

**IBM WebSphere Information Integrator  
OmniFind Edition**



## **エンタープライズ・サーチの管理**

バージョン 8.3



**IBM WebSphere Information Integrator  
OmniFind Edition**



## **エンタープライズ・サーチの管理**

バージョン 8.3

本書および本書で紹介する製品をご使用になる前に、『特記事項』に記載されている情報をお読みください。

本書には、IBM の専有情報が含まれています。その情報は、使用許諾条件に基づき提供され、著作権により保護されています。本書に記載される情報には、いかなる製品の保証も含まれていません。また、本書で提供されるいかなる記述も、製品保証として解釈すべきではありません。

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： SC18-9283-02  
IBM WebSphere Information Integrator  
OmniFind Edition  
Administering Enterprise Search  
Version 8.3

発 行： 日本アイ・ピー・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 2005.12

この文書では、平成明朝体™W3、平成明朝体™W7、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™W7、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 2004, 2005. All rights reserved.

© Copyright IBM Japan 2005

# 目次

本書のトピックについて . . . . .	vii	Windows での WebSphere II Content Edition 用ク ローラー・サーバーの構成 . . . . .	46
本書の対象読者 . . . . .	vii	DB2 クローラー . . . . .	48
<b>エンタープライズ・サーチとは? . . . . .</b>	<b>1</b>	DB2 クローラーのための WebSphere Information Integrator Event Publisher Edition の構成 . . . . .	50
エンタープライズ・サーチでサポートされているデー タ・ソース・タイプ . . . . .	2	DB2 クローラーのための WebSphere MQ の構成 . . . . .	52
エンタープライズ・サーチ・コンポーネントの概要 . . . . .	3	イベント・パブリッシングを使用するための UNIX でのクローラー・サーバーの構成 . . . . .	54
エンタープライズ・サーチ・クローラー . . . . .	4	イベント・パブリッシングを使用するための Windows でのクローラー・サーバーの構成 . . . . .	56
エンタープライズ・サーチ・パーサー . . . . .	5	DB2 Content Manager クローラー . . . . .	57
エンタープライズ・サーチ索引 . . . . .	7	UNIX での DB2 Content Manager 用クローラ ー・サーバーの構成 . . . . .	58
エンタープライズ・サーチ用検索サーバー . . . . .	8	Windows での DB2 Content Manager 用クローラ ー・サーバーの構成 . . . . .	60
エンタープライズ・サーチ管理コンソール . . . . .	9	Domino Document Manager クローラー . . . . .	61
エンタープライズ・サーチ・システムのモニター . . . . .	10	Exchange Server クローラー . . . . .	63
エンタープライズ・サーチ用ログ・ファイル . . . . .	10	セキュア Exchange Server 文書へのアクセスの検 証 . . . . .	64
エンタープライズ・サーチのカスタマイズ . . . . .	11	Nntp クローラー . . . . .	64
エンタープライズ・サーチ用サンプル検索アプリ ケーション . . . . .	12	Notes クローラー . . . . .	65
エンタープライズ・サーチのデータ・フロー . . . . .	13	Lotus Domino データベースのクローラーに関する ヒント . . . . .	67
<b>エンタープライズ・サーチ管理の概要 . . . . .</b>	<b>15</b>	Lotus Domino ソースをクローラーするための UNIX でのクローラー・サーバーの構成 . . . . .	68
管理コンソールへのログイン . . . . .	19	Lotus Domino ソースをクローラーするための Windows でのクローラー・サーバーの構成 . . . . .	70
シングル・サーバー構成のエンタープライズ・サー チ管理者パスワードの変更 . . . . .	20	DIIOp プロトコルを使用するサーバーの構成 . . . . .	72
マルチサーバー構成のエンタープライズ・サー チ管理者パスワードの変更 . . . . .	22	Lotus Domino ソースをクローラーするための AIX での I/O 完了ポートの構成 . . . . .	73
エンタープライズ・サーチ・システムのポート番号 の変更 . . . . .	26	QuickPlace クローラー . . . . .	75
<b>エンタープライズ・サーチ・コレクション 29</b>	<b>29</b>	ローカル・ユーザー・セキュリティーを使用する ための QuickPlace サーバーの構成 . . . . .	77
コレクション・ウィザードを使用したコレクション の作成 . . . . .	30	QuickPlace での Directory Assistance の構成 . . . . .	78
「コレクション」ビューを使用したコレクションの 作成 . . . . .	31	UNIX ファイル・システム・クローラー . . . . .	79
コレクションの編集 . . . . .	33	Web クローラー . . . . .	80
コレクションの削除 . . . . .	34	ユーザー・エージェントの構成 . . . . .	81
<b>エンタープライズ・サーチ・クローラーの 管理 . . . . .</b>	<b>37</b>	JavaScript のサポート . . . . .	82
クローラーの作成 . . . . .	40	Web クロール・スペースを制限する規則 . . . . .	82
クローラーのプロパティの編集 . . . . .	41	Web クローラーとの URL 接続のテスト . . . . .	87
クローラー・スペースの編集 . . . . .	41	Web クローラーにおける再クローラー間隔の設定 . . . . .	87
クローラーの削除 . . . . .	42	Web クローラーにおける URL アクセスのオプシ ョン . . . . .	87
Content Edition クローラー . . . . .	43	Web クローラーがソフト・エラーのページを処理 する方法 . . . . .	88
WebSphere II Content Edition リポジトリへのサ ーバー・モード・アクセス . . . . .	44	セキュア Web サイトのクローラーのサポート . . . . .	90
WebSphere II Content Edition リポジトリへの直 接モード・アクセス . . . . .	44	プロキシ・サーバーによってサービスされてい る Web サイト . . . . .	92
UNIX での WebSphere II Content Edition 用ク ローラー・サーバーの構成 . . . . .	45	Cookie の管理 . . . . .	93
		グローバル Web クロール・スペースの構成 . . . . .	95

リンクをたどらない、索引付けしないディレクティブ . . . . .	97
Web ページでのリンクをたどらないディレクティブおよび索引付けをしないディレクティブのオーバーライド . . . . .	97
WebSphere Portal クローラー . . . . .	98
WebSphere Portal クローラー用エンタープライズ・アプリケーションのデプロイ . . . . .	99
クロールする URL を WebSphere Portal からコピーする . . . . .	100
Windows ファイル・システム・クローラー . . . . .	101
データ・リスナー・アプリケーションのサポートの構成 . . . . .	102
カスタム・クローラー・プラグイン . . . . .	104
エンタープライズ・サーチ索引における URI フォーマット . . . . .	105

## エンタープライズ・サーチ・パーサーの管理 . . . . . 115

カテゴリに関する作業 . . . . .	116
ルール・ベースのカテゴリ . . . . .	117
モデル・ベースのカテゴリ . . . . .	120
カテゴリ・ツリー . . . . .	120
カテゴリ化タイプの選択 . . . . .	121
カテゴリの構成 . . . . .	122
XML 検索フィールドの作業 . . . . .	124
XML 検索フィールド . . . . .	124
XML エlementを検索フィールドにマップ . . . . .	125
HTML 検索フィールドの作業 . . . . .	127
HTML 検索フィールド . . . . .	127
検索フィールドへの HTML メタデータ・Elementのマッピング . . . . .	128
カスタム・テキスト処理 . . . . .	129
システムへのテキスト分析エンジンの追加 . . . . .	130
テキスト分析エンジンとコレクションとの関連付け . . . . .	131
共通分析構造への XML Elementのマッピング . . . . .	132
索引への共通分析構造のマッピング . . . . .	134
JDBC 表への共通分析構造のマッピング . . . . .	135
パーサー・サービス用スレッドの構成 . . . . .	136
複合語の拡張分析の使用可能化 . . . . .	136
ネイティブ XML 検索のサポートの使用可能化 . . . . .	137
中国語、日本語、韓国語の文書の言語分析 . . . . .	138
N-gram セグメンテーション . . . . .	138
非 ASCII 文字範囲からの改行文字の除去 . . . . .	138
コレクション・パーサーと Stellent セッションに関連付けられた文書タイプ . . . . .	139
文書タイプとコレクション・パーサーとの関連付け . . . . .	140
デフォルトのコレクション・パーサー・サービス規則 . . . . .	142
文書タイプと Stellent セッションとの関連付け . . . . .	142
Stellent セッション用デフォルト構文解析規則 . . . . .	145

## エンタープライズ・サーチ索引の管理 149

索引作成のスケジューリング . . . . .	150
索引スケジュールの変更 . . . . .	151
索引スケジュールの使用可能および使用不可設定 . . . . .	151
索引の並行作成の構成 . . . . .	152
索引の検索可能ビューに影響するオプション . . . . .	153
文書の検索に対する索引付きオプション . . . . .	154
照会内のワイルドカード文字 . . . . .	156
照会でのワイルドカード文字のオプションの構成 . . . . .	158
有効範囲 . . . . .	159
有効範囲の構成 . . . . .	160
縮小表示した URI . . . . .	161
検索結果内の URI の縮小表示 . . . . .	162
索引から URI を除去 . . . . .	163

## エンタープライズ・サーチの検索サーバー管理 . . . . . 165

検索キャッシュ . . . . .	166
検索キャッシュの構成 . . . . .	166
カスタム同義語辞書 . . . . .	167
システムへの同義語辞書の追加 . . . . .	168
同義語辞書とコレクションとの関連付け . . . . .	169
カスタムのストップワード辞書 . . . . .	169
システムへのストップワード辞書の追加 . . . . .	170
ストップワード辞書とコレクションとの関連付け . . . . .	171
動的要約 . . . . .	171
管理コンソールの文書サマリーのカスタマイズ . . . . .	172
プロパティの編集による文書サマリーのカスタマイズ . . . . .	172
クイック・リンクに関する作業 . . . . .	173
クイック・リンク . . . . .	173
クイック・リンクの構成 . . . . .	174

## エンタープライズ・サーチの文書ランキング . . . . . 177

テキスト・ベースのスコアリング . . . . .	177
静的ランキング . . . . .	178
カスタム・ランキング調整ワード辞書 . . . . .	178
ランキング調整ワード辞書のシステムへの追加 . . . . .	180
ランキング調整ワード辞書とコレクションとの関連付け . . . . .	180
URI パターンに基づいた文書ランキング . . . . .	181
URI パターンに一致する文書のスコアの操作 . . . . .	182
ランキング調整クラスに基づいた文書ランキング . . . . .	183
ランキング調整クラスへのフィールドのマッピング . . . . .	185
ランキング調整クラスのランキング調整要因の構成 . . . . .	186
ランキング調整クラスのデフォルト値 . . . . .	187

## エンタープライズ・サーチの検索アプリケーション . . . . . 189

コレクションとの検索アプリケーションの関連付け . . . . .	190
サンプル検索アプリケーションの機能 . . . . .	190
サンプル検索アプリケーションのプロパティ . . . . .	192

サンプル検索アプリケーション・プロパティの編集 . . . . .	201
サンプル検索アプリケーションへのアクセス . . . . .	203
サンプル検索アプリケーションに対するセキュリティの使用可能化 . . . . .	204
<b>  エンタープライズ・サーチ外部ソース . . . . .</b>	<b>207</b>
システムへの外部ソースの追加 . . . . .	207
外部ソースと検索アプリケーションの関連付け . . . . .	209
<b>エンタープライズ・サーチ・セキュリティ . . . . .</b>	<b>211</b>
管理役割 . . . . .	212
管理ユーザーの構成 . . . . .	214
認証とアクセス制御 . . . . .	214
WebSphere Application Server 内でのエンタープライズ・アプリケーションのセキュリティの使用不可化 . . . . .	216
コレクション・レベルのセキュリティ . . . . .	217
重複文書分析 . . . . .	218
アンカー・テキスト分析 . . . . .	218
禁止文書へのリンク内のアンカー・テキストの索引作成 . . . . .	219
検索アプリケーション ID を使用したセキュリティ . . . . .	220
文書レベル・セキュリティ . . . . .	221
保管されたセキュリティ・トークンによる検証	221
照会処理中の現行の信用証明情報の検証 . . . . .	222
Windows ファイル・システム文書に対する文書レベルのセキュリティの施行 . . . . .	225
Lotus Domino 文書に対する文書レベルのセキュリティの施行 . . . . .	228
文書レベル・セキュリティを使用不可にする . . . . .	230
<b>エンタープライズ・サーチの WebSphere Portal との統合 . . . . .</b>	<b>233</b>
ポータル検索エンジンでの文書レベル・セキュリティ . . . . .	235
検索ポートレットのデプロイ . . . . .	236
WebSphere Portal 検索ブラウザ・ポートレットのエンタープライズ・サーチのための構成 . . . . .	238
検索センターを使用するためのエンタープライズ・サーチ・アダプターのインストール . . . . .	239
検索センターを使用するためのエンタープライズ・サーチ登録ポートレットのインストール . . . . .	241
<b>WebSphere Portal からエンタープライズ・サーチへのマイグレーション . . . . .</b>	<b>243</b>
WebSphere Portal からのモデル・ベース分類構造のマイグレーション . . . . .	243
WebSphere Portal からのコレクションのマイグレーション . . . . .	245
マイグレーションされたコレクションの設定 . . . . .	248
マイグレーション・ウィザードのログ・ファイル . . . . .	249

<b>エンタープライズ・サーチ・サーバーの開始と停止 . . . . .</b>	<b>251</b>
エンタープライズ・サーチ・サーバーの開始 . . . . .	251
エンタープライズ・サーチ・サーバーの停止 . . . . .	253
<b>エンタープライズ・サーチ・アクティビティのモニター . . . . .</b>	<b>257</b>
単一コレクション内の文書数の見積もり . . . . .	258
システム・リソースの可用性のチェック . . . . .	258
コレクションのモニター . . . . .	259
URI についての詳細表示 . . . . .	260
クローラーのモニター . . . . .	262
Web クローラーのアクティビティ詳細の表示 . . . . .	264
Web クローラーのスレッドの詳細 . . . . .	265
Web クローラー・アクティブ・サイト . . . . .	265
Web クローラーのクロール速度 . . . . .	266
Web クローラー・レポートの作成 . . . . .	267
Web クローラーの HTTP 戻りコード . . . . .	268
パーサーのモニター . . . . .	272
コレクションの索引アクティビティのモニター . . . . .	273
エンタープライズ・サーチの索引キューのモニター . . . . .	274
検索サーバーのモニター . . . . .	275
データ・リスナーのモニター . . . . .	277
文書トラッキング . . . . .	278
文書トラッキングに対するログ・ファイルの構成 . . . . .	278
除去された文書に関するレポートの表示 . . . . .	279
<b>エンタープライズ・サーチのログ・ファイルおよびアラート . . . . .</b>	<b>281</b>
アラート . . . . .	281
コレクション・レベル・アラートの構成 . . . . .	282
システム・レベル・アラートの構成 . . . . .	284
ログ・ファイルの構成 . . . . .	285
SMTP サーバー情報の構成 . . . . .	286
ログ・メッセージについての E メール受信 . . . . .	287
ログ・ファイルの表示 . . . . .	289
<b>エンタープライズ・サーチ・システムのバックアップとリストア . . . . .</b>	<b>291</b>
エンタープライズ・サーチ・システムのバックアップ . . . . .	292
エンタープライズ・サーチ・システムのリストア . . . . .	293
エンタープライズ・サーチ・システム・ファイルを新規のサーバーにリストアする . . . . .	294
<b>  エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID . . . . .</b>	<b>297</b>
<b>エンタープライズ・サーチの資料 . . . . .</b>	<b>327</b>
<b>WebSphere II OmniFind Edition アクセシビリティ . . . . .</b>	<b>329</b>
<b>エンタープライズ・サーチの用語集 . . . . .</b>	<b>331</b>

<b>WebSphere Information Integration</b>	
に関する情報の入手 . . . . .	<b>343</b>
<b>IBM と連絡を取る.</b> . . . . .	<b>345</b>

商標 . . . . .	<b>347</b>
索引 . . . . .	<b>353</b>



---

## 本書のトピックについて

IBM® WebSphere® Information Integrator OmniFind™ Edition バージョン 8.3 システムを管理するときに、この情報を使用してください。

WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) には、エンタープライズ・サーチ という機能が搭載されています。エンタープライズ・サーチのコンポーネントは、WebSphere II OmniFind Edition 製品のインストール時に、インストールされます。エンタープライズ・サーチ という用語は、インストール・パスおよび製品パッケージ化ラベルを参照する場合をのぞき、WebSphere II OmniFind Edition の資料全体で使用されています。

エンタープライズ・サーチの管理資料では、以下のトピックについて説明します。

- エンタープライズ・サーチの概念、コンポーネント、および機能の紹介
- コレクションの作成、モニター、および管理の方法に関する説明
- データの解析方法、および検索や取り出しを最適化するために解析アクティビティをカスタマイズする方法
- 解析データの索引付けの方法と、索引付けアクティビティを管理する方法
- 検索サーバーがデータを検索する方法、ランク付けする方法、検索結果を戻す方法
- カスタム検索アプリケーションをコレクションと関連付ける方法
- エンタープライズ・サーチで使用可能なセキュリティー・レベルの概要
- エンタープライズ・サーチを IBM WebSphere Portal に組み込む方法
- WebSphere Portal の分類構造とコレクションをエンタープライズ・サーチのカテゴリ・ツリーとコレクションにマイグレーションする方法
- システムのバックアップとリストアの方法
- ログ・ファイルの作成と表示方法
- エンタープライズ・サーチ・コマンドの使用方法和、戻された情報の解釈方法

---

## 本書の対象読者

本書は、エンタープライズ・サーチ・コレクションの作成、モニター、および管理を担当するシステム管理者およびシステム・オペレーターを対象としています。

本書では、コレクションの作成、コレクションのコンテンツの選択、コンテンツを検索可能にするオプションの構成について説明します。また、コレクションおよびシステム・アクティビティのモニター、エンタープライズ・サーチ管理者としてのユーザーの登録、検索アプリケーションと、コレクションおよび外部検索可能ソースとの関連付けについても説明します。

本書を活用するためには、Web アプリケーションを十分理解していること、検索したいデータ・ソースについての経験があることが必要です。



---

## エンタープライズ・サーチとは?

エンタープライズ・サーチ・システムは、1回の照会により、構造化されているかどうかを問わず任意の数のデータ・ソースを検索できる幅広い機能を持っています。エンタープライズ・サーチ・システムは照会に対する応答が早く、結果セットを統合してランク付けする機能があるため、必要な情報を簡単に見つけることができます。

エンタープライズ・サーチ・コンポーネントは、IBM WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) と共にインストールされ、企業全体に渡って情報を収集し、それらの情報を検索で利用できるようにします。ユーザーは、Web ブラウザーに照会を入力するだけで、ローカルおよびリモート・データベース、コラボレーション・システム、コンテンツ・マネジメント・システム、ファイル・システム、および内外の Web サイトを同時に検索することができます。

エンタープライズ・サーチ・システムは、既存のシステムとシームレスに統合できるように設計されており、多種多様なソースからデータを収集して高速に検索するために必要な流れを処理します。また、エンタープライズ・サーチでは、言語分析その他の分析をデータに適用し、有用性の高い検索結果を提供することができます。異なるタイプのリポジトリを検索する場合でも、異なるインターフェースの使い方を習得する必要もありません。

エンタープライズ・サーチ索引に含めることを希望しないデータ・ソースを検索するサポートを追加できます。エンタープライズ・サーチのフェデレーテッド・サーチ機能を使用して、索引付けされたデータ・ソースを検索するときに同時に、これらの外部ソースを検索できます。

### 検索品質

探している情報をユーザーが確実に検出するために、WebSphere II OmniFind Edition は IBM Unstructured Information Management Architecture (UIMA) をサポートします。UIMA は、テキスト分析法の共通かつ標準インターフェースを定義するオープン・フレームワークです。大規模なテキスト分析にあたって、エンタープライズ・サーチは、概念、潜在的意味、関係、事実、および非構造化テキストではしばしば隠されている他の関連データを識別できます。分析中に抽出される情報は、検索結果の品質を拡大するために使用したり、ビジネス・インテリジェンスやデータ・マイニングなど、他のアプリケーションの品質を拡大するために使用したりすることができます。

### セキュリティ

セキュリティは、エンタープライズ・サーチに不可欠な要素です。システム管理は、権限のあるユーザーしか操作できないようになっています。IBM WebSphere Application Server で使用可能なセキュリティ・メカニズムを使用して、管理役割を構成して、さまざまな管理機能に対するアクセス権を、どのユーザーが持つかを制御できます。

