることができます。

共通システム管理

単一のクライアント・アプリケーションから Content Manager と Enterprise Information Portal 管理に別々にアクセスできるようになりました。

新しいコネクタ

- Content Manager バージョン 8 リリース 1 の ICM コネクタにより、Content Manager バージョン 8 の強力な文書ストレージ機能を利用できるようになりました。
- 新しい C++ 拡張検索バージョン 3.7 のコネクタは、AIX 環境で実行します。

コネクタの改善

- パラメトリック・テキスト検索が、統合層、および拡張検索の直接接続からサポートされるようになりました。
- OnDemand コネクタの機能上およびパフォーマンス上の改善点は次のとおりです。
 - OnDemand DDO の構造の変更
 - 非同期検索のサポート

新しい情報マイニング・サービス

- フィーチャーの抽出
- クラスタ化
- 言語識別

IBM Web Crawler

IBM Web Crawler は、Web 上の情報や Lotus Notes® データベース内の情報の検索および要約に役立つフィーチャーです。

ワークフローの拡張

AIX および Solaris でワークフローが完全サポートされるようになりました。

た。ワークフロー・ビルダー、API、および JavaBeans™ が提供するワークフロー機能が改善され、利用しやすくなりました。

Information Center

ブラウザー・ベースの Information Center に、Content Manager、Enterprise Information Portal、および IBM Content Manager VideoCharger for Multiplatforms™ についての文書が入っています。トピック・ベースの情報は、製品別およびタスク別 (たとえば、管理) に編成されています。付属しているナビゲーション機構と索引に加えて、検索機能も検索能力を補助します。

アクセス支援

アクセス支援機能は、運動機能に障害がある方や目が不自由な方など、身体に障害がある方にソフトウェア製品を支障なく使っていただくことを目的としたものです。この製品の主なアクセス支援機能は次のとおりです。

- マウスの代わりにキーボードを使用して、すべての機能を操作する能力
- 拡張表示プロパティのサポート
- ビデオおよびオーディオのアラート・キューのオプション
- 支援テクノロジーとの互換性
- オペレーティング・システムのアクセス支援機能との互換性
- アクセスを支援するドキュメンテーション・フォーマット

第 1 章 Enterprise Information Portal の紹介

保険会社や金融機関などの、書類を中心とした企業の多くは、業務関連の大量のコンテンツを管理します。業務情報の管理とアクセスのための社内ソリューションのニーズは、多くの業界に共通してあります。

コンテンツ・サーバー は、社内でコンテンツを処理したり利用したりするのに使えるメタデータと一緒に、マルチメディア、ビジネス・フォーム、文書、および関連データを保管するためのソフトウェア・システムです。別々のコンテンツ・サーバーを効率的に接続する方法がないと、情報の重複や、複数の検索を実行できるよう社員を研修するために、時間と経費を浪費することになりかねません。

Enterprise Information Portal (EIP) では、最先端のテクノロジーによって、すべての企業リソースをワークステーションのデスクトップ上で扱うことができます。EIP を利用すると、さまざまなコンテンツ・サーバーを単一のクライアントを通して接続することで、企業の情報およびマルチメディア・アセットの価値を最大化することができます。EIP クライアントを使用することで、ユーザーは、接続されたすべてのコンテンツ・サーバーへ素早く同時にアクセスできます。また、ユーザーは、情報マイニングを実行したり、複数のコンテンツ・サーバー (Web やイントラネットを含む) にまたがる拡張検索を実行することもできます。ユーザー定義のビジネス・プロセス内のワークフロー・タスクを実行することもできます。

EIP では、アプリケーションを企業の業務に合わせてカスタマイズできます。EIP サンプルを使用することにより、アプリケーション・プログラマーは、デスクトップ・アプリケーションおよび Web ベース・アプリケーションの両方を作成できます。

このセクションでは、EIP の概要を説明します。架空の保険会社である XYZ 保険のシナリオを用いて、EIP のフィーチャーと機能性をデモンストレーションします。

カスタマー情報の検索

XYZ 保険 (XYZ) は、大手の損害保険会社であり、写真、保険金支払請求、証券、査定人のメモ、専門家からの報告書、および他の書類を多量に蓄積しています。

XYZ 保険は、医療診断および判定のコンピューター・フォームと一緒に、保険契約者に送付したすべてのメモを Lotus® Domino™.Doc ファイル・キャビネットに保存しています。XYZ 社は、長期保存と迅速なアクセスのために、すべての保険申告書、通知書、および請求書を、Content Manager OnDemand サーバーに保存しています。XYZ 社は、保険契約者から受け取ったすべての保険金支払請求フォーム、写真、および書状を、Content Manager for iSeries システム・ホルダーに保管しています。XYZ 社は、DB2 Universal Database™ (DB2 UDB) データウェアハウスセンター・インフォメーション・カタログ・マネージャー内に、専門家からの報告書を保管しています。また XYZ 社は、広告、宣伝、および新規ビジネスの部署でも利用

できるよう、高解像度グラフィックスなどの企業メディア・アセットを Content Manager に保管しています。さらに、XYZ 社は、企業内手続きなどの情報をイントラネット上に保管しています。

ニーズ

保険金支払請求、顧客連絡、および保険契約者に対する一般サービスを 1 つのサーバーのコンテンツだけで満足に処理することはできません。担当者は、すべての顧客情報にアクセスする必要があるからです。顧客サービスを提供するには、担当者は、さまざまなコンテンツ・サーバーに同時にアクセスできなければなりません。XYZ 保険には、情報を検索して取り出せるようにコンテンツ・サーバーと企業イントラネットを接続するためのソリューションが必要です。さらに、ワークフロー・プロセッシングの使用を拡張することが必要です。

事務職から支払要求処理係や代理店にいたるまで、多種多様の社員が文書にアクセスする必要があります。XYZ 社は、ある項目ではアクセスを制限し、別の項目では無制限のアクセスを提供する必要があります。また XYZ 社には、研修の必要性を少なくするための使いやすいインターフェースも必要です。

ソリューション

XYZ 保険は、EIP を導入します。その理由は、その包括的な検索テクノロジーを利用して、すべてのコンテンツ・サーバーに接続して検索し、データを取り出すことができるからです。現在、XYZ 社の顧客対応センターの担当者は、電話連絡を受けてから、1 回の統合検索で保険契約者の必要情報をすべて取り出すことができます。

また XYZ 保険は、EIP 情報マイニング機能を使用して、社内イントラネットから情報を取り出します。さらに、ワークフロー・プロセスの使用を拡張することを検討しています。

概説

EIP は包括的な製品です。各コンポーネントが協働して、企業固有の要件に合わせたソリューションを提供します。複数層アーキテクチャーを中心とする EIP には、検索管理用の管理クライアント、検索実行用のクライアント、および IBM Content Manager、Content Manager ImagePlus[®] for OS/390[®]、Content Manager OnDemand、Lotus Domino.Doc、DB2 Universal Database、DB2 DataJoiner[®]、および DB2 データウェアハウスセンター・インフォメーション・カタログ・マネージャーなど、さまざまな別個のコンテンツ・サーバーに接続するためのコネクタが用意されています。EIP コネクタ・ツールキットおよびサンプルを使用することによって、追加のコンテンツ・サーバー用に、追加のコネクタを作成することもできます。

EIP のアーキテクチャーでは、クライアント・アプリケーションから 1 つまたは複数のコンテンツ・サーバーに対して単一検索を実行することができます。クライアントが検索を実行するには、EIP 管理者によって定義された検索テンプレートを使用します。

クライアントは、検索テンプレートを使って、統合検索を実行します。統合検索は、複数のコンテンツ・サーバーの固有属性を検索テンプレート内で使用される統合属性にマッピングした上で、それらのコンテンツ・サーバーに対して同時に実行

する検索です。EIP 検索テンプレートには、検索基準が入っています。これは、各コンテンツ・サーバーの固有の属性にマップされる統合属性を参照します。EIP 管理者は、検索テンプレートを作成します。EIP には、複数のコンテンツ・サーバーに保管されているデータにアクセスおよび検索するためのコネクタが用意されています。これによってコンテンツ・サーバーは、データ・オブジェクトをクライアントに戻します。

EIP アーキテクチャーには、以下の利点があります。

- 1 回の照会で、e-business トランザクションとカスタマー・サービス・アプリケーションをサポートする複数の異なるコンテンツ・サーバーにアクセスできます。
- 情報マイニング機能によって、Web を含む複数のコンテンツ・サーバーにアクセスできます。
- ワークフロー・プロセスによって、複数の異種コンテンツ・サーバーにわたって分散しているデータにアクセスできます。
- クライアント・アプリケーション開発サポート。クライアント・アプリケーション、索引、およびデータが別個になっているため、コンテンツ・サーバーでのデータの位置に依存していません。

Enterprise Information Portal コンポーネントの紹介

このセクションでは、EIP の各コンポーネントおよびインストール・オプションについて説明します。

3 ページの表 3 に、コンポーネントおよび互換性のあるオペレーティング・システムをリストします。

表 3. EIP コンポーネントと互換性のあるオペレーティング・システム

コンポーネント	Windows	AIX	Solaris	Notes
管理データベース	あり	あり	あり	データベースに、ワークフロー・ビルダーの機能を含む
管理クライアント	あり	なし	なし	クライアントは、Windows、AIX、または Solaris オペレーティング・システムにインストールされているデータベースに接続できる
コネクタ	あり	あり	あり	
情報マイニング	あり	あり	あり	
IBM Web Crawler	あり	あり	あり	
テキスト検索クライアント	あり	あり	あり	
イメージ検索クライアント	あり	あり	あり	

表3. EIP コンポーネントと互換性のあるオペレーティング・システム (続き)

コンポーネント	Windows	AIX	Solaris	Notes
コネクタ・ツール キットとサンプル	あり	あり	あり	<ul style="list-style-type: none"> Windows バージョンには、サンプル・クライアントをコンパイルするためのソース・コードが含まれる。AIX にはサンプル・クライアント・コードはインストールされていない ワークフロー・サンプルと API は統合コネクタ・サンプルと一緒にインストールされる
ビューアー	あり	なし	なし	OnDemand クライアントとビューアーをインストールする
Information Center	あり	あり	あり	

管理

管理コンポーネントは、管理データベースおよび管理クライアントのサブコンポーネントを提供します。管理データベースをインストールすると、ワークフロー・フィーチャーもインストールされます。

管理データベース: 管理データベースは、EIP ユーザーとグループについての情報、特権レベル、パスワード、ユーザー ID、およびその他の情報を管理する DB2 データベースです。データベースはまた、ワークフローも提供し、必要に応じて情報マイニング機能も提供します。複数のデータベースをインストールすることもできます。各データベースが EIP ワークフロー機能を提供します。Content Manager バージョン 8 がインストールされている場合は、EIP 管理データベースを Content Manager バージョン 8 のライブラリー・サーバー・データベースと共用できます。データベースを共用できる理由は、EIP によって要求される情報がすべてライブラリー・サーバー・データベースにあるためです。

管理クライアント: 管理クライアントをインストールできる環境は、Windows ワークステーションだけです。複数のクライアントをインストールすることもできます。Content Manager バージョン 8 システムがインストールされている場合は、同じクライアントから EIP と Content Manager バージョン 8 を管理できます。

クライアントが提供するインターフェースにより、管理者は次の作業が可能です。

- 各コンテンツ・サーバーを統合検索用に定義する

- コンテンツ・サーバー上で固有のエントティティと属性を識別し、それらを統合エントティティにマップする
- すべてのユーザー定義コンテンツ・サーバーに対する固有のエントティティと属性のインベントリーを保持する
- 検索テンプレートを作成する
- ユーザーおよびグループを識別し、管理する
- 特権と特権セットをユーザーに割り当てる
- 検索テンプレートへのアクセスを定義し、検索から得られた情報についてユーザーが実行できるアクションの条件を設定する
- ビジネス・ワークフロー・プロセスを設計し、管理する

コネクタ

コネクタは、EIP クライアント、コンテンツ・サーバー、および管理データベース間の通信インターフェースを提供します。Content Manager バージョン 7.1 コネクタなど、コンテンツ・サーバーのコネクタが提供する機能を利用することで、EIP はサーバーにログインし、情報を検索して、管理クライアントやエンド・ユーザー・クライアントにその情報を戻すことができます。統合コネクタは、管理クライアントを管理データベースに接続します。

EIP は、次のコネクタを提供します。

- 管理クライアントを管理データベースに接続する統合コネクタ。
- DB2 Universal Database 7.1、JDBC ドライバー 1.3 (Java™ のみ)、ODBC 3.0 (C++ のみ)、DataJoiner 2.1.1 用リレーショナル・データベース・コネクタ。
- Content Manager バージョン 7.1 サーバー用 Content Manager コネクタ。
- Content Manager バージョン 8.2 サーバー用 Content Manager コネクタ。
- Content Manager OnDemand バージョン 7.1 用 Content Manager OnDemand コネクタ。
- VisualInfo™ for 400® バージョン 4.3 およびバージョン 5.1 用 Content Manager。
- ImagePlus/390 Folder Application Facility バージョン 3.1、Image Plus/390 ODM バージョン 3.1 用 Content Manager ImagePlus for OS/390 コネクタ。
- Domino.Doc バージョン 3.0a および Desktop Enabler バージョン 3.0a 用 Lotus Domino.Doc コネクタ。
- バージョン 3.7 用拡張検索コネクタ。
- DB2 Universal Database Visual Warehouse™ バージョン 5.2 および DB2 Universal Database バージョン 7.2 用インフォメーション・カタログ・マネージャー・コネクタ。

フィーチャー

EIP には、4 つのオプション・フィーチャーがあります。

情報マイニング

情報マイニングは、コンテンツ・サーバー上のテキスト文書に隠された情報を検出するための、言語サービスを提供します。テキスト文書の処理時に、要約、カテゴリー化、および検索可能なメタデータを作成します。情報マイニングでは、WebSphere® Application Server 4.0 (スタンダード版またはアド

バinst版) がインストールされている必要があります。さらに、同類の文書をクラスター化したり、人名や会社名などの特徴を文書から抽出したり、文書の言語を判別したりすることができます。

イメージ検索クライアント

Content Manager バージョン 7 のコンテンツ・サーバー上のイメージ検索機能へのアクセスおよび管理に必要なインターフェースを提供します。

テキスト検索クライアント

テキスト検索サーバー上のテキスト検索機能へのアクセスおよび管理に必要なインターフェースを提供します。

IBM Web Crawler

Web Crawler は、Java ベースのコンテンツ・クローラーおよびマイナーです。Web Crawler は、イントラネット、エクストラネット、またはインターネットの Web 内、Lotus Notes データベース、データベース内のコンテンツ、あるいは Domino、ローカル・ファイル・システム、および FTP 接続を介したコンテンツのクロールが可能です。

Web Crawler は、さまざまなコンテンツのタイプからメタデータやテキストをマイニングすることができます。たとえば、HTML コンテンツは、URL、表題、本文、最後に修正した時刻、メタ・タグ (作成者、キーワード、説明など) でマイニングすることができます。ユーザーは、特定タイプのコンテンツに合わせて、事前定義されたマイナーのセットから選択できます。コンテンツおよび / またはマイニングされたメタデータは、ローカル・ディスクに保管されます。

コンテンツ・ビューアー

OnDemand ビューアーをインストールすると、OnDemand クライアント、および OnDemand サーバーから検索した文書の表示に必要なその他のファイルがインストールされます。

コネクター・ツールキットとサンプル

EIP はコネクター・ツールキットを提供しており、これには、以下のようなさまざまな EIP 機能を試したりテストするために使用できるサンプル・プログラムが含まれています。

- コンテンツ・サーバーへの接続および切断
- コンテンツ・サーバー上での SQL および他のサンプル照会の実行
- コンテンツ・サーバー MIME タイプなどの判別

Windows コネクター・ツールキット: Windows サーバー上にコネクター・ツールキットおよびサンプルをインストールするには、「開発用ワークステーション (Development Workstation)」マシン・タイプを選択する必要があります。次に、コネクター・ツールキットとサンプル・コンポーネントを選択します。すべてのコネクター用にサンプル・プログラムをインストールすることも、インストール済みコネクターに適合するサンプルを個別に選択することもできます。

Windows サーバーの場合は、以下のようにコネクター・ツールキットのサンプル・プログラムが編成されます。

```
c:%CMBROOT%\SAMPLES\activex\%xx
```

```
c:%CMBROOT%\SAMPLES\cpp\%xx
```

```
| c:¥CMBROOT¥SAMPLES¥java¥xx  
| c:¥CMBROOT¥SAMPLES¥jsp¥xx  
| c:¥CMBROOT¥SAMPLES¥server¥xx
```

ここで、*xx* は該当するコネクタごとにサンプル・プログラムが入っているディレクトリー名です (たとえば、db2、od、d1 など)。

AIX サーバーの場合は、以下のようにサンプル・プログラムが編成されます。

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

ここで、*xx* はサブディレクトリー名です (beans、servlets、など)。

Solaris サーバーの場合は、以下のようにサンプル・プログラムが編成されます。

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

ここで、*xx* はサブディレクトリー名です (beans、servlets、など)。

サンプル・プログラムは、プログラムを記述し、サンプル・コードを処理するために必要なサーバー設定値 (環境設定値、メモリー、など) を規定する文書を含みます。

Information Center

Information Center コンポーネントには、Enterprise Information Portal Information Center が含まれます。Information Center は、Enterprise Information Portal ライブラリーの Web ベースの検索可能なバージョンです。

第 2 章 管理クライアントの紹介

管理クライアントは、EIP 管理データベースと EIP 管理者の間のインターフェースを提供します。このセクションでは、クライアントが提供する、EIP システムの管理に役立つ多くのフィーチャーと機能について説明します。

クライアントの左側のペインにあるアイコンから、サーバー定義およびユーザー管理など、いくつかのフィーチャーと機能にアクセスできます。その他の機能にアクセスするには、「ツール (Tools)」メニュー・バーを使用します。

EIP First Steps の使用によるシステム管理クライアントの理解

First Steps は、EIP をインストールする際に必ず組み込まれるモジュールです。*First Steps* ではサンプル・データが提供され、オブジェクトが用意されるため、ユーザーは実データを使用する必要がありません。管理クライアントの基本構造やロック・アンド・フィールドなどの理解に役立てるために、サーバー定義、ユーザーおよびグループ、その他のフィーチャーなどを調べるには、*First Steps* を使用してください。

EIP の管理

システム管理者は、管理クライアントを介して、以下の 1 つまたは複数のタスクを実行することができます。

- コンテンツ・サーバーの定義
- ユーザーおよびグループの管理
- 特権およびアクセス・レベルの管理
- 統合検索テンプレートの作成
- 統合エンティティの作成
- サブドメインの作成 (管理可能ドメインが使用可能な場合)
- ワークフローの操作 (ワークフローが使用可能な場合)
- Content Manager バージョン 7 での統合テキスト・エンティティの作成

ユーザーおよびグループの管理

管理者がユーザー ID と特権を作成することで、ユーザーは複数のコンテンツ・サーバー上の文書を検索したり、操作したりできるようになります。システムに保管されているデータへのアクセスを制限するには、適切な特権を定義してそれをユーザーに割り当てる必要があります。

管理クライアント・ツールの使用

このセクションでは、管理クライアントによって提供されるツールについて説明します。

LDAP 構成

このオプションをクリックすると、EIP は、以下の 4 つのタブが入っているウィンドウを立ち上げます。

- 「LDAP」タブ - LDAP サーバーからのデータ・ソースのインポートを使用可能にしたり、LDAP ユーザーのインポートと認証を使用可能にしたり、あるいはこの両方を選択できます。
- 「サーバー (Server)」タブ - ホスト名、ユーザー名、参照タイプなどの、各種の LDAP サーバー仕様を定義するためのフィールドが含まれています。
- 「認証 (Authentication)」タブ - SSL 情報を定義するためのフィールドが含まれています。
- 「アドバンスド (Advanced)」タブ - 最大レコードおよびサーバー・タイムアウトに関する設定を定義します。

ユーザー・マッピング・オプション

このオプションを使用すれば、ユーザー・マッピングを使用可能にするデフォルト設定を、使用不可にすることができます。

統合ユーザー・マッピング・エディター

統合ユーザー・エディターにはユーザーのリストが表示されるので、特定のコンテンツ・サーバーにユーザーをマッピングすることができます。

検索テンプレート・ビューアー

検索テンプレート・ビューアーは、すべての検索テンプレートに関する詳細情報を提供します。このビューアーでは、検索テンプレートの詳細を表示する以下の 3 つのオプションが設けられています。

- 関連マッピング (デフォルト) - 統合エンティティに関する詳細情報や検索テンプレートに関するその他の詳細情報を提供します。
- 検索テンプレート - デフォルト演算子やデフォルト値などに関する詳細情報を提供します。
- 表示結果 - 表示名、表示幅、基準の順序などに関する詳細情報を提供します。

サーバー・インベントリー・ビューアー

選択したサーバー (複数も可) のインベントリーを表示します。

ログ・ビューアー

ログ・ビューアーを使用して、サーバー・インベントリーのリフレッシュ後に生成されたログを表示します。新しいインベントリーと前のインベントリーとが異なる場合は、ログにメッセージのリストが表示されます。

サービス

ワークフローおよび / または情報マイニングを使用可能にするサービスを選択します。

管理可能ドメイン

管理可能ドメインを使用可能にするために、管理可能ドメインを選択します。管理可能ドメインは、いったん使用可能にすると、使用不可にすることはできません。

MIME タイプ・エディター

MIME タイプ・エディターは、以下の情報を各コンテンツ・サーバーごとにリストします。

- コンテンツ・クラス
- ファイル拡張子
- リレーショナル・データベース (RDB) 列
- MIME タイプ

MIME タイプ・エディターにリストされているコンテンツ・サーバー名は省略形であり、新規のコンテンツ・サーバーを定義する際に表示されるコンテンツ・サーバー名のリストに対応しています。**ヒント:** DL は、Content Manager バージョン 7.1 コンテンツ・サーバーの省略語です。V4 は、Content Manager for AS/400® コンテンツ・サーバーの省略語です。

MIME タイプ・エディターでは、デフォルト情報への追加、デフォルト情報の除去、および編集が可能です。

MIME - アプリケーション・エディター

5 つのデフォルトの MIME - アプリケーション・アソシエーションへの追加、それらのアソシエーションの削除、または編集を行う場合は、MIME - アプリケーション・エディターを使用します。MIME - アプリケーション・エディターで定義した値や設定値は、エンド・ユーザー・クライアントが使用するビューアーに影響します。

サーバー・タイプ定義

システム・プログラマーが開発したあらゆるカスタム・サーバーを定義する場合は、このツールを使用します。

DB2 ID/ パスワードの変更

接続専用の DB2 ユーザー ID およびパスワードを変更する場合はこのオプションを選択します。ここで指定するのは、管理者のユーザー ID とはまったく別のユーザー ID です。

特権の紹介

このセクションでは、EIP の特権について説明します。4 つの特権にアクセスするには、「権限 (Authorization)」アイコンを展開します。**ヒント:** Content Manager バージョン 8 と EIP バージョン 8 は同じクライアントから管理できるため、クライアントでは、クライアントの Content Manager 側、EIP 側それぞれに、両方の特権についてのすべての特権が表示されます。

特権

管理クライアントは、定義済み特権、特権グループ、および特権セットを提供します。特権は、システム・ユーザー (管理者およびクライアントのエンド・ユーザーなど) に、ある種のオブジェクトをある種の方法で処理する権限を与えます。

特権 EIP は複数の特権を提供します。特権は、ある種のオブジェクトを特定の方法で処理する権限です。たとえば、クライアントのエンド・ユーザーに対して ItemAdd および ItemDelete 特権を割り当てることができます。これによって、コンテンツ・サーバーに項目を追加または削除する権限を与えられま

す。特権を表示するには、「認証 (Authentication)」アイコンを展開して、「特権 (Privileges)」をダブルクリックしてください。特権を作成するには、「特権 (Privileges)」を右マウス・ボタン・クリックして、「新規 (New)」をクリックしてください。

特権グループ

EIP はデフォルト特権グループを提供します。特権グループは、関連した特権の集合です。たとえば、Administer EIP という名前の特権グループには、EIP システムの管理に関連した、次のような典型的な 5 つの特権があります。

- EIPAdminServer
- EIPAdminEntity
- EIPAdminTextEntity
- EIPAdminTemplate
- EIPAdminInfoMining

特権グループに割り当て済みの特権を表示または変更するには、「認証 (Authentication)」アイコンを展開して、特権グループ名をダブルクリックしてください。特権グループを作成するには、「特権グループ (Privilege Groups)」を右マウス・ボタン・クリックして、「新規 (New)」をクリックしてください。

特権セット

EIP はいくつかのデフォルト特権セットを提供します。特権セットは、ユーザーの役割を定義する特権の集合です。たとえば、ClientUserCreateAndDelete という名前の特権セットには、クライアントのエンド・ユーザーの役割 (Delete (項目の削除)、ItemAdd (項目の追加)、など) に関連した 17 の特権があります。クライアント・ユーザー用にユーザー ID を作成して、特権セット ClientUserCreateAndDelete を割り当てた場合、そのユーザーはコンテンツ・サーバーにログオンして、特権セットに含まれる 17 のユーザーの役割をすべて実行できます。特権セットを表示または変更するには、「認証 (Authentication)」アイコンを展開して、「特権セット (Privilege Sets)」をクリックし、特権セット名をダブルクリックしてください。特権セットを作成するには、「特権セット (Privilege Sets)」を右マウス・ボタン・クリックして、「新規 (New)」をクリックしてください。

ユーザーやユーザー・グループを管理する場合、特権セットをユーザーやユーザー・グループに関連付けます。特権セットをユーザー・グループに割り当てると、そのグループ内のすべてのユーザーは、関連した特権セットに含まれるすべての役割を実行できます。

製品ビューおよびデータベースの切り替え

企業ソリューションの一部として Content Manager および Enterprise Information Portal を使用している場合は、1 つのユーザー・インターフェースから両方のシステム管理クライアントにアクセスすることができます。以前は、両方の製品がインストールされている場合、2 つのクライアントを別々に開かなければなりません。

した。1 つのクライアント・ビューから別のクライアント・ビューへ切り替えると、両方のクライアントに適用される情報の変更が容易になり、また、どちらの製品にも迅速にアクセスできます。

ログオフせずに EIP の管理から Content Manager の管理に切り替えるには、メイン・システム管理ウィンドウを表示し、左側のペインの上部にあるプルダウン・メニューで Content Manager を選択します。

統合データベース間で切り替えを行うには、クライアント・ウィンドウの左側のペインで、統合データベースのアイコンをダブルクリックします。

また、別のデータベースを管理する場合も、クライアントを終了して新しいデータベースにログインする必要はありません。管理クライアントには、cmbds.ini ファイルにリストされているすべての管理データベースのアイコンが表示されます。別のデータベースに切り替えるには、アイコンをクリックするだけです。クライアントにログインしたときに入力したユーザー ID と新しいデータベースのユーザー ID が異なる場合は、別のユーザー ID の入力を求めるプロンプトが表示されます。

管理クライアントの改善と機能強化

EIP バージョン 8.2 では、EIP 管理クライアントに対して、次のような重要な機能強化が行われています。

ウィザードおよびダイアログの改善

新しいダイアログは、ユーザー管理をさらに容易にします。新しいウィザードは、統合エンティティおよび検索テンプレートの定義と変更をさらに容易にします。ユーザーは、EIP バージョン 7.1 でサポートされているダイアログの使用を選ぶこともできます。

管理クライアントの共用

EIP バージョン 8.2 と Content Manager バージョン 8.2 を同じシステム上にインストールすると、これら 2 つの製品は 1 つの管理クライアントを共用します。両方の製品の管理者である場合は、クライアントに一度ログオンすると、クライアント内で 2 つのアプリケーションを切り替えることができます。管理データベースを切り替えるときも、いったんログアウトしてから再度ログインする必要はありません。

ドメイン管理者

ドメイン管理者を作成し、このドメイン管理者に、定義されたドメインに対する管理特権のみを持たせることができます。

シングル・サインオンと LDAP サポート

EIP が Windows Active Directory および LDAP を使用するようになったことにより、ユーザーは、シングル・サインオンで複数のコンテンツ・サーバーにアクセスできるようになりました。

ローカル管理データベースへの管理クライアントの接続

管理クライアントをインストールするサーバーと同じサーバー上に管理データベースをインストールする場合、ローカル・クライアントとサーバーを接続するために必要な情報は、データベース接続情報を保管する cmbds.ini ファイル内にすでに保管されています。ユーザーがインストール後に構成する必要はなく、このセクショ

ンの手順を使用して即時に接続できます。**要件:** EIP Database Install ユーティリティーを使用して追加のローカル・データベースを作成する場合、新しいデータベースに接続する前に、手作業で cmbds.ini の必要な情報を変更する必要があります。

1. 「スタート」 --> 「プログラム」 --> 「IBM Enterprise Information Portal for Multiplatforms 8.2」 --> 「管理 (Administration)」をクリックします。
2. 「サーバー (Server)」フィールド内のドロップダウン・リストから、ローカル・データベースを選択します。
3. 管理者のユーザー ID とパスワードを入力し、「OK」を押します。
4. システム管理クライアントがオープンされます。**ヒント:** EIP First Steps を使用した場合は、サンプル・データベースがクライアントの左側のペインに表示されます。

リモート管理データベースへの管理クライアントの接続

EIP 管理クライアントを AIX、Windows または Solaris データベースに接続するには、次の 2 つの方法があります。

- RMI サーバーを介して接続する (詳しくは、「*Information Integrator for Content* の計画とインストール」を参照)。
- DB2 Configuration Assistant を使用してデータベースをカタログし、次に EIP Server Configuration ユーティリティーを使用してサーバー接続パラメーターを定義することにより、接続を定義する。ユーティリティーは、データベース・スキーマ名、別名、オペレーティング・システムなどの情報を、cmbds.ini という名前のファイルにコピーします。ユーザーがシステム管理クライアントを起動すると、ログインできるサーバーのリストが cmbds.ini 内で定義済みのサーバーから取り出されます。

要件: リモート・データベースは、それぞれ別個にカタログする必要があります。管理クライアントから接続できるリモート・データベースは、cmbds.ini ファイルにリストされているデータベースのみです。

ヒント: 経験のあるユーザーは、Server Configuration ユーティリティーの手順をスキップして、cmbds.ini をテキスト・エディターで変更してもかまいません。cmbds.ini へのデフォルトのパスは、C:¥Program Files¥IBM¥CMgmt です。

重要: 接続したいリモート・データベースのデータベース・カタログ値を、この製品をインストールした担当者がすでに構成した場合は、そのデータベースの DB2 CCA ステップを実行する必要はありません。しかし、そのインストール担当者がデータベース・カタログ値を入力しなかった場合、または別のリモート・データベースにも接続したい場合は、DB2CCA を使用して cmbds.ini ファイルの接続パラメーターを追加のデータベース用に変更する必要があります。

ステップ 1 - DB2 Configuration Assistant を使用してリモート・データベースをカタログする

DB2 Configuration Assistant (CCA) は、リモート EIP データベースを DB2 内にカタログします。DB2CCA を使用してリモート・データベースをカタログするに

は、リモート・サーバーのホスト名、データベース名およびデータベース・インスタンス・ポート番号を知っている必要があります。また、リモート・データベースの別名を定義しておく必要があります。

ステップ 1a ~ 1f では、データベース名、スキーマ名および接続ポート番号を探す方法を説明します。名前とポート番号を構成し、管理クライアントとリモート・データベース間の接続を構成するには、名前と接続ポート番号を知っている必要があります。

1. リモート・データベース接続情報を探します。

- a. リモート AIX、Windows または Solaris サーバーに、DB2 管理権限をもつユーザー ID でログインする。
- b. `db2 list db directory` と入力する。
- c. 接続したい管理データベースの名前を選択する。インスタンスが異なると接続ポート番号も異なるので、データベースがインストールされている db2 インスタンスを書き留めておく。
- d. `db2 connect to <database> user <userID> using <password>` と入力する。
- e. `db2 list tables` と入力し、データベース・スキーマ名 (Server Configuration ユーティリティに必要) を書き留めておく。
- f. リモート管理データベースに関連した接続ポート番号を見つける。

Windows の場合

- 1) リモート Windows サーバー上で DB2 Control Center をオープンする。
- 2) ローカル・マシン用に使用可能なインスタンスの 1 つを右クリックする。
- 3) 「通信のセットアップ... (Setup Communications...)」を選択する。
- 4) TCP/IP 選択項目の右にある「プロパティ (Properties)」ボタンを選択する。ポート番号がウィンドウ上にリストされます。

AIX または Solaris の場合

- 1) `cd /usr/etc` を入力する。
- 2) `cat services` を入力する。
- 3) サービスのリストをスクロールして、リモート・データベースのデータベース・インスタンス用接続ポート番号を見つける。たとえば、データベースが db2inst1 上にインストールされている場合、接続ポートは 50000。

2. DB2 Configuration Assistant を使用して、リモート・データベースをカタログします。詳しくは、DB2CCA ヘルプ・ファイルを参照してください。

- a. 管理クライアントがインストールされている Windows サーバーにログインする。すべての DB2ADM 特権をもつユーザー ID でログインする必要があります。
- b. 「スタート」->「プログラム」メニューから、DB2 Configuration Assistant にナビゲートする。
- c. DB2 Configuration Assistant のプロンプトに従って、リモート・データベースをカタログし、接続をテストする。

- d. DB2 CCA 接続テストが成功した場合は、『ステップ 2 - Server Configuration ユーティリティを使用する』の手順に従うか、または cmbds.ini ファイルを直接変更して、cmbds.ini に保管されているリモート・データベース接続パラメータを定義する。

ステップ 2 - Server Configuration ユーティリティを使用する

Server Configuration ユーティリティのプロンプトでは、リモート・データベースについての接続情報 (ポート番号、ホスト名、など) の入力が求められ、それらのデータは cmbds.ini に保管されます。

1. 「スタート」 --> 「プログラム」 --> 「IBM Enterprise Information Portal for Multiplatforms」 --> 「Server Configuration ユーティリティ」をクリックします。
2. 各フィールドに情報を入力します (表 4 を参照)。

表 4. Server Configuration ユーティリティ

フィールド	情報	注
サーバー	データベース・タイプ (Content Manager または EIP) を選択する。	サーバーはデータベース・タイプを指し、データベースがインストールされているサーバーの名前ではない。 ヒント: Content Manager および EIP 管理クライアントが同じマシン上にある場合のみ、管理クライアントを使用して両方のデータベース・タイプを管理できる。
サーバー名	接続中のデータベースの別名を入力する。要件: DB2CCA で定義済みの別名と同じものを使用する必要がある。	別名は、ワークステーション上のリモート・データベースを識別する固有の名前を規定する。別名の長さは 8 文字以内という制限がある。たとえば、リモート・データベース名が ICMNLSDB の場合、別名を REMOTE1 にできる。
スキーマ名	リモート・データベースが作成されたときに割り当てられたスキーマ名を入力する。	ICMADMIN が EIP および Content Manager データベース用のデフォルトのスキーマ名。
ホスト名	リモート・データベースがインストールされているコンピューターの名前を入力する。	完全修飾ホスト名またはリモート・データベースがインストールされているコンピューターの IP アドレスを入力する。
オペレーティング・システム	ドロップダウン・ボックスからオペレーティング・システムを選択する。	AIX、Sun Solaris または Windows を選択する。OS/390 オプションは、EIP 8.2 では機能しない。
ポート番号	リモート・データベースに割り当てられているポート番号を入力する。	Windows、AIX および Solaris 上にインストールされている EIP および Content Manager データベースの場合、50000 がデフォルトの接続ポート番号。

表 4. Server Configuration ユーティリティー (続き)

フィールド	情報	注
リモート・データベース名	リモート・データベースの名前を入力する。大文字を使用する。	ICMNLSDDB が EIP および Content Manager データベース用のデフォルトの名前。
ノード名	リモート EIP または Content Manager データベースのノード名を入力する。	ノード名は、リモート・データベースに割り当てられている固有の名前で、リモート・データベース用に作成する別名と類似している。Windows、AIX または Solaris サーバー上にインストールされているデータベースのノード名は、以下の手順で見つけることができる。 a. DB2 コマンド行セッションを開く。 b. db2=> プロンプトで、LIST NODE DIRECTORY と入力する。 c. DB2 によって、リモート・サーバー上にインストールまたは定義されている全データベースのノード名と他のデータが表示される。
シングル・サインオンの使用可能化	シングル・サインオンがデータベースのインストール時に使用可能にされた場合は、クリックする。	デフォルトの設定では、チェックされない (使用不可)。
セキュリティ・オプション	このオプションがデータベース作成時に選択された場合は、クライアント認証をクリックする。	デフォルトの設定値は Server。

3. 「OK」をクリックします。

ステップ 3 - リモート・データベース接続をテストする

1. 管理クライアントがインストールされている Windows サーバーにログインします。
2. 「スタート」 --> 「プログラム」 --> 「IBM Enterprise Information Portal for Multiplatforms 8.2」 --> 「管理 (Administration)」をクリックします。
3. 「サーバー (Server)」フィールド内のドロップダウン・リストから、リモート・データベースの別名を選択します。その名前は、Server Configuration ユーティリティーおよび DB2 Configuration Assistant で定義した別名と一致します。
4. リモート・データベースに関連したユーザー ID とパスワードを入力します。
5. 「OK」をクリックします。管理クライアントがオープンされます。

文書タイプを定義する

EIP では、いくつかの文書タイプをサポートするビューアーを提供しています。サーバーに対して文書タイプを定義すると、そのネイティブ・アプリケーションで文書を立ち上げることができます。たとえば、Content Manager OnDemand サーバーで Lotus Word Pro® 文書を保管する場合、クライアント文書ビューアーではなく、Lotus Word Pro で .lwp 拡張子の文書を立ち上げるように EIP を設定することができます。

文書タイプを定義するには、CMBROOT ディレクトリー (通常、x:%Program Files%IBM%CMgmt) 内の cmbcc2mime.ini ファイルを変更します。このファイルには、カスタム MIME 定義を作成する方法の説明があります。このファイルは、コンテンツ・クラスを MIME タイプ・ストリームに変換し、クライアントがコンテンツ・サーバーからコンテンツを読み取れるようにします。

重要: MIME タイプに基づいてアプリケーションを立ち上げる場合、基本オブジェクトのみが表示されます。文書に加えられたマークアップは表示されません。文書に複数のパートがある場合は、最初のパートだけが表示されます。両方のファイルの MIME タイプは一致しなければなりません。

サーバーの MIME タイプ・ファイル (cmbcc2mime.ini) の変更

サーバー MIME タイプを追加する場合、追加する文書タイプがそのファイル用に作成された MIME タイプかどうか検査してください。詳しくは、Web サイト <ftp://ftp.isi.edu/in-notes/iana/assignments/media-types> を参照してください。

cmbcc2mime.ini ファイルに値を追加するには、次のステップを実行します。

1. テキスト・エディターで cmbcc2mime.ini を開きます。
2. ユーザー定義の値には、以下の形式を使用します。
 - コンテンツ・クラスは 4096 で開始
 - コンテンツ・クラスの値の後に等号 (=) が続く
 - MIME タイプが等号の後に続く。これがそのコンテンツ・クラス用の標準 MIME タイプでない場合、次のステップに従ってください。
 - a. MIME タイプはタイプとサブタイプから構成されます。有効なタイプは、アプリケーション、テキスト、イメージ、モデル、メッセージ、オーディオ、およびビデオです。
 - b. スラッシュ (/) がタイプに続きます。
 - c. サブタイプを作成するには、その文書に使用されるトークンの前にトークン (x-) を付けなければなりません。以下に例を示します。

```
x-mydocumentclass (4096=application/x-mydocumentclass)
```
 - 新しい MIME タイプごとに、2b と 2c を必要に応じて繰り返します。

ヒント: OnDemand コンテンツ・サーバーは、コンテンツ・クラスの数値ではなく、ファイル拡張子を MIME タイプ・ストリームにマップします。

第 3 章 EIP 管理クライアント・フィーチャーの使用

このセクションでは、EIP 管理者によって実行される共通タスクのいくつかについて説明します。

Federated Search の作成

統合検索は、クライアント・アプリケーションから発行される照会であり、1 つまたは複数のコンテンツ・サーバーを同時に検索します。EIP には、統合検索用の検索テンプレートを作成するツールが用意されています。コンテンツ・サーバーはそれぞれ異なる方法で情報の保管と編成を行っているため、検索テンプレートは各サーバーごとのこれらの違いを明らかにしなければなりません。検索テンプレートは、コンテンツ・サーバーを検索するために、統合エンティティとその統合属性を固有属性にマップします。

統合検索の作成には、次の作業が必要になります。

- EIP コネクターを使用するコンテンツ・サーバーへの接続を定義すること
- 統合エンティティを作成すること
 - 統合エンティティの定義
 - 統合属性の作成
 - 統合属性の固有属性へのマッピング
 - パラメーターの割り当て
- 検索テンプレートを作成すること
 - 検索テンプレートの定義
 - 検索基準の定義
 - テンプレート設定の定義
 - クライアント・ユーザーへのアクセスの割り当て

EIP バージョン 8.2 で提供されている 2 つのウィザードにより、統合エンティティおよび検索テンプレートの作成が、さらに容易になりました。統合エンティティ・ウィザードには、フィルター操作により固有属性の検索を容易にするサーバー・インベントリーが組み込まれています。統合エンティティ・ウィザードはまた、統合属性の有効なデフォルト・パラメーターを生成して、統合属性の構成を誤る可能性が低減します。検索基準の作成には、検索テンプレート・ウィザードが便利です。このウィザードでは、検索基準と結果を表示する場合の外観と動作を設計することもできます。また、クライアント・アプリケーションで検索テンプレートの外観をプレビューすることもできます。さらに、必要であれば、EIP バージョン 7.1 の統合エンティティと検索テンプレートの作成ダイアログも利用できます。

すべてのウィザード、ダイアログ、およびフィールドは、EIP オンライン・ヘルプに文書化されています。

サーバーの定義

事前にサーバーを定義しておかないと、サーバーに接続してサーバー・インベントリーを実行することはできません。サーバーのアイコンを右マウス・ボタン・クリックして「新規 (New)」をクリックすると、EIP でサポートされているすべてのコネクタがクライアントに表示されます。サーバーを定義する前に、以下のような、コネクタに関するいくつかの基本情報を把握しておく必要があります。

- インストーラーはどのコネクタを選択したか? インストールされたコネクタは、`cmbcs.ini` 構成ファイルにリストされます。Windows サーバーでのデフォルト・パスは `x:%Program Files%IBM%CMgmt` です。`cmbcs.ini` ファイルの位置については、AIX または Solaris の管理者にお問い合わせください。
- インストーラーはローカル・コネクタ・オプションまたはリモート・コネクタ・オプションのどちらを選択したか? `cmbcs.ini` ファイルに、ローカルまたはリモートのコネクタ・タイプが入っています。
- ユーザーのシステムが RMI 用に構成されている場合、RMI サーバーは開始されているか? ローカル RMI サーバーで RMI を始動するには、「スタート」→「プログラム」→「IBM Enterprise Information Portal for Multiplatforms 8.2」→「RMI サーバーの始動 (Start RMI servers)」を使用します。システムがリモート RMI を使用している場合は、`cmbsvclient.ini` を参照して、RMI コネクタがインストールされているリモート・サーバーを探してください。詳しくは、RMI サーバーの管理者にお問い合わせください。
- EIP をインストールしたマシンに CM for AS/400 コネクタが組み込まれている場合、`frnolint.tbl` というネットワーク・テーブルにどのような情報が組み込まれたか? AS/400 の `frnolint.tbl` は `%CMBROOT%` に入っています。
- リレーショナル・データベース (Content Manager バージョン 8 と DB2、DataJoiner と Information Catalog など) が含まれているリモート・コンテンツ・サーバーを定義する場合は、クライアントが実行されているワークステーションからそのデータベースをカタログする、つまり追加しなければなりません。

以下のリストは、サーバーを定義する際の一般的なステップです。

1. 「サーバー (Servers)」を右クリックし、「新規 (New)」を選択します。
2. リストからサーバーを選択します。「新規サーバー (New Server)」ウィンドウが表示されます。
3. 「一般 (General)」タブの「サーバー名 (Server Name)」フィールドにサーバー名と説明を入力します。一部のサーバーについては、データベース名のみを入力します。また、他の一部のサーバーについては、データベースがインストールされているサーバーの完全修飾名を入力します。
4. 必要に応じて、初期化パラメーターを指定します。一部のサーバーでは、接続ストリングや構成ストリングなどの初期化パラメーターが必要ですが、データベース名のみを指定すればよいサーバーもあります。
5. 「テスト・サーバー (Test Server)」接続をクリックします。EIP は、ユーザーが入力したユーザー ID とパスワードを使用してサーバーにログインし、管理クライアントを開始します。サーバーが別のユーザー ID とパスワードを要求する場合、EIP は、定義するコンテンツ・サーバーに固有なユーザー ID とパスワードの入力を求めるプロンプトを出します。