また、データ収集時に、セキュリティー・トークンを収集データに関連付けるオプションを指定することもできます。検索アプリケーションでセキュリティーを使用可能にすれば、索引内の文書と一緒に保管されている、これらのトークンを使用して、アクセス制御を施行し、適切な信用証明情報を持つユーザーのみがデータを照会し、検索結果を表示できるようにすることができます。

特定のタイプのデータ・ソースの場合、オプションを構成して、照会処理中に、現行のアクセス制御によってユーザーのログイン信用証明情報を妥当性検査することができます。この追加のセキュリティーの層は、ネイティブ・データ・ソースによって、リアルタイムでユーザーの権限を確実に妥当性検査するようにします。このオプションによって、文書およびそのセキュリティー・トークンが索引付けされた後、ユーザーの信用証明情報が変更になったインスタンスに対してプロテクトすることができます。

#### 関連概念

エンタープライズ・サーチ・セキュリティー

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』

「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

---

## エンタープライズ・サーチでサポートされているデータ・ソース・タイプ

データ・ソースを検索する場合に、さまざまなデータ・ソース・タイプが検索できるようにあらかじめサポート設定されています。

IBM WebSphere Information Integrator OmniFind Edition をインストールすると、次の各タイプのデータ・ソースからのデータの収集を開始できます。

- IBM DB2<sup>®</sup> Content Manager 項目タイプ (文書、リソース、および項目)
- IBM DB2 Universal Database<sup>™</sup> データベース
- IBM Domino<sup>®</sup> Document Manager (以前の Domino.Doc<sup>®</sup>) データベース
- IBM Lotus Notes<sup>®</sup> データベース
- IBM Lotus<sup>®</sup> QuickPlace<sup>®</sup> データベース
- Microsoft<sup>®</sup> Exchange Server パブリック・フォルダー
- Microsoft Windows<sup>®</sup> ファイル・システム
- NNTP ニュースグループ
- UNIX<sup>®</sup> ファイル・システム
- IBM WebSphere Information Integrator Content Edition リポジトリ:  
Documentum、FileNet Panagon Content Services、FileNet P8 Content  
Manager、Hummingbird<sup>®</sup> Document Management (DM)、OpenText Livelink、お  
よび Portal Document Manager (PDM) の項目クラス
- IBM DB2 Universal Database for z/OS<sup>®</sup>、IBM Informix<sup>®</sup>、Microsoft SQL  
Server、Oracle、および Sybase データベース用の IBM WebSphere Information  
Integrator ニックネーム表
- IBM WebSphere Portal のサイト
- インターネット上およびイントラネット内の Web サイト

これらのソースをエンタープライズ・サーチ索引に追加せずに、以下のタイプの外部ソースを検索するサポートを追加できます。

- Java™ Database Connectivity (JDBC) データベース表 (IBM DB2 Universal Database (DB2 UDB) および Oracle データベースのみ)。外部ソースは JDBC データベース内の各表ごとに個別に作成されます。
- Lightweight Directory Access Protocol (LDAP) サーバー。各 LDAP サーバーごとに 1 つの外部ソースが作成されます。

サポートされるデータ・ソース・タイプに関する情報については、WebSphere Information Integrator OmniFind Edition Web サイト を参照してください。

#### 関連概念

エンタープライズ・サーチ外部ソース

---

## エンタープライズ・サーチ・コンポーネントの概要

エンタープライズ・サーチでは、そのコンポーネントによって、企業全体からデータを収集し、収集した情報を分析、解析、カテゴリー化し、ユーザーが検索に使用する索引を作成します。

エンタープライズ・サーチ・コレクション は、ユーザーが 1 回の照会で検索できるデータ・ソースの集合を表します。コレクションを作成する場合、コレクションに入れるソースを指定し、ユーザーが索引付けされたデータを検索するときのオプションを構成します。

コレクションは複数作成することができ、各コレクションにはさまざまな種類のデータ・ソースのデータを入れることができます。例えば、IBM DB2 Universal Database、IBM Lotus Notes、および IBM DB2 Content Manager データベースの文書を含むコレクションを作成することができます。ユーザーがこのコレクションを検索すると、検索結果にはこれらの各データ・ソースの文書が含まれることになります。

フェデレーテッド検索のサポートにより、1 回の照会で複数のコレクションを検索することができます。場合によっては、検索結果にエンタープライズ・サーチ・システム内のすべてのコレクションと外部ソースの文書が組み込まれる可能性があります。

コレクションの作成および管理には、次のようなアクティビティがあります。

#### データの収集

クローラー・コンポーネントによって、連続的に、または指定したスケジュールに従ってデータ・ソースから文書が収集されます。頻繁にクロールを行うことにより、ユーザーが常に最新の情報にアクセスできるようになります。

#### データの分析

パーサー・コンポーネントによって、文書からテキストが抽出され、クローラーがクロールした各文書に対して言語分析その他の分析が行われます。この詳細なコンテンツ分析により、検索結果の品質が向上します。

#### データの索引作成

索引 コンポーネントが、定期的にスケジュールに従って実行され、新規の

文書、変更された文書に関する情報が索引に追加されます。索引コンポーネントでは、コレクション内の文書に対してグローバル分析を行い、検索結果の品質を高めます。

### データの検索

検索 コンポーネントによって、索引が検索され、検索アプリケーションと連携して検索要求を処理し、処理された検索結果を戻します。

他の WebSphere Information Integrator OmniFind Edition コンポーネントによって、セキュリティ設定の指定、システム・アクティビティのモニター、さらに問題が発生した場合のトラブルシューティングが行えるようになります。また、この製品では、独自の検索アプリケーションを作成するときのテンプレートとして使用できる検索アプリケーションの実用サンプルも提供されています。

### 関連概念

エンタープライズ・サーチ・クローラーの管理

エンタープライズ・サーチ・パーサーの管理

エンタープライズ・サーチ索引の管理

エンタープライズ・サーチの検索サーバー管理

## エンタープライズ・サーチ・クローラー

エンタープライズ・サーチ・クローラーは、データ・ソースから文書を収集し、それらの文書が分析、索引作成、および検索できるようにします。

WebSphere Information Integrator OmniFind Edition で提供されているクローラー・コンポーネントには、次のような機能があります。

- クローラーを構成すると、ディスカバリー・プロセスは、Lotus Notes データベースのすべてのビューとフォルダー名、または UNIX サーバーの全ファイル・システム名など、クローラー使用することができるソースに関する情報を探し出します。
- クローラーするソースを選択し、クローラーを開始すると、クローラー・コンポーネントはソースからデータを収集し、それらのデータが分析および索引作成できるようにします。

1 つのコレクションには複数のクローラーを入れることができ、各クローラーは特定タイプのデータ・ソースからデータを収集するように設計されています。例えば、ファイル・システム、Notes® データベース、およびリレーショナル・データベースのデータを結合するために、3 つのクローラーを同じコレクションに作成することができます。また、同一タイプのクローラーを複数作成し、各クローラーがクローラーしているデータの変更頻度に従って、それぞれ異なるクローラー・スケジュールを設定することもできます。

Web、WebSphere Portal、および Network News Transfer Protocol (NNTP) のソース用クローラーは継続的に稼働します。クローラー対象の URL または NNTP ニュースグループを指定すると、クローラーは、新規または変更されたデータがないか定期的にそれらの対象に戻ってチェックします。他のタイプのクローラーは、手動で開始および停止することもできますし、クローラー・スケジュールを設定することもで



きます。クローラーのスケジュールを設定する場合、初回実行の日時、および新規および変更された文書をクローリングするためにデータ・ソースにクローラーがアクセスする頻度を指定します。

クローラーのプロパティは、特定のクローラーがクローリングするときの動作を制御する規則の集合です。例えば、クローラーのシステム・リソースの使用を制御する規則を指定します。クローリング対象として適格なデータ・ソースの集合は、クローラーのクローリング・スペースを構成します。クローラーを作成後、クローラーのプロパティは、いつでも編集し、クローラーのデータ収集指定を変更することができます。また、クローリング・スペースを編集して、クローラー・スケジュールの変更、新規ソースの追加、または今後検索しないソースの除去を行うこともできます。

#### 関連概念

エンタープライズ・サーチ・クローラーの管理

#### 関連タスク

クローラーのモニター

## エンタープライズ・サーチ・パーサー

エンタープライズ・サーチ・パーサーは、クローラーによって収集された文書を分析し、文書の索引作成の準備を行います。

WebSphere Information Integrator OmniFind Edition で提供されているパーサー・コンポーネントは、文書コンテンツおよび文書のメタデータを分析します。分析結果は、索引作成コンポーネントがアクセスできるようにデータ・ストアに格納します。パーサーは、以下のようなタスクを実行します。

- さまざまなフォーマットによる文書からテキストを抽出する。例えば、パーサーは、XML および HTML 文書のタグ内のテキストを抽出します。Stellent for IBM WebSphere Information Integrator OmniFind Edition Outside In Viewer Technology の使用により、パーサーはまた、Microsoft Word や Adobe Acrobat PDF 文書などのバイナリー・フォーマットのテキストも抽出します。
- 各文書の文字セット・エンコードを検出する。パーサーは、言語分析を行う前処理として、検出したこの情報を使用してすべてテキストをユニコードに変換します。
- 各文書のソース言語を検出する。
- コレクションに管理者が指定した解析規則を適用する。パーサーの構成時、次の項目を構成できます。

#### XML 文書および HTML 文書のフィールド・マッピング規則

このオプションを使用して、ユーザーは XML 文書および HTML 文書内の構造化されたコンテンツ、および構造化されていないコンテンツを区別して検索することができます。XML エlementまたは HTML メタデータ・Elementをエンタープライズ・サーチ索引の検索フィールドにマップすると、ユーザーは照会でフィールド名を指定して、XML 文書および HTML 文書の特定の部分を検索することができます。(特定フィールドを検索する照会では、文書の全コンテンツを検索するフリー・テキスト照会より精度の高い検索結果が得られます。)

#### カテゴリー

このオプションで、ユーザーは、文書が属するカテゴリーを基準にして文

書を検索できます。また、ユーザーは検索結果でカテゴリを選択して、同じカテゴリに属する文書だけを参照することもできます。

コレクションの作成時に、カテゴリを使用する場合は、使用するカテゴリのタイプを選択します。「ルール・ベース」のカテゴリを使用すると、文書は管理者が定義した規則に従ってカテゴリに関連付けられます。ルール・ベースのカテゴリは、新規作成したエンタープライズ・サーチ・コレクションと IBM WebSphere Portal からマイグレーションしたコレクションで構成することができます。

「モデル・ベース」のカテゴリを使用すると、文書は WebSphere Portal システムにあるモデル・ベースのカテゴリに関連付けられます。このオプションを使用するには、エンタープライズ・サーチ索引サーバーに WebSphere Portal がインストールされていなければなりません。また、WebSphere Portal のカテゴリ化ツールを使用して、カテゴリを管理する必要があります。

### カスタム・テキスト分析

アプリケーション開発者は、ユーザーが検索するデータに複雑な言語分析を実行するために、カスタム分析プログラムを作成することができます。こうしたプログラムをエンタープライズ・サーチ・システムに埋め込み、これを使用してコレクションのコンテンツに注釈を付けることができます。注釈に索引付けをすることで、コレクションにセマンティック検索を行うことができます。

例えば、ユーザーは、互いに隣接して出現する照会用語や、同一文に出現する照会用語を検索したり、あるいは、照会用語間の関係を検索することができます (例: Smith という名前の IBM 技術者ではなく、Smith という名前の IBM 営業担当者についての文書、など)。

### N-gram セグメンテーションのサポート

中国語、日本語、韓国語で書かれた文書の検索性能を高めるためには、字句解析の N-gram セグメンテーション方式を使用可能にします。この分析方法では、ワードの区切りに空白文字を使用しません。(コレクション作成後に、セグメンテーション方式を変更することはできません。)

### ネイティブ XML 検索での XML 文書検索サポート

ネイティブ XML 検索は、XML マークアップを検索することで、より正確な検索結果を提供できます。例えば、照会で、あるワードが特定の XML エlement に出現しなければならないと指定することができます。

### フィールドの相対的重要度スコアをランキング調整するクラス

フィールドをランキング調整クラスにマップすると、検索結果でどのように文書をランク付けるかに影響を与えることができます。例えば、タイトル・フィールドのスコアをランキング調整したいとします。照会用語がタイトルに出現した場合、そのタイトルの用語を持つ文書が、検索結果でより高くランク付けされます。

- テキストを抽出し、トークンを追加してデータの検索機能を拡張する。このプロセスでは、パーサーは、以下のタスクを実行します。
  - 大文字の正規化、およびドイツ語のウムラウトなどの発音区別符号の正規化などの文字の正規化。



- 段落、文、単語、および空白文字の構造分析。パーサーは、このような言語分析により、複合語を分解し、辞書検索および同義語検索ができるようにするためのトークンの指定を行います。

### 関連概念

カテゴリに関する作業

XML 検索フィールド

HTML 検索フィールド

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』

「テキスト分析機能ガイド」の『エンタープライズ・サーチに組み込まれているテキスト分析』

### 関連タスク

パーサーのモニター

## エンタープライズ・サーチ索引

エンタープライズ・サーチの索引作成コンポーネントは、定期的なスケジュールに従って実行され、新規の文書、変更された文書に関する情報を索引に追加します。

ユーザーが、常に検索対象のソースの最新情報にアクセスできるようにするため、索引の作成には、次の 2 つの段階があります。

### 索引の再編成

索引を再編成すると、構成が最適な編成になるように索引全体が再作成されます。索引作成プロセスによって、クローラーが収集し、パーサーが分析したすべてのデータが読み込まれます。

### 索引のリフレッシュ

索引をリフレッシュすると、最後の索引の再編成の後にクロールされた情報が、索引に追加されます。

コレクションの索引オプションを構成する場合、索引の再編成およびリフレッシュのスケジュールを指定できます。索引の再編成やリフレッシュの頻度は、システム・リソースや、索引付けされるソースが静的コンテンツと動的コンテンツのどちらを含むかによって異なります。

確実に新しい情報を利用できるようにするため、索引のリフレッシュ・スケジュールは頻繁に行うようにしてください。また、定期的に索引の再編成をスケジュール設定して、新規の情報がすべて統合され、新規のコンテンツが分析されて、索引のパフォーマンスが最適化されるようにしてください。

また、索引作成プロセスは、スケジューリングせずに開始することもできます。例えば、特定の解析規則を変更し、その変更を使用している検索アプリケーションで有効になるようにしたい場合、スケジュールされた時に開始する索引の再編成を待たずにデータを再クロールし、解析した後、索引の再編成を開始することができます。

リソースの使用を制御するには、いくつかのコレクションが索引付けプロセスを共用し、索引作成要求を同時にサブミットするかを制御します。同時に索引作成を行うと、大規模な索引の再編成で、他の索引のリフレッシュを妨げないようにすることができます。索引作成は、リソースを集中的に使用するプロセスになりかねません

ので、大規模なシステムの場合、システム負荷をモニターして、索引の再編成およびリフレッシュ頻度を調整するようにしてください。

索引作成時、索引プロセスでは、グローバル文書分析を行います。この分析では、文書の重複を識別し、文書のリンク構造を分析し、アンカー・テキスト (ハイパーテキスト・リンクでターゲット・ページを記述するテキスト) に対する特殊処理を行うアルゴリズムが適用されます。

以下の索引付けアクティビティのオプションが指定できます。

- ユーザーがワイルドカード文字を指定できるように、照会用語を索引に拡張するサポートを構築するか、照会処理中に照会用語を拡張することを指定します。これを判断するときは、リソースの使用量と照会応答時間のトレードオフを考慮する必要があります。
- 有効範囲を構成できます。有効範囲を設定すると、ユーザーがコレクションで表示できる内容を制限することができます。例えば、技術サポート部の文書の URI を含む有効範囲を作成したり、人事部の文書の URI を含む別の有効範囲を作成したりすることができます。検索アプリケーションで有効範囲をサポートすると、ユーザーはコレクション中の該当するサブセットだけを対象として文書を検索、取得することができます。
- 同じ URI 接頭部をもつ検索結果文書を縮小表示するオプションが指定できます。また、URI 接頭部が異なる文書を検索結果にまとめて縮小表示できるようにグループ名を指定することもできます。
- 索引を作成後、ユーザーに検索させたくない URI を削除することができます。

#### 関連概念

エンタープライズ・サーチ索引の管理

照会内のワイルドカード文字

有効範囲

縮小表示した URI

URI パターンに基づいた文書ランキング

#### 関連タスク

索引作成のスケジューリング

索引の並行作成の構成

索引から URI を除去

コレクションの索引アクティビティのモニター

エンタープライズ・サーチの索引キューのモニター

## エンタープライズ・サーチ用検索サーバー

エンタープライズ・サーチ用の検索サーバーは、検索アプリケーションと連携して、照会の処理、索引の検索、および検索結果を戻す処理を行います。

エンタープライズ・サーチ用検索サーバーは、WebSphere Information Integrator OmniFind Edition のインストール時にインストールされます。コレクションに検索サーバーを設定すると、コレクションを検索する方法のオプションを指定できます。

- 検索キャッシュを構成して、頻繁に要求される検索結果を保持することができます。検索キャッシュを利用すると、検索効率が高まります。
- コレクション内の文書検索のデフォルト言語を指定できます。
- アプリケーション開発者がカスタム辞書を作成する場合、それら辞書をコレクションに関連付けることができます。
  - ユーザーが同義語辞書を使用するコレクションを照会するとき、照会用語の同義語を含む文書は、検索結果に含まれます。
  - ユーザーがストップワード辞書を使用するコレクションを照会するとき、ストップワードは、照会が処理される前に、照会から除去されます。
  - ユーザーがランキング調整ワード辞書を使用するコレクションを照会するとき、ランキング調整ワードを含む文書の重要度は、辞書のワードと関連付けられているランキング調整要因に従って調整されます。
- 特定の文書が特定の照会に関連付けられることが事前にわかっている場合、クイック・リンクを構成することができます。クイック・リンクでは、特定の URI と特定のキーワードおよび句を関連付けます。照会にクイック・リンク定義で指定したキーワードまたは句がある場合、関連付けられた URI が自動的に検索結果に戻されます。

マルチサーバー構成では、障害保護がサーバー・レベルではなく、コレクション・レベルで使用可能です。1つの検索サーバー上のコレクションがいかなる理由にする使用不可になった場合、その後そのコレクションの照会は、自動的に別の検索サーバーにルーティングされます。

#### 関連概念

エンタープライズ・サーチの検索アプリケーション

検索キャッシュ

カスタム同義語辞書

カスタムのストップワード辞書

カスタム・ランキング調整ワード辞書

クイック・リンク

#### 関連タスク

検索サーバーのモニター

## エンタープライズ・サーチ管理コンソール

エンタープライズ・サーチ管理コンソールは、ブラウザーで稼働します。すなわち、管理ユーザーは、いつでもどこからでもこのコンソールにアクセスできるようになります。同時に、セキュリティー機構によって、管理機能にアクセスすることを許可されたユーザーしかコンソールにアクセスできないことが保証されています。

エンタープライズ・サーチの管理コンソールは、WebSphere Information Integrator OmniFind Edition のインストール時に、検索サーバーにインストールされます。

管理コンソールには、さまざまな基本的管理用タスクの実行を助けるウィザードが付属しています。例えば、コレクション・ウィザードを使用すれば、簡単にコレクションを作成でき、ドラフト・モードで作業をする手間が省けます。クローラー・

ウィザードは、データ・ソース・タイプによって異なり、ユーザーが検索できるようにしたいソースを選択する作業を支援します。

他の管理タスクについては、管理する個別の項目を選択します。例えば、コレクションを編集する場合、索引スケジュールを変更するには「索引」ページを選択し、XML 文書の構文解析規則を変更するには「解析」ページを選択します。

#### 関連概念

エンタープライズ・サーチ管理の概要

管理役割

#### 関連タスク

管理コンソールへのログイン

## エンタープライズ・サーチ・システムのモニター

エンタープライズ・サーチ管理コンソールを使用すれば、システム・アクティビティをモニターし、必要に応じてその運用を調整することができます。

WebSphere Information Integrator OmniFind Edition をインストールし、最少でも 1 つのコレクションを作成すると、主要な各アクティビティ (クロール、解析、索引作成、検索) の詳細な統計データを表示することができます。統計データには、平均応答時間や、クロールまたは索引作成セッションでいくつの文書がクロールまたは索引作成されたかなどの進行状況データがあります。