ヒント: 事前定義されていないサーバー・タイプのコンテンツ・サーバー・タイプを定義することもできますが、その場合は、新しいサーバー・タイプの Java または C++ のコネクタ・クラスとサーバー定義クラスを指定する必要があります。また、サーバー・インベントリを実行するための Java コネクタも必要です。コンテンツ・サーバーの追加方法については、「ワークステーション・アプリケーション・プログラミング・ガイド」および「オンライン API 解説書」を参照してください。

コンテンツ・サーバーの構成に失敗した場合は、その状態でのトラブルシューティングの方法について、または、受け取ったエラー・メッセージの詳細について、「メッセージおよびコード」を参照してください。

また、接続したいサーバーの管理者に相談して、さらに支援を求めることもできます。

サーバー定義のガイドライン

このセクションでは、初期サーバー定義を行う上で役立つガイドラインを提供します。

DB2 (リレーショナル) データベースへの接続

このセクションでは、DB2、DataJoiner、JDBC、ODBC、Information Catalog、および Content Manager バージョン 7 とバージョン 8 サーバーについて説明します。

- **重要:** サーバーを定義する前に、各 DB2 データベースをカタログ化する必要があります。データベースをカタログ化するには、DB2 CCA を使用するか、または DB2 コマンド・プロンプトを使用します。詳細については、DB2 管理者に問い合わせてください。
- 「一般 (General)」タブの「サーバー名 (Server name)」フィールドに、接続するデータベースの名前を入力する必要があります。サーバー名の入力には大文字を使用します。
- DB2、DataJoiner、JDBC、ODBC、および Information Catalog を定義する際は、「初期化パラメーター (Initialization Parameters)」タブをクリックし、接続しようとしているデータベース・テーブルに関連付けられているスキーマ名 (たとえば、SCHEMA=ICMADMIN) を入力します。
- Content Manager バージョン 7.1 サーバーまたは Content Manager バージョン 8.2 サーバーを定義する場合は、データベース名を入力するだけで十分です。「初期化パラメーター (Initialization Parameters)」タブのデフォルト設定値は変更しないでください。
- Content Manager バージョン 7.1 サーバーを定義する場合は、ローカル・ドライブの x:%CMBROOT に frnolint.tbl というネットワーク・テーブルがなければなりません。このネットワーク・テーブルには、EIP が Content Manager バージョン 7.1 ライブラリー・サーバーを探してログオンするときに必要となるホスト名、ポート番号、およびサーバー・タイプ情報が入っています。複数の Content Manager バージョン 7.1 サーバーを定義する場合は、frnolint.tbl ファイル内にサーバーごとに項目を設けてから、各サーバーを定義してください。
- DB2 DataJoiner に接続する場合は、DB2 Universal Database で定義されているデータベース・インスタンスで、Enterprise Information Portal の認証メソッドがサーバーになっていることを確認してください。

- DataJoiner 2.1 に接続するには、DataJoiner Web サイトからバインド・プログラムをダウンロードして DataJoiner データベースにバインドする必要があります。これを行わないと、DataJoiner サーバーを定義することはできません。

Text Search server への接続

Text Search server を定義するには、まず、Text Search server に関連付けられている Content Manager バージョン 7.1 サーバーを定義する必要があります。

Text Search server の名前を、「関連する Content Manager Version 7.1 サーバーの名前を選択する (Select the name of the associated Content Manager Version 7.1 server)」ドロップダウン・ボックスから選択します。このドロップダウン・ボックスは、「関連するサーバー (Associated Server)」タブにあります。

Content Manager バージョン 7.1 サーバーおよび Text Search server を立ち上げて実行しておかないと、EIP はそれらに接続できません。

Content Manager for AS/400 の複数のサーバーへの接続

複数の AS/400 サーバーを使用する場合は、追加のサーバーをネットワーク・テーブルに定義する必要があります。ネットワーク・テーブル (frnolint.tbl) は、x:¥<cmbroot> に入っています。新規サーバーの場合は、サーバー名、接続タイプ (たとえば、TCP/IP)、ホスト名、ポート、およびサーバー・タイプを入力してください。最初のサーバーの場合は、インストーラーがインストール時にサーバー、ホスト名、ポートの値を入力して、frnolint.tbl を作成します。

次に、fronlint.tbl に保管されている情報の代表的な例を示します。

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

Extended Search コネクターの構成

Extended Search サーバーを定義するために入力する情報は、以下の 2 つの要素に応じて異なります。

- 拡張サーバーがインストールされている Web サーバーのタイプ - Domino Web Server、WebSphere、IIS。
- Extended Search がインストールされている Web サーバーに定義されているポート番号。

Extended Search コネクターを定義する際は、以下のステップを実行します。

1. 「一般 (General)」タブの「サーバー名 (Server Name)」フィールドに、Extended Search がインストールされている Web サーバーの完全修飾ホスト名を入力します。
2. Extended Search のインストール時にインストーラーが Web サーバーのポート番号のデフォルト設定を選択した場合は、「初期化パラメーター (Initialization Parameters)」タブの「ポート番号 (Port Number)」フィールドに 80 と入力します。
3. 「アプリケーション ID (Application ID)」フィールドに Demo と入力します。表示された名前を入力します。

4. 「パスワード (Password)」フィールドに Demo と入力します。
5. 「追加パラメーター (Additional Parameters)」フィールドでは、以下の点に留意してください。
 - a. Extended Search が Domino Web サーバーにインストールされていて、インストール時にインストーラーが Web サーバーおよび Extended Search のポート番号にデフォルトのポート番号設定を使用している場合は、2 つのセミコロンを変更しないでください。
 - b. カスタム設定を使用してインストールした Extended Search サーバーの「追加パラメーター (Additional Parameters)」フィールドを変更する方法については、次のセクションを参照してください。

Extended Search コネクタを WebSphere アプリケーション・サーバーで構成した場合、Extended Search ポートが 6001 ではない場合、または Extended Search サーバー名が Web サーバー名と同じではない場合は、コネクタを構成して、ES サブレット、ES ポート番号および ES サーバー名の正しい相対パスを見つける必要があります。

Extended Search を WebSphere で構成した場合、Extended Search ポートが 6001 ではない場合、または Extended Search サーバー名が Web サーバー名と同じではない場合は、以下の手順を完了することによって、構成ファイル (たとえば、desclient.cfg) を作成できます。

ディレクトリーを、ユーザーのアプリケーションまたはサンプルがあるディレクトリーに設定してください。構成ファイル (たとえば、desclient.cfg) を作成します。このファイルは Enterprise Information Portal では提供されません。

desclient.cfg 内で、DESHOSTNAME=(ES ホスト名) を追加します。

ES を Domino アプリケーション・サーバーで構成した場合は、DESREQURI=/servlet/ESAdmin を設定します。ES を WebSphere アプリケーション・サーバーで構成した場合は、DESREQURI=/lotuskms/ESAdmin (servlet が Extended Search コネクタをサポートする HTTP サーバー上のディレクトリー・パス) を設定します。

アプリケーション・サーバーが WebSphere である場合は、DESREQURI は、/servlet/ESAdmin ではなく、/lotuskms/ESAdmin でなければなりません。

シン・クライアントやファット・クライアントから ES ソースを検索する場合は、
"DESCFGPATH=<desclient.cfg の絶対パス>"

という追加パラメーターを、管理クライアントの DES サーバー定義ダイアログに定義します。

ES サンプルを実行する場合は、desclient.cfg の絶対パスをコマンド行引き数に渡します。

例 1:

```
TConnectDES es.stl.ibm.com user password
PORT=80;DESAPPID=Demo;DESAPPPW=password;DESCFGPATH
=<desclient.cfg の絶対パス>;
```

例 2:

```
java TConnectDES es.st1.ibm.com user password  
PORT=80;DESAPPID=Demo;DESAPPPW=password;DESCFGPATH=<desclient.cfg の絶対パス>;
```

Information Catalog サーバーの定義

Information Catalog サーバーは、定義する前にカタログに入れておいてください。

「サーバー名 (Server Name)」フィールドにサーバー名 (たとえば SAMPLE1) を入力します。「初期化パラメーター (Initialization parameters)」タブに、SCHEMA=<SAMPLE1 に関連したスキーマ名> と入力します。

OnDemand サーバーの定義

OnDemand サーバーとライブラリー・サーバー・デーモンが実行中でない限り、OnDemand サーバーの定義は行えません。OnDemand サーバーを EIP に定義する前に、OnDemand サーバーを PING して、サーバーとデーモンが実行中かどうかを調べることができます。

「一般 (General)」タブで、OnDemand がインストールされているサーバーの完全修飾ホスト名を入力します。

「初期化パラメーター (Initialization Parameters)」タブに、OnDemand サーバーのインストール時に割り当てたポート番号を入力します。OnDemand のインストール時にデフォルトのポート値 0 を選択した場合は、ポート番号フィールドに 0 と入力します。別のポート番号値を選択した場合は、ポート番号の先頭に # 記号を付けて入力します。たとえば、Windows サーバーの OnDemand に代替ポート番号 # 5000 を選択することができます。

バージョン 4 ソフトウェアを実行する AS/400 サーバーにインストールした OnDemand サーバーを定義する場合は、「追加パラメーター (Additional Parameters)」フィールドに STATECONNECT=#1 と入力する必要があります。

バージョン 2.1 ソフトウェアを実行する OS/390 サーバーにインストールした OnDemand サーバーを定義する場合は、OnDemand を OS/390 バージョン 2.1 サーバーにインストールする際に指定したカスタム・ポート番号を入力します。

OnDemand では、接続中、ソケットを活動状態にしておく必要があります。

Text Search server の定義

Text Search server を定義するには、まず、Text Search server に関連付けられている Content Manager バージョン 7 サーバーを定義する必要があります。

「一般 (General)」タブの「サーバー名 (Server Name)」フィールドに Text Search server の名前を入力します。「関連サーバー (Associated Server)」タブのドロップダウン・ボックスから、関連する Content Manager バージョン 7.1 サーバーを選択します。

Content Manager バージョン 7.1 サーバーおよび Text Search server を立ち上げて実行しておかないと、EIP はそれらに接続できません。

Domino.Doc サーバーの定義

「サーバー名 (Server Name)」フィールドに、サーバー名へのパスと Domino.Doc サーバーのライブラリー名を入力します。たとえば、oakley/DominoDoc1/Lib.nsf と入力します。

ローカル・コネクタを使用している場合は、EIP クライアントがインストールされているワークステーションに Domino Doc Desktop Enabler をインストールする必要があります。RMI を使用している場合は、RMI サーバーに Domino Doc Desktop Enabler をインストールする必要があります。Domino Doc Desktop Enabler は、Domino Doc サーバーと同じバージョンでなければなりません。

「初期化パラメーター (Initialization Parameters)」タブの 2 つのセミコロンを変更しないでください。

ImagePlus for OS/390 サーバーの定義

ImagePlus for OS/390 サーバーを定義する際は、以下のパラメーターを取得してサーバーに接続する必要があります。以下のリストにはサンプルの値が含まれています。

- FAF ポート番号: 3061
- FAF アプリケーション ID: 01
- FAF プロトコル: 4000
- FAF IP アドレス: 9.67.43.83
- Object Distribution Manager CICS: 4000
- Object Distribution Manager IP アドレス: 9.67.43.83
- Object Distribution Manager ポート番号: 3082
- Object Distribution Manager 端末 ID: このフィールドはブランクにしておいてください
- 追加パラメーター: *FAFSITE=CS61;*

Content Manager ImagePlus for OS/390 でのトレースの使用

Content Manager ImagePlus for OS/390 サーバーに接続できない場合は、トレースが問題の解決に役立ちます。Content Manager ImagePlus for OS/390 用のコネクタをインストールした場合は、cmbroot に入っている eypapi.ini ファイルを変更することで、ImagePlus for OS/390 のトレースをオンにすることができます。

eypapi.ini ファイルには、以下の行が含まれています。

```
; Path where the IPFAF files are stored
; (MUST NOT have a trailing '\')
; -- default is the <ROOT Directory>%
;
IPFAFPath=d:%cmbroot
; Flag for Logging (EYPLmdd.LOG files)
; -- default is Logging OFF (0)
; -- 0 All Logging OFF
; -- 1 Log files created only error conditions logged
; -- 2 Log files created all conditions logged
;
Logging = 0

;-----
;
```

```
; Flag for Logging the FAF Parameters Types created by APIs
;      -- default is Logging OFF (0)
;      -- 0 Parameter types Not logged
;      -- 1 Log Faf Parameter Types
;
FafTypeLogs = 0
```

IPFAFPath

ログを書き込むディレクトリーを指定します。ログ・ファイルの名前は、次のようになります。

EYPmdd.LOG

ここで、*mdd* は、ログが作成された月日です。

Logging

ログ・ファイルをいつ作成するかを指定します。

- 0 ログを行いません。デフォルトの設定値は 0 です。
- 1 作成されたログ・ファイルにはエラー状態のみが含まれます。
- 2 作成されたログ・ファイルにはすべての状態が含まれます。

FafTypeLogs

API によって作成された FAF パラメーター・タイプのロギングを指定します。

- 0 パラメーター・タイプのロギングを行いません。デフォルトの設定値は 0/ です。
- 1 FAF パラメーター・タイプのロギングを行います。

OnDemand コネクタの使用: TCP/IP チューニングとソケット

OnDemand サーバーに接続している場合、Windows の既知の問題がパフォーマンスに影響を与える可能性があります。OnDemand サーバー上で検索と取り出しが反復して行われている間に、Windows の多数のソケットが開閉されます。Windows の 2 つのデフォルト設定値が、EIP と OnDemand サーバー間の過密なトラフィックに影響を与えることがあります。

- アプリケーションが Windows ソケットを閉じると、Windows は、そのソケット・ポートを 240 秒間、TIME_WAIT 状況に置きます。この間は、このポートを再使用することができません。
- Windows は、アプリケーションが使用できるポート数を 5000 に制限しています。

この結果生じる可能性のある問題を回避するには、Windows のレジストリー・エディターを使用して、タイムアウト待機時間およびポート数のパラメーター値を変更します。

- タイムアウト待機時間の値を 240 秒より、低い数値 (有効範囲は 30 ~ 300 秒です) に変更する。キー名は、
HKEY_Local_Machine\System\CurrentControlSet\services\Tcpip\Parameters\TcpTimedWaitDelay です。
- 最大ポート数をデフォルトの 5000 より、高い数値 (有効範囲は 5000 ~ 65534 です) に増加する。キー名は、HKEY_Local_Machine\System\CurrentControlSet\services\Tcpip\Parameters\MaxUserPort です。

TcpTimedWaitDelay と MaxUserPort の詳細については、Windows のドキュメンテーションを参照してください。

Extended Search コネクタの使用

このセクションでは、EIP バージョン 8.2 の Extended Search コネクタの変更点について説明します。

ユーザー指定ロケールは、ロケール値を DESLOCALE キーで渡すことによってサポートされます。ES コネクタを直接起動する場合は、このペア値をコマンド行で渡すことができます。この値を、ES プロパティの **Additional** 引き数で設定することができます。

ヒント: Extended Search サーバー・ソフトウェアは、EIP バージョン 8.2 の製品ボックスに組み込まれています。

統合エンティティの作成

コンテンツ・サーバーへの接続を定義した後、統合検索の作成における次のステップは、検索テンプレートのビルディング・ブロックとなる統合エンティティを作成することです。このセクションでは、統合エンティティについて説明し、統合エンティティの作成ウィザードの使用方法についても説明します。

統合エンティティの理解

ほとんどの場合、クライアント・アプリケーション・ユーザーは、情報をサーバーごとに検索したいとは考えません。サーバーごとの検索ではなく、1 回の統合検索で済ませたいと考えます。検索テンプレートにより、クライアント・アプリケーション・ユーザーは、複数の検索を 1 回の照会にまとめることができます。EIP 管理者として、ユーザーは、これらの検索テンプレートをクライアント・アプリケーションで使用できるように作成することができます。検索テンプレートを作成する前に、統合属性をコンテンツ・サーバー上の固有属性にマップする統合エンティティを最初に作成する必要があります。

たとえば、DB2 は情報を表に保管します。この表の列は表内に保管される情報の属性を示します。Customer_Demographics という名前の表に、Name、Pol_Number、Address、Phone、Occupation などの列が含まれる場合などです。

これに対して、Content Manager は、表と列の代わりに、項目、項目タイプ、および属性を使用します。DB2 に保管されるのと同じ情報がエンティティ名 CustInfo 内に保管されます。その属性は、CustName、Acct、HomeAddress、HomePhone、Job などになります。これらはどちらの場合も、同一の情報がそれぞれ異なる方法で保管され、識別されます。

EIP を利用すれば、どのコンテンツ・サーバーにも同じ情報が保管されているのに、保管方法がそれぞれ異なることを考慮しなければならないという問題を解決することができます。統合エンティティがユーザーのためにこの情報の記録を続けています。統合エンティティは、実際にはデータを保管していません。統合エンティティが保管しているのは、各コンテンツ・サーバーがどのようにデータを保

管したかについてのメタデータです。統合エンティティを作成する場合は、そのすべての属性を、照会するコンテンツ・サーバー上の対応する固有属性にマップします。

上記の例の場合、Policy_Info という名前の統合エンティティを、統合属性 Policy_Name、Policy_Number、Home_Address、および Job_Title を指定して作成することができます。さらに、統合属性を、対応する固有属性のそれぞれにマップすることができます。

EIP は、この情報を入れるサーバー・インベントリを生成することができます。サーバー・インベントリにこの情報を入れて、統合エンティティの作成ウィザードでサーバー・インベントリを取得すれば、コンテンツ・サーバーでサーバー・インベントリをフィルターに掛けることができます。フィルター操作は、「統合エンティティの作成 (Create Federated Entities)」ウィザードを使用する場合のみ使用可能です。(ウィザードではなく) 手動で統合エンティティを作成する場合、フィルター操作はオプションではありません。サーバー・インベントリを生成した後、統合属性の固有属性へのマップを開始できます。

統合属性を固有属性にマップするには、まだ十分ではありません。それぞれの固有属性が、別のプロパティを持っている場合もあります。属性には、(1) ヌル可能、(2) 照会可能、(3) 更新可能、(4) テキスト検索可能、などがあります。また、選択したデータ型に応じ、データ長、精度、スケール、および最小値と最大値に関するオプションがあります。

これらのプロパティを定義する場合、統合属性にマップした固有属性によってすでに定義済みのプロパティよりも厳しく制限することはできません。ウィザードは、この基準を満たすデフォルトのプロパティを提供します。デフォルトの統合属性のプロパティをカスタマイズした後で、ウィザードによって提案されたデフォルト設定に復帰したい場合は、デフォルト設定を選択することができます。

要約すると、統合属性は、複数のコンテンツ・サーバー上の対応する固有属性にマップします。それぞれの統合属性のプロパティは、固有属性のすべてのプロパティを包含します。統合エンティティを作成することで、さまざまなコンテンツ・サーバー上に保管されている情報へのパスを持つことになります。これで、統合エンティティを使用して特定の照会のための検索テンプレートを作成できます。

統合エンティティの作成ウィザードの使用

統合エンティティの作成ウィザードは、EIP バージョン 8.2 で導入された新規機能です。統合エンティティの作成には、EIP バージョン 7.1 以前のものと同じダイアログを使用することもできますが、このウィザードを使用すると、統合エンティティの作成がさらに容易になります。

ウィザードを使用して統合エンティティを作成するには、以下にリストされたステップに従います。

1. 「**統合エンティティの定義 (Define Federated Entity)**」: 統合エンティティへの名前付け、および説明を行います。統合エンティティをテキスト検索可能にするかどうかも決定できます。

2. 「**統合属性の定義 (Define Federated Attributes)**」: 統合属性に名前を付けたり、変更したりします。
3. 「**統合属性のマッピング (Map Federated Attributes)**」: 統合属性を固有属性にマッピングします。用意されたツールによって、サーバー・インベントリを取得したり、マッピングしたい固有属性を選択したり、マッピングを後日変更したりすることができます。
4. 「**プロパティの定義 (Define Properties)**」: それぞれの統合属性ごとにプロパティを定義します。プロパティをカスタマイズするか、デフォルト設定を受け入れるかのいずれかが可能です。
5. 「**統合エンティティの確認 (Confirm Federated Entity)**」: 統合エンティティで選択した設定をチェックします。直前のパネルに戻って設定を変更することができます。終了する場合は、「**終了 (Finish)**」をクリックします。

これらのステップは、ウィザードの使用と対応しています。ウィザードの使用方法についての詳細は、EIP オンライン・ヘルプを参照してください。

統合テキスト索引の作成

テキスト検索エンジンを Content Manager バージョン 7.1 以前のコンテンツ・サーバーに統合すると、Content Manager に保管されたテキスト情報の索引付け、検索、および取り出しを自動的に行うことができます。ユーザーは、語または句を検索して文書を探し出すことができます。この Text Search Server は、単一バイトおよび 2 バイト文字セットをサポートします。

テキスト検索エンジンと Content Manager バージョン 7.1 以前のサーバーを併用している場合、統合テキスト索引を作成することができます。そして、その統合テキスト索引を Content Manager Text Search Server にある Content Manager テキスト検索索引にマッピングします。

統合テキスト検索索引を作成して、固有テキスト索引と固有属性の両方を検索する結合検索にそれを使えるようにすることができます。結合検索用に統合テキスト索引を使用可能にする場合、その索引を統合エンティティにマッピングします。それから、統合エンティティにマッピングされた固有属性、およびその統合属性を Text Search Server 上のテキスト検索索引にマッピングします。

検索テンプレートの作成

統合エンティティを作成した後に、検索テンプレートを作成します。コンテンツが保管されている場所へのマッピングとして、統合エンティティを検索テンプレートが使用するというのを思い出してください。検索テンプレートを作成する際、何を検索したいのか、検索結果を何に利用したいのか、およびこのテンプレートの使用をだれに許可するのかについて、引き続き定義する必要があります。各テンプレートに対して 1 つの統合エンティティだけを使用できると同時に、複数のテンプレートに対して 1 つの統合エンティティを使用することができます。統合エンティティの属性の任意の組み合わせを検索基準として検索することもできます。検索テンプレートを作成するには、検索テンプレート・ウィザードを使用して、次のステップを実行します。

1. 検索テンプレートの定義

2. 検索基準の定義
3. 検索設定の決定
4. アクセス権の割り当て

これらのステップは、検索テンプレート・ウィザードのステップと対応しています。検索テンプレートの作成プロセスの実行については、EIP オンライン・ヘルプを参照してください。

検索テンプレートの定義 (Define search template)

ウィザードを開始すると、検索テンプレートを定義するようプロンプトが出されます。次の作業について準備してください。

- 検索テンプレートの名前と記述を入力する
- 検索テンプレートの統合エンティティを選択する。**制約事項:** 各検索テンプレートごとに使用できる統合エンティティは 1 つのみです。
- 該当する場合、統合テキスト索引を選択する

ヒント: 統合テキスト索引のチェック・ボックスは、Content Manager バージョン 7.1 以前用のテキスト検索エンジンを使用している場合にのみ、適用できます。テキスト検索に DB2 TIE を使用する場合は、パラメトリック検索であるため、検索テンプレートでパラメトリック検索として構成することができます。

検索基準の定義 (Define search criteria)

検索テンプレートを定義すると、以下のことをするようウィザードがプロンプトを出します。

1. 検索タイプ (属性または文書のいずれか) を選択します。文書が有効なのは、前のステップで統合テキストの索引を選択した場合のみです。
2. 検索基準に名前を付ける
3. 統合属性を選択する
4. 使用可能な演算子を選択する
5. デフォルトの検索ストリングを入力する (文書検索のみ)

ウィザードは、選択された統合エンティティに関連付けられているすべての統合属性をリストしたドロップダウン・メニューを表示します。これらの属性が、検索テンプレートの検索基準となります。ウィザードは、使用可能な演算子のリストも提供します。

ヒント: テンプレートごとに複数の検索基準を作成できます。また、既存の基準をテンプレートから削除することもできます。

検索設定の決定 (Determine search settings)

このパネルにより、デフォルトの検索設定、基準の設定、および表示値の設定を定義することができます。これらの設定はそれぞれ、変更可能なデフォルト値を持っています。これらの設定を変更するには、それぞれの設定ごとに対応するボタンをクリックします。

「デフォルト設定 (Default Settings)」ウィンドウでは、次の作業が可能です。

- クライアント・アプリケーションのユーザーが検索テンプレートを使用するとき
にサーバーが使用不可であった場合の動作を制御する
- パラメトリック検索用のワイルドカード文字を定義する
- 検索結果を保管するフォルダーの名前を指定する
- 検索ですべての基準を使用するか (AND)、または一部を使用するか (OR) を選択
する

「基準設定 (Criteria Settings)」ウィンドウにより、検索基準の順序、結果表示列の順序、列見出し、および列幅を制御することができます。

「表示値の設定 (Display Value Settings)」ウィンドウは、検索結果の表示値を定義する方法を提供します。たとえば、あるサーバー上で `Weekday` の値が `Monday` であり、別のサーバー上では `Mon.` である場合、両方のサーバーの検索結果の表示値として `Monday` を使用することを指定できます。

特権の割り当て (Assign privileges)

どこを調べるか (統合エンティティの場合)、何を探すか (検索基準)、どのように結果を表示するか (設定) を定義するほかに、検索テンプレートへのアクセス権をだれに与えるかについても定義する必要があります。

「検索テンプレート (Search Template)」ウィザードの「特権の割り当て (Assign Privileges)」ウィンドウは、テンプレートへのアクセス権を既存のユーザーまたはユーザー・グループに割り当てるツールを提供します。

あるユーザーに検索テンプレートへのアクセス権を割り当てても、そのテンプレートにマップされたコンテンツ・サーバーへのアクセス権をそのユーザーに付与するわけではありません。ユーザーは、それぞれ個別のコンテンツ・サーバーごとにセキュリティ要件を満たす必要があります。検索テンプレートへのアクセスをユーザーに割り当てる前に、アクセス制御リストおよびユーザー管理を使用して、ユーザーに適切な特権を持たせるようにする必要があります。

ウィザードを使用してユーザーまたはユーザー・グループを検索すると、EIP は、要求されたコンテンツ・サーバーへの適切なアクセス権を持つユーザーのみを戻します。

第 4 章 ユーザー・アクセスの管理

ユーザーは、ユーザー ID、パスワード、特権セットがないと EIP システムにアクセスできません。しかし、ユーザーを作成して、それらのユーザーに特権を割り当てる前に、システムへのアクセス権をどのユーザーに与えるか、およびそれらのユーザーのジョブに必要なものが何であるかを決定しておく必要があります。あるオブジェクトを削除した場合に、それがどのような影響を及ぼすかを理解できないユーザーには、オブジェクトの削除権限を与えるべきではありません。他方、適切な権限をユーザーに与えないことにより、ユーザーのジョブを妨げるべきではありません。したがって、ユーザーに特権を割り当てる前に、各ジョブに必要なタスクのタイプを決定する必要があります。

ユーザーが EIP システムでオブジェクトを作成した場合、そのユーザーは、他のユーザーがそのオブジェクトにアクセスする権限を定義する必要があります。オブジェクトを作成するユーザーは、誰がそのオブジェクトにアクセスできるか、そのオブジェクトに対してどのような操作を行うことができるかを定義する必要があります。この定義は、アクセス制御リスト、つまり ACL として EIP システムに認識されます。

ユーザー ID とパスワードの作成

システム管理クライアントで定義するユーザー ID を DB2 認証にも使用したい場合、そのユーザー ID は DB2 命名規則に従っていなければなりません。DB2 命名規則は、スーパー・アドミニストレーターまたは接続ユーザー ID として使用したいユーザー ID に適用されます。以下のワードは使用できません。

- USERS
- ADMINS
- GUESTS
- PUBLIC
- LOCAL
- SQL 予約語は、「SQL リファレンス」にリストされています。

ユーザー ID の最初の 3 文字に、以下のものは使用できません。

- SQL
- SYS
- IBM

以下の文字は使用することができます。

- A ~ Z (制約事項: オペレーティング・システムによっては、ユーザー ID とパスワードの大文字小文字を区別します)。大文字小文字を区別する必要があるかどうかについては、ご使用のオペレーティング・システムの資料を調べてください。
- 0 ~ 9

- #
- \$

制約事項: ユーザー ID は 30 文字以下でなければなりません。

DB2 管理者権限について

システム管理クライアントにログオンする場合、データベース・レベルと製品レベルの 2 つの認証のレベルがあります。管理可能ドメイン・フィーチャーを使用可能にすると、管理者は、スーパー・アドミニストレーターとサブアドミニストレーターの 2 つの種別になります。一般に、スーパー・アドミニストレーターのみがシステム管理クライアントにアクセスできます。

スーパー・アドミニストレーターは、DB2 特権を持っていないければなりません。db2admin (つまり、DB2 への完全な管理特権) が必要です。このユーザー ID は、オペレーティング・システム内で db2admin 特権付きで定義されていなければなりません。このオペレーティング・システム ID 用パスワードは、DB2 へ接続してライブラリー・サーバーにログオンするために使用されます。ライブラリー・サーバー用に定義されたパスワードは使用されません。Content Manager 特権、このユーザー ID はライブラリー・サーバー内で定義され、完全な Content Manager 管理特権 ("AllPrivs") 付きで、すべての管理アクティビティーを実行できます。

サブアドミニストレーターには、DB2 特権は必要ありません。サブアドミニストレーターは、ライブラリー・サーバーの特定のセクションのみを管理します。そのため、サブアドミニストレーターは、次の 2 つの方法のどちらかでシステム管理クライアントにログオンします。

- ユーザー ID がオペレーティング・システム・ユーザー ID の場合、オペレーティング・システム内のパスワードは、DB2 に接続してライブラリー・サーバーにログオンするために使用されます。
- ユーザー ID がオペレーティング・システム・ユーザー ID ではない場合、ユーザー ID とパスワードのペア (Enterprise Information Portal では cmbfedenv.ini、Content Manager では cmbicmenv.ini 内で暗号化されている) が DB2 に接続するために使用され、「ログオン (Logon)」ウィンドウ内で指定されるユーザー ID とパスワードがライブラリー・サーバーにログオンするために使用されます。

ライブラリー・サーバーへのログオンの詳細については、次のセクションを参照してください。

サブアドミニストレーターには、EIP 特権も必要です。すべてのサブドメイン管理アクティビティーに対するドメイン管理特権が必要です。

INI ファイルを使用した DB2 への接続

INI ファイル内の各エントリーには、ライブラリー・サーバーの名前、および DB2 に接続するための暗号化されたユーザー ID とパスワードのペアが含まれます。この暗号化されたユーザー ID (接続ユーザー ID と呼ばれる) とパスワードは、製品のインストール時に定義されます。接続ユーザー ID は、システム管理者のユーザー ID とは異なるものでなければなりません。Enterprise Information Portal は cmbfedenv.ini を使用して DB2 に接続し、Content Manager は cmbicmenv.ini を使

用します。デフォルトの接続ユーザー ID は ICMCONCT です。インストール時に、ライブラリー・サーバーとリソース・マネージャー用のパスワードは 3 つの場所にあります。cmbicmenv.ini ファイルには、ライブラリー・サーバーに接続するためのユーザー ID とパスワードが含まれています。オペレーティング・システムは、ライブラリー・サーバーとリソース・マネージャーが常駐するデータベースへのアクセスを定義します。ICMRM.properties ファイルには、リソース・マネージャーのユーザー ID とパスワードが含まれています。

INI ファイルが使用されている場合、つまり、ユーザー ID がオペレーティング・システム・ユーザー ID ではない場合は、INI ファイル内のユーザー ID と接続ユーザー ID の両方がライブラリー・サーバー内に存在していなければなりません。

接続ユーザー ID は、ライブラリー・サーバーとオペレーティング・システム内で定義されていなければなりません。UserDB2Connect 特権が必要です。INI ファイル内の接続ユーザー ID とパスワードを変更するには、管理クライアント・ウィンドウから「ツール (Tools)」 --> 「データベース ID/パスワードの変更 (Change Database ID/password)」を選択してください。

リソース・マネージャーへのライブラリー・サーバーとシステム管理者のパスワードの変更

リソース・マネージャーへのパスワードを変更する必要がある場合は、リソース・マネージャーへのライブラリー・サーバーのログオン用パスワード、およびリソース・マネージャーへのシステム管理者のパスワードを変更する必要があります。**重要:** リソース・マネージャーへのライブラリー・サーバーとシステム管理者のログオン用パスワードを変更するときは、以下の手順を完了してください。

1. システム管理クライアントにログオンします。
2. 「リソース・マネージャー (Resource Manager)」ツリーを展開します。
3. 変更したいリソース・マネージャーをクリックして、そのツリーを展開します。
4. 「サーバー定義 (Server Definitions)」をクリックして、「プロパティー (Properties)」を選択します。「サーバー・パネル (Server Panel)」ウィンドウが開きます。
5. 「パスワード (Password)」フィールド内のパスワードを変更します。
6. 「OK」をクリックします。
7. (ステップ 2 から) 展開したリソース・マネージャーを右マウス・ボタン・クリックして、「プロパティー (Properties)」を選択します。「リソース・マネージャーのプロパティー (Resource Manager Properties)」ウィンドウが開きます。
8. 「パスワード (Password)」フィールド内のパスワードを変更して、「OK」をクリックします。

データベース・アクセス・パスワードの変更

データベース・アクセス・パスワードを変更する必要がある場合は、データベース接続用のオペレーティング・システム・パスワードと ICMRM.properties ファイルを変更して、リソース・マネージャーが新規パスワードを識別できるようにする必要があります。

データベース接続用のオペレーティング・システム・パスワードを変更するには、次の手順を実行してください。

1. 使用中のオペレーティング・システムに応じて、Users and Passwords ユーティリティーをナビゲートします。
2. 「ICMRM」をクリックします。
3. 「パスワードの設定 (Set Password)」を選択します。
4. 新規パスワードを入力します。

ICMRM.properties ファイルを変更するには、次の手順を完了してください。

1. ICMRM.properties ファイルを開きます。デフォルトの場所は、次のとおりです。 `X:\%WebSphere%\AppServer%\installedApps\%icrmr.ear\%icrmr.war\WEB-INF\classes\com\ibm\mm\%icrmr\ICMRM.properties` ここで、X は Content Manager をインストールしたドライブです。
2. DBPassword をオペレーティング・システム・パスワードと一致するように変更します。
3. ICMRM.properties ファイルを保管します。

データベース・パスワードを変更後に、データベースをリスタートする必要があります。リスタートしないと、2 ~ 3 回エラーが出た後でリセットされます。

システム管理クライアントでのリソース・マネージャー用のパスワードとその他のフィールドの変更についての詳細な説明は、システム管理オンライン・ヘルプを参照してください。

LDAP からのユーザーのインポート

LDAP は、ユーザー ID とパスワードの管理を、システムごとのベースではなく、企業レベルでサポートします。EIP は、IBM Directory (前のバージョンでは、IBM SecureWay Directory と呼ばれました)、Windows 2000 Active Directory、および Lotus Domino Directory Notes Address Book (NAB) の 3 つの LDAP テクノロジーを利用しています。ユーザー・パスワードは、LDAP サーバー上にあります。ユーザーが または Enterprise Information Portal にログオンすると、ユーザー ID とパスワードが認証され、そのユーザー ID に固有の特権が EIP データベースのユーザー・プロファイルによって検査されます。LDAP は、EIP インストール時に使用可能になっている可能性があります。LDAP がインストール時に使用可能になっている場合は、いつでもアクティブにすることができます。

LDAP を使用可能にするには、「スタート」→「プログラム」→「EIP for Multiplatforms」→「LDAP User ID Import Scheduler」を選択して、システム管理クライアントを起動してください。「LDAP の構成 (LDAP Configuration)」ウィンドウ (「ツール (Tools)」-->「LDAP の構成 (LDAP Configuration)」) を開きます。「LDAP ユーザー・インポートの使用可能化 (Enable LDAP user import)」と「認証チェック・ボックス (authentication check box)」を選択して、LDAP サーバー情報を「サーバー (Server)」ページ上で指定してください。

LDAP を使用可能にすると、「新規ユーザー (New User)」ウィンドウ内の LDAP ボタンをクリックしてユーザーをインポートすることができます。これによって、LDAP サーバーから選択的にユーザーを EIP にインポートできるようになります。

別の方法として、LDAP User ID Import Scheduler ユーティリティを使用して、グループ内にユーザーをインポートすることもできます。ログオン時に、ライブラリー・サーバーは自動的に LDAP サーバーに接続して、そのユーザーの認証を行います。LDAP サーバーが何らかの理由でユーザーのパスワードを検査できなかった場合は、認証が失敗します。

メインのシステム管理クライアント・ウィンドウで、「ツール (Tools)」->「LDAP の構成 (LDAP Configuration)」をクリックして、LDAP サーバーの構成を変更することができます。また、EIP 上の「スタート」メニューから LDAP User Registry Import ユーティリティで、現行 LDAP サーバーを変更することもできます。LDAP の計画の詳細については、「Content Management System の計画とインストール」を参照してください。システム管理ウィンドウ内での LDAP サーバー情報の構成方法については、システム管理クライアントのオンライン・ヘルプを参照してください。

LDAP の計画の詳細については、「Content Management System の計画とインストール」を参照してください。LDAP をインプリメントする方法については、システム管理クライアントのオンライン・ヘルプを参照してください。

特権の紹介

管理クライアントは、特権グループ、特権セット、および個別の特権を提供します。Content Manager/EIP 結合システムを管理する場合、特権は両方のクライアントに共通です。クライアントに組み込まれる特権は、システムの合理化に役立ちます。

特権グループ

特権グループは、管理者が新規の特権セットを作成するとき、または「特権セット (Privilege set)」ダイアログで役割を使用するとき、管理者を支援するためのユーザー・タスクの集合です。

特権セット

特権セットは、ユーザーの役割の集合です。

特権 特権は、ユーザー・アクションを表します。たとえば、以下のとおりです。

例 1 - 特権: 通常は文書をスキャンして Content Manager にインポートするのみのためにクライアントを使用するユーザーのグループに、特権 ClientScan および ClientImport を割り当てたいとします。複数のユーザーが通常はこのタスクを実行する場合、1 つのユーザー ID (たとえば、user1) を作成します。特権 ClientScan および ClientImport をユーザー ID User1 に関連付けます。次に、User1 を Group1 という名前のグループに割り当てます。任意のエンド・ユーザーが user1 と入力し、クライアントにログインして、Content Manager にアクセスすると、そのユーザーは、文書をスキャンしてインポートすることのみ可能です。

例 2 - 特権グループ: 典型的なクライアント・タスクすべてにアクセスする特権が必要な、経験豊富なエンド・ユーザーのグループがあるとします。ユーザー ID (たとえば、user2) を作成します。user2 をグループ (たとえば、group2) に割り当てます。次に、ClientTaskAll という名前の特権グループを user1 に関連付けます。任意のエンド・ユーザーが user2 と入力し、クライアントにログインして、Content

Manager にアクセスすると、そのユーザーは、ClientTaskAll という名前の特権グループに含まれるタスクすべてを実行することができます。

例 3 - 特権セット: 読み取り専用アクセスが必要なユーザーのグループがあるとします。ユーザー ID (たとえば、user3) を作成します。user3 をグループ (たとえば、group3) に割り当てます。次に、ClientUserReadOnly という名前の特権セットを user3 に関連付けます。任意のエンド・ユーザーが user3 と入力し、クライアントにログインして、Content Manager にアクセスすると、そのユーザーは、ClientUserReadOnly という名前の特権セットに含まれるタスクのみを実行することができます。

特権セットの作成

EIP システム構成を計画する際は、システムへのアクセス権を誰に与えるか、また、システム上のオブジェクトに対するアクセス権を、これらのユーザーにどの程度与えるかについても決定する必要があります。EIP システムは、特権によってアクセス権を定義します。

特権は、特定の 방법으로特定のオブジェクトにアクセスする権利を与えます。特権には、システムに保管されているオブジェクトの作成、削除、および選択を行うための権利が含まれます。ユーザーに割り当てられた特権のグループは特権セットと呼ばれます。

アクセス管理を行うにはまず、ユーザーに特権セットを作成する必要があります。特権セットは、ユーザーが実行できるタスクまたはアクションを識別します。特権セットは特権を結合し、特定のタイプのユーザー用に調整したものです。たとえば、ある管理者グループには文書ルーティング・サーバーを管理させ、別の管理者グループにはドメインを管理させることもできます。管理者がログオンすると、EIP はその管理者の特権セットを検査します。

システム管理クライアントには数多くの定義済み特権が用意されており、これらの特権をまとめて 1 つの特権セットを作成することができます。次に、作成した特権セットを個々のユーザーに割り当てます。特権セットをユーザー・グループに割り当てることはできません。

特権グループの作成

特権グループは、ユーザーのユーザー・グループと似ています。特権グループの作成は、特権セットに組み込む特権を簡単に見つけ出せるように類似した特権をまとめておくことが目的です。たとえば、2 つの特権を持っていて、システムを使用するほとんどすべてのユーザーにこれらの特権を割り当てる場合は、特権セットを作成するたびに多くの特権を検索するのではなく、これらの 2 つの基本の特権を BasicPrivs という 1 つの特権グループにまとめます。

ユーザーへの特権セットの割り当て

システム管理クライアントには数多くの定義済み特権が用意されており、これらの特権をまとめて 1 つの特権セットを作成することができます。次に、作成した特権セットを個々のユーザーに割り当てます。特権セットをユーザー・グループに割り当てることはできません。

特権名を作成することはできますが、特権そのものを作成することはできません。システム管理クライアントに定義されていない特権を作成する場合は、システム・プログラマーとの作業が必要になります。

EIP に組み込まれている特権セットを使用したり、独自の特権セットを作成したりすることができます。

ユーザー ID への認可特権セットの割り当て

ユーザーが現行の特権よりも多くの特権を持つユーザー ID を作成できないようにするために、EIP では、認可特権セットの使用を導入しています。認可特権セットを持つユーザー ID を割り当てると、ユーザーには、認可された特権の範囲内でユーザー ID を作成する権限が付与されます。たとえば、ドメインを管理するためのシステム管理特権のセットをユーザー ID に付与することができます。ただし、このユーザー ID にはユーザー作成特権を含めないようにすることをお勧めします。そのため、このユーザー ID を作成するときに、「認可特権セット (Grant Privilege Set)」フィールドで "Noprivs" を選択します。この結果、このユーザー ID はドメインを管理することはできますが、そのドメインのユーザーを作成することはできません。

リソース・マネージャーへのユーザーの割り当て

ユーザーが特定のリソース・マネージャーにアクセスできるようにするには、ユーザーがアクセスできるドメインにリソース・マネージャーを割り当てます。ドメインへのリソース・マネージャーの割り当てについての詳細は、43 ページの『ドメインへのリソース・マネージャーの割り当て』を参照してください。

コレクションへのユーザーの割り当て

ユーザーがコレクションにアクセスできるようにするには、ユーザーがアクセスできるドメインにリソース・マネージャーのコレクションを割り当てます。ドメインへのコレクションの割り当てについての詳細は、43 ページの『ドメインへのコレクションの割り当て』を参照してください。

ユーザー・グループの作成

多くの場合、同じ職務内容の複数のユーザーは、同一の作業または類似の作業を行うことから、システム上のオブジェクトに対して同一のアクセスを行います。共通のアクセス要件を持つ複数のユーザーを 1 つのユーザー・グループにまとめることができます。ユーザー・グループをネストすることはできません。

ユーザー・グループとは、同種のタスクを実行する個々のユーザーを便宜的にグループ化したものに過ぎません。ユーザー・グループに特権セットを割り当てないでください。ユーザー・グループ内の各ユーザーは、それぞれ独自の特権セットを持っています。ユーザー・グループを設定しておく、システム内のオブジェクトに関するアクセス制御リストを簡単に作成できます。

ドメインを使用可能にしている場合は、ユーザー ID をグループに割り当てる前に、そのユーザー・グループが特定のドメインに含まれているか、または PUBLIC ドメインに含まれているかを調べてください (ドメインの詳細については、41 ページの『ドメインの管理』を参照)。ユーザー・グループが、ユーザー ID を含めよう

としているドメインに入っていることを確認してください。個々のドメイン用のユーザー ID を作成したい場合は、「ユーザー・グループ (User Group)」ウィンドウで「**新規ユーザー (New User)**」をクリックします。次に、作成したユーザーをユーザー・グループに追加して、そのユーザーが同一ドメインに含まれていることを確認します。

アクセス制御リストの作成

ユーザーには、自分のタスクを実行するために必要な特権が付与されます。オブジェクトは、個々のベースで、それぞれ一定のアクセス制御を発行します。

アクセス制御リスト (ACL) は、1 つまたは複数の個別ユーザー ID またはユーザー・グループ、およびそれぞれに関連する特権から構成されるリストです。ACL を使用して、EIP システム内のオブジェクトへのユーザー・アクセスを制御します。アクセス制御リストに関連付けることのできるオブジェクトとしては、ユーザーが保管したデータ・オブジェクト、項目タイプおよび項目タイプ・サブセット、ワーク・リスト、プロセスなどがあります。

特権セットは、個々のユーザーがそのシステムで行うことのできる能力の上限を定義し、ACL は、個々のユーザーのオブジェクトへのアクセス権を制限します。あるユーザーの特権セットで定義されていない特権が ACL に含まれていれば、ACL は、そのユーザーにその特権を認可しません。その特権を持つユーザーのみが、その特権をオブジェクトに使用することができます。ACL はユーザー・アクセスを制限し、それより多くのアクセス権は認可しません。システムを管理する際には、アクセス制御リストは別のレベルのセキュリティを提供します。

アクセス制御リストへの特権セットの割り当て

アクセス制御リスト (ACL) に追加した各ユーザー ID には、特権セットを関連付ける必要があります。ユーザー ID と特権セットは、どのユーザーがオブジェクトにアクセスできるか、また、ユーザーが、そのオブジェクトに対してどのようなアクセスを行うことができるかを定義します。

ユーザーは、ACL に含まれていない限り、オブジェクトにアクセスできません。ユーザーまたはユーザー・グループを ACL に追加するには、その ACL へ追加するユーザー ID と特権セットを選択して、「**追加 (Add)**」をクリックする必要があります。「アクセス制御リスト (Access Control List)」ウィンドウには、定義された各 ACL ごとにユーザー ID とユーザー・グループがリストされます。このテーブルは、ユーザー ID とユーザー・グループの追加および除去を行うことで変更できます。ACL の作成または変更の詳細については、システム管理クライアントのオンライン・ヘルプを参照してください。

ドメインの作成

ドメインとは、1 人または複数の管理者が管理する管理データベースのセクションです。ドメインは、ユーザー ID、ユーザー・グループ、特権セット、アクセス制御リスト、リソース・マネージャー、および SMS コレクションから構成されます。ドメインはユーザーには見えないため、ドメインを管理するシステム管理者にとっては、ドメインに付けた名前のみが意味を持ちます。ユーザーは、管理データベー

スの一部のみに制限されていることを意識しません。つまり、ユーザーはそのドメイン内の項目についてしか認識しません。

ドメインは、管理データベースのサブセクションへの管理アクセスおよびユーザー・アクセスを制限します。管理データベースに対するすべての特権を持つ管理者は、一部の管理特権を別の管理者に委任することができます。すべての特権を持つ管理者、つまり上級管理者は、管理データベースのすべてのセクションにアクセスできますが、特権が限定されている管理者、つまり副管理者は、管理データベースの 1 つのセクションにしかアクセスできません。

ドメインは、副管理者がアクセス制御リスト (ACL) に対して行うアクセスを制限します。副管理者がユーザー ID やユーザー・グループを追加または削除するために使用できる ACL を作成することができるのは、上級管理者のみです。副管理者は、ACL の作成、更新、削除を行うことはできません。

副管理者は、上級管理者が担当している仕事のいろいろな組み合わせを分担できますが、分担できるのは自分のドメインのみに限られます。ドメインを作成し、それらのドメインの管理に管理者を割り当てることにより、上級管理者は、サブタスクを委任することでシステム全般に専念して管理を効率化できる一方、副管理者は、自分のドメインに固有のユーザーとタスクを管理することができます。

ドメインを使用可能にする前に、以下の条件を考慮してください。

- ドメインを使用不可能にすることはできない
- リソース・マネージャー、コレクション、ユーザー ID、およびユーザー・グループは、一度に 1 つのドメインにしか存在できない
- 特権セットとアクセス制御リストは、一度に複数のドメインに存在できる
- PUBLIC (共用) ドメインを除き、ドメインのオーバーラップはできない
- 上級管理可能ドメインで作成されたオブジェクトは、システム生成、ユーザー作成を問わず、移動できない

ドメインを使用可能にするには、ファイル・メニューを表示して、「ツール (Tools)」 → 「管理可能ドメイン (Administrative Domains)」を選択し、「管理可能ドメインを使用可能にする (Enable Administrative Domains)」を選択します。これを有効にするには、ドメインのシステム管理クライアントを再始動する必要があります。ドメインの管理データベースの構成方法については、システム管理クライアントのオンライン・ヘルプを参照してください。

ドメインの管理

所有する特権セットに応じて、管理データベース全体を管理するか、特定のドメインを管理するかが決まります。管理データベースへの全アクセス権を持つ管理者は上級管理者です。副管理者は、特定のドメイン内のオブジェクトへの全アクセス権を持ちます。

それぞれのタイプの管理者は、ユーザーおよびコレクションを含め、自分のドメイン内のオブジェクトを作成、取り出し、更新、および削除することができます。副管理者は、自分のドメイン内のオブジェクトの表示および検索と、PUBLIC (共用) ドメイン内のオブジェクトのリストおよび検索を行えます。

ドメインへのアクセス

副管理者はオブジェクトのドメインを変更することはできません。ただし、自分のドメインの内容にアクセスしたり、PUBLIC (共用) ドメイン内のオブジェクトをリスト、検索することは可能です。

上級管理者は、管理データベース上のすべてのドメインにアクセスできます。上級管理者は、オブジェクトを作成して、それをドメインに割り当てることができます。特権セットや ACL など、一部のオブジェクトは、上級管理者のみが副管理者のために作成することができます。

副管理者が実行できるのは、自分のドメイン内のオブジェクトの作成、検索、更新、削除 (CRUD) だけです。

ドメインへのユーザーの割り当て

ユーザー ID を作成する際、それをドメインに割り当てるか、またはデフォルト・ドメインのままにしておくことができます。後でユーザー・プロパティを使用して、ユーザー ID のドメインを変更することもできます。

ユーザー ID は、一度に 1 つのドメインにしかアクセスできません。ユーザーを PUBLIC ドメイン、つまり共用ドメインに追加することはできません。

ドメインを作成し、それらのドメインにユーザーを割り当てる権限を持っているのは、上級管理者のみです。1 つのドメインに複数の副管理者を割り当てることは可能ですが、ある特権セット内のシステム管理特権を与えることによってこれらの管理者を定義できるのは上級管理者のみです。「新規ユーザーまたはユーザー・プロパティ (New User or User Properties)」ウィンドウの「特権セットの付与 (Grant privilege set)」フィールドには、副管理者がドメイン内で所有している管理特権が示されます。

ドメインへのユーザー・グループの割り当て

あるユーザー・グループをドメインに割り当てると、そのユーザー・グループ内の各ユーザー ID に指定されているドメインが変更されます。ユーザー ID は、一度に 1 つのドメインにしかアクセスできません。したがって、割り当てたグループに含まれているすべてのユーザー ID も新規ドメインに移されます。

ユーザー・グループ名は、一度に 1 つのドメインにしか存在できません。ユーザー・グループは、PUBLIC ドメイン、つまり共用ドメインに割り当てることができます。

ドメインへの特権セットの割り当て

ドメインに追加するユーザー ID には、関連する特権セットも割り当てなければなりません。関連する特権セットが含まれていない場合、ユーザーはタスクを実行できません。ユーザーが利用できるように特権セットを保管する場所として一番良いのは、PUBLIC (共用) ドメインです。

ドメインへのリソース・マネージャーの割り当て

特定のリソース・マネージャーを特定のドメインに割り当てることにより、それらに対するユーザー・アクセスを制限することができます。管理データベースがアクセスする新規リソース・マネージャーを定義する場合に、ドメインを選択するオプションがあります。

すべてのリソース・マネージャーのデフォルトは PUBLIC です。リソース・マネージャーへのアクセスを制限したい場合は、リソース・マネージャーをドメインに割り当てる必要があります。リソース・マネージャーを割り当てるドメインがない場合は、リソース・マネージャーを定義してから、必要なドメインを作成することができます。該当するドメインを定義した後で、リソース・マネージャー・プロパティを開いてそのドメインを選択します。

ドメインへのコレクションの割り当て

リソース・マネージャー上の特定コレクションへのユーザー・アクセスを制限するには、そのコレクションを特定のドメインに割り当てます。リソース・マネージャーが PUBLIC ドメインに含まれている場合は、コレクションを他の任意の定義済みドメインに割り当てることができます。しかし、リソース・マネージャーがすでに特定のドメインに対して定義されている場合は、そのコレクションを別のドメインに割り当てることはできません (PUBLIC ドメインに対しても割り当てることはできません)。

ユーザーがリソース・マネージャー上のコレクションにアクセスするには、リソース・マネージャーへのアクセス権が必要であるため、リソース・マネージャーへのアクセスを制限すると、リソース・マネージャー上のコレクションに対して同じ制限を強要することになります。

ドメイン間のユーザーの移動

あるドメイン内の特定のユーザーを別のドメインに移動して、追加しなければならない場合もあります。「ユーザー定義 (User definition)」ウィンドウの「説明 (Description)」フィールドを使って、ユーザーがどのユーザー・グループに入っているかを記録しておくことをお勧めします。このようにしておけば、ドメイン間でのユーザーの移動が少し楽になるかもしれません。

重要: この作業は非常に時間がかかり、また正しく行わないとシステムへのアクセスで問題が発生することがあります。上級管理者でなければ、ユーザーのドメインを変更することはできません。

次のステップに従ってください。

1. ユーザーが属するすべてのグループを見付けます。
2. ユーザーが属するすべてのグループを PUBLIC ドメインに移動するか、すべてのグループからそのユーザーを除去します。
3. このユーザーに関連するすべてのリソース・マネージャーを PUBLIC ドメインに移動し、それが終わったら、ターゲット・ドメインに移動する各リソース・マネージャーのすべてのコレクションを PUBLIC ドメインに移動します。

4. ユーザーに関連する特権セットがターゲット・ドメイン内に揃っていない場合には、ターゲット・ドメインで、ユーザーに関連するすべての特権セットを作成します。移動するものではありません。
5. このユーザーに関連するアクセス制御リストがターゲット・ドメイン内に揃っていない場合には、ターゲット・ドメインで、ユーザーに関連するすべてのアクセス制御リストを作成します。移動するものではありません。
6. ユーザーの「プロパティ (Properties)」ウィンドウを開き、ユーザーのドメインを変更して、ユーザーをターゲット・ドメインに移動します。
7. **任意:** ステップ 1 (43 ページ)、 2、および 3 (43 ページ) で移動したグループとリソース・マネージャーは、 PUBLIC ドメインからターゲット・ドメインに移動できますが、これが行えるのは、移動するグループやリソース・マネージャーに関連するユーザーがソース・ドメイン内に残っていない場合に限りです。それ以外の場合は、グループとリソース・マネージャーを PUBLIC ドメインに残しておき、別のドメイン内のユーザーと共用できるようにしておく必要があります。

覚え書き: ユーザーを PUBLIC ドメインに入れておくことはできません。ユーザーを共用することはできません。

ドメイン間のユーザー・グループの移動

重要: この作業は、正しく行わないとシステムへのアクセスで問題が発生することがあります。上級管理者でなければ、ユーザー・グループのドメインを変更することはできません。

ユーザー・グループを別のドメインに移動する場合は、次のステップに従ってください。

- ユーザー・グループが空の場合は、現行のドメインからそのグループを削除してからグループを作り直し、それをターゲット・ドメインに割り当てます。
- ユーザー・グループが空ではない場合は、次のステップに従ってください。
 1. このグループに属するすべてのユーザーを見付けます。
 2. 現行のドメインからグループを削除することで、そのグループ内のすべてのユーザーを削除します。
 3. グループを作り直し、それをターゲット・ドメインに割り当てます。
 4. この新たに作成したグループにすべてのユーザーを追加します。

ドメイン間のリソース・マネージャーの移動

上級管理者でなければ、リソース・マネージャーのドメインを変更することはできません。

リソース・マネージャーを別のドメインに移動する場合は、次のステップに従ってください。

- リソース・マネージャーにコレクションが入っていない場合は、リソース・マネージャーのプロパティを開き、ドメインをターゲット・ドメインに変更して、そのリソース・マネージャーをターゲット・ドメインに移動します。
- リソース・マネージャーにコレクションが入っている場合は、次のステップに従ってください。

1. リソース・マネージャーを PUBLIC ドメインに移動します。
2. 「プロパティ (Properties)」ウィンドウを開いて、ターゲット・ドメインを選択し、ドメインをターゲット・ドメインに移動します。
3. 「プロパティ (Properties)」ウィンドウを開いて、ターゲット・ドメインを選択し、リソース・マネージャーをターゲット・ドメインに移動します。

ドメイン間のコレクションの移動

上級管理者でなければ、コレクションのドメインを変更することはできません。

コレクションを別のドメインに移動する場合は、次のステップに従ってください。

1. コレクションが属するリソース・マネージャーを見付けます。
2. 関連するリソース・マネージャーを PUBLIC ドメインに移動します。
3. 「プロパティ (Properties)」ウィンドウを開いて、ターゲット・ドメインを選択し、コレクションをターゲット・ドメインに移動します。
4. 「プロパティ (Properties)」ウィンドウを開いて、ターゲット・ドメインを選択し、リソース・マネージャーをターゲット・ドメインに移動します。

ドメイン間の特権セットの移動

特権セットは複数のドメインに配置できるので、移動しなくてもターゲット・ドメインに追加することができます。

ドメイン間のアクセス制御リストの移動

アクセス制御リストは複数のドメインに配置できるので、移動しなくてもターゲット・ドメインに追加することができます。

第 5 章 情報マイニングの管理

このセクションでは、始めに、情報マイニングの概要および業務環境での情報マイニングの使用方法について説明します。この後に、情報マイニング First Steps、情報マイニングにおける概念、Information Structuring Tool のセクションが続き、最後に、パフォーマンス・チューニングに関する注釈のセクションがあります。

情報マイニングの概要

Enterprise Information Portal 情報マイニングは、情報の抽出および分析の数多くの局面を自動化することによって、企業が、適切な情報へ低いコストで簡単にアクセスできる仕組みをユーザーに提供するうえで役立つ、主要なテクノロジーです。

情報マイニングの最初の課題は、コンピューターで使用する構造化されていないテキスト情報を簡単に利用できるようにすることです。現在のテクノロジーでは、制限のない自然言語で示される事実に基づいた知識だけは、完全に解釈しようとしてもできません。しかし、パターン認識手法および試行錯誤法を応用したツールを使用すれば、任意のフリー・テキストから有益な情報を抽出することができます。抽出される情報の範囲は、文書に記載されている、名前、機関、または場所などの、いわゆる重要ワードの識別から、文書の要約全体までにわたります。

情報マイニングには、単一文書から情報の破片をただ抽出する以外にもまださらに多くの機能があります。大量の文書コレクションを処理する際に、情報マイニングが効力を発揮します。「マイニング」メタフォーは、知識発見プロセス（つまり、単一の文書から体系化された情報を識別して抽出し、その情報をメタデータとして保管するプロセス）のほか、興味深い現象、パターン、または傾向を見いだす、こうした特徴の文書コレクション全体での分布の分析プロセスの両方で使用されます。

Enterprise Information Portal 情報マイニング・サービス

Enterprise Information Portal 情報マイニング・サービスは、個別の文書または文書コレクションに関連した情報の作成および維持管理を行うためのインフラストラクチャーを提供します。このような情報は、メタデータと呼ばれます。文書の内容を特徴付けて、メタデータとして保管される情報の例は、以下のとおりです。

- タイトル
- 要約またはまとめ
- 名前、用語、または表記方法
- 文書が属するカテゴリー

文書をメタデータに関連付ける従来のメタデータ・ストアと情報マイニングとの違いは、情報マイニングでは、明示的には使用可能になっていない場合でも、自動的にメタデータを作成する機能が提供されている点にあります。マイニング・アルゴリズムおよび検索アルゴリズムは、メタデータを使用して検索プロセスをガイドすることにより、あるいは、コレクション内の文書を個別に見ても分からない場合が

ある文書間の興味深い関係を、メタデータに対して統計モデルを実行して検出することにより、大量の文書コレクションの中の、関連する情報にアクセスすることができます。

マイニング操作および検索操作は、元の文書コンテンツは考慮せずに、明確に定義されたメタデータ・セットに基づいて行われるので、情報マイニング・データ・ストアと呼ばれる専用ストアにメタデータを保持して、これらのプロセスの処理速度をかなり速くすることができます。これにより、アプリケーションは、単一リポジトリからメタデータを取り出すことができ、しかも、任意のリモート・コンテンツ・サーバーに戻る必要がないため、メタデータへのアクセス速度はかなり速くなります。メタデータは、検索およびナビゲーションの場合に極めて重要であり、検索結果の絞り込みに使用されることが多いため、コンテンツ・サーバーに戻ることなく、数多くのステップを実行できます。

メタデータ・ストアを使用するもう 1 つの利点は、メタデータ・ストアでは、文書に関連付けられるデータが、実際の文書とは分離された状態となることです。Web 上にある外国語の文書などの、読み取りアクセス専用で使用可能な文書の場合には、明らかに、コンテンツとメタデータを同じリポジトリ内に保持することは、選択肢とはなり得ません。

Enterprise Information Portal 情報マイニング・サービスでは、メタデータを自動作成するための、以下の機構が提供されています。

- **カテゴリー化。** ユーザー定義の分類法に基づいて、文書に 1 つまたは複数のカテゴリーを割り当てます。カテゴリー化コンポーネントには、*Information Structuring Tool* と呼ばれる分類法の作成および維持を行うためのグラフィカル・ユーザー・インターフェースを備えたアプリケーションがあります。
- **要約。** 文書の最も重要なセンテンスを抽出するので、ユーザーが、文書全体を読むか、あるいは部分的に読むかを決定する場合に役に立ちます。ユーザーは、対象となる要約の長さを指定することができるので、これにより、直接、抽出されるメタデータの複雑さと文書内の情報量のバランスを取ることができます。
- **言語識別。** 文書の作成に使用されている言語を判別します。これは、その他の情報マイニング・サービスを適用する前の、便利なプリプロセス・ステップです。
- **情報抽出。** 名前、用語、および表記方法などの、テキスト文書内の有効な語い項目を自動的に認識します。
- **クラスター化。** 文書の集合を類似するグループまたはクラスターに分割します。クラスターは、文書コレクションから自動的に派生します。

情報マイニング・サービスのコンポーネント

情報マイニング・アプリケーションでは、一般に以下の作業が実行されます。

1. ブラウズおよびナビゲート可能な方法でのデータの編成
2. 異なるデータ・ソースへのアクセス
3. 拡張検索操作を使用しての、予測または傾向に必須なデータのフィルター操作

49 ページの図 1 に、これらの情報マイニング作業を示します。

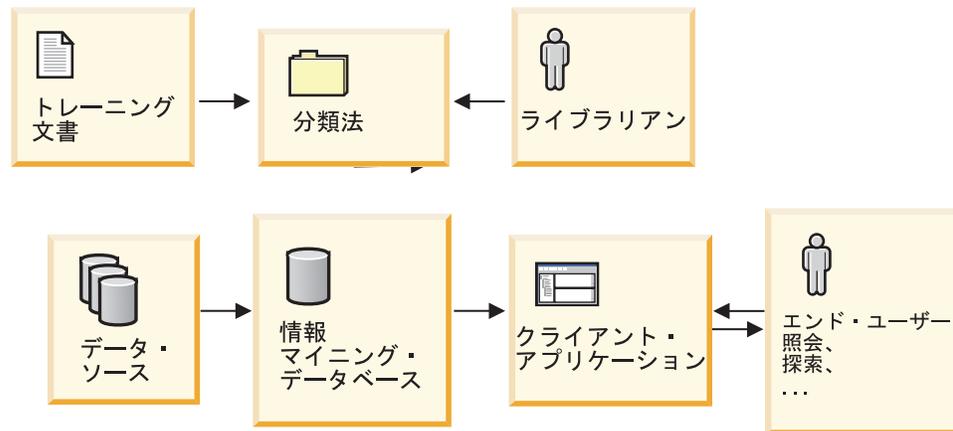


図1. 情報マイニング作業

情報マイニング機能を活用するには、文書を編成してブラウザおよびナビゲートできるようにする必要があります。この作業は通常、ライブラリアンまたは知識エンジニアが行います。ライブラリアンは、Information Structuring Tool (IST) を使用して、分類法の定義、つまり、さらなる調査が必要な文書内データについての、テーマの階層的な特徴付けを行います。IST は、グラフィカル・ユーザー・インターフェースを備えた、分類法の作成および維持が可能なアプリケーションです。カテゴリーが調整され、一度分類法が安定すると、異なるデータ・ソースにアクセスすることができます。

文書は、Web Crawler サービス (JavaBeans としてのみ使用可能) を使用して任意のコンテンツ・サーバー、または Web からインポートし、カテゴリーに割り当てることができます。

後に文書コンテンツに適用されるテキスト分析機能およびメタデータ作成機能は、非ビジュアルな JavaBeans および Java サービス API のレベルでプログラミング・インターフェースとして使用可能です。

情報マイニング JavaBeans は、迅速なアプリケーション開発のためのソフトウェア・コンポーネントであり、JavaBeans に準拠しています。Java サービス API には、アプリケーション作成のための個別の基本構成要素として、完全な情報マイニング機能が含まれています。アプリケーションの作成をサポートするために、JavaBeans ベースのサンプル、およびサンプル JSP が提供されています。

文書のコンテンツの識別は、文書のコンテンツを処理するすべての情報マイニング操作の前提条件です。このタスクのサブステップには、以下のものがあります。

1. 文書のコード・ページの識別
2. 処理する必要があるテキスト・セクションの識別。言い換えれば、マークアップ情報や、イメージなどのバイナリー・データは無視されます。

コンテンツ・サーバーの文書が構造化されてしまう場合もあるため、情報マイニング・サービスでは、文書フォーマットから、関係のあるテキスト部分の識別と抽出を行う、特定のモジュールを作成する手段が提供されています。情報マイニング・サービスでは、使用頻度の高い広範囲の文書フォーマットをカバーする、デフォルト・モジュールも使用可能です。サポートされるフォーマットのリストについては

133 ページの『第 9 章 文書フォーマット』を、また、デフォルト・モジュールのサンプルの使用方法の詳細については「アプリケーション・プログラミング・ガイド」を参照してください。

選択された各文書の文書メタデータの作成には、文書のコンテンツの処理と、知識リソース (辞書または頻度プロファイルなど) に基づいた統計方法または試行錯誤法の適用が伴います。

情報マイニング API は、以下の操作をサポートします。

- 要約
- カテゴリ化
- 言語識別
- 情報の抽出
- クラスタ化

作成された文書のメタデータは、すべて情報マイニング・データ・ストアに保管されます。

データ・ストアがいっぱいとなった場合は、文書選択を実行するもう 1 つのオプション、すなわち、このデータ・ストアの情報に基づいた文書の選択が使用可能になります。拡張検索操作により、テキスト照会とカテゴリが結合されます。これにより、検索が特定のカテゴリに属する文書に制限されます。

業務環境における情報マイニングの使用

何らかの情報マイニング・テクノロジーのインプリメンテーションをサポートしている組織のインフラストラクチャーは、通常、少なくとも、以下の役割から構成されています。

- 必ずしも情報マイニングだけに限定されない、一般的な IT のシステム管理者
- アプリケーション・プログラマー
- ライブラリアンまたは知識エンジニア
- 情報マイニング・アプリケーションを使用する人々 (エンド・ユーザー)

アプリケーションの性質によっては、上記にリストした役割のほかに、以下のようなより具体的な役割が存在する場合があります。

- Web 設計者
- 設計技術者またはコンサルタント

51 ページの図 2 に、これらの情報マイニングの役割とアクションを示します。

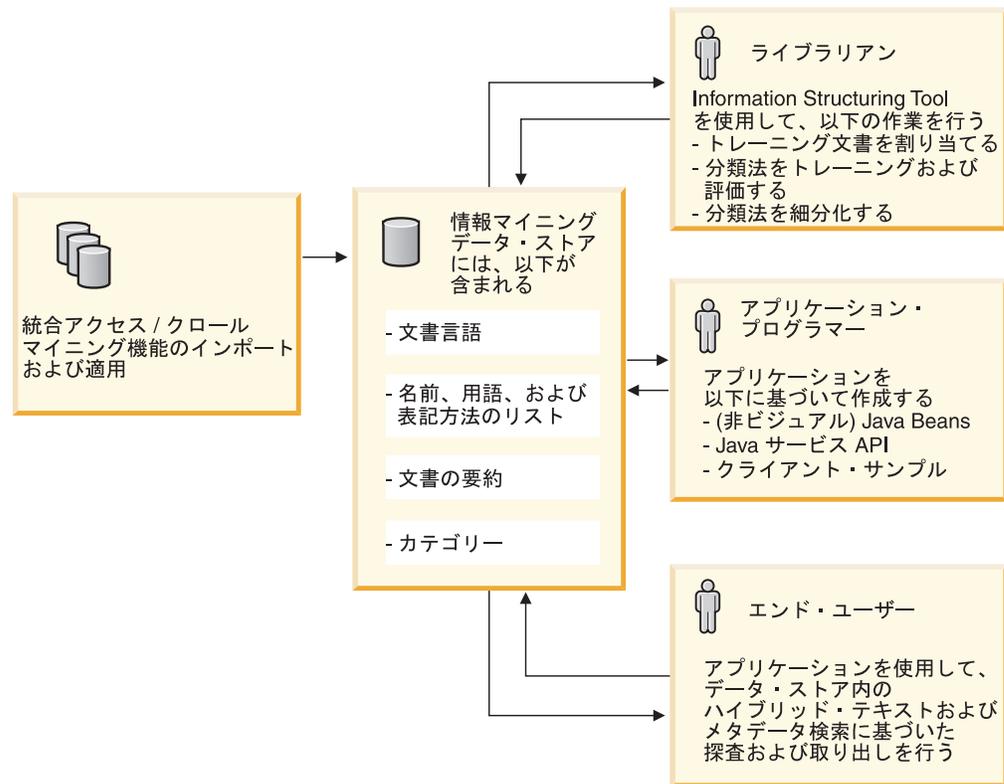


図2. 情報マイニングの役割およびアクション

システム管理者 は、ハードウェアおよびソフトウェア環境をセットアップして、たとえば、ファイル・システム・スペースおよびアクセス権限など、必要なリソースを維持管理します。システム管理者は、Enterprise Information Portal の必須コンポーネントをインストールして、各種リポジトリへのアクセスが使用可能となるように、コンテンツのソースおよび Enterprise Information Portal 管理アプリケーションを構成します。また、システム管理者は、データベース・レベルでの情報マイニング・メタデータ・ストアの管理も行います。

アプリケーション・プログラマー は、JavaBeans またはサービス API を使用してアプリケーションを作成します。JavaBeans サンプルについては「See the ワークステーション・アプリケーション・プログラミング・ガイド」、サービス API の説明については「アプリケーション・プログラミング・ガイド」を参照してください。アプリケーション・プログラマーは、Web デザイナー の支援のもとでこれを行う場合があります。

ライブラリアン または知識エンジニア は、マイニングおよび検索に使用する文書コレクションとリソースのセットアップおよび維持管理を行います。ライブラリアンは Enterprise Information Portal 管理アプリケーションを使用して、メタデータ・マッピングおよび検索テンプレートを作成し、Information Structuring Tool (69 ページの『分類法の作成』を参照) を使用して、カタログおよび分類法を定義します。アプリケーション・プログラマーが作成したアプリケーションを使用した、情報マイニング・メタデータ・ストアにデータ・ソースまたは Web からの文書の移植も、通常はライブラリアンが行います。

エンド・ユーザー は、アプリケーション・プログラマーが作成したアプリケーションを使用して、ライブラリアンまたは知識エンジニアが作成して維持管理するリソースに基づき、情報マイニングおよび検索のタスクを実行します。エンド・ユーザーとライブラリアンの間の作業配分によっては、エンド・ユーザーがコンテンツ・サーバーからの文書の選択および情報マイニング・メタデータ・ストアへの移植にかかわる場合もあります。

情報マイニングの使用例

Electro Corp. は、大量市場向けの電子装置を製造する企業です。同社のポートフォリオは、広範囲の個別構成を持つ、少なくとも 5 つの異なる製品を対象としています。

販売部門は、特定のデバイス・アプリケーション分野のカスタマー設定に関する情報を保有しています。これらの使用プロファイルでは、顧客が製品ならびに各種構成オプションをどのように使用するかを説明しています。各プロファイルは、特定のマーケティング、初公開、および関係管理戦略に関連するものです。

サービス部門は、デバイスを構成するパーツの種類、パーツの組み立て方法、これらのパーツのサプライヤー、ならびに個々のパーツの保守容易性および信頼性に関する情報を保有しています。

契約管理部門は、小売業者および下請け業者に関する情報を保持しています。ここでは、特定の種類の契約で有効な使用条件に関連する法律関係の文書にもアクセスすることができます。

最近、特定の顧客に対する売上の大幅な下落が報告されています。各種電子装置を適用する方法についての競争が移り変わり、顧客の期待は、これらの新しい技術的進歩に合わせて変化してきました。

このような変化を捕らえ、それを活用して独自の利益を得るために、Electro Corp. では、タスク・フォースを設立して、企業が再び利益を上げられるようになるための戦略を策定します。

この最初のステップは、立案者が迅速に、事情に通じた決定を下すことができるように、これに関連した情報にアクセスするための IT インフラストラクチャーを設計することです。

この種類の情報の例は、以下のとおりです。

- 競争相手の製品、フィーチャー、価格、および顧客の受容性に関するデータ
- 顧客から見た Electro Corps. の製造ラインの長所と短所についての認識
- これらの製品を主に使用する分野の傾向と見通し

この情報は、異なるハードウェアおよびソフトウェアのプラットフォーム、異なる編成レベル (たとえば、階層ファイル、索引付きファイル、またはフラット・ファイル)、および異なる文書タイプ (たとえば、データベース・レコードまたは HTML) を持つ異なるデータ・ソース上にあります。

53 ページの図 3 に、情報マイニングの使用例を示します。

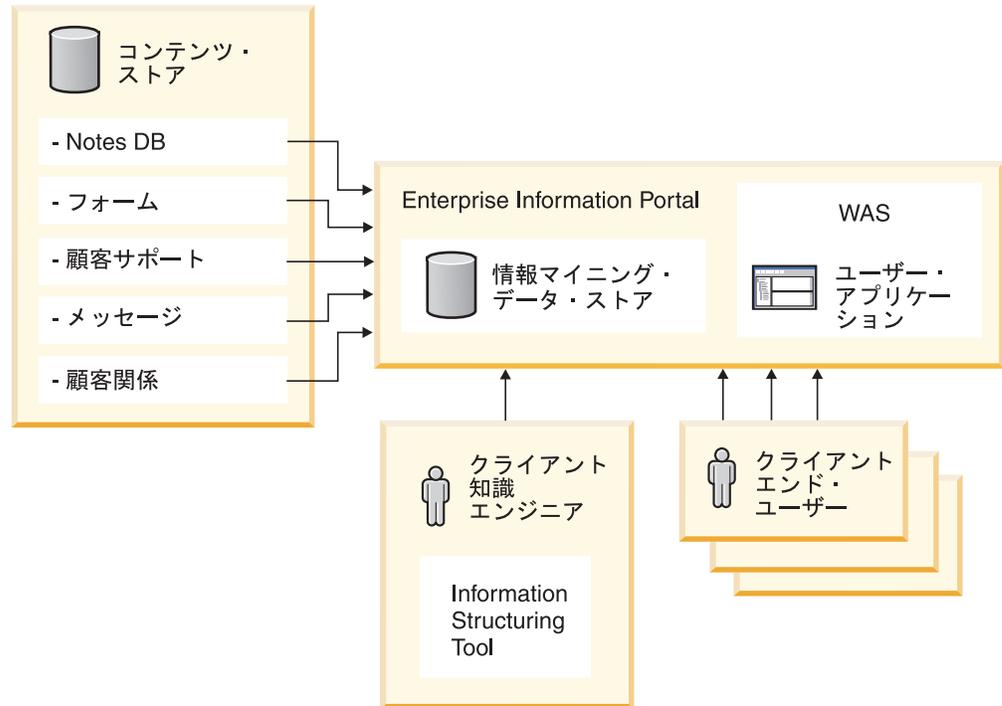


図3. 情報マイニングの使用例

タスク・フォースは、立案者のクライアント・マシンにあるシン・クライアント (ブラウザ) から、これらすべての別ソースにアクセスするポータルを作成することにします。このようなポータルを作成するためのビルディング・ブロックがすべて用意されており、業務アプリケーションの分野での標準が整っていないときに作成された既存のデータ・ソースへのカスタム接続の作成も可能であることから、Enterprise Information Portal を選択します。

以下のステップを実行する必要があります。

1. ハードウェアおよびソフトウェアのインフラストラクチャーをセットアップする
2. データ・ソースへのアクセス方式を定義し、必要な接続をセットアップして、関係のあるデータへのマッピングを作成する
3. 立案者がデータを参照およびナビゲートできるような形に、このデータを編成する
4. エンド・ユーザー・アプリケーションを作成する

ステップ 3 で、情報マイニングが実行されます。インフラストラクチャーが使用可能であり、かつ稼働中で、データ・ソースが識別されており、それに対応する接続が確立され、そして関係のあるマッピングが作成されている場合は、単一のポイントから必要なデータへアクセスすることが可能であり、統合検索を使用してこのデータのサブセットを定義することができます。次に問題となるのは、ある予測または傾向に合わせて必要なデータを取り出す方法、および、このデータを戦略計画プロセスに適合するように編成する方法です。

タスク・フォースでは、戦略計画情報の保守、編成、および更新を担当する人を知識エンジニア と定義します。さまざまなデータ・ソースにある大量の文書セットから関係のある情報を抽出するために、知識エンジニアは、過去の戦略計画にかかわ

ったスタッフに聞き取りを行い、プロセス間のやり取りの方法、および優れたやり方だとわかったものについて学びます。そして、顧客との関係およびサポート・データベースを検索します。

Enterprise Information Portal の検索機能を使用すると、顧客の名前、住所、あるいは、デバイス・プロパティにより、これらのデータベースの文書に簡単にアクセスすることができます。ただし、使用プロファイルを決定するために必要な情報は、テキスト内に隠されており、この情報を取り出す唯一の方法は、Enterprise Information Portal 情報マイニング・サービスを使用してインテリジェントな方法で文書コンテンツを分析することです。

これらのサービスにより提供される、情報の有効な種類は、たとえば、「これは、PDA に関する文書です」というように、カテゴリと呼ばれる文書コンテンツの主題別の特徴です。情報マイニングのカテゴリ化サービスは、コンテンツを分析して、文書をカテゴリに割り当てます。カテゴリは、分類法と呼ばれる項目別階層に構造化されます。明示的に使用可能で、しかも自動的に作成されたメタデータは、アクセスおよび検索の速度を上げるのに役に立つ、カタログと呼ばれるリポジトリにあります。このリポジトリは、情報マイニング・サービスにより保守されます。

Information Structuring Tool を使用して、知識エンジニアは、顧客がデバイスを使用する方法を示したカタログを定義します。

54 ページの図 4 にカタログの例を示します。

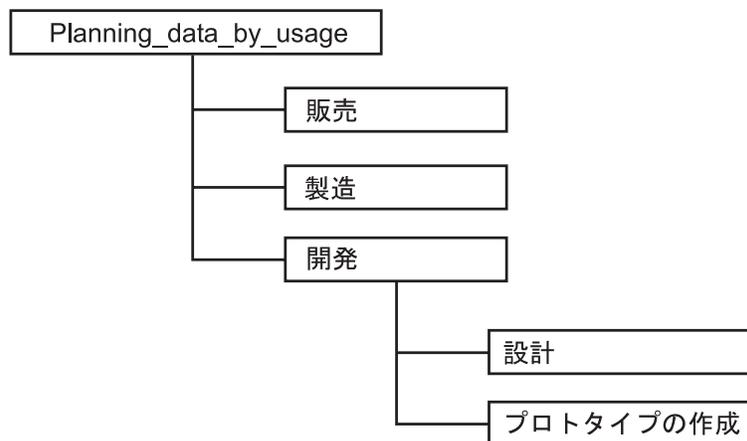


図 4. カタログの例

知識エンジニアは、全顧客およびそれらの顧客のデバイスの使用方法についての統合検索を販売データベースおよびサポート・データベースで実行して、分類法において各カテゴリを代表するトレーニング用文書セットを識別することができます。関係のあるデータをさらに詳しく調べると、新規カテゴリが現れる場合があるため、分類法の再編成につながる可能性があります。

分類法が安定しており、カテゴリごとに十分な数のトレーニング文書がある場合、知識エンジニアは、Information Structuring Tool を使用して分類法をトレーニング

グします。トレーニングにより、カテゴリー化モデルが作成されます。このモデルは、カテゴリー化サービスを利用してのカテゴリーの文書への割り当てに使用することができます。

一方、IT 部門のプログラマーは、Enterprise Information Portal シン・クライアントを使用する戦略立案者向けのエンド・ユーザー・アプリケーション、および JavaBeans または Java サービス API のいずれかを作成します。

このアプリケーションは、戦略計画で使用するようカスタマイズされた、数多くの検索テンプレートから構成されています。立案者は、これらのテンプレートを使用して、異なるバックエンドの組み合わせから検索した文書をカタログに移植することができます。カタログに移植する際に、情報マイニング・サービスは、カテゴリーを文書に自動的に割り当てます。新規使用プロファイルが識別されると、知識エンジニアは、立案者によりトレーニング資料として指定された新規文書を使用して、新規使用プロファイルに応じた分類法を再編成します。カタログは、リトレーニングされ、その新規結果が立案者に渡されます。

上記の例は、情報マイニング機能を結合して顧客の期待とニーズを反映させることで、Electro Corp. のような企業が、市場の移り変わりに後れを取らず、競争力を保持できることを示しています。

サポートされる言語およびフォーマット

Enterprise Information Portal 情報マイニングは、以下の言語をサポートします (表 5 を参照)。

表 5. サポートされる言語

言語	言語識別	情報抽出	要約	カテゴリー化	クラスター化
英語	x	x	x	x	x
ドイツ語	x		x	x	x
フランス語	x		x	x	x
デンマーク語	x				
フィンランド語	x				
イタリア語	x		x	x	x
ノルウェー語	x				
ポルトガル語	x		x	x	x
スペイン語	x		x	x	x
スウェーデン語	x				
韓国語	x		x	x	x
日本語	x	x	x	x	x
中国語 (繁体字および簡体字)	x		x	x	x

サポートされている文書フォーマットのリストについては、133 ページの『第 9 章 文書フォーマット』を参照してください。

概念

消費される情報の量は、常に増え続けています。ほとんどの組織では、潜在価値の高い情報を含むオンライン文書が大型化し、数も増えています。たとえば、顧客のフィードバック・データ、競争の激化している市場においての極めて重要な戦略的情報、または変化する新しいビジネス・チャンスへの洞察を提供する情報がその例です。情報マイニング・サービスは、大量のオンライン文書を処理するアプリケーションとして使用するよう設計されています。

システム体系

図5 は、情報マイニングのシステム体系を示しています。

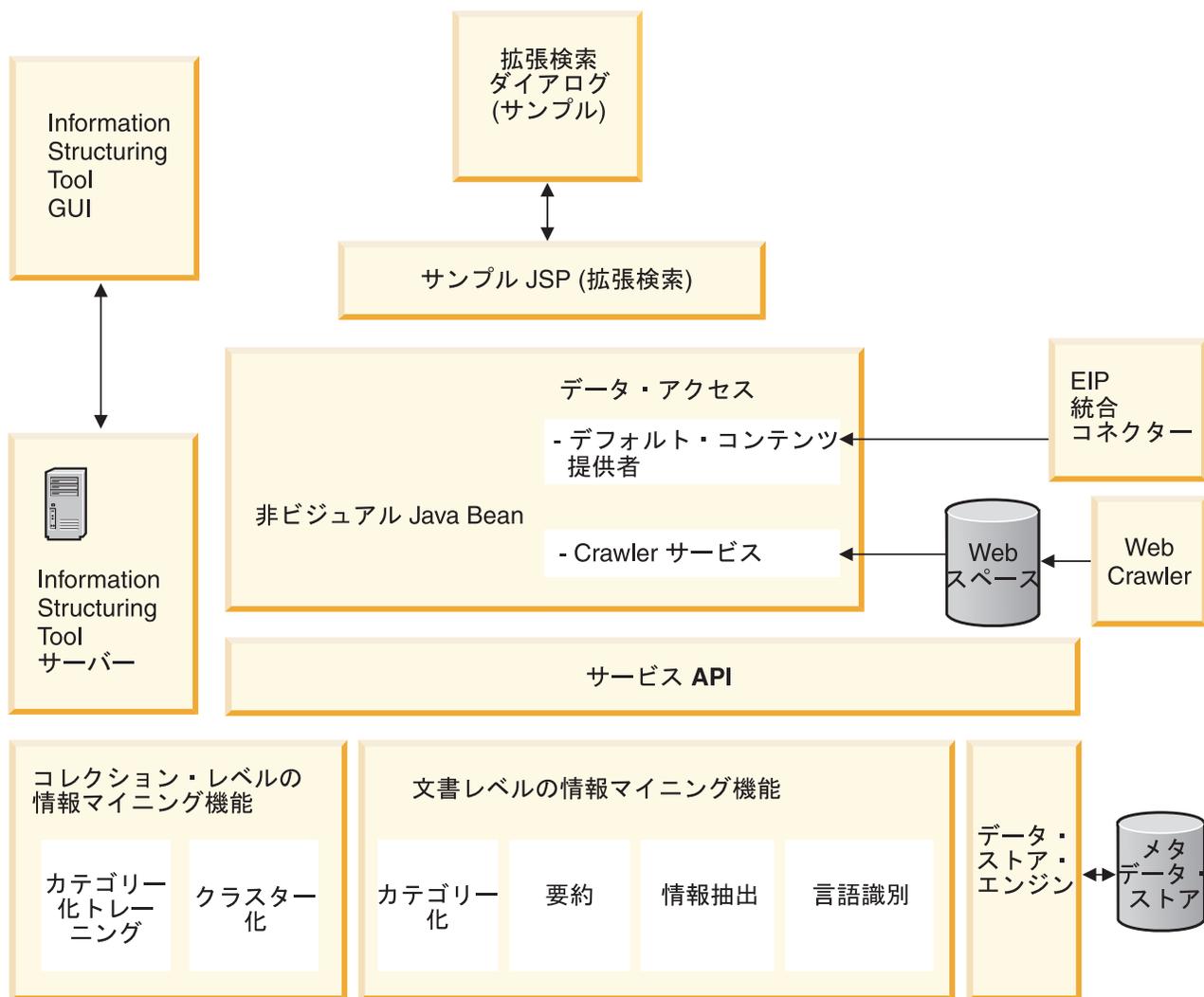


図5. 情報マイニングのシステム体系

図の右側にあるボックスは、情報マイニング・サービスが使用する以下のコンポーネントを示していますが、サービスの一部とは見なされません。

- 統合コネクター (OO API の一部)
- Web Crawler

以下のような、異なる情報マイニング機能の層があります。

1. Java サービス API

この層は、情報マイニング機能およびメタデータ持続性を、一貫性のある Java API として公開します。

2. 非ビジュアル JavaBeans

この層は、イベント・タイプと標準 Bean の規則を適用する JavaBeans 仕様に基づいた、いつでも使用可能なコンポーネントを使用します。

3. Java サーバー・ページのサンプル

このレベルは、非ビジュアル JavaBeans を使用するサンプル・コードからなり、拡張検索のアプリケーション、つまりカテゴリが制約されているテキスト検索を例示します。

4. Information Structuring Tool

分類法の作成および維持管理のための、グラフィカル・ユーザー・インターフェースを備えたアプリケーション。

情報マイニングの概念

情報マイニングの機能を完全に理解して効果的に使いこなすために、このセクションでは、その主な概念について説明します。

情報マイニング・サービスは、個別の文書または文書コレクションに関連した情報の作成および維持管理を行うためのインフラストラクチャーを提供します。文書に関するこのような情報を、文書の**メタデータ**といいます。

ライブラリーは、情報マイニング・データベースの内容の概念視点です。このライブラリーには、カタログのセットが含まれています。

カタログは、テキスト文書のメタデータ・ストアであり、以下のものを含みます。

- 各文書に保管される属性を定義する**カタログ・スキーマ**。
- **カテゴリ**の階層木構造である**分類法**。
- 文書にカテゴリを自動的に割り当てるために使用できる、文書トレーニングの結果に基づいた**カテゴリ化モデル**。このモデルは、**Information Structuring Tool** を使用して生成されます。このツールでは、分類法を作成し、トレーニングすることができます。このモデルは、カテゴリ化サービスへの入力として機能します。

スキーマにより、カタログ内の文書に対して生成または保管可能な属性の名前とタイプが指定されます。スキーマは事前定義されており、以下の属性を含みます。

- ストリング型の IKF_CONTENT
- ストリング型の IKF_TITLE
- ストリング型の IKF_AUTHOR
- ストリング型の IKF_CATEGORIES
- ストリング型の IKF_SUMMARY
- ストリング型の IKF_LANGUAGE
- ストリング型の IKF_FEATURES
- ストリング型の IKF_COMMENTS

- タイム・スタンプ型の IKF_DATE
- 整数型の IKF_IDNUMBER

カタログは、インポートされた文書から抽出または作成された情報を保管するために、カタログ・スキーマに従ってレコードを作成します。レコードには、固有 ID と名前値のペアのセットがあります。永続オブジェクト ID (PID) として知られる固有 ID により、作成されたレコードが元の文書ソースにリンクされます。

58 ページの図 6 に、サンプル・レコードを示します。

レコード	
IKF_TITLE	「Birds」
IKF_AUTHOR	「J. Smith」
IKF_SUMMARY	「これは Birds という本のサマリーである」
IKF_CATEGORIES	鳥/虫を食べるもの
IKF_DATE	07/01/2001

図 6. サンプル・レコード

JavaBeans またはサービス API を使用して情報マイニングのレコードの値をセットする場合は、使用する値が定義済みサイズ限界内であることを確認してください。さもないと、DKIKFSizeOutOfBoundsException の状態になります。定義済み限界は、次のとおりです。

表 6.