ほとんどのアクティビティは、管理者が開始および停止することができます。例えば、アクティビティを一時停止して、構成の変更または問題のトラブルシューティングを行い、アクティビティを進められる状況になった時点で処理を再開することができます。

また、アラートを設定することもでき、これによりモニターしている事象が発生したときはいつでも、特定のモニター・アクティビティに関する E メールを受信できるようになります。例えば、検索応答時間が指定したしきい値を超過した場合にアラートを受信することができます。

文書がエンタープライズ・サーチ・システムから除去された場合、文書をトラッキングして、いつ、どこで、なぜ、その文書が除去されたのかを判別できます。例えば、パーサーがその文書を構文解析できなかったか、または管理者が索引から文書を除去したことが考えられます。

#### 関連概念

エンタープライズ・サーチ・アクティビティのモニター

エンタープライズ・サーチ・サーバーの開始と停止

## エンタープライズ・サーチ用ログ・ファイル

ログ・ファイルは、個別のコレクションおよびシステム・レベル・セッションに対して作成されます。

特定のエンタープライズ・サーチ・コレクションまたはシステムに対してログイン・オプションを構成する場合、ログに記録するメッセージのタイプ (エラー・メッセージ、警告メッセージなど) を指定します。また、新しいメッセージのスペー

スを確保するためにシステムが古くなったログ・ファイルを循環使用する頻度を指定します。特定のメッセージ (アラートを含む) 、あるいはエラーが発生したときのすべてのエラー・メッセージについて、E メールを受信するオプションを選択できます。

ログ・ファイルを表示する場合、表示するログ・ファイル (ファイル名を見れば、ファイル作成日時、メッセージを発行したコンポーネントがわかります) を選択します。表示フィルターを指定することもできます。例えば、エラー・メッセージだけを表示したり、特定エンタープライズ・サーチ・セッションから出されたメッセージだけを表示することができます。

#### 関連概念

エンタープライズ・サーチのログ・ファイルおよびアラート

アラート

エンタープライズ・サーチのメッセージ

#### 関連タスク

ログ・ファイルの構成

SMTP サーバー情報の構成

ログ・メッセージについての E メール受信

ログ・ファイルの表示

## エンタープライズ・サーチのカスタマイズ

エンタープライズ・サーチ用のアプリケーション・プログラミング・インターフェース (API) を使用すれば、カスタム検索アプリケーション、コレクションのコンテンツを更新するカスタム・アプリケーション、テキスト分析用のカスタム・プログラム、および同義語、ストップワード、ランキング調整ワード用の辞書を作成することができます。

WebSphere Information Integrator OmniFind Edition をインストールすると、エンタープライズ・サーチ・コレクションの拡張に使用可能な、次の一連の API が使用できるようになります。

#### 検索および索引 API (SI-API)

この API を使用して、カスタム検索アプリケーションおよびカスタム管理インターフェースを作成します。

#### データ・リスナー API

この API は、外部クローラーからデータを検索する場合に使用します。外部クローラーは、エンタープライズ・サーチのデータ・リスナーに接続でき、その後コレクションにデータを追加したり、コレクションからデータを除去したりすることができます。

#### クローラー・プラグイン

Web および Web 以外のクローラー用 API を使用して、クロール中の文書にメタデータを追加したり、あるいは組織の業務を強化するセキュリティ・トークンとセキュリティ規則を関連付けたりします。

言語分析用カスタム・プログラムをエンタープライズ・サーチ・コレクションと統合することで情報の検索性を高めることができます。 カスタム・テキスト分析エンジンをシステムに追加すると、そのエンジンをコレクションに関連付けることがで



きます。ユーザーがコレクションを照会したときに、カスタム・プログラムで索引に構築したワードの関連付けが有効に働きます。例えば、ユーザーは単に用語そのものだけではなく、概念および用語と用語の間の関係を検索できます。

また、頭字語、省略語、業界固有の用語などを反映したカスタム辞書を統合することで、情報をより多く取得できるようになります。辞書をシステムに追加した後、それら辞書をコレクションに関連付けることができます。ユーザーがコレクションを照会したとき、次のような仕方で利点がもたらされます。

- 照会に同義語として定義されたワードが含まれる場合、照会用語の同義語を含む文書は、検索結果に含まれることとなります。
- 照会にストップワードが含まれる場合、そのストップワードは照会から除外されることになり、無関係な文書は検索結果に戻されません。
- 照会にランキング調整ワードが含まれる場合、ランキング調整ワードを含む文書は、検索結果において、辞書のワードと関連付けられたランキング調整値に従って調整されて、ランク付けされます。

#### 関連概念

エンタープライズ・サーチの検索アプリケーション

カスタム同義語辞書

カスタムのストップワード辞書

カスタム・ランキング調整ワード辞書

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API の概要』

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『データ・リスナー』

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『クローラー・プラグイン』

#### 関連タスク

データ・リスナー・アプリケーションのサポートの構成

## エンタープライズ・サーチ用サンプル検索アプリケーション

カスタム検索アプリケーションを作成する場合、エンタープライズ・サーチ用のサンプル検索アプリケーションを、提供されたまま、またはそれをテンプレートとして使用することができます。

サンプル検索アプリケーションは、WebSphere Information Integrator OmniFind Edition のインストール時に、併せてインストールされます。サンプル検索アプリケーションによって、エンタープライズ・サーチで使用可能な検索および取り出し機能のほとんどを実際に試すことができます。また、このアプリケーションは実用的なサンプルで、これによってエンタープライズ・サーチ・システム内のすべてのアクティブ・コレクションおよび外部ソースを検索できます。コレクションまたは外部ソースをユーザーに提供する前に、サンプル・アプリケーションを使用して、新規のコレクションおよび外部ソースをテストできます。

サンプル検索アプリケーションは、一度に複数のコレクションおよび外部ソースの検索を可能にして、フェデレーテッド検索の対応も実際に試すことができます。

コレクションに対するセキュリティーを使用可能にし、またシステムに対する WebSphere II OmniFind Edition ID 管理を使用可能にすると、ユーザーは検索アプリケーションを使用する際にプロファイルを作成できます。ユーザー・プロファイルは、さまざまなドメインにログインするためにユーザーが指定する信用証明情報を保管します。

照会の処理中、検索プロセスは、保管された信用証明情報 (WebSphere II OmniFind Edition によって管理されるセキュア保管に暗号化された) を使用して、セキュア・ドメインを検索する許可をユーザーが持っているかどうかを判別します。信用証明情報が欠如しているか、またはドメインに対して有効でない場合、そのドメインからの文書は、検索結果から除外されます。

サンプル検索アプリケーションの使用方法については、アプリケーションの使用中に「ヘルプ」をクリックしてください。カスタム検索アプリケーションを作成するには、エンタープライズ・サーチ用検索および索引 API を使用します。

#### 関連概念

エンタープライズ・サーチの検索アプリケーション

サンプル検索アプリケーションの機能

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API の概要』

#### 関連タスク

サンプル検索アプリケーションへのアクセス

サンプル検索アプリケーション・プロパティの編集

サンプル検索アプリケーションに対するセキュリティーの使用可能化

---

## エンタープライズ・サーチのデータ・フロー

WebSphere Information Integrator OmniFind Edition と共にインストールされるエンタープライズ・サーチ・コンポーネントは、互いに連携してシステム内のデータ・フローを確保しています。

クローラーは、企業全体のデータ・ソースから文書を収集します。パーサーは、クロール済みの文書から有用な情報を抽出し、さらに、例えば文書をカテゴリーに関連付け、検索要求にある用語と文書の適合性を判断するときに役立つトークンを生成します。そして、索引コンポーネントは、効率的な検索が行われるようにデータを保管します。

Web ブラウザーと検索アプリケーションを使用することにより、ユーザーは索引処理済みのコレクションおよび外部ソースを検索します。検索アプリケーションは、ユーザーがブラウザーでクリックして利用できる結果リストを表示することができます。また、より高度に使用して、さまざまなデータ・ソースの情報に基づいて動的に生成されたコンテンツを戻すこともできます。

例えば、カタログ検索アプリケーションを使用すれば、検索要求を満たす結果表示をカスタマイズすることができます。1 回の照会で、例えば IBM DB2 Content Manager および Lotus Notes リポジトリにある文書の組み合わせなど、異なるタイプのデータ・ソースから文書を検索することができます。

管理者は、収集対象のデータを指定し、それらのデータのクローラ、解析、索引作成、および検索に関する設定を行います。システム・アクティビティをモニターすることにより、管理者は、必要な調整を行ってデータのスループットを最適にします。

次のダイアグラムは、エンタープライズ・サーチ・システムにおけるデータの流れを示しています。

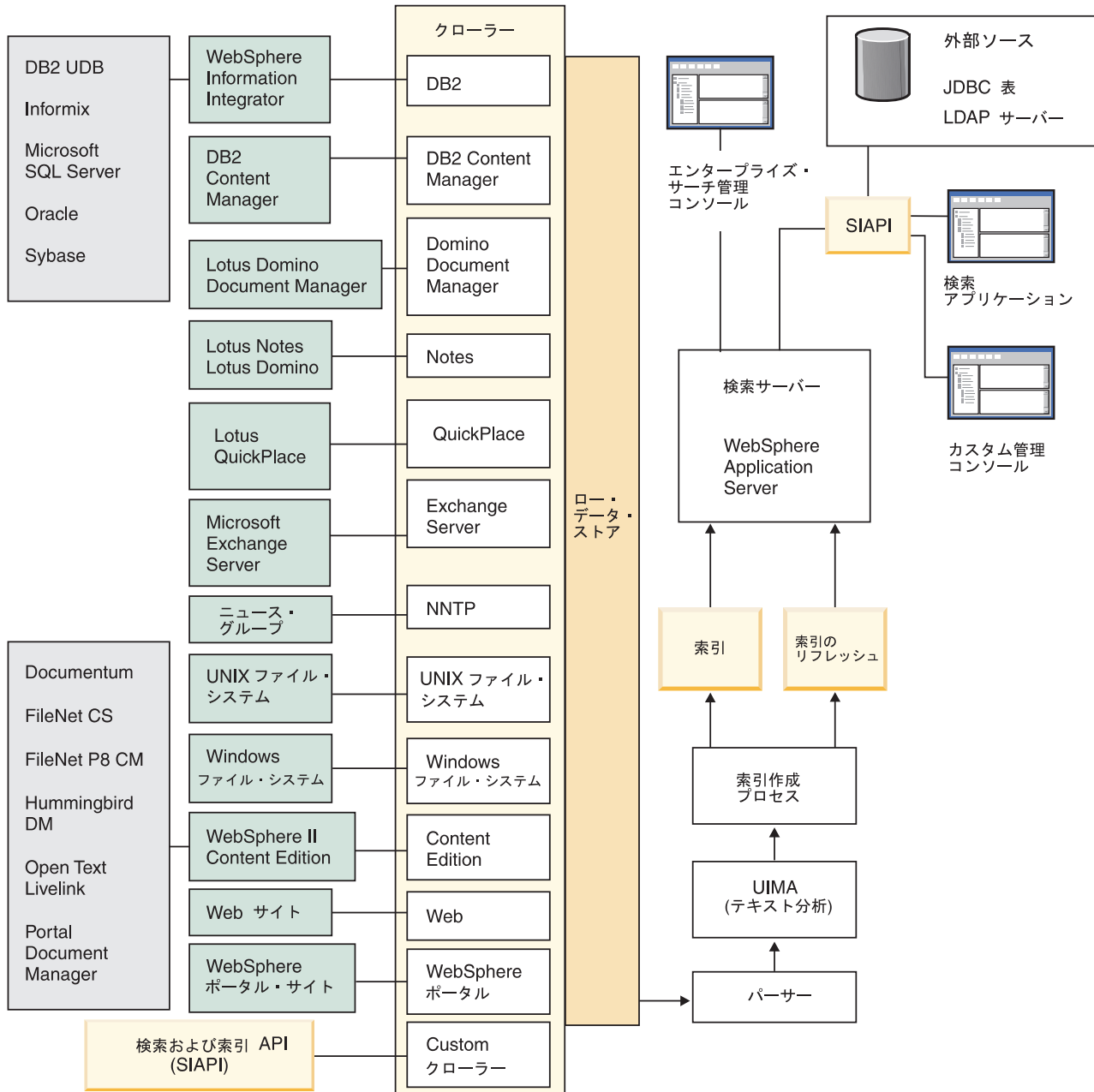


図 1. エンタープライズ・サーチ・システムにおけるデータ・フロー



---

## エンタープライズ・サーチ管理の概要


コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

### 「コレクション」ビュー

以下のガイドラインに従って、最初のコレクションを作成し、システムを管理します。

#### ログイン

エンタープライズ・サーチ管理コンソールにログインします。「コレクション」ビューは、コレクションの作成および管理のエントリー・ポイントです。

**ヒント:** 管理コンソールを使用する場合は、ツールバーの「ヘルプ」または「このページのヘルプ」をクリックします。メッセージに関する詳細情報および修正アクションがある場合は、メッセージの最後にある「 詳細」アイコンをクリックして、詳細を見ることができます。


#### コレクションの作成

以下のいずれかの方法を選択して、空の新規コレクションを作成します。

- コレクション・ウィザードを使用してコレクションを作成するには、「**コレクション・ウィザード**」をクリックし、ウィザードの指示に従ってください。
- 「コレクション」ビューを使用してコレクションを作成するには、「**コレクションの作成**」をクリックし、「コレクションの作成」ページの各フィールドに必要なデータを入力して「**OK**」をクリックします。

#### コレクションの構成

新規コレクションは空なので、編集してコンテンツを追加する必要があります。またデータをクロールするときやデータを検索可能にする場合のオプションを指定します。

作成した新規のコレクションに対して「 **編集**」をクリックし、続いてそのコレクションのオプションを指定するページを選択します。

**重要:** Web ブラウザーで「戻る」または「最新表示」ボタンをクリックすると、矛盾した結果が出たり、潜在的なデータ損失が起きる可能性があります。コレクション構成時に、前のページに戻ったり、管理コンソールの情報を最新表示する必要がある場合は、Web ブラウザーの「戻る」や「最新表示」ボタンではなく、エンタープライズ・サーチ管理コンソールの「**前へ**」および「**最新表示**」ボタンをクリックしてください。

- 「一般」 ページでは、コレクション全体に適用するオプションを指定します。
  - 一般オプションを編集して、コレクションの名前や説明を変更したり、あるいは、コレクションの見積もりサイズを変更することができます。
  - コレクション ID や検索結果で文書をランキングする静的ランキング方法などのコレクションに関する変更不可能な情報を表示することができます。
  - コレクション作成時にそのコレクションのセキュリティーを使用可能に設定している場合、文書レベルでセキュリティー制御を使用可能にしたり、使用不可にすることができます。
- 「クロール」 ページでは、クローラーを必ず 1 つは構成します。
 

1 つのコレクションには、さまざまな種類のデータ・ソースのデータを入れることができます。コレクションに組み込む各データ・ソースのタイプごとに、1 つ以上のクローラーを構成する必要があります。クローラーを作成する場合、クロール対象のデータ・タイプに固有のウィザードがあり、これを使用すると簡単にクローラーを構成できます。
- 「構文解析」 ページでは、クロール対象データの検索が効率的に行われるようにするための構文解析方法のオプションを構成します。
  - ネイティブ XML 検索で検索できるように XML 文書を構文解析するかどうかを指定することができます。
  - 文書をカテゴリーに関連付けることができます。こうすることでユーザーはコレクションのサブセットを検索したり、文書が含まれるカテゴリーを基準にして検索結果の文書を参照することができます。
  - XML エlementと HTML メタデータ・Elementを索引の検索フィールドにマップすると、ユーザーが照会にフィールド名を指定して文書の特定部分を検索することができるようになります。
  - エンタープライズ・サーチ・システムにカスタム・テキスト分析エンジンを追加した場合、コレクションで使用するものを 1 つ選択し、テキスト処理オプションを指定して情報の検索性を高めることができます。
  - フィールドをランキング調整クラスに関連付けて、照会用語に一致するフィールドの検索結果におけるランクを調整することができます。
- 「索引」 ページでは、索引の再編成およびリフレッシュのスケジュールを設定します。スケジュールは、ユーザーが常に最新の情報にアクセスできるようにするため、索引作成が頻繁に実行されるように設定してください。オプションで次のようなアクティビティーを実行することもできます。
  - ユーザーが照会用語にワイルドカード文字を指定できるようにします。
  - 有効範囲を設定し、ユーザーが索引のすべての文書ではなくコレクションの限定された一部分を検索できるようにします。
  - 検索結果を省略表示して、同じソースの文書をすべてグループ化して、検索結果にまとめて省略表示されるようにします。

- ランキング調整要因を割り当てて、URI パターンに一致する文書の検索結果におけるランクを調整します。
- 索引から URI を除去します。たとえば、コレクション作成後に、特定の文書をユーザーが表示できないようにする必要がある場合に行います。
- 「検索」ページでは、コレクションの文書を検索するためのオプションを指定します。
  - 検索結果用のキャッシュ・スペースを確保し、またコレクションのデフォルト言語を変更できます。
  - エンタープライズ・サーチ・システムに、同義語、ストップワード、ランキング調整ワードに関するカスタム辞書を追加した場合、コレクションの検索時に使用する辞書を選択できます。
  - 検索結果に表示する文書サマリーの長さを指定できます。
  - 照会に特定のキーワードまたは句が含まれているときに必ず特定の URI が自動的に表示されるように、クイック・リンクを構成することができます。
- 「ログ」ページでは、以下のアクティビティーを実行することができます。
  - ログに記録するメッセージのタイプと、ログ・ファイルのリサイクル頻度を指定します。
  - コレクション・アクティビティーに関するアラートを受信する場合のオプションを指定します。例えば、検索の平均応答時間が指定した限度を超える場合にアラート通知されるように設定することができます。
  - また、特定のメッセージあるいは特定タイプのメッセージがログに記録された場合に、必ず E メールを受信するようにするオプションを指定できます。
  - 文書がエンタープライズ・サーチ・システムから除去された時間や場所や理由を判別するためのロギング情報のオプションを指定します。

## コンポーネントの始動

クロール対象のデータ・ソース、およびデータ収集、検索に関するオプションを指定すると、コレクション作成プロセスを開始することができます。コンポーネントを開始する順序は重要です。クローラーはデータが構文解析される前にクロールする必要があり、パーサーはクロールされたデータが索引付けされる前に分析する必要があり、索引は検索サーバーが検索要求処理を開始する前にリフレッシュあるいは再編成される必要があります。

### 「外部ソース」ビュー

データ・ソースをクロールまたは索引付けせずに検索する場合は、ツールバーの「外部ソース」をクリックして、データ・ソースを検索可能にするオプションを指定できます。ご使用の Java Database Connectivity (JDBC) データベースおよび Lightweight Directory Access Protocol サーバーをエンタープライズ・サーチでアクセスできるように設定する情報を指定する必要があります。外部ソースと検索アプリケーションを関連付けると、クロール、構文解析、索引付けされたデータを使用

してコレクションを検索できると同時に、関連付けた外部ソースを検索することができます。

## 「システム」ビュー

エンタープライズ・サーチ管理者の役割を担当している場合、ツールバーの「システム」をクリックして、以下のアクティビティーを行うことができます (コレクション管理者、オペレーター、およびモニター担当者は、エンタープライズ・サーチ管理者によってアクセス権限が付与されている場合にのみ、このビューにアクセスできます)。

- システム・リソースのアベイラビリティを確認する。
- クライアント・データ・リスナー・アプリケーションがコレクションを更新できるように構成する。
- カスタム・テキスト分析エンジンをシステムに追加する。
- システムに、同義語、ストップワード、ランキング調整ワードに関するカスタム辞書を追加する。
- 並行して索引を作成できるコレクションの数を指定する。また、単一のコレクションに対するリフレッシュおよび再編成要求を並行して実行できるかどうかを指定する。
- システム・レベルのイベントに対するアラートを構成する。
- システム・レベル・セッションで生成されるロギング・メッセージのオプションを指定する。
- メール・サーバーに関する情報を指定して、エンタープライズ・サーチ・アクティビティーに関する E メールを受信できるようにする。