キー	最大サイズ (バイト)
IKF_CONTENT	209715
IKF_TITLE	2048
IKF_AUTHOR	2048
IKF_CATEGORIES	8192
IKF_SUMMARY	8192
IKF_LANGUAGE	8
IKF_FEATURES	524288
IKF_COMMENTS	8192

レコードは、作成された後、カテゴリーに割り当てられてカタログに保管されます。通常、カテゴリー化の結果を使用して、適切なカテゴリーを選択することができますが、カテゴリーは、レコード内に保管されている別の値によっても選択可能です。レコードは、テキスト検索を可能にする文書コンテンツの索引付けも含んでいるため、カテゴリーに割り当てる必要があります。各カタログには単一のテキスト索引があります。これは、すべての検索結果が常に自動的にカタログ検索の有効範囲内にあることを意味します。

データ・ストア・エンジンは、永続データ・ストアへのアクセスを維持するコンポーネントです。

情報マイニング・ツール

情報マイニング・サービスでは、オンライン文書进行处理する機能が提供されています。提供されている機能は次のとおりです。

- Information Structuring Tool。カタログを作成し、維持管理します。
- 言語識別サービス。文書作成に使用されている言語を自動的に検出します。
- カテゴリー化サービス。Information Structuring Tool を使用して事前に定義したカテゴリーに、文書を自動的に割り当てます。
- 要約サービス。文書内のワードとセンテンスを分析し、文書の要約を作成します。
- 情報抽出サービス。ドメインに依存する用語を定義しなくても、テキスト内の有効な項目を自動的に認識します。
- クラスタ化サービス。文書のセットをグループまたはクラスターに分割します。各クラスター内の文書は、共通の機能を共有します。クラスターは事前定義されておらず、自動的に派生します。
- 拡張検索。特定のカテゴリーに制限されたカタログに保管されている文書内のテキストを検索します。

Information Structuring Tool

Information Structuring Tool は、ライブラリーと呼ばれるカタログ・セットの作成および維持管理を行う手段を備えた Web ベース・アプリケーションです。カタログは、文書から抽出されたメタデータを保管するために使用され、事前定義された編成に関連する文書を編成するために使用される分類法に関連付けられます。分類法は、主題内容に従って文書を分類する、カテゴリーの階層構造です。

たとえば、Information Structuring Tool を使用すると、ライブラリアンは、鳥の捕食習性を説明するカタログを定義することができます。

60 ページの図 7 に、サンプル・カタログを示します。

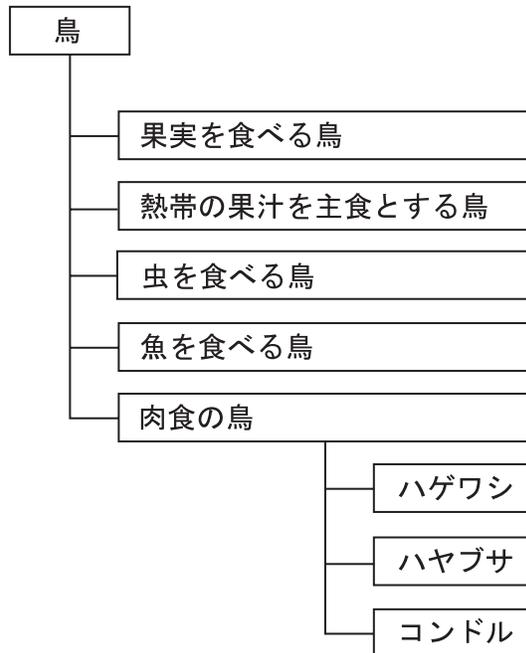


図7. サンプル・カタログ

正しく構成されたカテゴリーのシステムは、大量のデータの中にある関連する情報の検索に大いに役立ちます。カテゴリーを選択して文書のコレクションの用途に突き合わせ、あらかじめサンプル文書を使用してそれらのカテゴリーをトレーニングしておく必要があります。Information Structuring Tool を使用して作成されたカテゴリー化モデルは、その後、カテゴリー化サービスにより、文書にカテゴリーを自動的に割り当てるために使用されます。

Information Structuring Tool には、以下の機能があります。

- カタログの作成、名前変更、および削除。
- カタログの説明の編集。
- カテゴリーの作成、名前変更、および削除。
- トレーニング文書のカテゴリーへの追加、およびカテゴリーからの除去。
- トレーニング文書の内容の表示。
- カタログ・トレーニング・プロセスの開始および停止。
- カタログ内のトレーニング・データの品質に関するフィードバックの取得。

Information Structuring Tool のインストールおよび使用の方法についての詳細は、69 ページの『分類法の作成』を参照してください。

言語識別

言語識別サービスは、与えられた言語のセットのなかで、テキスト文書の作成に使用された可能性の最も高い言語を選択します。

言語識別サービスは、各文書に対して、ランクされた言語のリストと信頼性値を戻します。言語キーは、ISO 標準 639 に従って 2 文字のコードで指定されます。信頼性値は、文書がその言語にどの程度適合しているかを示す尺度であり、0 (不適合) から 1 (適合) までの範囲の浮動数で表されます。言語識別アルゴリズムは、単一の

言語で作成された文書の言語を判別するように設計されています。したがって、マルチリンガル文書の場合、信頼性値によるランキングは、必ずしも文書の正しい言語を反映しているとは限りません。

検出できる言語は、以下のとおりです。

- 英語 EN
- ドイツ語 DE
- フランス語 FR
- デンマーク語 DA
- フィンランド語 FI
- イタリア語 IT
- ノルウェー語 / ブークモール NB
- ノルウェー語 / ニーノシュク NO
- ポルトガル語 PT
- スペイン語 ES
- スウェーデン語 SV
- 韓国語 KO
- 日本語 JA
- 中国語 (簡体字および繁体字) ZH

以下のプロパティを設定することができます。

- **maxResults** (Java サービス API を使用する場合のみ):

各文書ごとに判別され、戻される言語の最大数。0 以上の整数値です。デフォルト値は 1 であり、これは、最高ランクの結果が戻されることを意味します。値が 0 に設定されていると、言語の信頼性値が 0.01 未満の場合を除き、認識されたすべての言語がランク順に戻されます。

言語認識は、他の情報マイニング・サービスのためのプリプロセス・ステップとして使用することができます。たとえば、情報抽出を行う前に、言語認識を使用して、英語または日本語の文書をすべて検索することもできます。

カテゴリー化

カテゴリー化は、文書にカテゴリーを割り当て、Information Structuring Tool を使用して作成された事前定義済みの編成モデルに関連する文書を編成する手段です。

したがって、カテゴリー化サービスを使用する前に、分類法を定義してトレーニングし、Information Structuring Tool を使用して、そのようなモデルを作成する必要があります。

カテゴリー化の結果には、カテゴリーと、文書がそのカテゴリーにどの程度適合しているかを示す信頼性値とが含まれます。そのような結果のセットが、それぞれの文書ごとに戻されます。リストは、戻された信頼性値に従ってランクされます。

以下のプロパティを設定することができます。

- **maxResults:**

それぞれの文書ごとに戻されるカテゴリーの最大数。デフォルト値は -1 であり、これは、すべてのカテゴリーが戻されることを意味します。結果リストはランク付けされます。

- **minConfidence:**

文書に割り当てられる信頼性値は、文書がそのカテゴリーにどの程度適合しているかを示します。パラメーター `minConfidence` は、0 (適合度が極めて低い) から 1 (適合度が高い) までの最小値を指定します。デフォルト値は 0 に設定されています。これは、文書に割り当てられたカテゴリーがすべて戻されることを意味します。結果リストには、適合度が高いものから適合度が極めて低いものまでがランクされます。

- **catalogName:**

このプロパティは、カテゴリー化に使用するカタログを指定します。カタログは、カタログ化を行うために、Information Structuring Tool を使用して作成およびトレーニングすることができます。

サポートされている言語のリストについては、55 ページの『サポートされる言語およびフォーマット』を参照してください。

要約

文書の要約は、文書コンテンツの特性である、文書から抽出された文のコレクションからなります。要約ツールは、たとえば、その文書には関連性があり、すべて読む必要があるかどうかを判断する際、あるいは照会結果の一部として戻された場合に、文書リンクをたどるかどうかを判断する際に役立ちます。

要約の結果には、要約が、単一ストリングおよびデータ構造 (マトリックス) として含まれます。アプリケーションは、この要約を使用して個々の文を選択し、それらのセンテンスが互いに隣接して出現するかどうかを判別することができます。

要約サービスは、各種のモードで使用することができます。モードにより、値 `maxLength` および `ratio` を使用して要約の長さを決定する方法が決まります。

- **maxLength:**

要約内のセンテンスの最大数。作成される要約は、`maxLength` より長くはなりません。デフォルト値は 3 です。

- **ratio:**

文書全体の長さに対する文の数。作成される要約の長さは、文書全体の長さによって決まります。デフォルト値は 0.1 です。

- **mode:**

要約の長さの設定に必要な `maxLength` と `ratio` の関係を決定します。各モードは、以下のとおりです。

- **MODE_LESS_THAN_MAXLENGTH:**

要約は、文の最大数が `maxLength` の値となります。これはデフォルト・モードです。

- **MODE_EQUALS_RATIO:**

要約は、文の数が `ratio` の数と同じになります。これは、`ratio` に文書内のセンテンスの合計数を掛けて決定されます。

- **MODE_EQUALS_RATIO_BUT_AT_MOST_MAXLENGTH:**

要約は、文の数が最低でも `ratio` の数 (`ratio` の値に文書内の文の合計数を掛けた値) となりますが、文の数は `maxLength` の値未満となります。

サポートされている言語のリストについては、55 ページの『サポートされる言語およびフォーマット』を参照してください。

情報抽出

文書を分析する際は、文書内容に関する情報をもたらす項目の抽出が重要な作業となります。以下のキー・エレメントを使用することができます。

- 文書に重要性があるかどうかの評価に役立つ重要な情報を示すために使用する。
- 照会の詳細化に使用する主な概念を検索し、保管するために使用する。
- 関係する文書の収集のための基準として使用する。

キー・エレメントの例としては、ワード、名前、またはマルチワード用語などの用語項目があります。

英語の場合、情報抽出サービスは、検索したキー・エレメントを正規化し、キー・エレメントが同じエンティティを示しているか、同じ概念を表している場合に、テキスト内のそのようなキー・エレメントの出現をグループにまとめます。たとえば、ある文書内で James J. Smith、Mr. Smith、James、および Smith が出現する場合、これらを同じ正規形にマッピングすることにより、すべて同一人物を示すものとしてマークされます。また、ワードの変化形も正規形に (たとえば、children は child に) マップされます。

しかし、日本語の場合は、すべてのキー・エレメントは文書内で出現したときに抽出されます。日付、時間、および通貨の表記方法を除き、正規化は行われません。この場合、正規化は ISO8601 および ISO4217 に準拠します。

情報抽出サービスを使用すると、以下の項目について文書を分析できます。

- 単一およびマルチワード用語項目。たとえば、発表、価格、製品サイクルなど。
- 場所名、人名、および組織名。たとえば、Washington、Bush、Data Management Academy など。
- 省略語。たとえば、MB (M バイト) など。
- 日付、通貨、および数値の用語。たとえば、11 Jan. 1958、01/11/58、\$30、30 ペンス、4.5、5000 など。

以下に示す 3 つのタイプの情報の抽出について指定することができます。

- 名前
- 用語
- 表記方法

Java サービス API を使用すると、以下にリストされているサブタイプと、サブタイプが抽出された用語項目にどの程度適合しているかを示す信頼性値を識別することもできます。信頼性値の範囲は 0 (不適合) から 1 (適合) までの間です。タイプ名、用語、および表記方法のサブタイプには、以下のものがあります。

- **名前 (Name)**
 - 場所 (たとえば、Montreal または London)
 - 人物 (たとえば、Tim Brown)
 - 組織 (たとえば、Smith and Son)

- 不明 (たとえば、Smashing Pumpkins、Silicon Valley、CCTV (完全な形のない省略形))
- その他 (たとえば、AIS Plan、ISO Conference、Internet、Privacy Act Officer、JCAHO Performance Report)

• **用語 (Term)**

- 未指定の用語 (たとえば、エンターテインメント複合企業、美術界、クラス変数、ソース・コード、データ定義、プロセス改善イニシアチブ)

• **表記方法 (Expression)**

- 基数 (たとえば、4、50、70)
- 序数 (たとえば、第 4、第 50)
- パーセント (たとえば、12%、60 パーセント)
- 日付 (たとえば、07/28/98)
- 時間 (たとえば、18 時間、4 時)
- 通貨 (たとえば、DM90、30 ポンド)
- 省略語 (たとえば、NY)

情報抽出が機能するのは、英語および日本語文書の場合のみです。言語識別サービスをプリプロセス・ステップとして使用すると、文書コレクション内の、英語および日本語のどちらでもない文書を識別することができます。情報抽出サービスと他のマイニング機能を結合することができます。たとえば、このサービスを要約サービスのプリプロセス・ステップとして使用し、テキサス州知事ではなく大統領としての Bush に関する文書のみを要約します。

クラスター化

クラスター化ツールは、同様の文書が互いにグループにまとめられ、異なるグループ (クラスター) 内の文書がその内容に関して相互に区別されるように、文書のコレクションを配置します。このように、クラスター化は、大規模な文書コレクションの概要を提供し、関連する文書を識別する手段として使用することができます。クラスター化を同じように使用して、アプリケーション領域内のトレーニング文書をクラスター化することによって、Information Structuring Tool を使用した分類法の構築をサポートすることができます。また、クラスター化は、コレクション内にあって新しい傾向または新しいテクノロジーを示すと考えられる同種の文書の検索、および競合の分析のために重要と考えられる、複製または極めてよく似た文書を検索する場合にも便利です。

クラスター化は、各クラスター内の文書がその内容に関して可能な限り相互に同じになり、また、クラスターができる限り相互に異なるように、文書を編成してクラスターにする反復プロセスです。クラスター化は、文書レベルで機能するカテゴリー化または要約のような上記の情報マイニング・サービスとは対照的に、文書のコレクション全体について機能します。クラスター化は、各文書の代表的な特徴を互いに比較し、その特徴の類似性に従って文書をグループ化することによって機能します。

クラスター化のフェーズでは、新規文書を文書セットに追加することはできません。

以下のプロパティを設定することができます。

- **maxClusterCount**
戻されるクラスタの最大数。
- **minClusterCount**
戻されるクラスタの最少数。
- **clusterFeatureCount**
クラスタごとに戻されるラベル (キーワード) の数。

ただし、これらの値は、クラスタ化機能にはバインディングされずに、単にガイドライン境界としてのみ機能します。クラスタ化サービスの出力は、結果リストです。

クラスタ化は、英語の文書でのみ機能します。言語識別サービスをプリプロセス・ステップとして使用して、文書コレクション内の、英語ではない文書を識別することができます。

拡張検索

Enterprise Information Portal コンテンツ・サーバー全体で実行される標準 Enterprise Information Portal 検索とは対照的に、いわゆる拡張検索は、Information Structuring Tool で作成されたカタログに ID が保管されている文書のみを検索します。検索をさらにしぼりこむために、拡張検索照会は、テキストを検索するだけでなく、この検索を特定のカテゴリ内の文書に制限することもできます。

以下のパラメーターを設定することができます。

- **catalogName:**
これは、検索に使用するカタログを指定します。カタログは、Information Structuring Tool を使用して作成およびトレーニングすることができます。
- **maxResults:**
それぞれの照会ごとに戻される検索結果の最大数。デフォルト値は 0 であり、これは、すべての結果が戻されることを意味します。

以下のタイプの照会をサブミットすることができます。

1. 純粋テキスト照会。この検索は、テキスト照会に一致するすべての文書を戻します。結果リストは、関連性により順序付けられています。
2. 純粋カテゴリ検索。この検索は、そのカテゴリに割り当てられているすべての文書を戻します。結果は、任意の順序になっています。
3. テキストとカテゴリの結合検索。この検索は、テキスト照会に一致し、かつ、そのカテゴリに割り当てられているすべての文書を戻します。結果リストは、関連性により順序付けられています。

システムに対してサブミットされる拡張検索照会は、常に特定のカテゴリに結合されます。これは、カタログ検索の有効範囲と呼ばれます。クロス・カタログ検索は不可能です。これは、カタログが、重視すべきインポート済み文書のビューを表しているためです。

照会ストリングの BNF (照会構文) は、次のとおりです。

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

```

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

```

- 端末、ストリングおよび数値は、共通の用語を表します。
- category_operator '=' により、検索の有効範囲は、1 つのカテゴリーのみに制限されます。
- category_operator '>=' により、検索はこのカテゴリーと、カテゴリー・ツリー内にあるそのすべてのサブカテゴリーに拡張されます。
- CONTAINS 文節内のストリング検索には、単一文字の場合には ('_')、および任意の数の文字の場合には ('%') というワイルドカードを含めることができます。たとえば、_LOB は BLOB と CLOB に一致する可能性があり、また、%name は filename に一致する可能性があります。ストリング演算子 CONTAINS を使用して照会できるのは、検索可能というフラグが立てられているスキーマ属性 (たとえば、IKF_CONTENT) のみです。
- LIKE 文節内の検索ストリングにも、SQL で使用されるようなワイルドカードを含めることができます。
- 現在サポートされている属性名の完全なリストについては、57 ページの『情報マイニングの概念』を参照してください。

以下に、照会の例を示します。

- 純粋テキスト照会:

```

("IKF_CONTENT" CONTAINS "'southern Africa'") AND NOT
("IKF_CONTENT" CONTAINS "'Cape'")

```

- 純粋カテゴリー照会:

```

("DKIKFCATEGORY" >= "birds/Fruit eaters")

```

- テキストとカテゴリーの結合照会:

```

("IKF_CONTENT" CONTAINS "'South Africa'") AND
("DKIKFCATEGORY" >= "birds/Birds of prey and scavengers/Falcon")

```

- 属性照会:

```

("IKF_SUMMARY" LIKE "humming birds in the tropics")

```

または

```

("IKF_FEATURES" LIKE "Goethe") AND ("IKF_TITLE" = "Faust")

```

プログラミング・インターフェース

情報マイニング機能は、アプリケーションを、次に示す 2 つのものとしてビルドするために使用可能です。

- Java サービス API
- 情報マイニング JavaBeans

Java サービス API は、Information Structuring Tool の一部であるカタログの維持を除き、すべての情報マイニング機能を Enterprise Information Portal サービスとして統合します。このサービスにより、Java RMI に基づくクライアント / サーバー通信が可能になります。

Java サービス API を使用するアプリケーションでは、以下のことが可能です。

- 文書の作成に使用されている言語の判別
- テキスト文書の要約の作成
- 文書へのカテゴリの割り当て
- テキスト文書からの情報 (たとえば、名前、用語、または表記方法) の抽出
- 類似する文書のグループ化
- カatalog内の文書のメタデータの保管および検索
- 特定のカテゴリに制限された文書および属性 (たとえば、要約) についてのテキスト検索の実行。

Java サービス API は、直接メソッド呼び出しを使用してローカル・モードで実行したり、Java リモート・メソッド呼び出し (RMI) を使用してリモート・モードで実行することができます。リモートで実行すると、1 つのサーバーを Web アプリケーションを実行するアプリケーション・サーバーとして構成し、別のサーバーをテキスト分析、索引付け、および検索を行う情報マイニング・サーバーとして構成することができます。サーバー・タスク機構は、情報マイニング・サーバー (リモート・マシン) にすべてのタスクを送信して、すべての処理をそのマシンで実行できるようにする手段です。

Web Crawler を使用している場合は、JavaBeans を使用して、対応するアクセス機構をインプリメントする必要があります。Web Crawler アクセスは、Java サービス API レベルでは利用不能です。

情報マイニング Java サービス API の詳細については、「ワークステーション・アプリケーション・プログラミング・ガイド」を参照してください。

68 ページの図 8 に、情報マイニングのリモート構成を示します。

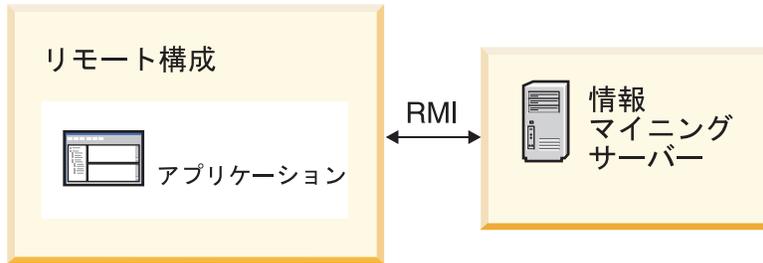


図8. 情報マイニングのリモート構成

情報マイニング Bean は、迅速なアプリケーション開発のための高水準の Java API であり、JavaBeans 仕様に従って作成されています。この Bean はサーバー・タスク処理をサポートしないため、パフォーマンス上の理由から、この Bean を使用するアプリケーションの開発はすべて同じマシン上で行う必要があります。

それぞれの Bean は、イベント・ドリブンで使用することができ、一部の Bean では、直接呼び出すことの可能なメソッドを提供します。既存の Enterprise Information Portal Bean との統合は、Enterprise Information Portal Bean が使用する結果イベントのサポートを介して行われます。つまり、統合検索および Web Crawler 結果イベントが情報マイニング・サービスと互換性があり、情報マイニングの結果を EIP で処理することができます。情報マイニング Bean の詳細については、「ワークステーション・アプリケーション・プログラミング・ガイド」を参照してください。

First Steps

Enterprise Information Portal 情報マイニングの First Steps は、ストーリーボード・チュートリアルです。このチュートリアルでは、知識エンジニア、管理者、またはアプリケーション・プログラマー向けに、実際のシナリオに基づき IBM 情報マイニング・テクノロジーをビジネス環境に応用する方法を示しています。チュートリアルは、以下のような構成になっています。

- 簡潔な概要
- First Steps を開始する前に行うこと
- サンプル・データの編成
- このデータへのアクセス
- サンプル・クライアントの使用
- サンプル・データの除去
- 追加の解説書

情報マイニングの First Steps にアクセスするには、
`<CMBROOT>%ikf%firststeps%first_steps.html` を実行してください。

情報マイニングの First Steps を Windows 上で実行するには、「スタート」ボタンをクリックして「**Enterprise Information Portal for Multiplatforms 8.1**→**<Information Mining>**→**First Steps**」を選択してください。

Enterprise Information Portal 情報マイニングの First Steps は、インストール検査としても機能します。

分類法の作成

Information Structuring Tool は、ライブラリーと呼ばれるカタログ・セットの作成および維持を行う手段を備えた Web ベース・アプリケーションです。カタログは、メタデータを保管するのに使用され、カタログ内の情報を編成する分類法に関連付けられます。分類法は、主題内容に従って、文書を分類するカテゴリーの階層構造です。たとえば、分類法のトップレベルには、ビジネス、文化、スポーツ などのカテゴリーを設け、さらに、スポーツ を、1 つ下のレベルで、チーム・スポーツ と アスレチック に細分することができます。さらにもう 1 つ下のレベルでは、チーム・スポーツ は、サッカー、野球、テニス に細分されます。

カテゴリーにトレーニング文書を割り当て、カタログをトレーニングすることにより、Information Structuring Tool は、カテゴリー化モデルを作成します。その後、カテゴリー化サービス Bean がこのモデルを使用して、文書にカテゴリーを割り当てることができます。

Information Structuring Tool のインストール

Information Structuring Tool は、たとえば、IBM WebSphere Application Server (WAS) サブレット・エンジンのような、サブレット・コンテナ内の Web アプリケーションとして配置する必要があります。

ただし、同じ情報マイニング・インスタンスに対して 2 つの Information Structuring Tool Web アプリケーション (たとえば、1 つは IST1、もう一方は IST2 と命名されたアプリケーション) を実行することはできない点に注意してください。

Information Structuring Tool を配置する前に、正しい WAS バージョンがインストールされ、構成済みであることを確認してください。詳細については、「*Enterprise Information Portal* の計画とインストール」を参照してください。

Information Structuring Tool の配置に必要なユーザーのアクセス権限は、以下のとおりです。

- Windows の場合: Administrator 権限
- AIX の場合: root ユーザー特権
- Sun の場合: root ユーザー特権

始めに

アプリケーションでは、Web ベース・インターフェースが提供されており、これによって、分類法の定義、維持、およびトレーニングを行うことができます。2 つのフレームがあります。左のフレームは、「カタログ」ビューと呼ばれ、分類法の作成および維持に使用されます。右のフレームは、「ノートブック」と呼ばれ、アプリケーションの使用に関する情報が表示されます。ノートブックには、カタログ用文書のトレーニングおよび評価を行うために使用される一連のタブが表示されます。アプリケーションを使用するには、最初に、左のカタログ・ビュー・フレームで、カタログおよびカテゴリーを作成する必要があります。

アクセス権限

Information Structuring Toolでは、ユーザー名およびパスワードの情報は Enterprise Information Portal により保持されます。Enterprise Information Portal で認識されるユーザー名とパスワードのみ入力することができます。

Information Structuring Tool を開始すると、セキュリティー警告メッセージが表示されます。これは、トレーニング文書のアップロードに使用される Java アプレットが、ユーザーのファイル・システムへの読み取りアクセスを必要とするためです。これを拒否した場合、トレーニング文書をアップロードすることができず、したがって、分類法をトレーニングすることができません。

Information Structuring Tool は、マルチユーザー環境で稼働させることができます。複数のユーザーが同じ分類法を表示できるようにするため、Information Structuring Tool では、カタログおよびそのカテゴリーへのアクセスを制御するロック機構を提供しています。

ユーザーは、カタログを明示的にロックする (カタログでの作業の開始前にカタログを選択してロックする) ことを選択するか、または、カタログでの作業を開始する (たとえば、トレーニング文書を追加する。この場合、カタログは自動的にロックされ、アクセス競合が回避される) ことができます。他のユーザーは、このカタログを表示できますが、カタログをロックしたユーザーが再度アンロックしない限り、カタログの内容を変更できません。

Information Structuring Tool が配置されているアプリケーション・サーバーがシャットダウンした場合、すべてのロックが削除されるという点に注意してください。

分類法の定義

カタログは、複数のカテゴリーで構成されるツリー構造の分類法のアンカー・ポイントです。

新規カタログを定義して、適切なカテゴリーを選択するステップは、以下のとおりです。

1. 定義するカテゴリーを決定し、新規の分類法を作成します。

「カタログ」ビューで、「**ライブラリー (Library)**」を選択して、右マウス・ボタンをクリックします。メニューが表示されます。「**新規カタログ (new catalog)**」を選択すると、カタログ・アイコンが作成されます。カタログ名を入力して、このアイコンの名前を変更し、「**実行 (Enter)**」をクリックします。同じ名前のフォルダーが作成されます。これは、ルート・カテゴリーです。ノートブックのコンテンツが変更されます。

ライブラリーに新規カタログを追加するもう 1 つの別の方法は、Information Structuring Tool 以外 (たとえば、ファイル・システムなど) で作成された、既存の分類法をインポートすることです。詳しくは、73 ページの『トレーニング文書のアップロード』を参照してください。

分類法でアクティブに作業するには、カタログをロックする必要があります。カタログを作成する際、これは自動的に行われます。既存のカタログをロックするには、そのカタログを選択して、右マウス・ボタンをクリックします。メニュー

が表示されます。「**カタログのロック (lock catalog)**」を選択すると、カタログ状況のアイコンが変化します。アイコンを使用して、以下の異なる種類のカタログ状況が表示されます。



分類法ツリーは縮小され、どのユーザーからもロックされていない。



分類法ツリーは展開され、どのユーザーからもロックされていない。



分類法ツリーは縮小され、現行ユーザーによってロックされている。



分類法ツリーは展開され、現行ユーザーによってロックされている。



分類法ツリーは縮小され、別のユーザーによってロックされている。



分類法ツリーは展開され、別のユーザーによってロックされている。

72 ページの図 9 に、2 つのカタログの例を示します。

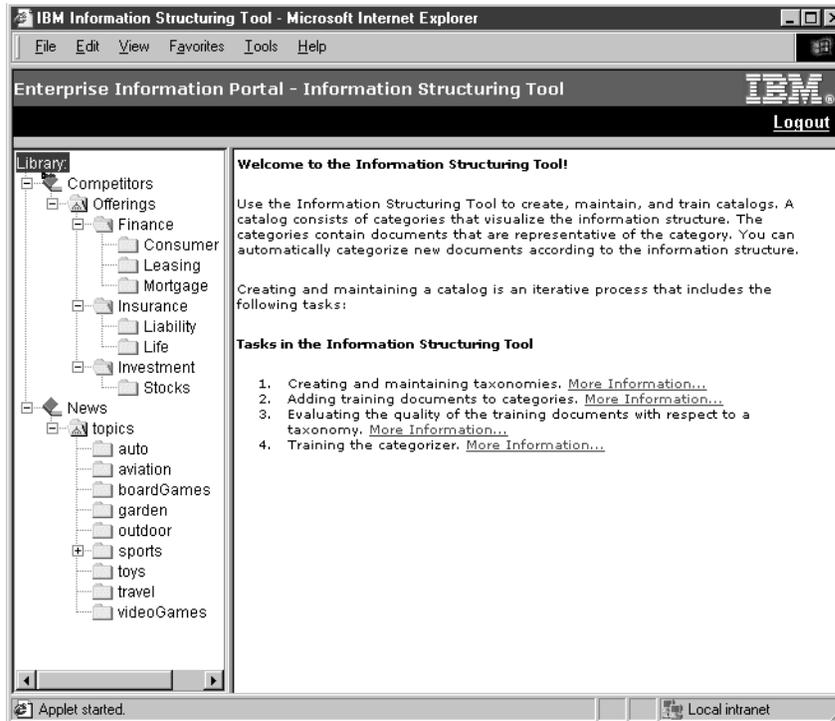


図9. カタログの例

2. カタログが作成されて、ロックされていると、カテゴリの追加、名前変更、または削除を行うことができます。新規カタログを最初に作成するとき、そのカタログと同じ名前を持つルート・カテゴリも作成されます。このカテゴリの名前は変更することができます。これを行うには、そのカテゴリを選択して、右マウス・ボタンをクリックします。「名前変更 (rename)」を選択して、新規名を入力し、「実行 (Enter)」を押します。

新規カテゴリを追加するには、新規サブカテゴリを追加するカテゴリを強調表示して、右マウス・ボタンをクリックします。「新規カテゴリ (new category)」を選択して、新規名を入力し、「実行 (Enter)」を押します。ツリー構造内の同じレベルのカテゴリ名は、固有のものでなければなりません。

ルート・カテゴリは、削除できません。これが削除されるのは、カタログが削除される場合のみです。カテゴリを削除すると、全サブカテゴリ (ツリー構造内の下位のカテゴリ)、カテゴリ内の全トレーニング文書および全レコードが除去されます。

3. 選択したカタログに関連した詳しい説明情報が、「プロパティ (Properties)」タブに表示されます。最初、説明フィールドは空です。カタログ説明の追加または編集を行うには、「説明の編集 (Edit Description)」ボタンをクリックします。ウィンドウが表示され、ここに説明を入力することができます。

トレーニング文書の選択

カテゴリ化モデルの品質は、各カテゴリに割り当てられたトレーニング文書の品質に大きく依存します。

トレーニング文書は、Enterprise Information Portal でサポートされている、いずれかの文書フォーマットで表されていない必要があります。サポートされているフォーマットのリストについては、133 ページの『第 9 章 文書フォーマット』を参照してください。

適切なトレーニング文書セットを選択することが、きわめて重要です。文書では以下のようなことを守る必要があります。

- カテゴリーを代表している。
- かなりの分量の記述テキストが収められていて、マークアップやワード・リストはあまり多くない。
- すべてが同じ文体で書かれている。たとえば、他の文書が報告書スタイルである場合には、散文で書き込まないようにしてください。
- すべてが、ほぼ同じ長さで、なるべく長すぎないようにする。トレーニング文書も、カテゴリー化サービスを使用してカテゴリー化する文書とほぼ同じ長さである必要があります。

カテゴリーごとに 40 個くらいのトレーニング文書のコレクションにすることをお勧めします。ただし、選択したカテゴリーが比較的一般的なものであれば、さらに多くの文書が必要になります。カテゴリーは慎重に選び、また、分かりやすいものにする必要があります。表示されるカテゴリーやトレーニング文書が人間のインデクサーにとってあいまいであったり、不明瞭であったりすると、自動処理中に必ず問題が起きます。

トレーニング文書のアップロード

トレーニング文書を追加するには、個々のカテゴリーを選択すると、「**トレーニング文書リスト (Training Document List)**」ウィンドウが表示されます。このリストに文書を追加するには、「**文書の追加 (Add Document)**」ボタンをクリックします。

「**トレーニング文書の追加 (Add Training Documents)**」ウィンドウが表示されます。ファイルをアップロードした後でこのウィンドウを閉じる必要はありません。このウィンドウは、引き続き、別のカテゴリーや別のカタログへの文書のアップロードに使用できます。

トレーニング文書を追加するには、「**参照 (Browse)**」ボタンをクリックし、「**開く (Open)**」ウィンドウで、該当のファイルまたはディレクトリーを選択します。

同じディレクトリーから 1 つ以上のファイルを選択してアップロードすることも、また、ディレクトリー全体をアップロードすることも可能です。選択したディレクトリーが空である場合は、通知されます。

これにより、Information Structuring Tool 以外 (たとえば、ファイル・システムなど) で作成された既存の分類法をインポートし、それを使用して作業することができます。

ファイル・システム内の選択されたディレクトリー (たとえば、Development) にサブディレクトリー Design が含まれ、分類法ツリー内のカテゴリー Development にもサブディレクトリー Design が含まれている場合、サブディレクトリー内のファ

イルはサブカテゴリに追加されます。サブカテゴリが存在しない場合には作成され、ファイルは新たに作成されたサブカテゴリに追加されます。

選択されているすべてのファイルについて、文書の言語およびフォーマットを選択します。プレーン・テキスト・ファイルを除いて、必ず、フォーマットの「自動検出」を使用してください。

トレーニング文書のリストにファイルを追加するには、「**サブミット (Submit)**」ボタンをクリックします。

図 10 に、トレーニング文書の追加を示します。

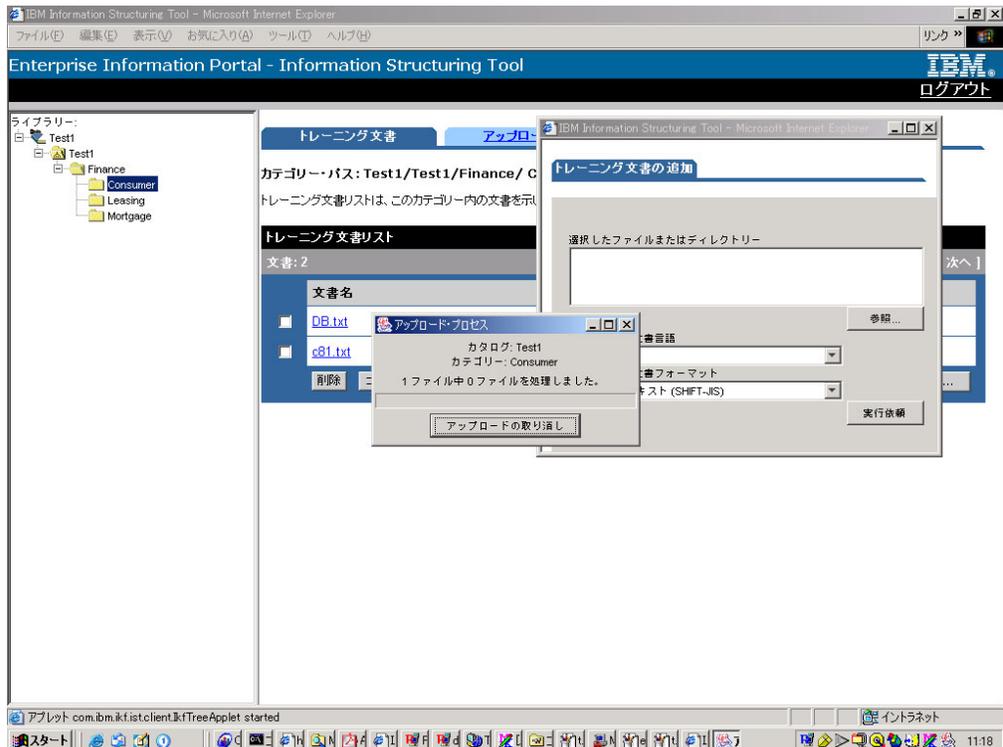


図 10. トレーニング文書の追加

「**アップロード・プロセス (Upload Process)**」ウィンドウが表示されます。このウィンドウから、文書のアップロード・プロセスをキャンセルすることができます。アップロード・プロセスがキャンセルされると、ファイルはそれ以上追加されなくなります。すでにトレーニング文書として追加されたファイルは、除去されません。ただし、アップロード中にサブカテゴリが追加された場合、それらのサブカテゴリはカタログから削除されません。

ファイルが正常にアップロードできない場合は、「**アップロード状況 (Upload Status)**」ウィンドウが自動的に表示され、ファイルがアップロードできなかった理由についての情報が表示されます。たとえば、次のとおりです。

- 同じ名前のファイルがすでに存在します (A file with the same name already exists)。
- ファイルが空です (The file is empty)。

- ファイルをサーバーにアップロードできませんでした (The file could not be uploaded to the server)。

「アップロード状況 (Upload Status)」ウィンドウで、「**トレーニング文書 (Training Documents)**」タブを押して、「**トレーニング文書 (Training Document)**」リストに戻ります。正常にカテゴリーにアップロードされたトレーニング文書は、すべてここに表示されます。リスト内のすべての文書を参照するには、「**戻る (Previous)**」および「**次へ (Next)**」ボタンを使用します。

複数のカテゴリーに同じ文書をトレーニング文書として追加する場合は、最初のカテゴリーにファイルをアップロードし、次に、そのファイルを他のカテゴリーにコピーします。このファイルを再度アップロードしないでください。コピーするには、「**トレーニング文書リスト (Training Document List)**」ウィンドウで 1 つまたは複数の文書を選択し、「**コピー (Copy)**」を押します。表示されたウィンドウで、「**参照 (Browse)**」をクリックして、文書のコピー先のカテゴリーを 1 つまたは複数選択し、「**サブミット (Submit)**」をクリックします。

ファイルのアップロード中、以下のアクションは許可されません。

- カタログのアンロック
- Information Structuring Tool からのログアウト
- カタログのトレーニングまたは評価の開始
- カタログの名前変更
- カタログのアップロード状況情報の削除

ただし、以下のアクションは許可されます。

- 別のファイル・アップロード・プロセスの開始 (同じカテゴリー内または異なるカテゴリー内へのいずれか)
- 別のカタログでの作業

76 ページの図 11 に、トレーニング文書のリストを示します。

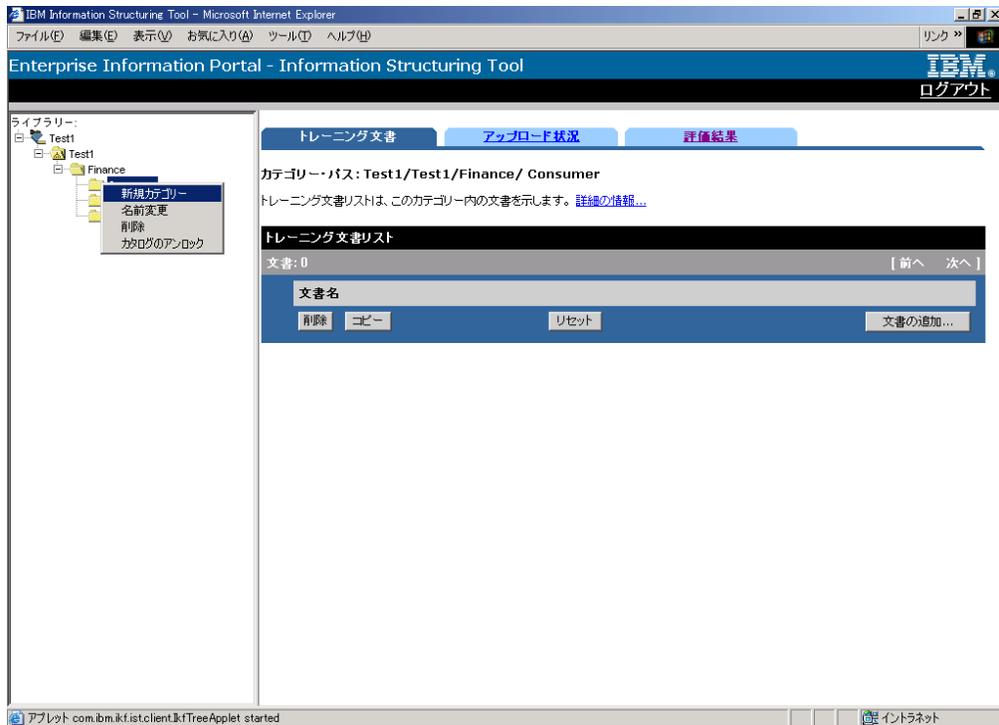


図 11. トレーニング文書のリスト

ファイルのアップロード中、ブラウザを閉じていても、ファイルがサーバーに転送済みであれば、それらのファイルはトレーニング文書として追加されます。転送済みでない場合は、ファイルは追加されません。

すべてのトレーニング文書をアップロードした後で、カタログを選択して評価を開始します。

カテゴリー化モデルの評価

分類法を定義し、トレーニング文書を各カテゴリーに割り当てた後、分類法を評価する必要があります。分類法の評価は、事前定義された分類法に関して、トレーニング文書がどの程度適切であるかを評価する際に役立ちます。これは、以下のステップからなる反復プロセスです。

1. 評価を開始する
2. 評価結果を評価する
3. 分類法またはトレーニング文書に変更を加える
4. 評価を戻す

各評価の反復中の評価プロセスは、以下のとおりです。

- トレーニング文書をトレーニング・セット (文書の約 80%) とテスト・セット (文書の約 20%) に分割します。このプロセスは、トレーニング・セットを使用してカタログをトレーニングし、テスト・セットには、カテゴリー化サービスを使用します。

- これらの文書が、正しいカテゴリーに割り当てられ、信頼性値が十分に高いかどうかを検査します。範囲は 0 から 1 までです。ここで、1 は、文書の適合度が最も高いことを意味します。信頼性値を設定することができます。デフォルト値は 0.5 です。

反復数は、3 回から 5 回までの中から選択できます。デフォルトの反復数は 3 回となっており、分類法の長所と短所を分かりやすく見通すことができます。5 回を選択すると、すべての文書がトレーニングおよびテスト・セットの両方に入られます。

「評価の開始 (Start Evaluation)」を押して、評価プロセスを開始します。

評価の反復中、それぞれのカテゴリーごとに以下が計算されます。

- 正しい 文書。評価中 c に割り当てられたカテゴリー c 内のトレーニング文書の数。
- アウトバウンド 文書。評価中、異なるカテゴリーに割り当てられたカテゴリー c 内のトレーニング文書の数。
- インバウンド 文書。評価中、カテゴリー c に割り当てられたが、元は別のカテゴリーにあったトレーニング文書の数。
- 割り当てられなかった 文書。評価中、どのカテゴリーにも割り当てられなかった、カテゴリー c 内のトレーニング文書の数。この中には、カテゴリーには割り当てられたが、指定された信頼性値よりも低くなった文書が含まれる可能性があります。

以下の項目について、カタログ・レベルでの評価プロセスの進行状況の概要が表示されます。

- 評価が引き続き実行されているか、停止されたかを示す評価状況
- 最終評価の日付がリストされている最終評価
- 完了した評価反復数
- 合計反復数
- 正しい文書 (すなわち、最初からカテゴリーに割り振られた文書) とインバウンド文書のパーセンテージを示す、精度全体の平均。
- カテゴリー内の正しい文書のパーセンテージを示す、再呼び出し全体の平均。
- 正しい文書、すなわち、正しく割り当てられた文書の数
- 誤って配置された文書、すなわち、インバウンド文書またはアウトバウンド文書の数
- 割り当てられなかった文書

精度および再呼び出しは、割り当てられた信頼性値と密接に関係しています。信頼性値が低いと、精度が低下して再呼び出しが増えます。逆もまた同様です。高い信頼性値は、多くのトレーニング文書が正しく割り当てられたことを意味します。これに対して、再呼び出し値が高いと、ほとんどのトレーニング文書がカテゴリーに割り当てられたことを意味します。言い換えれば、割り当てられなかった文書が全くないか、ほとんどないということを示します。

評価結果

「評価結果 (Evaluation Result)」タブを押すと、詳細な評価結果が得られます。カタログを選択してこのタブを押すと、カタログ全体の結果が表示されます。カテゴリを選択すると、カテゴリの結果が表示されます。

図 12 に、カタログ・レベルの評価結果を示します。

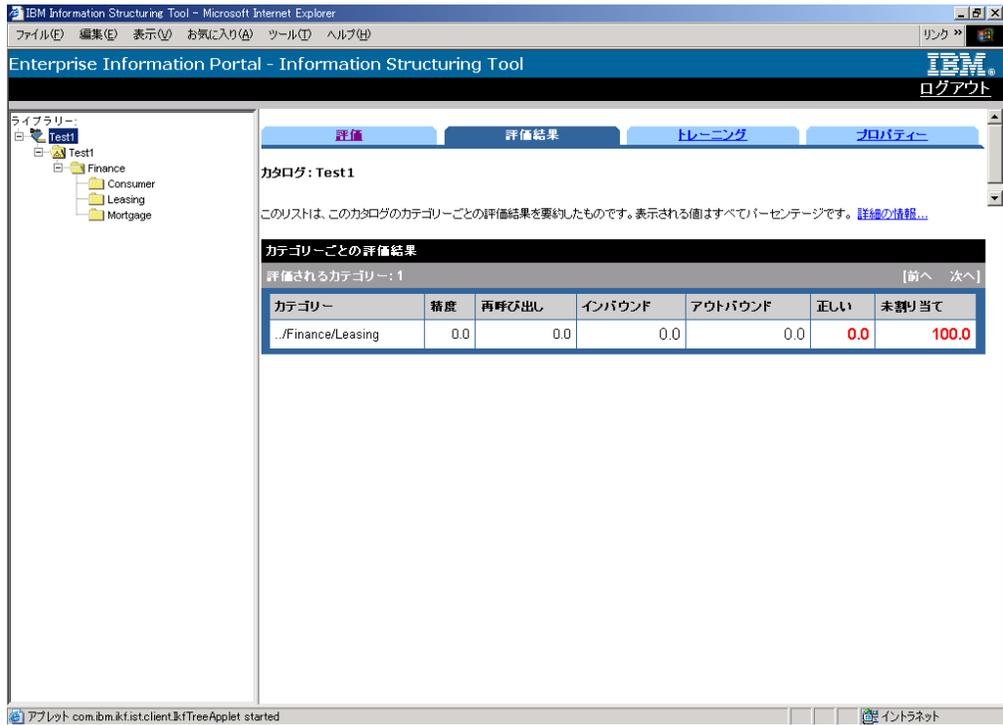


図 12. カタログ・レベルの評価結果

カタログ全体の結果から始めます。赤い値 (重大) と青い値 (やや重大) は、カテゴリまたはトレーニング文書のいずれかに何か誤りがあることを示します。カテゴリの品質は、トレーニング文書によって定義されます。したがって、カテゴリへの文書の移動 (インバウンド) と、カテゴリからの文書の移動 (アウトバウンド) を調べるのが重要です。

カテゴリまたはトレーニング文書に対する変更は、精度と再呼び出しの両方が等しく強調されているか、あるいは精度のみが強調されているかに大きく依存します。精度値が高いほど、分類法におけるそのカテゴリと他のカテゴリとの区別が明確になります。これに対して、再呼び出し値が高いほど、割り当てられなかったトレーニング文書は少なくなります。

評価結果は、以下の 2 つのレベルで表示されます。

1. カタログ・レベル:

カタログ内の各カテゴリについて:

- 精度と再呼び出しのパーセンテージ
- インバウンド文書のパーセンテージ
- アウトバウンド文書のパーセンテージ

- 正しい文書のパーセンテージ
 - 割り当てられなかった文書のパーセンテージ
2. **カテゴリ・レベル:**
- インバウンドおよびアウトバウンドの文書タイプについて:
- トレーニング文書、および発信カテゴリまたは宛先カテゴリ
- 正しい文書、および割り当てられなかった文書タイプの場合:
- トレーニング文書

評価結果の解釈

次のセクションでは、評価結果の解釈方法について説明しますが、分類法は、全体として機能することに常に留意してください。これは、分類法のあるセクションに変更を行うと、分類法の他の箇所に逆の結果が生じる可能性があることを意味します。

図 13 に、カテゴリ・レベルの評価結果を示します。



図 13. カテゴリ・レベルの評価結果

カタログ・レベルの評価結果から始める場合は、精度の値と再呼び出しの値が低い可能性があり、トレーニング文書の値に赤 (より重大) または青 (やや重大) のいずれかのマークが付いているカテゴリを選択します。各カテゴリに対して順番に (左側のフレームで選択可能)、以下の項目をチェックします。

- **インバウンド文書の数が多い:**
 - カテゴリが、他の多くのカテゴリから文書を取得する。
 - カテゴリの明確さが不十分である。これらの文書をこのカテゴリにコピーします。

- より適したトレーニング文書を選択するか、このカテゴリをサブカテゴリに分割することによって、このカテゴリをより明確にする。
- カテゴリは、以下のように 1 つまたは 2 つのカテゴリから文書を取得します。
 - これらの文書をこのカテゴリにコピーする。
 - 発信カテゴリをチェックする。このカテゴリを保持する意味がありますか？あるいは、カテゴリをマージする必要がありますか？マージする場合は、保持するカテゴリにすべてのトレーニング文書をコピーし、その他のカテゴリを削除します。
- **アウトバウンド文書の数が多い:**
 - カテゴリから他の多くのカテゴリへ文書が脱落した。
 - カテゴリの明確さが不十分である。さらに適したトレーニング文書を選択するか、カテゴリの削除を考慮する。
 - カテゴリから 1 つまたは 2 つのカテゴリへ文書が脱落した。
 - これらの文書を他のカテゴリにコピーする。
 - このカテゴリをチェックする。このカテゴリを保持する意味がありますか？あるいは、カテゴリをマージする必要がありますか？
- **割り当てられなかった文書の数が多い:**
 - 以下の項目について、割り当てられなかった文書と正しい文書とを比較する。
 - サイズ

割り当てられなかった文書の方が短い場合は、たとえば、2 つを結合して、正しい文書のサイズに一致するような、より長い新規の文書を作成することを考慮してください。
 - スタイル

スタイルが異なる場合は、文書を削除します。
 - トピック

トピックが若干異なっても、依然として正しい文書のトピックに関係している場合は、このトピックをカバーする追加のトレーニング文書を検索し、それらを同じカテゴリ、または新規に作成したカテゴリにアップロードすることを考慮します。

再呼び出しと精度の値が適正で、評価の反復を繰り返し行っている間に同じ文書が割り当てられないままである場合は、それらの文書を削除します。

分類法に対して変更を行った場合 (たとえば、カテゴリを削除したり、2 つをマージした場合、あるいは多くの文書をあちこちへ移動した場合) は、評価プロセスを再開するだけで済みます。評価プロセスを再開する前に、新規トレーニング文書の追加や文書のコピーなどの、カテゴリ・レベルのより小規模な変更を、いくつかのカテゴリに対して行うことができます。変更を取り消したい場合でも、元に戻す機能はないという点に注意してください。

ユーザーが関心のある赤または青の値がそれ以上ない場合、または精度および再呼び出しのレベルが 90% を超えた場合は、評価を停止します。

カタログのトレーニング

分類法の評価が終わり、その結果が満足のものであれば、特別なタイプのメタデータ、つまりカテゴリー化モデルを作成するために、すべてのトレーニング文書を使用して分類法をトレーニングする必要があります。このカテゴリー化モデルは、その後、情報マイニング・カテゴリー化サービスを使用して新規文書をカテゴリー化する際に使用することができます。

トレーニング・フェーズを開始するには、トレーニングに必要なカタログを選択して、「**トレーニング (Training)**」タブを押した後、「**トレーニングの開始 (Start Training)**」ボタンを押します。

右フレームの一番上に、選択したカタログの名前が表示されます。カタログのトレーニング状況も表示されます。トレーニング状況の種類は、以下のとおりです。

- 文書をカテゴリー化できない。カタログをトレーニングして、カテゴリー化機能を更新します。(これは、新規に作成されたカタログのデフォルトにもなります。この場合、「**最終トレーニング (Last training)**」の日付は空です。) カatalogが新しいためにまだトレーニングされていないか、あるいは、たとえば、カatalogが最後にトレーニングされてからカテゴリーが名前変更または削除されたなどの理由で、最新のトレーニング結果は無効となっています。カテゴリー化サービスを使用して、新規文書をカテゴリー化しようとすると、エラーになります。
- カatalog内のトレーニング文書のセットが変更されている。文書は、依然として、最終トレーニング結果に従ってカテゴリー化されています。カatalogをトレーニングして、カテゴリー化を更新します。
- トレーニングがカatalog上で実行されている。
- カatalogが最新で、トレーニングの必要がない。

トレーニング・プロセスを停止するには、「**トレーニングの停止 (Stop Training)**」ボタンをクリックします。トレーニングが停止されると、カatalog・トレーニング・プロセスを再実行しなければならないので注意してください。

新規トレーニング文書は、トレーニングされている間はカatalogへアップロードすることはできません。

パフォーマンスのチューニング

カatalog内にレコードを作成する場合、テキスト検索可能な属性 (たとえば、IKF_CONTENT など) 用のテキスト索引は、非常に大規模になり、検索パフォーマンスが著しく下がります。ストレージを最適化してパフォーマンスを向上させるには、テキスト索引を定期的に再編成する必要があります。特に、索引に対して大規模な更新を行った後は再編成が必要です。

テキスト索引の再編成を開始する時刻は、夜間などのコンピューターのオフピーク時が最適です。索引編成を実行するには、以下に切り替えます。

- Windows の場合: ...¥ikf¥IkfReorg.cmd
- AIX の場合: .../ikf/bin/IkfReorg.sh
- Solaris の場合: .../ikf/bin/IkfReorg

パラメーターには、IkfReorg <UserID><Password><DBName> が組み込まれます。

IBM Web Crawler の使用

このセクションでは、IBM Web Crawler フィーチャーの構成方法について説明します。このフィーチャーは、EIP インストール・プログラムで「フィーチャー (Features)」チェック・ボックスを選択するとインストールされます。

EIP バージョン 8.2 には、Web Crawler、Lotus Notes クローラー、クローリング済みファイルからデータを抽出する要約機能、HTML ベースの文書、構成サンプル、およびサポート・ユーティリティーが含まれています。IBM Web Crawler (GCS と呼ばれる) には、Java バージョン 1.3 以上が必要です。

IBM Web Crawler は、Java ベースのコンテンツ・クローラーおよびマイナーです。コンテンツがポイントされると、IBM Web Crawler はそのコンテンツを取得してマイニングします。

IBM Web Crawler は、イントラネット、エクストラネット、またはインターネットの Web 内のコンテンツのクローリング、自身または Domino を介しての Lotus Notes データベース内のコンテンツのクローリング、ローカル・ファイル・システム内にあるコンテンツのクローリングが可能です。IBM Web Crawler は、新規のプロトコルを容易に追加できるように作成されています。コンテンツは、HTML、Notes 添付ファイル、マルチメディアなど、あらゆるタイプが可能です。

IBM Web Crawler は、多くのタイプのコンテンツからメタデータおよびテキストをマイニングすることができます。たとえば、HTML コンテンツからは下記の項目をマイニングすることが可能です。

- URL
- タイトル
- 本文
- 最終更新時刻
- 作成者、キーワード、説明、などのメタ・タグ

ユーザーは、指定したタイプのコンテンツについて、一連の事前定義されたマイナーから選択します。コンテンツおよび / またはマイニングされたメタデータは、ローカル・ディスクに保管されます。IBM Web Crawler は、Network Solutions Outside In テクノロジーを使用して、200 タイプ以上のコンテンツからテキストを抽出することが可能であり、検索アプリケーションとの使用は、理想的な組み合わせです。また、IBM Web Crawler は、新規のマイナーを容易に追加できるようにも作成されています。

IBM Web Crawler は、Windows NT 4.0 および Windows 2000 オペレーティング・システムで使用可能です。IBM Web Crawler は、約 30 分でインストールおよび構成をして、使用可能状態にできます。500 MHz の PC では 1 秒間に約 10 個のファイルのコンテンツを取得およびマイニングします。オブジェクトが 100 万個 (200000 Notes) に達するまでテストされています。複数ユーザー構成、および 1 ユーザーにつき複数のクローリング構成およびマイニング構成をサポートしており、ユーザーが選択する各国語のサポートが可能です。

IBM Web Crawler の機能

インストール・プログラムは、以下の 2 つのファイルをインストールします。

`x:/<install directory>/run`

IBM Web Crawler Web 用バッチ・ファイルおよびサンプル構成。

`x:/<install directory>/notes-run`

Web Crawler Notes 用バッチ・ファイルおよびサンプル構成。

`x:/<install directory>/lib`

IBM Web Crawler .jar および .zip およびフィルター操作ファイル。

Web 用の IBM Web Crawler の構成および実行

このセクションでは、Web 用の IBM Web Crawler の構成および実行方法を説明します。Web 用の IBM Web Crawler は、HTTP サーバー、FTP サーバー、ニュース・サーバー、またはファイル・サーバーにアクセスして、HTML 文書およびその他のオブジェクトの要約を作成します。要約は、1 つの文書またはオブジェクトにつき 1 つのファイルになっており、メタデータおよび完全テキストが含まれていません。

基本構成

このセクションには、IBM Web Crawler 構成ファイルを XML フォーマットで編集する方法の説明が含まれています。作業の開始を支援するために、以下の 2 つのサンプル構成が提供されています。

- IBM Web Crawler を DB2 UDB とともに使用するための `config-db2.xml` ファイル。
- IBM Web Crawler を DB2 UDB なしで使用するための `config-sample.xml` ファイル。

1. コマンド・プロンプトを開く。
2. IBM Web Crawler をインストールした `run` サブディレクトリーへ、ディレクトリーを変更する。たとえば、IBM Web Crawler を Windows サーバー上にインストールした場合は、`cd x:<cmbrroot>%gcs%run` と入力します。IBM Web Crawler を AIX 上にインストールした場合は、`cd /usr/lpp/cmb/gcs` と入力します。

ヒント: (重要) 必ず元のファイルのコピーを保管しておいてください。ファイルのエラーが発生すると、IBM Web Crawler が壊れる場合があります。編集は、慎重に行ってください。

3. DB2 UDB データベースとともに IBM Web Crawler (よりスケーラブルだが、低速) を実行する場合は、`config-db2.xml` ファイルを編集する。たとえば、コマンド・プロンプトに `edit config-db2.xml` と入力します。
4. DB2 UDB データベースなしで IBM Web Crawler (拡張性は低いですが、速い) を実行する場合は、`config-sample.xml` ファイルを編集する。たとえば、コマンド・プロンプトに `config-sample.xml` と入力します。

データベースを使用せずに n 個の URL のクローリングを実行する場合は、クローリングした URL のメタデータを入れるために、マシン上におよそ $n/1000$ MB の RAM が必要です。たとえば、500,000 個の URL をクローリングするには、512 MB の RAM が必要となります。このメモリーを利用するには、`crawlweb.bat` ファイルを編集して、`JVMXmx` の値を大きくします。

Web Crawler 情報マイニング用の構成

Web Crawler を情報マイニングに使用する (つまり、クロール済み文書に情報マイニング機能を適用できるようにする) には、上述のサンプル構成ファイル (config-sample.xml および config-db2.xml) 内の設定値ではなく、以下の特定の構成設定値が必要です。

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

これらの設定値は、情報マイニング・コンポーネントのサンプル構成ファイル im-crawler-config-sample.xml の一部で、以下のディレクトリーにあります。

Windows の場合:

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

UNIX (AIX および Solaris) の場合:

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

IBM Web Crawler DB2 オプションの構成

DB2 オプションを構成する場合は、データベースを作成しなければなりません。データベースの作成には、DB2 管理者権限が必要です。場合によっては DB2 管理者アカウントに切り替える必要があります。データベースには、DB2 が許可する名前を付けることができますが、ユーザーのデータベース名が gcs でない場合は、Web Crawler の構成ファイルにある dbname を更新しなければなりません。

データベース管理者権限を持っている場合は、下記のコマンドを DB2 コマンド・プロンプトで実行して、データベースを作成することができます。

```
db -createdb <user><password>[database_name]
```

データベース名を指定しない場合は、gcs が使用されます。データベースが作成されたら、下記のコマンドを実行して IBM Web Crawler テーブルを追加します。

```
db -createtables<user><password>[database_name]
```

IBM Web Crawler で DB2 を使用する場合は、IBM Web Crawler データベースおよびテーブルの作成が完了していなければなりません。

新規のデータベース dbname を使用するには、以下の構成ファイルの設定 (urlpool-config セクションでの) が必要です。

- データベースの名前 (上記で作成したもの): 例: gcs
- ユーザー名: このユーザーの名前。例: db2admin
- パスワード: ユーザー・パスワード。例: db2admin

データベース、ユーザー名、パスワードの各プロパティを適切な値に設定します。キャッシュ・サイズまたはドライバーを変更してはいけません。ファイルの構成を続行し、ユーザーのシステムでの、クロールの有効範囲を設定します。

クロールの有効範囲の設定

これらの構成ファイルの設定は、DB2 の使用、不使用にかかわらず、クロールの有効範囲を設定する場合に必要です。

crawler-config セクションの以下の設定を確認し、これらの項目をユーザーの必要に応じた値に設定します。

seed list

1 つまたは複数の開始用絶対 URL。この URL は、使用可能でなければなりません。ブラウザを使用して検査してください。例:
http://www.<mysite>.com/

content-type-pattern-list

いずれかのファイル拡張子が、これらのパターンと一致する場合にのみ、ページ上で検出されるクロール URL (例: htm*)。

include-pattern-list

これらのパターンと一致する場合にのみ、ページ上で検出されるクロール URL (例: <mysite>.com)。

以下の項目も設定することができます。

recursion-depth

任意の開始点からクロールする、リンクの最大距離。深さを無制限にするには -1 を使用します。

exclude-pattern-list

これらのパターンと一致しない場合にのみ、ページ上で検出されるクロール URL (例: *cgi-bin*)。

system properties

Socks 化されていないマシンからファイアウォールを通過してクロールするには、このファイル内の socksProxy 値も設定する必要があります。

IBM Web Crawler の開始

.xml 構成ファイルを編集した場合は、それを保管します。

IBM Web Crawler の開始は、crawlweb バッチ・ファイルおよび構成ファイルを使用して行います。コマンド・プロンプトを開いて、以下を入力します。

- Windows の場合: crawlweb.bat<CONFIGFILE>
- AIX の場合: crawlweb.sh<CONFIGFILE>

DB2 UDB 付きで実行するには、`crawlweb config-db2.xml` と入力して「Enter」を押します。DB2 UDB なしで実行するには、`crawlweb config-sample.xml` と入力して「Enter」を押します。

ヒント: 定期的にクローलと要約の進行を報告するよう計画してください。ターゲットがクローलされると、`summaries-dir` に構成された場所に要約が書き込まれます。デフォルトの要約機能により、元のオブジェクトに加えてメタデータ・プロローグが、ツリー内に `.html` ファイルとして書き込まれます。クロール中またはクロール後に、追加情報をログ・ファイルで確認することができます。

拡張構成

ここでは、構成オプションについて学ぶことができます。構成例については、117 ページの『第 7 章 IBM Web Crawler サンプル・ファイル』にある `config-sample2.xml` ファイルを参照してください。サンプルでは、以下の構成について説明しています。

- クローラー・スレッドおよび要約機能スレッド
- グラフ用モニター
- ロギング・オプション
- SOCKS
- Lotus Domino のクロール
- 複数のコンテンツ・タイプ
- より多くの排他
- `.pdf` ファイルなどのオブジェクトの要約を取得するための、`InsoSummarizable` の使用。

構成ファイル内の使用可能なパラメーターの正式な定義については、`config.dtd` ファイルを参照してください。**勧告:** このファイルを編集してはなりません。このファイルのコピーを作成し、そのコピーの名前を変更してください。

IBM Web Crawler 構成ファイル

構成ファイルはどの Web ベース・リソースを集め、集めたりソースをどのように要約するかを IBM Web Crawler に通知する XML ファイルです。このセクションでは、`config.xml` で設定できる各エレメントおよび属性について説明します。Notes 用の IBM Web Crawler の使用方法に関する情報については、100 ページの『Notes 用の IBM Web Crawler』を参照してください。

IBM Web Crawler は、構成ファイルの内容が `gcs-config.dtd` に従っているかを検査します。クロールする URL がないなど重大なエラーが存在する場合は、IBM Web Crawler は終了し、エラー・メッセージをプリントします。小さな問題 (不明の属性または値) については、プログラムは警告をログ・ファイルに記録し、継続します。**勧告:** 構成ファイルは、編集する前にバックアップを取ってください。ファイルのエラーが発生すると、IBM Web Crawler が壊れる場合があります。

サンプル構成ファイルが IBM Web Crawler に同梱されています。

<gcs-config>

gcs-config ファイルには、**globals** と **group-list** という 2 つのセクションが含まれています。gcs-config ファイルの例については、117 ページの『第 7 章 IBM Web Crawler サンプル・ファイル』を参照してください。

globals

globals エlementは、ファイル・システム、パフォーマンス、およびネットワーク情報などの IBM Web Crawler の設定を取り込みます。

group-list

group-list エlementは、グループが業務またはネットワーク・ドメインなどの一連のリソースとなっているグループのクロールおよび要約を構成します。

<globals>

globals エlementは、IBM Web Crawler のグローバル設定を表します。この設定値は、global 属性および子Elementとしてエンコードされます。

以下のリストは、global 属性を定義します。global 子Elementの定義については、89 ページの『<logger-config>』を参照してください。

max-urls

クロールする URL の最大数。これは、正の整数でなければならず、デフォルトでは 100000 になっています。

summaries-dir

リソースの要約を書き込むディレクトリ。デフォルトでは、summaries/ ディレクトリが使用されています。

summaries-filepool-class

リソースの要約に使用されるファイル・プールのタイプ。これにより、要約ファイルにどのように名前を付けるか、およびどのサブディレクトリ構造(ある場合)を使用するかが決定されます。デフォルトでは、FullPathFilePool が使用されています。これは、ホスト用のディレクトリを作成し、その URL と同じサブディレクトリ構造およびファイル名を使用します。

num-crawlers

使用するクローラー・スレッドの数。これは、正の整数でなければならず、デフォルトでは 20 になっています。

num-summarizers

使用する要約機能スレッドの数。これは、正の整数でなければならず、デフォルトでは 5 になっています。以下のステップに従って、num-crawlers および num-summarizers を構成します。

1. クローラーを MHz/20 単位でユーザーのマシンの速度に設定する。たとえば、600 MHz のシステムでは、30 を使用します。
2. 要約機能の数をステップ 1 の数値の 1/4 に設定する (例、8)。
3. Windows 「タスク マネージャ」の「パフォーマンス (Performance)」パネルを監視しながら、実行を試行する。CPU が 1 秒間以上 100% になる

ことがある場合は、ステップ 1 に戻り、たとえば前回の値の 3/4 にするなど、より小さい数値を設定し、基本的に CPU が 100% になることがないようにします。

試行中に、text-monitor から要約機能に「処理待ち数が、構成されている要約機能の数をかなり下回っています (ToDo numbers well below the number of configured summarizers)」と絶えず報告される場合は、ステップ 3 に違反しない範囲で要約機能の数 (より少ない方がよい) を減らし、クローラーの数を増やすことができます。クローラーの数は多い方がよいでしょう。パフォーマンスを最大にするには、可能な範囲で最も速いネットワークを使用し、要約、データベース、一時スペース、およびログを、別のディスクへ分散します。

text-monitor

オンに設定されている場合は、text-monitor は 5 分ごとに IBM Web Crawler の状況を standard out にプリントします。10 進数値が指定されている場合は、text-monitor はテキスト出力の最新表示を行う間隔の時間を (秒単位で) 設定します。デフォルトでは、オフに設定されています。

graph-monitor

オンに設定されている場合は、graph-monitor は IBM Web Crawler にグラフ GUI を使用して状況を表示するよう通知します。10 進数値が指定されている場合は、graph-monitor はモニター GUI の最新表示を行う間隔の時間を (秒単位で) 設定します。デフォルトでは、オフに設定されています。

log-file

使用するメインのログ・ファイルを指定します。デフォルトは、log/log.txt です。

ヒント: logger-config エlementに追加の logger info を指定することができます。

log-priority

ログのデフォルトの優先順位を設定します。info、warn、または error の値を入力します。デフォルト値は warn です。

ヒント: logger-config エlementに追加の logger info を指定することができます。

temp-dir

一時ファイルを書き込むディレクトリー。**ヒント:** このディレクトリーのすべてのファイルは、IBM Web Crawler によって削除することができます。これは、デフォルト設定の x:/temp/gcs から変更する必要はありません。

temp-filepool-class

一時ファイルに使用するファイル・プールのタイプ。**警告:** これは、デフォルト設定の TempFilePool から変更してはなりません。

content-dir

IBM Web Crawler がコンテンツ・ファイルを書き込むディレクトリー。通常、content-dir は、temp-dir と同じです。

content-filepool-class

コンテンツ・ファイルに使用するファイル・プールのタイプ。通常、これは temp-filepool-class と同じです。

how-often-to-gc

ガーベッジ・コレクションの要求の間にクローलする URL の数。**勧告:** 整数 ≥ 50 を定義します。デフォルトでは 100 に設定されています。

max-resource-pool-size

要約されるのを待機しているリソースの最大キュー・サイズ。**勧告:** 整数 ≥ 10 を定義します。デフォルト設定では、1 つの要約機能に付き 10 個の待機リソースが許可されています。

connect-timeout

ネットワークでの接続がタイムアウトになるまで待機する時間をミリ秒で定義します。デフォルト値は 4000 です。有効範囲は 1000 ~ 60000 です。

read-timeout

ネットワークでの読み取りがタイムアウトになるまで待機する時間をミリ秒で定義します。デフォルト値は 6000 です。有効範囲は 1000 ~ 60000 です。

cookies

HTTP ヘッダーの cookies の検査と、データベースへの保管を行うかどうかを定義します。デフォルトでは、オフに設定されています。この値をオンに設定することにより、cookies を使用可能にすることができます。

locale 要約およびロギングに使用する言語を定義します。デフォルト値は en_US です。

global 子エレメントには、logger-config、urlpool-config、および system-properties が含まれます。

<logger-config>

logger-config ファイルは、ログに記録する内容、そのフォーマット設定の方法、およびログ・ファイルが書き込まれる場所についての拡張制御を提供します。default log-file および log-priority は、globals 属性として指定されます。ロギングについての詳細は、96 ページの『IBM Web Crawler でのロギング』を参照してください。

category

構成されているロガーのカテゴリ (例、gcs.crawler)。指定されていない場合は、デフォルトのロガーが構成されます。特定のカテゴリに対する設定は、すべての子カテゴリに影響することに留意してください。

priority

ログに記録しなければならないメッセージの最低の優先順位。指定されていない場合は、このロガーは、親カテゴリから優先順位を取得します (最終的には、globals のデフォルトの log-priority から取得します)。

log-file

ログ・ファイルを書き込む場所を定義します。「+」で始まる場合、そのログ・ファイルは他のいずれかの (親) ログ・ファイルと共に使用されます。指定されていない場合は、親ログ・ファイルが使用されます (最終的には、globals のデフォルトの log-file が使用されます)。

ヒント: 互いに上書きしてしまうため、複数のロガーに同じ log-file を指定しないよう注意してください。

log-layout

ログ・ファイルにプリントされるメッセージごとに使用されるレイアウトを定義します。

<urlpool-config>

urlpool-config ファイルは、URL が保管される IBM Web Crawler のコンポーネントを構成します。URL プールには、いくつかのオプションがあります。プールをメモリーに保管する、DB2 を使用する、または各 URL についての情報程度も保管しない専用の小さいメモリー・バージョンを使用することなどが可能です。urlpool-config エlementを指定しない場合は、URL プールはメモリー内に保管されます。urlpool-config には、データベース情報などを指定するための、子 urlpool-param Elementがある場合があります。

urlcontainer-class

使用する URL コンテナのタイプ。以下を指定します。

- DB2 UDB を使用してクロールする DB2URLContainer。
- DB2 UDB なしでクロールする MemoryURLContainer (デフォルト)。
- DB2 UDB なしでクロールし、追加のメモリーを使用する BigMemoryURLContainer (参照している URL およびその他の情報を保管する)。

urlcollection-class

使用する URL コレクションのタイプ。以下を指定します。

- DB2 UDB を使用してクロールする DB2URLCollection。
- DB2 UDB なしでクロールする MemoryURLCollection (デフォルト)。
- DB2 UDB なしでクロールし、追加のメモリーを使用する BigMemoryURLCollection (参照している URL およびその他の情報を保管する)。

<urlpool-param>

パラメーターを urlcollection-class に渡すために使用されます。一例として、117 ページの『第 7 章 IBM Web Crawler サンプル・ファイル』の、DB2 UDB を使用したサンプル構成のデータベース接続情報を参照してください。

name パラメーター名を定義します。

value パラメーター値を定義します。

ヒント: これらのパラメーターは、エラーの検査が行われなかったため、注意して使用してください。

<system-properties>

system-properties は、システム・プロパティ設定のリストを表します。

<property>

たとえば、拡張構成サンプルで SOCKS ゲートウェイを使用する場合の構成を参照してください。

name パラメーターの名前。

value パラメーターの値。

これに代わるものとして、PROXY ゲートウェイを介した外部サーバーへの IBM Web Crawler アクセスを、以下を使用して構成することができます。

```
<system-properties>
  <property name="proxySet" value="true"/>
  <property name="proxyHost" value="proxy.hostname"/>
  <property name="proxyPort" value="80"/>
</system-properties>
```

ヒント: これらのパラメーターは、エラーの検査が行われないため、注意して使用してください。

<group-list>

group-list は、1 つまたは複数の group エLEMENTのリストです。

<group>

group エLEMENTは、類似した方法でクロールおよび要約されるリソースの単一グループを表します。各グループには、1 つの固有の name 属性と、クロールする対象を通知する、少なくとも 1 つの crawler-config 子ELEMENTがあります。グループは、デフォルトの要約機能を使用しない場合、1 つの子 summarizer-config エLEMENTを持つことができます。**ヒント:** グループが重複している (同じ URL が 2 つ以上のグループにある) と、予期しない結果となることがあります。複数グループ URL は、検出された最初のグループにのみ関連付けられます。

name このグループに固有の名前 (必須)。

<crawler-config>

これらの規則を使用して、クロールの有効範囲を設定します。クローラーは、seed-list にある各 URL を取り出し、そのコンテンツからの URL の解析を行って、下記の条件を満たす URL を、クロールするもののリストに追加します。

- content-type-pattern-list の少なくとも 1 つの規則と一致する。
- および include-pattern-list の少なくとも 1 つの規則と一致する。
- および、exclude-pattern-list の規則とは一致しない。

crawler-config には、単一属性の **recursion-depth** も必要です。recursion-depth は、クローラーが移動可能な、各シードからのリンクの範囲の数を定義します。デフォルトは -1 で、これは、深さが無限であることを表します。

<seed-list>

これは、URL シードのリストです。認証情報が付いている場合もあります。

<seed>

seed は、クローラーを開始する URL シードを表し、URL 属性と、場合によっては認証情報が付いています。各シードは、絶対 URL (例、http://<your.server>.com/) でなければなりません。リダイレクトされているシード、使用不可能なシード、またはテキスト以外のページを指しているシードは避けてください。シードを含むよう編集したページを指すと効果的です。このようなページは、ブラウザでの更新、検討、およびテストが容易です。

URL クロールの開始点となるシード URL。

<authentication>

rfc2617 での定義に従って、基本認証 で保護されているシード URL に送信されるオプションの認証。

username

認証に使用されるユーザー名。

password

認証に使用されるパスワード。

たとえば、次のとおりです。

```
<seed url="http://your.server.com/"><authentication username="me"
password="mine"/></seed>
```

<content-type-pattern list>

これは、ファイル拡張子で識別されたとおりに、クローリングするコンテンツ・タイプを組み込むためのパターンのリストです。このリストのいずれかの `url-name-pattern` と一致する URL ファイル拡張子 (`.html`、`.gif`、`.doc` など) はすべて、このテストに合格します。拡張子を持たない URL は、デフォルトでテストに合格します。`content-type-pattern-list` が指定されていない、または空になっている場合は、ファイル拡張子のない URL のみが受け入れられます。

<include-pattern list>

これは、たとえば、サーバーまたはドメイン・ネームによってクローリングする URL を組み込むためのパターンのリストです。このリストの `url-obj-pattern`、`url-regex-pattern`、`url-name-pattern`、または `url-predicate-pattern` のいずれかと一致する URL はすべて、テストに合格します。`include-pattern-list` が指定されていない、または空になっている場合は、すべての URL が受け入れられます。

<exclude-pattern list>

これは、URL をクローリングする対象から除外するためのパターンのリストです。このリストの `url-obj-pattern`、`url-regex-pattern`、`url-name-pattern`、または `url-predicate-pattern` のいずれかと一致するすべての URL は、クローリングされません。`exclude-pattern-list` が指定されていない、または空になっている場合は、拒否される URL はありません。

<url-obj-pattern>

これは、URL の異なる部分 (プロトコル、ホストなど) をワイルドカードでマッチングするためのパターンです。このパターンは、`exclude-pattern-list` および `include-pattern-list` の両方で使用できます。それぞれの部分ごとのパターンは、先頭および / または末尾に、あらゆるものに一致するワイルドカード `*` を入れることができます。ただし、パターンの真中にワイルドカードを入れることはできません。マッチングは、大文字小文字を区別しません。省略された URL 部分のパターンはすべて、自動的にマッチングされます。

以下のリストには、Java および IBM Web Crawler が URL `http://www.ibm.com/products/index.html?query#ref` を分割する方法の例が記載されています。

- プロトコル: `http`

- ホスト: `www.ibm.com`
- ポート: `-1` (指定されていない)
- ファイル: `/products/index.html?query`
- パス: `/products/index.html`
- ディレクトリー: `/products/`
- ファイル名: `index.html`
- 拡張子: `.html`
- 照会: `query`
- 参照: `ref`

以下のリストは、`url-obj-pattern` の各エレメントについての詳細です。

プロトコル (protocol)

URL プロトコルが一致しなければならないワイルドカード・パターン (例、`http`)。

ホスト (host)

URL ホストが一致しなければならないワイルドカード・パターン (例、`*.ibm.com`)。

ポート (port)

URL ポートが一致しなければならないワイルドカード・パターン (例、`80`)。

ファイル (file)

URL ファイルが一致しなければならないワイルドカード・パターン (例、`*.htm*`)。URL のファイル部分は、ホストの後にまずスラッシュで始まり、`query` が含まれる場合がありますが、`ref` は含まれません。
`http://www.ibm.com/products/index.html?query#ref` の *file* 部分は、`/products/index.html?query` です。

パス (path)

URL パスが一致しなければならないワイルドカード・パターン (例、`*.html`)。URL のパス部分は、ホストの後にまずスラッシュで始まり、`query` または `ref` は含まれません。この例では、
`http://www.ibm.com/products/index.html?query#ref` の *path* 部分は、`/products/index.html` です。

ディレクトリー (dir)

URL 内のディレクトリーが一致しなければならないワイルドカード・パターン (例、`/products/`)。ディレクトリーは、パスの一部で、スラッシュで始まり最後のスラッシュで終わります。この例では、
`http://www.ibm.com/products/index.html?query#ref` の *dir* 部分は、`/products/` です。これには、`query` または `ref` は含まれません。最後のスラッシュが削除されている正しくない URL (例、`http://www.ibm.com/products`) は、*dir* に正確に一致しない点に注意してください。正しくない URL の例では、URL *dir* は `/` となります。

ファイル名 (filename)

URL 内のファイル名が一致しなければならないワイルドカード・パターン (例、`index.html`)。ファイル名は、パスの一部で、最後のスラッシュの後に

続きます。この例では、
`http://www.ibm.com/products/index.html?query#ref` の *filename* の部分
は、`index.html` です。これには、`query` または `ref` は含まれません。

拡張子 (extension)

URL ファイル拡張子が一致しなければならないワイルドカード・パターン
(例、`htm*`)。可能であれば `content-type-pattern-list` を使用することをお
勧めします。

照会 (query)

URL 照会が一致しなければならないワイルドカード・パターン。

参照 (ref)

URL 参照が一致しなければならないワイルドカード・パターン (HTTP で
は使用されていません)。たとえば、`<url-obj-pattern host="*.ibm.com"/>`
は、あらゆる IBM サイトの HTML ページと一致します。

<url-regex-pattern>

`url-regex-pattern` は、正規表現を使用している URL と一致するパターンです。
このパターンは、`exclude-pattern-list` または `include-pattern-list` のどちらで
も使用できます。これは、`com.ibm.regex` パッケージ (`regex4j`) を使用しており、
Perl 5 の正規表現の機能のほとんどを備えています。また、2 つの正規表現を持つ
ことができ、1 つは URL が必ず一致しなければならない、もう 1 つは URL が一致
してはなりません。大文字小文字を区別しない場合に指定する `i` など、その他のオ
プションを指定することができます。詳しくは、「`Regex4j Regular Expressions`」を
参照してください。

match URL が一致しなければならない Perl 5 スタイルの正規表現。

no-match

URL が一致してはならない Perl 5 スタイルの正規表現。

options

大文字小文字を区別しない場合に指定する `i` などのオプションの修飾子。

たとえば、`<url-regex-pattern`

`match="^http://www%.ibm%.com/.*%.html?$"/>` は、メインの IBM Web サ
イトの HTML ページと一致します。

<url-name-pattern>

これは、完全 URL または URL ファイル拡張子をマッチングするための、ワイル
ドカード付きの単純なパターンです。`content-type-pattern-list`、
`include-pattern-list`、および `exclude-pattern-list` で使用することができま
す。これは、パターン・ストリングの先頭および / または末尾に、あらゆるものに
一致する `*` を入れることができます。ただし、名前の真中にワイルドカードを入
れることはできません。マッチングは、大文字小文字を区別しません。

たとえば、`<url-name-pattern name="*.ibm.com/*"/>` は、IBM サイト上のすべて
のファイルと一致しますが、`<url-name-pattern name="*.ibm.com/*.html"/>` は、
真中にワイルドカードがあるため、無効です。

name オプションの `*` ワイルドカードが先頭および / または末尾にある、URL
ストリングが一致しなければならないワイルドカード・パターン。

<url-predicate-pattern>

このパターンは、URL をマッチングするための Java UnaryPredicate クラスをロードします。これは、exclude-pattern-list または include-pattern-list のどちらでも使用できます。クラスには、URL が述部に一致した場合に true を戻す public boolean execute(URL url) メソッドがなければなりません。

class 完全修飾の UnaryPredicate クラス名。

<summarizer-config>

これは、要約機能用の構成であり、子 resource-handlers のリストが付いています。現在は、1 つのグループにつき 1 つの summarizer-config しか存在できません。

<resource-handler>

リソース (Web ページまたはニュースグループ記事など) に対して作成する要約のタイプを、コンテンツ・タイプ (text/html など) か、またはファイル名の拡張子 (htm) を基に決定します。リソースを要約する準備が整うと、IBM Web Crawler は、順に resource-handler を検査し、コンテンツ・タイプまたはファイル拡張子に一致する最初のリソースを使用します。一致するリソースがない場合は、デフォルトの Copy2RdfSummarizable および Copy2RdfSummaryMaker が使用されます。これは、リストの下部に content-type または file-extension のない resource-handler を追加することにより、指定変更することができます。

resource-handler は、子の summarizer-param を持つこともできます。これは、特別なパラメーターをその SummaryMaker クラスに渡します。

content-type

リソースの content-type が一致しなければならないワイルドカード・パターン (例、*htm*)。

file-extension

リソースのファイル拡張子が一致しなければならないワイルドカード・パターン (例、htm*)。

summarizable

リソースの Summarizable クラスの名前 (例、HtmlRawSummarizable)。

summary-maker

リソースの SummaryMaker クラスの名前 (例、HtmlRawSummaryMaker)。

content-type パターンおよび file-extension パターンでは、ワイルドカードを使用できます。パターンは、パターン・ストリングの先頭および / または末尾に、あらゆるものに一致する * を入れることができます。ただし、真中にワイルドカードを入れることはできません。マッチングは、大文字小文字を区別しません。

resource-handler は、content-type パターンおよび file-extension パターンの両方が一致すれば一致し、指定されていないパターンは、常に一致します。下記は、その例です。

```
<resource-handler content-type="*htm*"
summarizable="*HtmlRawSummarizable" summary-maker=
"HtmlRawSummaryMaker"/>
```

この場合、ファイル拡張子に関係なく、コンテンツ・タイプが text/html のファイルはすべて、一致します。

summarizable および summary-maker については、クラスが com.ibm.IBM Web Crawler.summarizer.resource パッケージにある場合、クラスの絶対パスを指定する必要はありません。

<summarizer-param>

これらは、SummaryMaker クラスに渡される特別なパラメーターです。使用は、このクラスに特定されています。

name パラメーターの名前。

value パラメーターの値。

ヒント: これらのパラメーターは、エラーの検査がない場合があるため、注意して使用してください。

IBM Web Crawler でのロギング

これは、IBM Web Crawler のロギング・フィーチャーの概要です。

IBM Web Crawler は、ログに記録する対象、ログに記録する場所、およびそのフォーマット方法についての強力な制御を提供します。たとえば、クロールした各ページの応答コードを 1 つのファイルに、IBM Web Crawler の状況 (クロールした URL の数、作動中のスレッドの数など) を別のファイルに、要約された URL を 3 番目のファイルに、すべての IBM Web Crawler 警告を 4 番目のファイルに、そしてネット・ユーティリティー・パッケージ内のすべてのログ・メッセージを別のファイルにデバッグのために書き込むよう、選択することができます。

ログ分析ファイルのサンプルについては、119 ページの『IBM Web Crawler ログ分析ファイルの例』を参照してください。

ログの利用

ロギングは、network/web/crawl/summarization アカウンティング、他のアプリケーション・コンポーネントとの通信、および IBM Web Crawler のデバッグに役立ちます。

クロールおよびマイニングのアカウンティングは、たとえば、構成に誤りがあるサーバー、ページの欠落、および 1 つの content-type あたりのオブジェクトの数など、広範囲に渡る興味深い特徴を明らかにすることができます。loganalysis.pl Perl スクリプトでは、ログの要約アカウンティングのサンプルを提供しています。アプリケーションは、コンテンツが除去されたタイミングなどの、IBM Web Crawler からの情報が必要な場合があります。

ロガーの構成

IBM Web Crawler の構成ファイル内の 1 つまたは複数のロガーの構成を指定することができます。globals エレメントの log-priority 属性および log-file 属性は、デフォルトのロギング・ポリシーを確立します。

ロギング・ポリシーを拡張するには、logger-config ステートメントを globals エレメントの子として作成します。各ステートメントは、IBM Web Crawler のログ・メッセージのサブセットを選択し、それらを特定のファイルに経路指定して、特定のフォーマットを使用して書き込みます。ログに記録されたメッセージのサブセッ

トは、priority 属性および category 属性を使用して選択されます。規則上の優先順位の値は、trace、debug、info、および warn (大文字小文字は区別されない) です。

- 優先順位の値を設定すると、最も冗長なトレースを基に、ロガーの冗長度が決定されます。
- trace および debug は、保守レベルです。メッセージは、英語でハードコーディングされています。
- info および warn は、ユーザー・レベルであり、各国語がサポートされています。
- info は、多くのメッセージを作成します。warn の優先順位を指定して、メッセージ出力を削減してください。

構成例のロギング

ファイル `log/fromto.txt` への、ハイパーリンクからの、またはハイパーリンクへのログ (date/time/thread の情報なし)

```
<logger-config category="gcs.url.fromto" priority="info"
  log-layout="%m%n" log-file="log/fromto.txt"/>
```

ファイル `log/resources.txt` への、要約済みオブジェクトのログ

```
logger-config category="gcs.summaries.list.resource"
  priority="info" log-file="log/resources.txt"/
```

スキップされた URL およびスキップの理由のログ

```
<logger-config category="gcs.url.skipped"
  priority="info" log-file="log/urls_skipped.txt"/>
```

特別に処理された HTTP 応答コードのログ

```
<logger-config category="gcs.http.302"
  priority="info" log-file="log/urls_redirected.txt"/>
<logger-config category="gcs.http.404"
  priority="info" log-file="log/urls_not_found.txt"/>
```

要約機能カテゴリー内のすべてのメッセージのログ (優先順位を含む)

```
<logger-config category="gcs.summarizer"
  priority="TRACE" log-file="summarizer_trace.txt"
  log-layout="%d: %t: %c: %p: %m%n"/>
```

トラブルシューティング

問題が発生した場合には、以下の項目を最初に検査してください。

ユーザーのシード・リスト上のページは、到達可能ですか?