## 「セキュリティ」ビュー

エンタープライズ・サーチ管理者の役割を担当している場合、「セキュリティ」をクリックしてセキュリティ・オプションを指定することができます。コレクション管理者、オペレーター、およびモニター担当者は、このビューにはアクセスできません。

IBM WebSphere Application Server でセキュリティを使用可能に設定している場合、「セキュリティ」ビューを使用して管理役割を構成することができます。管理役割を構成することで、システム管理を実施できるユーザーを増やすと同時に、それぞれのユーザーがアクセスできる対象を、特定の機能およびコレクションに制限することができます。

独自の検索アプリケーションを作成するまで、サンプル検索アプリケーションを使用して、あらゆるコレクションおよび外部ソースの検索ができます。カスタム検索アプリケーションの作成後は、「セキュリティ」ビューを使用して、アプリケーションを、アプリケーションで検索可能なコレクションおよび外部ソースに関連付けてください。

## 「モニター」ビュー

 「モニター」をクリックすると、システムまたはコレクション・コンポーネントをいつでもモニターできます。管理役割で権限が与えられていれば、モニターするだけでなく、コンポーネント・プロセスの開始および停止もできます。

## 関連概念

37 ページの『エンタープライズ・サーチ・クローラーの管理』

コレクションに入れるデータ・タイプに応じて、さまざまなタイプのクローラーを構成できます。1 つのコレクションに任意の数のクローラーを入れることができます。

257 ページの『エンタープライズ・サーチ・アクティビティのモニター』

システム・アクティビティおよびコレクション・アクティビティをモニターすると、さまざまなプロセスの状況の表示、問題発生の可能性の監視、または、パフォーマンス改善のための構成設定の調整を行うことができます。

## 関連タスク

251 ページの『エンタープライズ・サーチ・サーバーの開始』

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

253 ページの『エンタープライズ・サーチ・サーバーの停止』

エンタープライズ・サーチ・コンポーネントの構成に変更を加えた場合や、問題のトラブルシューティングが必要な場合などには、エンタープライズ・サーチ・サーバーを停止して再始動する必要があります。

コレクション・ウィザードを使用したコレクションの作成

エンタープライズ・サーチにまだ慣れていなければ、ウィザードのガイドに従いながら、コレクションを作成できます。ウィザードは、処理の各ステップの詳細を提供し、途中で設定を保管できます。

31 ページの『「コレクション」ビューを使用したコレクションの作成』

「コレクション」ビューを使用して、空のコレクションを作成します。その後、このコレクションを編集して、コレクションにデータを追加したり、コレクションを検索可能にするオプションを指定できます。

---

## 管理コンソールへのログイン

エンタープライズ・サーチ・システムを管理するには、Web ブラウザーで URL を指定した後で、管理コンソールにログインします。

### 始める前に

ログインに使用するユーザー ID は、エンタープライズ・サーチ管理コンソールへのアクセスを認可されている必要があります。

- WebSphere Application Server でグローバル・セキュリティーを使用可能にしていない場合、管理コンソールにアクセスできるのは、WebSphere II OmniFind Edition のインストール時に指定されたエンタープライズ・サーチ管理者だけです。
- WebSphere Application Server でグローバル・セキュリティーを使用可能に設定してある場合、エンタープライズ・サーチ管理コンソールを使用して、管理役割を構成できます。構成するユーザー ID は、WebSphere Application Server ユーザー・レジストリー内に存在していなければなりません。管理役割を構成すると、より多くのユーザーが管理コンソールにログインするのを許可することになりますが、各管理ユーザーがアクセス可能な機能およびコレクションは制御することができます。

## 手順

エンタープライズ・サーチ管理コンソールにログインするには、次のようにします。

1. Web ブラウザーで、管理コンソールの URL を入力します。以下に例を示します。

`http://SearchServer.com/ESAdmin/`

`SearchServer.com` は、エンタープライズ・サーチ用の検索サーバーのホスト名です。

ご使用の Web サーバー構成によっては、ポート番号を指定する必要がある場合もあります。以下に例を示します。

`http://SearchServer.com:9080/ESAdmin/`

2. 「ようこそ」ページで、ユーザー ID とパスワードを入力し、「**ログイン**」をクリックします。

システムおよびコレクションの管理の入り口点である「コレクション」ビューが表示されます。管理役割を使用する場合、実行できるアクションと表示されるコレクションは、使用する管理役割によって決まります。

一定の時間セッションが非アクティブな状態が続くと、システムによって自動的にログアウトされます。システムの管理を続けるには、もう一度ログインしてください。

コレクションの管理の終了後、「**ログアウト**」をクリックすると、コンソールをログアウトできます。その後、別の ID とパスワードを使ってログインすることもできますし、Web ブラウザーを閉じて管理コンソールを終了することもできます。

### 関連概念

212 ページの『管理役割』

エンタープライズ・サーチは管理コンソールのさまざまな機能へのアクセスを制御するために、役割の概念を使用します。

### 関連タスク

251 ページの『エンタープライズ・サーチ・サーバーの開始』

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

---

## シングル・サーバー構成のエンタープライズ・サーチ管理者パスワードの変更

エンタープライズ・サーチ管理者のパスワードは、暗号化されたフォーマットで保管されます。パスワードを変更するには、`eschangepw` スクリプトを使用してください。

### 始める前に



エンタープライズ・サーチ管理者の ID とパスワードは、ご使用のオペレーティング・システムで有効であり、さらに DB2 Universal Database にアクセスして構成する権限がなければなりません。

### このタスクについて

エンタープライズ・サーチ管理者 ID の初期パスワードは、WebSphere II OmniFind Edition のインストール時に指定されます。

このパスワードを変更するためには、`eschangepw` スクリプトを実行して、エンタープライズ・サーチ・システム全体に変更を周知させる必要があります。インストール・プログラムでは、`eschangepw` スクリプトで使用できる、次の 2 つの環境変数を生成します。

#### ES\_INSTALL\_ROOT

エンタープライズ・サーチのインストール・ディレクトリー

#### ES\_NODE\_ROOT

エンタープライズ・サーチのデータ・ディレクトリー。エンタープライズ・サーチ管理者 ID のパスワードは、このディレクトリーの `es.cfg` ファイルに保管されます。

`eschangepw` スクリプトは、`ES_INSTALL_ROOT/bin` ディレクトリーにインストールされるため、システムのどこからでも、それを実行できます。

### 手順

シングル・サーバー構成のエンタープライズ・サーチ管理者パスワードを変更するには、以下のようにします。

1. エンタープライズ・サーチ管理者としてログインします。
2. 次のコマンドを入力してサーバーを停止します。 `esadmin stop`
3. WebSphere Application Server 管理コンソールを開いて、エンタープライズ・アプリケーション `server1` および `ESSearchServer` を停止します。
4. UNIX オペレーティング・システムのコマンド、または Microsoft Windows のパスワード変更機能を使用して、エンタープライズ・サーチ管理者のユーザー ID のシステム・パスワードを変更します。
5. 以下のスクリプトを実行します。ここで、*newValue* は、ステップ 4 で指定したパスワードです。

オペレーティング・システム	コマンド
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

6. WebSphere Application Server 管理コンソールで、アプリケーション `server1` および `ESSearchServer` を開始します。
7. 以下のコマンドを入力することによって、WebSphere II OmniFind Edition 共通通信層 (CCL) を再生します。

オペレーティング・システム	コマンド
UNIX	<code>stopccl.sh</code> 、次に <code>startccl.sh -bg</code>

オペレーティング・システム	コマンド
Windows コマンド・プロンプト	stopccl、次に startccl
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>1. Windows Services を起動する。</li> <li>2. 「WebSphere Information Integrator OmniFind Edition」を右クリックし、「停止」を選択する。</li> <li>3. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「プロパティ」を選択する。</li> <li>4. 「ログオン」タブをクリックする。</li> <li>5. <i>newValue</i> を指定してパスワードを変更してから、「OK」をクリックする。</li> <li>6. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「開始」を選択する。</li> </ol>

8. 次のコマンドを入力して、エンタープライズ・サーチを再始動します。 `esadmin start`

#### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

## マルチサーバー構成のエンタープライズ・サーチ管理者パスワードの変更

エンタープライズ・サーチ管理者のパスワードは、暗号化されたフォーマットで保管されます。このパスワードを変更するには、ご使用のエンタープライズ・サーチ・システムのすべてのコンピューターで `eschangepw` スクリプトを実行して変更してください。

#### 始める前に

エンタープライズ・サーチ管理者の ID とパスワードは、ご使用のオペレーティング・システムで有効であり、さらに DB2 Universal Database にアクセスして構成する権限がなければなりません。

エンタープライズ・サーチ管理者のパスワードは、WebSphere II OmniFind Edition のこのインストール環境に属するすべてのコンピューターで同一である必要があります。

#### このタスクについて

エンタープライズ・サーチ管理者 ID の初期パスワードは、WebSphere II OmniFind Edition のインストール時に指定されます。



パスワードを変更し、その変更をエンタープライズ・サーチ・システム全体に周知させるには、マルチサーバー構成の各コンピューターで、`eschangepw` スクリプトを実行する必要があります。インストール・プログラムでは、`eschangepw` スクリプトで使用できる、次の 2 つの環境変数を生成します。

#### **ES\_INSTALL\_ROOT**

エンタープライズ・サーチのインストール・ディレクトリー

#### **ES\_NODE\_ROOT**

エンタープライズ・サーチのデータ・ディレクトリー。エンタープライズ・サーチ管理者 ID のパスワードは、このディレクトリーの `es.cfg` ファイルに保管されます。

`eschangepw` スクリプトは、`ES_INSTALL_ROOT/bin` ディレクトリーにインストールされるため、システムのどこからでも、それを実行できます。

#### **手順**

マルチサーバー構成のエンタープライズ・サーチ管理者パスワードを変更するには、以下のようにします。

1. 索引サーバーをインストールしたコンピューターで、エンタープライズ・サーチ管理者としてログインします。
2. 次のコマンドを入力して、システム保守モードを開始します。 `esadmin system maintenance`
3. UNIX オペレーティング・システムのコマンド、または Microsoft Windows のパスワード変更機能を使用して、エンタープライズ・サーチ管理者のユーザー ID のシステム・パスワードを変更します。
4. 以下のスクリプトを実行します。ここで、*newValue* は、ステップ 3 で指定したパスワードです。

オペレーティング・システム	コマンド
<b>UNIX</b>	<code>eschangepw.sh newValue</code>
<b>Windows</b>	<code>eschangepw newValue</code>

5. 索引サーバーで以下のコマンドを入力することによって、WebSphere II OmniFind Edition 共通通信層 (CCL) を再生します。

オペレーティング・システム	コマンド
<b>UNIX</b>	<code>stopccl.sh</code> 、次に <code>startccl.sh -bg</code>
<b>Windows</b> コマンド・プロンプト	<code>stopccl</code> 、次に <code>startccl</code>

オペレーティング・システム	コマンド
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>1. Windows Services を起動する。</li> <li>2. 「WebSphere Information Integrator OmniFind Edition」を右クリックし、「停止」を選択する。</li> <li>3. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「プロパティ」を選択する。</li> <li>4. 「ログオン」タブをクリックする。</li> <li>5. <i>newValue</i> を指定してパスワードを変更してから、「OK」をクリックする。</li> <li>6. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「開始」を選択する。</li> </ol>

6. クローラー・サーバーをインストールしたコンピュータで、エンタープライズ・サーチ管理者としてログインします。
7. UNIX オペレーティング・システムのコマンド、または Windows パスワード変更機能を使用して、エンタープライズ・サーチ管理者のユーザー ID のシステム・パスワードを変更します。このパスワードは、ステップ 3 (23 ページ) で指定したパスワードと一致している必要があります。
8. 以下のスクリプトを実行します。ここで、*newValue* は、ステップ 3 (23 ページ) で指定したパスワードです。

オペレーティング・システム	コマンド
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

9. 以下のコマンドを入力してクローラー・サーバーの CCL を再生します。

オペレーティング・システム	コマンド
UNIX	<code>stopccl.sh</code> 、次に <code>startccl.sh -bg</code>
Windows コマンド・プロンプト	<code>stopccl</code> 、次に <code>startccl</code>
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>1. Windows Services を起動する。</li> <li>2. 「WebSphere Information Integrator OmniFind Edition」を右クリックし、「停止」を選択する。</li> <li>3. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「プロパティ」を選択する。</li> <li>4. 「ログオン」タブをクリックする。</li> <li>5. <i>newValue</i> を指定してパスワードを変更してから、「OK」をクリックする。</li> <li>6. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「開始」を選択する。</li> </ol>

10. 検索サーバーをインストールしたコンピューターの 1 つで、エンタープライズ・サーチ管理者としてログインします。
11. IBM HTTP Web サーバーを停止します。(このステップにより Network Dispatcher が照会を該当サーバーにルーティングしないようにします。)
12. 以下のコマンドを入力して検索サーバーの CCL を停止します。

オペレーティング・システム	コマンド
UNIX	stopccl.sh
Windows コマンド・プロンプト	stopccl
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>1. Windows Services を起動する。</li> <li>2. 「WebSphere Information Integrator OmniFind Edition」を右クリックし、「停止」を選択する。</li> </ol>

13. WebSphere Application Server 管理コンソールを開いて、アプリケーション server1 および ESSearchServer を停止します。
14. UNIX オペレーティング・システムのコマンド、または Windows パスワード変更機能を使用して、エンタープライズ・サーチ管理者のユーザー ID のシステム・パスワードを変更します。このパスワードは、ステップ 3 (23 ページ) で指定したパスワードと一致している必要があります。
15. 以下のスクリプトを実行します。ここで、*newValue* は、ステップ 3 (23 ページ) で指定したパスワードです。

オペレーティング・システム	コマンド
UNIX	eschangepw.sh <i>newValue</i>
Windows	eschangepw <i>newValue</i>

16. 以下のコマンドを入力して検索サーバーの CCL を再始動します。

オペレーティング・システム	コマンド
UNIX	startccl.sh -bg
Windows コマンド・プロンプト	startccl
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>1. Windows Services を起動する。</li> <li>2. 「WebSphere Information Integrator OmniFind Edition」を右クリックし、「プロパティ」を選択する。</li> <li>3. 「ログオン」タブをクリックする。</li> <li>4. <i>newValue</i> を指定してパスワードを変更してから、「OK」をクリックする。</li> <li>5. 再度「WebSphere Information Integrator OmniFind Edition」を右クリックし、「開始」を選択する。</li> </ol>

17. 2 番目の検索サーバーで、ステップ 10 から 16 までを繰り返します。
18. IBM HTTP Web サーバーを開始します。

19. WebSphere Application Server 管理コンソールで、エンタープライズ・アプリケーション server1 および ESSearchServer を開始します。
20. 索引サーバーをインストールしたコンピュータで、エンタープライズ・サーチ管理者としてログインし、次のコマンドを入力してエンタープライズ・サーチを開始します: `esadmin start`
21. エンタープライズ・サーチ管理コンソールにログインして、コレクションをモニターし、「検索」ページをクリックしてから、「停止」と「開始」をクリックして検索サーバーを再始動します。ご使用のエンタープライズ・サーチ・システムのそれぞれのコレクションごとにこのステップを繰り返してください。

#### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## エンタープライズ・サーチ・システムのポート番号の変更

エンタープライズ・サーチ・システムが通信用に使用するポート番号が、別の製品が使用するポート番号と競合する場合、エンタープライズ・サーチのポート番号を変更する必要があります。

#### このタスクについて

エンタープライズ・サーチ・システムのポート番号は、WebSphere II OmniFind Edition のインストール時に指定されます (マルチサーバー構成の場合、すべてのサーバーで同じポート番号が指定されます)。

ポート番号が使用不能の場合 (例えば、そのポート番号が同じサーバー上の別の製品に割り当てられている場合など)、競合が原因となって、`CCLServer_date.log` ファイルに以下のエラー・メッセージが出力されます (ここで、`date` は、ファイルが作成された日付を指定しています)。

```
FFQ00273W 内部警告 java.net.BindException: アドレスは既に使用されています: NET_Bind
at java.net.PlainSocketImpl.socketBind(Native Method)
at java.net.PlainSocketImpl.bind(PlainSocketImpl.java:357)
at java.net.ServerSocket.bind(ServerSocket.java:341)
at java.net.ServerSocket.<init>(ServerSocket.java:208)
at java.net.ServerSocket.<init>(ServerSocket.java:120)
```

#### 手順

エンタープライズ・サーチが使用するポート番号を変更するには、次のようにします。

1. 以下のコマンドを入力して、エンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```

2. ポート番号を変更する必要があるサーバーに移動し、以下のコマンドを入力して共通通信層 (CCL) を停止します。

オペレーティング・システム	コマンド
UNIX	stopccl.sh,
Windows コマンド・プロンプト	stopccl
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>Windows Services を起動する。</li> <li>「WebSphere Information Integrator OmniFind Edition」を右クリックし、「停止」を選択する。</li> </ol>

- \$ES\_NODE\_ROOT/nodeinfo/es.cfg ファイル (UNIX) または %ES\_NODE\_ROOT%\nodeinfo\es.cfg ファイル (Windows) を編集し、以下のプロパティを見つけ、新しいポート番号値を指定し、ファイルを保管してクローズします。

```
CCLPort=new_port_number
```

- 以下のコマンドを入力して CCL を再始動します。

オペレーティング・システム	コマンド
UNIX	startccl.sh -bg
Windows コマンド・プロンプト	startccl
Windows Services 管理ツール	<ol style="list-style-type: none"> <li>Windows Services を起動する。</li> <li>「WebSphere Information Integrator OmniFind Edition」を右クリックし、「開始」を選択する。</li> </ol>

- 索引サーバーに移動し、ステップ 2 (26 ページ) の手順に従って CCL を停止します。

- \$ES\_NODE\_ROOT/nodeinfo/es.cfg ファイル (UNIX) または %ES\_NODE\_ROOT%\nodeinfo\es.cfg ファイル (Windows) を編集します。

- 以下のプロパティを見つけます。ここで、*computer\_name* は、ステップ 3 でポート番号を変更したサーバーの名前です。node*N* プロパティ内の *N* は、サーバーを識別する番号です。

```
nodeN.destination=computer_name
```

- 以下のサブプロパティを見つけ、ステップ 3 でサーバーに指定したのと同じポート番号を指定し、ファイルを保管してクローズします。

```
nodeN.port=new_port_number
```

- ステップ 4 の手順に従って、CCL を再始動します。
- 以下のコマンドを入力して、エンタープライズ・サーチ・システムを再始動します。

```
esadmin start
```

このコマンドが完了したら、すべてのエンタープライズ・サーチ・サーバーで新しいポート番号が更新されます。

#### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

| コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。  
|  
|

---

## エンタープライズ・サーチ・コレクション

エンタープライズ・サーチ・コレクションには、ユーザーが 1 回の照会で検索できる全ソースが格納されています。フェデレーションによって、ユーザーは複数のコレクションを 1 回の照会で検索することができます。

コレクションを作成する場合、コレクション全体に適用されるオプションを指定します。コレクションは、コンテンツを追加するまでは空の状態です。

コレクションをエンタープライズ・サーチ・システムに追加する方法には、次の 2 つの方法があります。

- エンタープライズ・サーチ管理コンソールの操作に不慣れな場合、また、さまざまなコレクション・コンポーネントが連携してどのように機能するかが、まだよくわからない場合は、コレクション・ウィザードを使用してコレクションを作成してください。コレクション・ウィザードを使用すれば、作成タスクを順序良く実行することができ、ドラフト・コレクションで作業する手間が省けます。
- 管理コンソールの操作に慣れている場合は、「コレクション」ビューで管理を行う特定のページを選択しながらコレクションを作成する方法もあります。

コレクションの作成後は、「コレクション」ビューのコントロールを使用して、コレクション、エンタープライズ・サーチ・システム、およびセキュリティー・オプションの編集とモニターをすることができます。

### コレクション・フェデレーション

フェデレーションのサポートが検索アプリケーションに組み込まれて構築されている場合、ユーザーは同時に複数のコレクションを検索できます。また、フェデレーションにより、コレクションのサイズ制限 (コレクションあたり 20 000 000 文書) を拡大することができます。例えば、それぞれ 20 000 000 個の文書を含む 2 つのコレクションを検索できます。

検索の質は、マージされて最終的な結果セットを構成する前の、個別のコレクションで生成されたスコアに依存します。結果は、2 つの別々の検索を実行してから、その結果をマージしてランキングした場合と同じになります。