このページは、存在していなければならず (シードのリダイレクトを回避する)、システムから、および SOCKS が使用されている場合は SOCKS を通して、到達可能でなければなりません。

ユーザーのシード・リスト上のページは正規の HTML ですか?

フレーム、Flash、JavaScript、およびその他のこの種のアイテムは、シードとして選択するアイテムとしては適していません。正規の HTML ページを選択してください。

DB2 UDB を使用している場合、すでにクロールしましたか?

DB2 UDB は、ユーザーがクロールした対象を追跡します。すべてのページ

が処理された場合は、DB2 UDB はメッセージを出しません。db
-emptytables コマンドを使用して、新規のクロールを開始してください。

DB2 UDB を使用している場合、構成ファイルのデータベース・アクセス項目は正しいですか？

データベース接続が失敗すると、クロールも失敗します。

ユーザーの構成ファイルの編集を慎重に検査しましたか？

エラーが発生すると、IBM Web Crawler が壊れる場合があります。制限された max-urls または recursion-depth の値を確認しましたか？

まだ問題がありますか？

構成ファイルを編集して、ログの優先順位を「debug」に変更してください。Web Crawler を再度開始し、停止後にログ・ファイルを確認してください。

要約機能の選択

要約機能の目的は、リソース (Web ページなど) またはホスト (Web サーバーなど) を対象に、ユーザーに関心のある情報が、使用しやすいフォーマットで含まれているファイルを作成することです。

IBM Web Crawler には、異なるコンテンツ・タイプの処理、リソースからの異なる種類のデータの抽出、および各種のファイル・フォーマットでの出力を行うさまざまな要約機能が組み込まれています。このセクションでは、使用可能な要約機能のフィーチャーおよび要件を説明します。必要とする機能がこれらの中にはない場合は、独自の要約機能を作成することもできます。

IBM Web Crawler には、2 種類の要約機能があります。リソース要約機能は、Web ページなどの単一リソースの要約を作成し、ホスト要約機能は、Web サーバーなどのホストの要約を作成します。現在は、リソース要約機能のみが構成可能です。

リソース要約機能の選択

リソース要約機能を選択または作成する場合に考慮する事項は、次のとおりです。

- 入力フォーマットは何ですか (HTML Web ページ、PDF、WordPro 文書、XML ファイル)?
- どのメタデータ (HTTP ヘッダー、タイトル、注釈付きリンク、本文テキスト) を抽出しますか?
- どの出力フォーマット (XML、HTML、RDF) を希望しますか?

IBM Web Crawler config ファイルで、resource-handler エレメントを使用している特定のタイプのリソースに対して、どの要約機能を使用するかを指定します。まず、その要約機能を使用するコンテンツ・タイプおよび / またはファイル名の拡張子を指定します。次に、その作業を実行する Java Summarizable クラスおよび SummaryMaker クラスを指定します。summarizable クラスは、要約されるリソースを表し、summary-maker クラスは作成される要約のタイプを表します。

デフォルトの要約機能 (コピー + RDF 要約機能)

コピー + RDF の要約機能は、別の要約機能によって処理するよう明示的に構成されていない content-type を持つすべてのオブジェクトに適用される要約機能です。この要約機能は、どのようなタイプのリソースでも使用でき、また、2 つのファイ

ルを作成します。最初のファイルは元のリソースの正確なコピーで、2 番目のファイルは元の URL、保管されたファイルのファイル名、および HTTP ヘッダー情報を含む RDF の要約です。これは、DefaultSummarizable および Copy2RdfSummaryMaker を使用して明示的に構成することもできます。

HTML ページ用の要約機能 (未加工の HTML 要約機能)

HTML リソースの場合、未加工の HTML 要約機能は、単に、先頭に URL および HTTP ヘッダー情報がコメント形式で入れられた、ファイルのコピーを作成します。これは、HtmlSummarizable および HtmlRawSummaryMaker を使用して、構成されます。

```
<resource-handler content-type="*htm*"
    summarizable="HtmlSummarizable"
    summary-maker="HtmlRawSummaryMaker" />
```

HTML ページ用の要約機能 (EIP HTML 要約機能)

HTML リソースの場合、未加工の HTML 要約機能は、単に、先頭に URL および HTTP ヘッダー情報がコメント形式で入れられた、ファイルのコピーを作成します。これは、EIPHtmlSummarizable および EIPHtmlRawSummaryMaker を使用して、構成されます。

```
<resource-handler content-type="*htm*"
    summarizable="EIPHtmlSummarizable"
    summary-maker="EIPHtmlRawSummaryMaker" />
```

HTML ページ用の要約機能 (ファイルの要約を作成しない HTML 要約機能)

この要約機能は、HTML をクローリングしてリンクをたどりますが、ファイルの要約は作成しません。これは、たとえば、(INSO から XML への要約機能を使用して) サイト上のすべての PDF ファイルをクローリングしたいが、HTML ファイルは保管したくない、というような場合に役立ちます。これは、HtmlSummarizable および NoWriteSummaryMaker を使用して、構成されます。

```
<resource-handler content-type="*htm*"
    summarizable="InsoSummarizable"
    summary-maker="InsoSummaryMaker" />
```

その他のコンテンツ・タイプ用の要約機能 (INSO から XML への要約機能)

この要約機能は、Microsoft Word 文書、PDF ファイル、PowerPoint プレゼンテーション、およびその他の 200 タイプを超えるリソースの XML 要約を作成します。この要約機能は、一部のメタ情報を持っており、本文テキストは Network Solutions INSO フィルター (INSO ライセンスが必要) によって抽出されます。これは、InsoSummarizable および InsoSummaryMaker を使用して、構成されます。

```
<resource-handler content-type="pdf"
    summarizable="InsoSummarizable"
    summary-maker="InsoSummaryMaker" />
```

その他の要約機能

その他のリソース・タイプを要約する必要がある場合は、その他のデータを取り出すか、またはその他のフォーマットで出力して、IBM に問い合わせるか、またはカスタムの要約機能を作成してください。

Notes 用の IBM Web Crawler

このセクションでは、Notes 用の IBM Web Crawler の構成および実行方法を説明します。Notes 用の IBM Web Crawler は、Notes データベースにアクセスし、Notes 文書および添付ファイルの要約を作成します。要約は、1 つの文書または添付ファイルにつき 1 つの XML フォーマットのファイルになっており、オブジェクトおよび完全テキストが含まれています。

前提条件

Notes 用の IBM Web Crawler を実行する前に、以下の前提条件が必要です。

- Lotus Notes バージョン 5.0.5 以降。
- PKZIP バージョン 2.50 (自己解凍形式の ZIP ファイルになっている添付ファイルを扱う場合)。

テスト・クロールの実行

「スタート」→「プログラム」→「コマンド プロンプト」の順に選択します。新規のウィンドウで Notes 用の IBM Web Crawler をインストールしたディレクトリーに移動し、notes-run サブディレクトリーへ移動します。たとえば、次のとおりです。

```
cd c:;%<install directory>%gcs%notes-run
```

Notes のクロールは、以下の 2 つのファイルにより制御されます。

- ユーザーがクロールできる Notes データベースを識別するために編集したソース・リスト。これには、Notes サーバー名、IP アドレス、.nsf ファイル名などが含まれています。たとえば、1 つのソース・リストが 34 の Notes データベースを指定するとします。これによって、どのデータベースをクロールするかが、以下に記載されている構成に設定されます。ソース・リストは、.xml ファイルまたは Notes データベース (.nsf ファイル) のいずれかとすることができます。
- ソース・リスト、どのソースをクロールするか、どの添付ファイル・タイプを処理するか、および出力するフォーマットなどを指定する構成ファイル。構成ファイルは、常に .xml ファイルです。

IBM Web Crawler が正常にインストールされていることを確認するには、IBM Web Crawler のテスト・データベースをクロールします。エディターを使用して、testSources.xml ソース・リストが、test.nsf データベースへの正しいパスおよびファイル名を持っているかどうかを確認します。このデータベースは、IBM Web Crawler をインストールした場所の notes-run サブディレクトリーにあります。元のファイルのバックアップ・コピーを作成してください。**勧告:** 注意して編集してください: このファイルのエラーは、IBM Web Crawler が故障する原因となります。すべての変更を保管してください。

組み込まれている test.nsf データベースをクロールして、インストールをテストします。次のように入力してください。crawlNotes crawlTestXml

crawlNotes.bat ファイルが、構成ファイルに crawlTestXml を使用して IBM Web Crawler を開始します。構成ファイル名に .xml が自動的に追加されます。IBM Web Crawler は、クロールを報告し、それぞれに添付ファイルのある 2 つの文書を要約する必要があります。

IBM Web Crawler が終了すると、構成ファイルで指定した、summary ディレクトリーにある要約、および log summary ディレクトリーにあるクローल・ログ・ファイルを表示することができます。

カスタム Notes クロールの構成

テスト・クロールが正常であれば、他のデータベースをクロールできます。

1. データベース・ソース・リストを作成する。クロールする Notes データベースをソース・ファイルに追加します。

XML ファイルでクロールする Notes データベースを識別するには、testSources.xml ファイルを編集することから始めます。Notes データベースでクロールする Notes データベースを識別するには、Notes を使用して testSources.nsf Notes データベースを開き、更新します。ソース・ファイル内で設定できるパラメーターは、『Editing Source Lists』に説明されています。
2. クローラー構成を設定する。構成ファイルを XML 形式で編集する必要があります。
 - ソースが XML ファイル内にリストされている場合は、crawlTestXml.xml を編集することから始めて、sourcesInXmlFile を、ソース・ファイルを指すよう設定する。
 - ソースが Notes データベース内にある場合は、crawlTestNsf.xml を編集することから始めて、sourcesInNotesDB を、ソース・データベースを指すよう設定する。構成ファイル内で設定できるパラメーターは、86 ページの『IBM Web Crawler 構成ファイル』に説明されています。

ソース・リストおよび構成が完了した後、以下のように IBM Web Crawler を起動します。crawlNotes your_config

あるいは、ソース・リストが Notes データベース内にある場合は、IBM Web Crawler を以下の作業例のように開始します。crawlNotes crawlTestNsf

IBM Web Crawler Notes クローラーが終了すると、構成ファイルで指定した、summary ディレクトリーにある要約、および log summary ディレクトリーにあるクロール・ログ・ファイルを表示することができます。

ソース・リスト・パラメーター

ソース・リストには、クロールできる Notes データベースの説明が含まれていません。file.xml 形式のソース・リストには、1 つまたは複数の oneDBInfo エレメントと共に、1 つの notesDataSources エレメントが含まれています。各 oneDBInfo エレメントには、以下が含まれています。

id このデータベース用の数字の ID。これは、構成ファイルの範囲パラメーターによって参照されます。

serverName

データベースをサービスするサーバーの名前。ローカル・データベースにはヌル・ストリング "" を使用します。

pathAndFileName

サーバー上のデータベースの絶対パスおよびファイル名。パスおよびファイル名の末尾は、.nsf とします。

viewName

クローリングするデータベースの Notes ビューの名前。

ipAddress

オプション。サーバーの IP アドレス (指定されている場合)。DNS は使用されません。DNS がターゲット・サーバー名を解決できない場合は、その IP アドレスをここに指定することができます。Windows では、`nslookup server_name` コマンドを使用して、IP アドレスを判別することができます。

dateLastCrawled

オプション。データベースが最後にクローリングされた日付。構成ファイル内の `update Date Last Crawled` を `no` に設定しない限り、これは自動的に変更されます。

tries オプション。クローリングが失敗 (タイムアウト) した場合には、そのデータベースの再クローリングを試行する回数。

fieldSubstitutions

Notes データベースのフィールド名を出力 XML 文書でどのように置き換えるかを指定するマッピング。これには、以下のように、それぞれ 2 つの属性を持つ、1 つまたは複数の置換エレメントが含まれています。

- オリジナル: 存在する場合は、出力 XML 文書に置き換えられるフィールド名。
- 置き換え: 出力 XML 文書でオリジナルのフィールド名と置き換えられる新規のフィールド名。

Notes データベース内のソース・リストは、Notes Client を使用して検査および更新することができます。Notes を開始して、「**ファイル (File)**」→「**データベース (Database)**」→「**開く (Open)**」の順に選択します。「**参照 (Browse)**」をクリックし、`x:\%<install directory>\gcs\notes-run` ディレクトリーにある `testSources.nsf` データベースを見つけて開きます。同梱のテスト・データベースもまた、Notes Client を使用して検査および更新することができます。Notes を開始して、「**ファイル (File)**」→「**データベース (Database)**」→「**開く (Open)**」の順に選択します。「**参照 (Browse)**」をクリックし、`x:\%<install directory>\gcs\notes-run` ディレクトリーにある `test.nsf` データベースを見つけて開きます。

構成ファイル・パラメーター

構成ファイルに設定できるパラメーターについて、以下で説明します。リストされているデフォルトのパラメーターは、省略することができます。

`sourcesInXml` ファイルまたは `sourcesInNotesDB` エレメントは、ソース・リストを識別します。ソースは、それぞれ XML 形式のファイルまたは Notes データベースで、クローリングするデータベースを識別する情報が含まれています。

`runInfo` エレメントには単一の `run` を制御するパラメーターが含まれています。つまり、以下のパラメーターは、クローラーを指定通りに使用してクローリングされるすべての Notes データベースに適用されます。

rangeSpecify

クローリングするデータベースの ID。この ID は、ソース・リストの ID フィ

ールドに指定した数字です。これは、個々の ID のコンマ区切りのリストとして指定するか、または 1-4、15、25-31 のようにハイフンで指定した範囲として指定するか、あるいはその両方として指定することができます。

SummaryDirectory

出力要約のルート・ディレクトリーを指定します。要約は、このディレクトリーのサブディレクトリーに書き込まれます。

MaxThreads

並列クロール・スレッドの数を指定します。各 Notes データベースは、単一のスレッドによってクロールされます。データベースが複数の場合は、並行してクロールされます。

doIncrementalCrawl

デフォルトは、no です。yes の場合は、新規の Notes 文書、または 'summarizeThisDateAndLater' 以降に変更された Notes 文書のみを処理します。'summarizeThisDateAndLater' が指定されていない場合は、クローラーはデータベース・ソース・リストに指定されている各データベース自身の DateLastCrawled フィールドを使用します。doIncrementalCrawl=no の場合は、日付に関係なくすべての文書が処理されます。

summarizeThisDateAndLater

このフィールドのフォーマットは、MM/dd/yyyy hh:mm a tz (たとえば、01/01/2000 01:11 PM PDT) です。日付および時刻が指定されていない場合は、ソース・リストに記録されている最後のクロール以降 (doIncrementalCrawl が yes に設定されている場合)、または、最初から (doIncrementalCrawl が no に設定されている場合) のすべての文書を要約します。

detachAttachments

デフォルトは、yes です。yes の場合は、添付ファイルを切り離して要約します。処理する添付ファイルのタイプは、構成ファイルにリストされています。no の場合、添付ファイルは無視されます。

attachmentFilenameFormat

デフォルトは l (long) です。s (short) とすることも可能です。長いファイル名は、タイプ、サーバー、データベース名、および Notes ID をエンコードします。短いファイル名は、タイプおよび Notes ID をエンコードします。

processAttachmentsAfterwards

デフォルトは、no です。yes の場合、Notes データベースの要約中に添付ファイルは要約されません。代わりに、それぞれの添付ファイルごとに、記録が notesCrawl-attachments.bat に書き込まれ、ファイルを要約するためのコマンドが指定されます。要約後にバッチ・ファイルの作成および実行を行ってから、添付ファイルを削除します。後で添付ファイル进行处理する場合には、通常、十分なディスク・ストレージが必要です。

saveAttachmentFiles

デフォルトは、no です。yes の場合、元の添付ファイルは処理後に消去されません。このオプションは、processAttachmentsAfterwards が no の場

合にのみ有効です。processAttachmentsAfterwards が yes に設定されている場合は、保管されている添付ファイルを保管するためのディスク・スペースが必要です。

MaximumNumberOfDetachingErrors

デフォルトは 10 です。クローラーがクローリングを打ち切るまえに許容する、添付ファイルの処理におけるエラーの最大数 (たとえば、添付ファイルの保管時にディスク・スペースが不足するなど)。

saveURLsToFile

デフォルトは no です。指定した場合、Notes 文書項目で検出された URL は、次のような形式の名前のファイルに書き込まれます。

```
databaseName(without path and .nsf) + ".html"
```

updateDateLastCrawled

デフォルトは、yes です。no の場合は、ソース・ファイル内の dateLastCrawled を更新しません。

tempDirectory

デフォルトは、c:\temp です。このディレクトリーは、すべての一時ファイルの書き込みに使用されます。

logSummaryDirectory

デフォルトは、log です。ログ・ファイルが保管されるディレクトリーです。

loggerPriority

デフォルトは、info です。これらの設定は、ロガーの優先順位を定義します。設定値は、最高から最低まで、error、warn、または info が可能です。たとえば、ロガーの優先順位が warn に設定されている場合は、優先順位が warn と error のログ・メッセージのみ、ログに記録されます。

ここに含まれている attachments エレメントには、処理する添付ファイルの拡張子 (例、.prz) を識別するエレメントが組み込まれています。

サーバーからの IBM Web Crawler の除外

セキュリティおよびパフォーマンス上の理由で、EIP 管理者は、特定のサーバーまたはページをクローリングの対象から除外したい場合があります。サーバーやページのクローリング・アクティビティーを制限できるようにする必要が生じることもあります。

サーバーの管理者は、アクセス・ポリシー・ファイルを使用して、サーバーやページを回避するよう IBM Web Crawler に指示できます。このファイルは、*A Standard for Robot Exclusion*

(<http://info.webcrawler.com/mak/projects/robots/norobots.html> を参照) に公表されている指針に従って構築されています。

- IBM Web Crawlerにより、サーバーのクローリング前にアクセス・ポリシー・ファイル `http://yourserver/robots.txt` が要求され、その後も定期的に要求される。
- このファイルは次の形式の行で構成される。

```
field:<optionalspace>value<optionalspace>
```

field が User-Agent で、*value* が IBM-WebCrawler または * の場合、その後に Disallow 行 (その次の User-Agent の行の前まで有効) を使用すると、回避するアドレスの一部を指定できます。絶対パスも部分パスも指定できます。この値で始まるアドレスはすべて取り出されません。

たとえば、次のとおりです。

```
Disallow: /help
```

これは /help.html と /help/index.html の両方を回避します。

```
Disallow: /help/
```

これは /help/index.html を回避しますが /help.html は取り出します。

空値を指定すると、アドレスはすべて取り出されます。

- 上記の行は、ブランク行で区切ることができます。
- # 文字を入力して、コメントを含めることができます。その行のそれ以降の部分はコメントと見なされます。

次に数例を示します。

- 次の /robots.txt は、すべてのロボットがこのサーバーを回避するよう指定します。

```
# disallow everybody
```

```
User-agent: *  
Disallow: /
```

- 次の /robots.txt は、IBM Web Crawler のみがこのサーバーをクロールでき、Web Crawler には制限がないことを指定します。

```
# allow only IBM
```

```
User-agent: *  
Disallow: /  
User-agent: IBM-WebCrawler  
Disallow: # disallow nothing
```

- 次の /robots.txt は、すべてのロボットが htmldocs の temp、development、および testing ツリー中のアドレスを回避し、IBM-WebCrawler というユーザー・エージェントが development ツリーと testing ツリーを回避し (ここでは許可されている)、xyz ロボットと wxyz ロボットが全体を回避するよう指定します。

```
# 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-WebCrawler  
Disallow: /htmldocs/temp
```

第 6 章 ワークフローの紹介

EIP ワークフローを使用すると、業務における作業の流れとパフォーマンスを制御できます。ユーザーは、統合検索の結果を処理する際に、どのアクションを実行するかを決めなければならないことがよくあります。EIP ワークフローを使用すると、ユーザーが作業を実行する方法を前もって決めておくことができます。

ワークフローは、各ワークフロー・コンポーネントが協働する方法を制御するプロファイルおよび規則を設定することによって、自動化することができます。また、特権セットとアクセス制御リストを介してユーザー・アクセスおよび権限を制御することによって、システムにどの程度の制限を加えるかを選択することができます。

ワークフローの理解

業務の操作の大半は、相互に関連付けられた一連のプロセスを特徴としています。作業はある従業員から別の従業員へ、また部門から部門へと流れていきます。単純なプロセスであればいくつかのステップで済みますが、プロセスが複雑になっていくにつれて、各種の部門にわたって何人もの従業員がかかわることになります。

ワークフローを使用することにより、ユーザーは、業務がプロセスを通るようにし、プロセスの全体を通して業務に関する決定を行うことができます。たとえば、XYZ 保険会社では大量の保険金支払請求の用紙を郵便で受け取ります。検査プロセスの際に、保険会社の査定人は写真や査定記録、および専門家による報告書などの資料を収集する必要があります。従業員は、毎日何時間もかけて、情報を開き、ソートし、ファイリングし、管理して、さらには最終的に承認するための関連文書を集めます。

この情報は、情報が受け取られ検査されるたびに従業員の間で移動します。保険金支払請求の処理が完了するまでには、複数の部門の従業員がこの処理にかかわることになります。

ワークフローの使用法

XYZ 保険会社の例と同様に、文書を処理する大半の企業では、以下のタスクの一部またはすべてを実行します。

- 後で取り出せるように、文書をファイリングする
- あらゆるソースから文書、書式、報告書、および情報を収集し、その後、これらの文書を処理するための場所へ送達する
- 現在処理中の文書と届いた郵便を付き合わせる

作業プロセスとは、作業の流れを表します。この中では 1 つ以上の文書またはコンテンツのグループで実行できるアクションと、このグループがワークフロー・プロセスで経由するパスを記述しています。ワークフローは、有効範囲および境界を明確に定義して、作業の実行方法を反映します。ここでは、アクティビティーおよびタスクのシーケンス、さらにこれらのアクティビティーやタスクの接続および関

係を定義します。また、ワークフローは作業の流れで決定を下すために使用する基準を判別します。ワークフロー作成プロセスの詳細については、「ワークステーション・アプリケーション・プログラミング・ガイド」を参照してください。ワークフローでのクライアントの使い方については、「eclient のインストール、構成と管理」を参照してください。

ユーザー ID とグループの同期化

このセクションでは、EIP、Content Manager および MQ Series のワークフロー間のユーザー ID とグループを同期させる方法について説明します。

ご使用のシステムの一部が MQSeries Workflow である場合、Content Manager または Enterprise Information Portal 内のユーザーを管理するときに、MQSeries Workflow 内のユーザーも管理することになります。そのため、Content Manager または Enterprise Information Portal 内のユーザー ID またはグループを作成、変更、または削除する場合、MQSeries Workflow サーバーについても実行する必要があります。

Content Manager と Enterprise Information Portal はユーザー ID とグループを共用するため、ワークフロー・サービスが使用可能な状態で、システム管理クライアント内にユーザー ID またはグループを作成し、しかも MQSeries Workflow サーバーが実行されていないと、「ユーザーまたはグループが MQSeries Workflow サーバー上に作成されていません」というエラーを受け取ります。このユーザー ID は、Content Manager または Enterprise Information Portal 上のユーザーとしてまだ存在していますが、MQSeries Workflow サーバー上には存在していません。

コンテンツ管理ユーザー ID とグループを MQSeries Workflow のものと同期させるには、Workflow ユーザー同期ユーティリティを実行する必要があります。MQSeries Workflow サーバーを手動でインストールした場合は、MQSeries Workflow サーバーが開始済みであることを確認してください。「サービス (Services)」をオープンし、MQSeries Workflow X.X - FMC (ここで、X.X はインストールしたワークフローのバージョン) の状況を検査してください。MQSeries Workflow が稼働中ではない場合はサービスを開始するか、または自動方式のインストールを行った場合は、WFInstall ディレクトリーでバッチ・ファイル CMBWFStart.bat を実行します。同期ユーティリティを実行するには、以下の手順を完了してください。

1. ディレクトリーを、Enterprise Information Portal をインストールした場所に変更します。デフォルトのディレクトリーは C:\CMBROOT です。
2. EIPUser2WF.bat を入力します。

重要: リモート・ワークフロー・サーバーがある場合は、RMI サーバー・バッチまたはシェル・ファイル (cmbsvregist81.bat、または cmbsvregist81.sh) を始動してから、EIPUser2WF.bat と入力してください。また、EIPUser2WF.bat は 2 つの INI ファイルからの情報を使用して、ローカルまたはリモート・サーバー (cmbsvcs.ini および cmbsvclient.ini) を見つけます。cmbsvcs.ini が、サーバーがローカルの場合は LOCAL、またはサーバーがリモートの場合は REMOTE になっていることを確認してください。cmbsvclient.ini ファイルには、サーバー・ロケーションが含まれています。

3. EIP データベース名、ユーザー ID、パスワード、およびスキーマについて、必要な情報を入力してください。

必要な情報を入力すると、同期ユーティリティーは、Content Manager または Enterprise Information Portal サーバーから MQSeries Workflow サーバーに任意のユーザーをコピーします。このユーティリティーが完了した後は、ユーザーまたはグループが MQSeries Workflow サーバーには分からないというエラーを受け取らなくなります。

ユーザー ID またはグループを Content Manager または Enterprise Information Portal から削除する場合は、そのユーザー ID またはグループを MQSeries Workflow サーバーからも削除する必要があります。ユーザー ID またはグループが MQSeries Workflow サーバー上に存在しない場合は、Content Manager または Enterprise Information Portal から削除できません。たとえば、ユーザー ID またはグループを Content Manager または Enterprise Information Portal 内に、EIP ワークフロー・サーバー・オプションが使用不可の状態で作成するとします。その場合は、そのユーザー ID またはグループを削除しようとするときに、ワークフロー・サーバー・オプションを使用可能にします。そのユーザー ID またはグループが MQSeries Workflow 内に存在しないため、ユーザー ID が MQSeries Workflow サーバー内に存在しないというエラーを受け取ります。このエラーを訂正するには、EIPUser2WF ユーティリティーを実行してユーザー ID およびグループを同期させ、次にそのユーザー ID およびグループを削除します。

EIP サーバー (Workflow 使用可能) の再インストール

EIP サーバー内で EIP Workflow が使用可能化されていて、新規 EIP サーバーを再インストールしたい場合は、すべての EIP ワークフロー・データを MQSeries Workflow サーバーから除去する必要があります。

EIP サーバーを除去する前に、すべての EIP ワークフロー・データを MQSeries Workflow サーバーから除去するために、以下の手順を完了してください。 **重要:** 以下の順序で実行してください。

1. eClient または Workflow API を使用して、すべてのワークフロー・インスタンスを終了します。ワークフロー・インスタンスを終了することによって、すべての作業項目が削除されます。
2. EIP システム管理クライアントまたはシステム管理 API を使用して、ワークリストおよびワークフロー・テンプレートを削除します。
3. ワークフローが使用可能なときに、すべてのユーザーとユーザー・グループを EIP から削除し、MQSeries Workflow サーバーからも、すべてのユーザー ID とユーザー・グループが削除されます。
4. MQSeries Workflow server for EIP を再構成します。
 - a. EIP インストール CD-ROM を挿入します。コマンド・プロンプトを使用して、WfInstall ディレクトリに変更してください。
 - b. コマンド・プロンプトで、次のように入力します。

```
fmcibie -iCMBWFAdmin.fdl -uadmin -ppassword -o -f
```

重要: EIP および MQSeries Workflow の両方からユーザー ID およびグループを削除する前に EIP データベースを除去し、次に、削除した EIP データベース

内に存在していたユーザー ID およびグループと同じものを作成しようとする
と、そのユーザーとグループを EIP に追加できないというエラーになります。
そのエラー・メッセージは、次のとおりです。

```
DGL2616A: Fail to add the user: XXX -DGL2485A:  
This workflow user already exists.
```

この問題を解決するために、次の手順を完了してください。

1. EIP システム管理クライアント内の EIP ワークフロー・サービスを使用不可に
します。
2. EIP システム管理クライアントからログオフし、ログオンし直します。ワークフ
ロー・サービスを使用不可にした後ログインすれば、以前 EIP 内に存在してい
て MQSeries Workflow サーバー内にまだ存在しているユーザー ID およびグル
ープと同じものを作成することができます。
3. MQSeries Workflow サーバー内にすでに存在しているユーザー ID およびグル
ープを作成した後で、EIP ワークフロー・サービスを使用可能にしてください。

MQSeries Workflow と EIP データベース間のユーザー ID とグ ループの更新

MQSeries Workflow サーバー内に存在し、EIP データベース内には存在しないユー
ザー ID またはグループそれぞれについて、ユーザー ID およびグループを同期さ
せるには、以下の手順を完了してください。

1. ファイルを作成し、以下のテキストを入力します。
 - CODEPAGE 1252
 - FM_RELEASE V3R3 2
 - DELETE PERSON 'User1'
 - DELETE PERSON 'User2'
 - DELETE PERSON 'User3'

ここで、User1、User2、および User3 は、削除したいユーザーです。必要な数の
ユーザーをリストすることができます。

2. このファイルを保管し、コマンド・プロンプトで以下のコマンドを実行します。

```
fmcibie -u admin -ppassword -i DeletePersons.fdl -f -o
```

ここで、DeletePersons.fdl は前のステップで作成したファイルの名前です。

MQSeries Workflow サーバー内に存在し、EIP データベース内には存在しないグル
ープそれぞれについては、以下の手順を完了してください。

1. テキスト・ファイルを作成し、以下の情報を入力します。
 - CODEPAGE 1252
 - FM_RELEASE V3R3 2
 - DELETE ROLE 'Group1'
 - DELETE ROLE 'Group2'
 - DELETE ROLE 'Group3'

ここで、Group1、Group2、および Group3 は削除したいグループです。必要な数のグループをリストすることができます。

2. このファイルを保管し、コマンド・プロンプトで以下のコマンドを入力します。

```
fmcibie -u admin -p password -i DeleteGroups.fdl -f -o
```

ここで、DeleteGroups.fdl は前のステップで作成したファイルの名前です。

ユーザー ID を EIP から削除し、ワークフロー・サポートが使用可能な場合は、次のエラーが表示されます。

```
Failed to delete a user [RC=12]
```

ユーザーを削除する操作が失敗すると、システムは temp.log を生成し、詳細な情報が提供されます。temp.log ファイルは、x:%CMBROOT に書き込まれます。**ヒント:** MQSeries ワークフロー・クライアントへのログインに使用していたユーザー ID を削除しようとする、その操作は失敗します。

ワークフローの計画

ワークフローを定義するには、業務で実行する作業、それを実行する場所と方法、そしてその作業の担当者を前もって分析しておく必要があります。管理者または業務アナリストがこの計画ステップを行います。

最終成果物は何でしょうか？ 最終成果物は、業務、業務の 1 つの部門、または異なる部門の特定の従業員が実行したすべての作業の結果です。たとえば、XYZ 保険会社の保険金支払請求に対する補償プロセスの成果物は、保険契約者に保険金支払請求が承認されたか拒否されたかを通知する手紙になります。

まず、最終成果物を生成するために処理しなければならない情報を分析します。次に、実行するアクションと実行する場所を判別し、最後に、ワークフローでこの情報をどのように流すかを決定します。

処理する情報

企業内のユーザーが扱う情報について考えてみます。最終成果物をサポートする入力タイプは何でしょうか。処理しなければならない文書はどのようなものか？

作業項目 は、実際の文書ではありません。作業項目 には、文書への参照が含まれ、また、文書の状況、作成日などの、文書についての他の情報も含まれています。コンテンツ・サーバーからのいずれのコンテンツも (文書またはオブジェクト)、作業項目 になります。たとえば、XYZ 保険会社では、最初に保険金支払請求の用紙を受け取り、その後で、写真や査定記録、および専門家による報告書などの追跡文書を受け取ります。

情報を処理する方法

プロセスの各ステップを処理する適任者は誰でしょうか？ たとえば、管理部門のアシスタントは保険金支払請求の用紙が完全に記入されていることを確認し、保険契約者から特定の文書が送られてくるまで、この保険金支払請求の用紙をファイリングします。文書が届くと、査定人は、保険金支払請求の用紙と文書を突き合わせ、その文書を承認します。

ワーク・リストは、1人または複数の従業員が使用するために作成された作業のキューと見なすことができます。保険金支払請求の用紙は、複数の査定人が使用できるワーク・リストにグループ化できます。ワーク・リストは、フィルターを掛けた後の作業項目のビューです。従業員は、見ることを許可されたワーク・リストの中の項目のみを、見ることができます。

ワーク・リストは、写真や査定、報告書の収集などの保険金支払請求処理の各部分を処理するように、作業項目をフィルターに掛けて定義することができます。ワーク・リストは、それぞれ異なるワークフローの作業で構成することもできます。たとえば、1人の査定人のワーク・リストに、1件目の支払要求に対する査定、2件目の支払要求の写真、3件目の支払要求に対する専門家の報告書を含めることができます。査定人がこのワーク・リスト内の各項目に対して行う処理は異なります。査定人は査定を検討し、1件目の保険金支払要求を承諾するかもしれません。2件目の写真の処理については、その前にもっと多くの情報を待つ必要があると感じるかもしれません。また、3件目については、他の従業員に専門家の報告書を送って、処理してもらうこともできます。

実行するアクション

ワークフローの中において、作業項目のコンテンツでどのようなアクションを実行するかを考えてみます。たとえば、査定人は、保険金支払請求を受け入れたり、不備があるとして拒否したりできます。アクション・リストは、作業でユーザーが実行できるアクションを定義します。

たとえば、保険金支払請求が最初の要件に適合するかどうかにしたがって、査定人は、ワークフローで保険金支払請求を継続するか、あるいは保険金支払請求を拒否するかを選択できます。

プロセスでの情報の流れ

情報やアクティビティがどのように流れるかを考えてみます。たとえば、最初の保険金支払請求の用紙が検討されるのはいつか？ プロセスの次のステップに進むには、どのような追跡文書が必要か？ どのような基準に基づいて保険金支払請求を承認または拒否するかを判断するか？ この情報の流れがワークフローの基本となります。

ワークフローは、処理が終了するまで作業を導くパスで構成されています。入力発信源はどこか？ ワークフローの開始点を決めなければなりません。XYZ 保険会社の場合、保険契約者が提出した保険金支払請求の用紙が、ワークフローの開始点となる文書です。

すべての文書を受け取ったら、作業項目はパスに従って最終処理（たとえば、保険金支払請求の承認）まで進むことができます。

作業プロセス全体の統合

処理する情報を分析し、実行するアクションを判別して、さらに情報をどのように流すかを決定した後、次は、ワークフロー・ダイアグラムを作成します。これは、ワークフローを図形によって表したものです。EIP ワークフロー・ビルダー・フィチャーを使用してダイアグラムを作成します。

ワークフロー・ダイアグラムは、作業が、プロセス内の各種のアクティビティーをどのように移動するかを示し、各アクティビティーに関係するタスクを説明します。この図では、ワークフローの流れ、主なエレメント、およびキー・ポイントを説明します。

ワークフロー・ダイアグラム内の各記号は、作業の実行ポイントを示します。保険金支払請求が検討され、関係する文書が集められたら、一定の基準にしたがってこの保険金支払請求が承認されるか拒否されるかが決まります。ワークフロー・ビルダーで使用されるプロセスの記号について詳しくは、115 ページの『ワークフローの作成』を参照してください。

Enterprise Information Portal ワークフロー・コンポーネントの使用

このセクションでは、ワークフローのコンポーネントについて説明します。コンポーネントには、すべて、管理クライアントを介してアクセスします。**ヒント:** EIP バージョン 8 のワークフローでは、新しい Content Manager バージョン 8 のアーキテクチャーに適合するように、バージョン 7.1 コンテナの変更点を含め、いくつかの変更が加えられています。

ワークフロー・ビルダーの使用

ワークフロー・ビルダーを使用すれば、ワークグループ、部門、企業のいずれにおいても、ワークフローを図で定義し、構築することができます。**制限:** EIP のマイグレーション・プロセスは、バージョン 7.1 データベースからユーザーをマイグレーションします。EIP バージョン 8.2 には、ワークフロー・データの自動マイグレーション機能はありません。バージョン 7.1 のワークフロー・ダイアグラムは、EIP バージョン 8.2 のワークフロー・ビルダーを使用して作成し直し、EIP バージョン 7.1 のワークフロー・プロセスを再展開してください。

ワークフロー・ビルダーを使用してワークフローのモデルを作成する前に、特権セット、アクセス制御リスト、ユーザー、ユーザー・グループ、アクション、アクション・リスト、およびワーク・リストを定義する必要があります。管理クライアントでワークフローを定義する際には、ワークフロー全体のデフォルト・アクション・リストを設定することができます。さらに、ワークフローの各ノードに、それぞれ異なるアクション・リストを割り当てることもできます。これらのタスクについての詳細は、115 ページの『アクション・リストの定義』、114 ページの『ワーク・リストの定義』、およびオンライン・ヘルプを参照してください。

ユーザーは、ワークフロー・ビルダーを使用してワークフローを作成しますが、ワークフロー・ビルダーを使用してこのワークフローを実行することはできません。ユーザーは、クライアントを使用して、ワーク・リストおよび作業項目の処理を表示し、実行します。Enterprise Information Portal ワークフローを処理するようにクライアントをプログラミングする方法については、「ワークステーション・アプリケーション・プログラミング・ガイド」および「オンライン API 解説書」を参照してください。114 ページの表 7 に、すべてのワークフローに共通の 3 つのワークフロー・アイコンを示します。サブワークフロー、イベント、コレクション、およびユーザー出口など、他のワークフロー・アイコンもあります。ツールバー・アイコンの詳細については、EIP オンライン・ヘルプを参照してください。

表7. 基本ワークフロー・アイコン

アイコン	説明
	ワークフロー・プロセスは開始ノードから始まる。開始ノードは、ワークフロー・プロセス・ダイアグラム内に 1 つしかない。
	ワークフロー・プロセスは停止ノードで終了する。ワークフロー・プロセス・ダイアグラムごとに 1 つの停止ノードがある。プロセスを作成すると、自動的に停止ノードが生成される。停止ノードは、画面上で自由に移動できる。停止ノードは、ワークフロー・プロセス・ダイアグラム内に 1 つしかない。
	作業ノードは、ワークフロー・プロセス内の特定のポイントについて、ワーク・リストとアクション・リストを関連付ける。作業ノードは、作業が実行されるワークフロー・プロセス内のポイントを表す。各ノード (開始ノードおよび停止ノードを含む) には、アクションの集合およびそれらのアクションの実行を割り当てられた誰かが必要。

ワークフロー・サービスの使用

Enterprise Information Portal にはワークフロー情報を保持するワークフロー・サービスがあります。ワークフロー・ビルダーを使用して作成するワークフローおよびアクション・リストの定義は、Enterprise Information Portal 管理データベースおよび IBM MQSeries® Workflow データベースに保持されます。

システム管理者がワーク・リストを作成すると、そのワーク・リストに関連した情報は永久に管理データベースに保存されます。システム管理者は、EIP を使用してワーク・リストを更新、削除、追加することができます。システム管理者がワークフローをチェックアウトすると、ワークフローは Enterprise Information Portal データベース内でロックされ、チェックアウトの印が付けられて、ユーザーの作業が終了するまで別のユーザーはこのワークフローを更新することができなくなります。

ワーク・リストの定義

ワーク・リストは、使用可能な作業のフィルターと考えることができます。ワーク・リストは、特定のユーザーまたはユーザー・グループに割り当てられた項目をフィルター操作したリストです。ユーザーが Enterprise Information Portal にログインすると、割り当てられた作業項目のフィルター操作済みリストが表示されます。ワーク・リストを定義するには、Enterprise Information Portal 管理クライアントを使用します。

ワーク・リストの定義には、その作業項目の表示、状況、およびセキュリティーを管理する規則が含まれています。ワーク・リストを作成する時点で、それぞれのワーク・リストごとに規則を指定します。ワーク・リストへのアクセスを管理するには、そのワーク・リスト用のアクセス制御リストを作成します。ワーク・リストの定義方法についての完全な説明は、オンライン・ヘルプを参照してください。ワーク・リストの定義には、以下が含まれます。

アクセス制御リスト

アクセス制御リストは、1 つまたは複数の個人のユーザー ID やユーザー・グループ、およびそれぞれのユーザー ID またはユーザー・グループと関連する特権セットで構成されます。特権セットは、ユーザーが作業で特定のタスクにアクセスしたり実行したりするための権限を定義するのに使用されます。アクセス制御リストは、ワーク・リストの項目へのユーザー・アクセスを制御するのに使用します。

ワーク・リストのフィルター操作およびソート

フィルター操作とソートを実行したワーク・リストを表示するための基準。

ワーク・リストでの項目の最大数

ワーク・リストに入れる項目の最大数。

アクション・リストの定義

アクション・リストは、ワークフローでの作業においてユーザーが実行できるすべてのアクションの包括的なリストです。

アクションおよびアクション・リストを定義する方法の詳細については、「管理クライアントのオンライン・ヘルプ」を参照してください。

ワークフローの作成

アクション、アクション・リスト、およびワーク・リストを定義したら、ワークフロー・ビルダーを使用してワークフローのモデルを作成します。アクションおよびアクション・リストを定義する方法の詳細については、「管理クライアントのオンライン・ヘルプ」を参照してください。ワークフロー・ビルダーでは、ワークフローを作成する際に、視覚的合図があります。

ワークフロー・ビルダーの使用可能化

このステップでは、管理データベース上でワークフローを開始します。**制限:** ワークフロー用に選択するデータベースは、MQ Series をインストールしたサーバーと同じサーバー上になければなりません。また MQSeries サービスが開始されていなければなりません。

EIP ワークフローを使用可能にしてワークフロー定義を作成する場合は、次のステップに従ってください。

1. 管理クライアントにログインします。
2. 複数の管理データベースがある場合は、ワークフローを使用可能にしたいデータベースのアイコンをクリックします。
3. 「ツール (Tools)」 → 「サービス (Services)」 とクリックします。「ワークフローの使用可能化 (Enable Workflow)」 をクリックします。
4. クライアントからログアウトし、再度ログインします。複数のデータベースがある場合は、EIP ワークフローを使用可能にしたデータベースのアイコンを選択します。Workflows フォルダーのアイコンが表示されます。
5. Enterprise Information Portal 管理メイン・ウィンドウの左側のペインで、Workflows フォルダー「**Workflows**」をダブルクリックします。

6. 「ワークフロー定義 (Workflow Definitions)」アイコンを右マウス・ボタン・クリックし、「**新規 (New)**」を選択してワークフロー定義を作成します。

要件: ワークフローを定義する前に、少なくとも 1 つのアクセス制御リスト、アクション、およびアクション・リストを作成する必要があります。

MQSeries Workflow サーバーの開始

MQSeries Workflow サーバーを開始するには、コマンド・プロンプトで `cmbwfstart` を入力します。MQSeries Workflow サーバーの 2 つのウィンドウが開きます。サーバーを引き続き実行する場合は、これらのウィンドウを開いたままにしておきます。

Enterprise Information Portal の初期インストールを行ってからワークフローをインストールする場合は、ワークフロー・フィーチャー用に Enterprise Information Portal システムを構成する必要があります。管理クライアントがインストールされているワークステーションとは別のワークステーションにワークフロー・フィーチャーをインストールする場合にも、構成を変更する必要があります。

1. 管理ウィンドウで、ファイル・メンバー「**ツール (Tools)**」をクリックします。
2. メニューで、「**サービス (Services)**」をクリックします。
3. 「**ワークフロー (Workflow)**」チェック・ボックスを選択します。
4. 構成が完了したら、Enterprise Information Portal 管理クライアントをログオフしてから、ワークフロー・フィーチャーを初期化するために再度ログオンします。Enterprise Information Portal 管理クライアントにログオンすると、左側のペインに「**ワークフロー定義 (Workflow Definitions)**」アイコンが現れます。

ヒント: 管理者がワークフロー・フィーチャーの管理権限を持っていない場合、「**ワークフロー定義 (Workflow Definitions)**」アイコンは表示されません。ワークフロー・フィーチャーへのアクセス権を制限する場合は、各コンテンツ・サーバーの該当するシステム管理関連の資料を参照してください。ワークフロー・フィーチャーの管理権限を管理者に与える方法の詳細は、「**オンライン・ヘルプ**」を参照してください。

クライアントは、EIP コネクター・ツールキットおよびサンプルを使用して、カスタム・アプリケーションから作成することができます。または、EIP サンプル・クライアントを使用することができます。

第 7 章 IBM Web Crawler サンプル・ファイル

このセクションでは、コード・サンプルを 2 つ示します。 config-sample2.xml サンプル・ファイルには、 <gcs-config> 構成パラメーターの例が含まれています。ログ分析のサンプルは、完了したクローラからの情報を含んだレポートの例を提供します。

config-sample2.xml サンプル

このセクションのコード・サンプルは、gcs-config ファイルの 1 つの例です。

```
<!DOCTYPE gcs-config SYSTEM "config.dtd">
<gcs-config>
  <!-- Global settings: -->
  <globals max-urls="1000000"
    num-crawlers="30"
    num-summarizers="8"
    summaries-dir="summaries"
    log-file="log/LOG.txt"
    temp-dir="temp"
    log-priority="warn"
    text-monitor="60"
    graph-monitor="2"
    connect-timeout="120"
    read-timeout="100">

    <!-- sample logger settings -->
    <logger-config category="gcs.summaries.list.resource"
priority="info" log-file="log/resources.txt"/>
    <logger-config category="gcs.summaries.list.host" priority="info"
log-file="log/hosts.txt"/>
    <logger-config category="gcs.url.skipped" priority="info"
log-file="log/skipped_urls.txt"/>
    <logger-config category="gcs.url.fromto" priority="info"
log-layout="%m%n" log-file="log/fromto.txt"/>
    <logger-config category="gcs.http" priority="info"
log-file="log/http.txt"/>
    <logger-config category="gcs.http.connect" priority="info"
log-file="log/connecterrs.txt"/>

    <!--use this to specify a database
    <urlpool-config urlcontainer-class="DB2URLContainer"
urlcollection- class="DB2URLCollection">
      <urlpool-param name="dbname" value="gcs"/>
      <urlpool-param name="user" value="xxxxxx"/>
      <urlpool-param name="password" value="xxxxxx"/>
      <urlpool-param name="cachesize" value="1000"/>
      <urlpool-param name="driver"
value="COM.ibm.db2.jdbc.app.DB2Driver"/>
    </urlpool-config> -->

    <!--use this to specify a SOCKS proxy
    <system-properties>
      <property name="socksProxySet" value="true"/>
      <property name="socksProxyHost" value="socks2.server.ibm.com"/>
      <property name="socksProxyPort" value="1080"/>
    </system-properties> -->

  </globals>
```

```

<group-list>
  <group name="ibm">

    <crawler-config recursion-depth="-1">
      <seed-list>
        <!-- URLs to start crawling at: -->
        <seed url="http://gcs.stl.ibm.com/gcs/testurl.html"/>
        <seed url="http://gcs.stl.ibm.com/gcs/stl.html"/>
        <seed url="http://gcs.stl.ibm.com/gcs/ibm.html"/>
      </seed-list>

      <content-type-pattern-list>
        <!-- URL file extensions that don't match these patterns
        won't be crawled: -->
        <url-name-pattern name="htm*"/>
        <url-name-pattern name="pdf"/>
        <url-name-pattern name="gif"/>
        <url-name-pattern name="zip"/>
        <url-name-pattern name="txt"/>
      </content-type-pattern-list>

      <include-pattern-list>
        <!-- URLs that don't match these patterns won't be crawled: -->
        <url-obj-pattern host="*.ibm.com"/>
        <!-- sbo - url-obj-pattern query="*OpenDocument*" / -->
        <!-- sbo - url-obj-pattern query="*OpenView*" / -->
        <!-- url-obj-pattern query="*OpenDocument =>
OpenDocument&amp;ExpandAll*" / -->
        <!-- url-obj-pattern query="*OpenView =>
OpenView&amp;ExpandAll&amp;Count=999999*" / -->
      </include-pattern-list>

      <exclude-pattern-list>
        <!-- URLs that match these patterns won't be crawled: -->
        <!-- skip these common patterns in our intranet -->
        <url-obj-pattern file="*news*"/>
        <url-obj-pattern file="*search*"/>
        <url-obj-pattern file="*/afs*/>
        <url-obj-pattern file="*/...*/>
        <url-obj-pattern file="*bluepages*"/>
        <!-- skip personal home pages -->
        <url-obj-pattern file="*/~*"/>
        <!-- skip SOCKS: no URL should specify this directly -->
        <url-regex-pattern match=".*:1080/.*"/>
        <!-- skip gateways: recommended for mere mortals -->
        <url-regex-pattern match=".*[¥?¥=¥+¥;¥%&quot;&amp;].*"/>
        <!-- else crawl gateways as configured here... -->
        <!-- skip Domino -->
        <url-obj-pattern file="*.nsf*"/>
        <!-- else crawl Domino: allow only OpenDocument -->
        <url-obj-pattern query="*OpenServer*"/>
        <url-obj-pattern query="*OpenDatabase*"/>
        <url-obj-pattern query="*OpenElement*"/>
        <url-obj-pattern query="*OpenView*"/>
        <url-obj-pattern query="*OpenAbout*"/>
        <url-obj-pattern query="*OpenHelp*"/>
        <url-obj-pattern query="*OpenIcon*"/>
        <url-obj-pattern query="*OpenForm*"/>
        <url-obj-pattern query="*OpenNavigator*"/>
        <url-obj-pattern query="*OpenAgent*"/>
        <url-obj-pattern query="*CreateDocument*"/>
        <url-obj-pattern query="*DeleteDocument*"/>
        <url-obj-pattern query="*EditDocument*"/>
        <url-obj-pattern query="*SaveDocument*"/>
        <url-obj-pattern query="*SearchSite*"/>
        <url-obj-pattern query="*SearchView*"/>
        <url-obj-pattern query="*&login*"/>

```

```

        <url-obj-pattern    query="*Command*"/>
        <!-- crawl Domino: avoid OpenDocument permutations -->
        <url-obj-pattern    query="*ExpandSection*"/>
        <url-obj-pattern    query="*Navigate*"/>
        <url-obj-pattern    query="*Start*"/>
        <!-- -->

        </exclude-pattern-list>
    </crawler-config>

    <sumarizer-config>
    <!-- Copy2Rdf is the default summarizer. For these types, use: -->

        <resource-handler content-type="*htm*"
            summarizable="EipHtmlSummarizable"
            summary-maker="EipHtmlRawSummaryMaker" />
        <resource-handler content-type="*pdf"
            summarizable="InsoSummarizable"
            summary-maker="InsoSummaryMaker" />

    </sumarizer-config>
v    </group>
    </group-list>
</gcs-config>

```

IBM Web Crawler ログ分析ファイルの例

```
D:%gcs%run%log>perl loganalysis.pl log.txt
```

Log.txt 内の 7710 行の経過時間は、1.84 分でした。

```

GCS was configured for 20 crawlers
999 total crawls attempted
137 - total crawl failures:
      21  GCSHttpConnection.ABANDONING
      12  GCSHttpConnection.CONNECT_ERROR
      16  GCSHttpConnection.UNKNOWN_HOST
       4  HTTP 403
      29  HTTP 404
       2  HTTP 500
       8  HTTP 599
       1  Read timed out
      39  Robots not allowed
       4  over max redirects
       1  unknown protocol

```

```

-----
862 = successfully crawled
  0 - unchanged since earlier crawl
-----

```

```

862 = new or changed
468 crawled per minute

```

```

GCS was configured for 5 summarizers
855 total summaries attempted
  0 - total summary failures:
-----
855 = successfully summarized
144 gcs.summaries.list.host
855 gcs.summaries.list.resource
465 summarized per minute

```

```

GCS successfully crawled 134 servers to obtain 862 URL:
afqa0854.mop.ibm.com: 15
als1f1.yamato.ibm.com: 1
apache.btv.ibm.com: 1

```

apc.endicott.ibm.com: 2
as400service.ibm.com: 1
atlas.bocaraton.ibm.com: 1
autoproxym.ibm.com: 1
cer.si.ibm.com: 1
commerce.www.ibm.com: 1
crmweb.boulder.ibm.com: 3
d02ntcl01.ibm.com: 1
dacs.endicott.ibm.com: 1
duke.toraix.can.ibm.com: 1
ebcweb.austin.ibm.com: 1
ecspubs.ibmus2.ibm.com: 5
edaw3.fishkill.ibm.com: 1
endwww.endicott.ibm.com: 1
gcs.stl.ibm.com: 1
gustwick.austin.ibm.com: 1
ibmfnsys.somers.hqregion.ibm.com: 1
ibmpnyil.somers.hqregion.ibm.com: 2
ifw-www.mul.ie.ibm.com: 1
iplswww.nas.ibm.com: 2
itirc.ibm.com: 1
logosite.services.ibm.com: 1
lt.lahulpe.ibm.com: 17
messaging.ibm.com: 1
mrsmrn04.leeds.uk.ibm.com: 1
online.lahulpe.ibm.com: 1
page.sg.ibm.com: 1
procure.sby1.ibm.com: 1
reso.somers.hqregion.ibm.com: 1
ristal.leipzig.de.ibm.com: 1
rrhhar.argentina.ibm.com: 1
seashore.stl.ibm.com: 1
secureway.raleigh.ibm.com: 15
service.software.ibm.com: 1
software.ibmus2.ibm.com: 1
techcenter.austin.ibm.com: 1
tr2.fishkill.ibm.com:8080: 1
ucd.torolab.ibm.com: 1
usmweb.boulder.ibm.com: 1
w3-1.ibm.com: 32
w3-2.ibm.com: 3
w3-3.ibm.com: 108
w3-5.ibm.com: 4
w3.a-nz.au.ibm.com: 1
w3.academy.ibm.com: 1
w3.almaden.ibm.com: 2
w3.alphaworks.ibm.com: 1
w3.ap.ibm.com: 1
w3.asca.ibm.com: 7
w3.austin.ibm.com: 3
w3.boulder.ibm.com: 1
w3.br.ibm.com: 1
w3.btv.ibm.com: 1
w3.can.ibm.com: 40
w3.chq.ibm.com: 4
w3.coc.ibm.com: 1
w3.corporatetechnology.ibm.com: 1
w3.cupertino.ibm.com: 1
w3.dds.dfw.ibm.com: 17
w3.demopkg.ibm.com: 4
w3.design.ibm.com: 1
w3.developer.ibm.com: 3
w3.education.ibm.com: 1
w3.emea.ibm.com: 14
w3.enterlib.ibm.com: 7
w3.finsys.ibm.com: 1
w3.gcg.ibm.com: 1

w3.globalfinancing.de.ibm.com: 1
w3.hakozaki.ibm.com: 1
w3.houston.ibm.com: 1
w3.hursley.ibm.com: 5
w3.iabc.ibm.com: 1
w3.ibm.com: 180
w3.ibmfax.ibm.com: 1
w3.ibmmla.ibm.com: 14
w3.isicc.de.ibm.com: 1
w3.itso.ibm.com: 1
w3.japan.ibm.com: 1
w3.knowledge.raleigh.ibm.com: 1
w3.linux.ibm.com: 1
w3.marketiq.ibm.com: 1
w3.micro.ibm.com: 2
w3.mtlisc.can.ibm.com: 1
w3.munich.ibm.com: 1
w3.ode.raleigh.ibm.com: 1
w3.paylink.au.ibm.com: 1
w3.pisc.uk.ibm.com: 1
w3.pl.ibm.com: 1
w3.printers.ibm.com: 1
w3.pssc.mop.ibm.com: 1
w3.pssed.au.ibm.com: 1
w3.raleigh.ibm.com: 3
w3.rchland.ibm.com: 1
w3.research.ibm.com: 3
w3.reserve.ibm.com: 1
w3.rs6000.ibm.com: 1
w3.security.ibm.com: 1
w3.software.ibm.com: 6
w3.ssd.ibm.com: 1
w3.stl.ibm.com: 1
w3.techline.ibm.com: 1
w3.techsupp.yamato.ibm.com: 1
w3.torolab.ibm.com: 2
w3.usergroup.ibm.com: 1
w3.vendor.pok.ibm.com: 1
w3.viewblue.ibm.com: 1
w3.watson.ibm.com: 2
w3.wdg.uk.ibm.com: 1
w3.ytal.yasu.ibm.com: 1
w3.zurich.ibm.com: 1
w3chq.disbursements.ibm.com: 1
w3is.lagaude.ibm.com: 1
w3md.btv.ibm.com: 1
w3ssd.mainz.de.ibm.com: 1
w3vm.demopkg.ibm.com: 1
widweb.raleigh.ibm.com: 1
wtscpok.itso.ibm.com: 1
wwas.raleigh.ibm.com: 1
www-1.ibm.com: 63
www-3.ibm.com: 4
www-4.ibm.com: 86
www.almaden.ibm.com: 1
www.as400.ibm.com: 1
www.chips.ibm.com: 1
www.ibm.com: 52
www.ieg.ibm.com: 1
www.patents.ibm.com: 1
www.pc.ibm.com: 2
www.rs6000.ibm.com: 23
www.software.ibm.com: 9
www.storage.ibm.com: 1
www.watson.ibm.com: 1

GCS timed out 1 times:

w3-3.ibm.com: 1

GCS ignored 42 URL prohibited by robots.txt:

reso.somers.hqregion.ibm.com: 1
w3.education.ibm.com: 1
w3.rchland.ibm.com: 34
w3.zurich.ibm.com: 1
www.ibm.com: 5

GCS skipped 3846 URL (requires gcs.url logging)

59 specified an unsupported protocol:

protocol not supported gopher: 6
protocol not supported mailto: 53

1206 had content-types (lower or UPPER case, > 10) that were not included

.2: 12
.faq: 13
.1: 14
.asp: 16
.cgi: 21
.shtml: 90
.pl: 92
.gif: 157
.nsf: 160
.jpg: 214
.css: 240

516 URL were on servers and/or paths that were not included

2065 were excluded for these reasons:

URL longer than 254: 1
excluded by rule 1: 1210
excluded by rule 2: 854

第 8 章 テキスト検索および QBIC[®] の使用

この付録の最初のセクションでは、テキスト検索およびイメージ・コンテンツによる照会 (Query by Image Content (QBIC)) の構成と使用の方法を説明します。このテキスト検索および QBIC の 2 つのフィーチャーは、EIP のインストール時に、Content Manager バージョン 7.1 コネクターを選択した場合にのみ使用可能になります。この付録の 2 番目のセクションでは、サンプル・アプリケーションで使用するサンプル・テキストとイメージ・データのロードについての情報を提供します。

テキスト検索エンジンを用いた文書検索

テキスト検索は Content Manager バージョン 7.1 サーバーに統合することができ、これにより、Content Manager に保管されている文書を自動的に索引付けして、検索し、取り出すことができます。ユーザーは、語または句を検索して文書を探し出すことができます。この Text Search Server は、単一バイト文字セットと 2 バイト文字セットの両方をサポートしており、AIX と Windows の両方で稼働します。

テキスト検索には、XML、HTML、およびタグ付きの ASCII 文書の構造化された文書サポートが組み込まれています。このサポートを使って、文書内の指定セクションで用語を検索することができます。ネストされたセクションでデータを検索することができます。XML のフル・コンテキストで検索することができます。たとえば、複数のタイトルの中で IBM を検索したり、特定のセクション内の 1 つのタイトルの中で IBM を検索したりできます。DTD パスを指定すると、テキスト検索は、各文書ごとに該当する DTD を動的に使うことができます。その場合、DTD の参照は、その文書のメタデータとして保管されるものと見なされます。

テキスト検索を用いる EIP システムの計画とインストールについての詳細は、「Enterprise Information Portal の計画とインストール」を参照してください。

Text Search Server の使用可能化

Text Search Server を使用するには、IBM Content Manager for Multiplatforms 管理クライアントを始動する前に、このサーバーの管理を使用可能にしておかなければなりません。管理を使用可能にするには、次の手順を実行します。

1. IBM Content Manager for Multiplatforms ライブラリー・サーバーを始動します。
ライブラリー・サーバーに、索引クラスの作成を完了させます。
2. 次のように入力して、これがインストールされているワークステーションで Text Search server を始動します。

```
imlss -start dlinst
```

dlinst は、インストール時または *imlcfgsv* コマンド・ユーティリティーの使用時に選んだ Text Search Server インスタンスの名前です。

イメージ・コンテンツによる照会 (Query by Image Content (QBIC)) を用いたイメージ検索

このセクションでは、イメージ・コンテンツによる照会 (Query by Image Content (QBIC)) を紹介し、QBIC の構成と使用の方法を説明します。QBIC フィーチャーは、Content Manager バージョン 7.1 コネクターをインストールした場合にのみ使用可能になります。QBIC は、Windows および AIX オペレーティング・システムと互換性があります。

イメージ検索の紹介

イメージ検索サーバーは、IBM の QBIC (イメージ・コンテンツによる照会) テクノロジーを使用して、色やテクスチャーなどの特定の視覚的な特性によるオブジェクトの検索を支援します。イメージ検索サーバーはイメージを分析し、そのイメージ情報をデータベースに格納します。その後イメージ照会を実行しますが、イメージの視覚的な特性を使用します。これにより、言語で特性を記述せずに色、テクスチャー、位置を突き合わせるすることができます。内容に基づいた照会とテキストおよびキーワードの検索を組み合わせることにより、より強力なイメージ検索やマルチメディア・データの検索が可能になります。

各イメージ検索サーバーには、1 つ以上のイメージ検索データベースを含むデータ・ディレクトリーが備えられています。イメージ検索カタログはここに格納されます。イメージ検索カタログには、イメージのコレクションの視覚特性についてのデータが収められています。実際のイメージ・オブジェクトは、IBM Content Manager for Multiplatforms システム内のオブジェクト・サーバーに格納されます。このイメージ検索サーバーは、AIX と Windows で稼働します。

イメージ検索のインストールについての詳細は、「*Enterprise Information Portal* の計画とインストール」を参照してください。

イメージ検索の設定

次の説明は、Content Manager バージョン 7.1 コネクターを選択した場合に自動的にインストールされるイメージ検索をインストールした後で、適用できます。イメージ検索の設定は次のステップから構成されます。

1. 環境の設定
2. イメージ検索サーバーの構成
3. イメージ検索クライアントの構成
4. サンプル・イメージ・データのロード

AIX でインストール・ウィザードを使用する場合: 構成およびセットアップ・スクリプトを実行したり、サーバー構成コマンドを発行したりする必要はありません。ウィザードが、これらのタスクを自動的に完了します。

Windows でインストールを行う場合: これらのタスクを自分で実行する必要があります。

環境の設定

このセクションでは、サーバーとクライアントの両方のマシンで、環境セットアップ・タスクを完了します。イメージ検索サーバーには、次の環境変数が必要です。

QBICTOP

イメージ検索構成中にファイル名を解決します

QbicImagePath

サーバー・イメージ・ファイルでファイル名を解決します

QbicMaskPath

サーバー・マスク・ファイルでファイル名を解決します

QbicSketchPath

サーバー・スケッチ・ファイルでファイル名を解決します

QbicTextPath

サーバー・テキスト・ファイルでファイル名を解決します

イメージ検索クライアントに必要なのは、QBICTOP 環境変数だけです。

AIX の例: AIX では、セットアップ・スクリプトを生成する構成スクリプトを実行してから、そのセットアップ・スクリプトを実行して環境をセットアップします。

1. 次の構成スクリプトを実行します。

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

ここで QBICTOP は、制御ファイル (*.ini) へのディレクトリー・パスです。QBICTOP を /user1/cmb/qbic に設定します。ここで /user1 はイメージ検索管理ユーザー ID のホーム・ディレクトリーです。イメージ検索ユーザー ID は、このディレクトリーへの読み取り/書き込みアクセスも持っていなければなりません。

このスクリプトは、セットアップ・スクリプト frnsetup.iss を生成します。

2. イメージ検索ユーザー ID のホーム・ディレクトリーから、以下を実行します。

```
./frnsetup.iss
```

このスクリプトは、環境変数をイメージ検索サーバーおよびクライアントに移植します。

Windows の例: 環境変数を設定する方法は次のとおりです。

1. 「スタート」→「設定」→「コントロール パネル」と選択します。
2. 「システム」をダブルクリックします。
3. 「環境 (Environment)」タブをクリックします。
4. 表 8 に示される変数と値を該当するフィールドに入力して設定し、「設定」ボタンをクリックします。

要件: イメージ検索クライアントに必要なのは、QBICTOP 変数だけです。クライアント環境の場合、QBICTOP 変数だけを設定します。

表 8. イメージ検索環境変数

変数	値
QBICTOP	d:%cmbroot%iss
QbicImagePath	d:%cmbroot%iss
QbicMaskPath	d:%cmbroot%iss

表 8. イメージ検索環境変数 (続き)

変数	値
QbicSketchPath	d:%cmbroot%iss
QbicTextPath	d:%cmbroot%iss

ここで、d: はイメージ検索がインストールされるドライブです。

イメージ検索サーバーの構成

イメージ検索サーバーの開始前に、それを構成する必要があります。サーバーの構成は、初期構成の完了と接続の検査から成り立ちます。

サーバーを構成する方法は次のとおりです。

1. qbicadm と入力してコマンド・インタープリターを開始します。
2. **config server** コマンドを入力します。たとえば、以下のようになります。

```
config server LIBSRVRN FRNADMIN PASSWORD 9999
```

ここで LIBSRVRN はライブラリー・サーバー名、FRNADMIN は Content Manager ユーザー ID、PASSWORD は Content Manager パスワード、および 9999 はイメージ検索サーバーのポート番号です。

詳細については、『接続の検査』を参照してください。

イメージ検索クライアントの構成

イメージ検索システム管理プログラムを含め、イメージ検索クライアントは開始前に構成する必要があります。Content Manager システム管理では、1 つの別名を割り当てる必要があります。接続を検査することにより、構成をテストしてください。

別名の割り当て: イメージ検索クライアントとして作動するイメージ検索 Content Manager システム管理を使用する前に、最低 1 つのサーバー別名を割り当てる必要があります。

別名を割り当てる方法は次のとおりです。

1. qbicadm と入力してコマンド・インタープリターを開始します。
2. **add alias** コマンドを入力します。たとえば、以下のようになります。

```
add alias QBICSRV HOSTNAME 9999
```

ここで QBICSRV は別名、HOSTNAME はイメージ検索サーバーのホスト名、9999 はイメージ検索サーバーのポート番号です。

接続の検査:

重要:

1. イメージ検索サーバーに接続するには、ライブラリー・サーバーが稼働状態である必要があります。
2. イメージ検索システム管理には、既存の Content Manager ユーザー ID が必要です。ライブラリー・サーバーに正常に接続するには、イメージ検索ユーザー ID およびライブラリー・サーバー・ユーザー ID が同じでなければなりません。こ

のユーザー ID のデフォルト値は frnadmin です。この値を変更する場合は、必ず ID を一致させるようにしてください。

接続を検査する方法は次のとおりです。

1. イメージ検索サーバーを構成し、別名を追加したら、サーバーのコマンド行から `commsrv` と入力してサーバーを開始します。
2. コマンド・インタプリターを開始するには、`qbicadm` と入力します。
3. コマンド・インタプリターで、`connect` コマンドを入力します。

```
connect QBICSRV FRNADMIN PASSWORD
```

ここで、QBICSRV は別名、FRNADMIN は Content Manager ユーザー ID、PASSWORD は Content Manager パスワードです。

接続に成功すると、メッセージ `Library Server is LIBSRVRN` が表示されます。

4. サーバーから切断するには、`disconnect` と入力します。
5. コマンド・インタプリターを終了するには、`quit` と入力します。

サンプル・データのロードおよび索引作成

このセクションでは、サンプル・アプリケーションで使用するサンプル・テキストとサンプル・イメージ・データのロード、および索引作成の方法を説明します。このセクションは、Content Manager バージョン 7.1 コネクターをインストールしてテキスト検索オプションを選択した場合にのみ、適用されます。

Enterprise Information Portal CD には、いくつかのサンプル・ローダーが用意されています。このセクションでは、サンプル・ローダーの `LoadSampleTSQBICDL` を使用して、イメージ・データとテキスト・データの両方についてロードの方法を説明します。それぞれのフィーチャーが正常に動作することを確認するため、テキスト、およびイメージの検索データを別々にロードすることができます。

サンプル・データをロードする前に

ローダー・プログラムを実行する前に次のことを行う必要があります。

1. EIP 管理クライアントにログインします。「スタート」→「プログラム」→「**Enterprise Information Portal for Multiplatforms 8.2**」→「管理 (Administration)」の順にクリックします。
2. データベースを選択し、正しいユーザー ID とパスワードを使用してログインします。デフォルト・データベース `icm1sdb` を選択する場合は、ユーザー ID に `icmadmin` と入力し、パスワード・フィールドに `password` と入力します。別のデータベースを使用する場合は、適切なユーザー ID を入力します。
3. Content Manager システム管理プログラムを使用してライブラリー・サーバー構成を作成します。このタスクについて不明な点がある場合は、システム管理プログラムのオンライン・ヘルプを参照してください。
4. 以下のステップを実行して、ライブラリー・サーバー構成の「**アクセス (Access)**」プロパティーを変更します。
 - a. 新しい構成を右クリックして「**プロパティー (Properties)**」をクリックし、「プロパティー (Properties)」ノートブックをオープンします。

- b. 「アクセス (Access)」タブをクリックします。
- c. 「任意のワークステーションからの無制限セッション (Unlimited sessions from any workstation)」ラジオ・ボタンをクリックします。

テキスト検索索引の作成

データをロードする前に、空のテキスト検索の索引を作成する必要があります。これによって、テキストのサンプルに索引を付けることができます。**ヒント:** テキスト検索の索引は、Content Manager バージョン 6.1 またはバージョン 7.1 のサーバーでしか作成できません。

テキスト検索索引を作成するには次のようにします。

1. 次のコマンドを使って、インストール先のワークステーションで Text Search Server を始動します。

```
imlss -start dlinst
```

dlinst は、インストール時または imlcfgsv コマンド・ユーティリティーの使用時に選んだ Text Search Server インスタンスの名前です。
2. Content Manager システム管理プログラムを開始して、ログインします。
3. 左の一番上にあるペインのリストから、「**テキスト検索 (Text Search)**」を選択します。
4. 左のペインで「**検索サーバー (Search Servers)**」をダブルクリックします。
5. 「**TM**」をダブルクリックします。TM というのは、Text Search Server の検索サーバーの別名です。
6. 左のペインから「**Indexes**」フォルダーをダブルクリックします。メッセージ RC_EMPTY_LIST が表示されたら、メニュー・バーから「**選択 (Selected)**」→「**新規 (New)**」とクリックして索引を作成します。
7. 「新しい索引 (New Index)」ウィンドウで、索引を定義します。各フィールドの詳細な説明については、「**ヘルプ (Help)**」を参照してください。
たとえば、以下のようになります。

Windows の場合:

名前 (Name)

TMINDEX

タイプ (Type)

Precise

索引ファイル (Index files)

x:%cmbroot%ts%index%tmindex。ここで、x はインストール・ドライブを表します。パスが存在しない場合は作成されます。

索引作業ファイル (Index work files)

x:%cmbroot%ts%work%tmindex。ここで、x はインストール・ドライブを表しています。パスが存在しない場合は作成されます。

情報項目 (Information entry)

Content Manager ライブラリー・サーバーの名前。

クライアントとサーバーのデフォルトの DLL 名は変更しないでください。

AIX の場合:

名前 (Name)

TMINDEX

タイプ (Type)

Precise

索引ファイル (Index files)

/home/cltadmin/tsindex/index/tmindex。パスが存在しない場合は作成されます。

索引作業ファイル (Index work files)

/home/cltadmin/tsindex/work/tmindex。パスが存在しない場合は作成されます。ユーザーにディレクトリーへの書き込みの権限があることを確認してください。

情報項目 (Information entry)

Content Manager ライブラリー・サーバーの名前。

8. 「**OK**」をクリックします。
9. 「**TMINDEX**」をダブルクリックして、「**TMINDEX 管理 (TMINDEX Administration)**」ノートブックをオープンします。

イメージ検索データベース、カタログ、およびフィーチャーの作成

サンプルのテキスト検索データ用のテキスト検索索引を作成した後に、サンプル・イメージ・データ用のイメージ検索データベースおよびカタログを作成する必要があります。

イメージ検索データベース、カタログ、フィーチャーを作成するには次のようにします。

1. 以下のコマンドを入力して、インストール先のワークステーションでイメージ検索サーバーを始動します。

```
commsrv
```
2. Content Manager システム管理プログラムを開始して、ログインします。
3. 左の一番上にあるペインのリストから、「**イメージ検索 (Image Search)**」を選択します。
4. 左のペインで「**イメージ検索サーバー (Image Search Servers)**」をクリックします。
5. 「**QBICSRV**」をクリックします。

ここで QBICSRV は、インストール時に指定したイメージ検索サーバー名です。
6. 左のペインで「**データベース (Databases)**」を右クリックし、「**新規データベース (New Database)**」を選択します。
7. 「**新規データベース (New Database)**」ウィンドウで、「**名前 (Name)**」フィールドに SAMPLEDB と入力し、「**OK**」をクリックします。
8. 左のペインで「**データベース (Databases)**」をクリックして、「**SAMPLEDB**」アイコンを左のペインに表示させます。
9. 「**SAMPLEDB**」をクリックします。

10. 左のペインで、「**カタログ (Catalogs)**」を右クリックして、「**新規カタログ (New Catalog)**」をクリックします。
11. 「**新規カタログ (New Catalog)**」ウィンドウで、「**名前 (Name)**」フィールドに **SAMPLECAT** と入力し、「**OK**」をクリックします。
12. 左のペインで、「**カタログ (Catalogs)**」をクリックして「**SAMPLECAT**」アイコンを表示します。
13. 「**SAMPLECAT**」をクリックします。
14. 左のペインで「**フィーチャー (Features)**」を右クリックして、「**新規フィーチャー (New Features)**」をクリックします。
15. 「**新規フィーチャー (New Features)**」ウィンドウの「**名前 (Name)**」フィールドで各フィーチャーを選択し、「**適用 (Apply)**」をクリックします。4 つのフィーチャーをすべて選択すると、次のメッセージが表示されます。
「使用可能なすべてのフィーチャーがカタログに追加されました。
(All possible features have been added to catalog.)」
16. 「**OK**」をクリックします。
17. 「**キャンセル (Cancel)**」をクリックします。

ローダー・プログラムの実行

テキストおよびイメージ検索をテストするためのサンプル・データをロードすることができます。

サンプル・イメージ・データは、以下のファイルにあります。

Windows の場合:

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

AIX の場合:

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

サンプル・ローダー・プログラムは、データを **Content Manager** にロードし、索引を付けます。プログラムを実行するための構文についての説明は、ソース・プログラムのプロローグを参照してください。サンプル・ローダー・プログラムは次のファイルです。

Windows の場合:

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

AIX の場合:

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

サンプル・データ・ローダー・プログラムを実行するには、次のようにします。

1. 次のように入力して **.jar** ファイルを解凍します。

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

ファイルは、適切なディレクトリーに解凍されます。

2. ワークステーションの環境変数を設定し、以下のタスクを行うことによってサンプル・ローダー・プログラムをコンパイルします。

Windows の場合:

- a. テキスト・エディターで `x:%cmbroot%cmbenv71.bat` を開き、次のように最初の 3 行を変更して、ワークステーションの環境変数を設定します。

```
set CMBROOT = e:%cmbroot
set DB2HOME = e:%sqllib
set JAVAHOME = d:%jdk117
```

- b. `cmbenv71.bat` を保管し、以下のように入力して環境変数を設定します。

```
cmbenv71
```

AIX の場合:

- a. `/usr/lpp/cmb/bin/` に進み、以下のように入力してセットアップを実行します。

```
./cmbenv71.sh
```

- b. `/usr/lpp/cmb/samples/java/d1/` の下にあるすべてのサブディレクトリー、およびサンプル・ファイルがすべてのユーザーから書き込めるようにします。

3. 次のコマンド (大文字小文字の区別あり) を入力して、ローダー・プログラムをコンパイルします。

```
javac LoadSampleTSQBICDL.java
```

4. **要件:** ローダー・プログラムを始動するには、事前に以下のサーバーを稼動させておく必要があります。

- ライブラリー・サーバー
- オブジェクト・サーバー
- テキスト検索サーバー
- イメージ検索サーバー

各国語バージョンの Content Manager を実行する場合、ローダー・プログラムを実行する前に `FRNDEFLANG` 変数を `ENU` に設定します。この環境変数を設定する AIX コマンドは `export FRNDEFLANG=ENU` です。

5. 次のように入力して、ローダー・プログラムでサンプル・データをロードします。

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

ここで、ユーザー ID は `frnadmin`、パスワードは `password`、およびライブラリー・サーバーは `LIBSRVRN` です。

6. `load.log` を検査して、サンプル・データが正常にロードされたことを確認してください。

サンプル・データをロードし終わったら、Content Manager システム管理プログラムまたはテキスト検索コマンド行ツールを使って、サンプル・テキスト・データに索引を付けます。

サンプル・テキスト・データの索引作成

サンプル・データに索引を作成するには、次のようにします。

1. Content Manager システム管理プログラムを開始して、ログオンします。

2. 左の一番上にあるペインのリストから、「**テキスト検索 (Text Search)**」を選択します。
3. 「**検索サーバー (Search Servers)**」をダブルクリックします。
4. 「**TM**」をダブルクリックします。TM というのは、Text Search Server の検索サーバーの別名です。
5. 「**新しいテキスト索引 (new text index)**」を右クリックし、「**プロパティ (Properties)**」をクリックします。
6. 「プロパティ (Properties)」ノートブックの「**明示 (Explicit)**」ページで、「**最新表示 (Refresh)**」をクリックします。
7. 「**索引カウント (Index count)**」フィールドには、ローダー・プログラムでロードされた文書の数が表示されます。
8. 「**索引 (Index)**」をクリックして、ファイルの索引を作成します。
9. しばらくしてから、「**最新表示 (Refresh)**」をクリックし、「**1 次文書索引 (Primary document index)**」フィールドで、正常に索引が作成された文書の数を確認します。

データに索引を付け終わったら、サンプルの Java アプリケーションで、imlsrch コマンド行ツールを使って、コレクションを照会したり、単純な照会を実行したりできます。

第 9 章 文書フォーマット

情報マイニングの文書フォーマット

この付録では、情報マイニングによってサポートされている文書フォーマットについて説明します。

ワード・プロセッシング: 一般

ANSI テキスト (7 & 8 ビット)	全バージョン
ASCII テキスト (7 & 8 ビット・バージョンが利用可能)	全バージョン
HTML	バージョン 3.0 まで (一部の制約あり)
IBM FFT	全バージョン
IBM 変更可能テキスト	全バージョン
Microsoft RTF (リッチ・テキスト・フォーマット)	全バージョン
ユニコード・テキスト	全バージョン

ワード・プロセッシング: DOS

DEC WPS Plus (DX)	バージョン 4.0 まで
DEC WPS Plus (WPL)	バージョン 4.1 まで
DisplayWrite® 2 & 3 (TXT)	全バージョン
DisplayWrite 4 & 5	バージョン、リリース 2.0 まで
Enable	バージョン 3.0、4.0、および 4.5
First Choice	バージョン 3.0 まで
Framework	バージョン 3.0
IBM Writing Assistant	バージョン 1.01
Lotus Manuscript	バージョン 2.0 まで
MASS11	バージョン 8.0 まで
Microsoft Word	バージョン 6.0 まで
Microsoft Works	バージョン 2.0 まで
MultiMate	バージョン 4.0 まで
Navy DIF	全バージョン
Nota Bene	バージョン 3.0

Office Writer	バージョン 4.0 から 6.0 まで
PC-File Letter	バージョン 5.0 まで
PC-File+ Letter	バージョン 3.0 まで
PFS:Write	バージョン A、B、および C
Professional Write	バージョン 2.1 まで
Q&A	バージョン 2.0
Samna Word	バージョン 4.0 まで
SmartWare II	バージョン 1.02
Sprint	バージョン 1.0
Total Word	バージョン 1.2
Volkswriter 3 & 4	バージョン 1.0 まで
Wang PC (IWP)	バージョン 2.6 まで
WordMARC	バージョン Composer Plus まで
WordPerfect	バージョン 6.1 まで
WordStar	バージョン 7.0 まで
WordStar 2000	バージョン 3.0 まで
XyWrite	バージョン III Plus まで

ワード・プロセッシング: インターナショナル

JustSystem 一太郎	バージョン 5.0、6.0、8.0、9.0、および 10.0
-----------------------	--------------------------------

ワード・プロセッシング: Windows

AMI/AMI Professional	バージョン 3.1 まで
Corel WordPerfect for Windows	バージョン 9.0 まで
JustWrite	バージョン 3.0 まで
Legacy	バージョン 1.1 まで
Lotus WordPro (Win32 / Intel プラットフォーム)	SmartSuite® 96、97、および Millennium
Lotus WordPro (Unix プラットフォーム - テキストのみ)	SmartSuite 97 および Millennium
Microsoft Windows Works	バージョン 4.0 まで
Microsoft Windows Write	バージョン 3.0 まで
Microsoft Word 97	Word 97
Microsoft Word 2000	Word 2000
Microsoft Word for Windows	バージョン 7.0 まで

Microsoft WordPad	全バージョン
Novell Perfect Works	バージョン 2.0
Novell WordPerfect for Windows	バージョン 7.0 まで
Professional Write Plus	バージョン 1.0
Q&A Write for Windows	バージョン 3.0
WordStar for Windows	バージョン 1.0

ワード・プロセッシング: Macintosh

Microsoft Word	バージョン 4.0 から 6.0 まで
Microsoft Word 98	Word 98
WordPerfect	バージョン 1.02 から 3.0 まで
Microsoft Works	バージョン 2.0 まで
MacWrite II	バージョン 1.1

スプレッドシート・フォーマット

VP Planner 3D	バージョン 1.0
Enable	バージョン 3.0、4.0、および 4.5
First Choice	バージョン 3.0 まで
Framework	バージョン 3.0
Lotus 1-2-3® (DOS & Windows)	バージョン 5.0 まで
Lotus 1-2-3 for SmartSuite	SmartSuite 97 および Millennium
Lotus 1-2-3 Charts (DOS & Windows)	バージョン 5.0 まで
Lotus 1-2-3 (OS/2®)	バージョン 2.0 まで
Lotus 1-2-3 Charts (OS/2)	バージョン 2.0 まで
Lotus Symphony	バージョン 1.0、1.1、および 2.0
Microsoft Excel 97	Excel 97
Microsoft Excel 2000	Excel 2000
Microsoft Excel Macintosh	バージョン 3.0 から 4.0、および 98
Microsoft Excel Windows	バージョン 2.2 から 7.0 まで
Microsoft Excel Charts	バージョン 2.x から 7.0 まで
Microsoft Multiplan	バージョン 4.0
Microsoft Windows Works	バージョン 4.0 まで
Microsoft Works (DOS)	バージョン 2.0 まで
Microsoft Works (Mac)	バージョン 2.0 まで

Mosaic Twin	バージョン 2.5
Novell Perfect Works	バージョン 2.0
QuattroPro for DOS	バージョン 5.0 まで
QuattroPro for Windows	バージョン 9.0 まで
PFS:Professional Plan	バージョン 1.0
SuperCalc 5	バージョン 4.0
SmartWare II	バージョン 1.02

データベース・フォーマット

SmartWare II	バージョン 1.02
Access	バージョン 2.0 まで
dBase	バージョン 5.0 まで
DataEase	バージョン 4.x
dBXL	バージョン 1.3
Enable	バージョン 3.0、4.0、および 4.5
First Choice	バージョン 3.0 まで
FoxBase	バージョン 2.1
Framework	バージョン 3.0
Microsoft Windows Works	バージョン 4.0 まで
Microsoft Works (DOS)	バージョン 2.0 まで
Microsoft Works (Mac)	バージョン 2.0 まで
Paradox (DOS)	バージョン 4.0 まで
Paradox (Windows)	バージョン 1.0 まで
Personal R:BASE	バージョン 1.0
R:BASE 5000	バージョン 3.1 まで
R:BASE System V	バージョン 1.0
Q & A	バージョン 2.0 まで
Reflex	バージョン 2.0

標準グラフィック・フォーマット

PNG - Portable Network Graphics Internet Format	バージョン 1.0
Binary Group 3 Fax	全バージョン
BMP(RLE、ICO、CUR、および os/2 DIB を含む)	Windows
CDR (TIFF イメージが CDR に組み込まれている場合)	Coral Draw バージョン 2.0 から 9.0 まで

CGM - Computer Graphics Metafile	ANSI、CALs、NIST、バージョン 3.0
CMX - Corel Clip Art Format	バージョン 5 から 6 まで
DCX (マルチページ PCX)	Microsoft Fax
DRW - Micrografx Designer	バージョン 3.1
DRW - Micrografx Draw	バージョン 4.0 まで
DXF (バイナリーおよび ASCII) AutoCAD Drawing Interchange Format	バージョン 14 まで
EMF	Windows Enhanced Metafile
EPS Encapsulated PostScript	TIFF イメージが EPS に組み込まれている場合
FMV - FrameMaker グラフィックス	ベクトルおよびラスター・フォーマットでバージョン 5.0 まで
FPX - Kodak Flash Pix	特定フォーマットの指定はない
GDF - IBM グラフィックス・データ・フォーマット	バージョン 1.0
GEM - Graphics Environment Manager Metafile	ビットマップおよびベクトル
GIF - グラフィック交換形式	CompuServe
GP4 - Group 4 CALS Format	タイプ I およびタイプ II
HPGL - Hewlett Packard Graphics Language	バージョン 2.0
IMG - GEM Paint	特定バージョンの指定はない
JFIF (TIFF フォーマットではない JPEG)	全バージョン
JPEG - Joint Photographic Experts Group フォーマット	全バージョン
MET - OS/2 PM Metafile	バージョン 3.0
PBM - Portable Bitmap	特定バージョンの指定はない
Kodak Photo CD	バージョン 1.0
PCD - PCX Bitmap	PC Paintbrush
Perfect Works (Draw)	Novell バージョン 2.0
PGM - Portable Graymap	特定バージョンの指定はない
PIC - Lotus 1-2-3 Picture File Format	特定バージョンの指定はない
PICT1 & PICT2 (ラスター)	Macintosh 標準

PIF - IBM 描画交換形式	バージョン 1.0
PNTG	MacPaint
PPM - Portable Pixmap	特定バージョンの指定はない
Progressive JPEG	特定バージョンの指定はない
PSP - Paintshop Pro (Win32 のみ)	バージョン 5.0、5.0.1
RND - AutoShade Rendering File Format	バージョン 2.0
SDW Ami Draw Snapshot (Lotus)	全バージョン
SRS - Sun Raster File Format	特定バージョンの指定はない
Targa	Truevision
TIFF	バージョン 6 まで
TIFF CCITT Group 3 & 4	Fax Systems
VISO (ページ・プレビュー・モードはバージョン 4 のみ) Visio 4、5、2000	Visio 4、5、2000
WMF	Windows Metafile
WordPerfect Graphics [WPG および WPG2]	バージョン 2.0 まで
XBM - X-Windows Bitmap	x10 互換
XPM - X-Windows Pixmap	x10 互換
XWD - X-Windows Dump	x10 互換