#### 関連タスク

259 ページの『コレクションのモニター』

コレクション内の各コンポーネントの状態に関する一般情報を表示するか、または、個々のコンポーネントおよび URI に関する詳細情報を表示するオプションを選択することができます。

245 ページの『WebSphere Portal からのコレクションのマイグレーション』

WebSphere Portal からエンタープライズ・サーチにコレクションをマイグレーションするには、WebSphere Portal にコレクションを準備し、マイグレーション・ウィザードを使用して準備したコレクションをマイグレーションします。

---

## コレクション・ウィザードを使用したコレクションの作成

エンタープライズ・サーチにまだ慣れていなければ、ウィザードのガイドに従いながら、コレクションを作成できます。ウィザードは、処理の各ステップの詳細を提供し、途中で設定を保管できます。

### 始める前に

コレクションを作成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

コレクションにコンテンツを追加したり、コレクション内のコンテンツの構文解析、索引付け、または検索方法のオプションを指定したりするには、エンタープライズ・サーチ管理者であるか、そのコレクションのコレクション管理者でなければなりません。

### このタスクについて

コレクションの作成途中で、それをドラフト状態として保管できます。それがドラフト状態にある間、コレクションを管理する権限を持つ管理者は、それを変更することができます。例えば、Lotus Notes ソースに詳しいコレクション管理者が Lotus Notes クローラーを構成します。その後、ドラフト状態のそのコレクションを、UNIX システムに詳しい管理者が編集して UNIX ファイル・システム クローラーを構成します。


### 手順

コレクション・ウィザードを使用してコレクションを作成するには、次のようにします。

1. 「**コレクション**」をクリックして「**コレクション**」ビューを開きます。
2. 「**コレクション・ウィザード**」をクリックします。
3. ウィザードの指示に従って、空のコレクションを作成し、それにコンテンツを追加します。

コレクションについての一般情報を構成し、少なくとも 1 つのクローラーを作成しなければなりません。その他の構成オプションについては、デフォルト値を受け入れることも、あるいは、新規コレクションのオプションを指定することもできます。

4. コレクションの作成が完了する前にコレクションを保管するには、「**ドラフトとして保管**」をクリックします。

そのコレクションが他のドラフト・コレクションと一緒に「**コレクション**」ビューにリストされます。コレクションのセキュリティーを使用可能にしてある場合、「 **コレクション・セキュリティー使用可能**」アイコンがコレクション名の横に表示されます。

5. 作成途中のコレクションに戻るには、「**コレクション**」ビューで、「**ウィザードに戻る**」をクリックします。
6. 「**完了**」をクリックして、コレクションを作成します。



新規コレクションが他のコレクションと一緒に「コレクション」ビューにリストされます。

コレクションを作成したあと、コレクションのクローल、構文解析、索引付け、および検索のプロセスを開始しなければなりません。コレクションを検索できる検索アプリケーションにコレクションを関連付ける準備ができるまで、サンプル検索アプリケーション (名前は Default) を使用して、新規コレクションを検索できます。

---

## 「コレクション」ビューを使用したコレクションの作成

「コレクション」ビューを使用して、空のコレクションを作成します。その後、このコレクションを編集して、コレクションにデータを追加したり、コレクションを検索可能にするオプションを指定できます。

### 始める前に

コレクションを作成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

コレクションにコンテンツを追加したり、コレクション内のコンテンツの構文解析、索引付け、または検索方法のオプションを指定したりするには、エンタープライズ・サーチ管理者であるか、そのコレクションのコレクション管理者でなければなりません。

### このタスクについて

新規コレクションに指定できる値の情報は、コレクションを作成している間に「ヘルプ」をクリックすると表示されます。


### 手順

「コレクション」ビューからコレクションを作成するには、次のようにします。

1. 「コレクション」ビューで、「**コレクションの作成**」をクリックします。
2. 「コレクションの作成」ページで、次のフィールドに情報を入力するか、選択を行います。
  - **コレクション名**。コレクションのコンテンツまたは目的の記述名を指定します。
  - **コレクション・セキュリティー**。コレクションのセキュリティーを使用可能にするかどうかを指定します。コレクションを作成した後、この値を変更することはできません。コレクション・セキュリティーが使用可能であれば、あとで、文書レベルのアクセス制御を実行するためのオプションを指定できます。
  - **文書の重要度 (静的ランキング・モデル)**。検索結果での文書のランク付けに使用される静的ランキング要素を割り当てるための方法を指定します。コレクションの作成後、この値を変更することはできません。
  - **カテゴリー化タイプ**。文書を、所属するカテゴリーで検索できるようにするかどうかを指定します。
  - **使用する言語**。コレクション内の文書検索のデフォルト言語を指定します。
3. 次のフィールドについては、デフォルト値を受け入れることも、あるいは、このコレクションで使用するオプションを指定することもできます。

- **説明。**デフォルトでは、説明は作成されません。
- **見積もり文書数。**コレクションのデフォルトの見積もり文書数は 1 000 000 です。システムは、この値を使用して、コレクションのメモリーとディスク・リソースを見積もります。コレクションのサイズを制限することはありません。
- **コレクション・データの場所。**コレクション関連ファイルのデフォルトの場所は、索引サーバー上にあります。コレクションの作成後、この値を変更することはできません。
- **コレクション ID。**デフォルトのコレクション ID は、コレクション名に基づきます。コレクションの作成後、この値を変更することはできません。(カスタムのコレクション ID を指定すると、検索アプリケーションは、システムが作成する意味の分かりにくい ID の代わりにこの ID を使用してコレクションを呼び出します。)
- **N-gram セグメンテーション。**デフォルトのセグメンテーション方式は、Unicode ベースの、空白文字セグメンテーションです。このオプションは、コレクションに中国語、日本語、または韓国語の文書が含まれており、ワードを区切るのにパーサーに N-gram セグメンテーションを使用させたい場合のみ、選択します。コレクションの作成後、この値を変更することはできません。

#### 4. 「OK」をクリックします。

「コレクション」ビューに、新規コレクションが、エンタープライズ・サーチ・システム内の他のコレクションと一緒にリストされます。コレクションのセキュリティを使用可能にしてある場合、「 **コレクション・セキュリティ使用可能**」アイコンがコレクション名の横に表示されます。

コレクションは、コンテンツを追加するまでは空の状態です。新規コレクションにコンテンツを追加するには、「コレクション」ビューでコレクションを選択し、それを編集し、少なくとも 1 つのクローラーを作成します。そして、データの構文解析、索引付け、および検索の方法を指定します。

そして、コレクションのクロール、構文解析、索引付け、および検索のプロセスを開始しなければなりません。カスタム検索アプリケーションでユーザーが新規コレクションを検索できるようにする準備を終えるまでは、新規コレクションの検索にサンプル検索アプリケーションを使用できます。

#### 関連概念

15 ページの『エンタープライズ・サーチ管理の概要』

コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティーおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

---

## コレクションの編集

コレクションに入れる文書に関する情報を指定するために、コレクションを編集します。

### 始める前に


コレクションを編集するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

コレクションを編集するときに、データ・ソースのクロール、文書の構文解析、索引の再編成とリフレッシュ、索引付きコンテンツの検索、およびエラー・メッセージのログ記録のオプションを指定します。コレクションを作成するときに、コレクションを編集して、それにコンテンツを追加しなければなりません。あとでコレクションを編集して、コンテンツを更新したり、情報のクロール、構文解析、索引付け、検索、またはログの方法を変更したりできます。

### 手順

コレクションを編集するには、次のようにします。

1. ツールバーの中の「**コレクション**」をクリックして「コレクション」ビューを開きます。
2. コレクションのリストで編集したいコレクションを指して、「 **編集**」をクリックします。
3. 次のページで変更を行います。

**一般** コレクションに関する一般情報を指定し、変更できない設定を確認します。コレクションの作成時にそのコレクションに対してセキュリティーを使用可能にすると、文書レベルのセキュリティーのオプションを構成することができます。

### クロール

クロールするデータ・ソースを指定し、コンテンツのクロール方法のオプションを指定します。すべてのコレクションは、それぞれ少なくとも 1 つのクロールを含む必要があり、1 つのコレクションに、複数のタイプのデータ・ソースからのデータを含めることができます。コレクションに含めるデータ・ソースのタイプごとに、少なくとも 1 つのクロールを構成しなければなりません。

### 構文解析

クロールされた文書の構文解析および分析方法のオプションを指定します。カテゴリーを構成することで、ユーザーがコレクションのサブセットを検索できるようにすることができます。また、XML 文書および HTML 文書の特定の部分をユーザーが検索できるようにする規則を構成できます。カスタム・テキスト分析エンジンをエンタープライズ・サーチ・システムに追加する場合、その 1 つをこのコレクション内のコンテンツの分析と注釈付けに使用するよう選択できます。また、フィールドをランキング調整クラスと関連付けて文書ランキングを操作することができます。

- 索引** 全体の索引を再編成するスケジュールと、新規および変更されたコンテンツで索引をリフレッシュするスケジュールを指定します。照会でワイルドカード文字を使用するためのオプション、コレクションのビューを一定範囲の URI に制限するためのオプション、同じ Web サイトからの検索結果を縮小表示するためのオプション、索引から URI を除去するためのオプションも構成できます。
- 検索** 検索キャッシュの構成、検索言語の選択といった、コレクションの検索処理オプションを指定します。また、クイック・リンクを構成することもできます。これは、ユーザーが照会の中に特定の語または句を含めたときは、必ず事前に決められた URI が戻されるようにする機能です。エンタープライズ・サーチ・システムにカスタム同義語辞書を追加した場合、それらの辞書をこのコレクションの検索に使用するよう選択できます。
- ログ** ログに記録するメッセージのタイプと、ログ・ファイルの作成と循環のオプションを指定します。特定のイベントが発生した時に通知されるようにアラートを構成でき、特定のメッセージまたは特定のタイプのメッセージがログに記録されたら必ず E メールを受信するオプションを指定できます。また、エンタープライズ・サーチ・システムからドロップされた文書に関する情報をログに記録するオプションを指定することもできます。

---

## コレクションの削除

コレクションを削除すると、そのコレクションに関するすべての情報がエンタープライズ・サーチ・システムから除去されます。

### 始める前に

コレクションを削除するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

コレクションを削除するには、その前に、そのコレクションに関連付けられたすべてのプロセスを停止しなければなりません。


### このタスクについて

コレクションの削除は、処理に時間がかかることがあります。コレクションの削除を確認すると、システムは、システムにあるコレクションに関連するすべてのデータを削除します。

**ヒント:** プロセスがバックグラウンドでまだ実行されているにもかかわらず、要求された操作がタイムアウトになったというメッセージが表示される場合があります。タスクが完了したかどうかを判断するには、管理コンソールで「リフレッシュ」をクリックします (Web ブラウザーの「リフレッシュ」はクリックしないでください)。削除処理は、コレクション名がコレクション・リストに表示されなくなったときに完了します。

### 手順

コレクションを削除するには、次のようにします。

1. 「コレクション」をクリックして「コレクション」ビューを開きます。
2. コレクションのリストで、削除したいコレクションを指して、「 削除」をクリックします。



---

## エンタープライズ・サーチ・クローラーの管理

コレクションに入れるデータ・タイプに応じて、さまざまなタイプのクローラーを構成できます。1つのコレクションに任意の数のクローラーを入れることができます。

### クローラーの構成

クローラーの作成、編集、および削除には、エンタープライズ・サーチ管理コンソールを使用します。クローラーの構成は、通常、クロール対象のデータ・タイプに精通した担当者が行います。例えば、Lotus Notes データ・ソースをクロールするクローラーをセットアップする場合、コレクション管理者には Notes 管理者がなるか、または、クロール対象のデータベースを十分理解している担当者と緊密な協力を得ながら作業を行います。

クローラーを作成する場合、クロール対象のデータ・タイプ専用のウィザードがあり、これを使用するとクローラーによるシステム・リソースの使用を制御するプロパティを指定できます。また、このウィザードによって検索対象のソースを選択することもできます。

既存のクローラーは、いつでも変更できます。クローラー・プロパティの編集またはクロール・スペースの一部変更を必要に応じて行うことができます。クローラー・ウィザードを使用すれば、これらの変更を行うこともできます。

### 新規のクローラーにベースとなる値を移植する

クローラーは、システム・デフォルト値を使用するか、または同じタイプの別のクローラーに指定されている値をコピーして作成することができます。既存のクローラーを新規クローラーのベースとして使用すると、類似したプロパティを持つ複数のクローラーを手早く作成でき、それらのクローラーを構成して異なるソースをクロールするようにしたり、また、例えば、異なるクロール・スケジュールで運用したりすることができます。

同一のクローラーをコピーすれば、同じクロール規則に従う複数のクローラーにクロールの負荷を分散することができます。例えば、異なる Lotus Notes サーバーで同じプロパティおよびフィールド・クロール規則を使用する場合、同一の Lotus Notes クローラーをコピーすることが考えられます。この場合、各クローラーで異なるのは、クロール対象のデータベースと文書レベルのセキュリティ設定だけとなります。

### 単一コレクション内で異なるクローラー・タイプを組み合わせる

エンタープライズ・サーチ・クローラーは、特定タイプのデータ・ソースから情報を収集するように設計されています。コレクションに対してクローラーを構成する場合、さまざまな異なるデータ・ソース・タイプをどのように組み合わせればユーザーが簡単にエンタープライズ・データを検索できるか、その組み合わせを決定する必要があります。例えば、ユーザーが Microsoft Windows ファイル・システムと Microsoft Exchange Server のパブリック・フォルダーを 1 回の照会で検索できるよ



うにするには、Windows ファイル・システム・クローラーと Exchange Server クローラーを含むコレクションを作成します。

1 つのコレクションに複数のタイプのクローラーを組み合わせる場合、必ずすべてのクローラーが同一の静的ランキング方式を使用できることを確認してください。(静的ランキング方式は、コレクション作成時に指定します。) 例えば、Web ソース (ランキング係数として文書リンクを使用) と NNTP ソース (ランキング係数として、通常、文書日付を使用) を組み合わせてしまうと、検索結果の品質が低下する可能性があります。

## 文書レベル・セキュリティ

コレクションの作成時に、そのセキュリティを使用可能にすると、文書レベルのセキュリティ・オプションを構成することができます。各クローラーは、クローラー対象の文書にセキュリティ・トークンを関連付けることができます。クローラーの構成時に文書レベルのセキュリティを使用することを指定すると、そのクローラーは指定したセキュリティ・トークンを各文書に関連付けます。そして、これらのトークンは文書の索引に追加されます。

カスタム検索アプリケーションでセキュリティを使用可能にした場合、そのアプリケーションはクローラーが文書に関連付けたセキュリティ・トークンを使用してユーザーを認証します。この機能を使えば、コレクション内の一部の文書をアクセス制限し、他の文書はすべてのユーザーが検索できるように設定することができます。例えば、あるコレクションで、Microsoft Exchange Server のパブリック・フォルダー内の全文書をすべてのユーザーがアクセスできるようにし、Lotus Notes データベース文書にアクセスできるのは特定のユーザー ID のユーザーだけに限定するように設定できます。

規則を Java クラスでエンコードすることにより、セキュリティ・トークンの値の判別にカスタム・ビジネス・ルールを適用することができます。クローラー・プロパティを構成するときに、文書をクローラーする際にクローラーに使用させたいプラグインの名前を指定します。プラグインが追加するセキュリティ・トークンは、索引に保管され、文書へのアクセス制御に使用されます。

特定のタイプのクローラーを構成する場合、さらにセキュリティ制御を追加指定することができます。例えば、照会処理中にユーザーの検証を行うように指定することができます。このオプションを設定すると、ユーザーの信用証明情報は検索対象のデータ・ソースで維持されている現在のアクセス制御リストと比較されます。この現行の信用証明情報に関する検証は、エンタープライズ・サーチ索引に保管されているセキュリティ・トークンに基づく検証の代わりに、あるいはその検証に加えて行うことができます。

## クローラーのスケジューリング

Web、NNTP、および WebSphere Portal ソースに対して作成したクローラーは、連続的に実行されます。このようなクローラーについては、クローラーをいったん始動後は、通常、クローラーの構成を変更しない限り特に停止する必要はありません。

それ以外のすべてのクローラー・タイプについては、クローラーの構成時にクローラー・スケジュールを指定します。データ・タイプによっては、1 つのスケジュール

で、クローラーがクロール・スペース内のすべてのデータ・ソースにアクセスする日時が管理されます。また、他のデータ・ソース・タイプの場合は、特定のデータ・ソースに対して複数の異なるスケジュールを指定できます。例えば、Lotus Notes データベースについて、クローラーがクロールする各データベースに異なるクロール・スケジュールを指定できます。

スケジュールを構成する際に、行われるクロールのタイプを指定します。クロール・スペース内の全文書の完全クロールをスケジュールしたり、クロール・スペースに対するすべての更新 (新規文書、変更された文書、削除された文書) を対象とするクロールをスケジュールしたり、新規および変更された文書のみを対象とするクロールをスケジュールすることができます。完全クロールは、最も時間がかかります。削除された文書を除去するクロールは、削除された文書を無視するクロールに比べて時間がかかります。

クローラーのクロール・スペースを編集する際に、2 番目のクロール・スケジュールを指定することができます。例えば、毎週土曜日の夜にクロール・スペース内のすべての文書をクロールするスケジュールを構成し、新規および変更された文書のクロールをもっと頻繁に実行する 2 番目のスケジュールを構成することができます。

複数のクローラー・スケジュールを作成すれば、クローラーのターゲット・ソースへのアクセス日時をより適切に管理できます。例えば、異なるタイム・ゾーンにある複数のデータベースをクロールする場合、ユーザーが日中の仕事を終了していると思われる時間にクローラーがクロールするように、タイム・ゾーンに合わせて複数のスケジュールを設定できます。

### 関連概念

15 ページの『エンタープライズ・サーチ管理の概要』

コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティーおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

221 ページの『文書レベル・セキュリティ』

コレクションが作成されたときに、そのコレクションに対するセキュリティを使用可能にすると、文書レベルのセキュリティ制御を構成することができます。文書レベルのセキュリティによって、コレクションを検索するユーザーは、見ることが許される文書にしかアクセスできないようになります。

### 関連タスク

262 ページの『クローラーのモニター』

コレクション内の各クローラーの状態に関する一般情報を表示するか、または、クローラーのアクティビティーに関する詳細情報を表示するオプションを選択することができます。

### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』

エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

---

## クローラーの作成

クローラーを作成するとき、作成するクローラーのタイプを指定します。コレクションに含めるデータについて、指定する情報をウィザードがガイドします。

### 始める前に

クローラーを作成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのクローラーを所有するコレクションのコレクション管理者でなければなりません。

### このタスクについて

1 つのコレクションに対して、クローラーを少なくとも 1 つ作成しなければなりません。作成するクローラーのタイプは、コレクションに含めるデータのタイプによって異なります。作成するクローラーのタイプに対応するウィザードが、クローラーのオプションの指定をガイドします。ウィザードは、例えば、クローラーがシステム・リソースをどのように使用するかについてのオプションを指定する手助けをします。また、ウィザードは、コレクションに含めるデータ・ソースの選択の手助けもします。

### 手順

クローラーを作成するには、次のようにします。

1. コレクションを編集して「クロール」ページを選択し、「クローラーの作成」をクリックします。
2. クローラー・タイプと、クローラーの基本値を選択します。
  - a. Web サイト、Lotus Notes データベース、または UNIX ファイル・システムなど、クロールするデータのタイプをサポートするクローラーのタイプを選択します。

クローラー・タイプを選択したあと、その作成方法についてのオプションが表示されます。

- b. クローラーの基本値を選択します。

#### 「システム・デフォルト値を新規クローラーに使用」

クローラー初期設定値に、インストールのデフォルト値を設定します。

このオプションを選択した場合、「次へ」をクリックすると、新規クローラーの構成が開始されます。

#### 「新規クローラーに既存のクローラーの値を複製」

クローラー初期設定値に、このタイプの別のクローラーに構成されている値を設定します。

このオプションを選択すると、このクローラー・タイプに一致するクローラーのリストが表示されます。新規クローラーとして使用するクローラーを選択し、「次へ」をクリックして、新規クローラーの構成を開始します。

作成しているクローラー・タイプ用のウィザードが開きます。ウィザードのプロンプトに従って、クローラーを作成します。ウィザードのどのページでも、「ヘルプ」をクリックすると、そのタイプのクローラーに指定できるオプションの説明が表示されます。