ハイエンド・グラフィックス・フォーマット

PSD - Adobe Photoshop File Format	バージョン 4.0
AI - Adobe Illustrator File Format	バージョン 7.0 まで
CDR - Corel Draw	バージョン 8.0 まで
DSF - Micrografx Designer	Windows 95、バージョン 6.0
DWG - AutoCAD Native Drawing Format	バージョン 12 から 14 まで
IGES - Initial Graphics Exchange Specification	バージョン 5.1
PDF - Portable Document Format	Acrobat バージョン 2.1、3.0、4.0 (日本語版 PDF を含む)
PS - Postscript	レベル 2

プレゼンテーション・フォーマット

Microsoft PowerPoint for Macintosh

バージョン 4.0、98

Corel Presentations

バージョン 8.0 および 9.0

Novell Presentations

バージョン 3.0 および 7.0

Harvard Graphics for DOS

バージョン 2.x および 3.x

Harvard Graphics

Windows バージョン

Freelance 96

Freelance 96

Freelance for Windows 95

SmartSuite 97 および Millennium

Freelance for Windows

バージョン 1.0 および 2.0

Freelance for OS/2

バージョン 2.0 まで

Microsoft PowerPoint for Windows

バージョン 7.0 まで

Microsoft PowerPoint 97

PowerPoint 97

Microsoft PowerPoint 2000

PowerPoint 2000

圧縮およびエンコードのフォーマット

ZIP PKWARE

バージョン 2.0g まで

GZIP

特定バージョンの指定はない

LZA Self Extracting Compress

特定バージョンの指定はない

LZH Compress

特定バージョンの指定はない

Microsoft Binder

バージョン 7.0、Binder 97

MIME (テキスト・メール)

特定バージョンの指定はない

UUEncode

特定バージョンの指定はない

UNIX[®] Compress

特定バージョンの指定はない

UNIX TAR

特定バージョンの指定はない

その他

vCard Electronic Business Card

バージョン 2.1

実行可能ファイル (EXE、DLL)

特定バージョンの指定はない

Windows NT 用実行可能ファイル

特定バージョンの指定はない

MSG (テキストのみ)

Microsoft Outlook メール・フォーマット

Microsoft Project (テキストのみ)

Project 98

第 10 章 権限の管理

この章では、EIP 権限の管理フィーチャーを紹介し、権限管理の概念について説明します。また、財産の保護に役立つマーク付け方法についても説明します。

知的財産の保護

デジタル形式で保管されているマルチメディア・オブジェクトは知的財産です。これらのオブジェクトを保護することは、特にこれらのオブジェクトが Web 上にある場合はコピーが比較的簡単なので、ビジネスにとって非常に重要です。Content Manager に組み込まれているマーク付けテクノロジーを使用すれば、マルチメディア・デジタル・オブジェクトに保護用のマーキングを付けたり、識別用の指紋を付けたりすることで、知的資産を権限のない使用から守ることができます。

次の場合に、貴重なオブジェクトにマークを付けることができます。

- 送信元を識別し、権限のないコピーや使用を防ぐ。これをウォーターマークと言います。通常は表示されます。
- コンテンツの宛先を識別し、権限のないコピーや使用を防ぐ。これを指紋と言います。通常は表示されません。
- 追加情報を入手するための連絡先を提供する。
- 付加価値流通チェーンで使用するために、日時などの情報を与える。

デジタル・オブジェクトは、顧客に送る前にマークを付けることができます。送信前に、ウォーターマークと指紋の両方を適用できます。ただし、送信前に指紋を適用する場合は、宛先が分かっていることが前提なので、送信プロセス時にマークを動的に適用しなければならない場合もあります。送信前に独自の制御下の環境からマークを適用すれば、改ざんの危険性が減るため、セキュリティーが向上します。

マークは、管理および送信プロセスのいくつかの段階で適用することができます。どの段階が適しているかは状況によって異なります。マークを適用できるプロセスの段階は次のとおりです。

- オブジェクトを保管する前
オブジェクトに共通マークを使う場合（たとえば、所有者を識別するための可視ウォーターマークなど）は、オブジェクトを保管する前または保管するときにマークを適用できます。Content Manager システムには、オリジナルのマークなしオブジェクトとマーク付きオブジェクトの両方を保管できます。また、マーク付きオブジェクトだけ保管して、マークなしオブジェクトは別のリポジトリに入れておくこともできます。
- オブジェクトを保管した後
Content Manager システムに保管したオブジェクトにマークを付ける場合は、オブジェクトを取り出し、マークを付け、マークなしオブジェクトとそのマーク付きオブジェクトを交換するか、あるいはマーク付きオブジェクトを新しい項目として保管することができます。

- オブジェクトを取り出すとき
適用するマークが宛先によって異なる場合は、オブジェクトを取り出した後でマークを動的に適用することができます。このようにすれば、オリジナル・オブジェクトの代わりにマーク付きオブジェクトを送ることができます。
システム内にマークが付いていない従来のオブジェクトが数多くあり、わざわざ戻って、各オブジェクトにマークを付ける時間もリソースも使いたくない場合は、オブジェクトを取り出すときにマークを付けることができます。

各種のマーク付け方法

コンテンツのマーク付け方法はいろいろあり、適したものを選ぶことができます。方法ごとに、対応する問題が異なり、除去や修正に対する耐久性も異なります。

マーク付けは、次のように分類されます。

- 情報の伝達

ウォーターマーク

コンテンツの送信元を識別します。オブジェクトの所有者やバージョンなどの情報を入れることができます。

指紋 コンテンツの宛先を識別します。オブジェクトの宛先についての情報を入れることができます。

- 可視性

可視 マークは表示されるので、認識できます。

非可視 マークはイメージの中に隠されます。

- 保全性

脆弱 マークは変更されると壊れます。

堅固 マークは、マーク付きオブジェクトに変更 (サイズ変更、圧縮、回転、クリッピングなど) が加えられても影響を受けません。

- マークを付けるとき

- オブジェクトを取り込むとき
- オブジェクトを保管するとき
- 顧客に配布するためにオブジェクトを取り出すとき
- オブジェクトが宛先のワークステーションに届いた時刻

- 場所

- 可視マークの目的が権限のない再使用を防ぐことである場合は、可視マークをイメージの大部分に適用することができます。さらにイメージのテキストチャ域に重なるようにマークを移動すれば、除去するのはもっと難しくなります。
- 非可視マークをイメージのテキストチャ域に利用すれば、イメージに影響をほとんど与えずにデータを埋め込むことができます。
- 可視マークの目的が所有者を示すことである場合は、邪魔にならないようにイメージの隅に配置することができます。
- 可視マーク、非可視マークの両方を使用する場合は、可視マークの方を先に適用してください。

- フォーマット

バイナリー

ビットのランダム・シーケンスをイメージ全体で繰り返して、マークとして使用することができます。このランダム・シーケンスが、イメージにマークを付けるまたはマークを外す場合に使うキーとなります。

また、イメージをマークとして使用することもできます。

構造化データ

組み込みテキスト・データをマークとして使用することができます。

可視マークの適用

可視マークは、マスクとイメージの両方が表示されるようにイメージの上に置く透明のマスクです。可視マークは除去するのが難しいため、オブジェクトの不正使用の防止に効果があります。

次の場合には可視マークを使用してください。

- 顧客が再表示できるイメージを提供するが、その再表示コピーを顧客が再使用することは認めない場合
- イメージを Web 上での宣伝に使用する場合

非可視マークの適用

非可視マークは、イメージが変更されていないことが分かるようにイメージの中に隠すデータです。マークを適用し、検出し、復号化するアプリケーションが必要です。

次の場合には非可視マークを使用してください。

- 所有者を識別するための情報を埋め込み、マーク付きオブジェクトの不正コピーを防止する場合 (ウォーターマーク)。
- 配布パスを追跡するための情報を埋め込む場合 (指紋)
- イメージに注釈やキャプションを埋め込む場合

第 11 章 アクセス支援機能

本製品には、障害のある人のアクセスを支援する機能が数多く含まれています。これらの機能としては、次のものがあります。

- マウスの代わりにキーボードを使用して、すべての機能を操作する機能
- 拡張表示プロパティのサポート
- 支援テクノロジーとの互換性
- オペレーティング・システムのアクセス支援機能との互換性
- アクセスを支援するドキュメンテーション・フォーマット

キーボード入力およびナビゲーション

キーボード入力とナビゲーションのために、次の機能が使用可能です。

キーボード入力

マウスの代わりにキーボードを使用して、この製品を操作できます。

メニュー項目およびコントロールはアクセス・キーを提供しており、これにより、直接キーボードからコントロールをアクティブにしたり、メニュー項目を選択したりすることができます。これらのキーは、セルフ・ドキュメンテーションとなっています。つまり、これらのアクセス・キーは、自らが表示されるコントロールやメニュー上で下線付きとなります。

キーボード・フォーカス

Windows ベースのシステムでは、キーボード・フォーカスの位置が強調表示されて、ウィンドウのアクティブなエリア、およびキー・ストロークが有効である場所を示します。

応答時間の調節

Windows ベースのシステムでは、「コントロール パネル」を使用して応答時間を調節できます。

アクセス支援表示機能

クライアントには、数多くの機能が用意されており、これらがユーザー・インターフェースを強化し、視覚障害のあるユーザーのためのアクセス支援を向上します。これらの機能強化としては、ハイコントラスト設定およびカスタマイズ可能なフォント・プロパティのサポートがあります。

ハイコントラスト・モード

クライアントは、オペレーティング・システムによって提供されるハイコントラスト・モード・オプションをサポートしています。この機能は、背景色と前景色との、より強いコントラストをサポートしています。

フォントの設定

Windows ベースのシステムでは、ディスプレイの設定を指定でき、これにより、メニューおよびダイアログ・ウィンドウ内のテキストの色、サイズ、

およびフォントを決定することができます。クライアントでは、文書リストのフォントを選択することができます。

カラーに依存しない

本製品の機能を使用する上で、カラーを識別する必要はありません。

支援テクノロジーとの互換性

クライアントは、Narrator や Via Voice などのスクリーン・リーダー・アプリケーションと互換性があります。クライアントには、これらのアクセス支援アプリケーションに必要なプロパティが用意されており、視覚に障害のあるユーザーがスクリーン内の情報を利用できるようにします。

アクセス支援のドキュメンテーション

本製品のドキュメンテーションは、PDF 形式で提供されています。Adobe のアドレス access.adobe.com から入手可能なフリー・ツールを使用して、PDF ファイルを HTML またはテキストに変換することができます。これにより、ユーザーは、自分のブラウザーに設定された表示設定に応じてドキュメンテーションを表示することができます。また、スクリーン・リーダーやその他の支援テクノロジーを使用することもできます。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものであり、本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒106-0032
東京都港区六本木 3-2-31
IBM World Trade Asia Corporation
Licensing

以下の保証は、国または地域の法律に沿わない場合は、適用されません。IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

本書は定期的に見直され、必要な変更 (例えば、技術的に不適切な記述や誤植など) は、本書の次版に組み込まれます。IBM は、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Corporation

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

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができませんが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性がありますが、その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者をお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

著作権使用許諾:

本書には、さまざまなオペレーティング・プラットフォームでのプログラミング手法を例示するサンプル・アプリケーション・プログラムがソース言語で掲載されています。お客様は、サンプル・プログラムが書かれているオペレーティング・プラットフォームのアプリケーション・プログラミング・インターフェースに準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。このサンプル・プログラムは、あらゆる条件下における完全なテストを経ていません。従って IBM は、これらのサンプル・プログラムについて信頼性、利便性もしくは機能性があることをほのめかしたり、保証することはできません。お客様は、IBM のアプリケーション・プログラミング・インターフェースに準拠したアプリケーション・プログラムの開発、使用、販売、配布を目的として、いかなる形式においても、IBM に対価を支払うことなくこれを複製し、改変し、配布することができます。

商標

以下は、IBM Corporation の商標です。

IBM	DisplayWrite	PowerPC
400	e-business	PTX
Advanced Peer-to-Peer Networking	HotMedia	QBIC
AIX	Hummingbird	RS/6000
AIXwindows	ImagePlus	SecureWay
APPN	IMS	SP
AS/400	Micro Channel	VideoCharger
C Set ++	MQSeries	Visual Warehouse
CICS	MVS/ESA	VisualAge
DATABASE 2	NetView	VisualInfo
DataJoiner	OS/2	WebSphere
DB2	OS/390	
DB2 Universal Database	PAL	

Approach、Domino、Lotus、Lotus 1-2-3、Lotus Notes、および SmartSuite は、Lotus Development Corporation の商標です。

Intel、および Pentium は、Intel Corporation の米国およびその他の国における商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

Java およびすべての Java 関連の商標およびロゴは、Sun Microsystems, Inc. の米国およびその他の国における商標または登録商標です。

UNIX は、The Open Group がライセンスしている米国およびその他の国における登録商標です。

他の会社名、製品名およびサービス名などは、それぞれ各社の商標または登録商標です。

用語集

この用語集では、このシステムに固有の用語および略語を定義します。イタリック体の用語は、この用語集の他の場所で定義されています。

[ア行]

アクション・リスト (action list). ユーザーがワークフロー または文書経路指定プロセス内での実行を承認されているアクションのリスト。システム管理者またはその他のワークフロー・コーディネーター が、これらのアクションを定義する。

アクセス制御リスト (access control list). 1 つまたは複数のユーザー ID やユーザー・グループ、およびそれらに関連する特権 で構成されるリスト。アクセス制御リストを使用すると、Enterprise Information Portal システム内の検索テンプレート へのユーザー・アクセスを制御できる。

アクセス制御(access control). 許可ユーザーが、許可された方法でアクセスする場合に限り、特定の機能および保管オブジェクト にアクセスできるようにするプロセス。

アプリケーション・プログラミング・インターフェース (application programming interface (API)). アプリケーション相互の通信を実行できるようにするソフトウェア・インターフェース。API はプログラム言語の構造体とステートメントの集合で、アプリケーション・プログラム中にコード化することにより、基礎を成すライセンス・プログラムに備えられている特定の機能やサービスを利用できる。

イテレーター (iterator). オブジェクトのコレクションを、一度にステップスルーするために使用するクラスまたは構成。

イメージ・オブジェクト・コンテンツ・アーキテクチャー (Image Object Content Architecture (IOCA)). イメージの交換と表示に使われる構成の集合。

イメージ・コンテンツによる照会 (query by image content (QBIC)). プレーン・テキストではなく、フィーチャーというビジュアル・コンテンツに基づく検索を可能にする照会のテクノロジー。QBIC を使用すると、色、テキストチャーなど、視覚特性に基づいてオブジェクトを検索できる。

永続 ID (persistent identifier (PID)). 保管場所に関係なくオブジェクト を一意的に識別する ID。PID は項目 ID と位置の両方から成る。

延期 (suspend). オブジェクト をワークフロー から取り除き、それを活動化させるのに必要な延期基準を定義すること。そのオブジェクトを後で活動化させると、処理を再開する。

オーバーレイ (overlay). 印刷時にページ上の変数データと組み合わせることができる、事前定義されたデータの集まり (線、陰影、テキスト、枠、ロゴなど)。

オブジェクト (object). ユーザーが単一のユニットとして保管、検索、および操作できるすべてのデジタル・コンテンツ。たとえば、JPEG イメージ、MP3 オーディオ、AVI ビデオ、およびブックからのテキスト・ブロックなど。

オブジェクトのリンクと組み込み (Object Linking and Embedding (OLE)). アプリケーションをリンクしたり埋め込んだりすることで、他のアプリケーションからそのアプリケーションを呼び出せるようにする、Microsoft® 社による仕様。

オブジェクト・サーバー(object server). 「リソース・マネージャー (resource manager)」を参照。

オブジェクト・サーバー・キャッシュ (object server cache). 「リソース・マネージャー・キャッシュ (resource manager cache)」を参照。

[カ行]

カーソル (cursor). アプリケーション・プログラムで、順序付けられた行の集まりの中の特定の行を指すのに使用される名前付きの制御構造。カーソルは、集まりから行を取り出すのに使用される。

解放 (release). 延期基準を項目 から除去すること。延期された項目が解放されるのは、基準が満たされた場合か、または適切な権限を持つユーザーが基準をオーバーライドして手動で解放する場合である。

拡張データ・オブジェクト (extended data object (XDO)). アプリケーション・プログラムで、保管されている複雑なマルチメディア・オブジェクト がストレ

ージから出し入れされる場合に、そのオブジェクトを総称した表現。XDO は DDO に含まれることが最も多い。

キー・フィールド (key field). 「属性 (attribute)」を参照。

基数 (cardinality). データベース・テーブルの行数。

キャッシュ (cache). 特定の目的を持つバッファで、主記憶域よりも小さく、高速。頻繁にアクセスされるデータのコピーを保持するために使用される。キャッシュを使用することにより、アクセス時間が短縮されるが、メモリー所要量が増える場合がある。

クライアント / サーバー (client/server). 通信で、あるサイトのプログラムが別のサイトのプログラムへ要求を送信して応答を待つ、分散データ処理における対話のモデル。要求側プログラムはクライアントと呼ばれ、応答側プログラムはサーバーと呼ばれる。

クライアント・アプリケーション (client application). Enterprise Information Portal からコンテンツ・サーバーにアクセスするための、オブジェクト指向またはインターネット API で作成されたアプリケーション。

クラス (class). オブジェクト指向の設計やプログラミングで、共通の定義でオブジェクトを作成するようインスタンス化できるモデルまたはテンプレート。したがって、プロパティ、操作、および動作も共通になる。オブジェクトはクラスのインスタンスに当たる。

結合サーチ (combined search). パラメトリック、テキスト、またはイメージのいずれかのタイプの検索 (サーチ) を 1 つまたは複数組み合わせ合わせた照会。

検索基準 (search criteria). Enterprise Information Portal で、管理者が定義する検索テンプレートの特定のフィールドのこと。ユーザーが使用できる選択項目を制限したり定義を詳細化したりする。

検索テンプレート (search template). 管理者が設計した検索基準から成る、特定のタイプの統合検索のフォーム。管理者は、個々の検索テンプレートにアクセスできるユーザーとユーザー・グループも指定する。

交換 (interchange). イメージとその索引を 1 つの Content Manager ImagePlus for OS/390 システムから別の ImagePlus システムへ、共通交換ファイルや共通交換単位を利用してインポートまたはエクスポートする機能。

項目 (item). Enterprise Information Portal が管理する最小の情報単位を総称した用語。個々の項目には ID がある。項目の例としては、フォルダー または文書 がある。

項目タイプ (item type). 同種の項目を定義し、後で位置決めを行うためのテンプレート。ルート・コンポーネント、ゼロまたはそれ以上の子コンポーネント、および種別で構成される。

項目タイプ種別 (item type classification). 項目タイプ内でのカテゴリー化。その項目タイプの項目をより詳しく特定する。同じ項目タイプの項目はすべて、同じ項目タイプ種別を持っている。

Content Manager が提供する項目タイプ種別には、フォルダー、文書、オブジェクト、ビデオ、イメージ、およびテキストがある。ユーザーが、独自の項目タイプ種別を定義することもできる。

子コンポーネント (child component). 階層項目タイプのオプションの 2 次または下位レベル。それぞれの子コンポーネントは、その上のレベルに直接関連付けられている。

コネクター・クラス (connector class). 特定のコンテンツ・サーバー固有の API への標準アクセスを提供するオブジェクト指向のプログラミング・クラス。

コモン・ゲートウェイ・インターフェース (Common Gateway Interface (CGI)). Web サーバーとその外部のプログラムとの間で情報を交換するための規格。外部プログラムを作成する際のプログラム言語は、Web サーバーを実行しているオペレーティング・システムでサポートされているものであればどれでもよい。「CGI スクリプト (CGI script)」を参照。

固有エンティティ (native entity). 特定のコンテンツ・サーバー上で管理され、固有属性から成るオブジェクト。たとえば、Content Manager 索引クラスは、Content Manager キー・フィールドから成る固有エンティティである。

固有属性 (native attribute). 特定のコンテンツ・サーバー上で管理され、そのコンテンツ・サーバーにとって固有のオブジェクトの特性。たとえば、キー・フィールド policy num は Content Manager コンテンツ・サーバーの固有属性で、フィールド policy ID は Content Manager OnDemand コンテンツ・サーバーの固有属性である。

固有テキスト索引 (native text index). 特定のコンテンツ・サーバー上で管理されるテキスト項目の索引。たとえば、Content Manager コンテンツ・サーバーにある単一のテキスト検索索引など。

コレクション (collection). 類似した一管理規則のセットを持つオブジェクトのグループ。

コンストラクター (constructor). プログラム言語では、クラスと同じ名前のメソッドのことで、そのクラスのオブジェクトの作成と初期化に使用される。

コンテンツ・サーバー (content server). マルチメディア・データや業務データ、またそのデータの処理に必要な関連メタデータを保管するためのソフトウェア・システム。コンテンツ・サーバーの例としては、Content Manager や Content Manager ImagePlus for OS/390 がある。

コンポーネント (component). ルート・コンポーネント または子コンポーネント を総称した用語。

[サ行]

サーバー定義 (server definition). Enterprise Information Portal で一意的に識別できるようにするための、特定のコンテンツ・サーバー の特性。

サーバー・インベントリー (server inventory). 指定されたコンテンツ・サーバー の固有エンティティ および固有属性 の包括的なリスト。

サーバー・タイプ定義 (server type definition). 特定のタイプのカスタム・サーバーが、Enterprise Information Portal で一意的に識別されるように、管理者が指定した特性のリスト。

作業項目 (work item). 旧バージョンの Content Manager ワークフローおよび Enterprise Information Portal 拡張ワークフローにおいて、ワークフロー 内でアクティブになっている作業アクティビティ。

作業状態 (work state). 個々の作業項目、文書、またはフォルダー の状況。

作業ステップ (work step). 個々の作業項目、文書、またはフォルダー が通過しなければならないワークフロー または文書ルーティング・プロセス の離散ポイント。

索引 (index). 特定の項目 またはオブジェクト を識別する属性値を追加または編集し、後で検索できるようにすること。

索引クラス (index class). 「項目タイプ (item type)」を参照。

索引クラス・サブセット (index class subset). 以前の Content Manager では索引クラスのビューで、アプ

リケーションでこれを使用してフォルダーやオブジェクトの保管、取り出し、表示が行われる。

索引クラス・ビュー (index class view). 初期の Content Manager では、索引クラス・サブセット の API で使用された用語。

サブクラス (subclass). 別のクラスから派生するクラス。クラスとサブクラスの間には 1 つ以上のクラスがある場合がある。

参照 (reference). ルートまたは子コンポーネント と別のルート・コンポーネント 間における、単一方向で 1 対 1 の関連。「リンク (link)」と対比。

照会ストリング (query string). 照会のプロパティおよびプロパティ値を指定するための文字ストリング。アプリケーションの中で照会ストリングを作成して、照会に渡すことができる。

情報マイニング (information mining). テキストからの重要な情報の抜粋 (要約)、文書の集合からの主要なテーマの検出 (カテゴリー化)、および強力で柔軟性のある照会を使用して関連文書の検索を実行するための自動化されたプロセス。

シン・クライアント (thin client). ソフトウェアがほとんど、またはまったくインストールされていないが、接続先のネットワーク・サーバーで管理され送達されるソフトウェアに対するアクセス権はあるクライアント。シン・クライアントは、ワークステーションなどの全機能クライアントの代替になる。

スーパークラス (superclass). クラスの派生元のクラス。クラスとスーパークラスの間には 1 つ以上のクラスがある場合がある。

ステージング (staging). 保管オブジェクト を、オフラインまたは優先順位の低いデバイスから、オンラインまたは優先順位の高いデバイスに戻すためのプロセス。普通はシステムまたはユーザーの要求に基づいて行われる。ユーザーが永続記憶装置に保管されているオブジェクトを要求すると、作業用コピーがステージング域 に書き込まれる。

ステージング域 (staging area). リソース・マネージャー の作業用ストレージ域。リソース・マネージャー・キャッシュ ともいう。

ストリーム化データ (streamed data). ネットワーク接続を介して一定の速度で送信されるすべてのデータ。ストリームのデータ型は、1 つでも、あるいは型を組み合わせたものでも可能。データ転送率 (ビット / 秒で示される) は、ストリームやネットワークの種類によって異なる。

セマンティック・タイプ (semantic type). 項目の使用法または規則。基本、注釈、および注は、Content Manager が提供するセマンティック・タイプである。ユーザーが、独自のセマンティック・タイプを定義することもできる。

属性 (attribute). 項目のある特定の特性またはプロパティ (たとえば、名前、住所、年齢など) を記述するデータの群。項目を見つけるために使用できる。属性には、その属性が保管している情報の範囲を示す型と、その範囲内の値がある。たとえば、タイトル、実行時間、またはエンコード・タイプ (MPEG1、H.263 など) など、マルチメディア・ファイル・システム内のファイルについての情報が考えられる。Enterprise Information Portal の場合は、「**統合属性 (federated attribute)**」および「**固有属性 (native attribute)**」も参照。

[タ行]

抽象クラス (abstract class). 概念を表すオブジェクト指向プログラミングのクラス。抽象クラスから派生したクラスは、概念のインプリメンテーションを表す。抽象クラスのオブジェクトは構成できない。すなわち、インスタンス化することはできない。

データ・ストア (datastore). (1) データを保管する場所 (データベース・システム、ファイル、ディレクトリーなど) を示す総称用語。(2) アプリケーション・プログラムで、コンテンツ・サーバー のこと。

データ・フォーマット (data format). 「**MIME タイプ (MIME type)**」を参照。

統合エンティティ (federated entity). 統合属性 から成る Enterprise Information Portal メタデータ・オブジェクト。統合エンティティは、1 つ以上の統合テキスト索引 と関連付けることもできる。

統合検索 (federated search). 1 つ以上のコンテンツ・サーバー (異機種混合可) でデータを同時に検索する、Enterprise Information Portal から発行される照会。

統合コレクション (federated collection). 統合検索の結果であるオブジェクトのグループ。

統合属性 (federated attribute). 1 つ以上のコンテンツ・サーバー で固有属性 にマップされる Enterprise Information Portal メタデータ・カテゴリー。たとえば、統合属性 policy number は、Content Manager では属性 policy num に、Content Manager ImagePlus for OS/390 では属性 policy ID にマップされる。

統合データ・ストア (federated datastore). 任意の数の特定のコンテンツ・サーバー (Content Manager など) のこと。

統合テキスト索引 (federated text index). 1 つ以上のコンテンツ・サーバー で 1 つ以上の固有テキスト索引 にマップされる Enterprise Information Portal メタデータ・オブジェクト。

動的データ・オブジェクト (dynamic data object (DDO)). アプリケーション・プログラムで、ストレージから保管オブジェクトを出し入れする際に、その保管オブジェクトを表す総称表現。

特権 (privilege). 特定のオブジェクト に特定の方法でアクセスする権利。特権には、システムに保管されているオブジェクトの作成、削除、および選択を行う権利が含まれる。特権は、管理者が割り当てる。

特権セット (privilege set). システムのコンポーネントや機能を使用するための特権 の集まり。管理者は特権セットをユーザー (ユーザー ID) およびユーザー・グループ に割り当てる。

[ナ行]

ネットワーク・テーブル・ファイル (network table file). Content Manager システムのそれぞれのノードのシステム固有の構成情報が含まれているテキスト・ファイル。システムのそれぞれのノードには、ノードを識別し、接続する必要のあるノードのリストを表示するネットワーク・テーブル・ファイルがある。

ネットワーク・テーブルの名前は FRNOLINT.TBL。

[ハ行]

パーツ (part). 「オブジェクト (object)」を参照。

バイナリー・ラージ・オブジェクト (binary large object (BLOB)). サイズが 0 バイト ~ 2 ギガバイトの、バイトのシーケンス。このストリングには関連したコード・ページや文字セットはない。イメージ、オーディオ、およびビデオの各オブジェクトは BLOB の形で保管される。

ハイパーテキスト・マークアップ言語 (Hypertext Markup Language (HTML)). SGML 規格に準拠したマークアップ言語で、主にハイパーテキスト・リンクを含むテキスト情報とグラフィック情報のオンライン表示をサポートするために設計された。

パッケージ (package). アクセス保護およびネーム・スペース管理を提供する、関連するクラス およびインターフェースの集合。

パラメトリック・サーチ (parametric search). オブジェクトのプロパティに基づき、オブジェクトの照会。

ハンドル (handle). オブジェクトを表す文字ストリング。オブジェクトを検索する場合に使用する。

ヒストリー・ログ (history log). ワークフローのアクティビティを保持するファイル。

ファイル・システム (file system). AIX で、ハード・ディスクのパーティションを区切ってストレージを作る方法。

フィーチャー (feature). イメージ検索サーバーに格納されたビジュアル・コンテンツ情報。さらに、イメージ検索アプリケーションが突き合わせに使用する可視特性。QBIC フィーチャーには、平均色、ヒストグラム色、定位置色、およびテクスチャーの 4 種類がある。

フォルダー (folder). 種別に関係なく、フォルダー・セマンティック・タイプをもつ、任意の項目タイプの項目。フォルダー・セマンティック・タイプの項目には、すべての非リソース項目機能、および文書 またはリソース・タイプなどの項目タイプ種別から使用可能な追加の機能に加えて、Content Manager によって提供される特定のフォルダー機能が含まれます。フォルダーには、文書およびサブフォルダーをはじめ、任意のタイプの項目をいくつでも含めることができます。フォルダーは、属性別に索引が付けられます。

フォルダー・マネージャー (folder manager). データをオンライン文書とフォルダーとして管理する Content Manager モデル。フォルダー・マネージャー API は、アプリケーションと Content Manager コンテンツ・サーバーとの間の基本インターフェースとして使用できる。

プロパティ (property). オブジェクトについて記述する、オブジェクトの特性。プロパティは変更や修正が可能。プロパティの例としては、たとえば、書体などがある。

文書 (document). 単独のユニットとして、Content Manager システムとユーザーとの間で保管、検索、および交換可能な項目。文書セマンティック・タイプの項目は、文書を形成する情報を含んでいると予想されるが、必ずしも Content Manager 文書モデルのインプリメンテーションであることを意味しない。

文書分類項目タイプから作成された項目 (Content Manager 文書モデルの特定のインプリメンテーション)

は、文書パートを含んでいなければならない。文書分類項目タイプを使用して、文書セマンティック・タイプまたはフォルダー・セマンティック・タイプのどちらかの項目を作成できる。

文書パートには、テキスト、イメージ、およびスプレッドシートなど、異なるタイプのコンテンツを含めることができる。

文書型定義 (document type definition (DTD)). XML 文書の特定のクラスの構造を指定する規則。DTD は、エレメント、属性、および表記法で構造を定義し、文書の特定のクラス内でのそれぞれのエレメント、属性、および表記法の使用方法について制約を設定する。DTD は、DTD が特定のマークアップ言語の構造を完全に記述しているという点において、データベース・スキーマに類似している。

ボリューム (volume). システムのオブジェクトが保管されている実際の物理ストレージ・デバイスまたは物理記憶装置を表す。

[マ行]

マルチメディア (multimedia). コンピューターから表示および制御するための、異なるメディア・エレメント (テキスト、グラフィックス、オーディオ、静止画像、ビデオ、アニメーション) の組み合わせ。

マルチメディア・ファイル・システム (multimedia file system). ビデオおよびオーディオのストレージおよびデリバリーに最適化されたファイル・システム。

メソッド (method). Java の設計とプログラミングで、操作によって指定される動作をインプリメントするソフトウェア。C++ のメンバー関数と同義。

メディア・アーカイバー (media archiver). オーディオおよびビデオのストリーム・データの保管に使用する物理デバイス。VideoCharger はメディア・アーカイバーの一種である。

メディア・サーバー (media server). Content Manager システムの AIX ベースのコンポーネント。ビデオ・ファイルを保管し、これらのファイルにアクセスするときに使う。

[ヤ行]

ユーザー (user). Enterprise Information Portal で、Enterprise Information Portal 管理プログラムによって識別される人物。

ユーザー出口 (user exit). IBM 提供のプログラムのうち、ユーザー出口ルーチンが制御を受け取ることができる点。

ユーザー出口ルーチン (user exit routine). 事前定義のユーザー出口で制御を受け取るユーザー作成ルーチン。

ユーザー・グループ (user group). 1 つ以上の定義済み個別ユーザーから構成されるグループ。単一グループ名によって識別される。

ユーザー・マッピング (user mapping). Enterprise Information Portal のユーザー ID とパスワードを、1 つ以上のコンテンツ・サーバー中の対応するユーザー ID とパスワードに関連付けること。ユーザー・マッピングにより、Enterprise Information Portal と複数のコンテンツ・サーバーにシングル・ログオンできる。

[ラ行]

ライブラリー・クライアント (library client). ライブラリー・システムに低レベルのプログラミング・インターフェースを提供する Content Manager システムのコンポーネント。ライブラリー・クライアントには、ソフトウェア開発者キットの一部を成す API が含まれている。

ライブラリー・サーバー (library server). 項目の照会を保管、管理、および処理する Content Manager システムのコンポーネント。

ランク (rank). 指定されたパートと照会結果の関係を示す整数値。ランクが高いほど一致の度合いが高い。

リソース・マネージャー. オブジェクトを管理する Content Manager システムのコンポーネント。これらのオブジェクトは、ライブラリー・サーバー上に保管された項目によって参照される。

リソース・マネージャー・キャッシュ (resource manager cache). リソース・マネージャーの作業用ストレージ域。「ステー징域 (staging area)」ともいう。

リモート・メソッド呼び出し (Remote Method Invocation (RMI)). 分散プログラミングを可能にする API の集合。ある Java 仮想マシン (JVM) 中のオブジェクトが、別の JVM 中のオブジェクトのメソッドを呼び出せる。

リンク (link). ソースとターゲットという 2 つの項目間の指示関係。一連のリンクを使用して 1 対多の関連を作ることができる。「参照 (reference)」と対比。

ルート・コンポーネント (root component). 関連するシステム定義およびユーザー定義の属性から成る、最初の、または唯一のレベルの階層項目タイプ。

レンダリング (render). 通常であればイメージ指向ではないデータを、イメージとして描画または表示すること。Content Manager では、ワープロ文書を表示目的でイメージとしてレンダリングできる。

ローカル・エリア・ネットワーク (local area network (LAN)). 一連の装置が通信のために互いに接続されたネットワーク。より大規模なネットワークに接続することができる。

[ワ行]

ワークフロー (workflow). Enterprise Information Portal において、作業パケット、文書、またはフォルダーが処理の過程で通過する一連の作業ステップ、およびこれらのステップを管理する規則のこと。

たとえば、claims approval の場合は、承認を得るために個々の保険金支払請求がたどらなければならないプロセスを表します。

ワークフローの状態 (workflow state). ワークフロー全体の状況。

ワーク・パケット (work packet). Enterprise Information Portal バージョン 7.1 における、ある場所から別の場所に経路指定される文書の集まり。ユーザーはワーク・リストを介してワーク・パケットにアクセスし、処理する。

ワーク・リスト (worklist). ユーザーに割り当てられた作業項目、文書、またはフォルダーの集まり。

ワイルドカード文字 (wildcard character). 1 つまたは複数の文字を示す場合に使用可能な、アスタリスク (*) または疑問符 (?) などの特殊文字。任意の文字 (単数) または一連の文字 (複数) を、ワイルドカード文字に置き換えることができる。

A

ADSM. 「Tivoli® Storage Manager」を参照。

API. 「アプリケーション・プログラミング・インターフェース (application programming interface)」を参照。

AVI. 「Audio/Video Interleaved」を参照。

AVI (Audio Video Interleaved). オーディオおよびビデオ・データをファイル内でインターリーブできるようにする RIFF (Resource Interchange File Format) ファイ

ル仕様。ファイル装置上での順次アクセスを維持したまま、再生または記録用の代替チャンク内で、別々のトラックにアクセスできる。

B

BLOB. 「バイナリー・ラージ・オブジェクト (*binary large object*)」を参照。

C

CGI. 「コモン・ゲートウェイ・インターフェース (*Common Gateway Interface*)」を参照。

CGI スクリプト (CGI script). Web サーバー上で稼働するコンピューター・プログラムで、コモン・ゲートウェイ・インターフェースを使用して、普通は Web サーバーでは行われないタスク (データベース・アクセスやフォーム処理など) を実行する。CGI スクリプトは、Perl などのスクリプト記述言語で書かれた CGI プログラムである。

CIF. 「*Common Interchange File (CIF)*」を参照。

CIU. 「*Common Interchange Unit (CIU)*」を参照。

Common Interchange File (CIF). ImagePlus Interchange Architecture (IPIA) データ・ストリームを 1 つ含んでいるファイル。

Common Interchange Unit (CIU). CIF で使われる、独立した転送単位。これは CIF の一部であり、受信データベースとの関係を示している。1 つの CIF に CIU が複数含まれることもある。

D

DDO. 「動的データ・オブジェクト (*dynamic data object*)」を参照。

DTD. 「文書型定義 (*document type definition*)」を参照。

H

HTML. 「ハイパーテキスト・マークアップ言語 (*Hypertext Markup Language*)」を参照。

I

IOCA. 「イメージ・オブジェクト・コンテンツ・アーキテクチャー (*Image Object Content Architecture*)」を参照。

J

JavaBeans. 「beans」と呼ばれる再使用可能な Java コンポーネントを作成するための、プラットフォームに依存しないソフトウェア・コンポーネント・テクノロジー。Bean を構築すると、他のソフトウェア・エンジニアがそれを使用したり、Java アプリケーションで使用したりできる。JavaBeans を使用すると、ソフトウェア・エンジニアはグラフィカルなドラッグ・アンド・ドロップ開発環境で bean の操作やアセンブルを行える。

Joint Photographic Experts Group (JPEG). (1) デジタル化された連続トーン・イメージの規格を制定したグループ。(2) このグループによって開発された静止画像の規格。

JPEG. 「*Joint Photographic Experts Group*」を参照。

L

LAN. 「ローカル・エリア・ネットワーク (*local area network*)」を参照。

M

MIME タイプ (MIME type). インターネットを通して転送されるオブジェクトのタイプを定義するインターネット規格。MIME タイプには、オーディオ、イメージ、およびビデオのいくつかのタイプが含まれます。各オブジェクトは、MIME タイプを持つ。

Multipurpose Internet Mail Extensions (MIME). 「*MIME タイプ (MIME type)*」を参照。

O

OLE. 「オブジェクトのリンクと組み込み (*Object Linking and Embedding*)」を参照。

P

PID. 「永続 ID (*persistent identifier*)」を参照。

Q

QBIC. 「イメージ・コンテンツによる照会 (*query by image content*)」を参照。

R

README ファイル (README file). このファイルに関連したプログラムのインストールまたは実行前に、読

む必要のあるファイル。README ファイルには、通常、最新の製品情報、インストール情報、または製品を使用する上でのヒントが含まれている。

Resource Interchange File Format (RIFF). 異なるタイプのコンピューター機器で再生するための音声およびグラフィックスの保管に使用される。

RIFF. 「*Resource Interchange File Format*」を参照。

RMI サーバー (RMI server). Java リモート・メソッド呼び出し (RMI) 分散オブジェクト・モデルをインプリメントしたサーバー。

T

Tivoli StorageManager (TSM). 異機種混合環境でのストレージ管理とデータ・アクセス・サービスを備えたクライアント / サーバー 製品。さまざまな通信方式をサポートし、ファイルのバックアップとストレージを管理する機能を備え、バックアップ操作のスケジューリング機能を備えている。

TSM. 「*Tivoli Storage Manager*」を参照。

TSM ボリューム (TSM volume). *Tivoli Storage Manager* によって管理される、ストレージの論理領域。

U

uniform resource locator (URL). コンピューター上や、インターネットなどのネットワーク内にある情報リソースを示す文字シーケンス。この文字シーケンスには、情報リソースにアクセスするために使用するプロトコルの省略名、およびプロトコルが情報リソースを見つけるために使用する情報が含まれている。たとえば、インターネットのコンテキストでは、http、ftp、gopher、telnet、および news などは、さまざまな情報リソースにアクセスするために使用されるプロトコルの省略名である。

X

XDO. 「拡張データ・オブジェクト (*extended data object*)」を参照。

XML. 「*Extensible Markup Language*」を参照。

XML (Extensible Markup Language). SGML から派生し、そのサブセットでもある標準のメタ言語マークアップ言語を定義する。XML は、SGML の比較的複雑であまり使用されない部分を省略し、文書タイプを処理するアプリケーションの作成、構造化された情報の作成

および管理、異なるコンピューター・システム間での構造化された情報の伝送および共用をより容易にする。XML を使用する場合、SGML に必要な堅固なアプリケーションや処理は必要としない。XML は、W3C (World Wide Web Consortium) の主導の下に開発されている。

索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

[ア行]

アクションの定義 115
アクション・リスト
事前定義アクション 115
定義 112, 115
アクセス支援 145
アクセス制御リスト
ドメイン間の移動 45
イメージ検索
構成 124
接続の検査 126
セットアップ 124
別名の割り当て 126
イメージ検索オプション 6
イメージ検索の構成 124

[カ行]

開始
ワークフロー・ビルダー 115
カタログ
削除 70
追加 70
トレーニング 80
トレーニング文書の追加 72
名前変更 70
評価 76
管理可能ドメイン 39
管理クライアント
作成
検索基準 30
検索テンプレート 29
ワークフロー 115
定義
アクション 115
アクション・リスト 115
ワーク・リスト 114
キーボード 145
計画
Enterprise Information Portal 2
検索基準
定義とマッピング 30
検索テンプレートの作成 29
コネクタ 5

コレクション
ドメイン間の移動 45
割り当て、ドメインへの 43
コンテンツ・サーバー
定義 19
コンテンツ・ビューアー・オプション 6

[サ行]

サーバー・インベントリ 19
作業バケットの説明 111
サンプルのローダー・プログラムの実行 130
情報マイニング
コンポーネント 48
サービス 47
サポートされる言語 55
サポートされる文書フォーマット 55
説明 47
ターゲット・グループ 50
ビジネス環境での動作 50
分類法の作成 68
例 52
身体障害 145

[タ行]

テキスト検索
セットアップ 123
XML サポート 123
特権グループ 38
特権セット 33, 38
作成 38
ドメイン間の移動 45
ドメイン 44
作成 40
上級管理者特権 41, 42
副管理者特権 41, 42
理解 40

[ナ行]

認可特権セット 39

[ハ行]

分類法
Information Structuring Tool の使用 68

[マ行]

メタデータ・ストア
情報マイニングの使用 47

[ヤ行]

ユーザー 33
特権セット 39
ドメイン間の移動 43
ユーザー ID 33
ユーザー・グループ 39
ドメイン間の移動 44

[ラ行]

リソース・マネージャー
割り当て、ドメインへの 43
割り当て、ユーザーの 39
リソース・マネージャー、ドメイン間の移動 44
ロード、サンプル・データの 127
ロード、テキストおよびイメージ検索文書の 130

[ワ行]

ワークフロー
概念 107
計画 111
作成 115
ワークフロー・ビルダー
開始 115
説明 113
ワークフローの作成 115
ワークフロー・フィーチャー
構成 107
コンポーネント 113
ワーク・リスト
説明 111
定義 114

C

cmbcc2mime.ini 18

E

EIP
イメージ検索クライアント 6

EIP (続き)

- 管理コンポーネント 4
- コネクタ 5
- コネクタ・ツールキット 6
- コンテンツ・ビューアー・クライアント 6
- 情報マイニング・オプション 5
- テキスト検索クライアント 6
- Information Center コンポーネント 7
- Web Crawler オプション 6

EIP コンポーネント

- イメージ検索 6
- オペレーティング・システムの互換性 3, 4
- 管理 4
- コネクタ 5
- コンテンツ・ビューアー 6
- 情報マイニング 5
- テキスト検索 6
- Information Center 7
- Web Crawler 6

Enterprise Information Portal

- 作成
 - 検索基準 30
 - 検索テンプレート 29
 - ワークフロー 115
- 定義
 - アクション 115
 - アクション・リスト 115
 - ワーク・リスト 114

I

IBM Enterprise Information Portal for Multiplatforms

- コンポーネント 3

Information Structuring Tool

- インストール 69
- 説明 68
- トレーニング文書の選択 72
- 始めに 69
- 分類法の定義 70
- 分類法のトレーニング 80
- 分類法の評価 76
- ロック機構 69
- WAS の使用 69

L

LDAP

- インポート 36
- 構成 36

M

- MIME タイプのカスタマイズ 18
- MIME タイプ・ファイル
サーバーでの変更 18

W

- Web Crawler
EIP オプション 6



プログラム番号: 5724-B43

Printed in Japan

SC88-9204-01



日本アイ・ビー・エム株式会社
〒106-8711 東京都港区六本木3-2-12