新規クローラーは、そのコレクションに属する他のクローラーと一緒に「クローラール」ページにリストされます。クローラーを変更する必要があるときはいつでも、クローラーのプロパティおよびクローラール・スペースを編集するオプションをクリックできます。

---

## クローラーのプロパティの編集

クローラーについての情報と、クローラーがデータをクローラールする方法を変更できます。例えば、クローラーがシステム・リソースを使用する方法を変更できます。

### 始める前に


クローラー・プロパティを編集するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのクローラーを所有するコレクションのコレクション管理者でなければなりません。

### このタスクについて

クローラーのプロパティを編集するときに、「ヘルプ」をクリックすると、変更できるタイプを知ることができます。編集できるプロパティは、クローラー・タイプによって異なります。

### 手順

クローラーのプロパティを編集するには、次のようにします。

1. コレクションを編集して「クローラール」ページを選択し、編集するクローラーを指して「 クローラーのプロパティ」をクリックします。
2. クローラーのプロパティを変更して、「OK」をクリックします。
3. 変更を有効にするために、クローラーを停止し、再開します。(クローラーの説明のみを変更した場合、クローラーを再始動する必要はありません。)

---

## クローラール・スペースの編集

クローラーがクローラールするデータ・ソースについての情報を変更できます。例えば、データ・ソースの追加、データ・ソースの除去、クローラール・スケジュールの変更、および、特定のデータ・ソース内の文書のクローラール規則の変更を行うことができます。

### 始める前に


クローラール・スペースを編集するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのクローラーを所有するコレクションのコレクション管理者でなければなりません。

### このタスクについて

管理するクローラーのタイプで行える変更については、クロール・スペースを編集している間に「ヘルプ」をクリックすると説明が表示されます。

### 手順

クロール・スペースを編集するには、次のようにします。

1. コレクションを編集して「クロール」ページを選択し、編集するクローラーを指して「 クロール・スペース」をクリックします。
2. 変更するオプションを選択して、クロール・スペースを変更します。

選択できるオプションは、クローラー・タイプによって異なります。データ・ソースをコレクションに追加するといった一部のオプションについては、クローラー・タイプに応じたウィザードが開き、クロール・スペースの変更を手助けします。

3. 変更を有効にするために、クローラーを停止し、再開します。

---

## クローラーの削除

クローラーを削除すると、そのクローラーに関するすべての情報がエンタープライズ・サーチ・システムから除去されます。前にそのクローラーでクロールされた情報は、索引を再編成するまで、索引の中に残ります。

### 始める前に

クローラーを削除するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのクローラーを所有するコレクションのコレクション管理者でなければなりません。


### このタスクについて

クローラーの削除は、時間のかかるプロセスです。クローラーの削除を確認すると、システムは、システムにあるクローラーに関連するすべてのデータを削除します。

**ヒント:** このタスクは完了するのに時間がかかるため、プロセスがバックグラウンドでまだ実行中であっても、要求された操作のタイムアウトに関するメッセージが表示されます。タスクが完了したかどうかを確認するには、時々、管理コンソールで「リフレッシュ」をクリックします (Web ブラウザーの「リフレッシュ」をクリックしないでください)。クローラー名がクローラーのリストに表示されなくなったら、削除プロセスは終了です。

### 手順

クローラーを削除するには、次のようにします。

1. コレクションを編集して「クロール」ページを選択します。
2. 削除するクローラーを指して「 削除」をクリックします。

## Content Edition クローラー

IBM WebSphere Information Integrator Content Edition リポジトリをエンタープライズ・サーチ・コレクションに入れる場合は、Content Edition クローラーを構成する必要があります。

Content Edition クローラーを使用して、Documentum、FileNet Panagon Content Services、FileNet P8 Content Manager、Hummingbird Document Management (DM)、OpenText Livelink、および Portal Document Manager (PDM) のリポジトリをクロールすることができます。

クローラーを構成する場合、クローラーがクロール・スペースのすべてのリポジトリをクロールするときのオプションを指定します。また、各リポジトリでクロールする項目クラスも選択します。

Content Edition クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティで、クローラーがクロール・スペース内のすべてのリポジトリをクロールする方法を制御します。
- クローラーがリポジトリへのアクセスに直接モードまたはサーバー・モードのいずれを使用するかを指定します。サーバー・モードの場合は、クローラーが Web アプリケーション・サーバーにアクセスするための情報を指定する必要があります。
- クロールするリポジトリを選択します。
- 選択されたリポジトリのコンテンツにクローラーがアクセスできるようにするためのユーザー ID とパスワードを指定します。
- リポジトリをクロールするスケジュールを設定します。
- 各リポジトリでクロールする項目クラスを選択します。
- 項目クラスのプロパティを検索可能にするときのオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外したり、特定リポジトリの特定バージョンをクロールすると指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、アクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

Documentum、FileNet Panagon Content Services、Portal Document Manager 項目クラスの場合は、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・



ソースで維持されている現行のアクセス制御リストと比較します。(このタイプの現行信用証明情報の妥当性検査は、他のリポジトリ・タイプには使用できません。)

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローリングに必要な情報をクローラーに指定する方法が表示されます。

## WebSphere II Content Edition リポジトリへのサーバー・モード・アクセス

Content Edition クローラーを構成して、サーバー・モードでリポジトリにアクセスすることができます。

サーバー・モードでは、クローラーがデータのアクセスに使用する WebSphere Information Integrator Content Edition コネクタが WebSphere Application Server にエンタープライズ・アプリケーションとしてインストールされ、クローラーはサーバーを介してリポジトリにアクセスします。この方法により、J2EE アプリケーション・サーバー環境の利点を活用することができます。

サーバー・モードで WebSphere Information Integrator Content Edition リポジトリにアクセスできるようにクローラーを構成する前に、クローラー・サーバーでスクリプトを実行する必要があります。このスクリプトは、WebSphere Information Integrator OmniFind Edition で提供されており、これによって、Content Edition クローラーがサーバー上のリポジトリにアクセスできるようになります。

エンタープライズ・サーチ管理コンソールを使用して、サーバー・モードを使用するように Content Edition クローラーを構成する前に、ご使用の環境に該当するタスクを実行する必要があります。

- 45 ページの『UNIX での WebSphere II Content Edition 用クローラー・サーバーの構成』
- 46 ページの『Windows での WebSphere II Content Edition 用クローラー・サーバーの構成』

## WebSphere II Content Edition リポジトリへの直接モード・アクセス

Content Edition クローラーを構成して、直接モードでリポジトリにアクセスすることができます。

直接モードでは、クローラーは、WebSphere II OmniFind Edition のインストール時にクローラー・サーバーにインストールされる WebSphere Information Integrator Content Edition コネクタを使用します。

### 手順

リポジトリが直接モードでアクセスされるようにシステムを構成するには、次のようにします。



1. `iice_install_root/bin/config.sh` ファイル (UNIX) または `iice_install_root/bin/config.bat` ファイル (Microsoft Windows) の `VBR_HOME` および `JAVA_HOME` 環境変数で正しいディレクトリが指定されているか確認します。
2. WebSphere Information Integrator Content Edition 管理コンソールが直接モードで実行されるように構成するために、`vbr.as.operationMode=direct` Java システム・プロパティを `iice_install_root/bin/Admin.bat` ファイル (UNIX) または `iice_install_root/bin/Admin.bat` ファイル (Windows) に追加します。
3. WebSphere Information Integrator Content Edition 管理コンソールを直接モードで開始して、WebSphere II OmniFind Edition クローラー・サーバーのコネクタを構成します。(手順については、WebSphere Information Integrator Content Edition 文書を参照してください。)
4. WebSphere II OmniFind Edition 管理コンソールの使用時に直接モードを選択して、Content Edition クローラーを構成します。

## UNIX での WebSphere II Content Edition 用クローラー・サーバーの構成

IBM AIX<sup>®</sup>、Linux<sup>®</sup>、または Solaris オペレーティング環境を稼働するコンピュータに、WebSphere II OmniFind Edition をインストールし、リポジトリにアクセスする際にサーバー・モードを使用する Content Edition クローラーを構成する場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。このスクリプトによって、Content Edition クローラーは、WebSphere Information Integrator Content Edition リポジトリにアクセスできるようになります。

### このタスクについて

Content Edition クローラーは、Java クライアントとして WebSphere Information Integrator Content Edition Java ライブラリーを使用します。サーバー・モードで、これらの Java ライブラリーは、WebSphere Application Server の EJB 関連 Java ライブラリーを必要とします。Content Edition クローラーが Java ライブラリーと確実に連動できるようにするため、WebSphere Application Server をインストールした後、クローラー・サーバー上に WebSphere II OmniFind Edition が提供するセットアップ・スクリプトを実行する必要があります。

WebSphere Information Integrator Content Edition は、WebSphere II OmniFind Edition のインストール時に、クローラー・サーバーにインストールされます。Content Edition クローラーをサーバー・モードで使用できるようにするには、`vbr_access_services.jar` ファイルを WebSphere Information Integrator Content Edition サーバーからクローラー・サーバーにコピーする必要があります。

### 手順

クローラー・サーバーが WebSphere Information Integrator Content Edition リポジトリをクローラーできるように構成するには、次のようにします。

1. WebSphere II OmniFind Edition がマルチサーバー構成でインストールされる場合、WebSphere Application Server Java ライブラリーをインストールおよびバインドします。

2. クローラー・サーバー上で、Content Edition クローラー用のセットアップ・スクリプトを実行します。
  - a. エンタープライズ・サーチ管理者としてログインします。
  - b. 以下のスクリプト (これは、`$ES_INSTALL_ROOT/bin` ディレクトリーにインストールされています) を開始し、プロンプトに応えます。

```
escrvbr.sh
```

3. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

4. `vbr_access_services.jar` ファイルを WebSphere Information Integrator Content Edition サーバーからクローラー・サーバーにコピーします。

#### コピー元:

`vbr_access_services.jar` ファイルは、次のデフォルト場所にあります。

```
was_install_root/installedApps/server_name/application_name
```

*was\_install\_root* は WebSphere Application Server インストール・ディレクトリーであり、*server\_name* はサーバーに指定した名前であり、そして *application\_name* は WebSphere Application Server の WebSphere Information Integrator Content Edition アプリケーションに指定した名前です。

#### コピー先:

クローラー・サーバー上のターゲット・ディレクトリーは `iice_install_root/lib` です。ここで、*iice\_install\_root* は、クローラー・サーバー上の WebSphere Information Integrator Content Edition インストール・ディレクトリーです。

## Windows での WebSphere II Content Edition 用クローラー・サーバーの構成

Microsoft Windows コンピューターに、WebSphere II OmniFind Edition をインストールし、リポジトリーにアクセスする際にサーバー・モードを使用する Content Edition クローラーを構成する場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。このスクリプトによって、Content Edition クローラーは、WebSphere Information Integrator Content Edition リポジトリーにアクセスできるようになります。

#### このタスクについて

Content Edition クローラーは、Java クライアントとして WebSphere Information Integrator Content Edition Java ライブラリーを使用します。サーバー・モードで、これらの Java ライブラリーは、WebSphere Application Server の EJB 関連 Java ライブラリーを必要とします。Content Edition クローラーが Java ライブラリーと確実に連動できるようにするため、WebSphere Application Server をインストールした

後、クローラー・サーバー上に WebSphere II OmniFind Edition が提供するセットアップ・スクリプトを実行する必要があります。

WebSphere Information Integrator Content Edition は、WebSphere II OmniFind Edition のインストール時に、クローラー・サーバーにインストールされます。Content Edition クローラーをサーバー・モードで使用できるようにするには、vbr\_access\_services.jar ファイルを WebSphere Information Integrator Content Edition サーバーからクローラー・サーバーにコピーする必要があります。

## 手順

クローラー・サーバーが WebSphere Information Integrator Content Edition リポジトリをクローリングできるように構成するには、次のようにします。

1. WebSphere II OmniFind Edition がマルチサーバー構成でインストールされる場合、WebSphere Application Server Java ライブラリーをインストールおよびバインドします。
2. クローラー・サーバー上で、Content Edition クローラー用のセットアップ・スクリプトを実行します。
  - a. エンタープライズ・サーチ管理者 ID でログインします (このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです)。
  - b. 以下のスクリプト (これは、%ES\_INSTALL\_ROOT%\bin ディレクトリーにインストールされています) を開始し、プロンプトに応答します。

```
escrvbr.vbs
```

3. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。
  - a. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```

- b. 「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」を選択して、IBM WebSphere Information Integrator OmniFind Edition サービスを再始動します。
- c. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを始動します。

```
esadmin start
```

4. vbr\_access\_services.jar ファイルを WebSphere Information Integrator Content Edition サーバーからクローラー・サーバーにコピーします。

### コピー元:

vbr\_access\_services.jar ファイルは、次のデフォルト場所にあります。

```
was_install_root\installedApps\server_name\application_name
```

was\_install\_root は WebSphere Application Server インストール・ディレクトリーであり、server\_name はサーバーに指定した名前であり、そして

`application_name` は WebSphere Application Server の WebSphere Information Integrator Content Edition アプリケーションに指定した名前です。

#### コピー先:

クローラー・サーバー上のターゲット・ディレクトリーは `iice_install_root\lib` です。ここで、`iice_install_root` は、クローラー・サーバー上の WebSphere Information Integrator Content Edition インストール・ディレクトリーです。

---

## DB2 クローラー

DB2 クローラーを使用して、IBM DB2 Universal Database データベースをコレクションに入れます。また、DB2 クローラーを使用して、IBM DB2 Universal Database for z/OS、IBM Informix、Oracle、Microsoft SQL Server データベース用に作成したニックネーム表を入れることもできます。

クローラーは、クロール対象の各データベース・サーバーごとに独立したクローラーを構成する必要があります。クローラーを構成する場合、クローラーが同じサーバー上のすべてのデータベースをクロールするときのオプションを指定します。また、各データベースでクロールする特定の表も選択します。

クロール用に選択する表は、データベース表、ニックネーム表、ビューのいずれかでなければなりません。DB2 クローラーは結合表をサポートしません。

### イベント・パブリッシング

WebSphere Information Integrator Event Publisher Edition を使用している場合に、クロールするデータベースをパブリッシング・キュー・マップに関連付けると、DB2 クローラーは、マップを使用してデータベース表への更新をクロールできます。

パブリッシング・キュー・マップは、データベース表に対する更新が行われると、XML メッセージを受け取る WebSphere MQ キューを識別します。クローラーは、表が更新されると、発行されたイベントに関する情報のキューを `listen` し、クロール・スペースを更新します (クローラーは、最初に表をクロールする場合、すべての文書をクロールします)。

イベント・パブリッシングにより、新規文書および変更された文書は、クローラーがクローラー・スケジュールに従ってクロールする文書よりも速く検索されるようになります。

一部またはすべての表がイベント・パブリッシングを使用するように構成されている場合は、クローラーの構成時に、クローラーが WebSphere MQ およびパブリッシング・キュー・マップにアクセスできるように指定することができます。

また、クロールされるサーバーで WebSphere MQ および WebSphere Information Integrator Event Publisher Edition が構成されており、クローラー・サーバーで WebSphere MQ クライアント・モジュールが構成されていることを確認する必要があります。イベント・パブリッシングを DB2 クローラーで使用するには、以下のタスクを実行します。

- 52 ページの『DB2 クローラーのための WebSphere MQ の構成』

- 50 ページの『DB2 クローラーのための WebSphere Information Integrator Event Publisher Edition の構成』
- 54 ページの『イベント・パブリッシングを使用するための UNIX でのクローラー・サーバーの構成』
- 56 ページの『イベント・パブリッシングを使用するための Windows でのクローラー・サーバーの構成』

## 構成の概要

DB2 クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによる特定のデータベース・サーバーのすべてのデータベースに対するクロールを制御します。
- クロールするデータベースのタイプに関する情報を指定します。

ローカル・データベース・サーバーでカタログされていないリモート・データベースをクロールする予定がある場合、まず、リモート・サーバーで DB2 Administration Server を始動してから、DB2 クローラーを使用してそれらのデータベースをクロールします。また、クローラーの構成時にそのリモート・データベースのホスト名とポートを指定する必要があります。

- クロールするデータベースを選択します。
- アクセス制御をしているデータベースにクローラーがアクセスできるようにするためのユーザー ID とパスワードを指定します。
- データベースをクロールするスケジュールを設定します。
- 各データベースでクロールする表を選択します。

**重要:** ディスカバリー・プロセスのパフォーマンスを最適化するため (そしてクローラー構成プロセスがタイムアウトになるのを防止するため)、すべての表をクロールする選択は、データベース中の表の数が少ない場合か、各表に含まれる列の数が少ない場合のみ行うようにしてください。一部の表についてすぐにクロールする選択をした場合、後からクロール・スペースを編集して、さらに多くの表をコレクションに追加できます。

- イベント・パブリッシング・キューで表に対する更新が行われた際にクロールする表を選択し、クローラーがイベント・パブリッシング・キューにアクセスするための情報を指定します。
- 特定の表の列を検索可能にするときのオプションを指定します。例えば、特定の列をパラメトリック照会で使用できるようにしたり、検索結果に戻すことのできる列を指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローリングに必要な情報をクローラーに指定する方法が表示されます。

## DB2 クローラーのための WebSphere Information Integrator Event Publisher Edition の構成

イベント・パブリッシングを使用するように DB2 クローラーを構成する前に、クローリング対象のサーバー上に IBM WebSphere Information Integrator Event Publisher Edition が構成されていることを確認してください。

### このタスクについて

DB2 クローラーと共に使用するために WebSphere Information Integrator Event Publisher Edition を構成する際には、以下のガイドラインを使用してください。

- ソース表内の変更された列と未変更の列の両方がパブリッシングに選択されなければなりません。
- ソース表内の削除された行がパブリッシングに選択されなければなりません。
- 1 つのイベント・パブリッシング・キューを複数のデータベースで共用することはできません。
- 単一のデータベースが複数のキュー・マップおよびキューを持つことはできません。
- 1 つの表は、1 つのパブリッシング・キュー・マップに関連付けられた 1 つの XML パブリケーションを持っているべきです。(1 つの表が、1 つのパブリッシング・キュー・マップに関連付けられた複数の XML パブリケーションを持つべきではありません。各 XML パブリケーションがそれぞれ異なる 1 つのパブリッシング・キューに関連付けられている場合は、1 つの表が複数の XML パブリケーションを持つことができます。)

### 手順

DB2 クローラーがイベント・パブリッシング・キューにパブリッシュされる表更新にアクセスできるようにデータベース・サーバーを構成するには、次のステップを実行します (これらのステップについての不明点は、WebSphere Information Integrator Publisher Edition 資料を参照してください)。

1. クローリングされるデータベース・サーバー上に WebSphere Information Integrator Event Publisher Edition をインストールします。
2. レプリケーション・センターのランチパッドを開始します。

オペレーティング・システム	コマンド
UNIX	db2rc
Windows	「スタート」 → 「IBM DB2 Replication Center」を選択

3. Q キャプチャー制御テーブルを作成します。
  - a. ランチパッドのビューとして「イベント・パブリッシング」を選択し、「Q キャプチャー制御テーブルの作成」を選択して「次へ」をクリックします。



- b. 「**Q キャプチャー・サーバー**」フィールドで、選択可能なデータベース・サーバーのリストから、**Q キャプチャー・サーバー**として使用するサーバーを選択し、「**OK**」をクリックします。
  - c. 選択した **Q キャプチャー・サーバー**へのアクセスを認可されているユーザーID とパスワードを指定します。**Q キャプチャー**のスキーマを変更するか、デフォルトのスキーマ名を受け入れ、「**次へ**」をクリックします。
  - d. このデータベース・サーバー上に **WebSphere MQ** を構成したときに指定した、キュー・マネージャー、管理キュー、および再始動キューの名前を指定し、「**次へ**」をクリックします。
  - e. 「**完了**」をクリックします。メッセージと **SQL** スクリプトを示すページが表示された後、「**クローズ**」をクリックします。
  - f. 処理オプションに「**すぐに実行**」を選択し、「**OK**」をクリックします。**SQL** スクリプトが完了したことを示すメッセージが表示された後、「**クローズ**」をクリックします。
4. **XML** パブリケーションを作成します。
- a. レプリケーション・センターのランチパッドで、ランチパッドのビューとして「**イベント・パブリッシング**」を選択し、「**XML** パブリケーションの作成」を選択して「**次へ**」をクリックします。
  - b. 「**開始**」ページで、「**次へ**」をクリックします。
  - c. 「**サーバーおよびキュー・マップ**」ページで、**Q キャプチャー・サーバー**と **Q キャプチャー・スキーマ**が正しいことを確認し、「**パブリッシング・キュー・マップ**」フィールドの横のボタンをクリックし、**パブリッシング・キュー・マップ**を作成するため「**新規**」をクリックします。
  - d. 「**一般**」ページで、キュー・マップの名前を入力します。
  - e. 「**プロパティ**」ページで、送信キューの名前 (例えば、このサーバー上に **WebSphere MQ** を構成したときに指定したデータ・キューの名前) を指定し、メッセージ内容のタイプに「**行操作**」または「**トランザクション**」を選択し、ハートビート・メッセージ送信のチェック・ボックスと **JMS** メッセージ・ヘッダー追加のチェック・ボックスをクリアし、「**OK**」をクリックします。
  - f. メッセージと **SQL** スクリプトを示すページが表示された後、「**クローズ**」をクリックします。
  - g. 処理オプションに「**すぐに実行**」を選択し、「**OK**」をクリックします。**SQL** スクリプトが完了したことを示すメッセージが表示された後、「**クローズ**」をクリックします。
  - h. 「**パブリッシング・キュー・マップの選択**」ページで、作成したキューを選択して「**OK**」をクリックします。
  - i. 「**サーバーおよびキュー・マップ**」ページで、キュー・マップ名が正しいことを確認して「**次へ**」をクリックします。
  - j. 「**ソース表**」ページで、「**追加**」をクリックし、「**すべて取得**」をクリックし、**イベント・パブリッシング**に使用可能化したい表を選択して「**OK**」をクリックし、「**次へ**」をクリックします。
  - k. 「**列と行**」ページで、**DB2** クローラーがクロールする対象にする列 (またはすべての列) を選択し、キー列を選択します。クロールする行 (またはすべて



の行) を選択したページで、ソース表の削除をパブリッシュするオプションを選択します。これらのオプションの構成を完了した後、「次へ」をクリックします。

- l. 「メッセージ内容」 ページで、列データには変更された列と未変更の列の両方を組み込むオプションを選択し、新規データ値のみのオプションを選択します。XML パブリケーションを自動的に開始するよう指定するチェック・ボックスが選択されていることを確認し、「次へ」をクリックします。
  - m. 「XML パブリケーションの検討」 ページで、「次へ」をクリックします。
  - n. 「サマリー」 ページで、「完了」 ページをクリックします。メッセージと SQL スクリプトを示すページが表示された後、「クローズ」をクリックします。
  - o. 処理オプションに「**すぐに実行**」を選択し、「**OK**」をクリックします。SQL スクリプトが完了したことを示すメッセージが表示された後、「クローズ」をクリックします。
5. Q キャプチャー・サーバーを開始します。
- a. レプリケーション・センターのランチパッドをクローズし、レプリケーション・センターを開始します。
  - b. オブジェクト・ツリーで、「**Q レプリケーション**」 → 「**定義**」 → 「**Q キャプチャー・サーバー**」を順にクリックします。
  - c. 構成した Q キャプチャー・サーバーのアイコンを右クリックし、「**データベースを Q レプリケーションに使用可能にする**」を選択します。
  - d. 警告メッセージが表示された後、「**OK**」をクリックします。
  - e. DB2 メッセージを示すページが表示された後、「クローズ」をクリックします。
  - f. オブジェクト・ツリーで、Q キャプチャー・サーバーのアイコンを右クリックし、「**Q キャプチャー・プログラムの開始**」を選択します。
  - g. 処理オプションに、「**すぐに実行**」を選択し、システム名、DB2 ユーザーのユーザー ID とパスワード、ログを保管するディレクトリーのパス、および DB2 インスタンス名を指定し、「**OK**」をクリックします。
  - h. 要求がサブミットされたことを示すメッセージが表示された後、「クローズ」をクリックします。
  - i. オブジェクト・ツリーで、Q キャプチャー・サーバーのアイコンを右クリックし、「**状況のチェック**」を選択します。

Q キャプチャー・サーバーの状況が表示されます。エラーが発生した場合、サーバーがダウンしたと推定されることを示す状況メッセージが表示されません。ログを検討し、エラーの原因を判別するには、コマンド行で以下のコマンドを入力します。

```
asnlcap Capture_Server=capture server name LOGSTDOUT=y
```

## DB2 クローラーのための WebSphere MQ の構成

イベント・パブリッシングを使用するように DB2 クローラーを構成する前に、クローラーが listen するサーバー上に IBM WebSphere MQ が構成されていることを確認してください。

## 始める前に

DB2 UDB、WebSphere Information Integrator Event Publisher Edition、および WebSphere MQ がターゲット・データベース・サーバーにインストール済みであることを確認してください。

## 制約事項

ターゲット・データベース・サーバーが Linux コンピューターにインストールされている場合、すべての DB2 Universal Database (DB2 UDB) ユーザー、WebSphere MQ ユーザー、および WebSphere II OmniFind Edition ユーザーは、以下の環境変数を設定する必要があります。

```
export LD_ASSUME_KERNEL=2.4.19
```

この環境変数は、LinuxThread スレッド化インプリメンテーションのエクスポートが、インストールが実行されるシェル、WebSphere MQ 制御コマンドが発行されるシェル、または WebSphere MQ アプリケーションが実行される任意のシェルから行われることを可能にします。WebSphere MQ は、この環境変数がエクスポートされることを必要とします。

## このタスクについて

DB2 クローラーは、WebSphere MQ サーバーへのクライアント接続モードをサポートします。このクローラーは、イベント・パブリッシング・キューにパブリッシュされる XML メッセージを listen します。クローラーは、複数のキューを経由して移送される XML メッセージを listen できません。

WebSphere MQ を構成すると、それ以降、DB2 クローラーはキュー・マネージャー名、キュー名、サーバー・ホスト名、サーバー・ポート番号、およびサーバー・チャンネル名を使用して、パブリッシング・キューから XML メッセージを取得します。クローラーは、メッセージを構文解析し、更新された表についての情報でクローラー・スペースを更新します。

## 手順

DB2 クローラーがイベント・パブリッシング・キューを listen できるようにデータベース・サーバーを構成するには、次のステップを実行します (これらのステップについての不明点は、WebSphere MQ 資料を参照してください)。

1. WebSphere MQ 管理者役割としてログインし、キュー・マネージャーおよびキューを作成するため以下のコマンドを入力します。
  - a. コマンド行で、以下のコマンドを入力します。

```
crtmqm QM1
```

- b. Setup completed メッセージが表示されたら、以下のコマンドを入力します。

```
strmqm QM1
```

- c. 'QM1' started メッセージが表示されたら、以下のコマンドを入力します。

```
runmqsc QM1
```

d. Starting MQSC for queue manager QM1 メッセージが表示されたら、以下のコマンドを入力して管理キューを作成します。

```
DEFINE QLOCAL('ASN.QM1.ADMINQ')
```

e. WebSphere MQ queue created メッセージが表示されたら、以下のコマンドを入力して再始動キューを作成します。

```
DEFINE QLOCAL(' ASN.QM1.RESTARTQ')
```

f. WebSphere MQ queue created メッセージが表示されたら、以下のコマンドを入力してデータ・キューを作成します。

```
DEFINE QLOCAL(' ASN.QM1.DATAQ')
```

g. WebSphere MQ queue created メッセージが表示されたら、以下のコマンドを入力して終了します。

```
end
```

2. 以下のコマンドを入力してデータベース・サーバー上で MQ Listener を開始します (MQ Listener は、イベント・パブリッシングを使用する DB2 クローラー作成時に実行中でなければなりません)。この例では、サーバーのポート番号は 1414 であり、デフォルトのチャンネル SYSTEM.DEF.SVRCONN が使用されています。

```
runmqtsr -m QM1 -t TCP -p 1414 &
```

3. 以下のコマンドを入力して、イベント・パブリッシングのために Message Queuing Interface (MQI) を経由してキュー・マネージャーおよびキューにアクセスすることを DB2 UDB ユーザーに認可します (この例では、ユーザー ID は db2inst1 です)。

```
setmqaut -m QM1 -t qmgr -p db2inst1 +allmqi
```

```
setmqaut -m QM1 -t queue -n ASN.QM1.DATAQ -p db2inst1 +allmqi
```

```
setmqaut -m QM1 -t queue -n ASN.QM1.ADMINQ -p db2inst1 +allmqi
```

```
setmqaut -m QM1 -t queue -n ASN.QM1.RESTARTQ -p db2inst1 +allmqi
```

4. DB2 クローラーを作成し、イベント・パブリッシングと共にこのクローラーを実行するのに使用されるユーザー ID に対して、以下のコマンドを入力します。これらのコマンドは、イベント・パブリッシングのために Message Queuing Interface (MQI) を経由してそのユーザー ID がキュー・マネージャーおよびキューにアクセスすることを認可します。この例では、ユーザー ID は esuser です。

```
setmqaut -m ASN.QM1.QM2 -t qmgr -p esuser +allmqi
```

```
setmqaut -m ASN.QM1.QM2 -t queue -n ASN.QM1.DATAQ -p esuser +allmqi
```

## イベント・パブリッシングを使用するための UNIX でのクローラー・サーバーの構成

IBM AIX、Linux、または Solaris オペレーティング環境が稼働するコンピューターに WebSphere II OmniFind Edition をインストールし、イベント・パブリッシングを使用するよう DB2 クローラーを構成する場合、クローラー・サーバーを構成す

るスクリプトを実行する必要があります。そのスクリプトは、クローラーが WebSphere MQ キュー・マネージャーおよびキューにアクセスできるようにします。

### このタスクについて

DB2 クローラーは、Java Messaging 用の WebSphere MQ 5.3 モジュールを使用して WebSphere MQ キュー・マネージャーおよびキューにアクセスします。これらのモジュールをクローラー・サーバーにインストールする必要があります。

WebSphere MQ モジュールをインストールした後、DB2 クローラーがイベント・パブリッシングを確実に使用できるようにするため、WebSphere II OmniFind Edition によって提供されているセットアップ・スクリプトをクローラー・サーバーで実行する必要もあります。

### 手順

イベント・パブリッシングを使用するようクローラーを構成するには、次のようにします。

1. Java Messaging 用の WebSphere MQ 5.3 モジュールをクローラー・サーバーにインストールします。

- a. root ユーザーとしてログインし、以下のコマンドを入力します。

```
export LD_ASSUME_KERNEL=2.4.19
```

- b. WebSphere MQ CD を挿入します。

- c. Java Messaging 用の MQ モジュールが入っているディレクトリーに移動します。

- d. これらのモジュールをインストールするため、以下のコマンドを入力します。

```
rpm -i MQSeriesJava-5.3.0-1.i386.rpm
```

2. クローラー・サーバー上で DB2 クローラー用のセットアップ・スクリプトを実行します。

- a. エンタープライズ・サーチ管理者としてログインします (このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです)。

- b. 以下のスクリプト (これは、\$ES\_INSTALL\_ROOT/bin ディレクトリーにインストールされています) を開始し、プロンプトに応答します。

```
escrdb2.sh
```

3. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

```
esadmin stop
```

```
stopccl.sh
```

```
startccl.sh -bg
```

```
esadmin start
```

## イベント・パブリッシングを使用するための Windows でのクローラー・サーバーの構成

Microsoft Windows が稼働するコンピュータに WebSphere II OmniFind Edition をインストールし、イベント・パブリッシングを使用するよう DB2 クローラーを構成する場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。そのスクリプトは、クローラーが WebSphere MQ キュー・マネージャーおよびキューにアクセスできるようにします。

### このタスクについて

DB2 クローラーは、Java Messaging 用の WebSphere MQ 5.3 モジュールを使用して WebSphere MQ キュー・マネージャーおよびキューにアクセスします。これらのモジュールをクローラー・サーバーにインストールする必要があります。

WebSphere MQ モジュールをインストールした後、DB2 クローラーがイベント・パブリッシングを確実に使用できるようにするため、WebSphere II OmniFind Edition によって提供されているセットアップ・スクリプトをクローラー・サーバーで実行する必要があります。

### 手順

イベント・パブリッシングを使用するようクローラーを構成するには、次のようにします。

1. Java Messaging 用の WebSphere MQ 5.3 モジュールをクローラー・サーバーにインストールします。
  - a. WebSphere MQ CD を挿入します。
  - b. WebSphere MQ インストーラーを開始します。
  - c. 「製品フィーチャーの選択」ウィンドウで、インストール・オプションに「**Java Messaging**」を選択します。
2. クローラー・サーバー上で DB2 クローラー用のセットアップ・スクリプトを実行します。
  - a. エンタープライズ・サーチ管理者 ID でログインします (このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです)。
  - b. 以下のスクリプト (これは、%ES\_INSTALL\_ROOT%\bin ディレクトリーにインストールされています) を開始し、プロンプトに応答します。

```
escrdb2.vbs
```
3. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。
  - a. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```
  - b. 「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」を選択して、IBM WebSphere Information Integrator OmniFind Edition サービスを再始動します。

- c. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを始動します。

```
esadmin start
```

---

## DB2 Content Manager クローラー

IBM DB2 Content Manager 項目タイプをエンタープライズ・サーチ・コレクションに入れる場合は、DB2 Content Manager クローラーを構成する必要があります。

### クローラー・サーバーの構成

DB2 Content Manager サーバーをクロールするには、あらかじめクローラー・サーバーで特定のスクリプトを実行する必要があります。このスクリプトは、WebSphere Information Integrator OmniFind Edition で提供されており、これによって DB2 Content Manager クローラーが DB2 Content Manager サーバーと通信できるようになります。

エンタープライズ・サーチ管理コンソールを使用して、DB2 Content Manager クローラーを構成する前に、ご使用の環境に該当するタスクを実行する必要があります。

- 58 ページの『UNIX での DB2 Content Manager 用クローラー・サーバーの構成』
- 60 ページの『Windows での DB2 Content Manager 用クローラー・サーバーの構成』

### 構成の概要

DB2 Content Manager クローラーを使用して、任意の数の DB2 Content Manager サーバーをクロールすることができます。クローラーを構成する際に、クローラーがクロール・スペースのすべての DB2 Content Manager サーバーをクロールするためのオプションを指定します。また、各サーバーでクロールする特定の項目タイプも選択します。

DB2 Content Manager クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクロール・スペース内のすべての DB2 Content Manager サーバーのあらゆる項目タイプのクロールを制御します。
- クロールする DB2 Content Manager サーバーを選択します。
- クローラーが DB2 Content Manager サーバーのコンテンツにアクセスできるようにするためのユーザー ID とパスワードを指定します。
- サーバーをクロールするスケジュールを設定します。
- 各 DB2 Content Manager サーバーでクロールする項目タイプを選択します。



- 一部の項目タイプの属性を検索可能にするときのオプションを指定します。例えば、特定タイプの文書をクローラー・スペースから除外したり、検索結果に戻すことのできる属性を指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローリングに必要な情報をクローラーに指定する方法が表示されます。

## UNIX での DB2 Content Manager 用クローラー・サーバーの構成

IBM AIX、Linux、または Solaris オペレーティング環境が稼働するコンピューターに WebSphere II OmniFind Edition をインストールする場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。そのスクリプトは、DB2 Content Manager クローラーが IBM DB2 Content Manager サーバーと通信できるようにします。

### このタスクについて

DB2 Content Manager クローラーは、DB2 Content Manager バージョン 8 用の Java コネクタを使用して、DB2 Content Manager サーバーにアクセスします。このコネクタは、クローラー・サーバーに IBM DB2 Information Integrator for Content バージョン 8.2 以降をインストールすると、インストールされます。DB2 Content Manager クローラーが DB2 Content Manager と連動して機能できるようにするには、コネクタをインストールした後で、WebSphere II OmniFind Edition が提供するセットアップ・スクリプトをクローラー・サーバー上で実行します。

### 手順

クローラー・サーバーが DB2 Content Manager サーバーをクローリングできるように構成するには、次のようにします。

1. クローラー・サーバーに DB2 Content Manager バージョン 8 の Java コネクタをインストールします。
  - a. クローラー・サーバーに、root ユーザーとしてログインします。

```
su - root
```

- b. db2profile ファイルを実行します。以下に例を示します。

```
./home/db2inst/sql1lib/db2profile
```

- c. JAVAHOME 環境変数をエクスポートします。以下に例を示します。

```
export JAVAHOME=/usr/IBMJava2-141
```

- d. Java ディレクトリを PATH 環境変数に追加します。

```
export PATH=$PATH:$JAVAHOME/bin
```



- e. DB2 Information Integrator for Content インストール CD を挿入し、インストール・ウィザードを実行します。
  - f. 「コンポーネントの選択」ウィンドウで、次の操作を行います (Information Integrator for Content バージョン 8.3 を使用して作業している場合、「コンポーネントの選択」ウィンドウには、カスタム・インストールのオプションが表示されます。)
    - 1) 「コンポーネント」リストから「ローカル・コネクタ」を選択し、「サブコンポーネント」リストから「Content Manager V8 コネクタ」を選択します。
    - 2) 「コンポーネント」リストから「コネクタ・ツールキットとサンプル」を選択し、「サブコンポーネント」リストから「Content Manager V8 コネクタ」を選択します。
  - g. DB2 Content Manager ライブラリーのデータベース名、ユーザー名、およびパスワードを指定し、残りのウィンドウでは、デフォルトの設定値を受け入れます。
2. クローラー・サーバーに、DB2 管理グループの中のユーザー ID でログインします。
  3. リモート DB2 Content Manager ライブラリー・サーバー・データベースをカタログし、クローラー・サーバーが DB2 Content Manager サーバーに接続できることを確認します。

```
db2 catalog tcpip node node_name remote hostname server port
db2 catalog database database_name as alias at node node_name
```

4. オプション: root ユーザーとしてログインし、データベース接続をテストします。

```
. Information_Integrator_for_Content_install_directory/bin/cmbenv81.sh
cd Information_Integrator_for_Content_install_directory/samples/java/icm
javac *.java
java SConnectDisconnect ICMdatabase_name CMadmin_ID CMadmin_password
```

5. クローラー・サーバー上で DB2 Content Manager クローラー用のセットアップ・スクリプトを実行します。
  - a. 次のようにして、ES\_INSTALL\_ROOT/bin ディレクトリーに移動します。

```
cd $ES_INSTALL_ROOT/bin
```

- b. 以下のスクリプトを開始し、プロンプトに応答します。

```
escrcm.sh
```

6. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

## Windows での DB2 Content Manager 用クローラー・サーバーの構成

Microsoft Windows コンピューターに WebSphere II OmniFind Edition をインストールする場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。そのスクリプトは、DB2 Content Manager クローラーが IBM DB2 Content Manager サーバーと通信できるようにします。

### このタスクについて

DB2 Content Manager クローラーは、DB2 Content Manager バージョン 8 用の Java コネクターを使用して、DB2 Content Manager サーバーにアクセスします。このコネクターは、クローラー・サーバーに IBM DB2 Information Integrator for Content バージョン 8.2 以降をインストールすると、インストールされます。DB2 Content Manager クローラーが DB2 Content Manager と連動して機能できるようにするには、コネクターをインストールした後で、WebSphere II OmniFind Edition が提供するセットアップ・スクリプトをクローラー・サーバー上で実行します。

### 手順

クローラー・サーバーが DB2 Content Manager サーバーをクロールできるように構成するには、次のようにします。

1. クローラー・サーバーに DB2 Content Manager バージョン 8 の Java コネクターをインストールします。
  - a. DB2 Information Integrator for Content インストール CD を挿入します。インストール・プログラムが自動的に始動します。

DB2 Content Manager Enterprise Information Portal インストール・ウィザードが開きます。

- b. 「コンポーネントの選択」ウィンドウで、次の操作を行います (Information Integrator for Content バージョン 8.3 を使用して作業している場合、「コンポーネントの選択」ウィンドウには、カスタム・インストールのオプションが表示されます。)
    - 1) 「コンポーネント」リストから「ローカル・コネクター」を選択し、「サブコンポーネント」リストから「Content Manager V8 コネクター」を選択します。
    - 2) 「コンポーネント」リストから「コネクター・ツールキットとサンプル」を選択し、「サブコンポーネント」リストから「Content Manager V8 コネクター」を選択します。
  - c. DB2 Content Manager ライブラリーのデータベース名、ユーザー名、およびパスワードを指定し、残りのウィンドウでは、デフォルトの設定値を受け入れます。
2. リモート DB2 Content Manager ライブラリー・サーバー・データベースをカタログし、クローラー・サーバーが DB2 Content Manager サーバーに接続できることを確認します。クローラー・サーバー上のコマンド・プロンプトで、次のコマンドを入力します。

```
db2 catalog tcpip node node_name remote hostname server port  
db2 catalog database database_name as alias at node node_name
```

3. オプション: データベース接続をテストするため、コマンド・プロンプトをオープンし、以下のコマンドを入力します。

```
cmbenv81.bat
cd Information_Integrator_for_Content_install_directory\samples\java\icm
javac *.java
java SConnectDisconnect ICMdatabase_name CAdmin_ID CAdmin_password
```

4. クローラー・サーバー上で DB2 Content Manager クローラー用のセットアップ・スクリプトを実行します。
  - a. 次のようにして、ES\_INSTALL\_ROOT¥bin ディレクトリーに移動します。

```
cd %ES_INSTALL_ROOT%bin
```

- b. 以下のスクリプトを開始し、プロンプトに応答します。

```
escrcm.vbs
```

5. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

- a. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```

- b. 「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」を選択して、IBM WebSphere Information Integrator OmniFind Edition サービスを再始動します。

- c. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを始動します。

```
esadmin start
```

---

## Domino Document Manager クローラー

Domino Document Manager ライブラリーおよびキャビネットをエンタープライズ・サーチ・コレクションに入れる場合は、Domino Document Manager クローラーを構成する必要があります。

### クローラー・サーバーの構成

クロールする Domino Document Manager サーバーが Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用している場合は、クローラー・サーバーで特定のスクリプトを実行する必要があります。このスクリプトは、WebSphere Information Integrator OmniFind Edition で提供されており、これによって、Domino Document Manager クローラーが NRPC を使用するサーバーと通信できるようになります。

クロールする Domino Document Manager サーバーが Domino Internet Inter-ORB Protocol (DIIOP) を使用している場合、クローラー・サーバーでセットアップ・スクリプトを実行する必要はありません。ただし、Domino Document Manager クローラーが Domino Document Manager サーバーにアクセスできるようにサーバーを構成する必要があります。

WebSphere II OmniFind Edition が IBM AIX システムにインストールされている場合は、I/O 完了ポート・モジュールがクローラー・サーバーにインストールされており、使用可能であることを確認してください。

エンタープライズ・サーチ管理コンソールを使用して、Domino Document Manager クローラーを構成する前に、ご使用の環境に該当するタスクを実行する必要があります。

- 68 ページの『Lotus Domino ソースをクロールするための UNIX でのクローラー・サーバーの構成』
- 70 ページの『Lotus Domino ソースをクロールするための Windows でのクローラー・サーバーの構成』
- 72 ページの『DIOP プロトコルを使用するサーバーの構成』
- 73 ページの『Lotus Domino ソースをクロールするための AIX での I/O 完了ポートの構成』。

## 構成の概要

Domino Document Manager クローラーを使用して、任意の数の Domino Document Manager ライブラリーをクロールすることができます。クローラーを作成する際に、単一の Domino Document Manager サーバーからクロールするライブラリーを選択します。後でクロール・スペースを編集する際に、同じクロール・スペースに別の Domino Document Manager サーバーの文書を追加できます。クローラーを作成または編集する際に、クロール対象として選択したライブラリー内のすべてのキャビネットをクロールするか、または、特定のキャビネットをクロールするかを指定できます。

Domino Document Manager クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティで、クローラーがクロール・スペース内のすべての文書をクロールする方法を制御します。
- Domino Document Manager サーバーおよび通信プロトコルを識別します。
- クロールするライブラリーを選択します。
- ライブラリーをクロールするスケジュールを設定します。
- クロールする文書を選択します。クローラーは、ライブラリー内のすべてのキャビネットをクロールするか、またはユーザーが選択したキャビネット内のある特定の文書のみをクロールすることができます。
- さまざまなライブラリーおよびキャビネットのフィールドを検索可能にするオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外したり、添付ファイル検索のオプションを指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケ

ーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

また、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・ソースで維持されている現行のアクセス制御リストと比較します。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

---

## Exchange Server クローラー

Microsoft Exchange Server のパブリック・フォルダーをエンタープライズ・サーチ・コレクションに入れる場合は、Exchange Server クローラーを構成する必要があります。

Exchange Server クローラーを使用して、Exchange Server のパブリック・フォルダー・サーバー上の任意の数のフォルダーおよびサブフォルダーをクロールすることができます。クローラーの作成時に、パブリック・フォルダー・サーバーでクロールする対象コンテンツを選択します。クロール対象については、後からクロール・スペースを編集して、別のパブリック・フォルダー・サーバーのコンテンツを追加することができます。

Exchange Server クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクロール・スペース内の全サーバーの全サブフォルダーに対するクロールを制御します。
- クロールする Exchange Server パブリック・フォルダー・サーバーに関する情報を指定します。

クローラーがサーバーのコンテンツにアクセスできるようにするためのユーザー ID とパスワードを指定する必要があります。サーバーが Secure Sockets Layer (SSL) プロトコルを使用している場合、クローラーがクローラー・サーバーの鍵ストア・ファイルにアクセスできるようにするためのオプションを指定することができます。

- パブリック・フォルダー・サーバーをクロールするスケジュールを設定します。
- クロールするフォルダーおよびサブフォルダーを選択します。
- サブフォルダー内の文書を検索可能にするときのオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外することができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケ

ーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローリングに必要な情報をクローラーに指定する方法が表示されます。

## セキュア Exchange Server 文書へのアクセスの検証

ファイアウォールによって保護された文書を Exchange Server クローラーを使用してクローリングするには、クローラー・サーバーが Microsoft Exchange Server パブリック・フォルダー・サーバーにアクセスできることを検証する必要があります。

### このタスクについて

クローラー・サーバーがセキュア Exchange Server サーバーにアクセスできない場合、サーバーから HTTP コード 501 (インプリメントされていない) を受け取ります。予期しない HTTP 応答を受け取ったことを示すメッセージも表示されます。

### 手順

ファイアウォールで保護された文書にクローラー・サーバーがアクセスできることを検証するには、次のようにします。

1. クローラー・サーバー上で Web ブラウザーを起動します。
2. クローリングしたい Exchange Server パブリック・フォルダー・サーバーの URL に移動します。例: `http://exchange.yourCompany.com/public/`
3. Exchange Server ページをオープンできることを検証します。

Exchange Server サーバーにアクセスできない場合、社内のサーバー管理者に連絡してください。

---

## NNTP クローラー

NNTP ニュース・グループの記事をエンタープライズ・サーチ・コレクションに入れる場合は、NNTP クローラーを構成する必要があります。

NNTP クローラーを使用して、任意の数の NNTP サーバーをクローリングすることができます。クローラーの構成時に、各サーバーごとにクローリングするニュースグループを選択します。また、対象から除外するニュースグループのパターンを指定することもできます。この設計により、特定サーバーの大多数のニュースグループをクローラーのクローリング対象として容易に許可することができ、また、ユーザーに検索してほしくない少数のニュースグループをクローリング禁止にすることができます。

例えば、特定の NNTP サーバーのすべてのニュース・グループを含める規則を指定して、そのサーバーのニュース・グループのうち、`private` というストリングが名前に含まれているニュース・グループは除外する規則を指定できます。

NNTP クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。



クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクロール・スペース内の全ニュースグループに対するクロールを制御します。
- クロール対象のニュースグループを指定し、さらにクロール・スペースから除外する一定のニュースグループのパターンを指定します。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

---

## Notes クローラー

IBM Lotus Notes データベースをエンタープライズ・サーチ・コレクションに入れる場合は、Lotus Notes クローラーを構成する必要があります。

### クローラー・サーバーの構成

クロールする Lotus Notes サーバーが Notes リモート・プロシージャ・コール (NRPC) プロトコルを使用している場合、クローラー・サーバーで特定のスクリプトを実行する必要があります。このスクリプトは、WebSphere Information Integrator OmniFind Edition で提供されており、これによって、Lotus Notes クローラーが NRPC を使用するサーバーと通信できるようになります。

クロールする予定の Lotus Notes サーバーが Domino Internet Inter-ORB Protocol (DIIOP) を使用している場合、クローラー・サーバーに対してセットアップ・スクリプトを実行する必要はありません。ただし、Lotus Notes クローラーが Lotus Notes サーバーにアクセスできるようにサーバーを構成する必要があります。

WebSphere II OmniFind Edition が IBM AIX システムにインストールされている場合は、I/O 完了ポート・モジュールがクローラー・サーバーにインストールされており、使用可能であることを確認してください。

エンタープライズ・サーチ管理コンソールを使用して、Lotus Notes クローラーを構成する前に、ご使用の環境に該当するタスクを実行する必要があります。

- 68 ページの『Lotus Domino ソースをクロールするための UNIX でのクローラー・サーバーの構成』
- 70 ページの『Lotus Domino ソースをクロールするための Windows でのクローラー・サーバーの構成』
- 72 ページの『DIIOP プロトコルを使用するサーバーの構成』
- 73 ページの『Lotus Domino ソースをクロールするための AIX での I/O 完了ポートの構成』



## 文書レベル・セキュリティー

コレクションのセキュリティーが使用可能で、クロールする Lotus Notes サーバーが NRPC プロトコルを使用している場合は、クローラー・サーバーで Lotus Domino トラステッド・サーバーを構成する必要があります。トラステッド・サーバーは、文書レベルのアクセス制御を行うために使用されます。ユーザーに対してコレクションを検索可能にする前に、以下のタスクを行います。

- 228 ページの『ユーザーの信用証明情報を検証するための Lotus Domino トラステッド・サーバーの構成』
- セキュリティーを使用するために、WebSphere Application Server でグローバル・セキュリティーを使用可能にし、検索アプリケーションを構成します。このステップにより、ユーザーが検索アプリケーションを使用する際に、信用証明情報を指定するように指示するプロンプトが出されます。検索サーバーは、指定された信用証明情報を使用して、Lotus Notes 文書に対するユーザーのアクセス権限を checks します。

## 構成の概要

Lotus Notes クローラーを使用して、任意の数の標準 Lotus Notes データベース (.nsf ファイル) をクロールすることができます。クローラーを作成するときに、単一の Lotus Notes サーバーからクロールするデータベースまたはディレクトリーを選択します。後でクロール・スペースを編集して、同じクロール・スペースに別の Lotus Notes サーバーから文書を追加できます。クローラーを作成または編集するときに、サーバー上のすべてのデータベースまたはディレクトリーをクロールするか、あるいは、特定のデータベース、ビューまたはフォルダーをクロールするかを指定できます。

Lotus Notes クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティーを指定します。クローラーのプロパティーで、クローラーがクロール・スペース内のすべての文書をクロールする方法を制御します。
- Lotus Notes サーバーのホスト名、ポート、および通信プロトコルを指定します。
- クロールするデータベースまたはディレクトリーを選択します。
- データベースまたはディレクトリーをクロールするスケジュールを設定します。
- クロールする文書を選択します。ディレクトリー内のすべての文書、データベース内のすべての文書、またはデータベース内の選択したビューおよび選択したフォルダー内の文書をクロールすることができます。
- さまざまなデータベース、ビュー、およびフォルダーのフィールドを検索可能にするオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外したり、添付ファイル検索のオプションを指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー

ー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

また、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・ソースで維持されている現行のアクセス制御リストと比較します。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローリングに必要な情報をクローラーに指定する方法が表示されます。

#### 関連概念

228 ページの『Lotus Domino 文書に対する文書レベルのセキュリティーの施行』

クローリングされる Lotus Notes サーバーが、Notes リモート・プロシージャ・コール (NRPC) プロトコルを使用する場合、文書レベルのアクセス制御を施行するように、クローラー・サーバーを構成する必要があります。

#### 関連タスク

228 ページの『ユーザーの信用証明情報を検証するための Lotus Domino トラストド・サーバーの構成』

Notes リモート・プロシージャ・コール (NRPC) プロトコルを使用する Lotus Notes クローラーによってクローリングされたドキュメントのためにセキュリティーを施行するには、クローリングされる Domino サーバーを、Lotus Domino トラストド・サーバーとして構成する必要があります。

## Lotus Domino データベースのクローリングに関するヒント

Lotus Notes クローラーを構成する前に、Lotus Domino データベースのクローリングに関するガイドラインを示します。

- 標準テンプレートに基づく Notes データベース (ディスカッション・データベースなど) は、クローリングするのに最適なタイプのデータベースです。
- Lotus Notes クローラーは、以下のフィールド・マッピング規則を適用します。
  - Domino 標準テンプレートの主フィールド名が最初に登録されます。
  - マッピング規則表に指定されている Notes フィールドの値は、検索結果で文書サマリーとして使用されます。
  - マッピング規則表に指定されていない Notes フィールドの値は、文書サマリーで使用されません。
  - Title フィールドにマップされる Notes フィールドの値は、検索結果で文書タイトルとして使用されます。
  - 以下の表のフィールドは、デフォルトで、以下の検索フィールド名にマップされます。

表 1. デフォルトのフィールド・マッピング規則

Notes データベースのフィールド名	検索フィールド名
Title	Title
EventTitle	Title

表1. デフォルトのフィールド・マッピング規則 (続き)

Notes データベースのフィールド名	検索フィールド名
Subject	Title
Body	Body
Mission	Body
From	Creator
Author	Creator
Keywords	Categories
Categories	Categories
TeamRoomName	Organization
TeamName	Organization
Department	Organization

- Lotus Notes クローラーは、computed for display フィールドを除くすべてのタイプのフィールドをクロールすることができます。
- Notes フォーム上の静的テキストおよびイメージは、クロールされません。
- クローラーを構成する際に、「すべてのフィールドをクロール」チェック・ボックスを選択して、すべてのフィールドをクロールし、クロールされるフィールド・データを最大限にします（「以下を除くすべてのフィールドをクロール」フィールドを使用して、クロールされるフィールドを制限することができます）。

不要なフィールドのクロールを最小限にとどめるには、検索フィールドにマップされているフィールドを除くすべてのフィールドについて「クロール」チェック・ボックスのチェックを外します。

## Lotus Domino ソースをクロールするための UNIX でのクローラー・サーバーの構成

IBM AIX、Linux、または Solaris オペレーティング環境が稼働するコンピューターに WebSphere II OmniFind Edition をインストールした場合で、Notes リモート・プロシージャ・コール (NRPC) プロトコルを使用するサーバーをクロールする予定の場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。そのスクリプトの実行によって、Lotus Notes クローラー、QuickPlace クローラー、および Domino Document Manager クローラーがデータベース・サーバーと通信できるようになります。

### 制約事項

Domino Server は、NRPC プロトコルを使用するよう構成された Lotus Notes クローラー、QuickPlace クローラー、または Domino Document Manager クローラーとは、同じコンピューター上で同時に稼働できません。Domino Server が稼働中にこれらのクローラーのいずれかを開始しようとする、エラーが発生し、クローラーは停止します。

### このタスクについて

NRPC プロトコルを使用するクローラーは、Domino ライブラリーをクライアントとして使用します。これらのライブラリーは、クローラー・サーバーに Lotus

Domino Server バージョン 6.0.2 以降をインストールすることによってインストールされます。クローラーが Domino ライブラリーと連動できるようにするため、Domino ライブラリーをインストールした後、WebSphere II OmniFind Edition によって提供されているセットアップ・スクリプトをクローラー・サーバー上で実行してください。

## 手順

Lotus Notes サーバー、Lotus QuickPlace サーバー、および Domino Document Manager サーバーをクローラーできるようにクローラー・サーバーを構成するには、次のようにします。

1. クローラー・サーバー上にユーザー notes とグループ notes を作成します。

- a. root ユーザーとしてログインします。

```
su - root
```

- b. ユーザーを追加します。

```
useradd notes
```

- c. このユーザーのパスワードを追加します。

```
passwd notes
```

パスワードを変更するようプロンプトが出されます。

2. クローラー・サーバーに Lotus Domino Server をインストールします。
  - a. Domino Server CD バージョン 6.0.2 以降を挿入し、マウントします。(CD がない場合、イメージをダウンロードできます。)
  - b. ご使用のオペレーティング・システムに応じたフォルダーに移動します。

```
AIX: cd /mnt/cdrom/aix
```

```
Linux: cd /mnt/cdrom/linux
```

```
Solaris: cd /mnt/cdrom/solaris
```

- c. インストール・プログラムを開始します。

```
./install
```

- d. プロンプトに回答し、デフォルト値を受け入れるか、希望するインストール設定値 (インストール・ディレクトリーおよびデータ・ディレクトリーのパスなど) を指定します。

Domino Server のインストールで不明な点があれば、Domino 資料を参照してください。

3. クローラー・サーバー上で、WebSphere II OmniFind Edition によって提供されるセットアップ・スクリプトを実行します。

- a. エンタープライズ・サーチ管理者としてログインします (このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです)。

- b. 以下のスクリプトを開始します。これは、\$ES\_INSTALL\_ROOT/bin ディレクトリーにインストールされています。

escrnote.sh

c. プロンプトに応答します。

- 次のプロンプトに対して、Domino Server がデフォルトのディレクトリーにインストールされている場合は Y、そうでない場合は N と応答します。

Lotus Notes ディレクトリー・パス /opt/lotus/notes/latest/linux  
が見つかりました。  
これは正しい Lotus Notes ディレクトリー・パスですか?

AIX の場合のデフォルトのパスは /opt/lotus/notes/latest/ibmpow です。

Linux の場合のデフォルトのパスは /opt/lotus/notes/latest/linux です。

Solaris の場合のデフォルトのパスは /opt/lotus/notes/latest/sunspa です。

- Domino Server がクローラー・サーバー上のデフォルトのディレクトリーにインストールされていない場合、次のプロンプトに対して Domino がインストールされている場所を指定します。

Lotus Notes ディレクトリーのパスを入力してください

例えば、Linux コンピューターで /opt/lotus/notes/latest/linux のように指定します。

- 次のプロンプトに対して、Domino Server データ・ディレクトリーがデフォルトのディレクトリーにインストールされている場合は Y、そうでない場合は N と応答します。

Lotus Notes データ・ディレクトリー・パス /local/notesdata  
が見つかりました。  
これは正しい Lotus Notes データ・ディレクトリー・パスですか?

デフォルトのパスは /local/notesdata です。

- Domino Server データ・ディレクトリーがクローラー・サーバー上のデフォルトの場所にデプロイされていない場合、次のプロンプトに対して、Domino データ・パスを指定します。

Lotus Notes データ・ディレクトリーのパスを入力してください。

4. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

```
esadmin stop
stopccl.sh
startccl.sh -bg
esadmin start
```

## Lotus Domino ソースをクローラーするための Windows でのクローラー・サーバーの構成

Microsoft Windows コンピューターに WebSphere II OmniFind Edition をインストールした場合で、Notes リモート・プロシージャ・コール (NRPC) プロトコルを使用するサーバーをクローラーする予定の場合、クローラー・サーバーを構成するスクリプトを実行する必要があります。そのスクリプトの実行によって、Lotus Notes クローラー、QuickPlace クローラー、および Domino Document Manager クローラーがデータベース・サーバーと通信できるようになります。

### 制約事項

Lotus Domino Server および Lotus Notes クライアントは、NRPC プロトコルを使用するよう構成された Lotus Notes クローラー、QuickPlace クローラー、または Domino Document Manager クローラーとは、同じコンピューター上で同時に稼働できません。Domino Server が稼働中にこれらのクローラーのいずれかを開始しようとすると、エラーが発生し、クローラーは停止します。

### このタスクについて

NRPC プロトコルを使用するクローラーは、Lotus Notes クライアント・ライブラリーを使用します。これらのライブラリーは、クローラー・サーバーに Lotus Notes リリース 6.0.2 以降をインストールすることによってインストールされます。クローラーが Lotus Notes クライアント・ライブラリーと連動できるようにするため、Lotus Notes クライアントをインストールした後、WebSphere II OmniFind Edition によって提供されているセットアップ・スクリプトをクローラー・サーバー上で実行してください。

### 手順

Lotus Notes サーバー、Lotus QuickPlace サーバー、および Domino Document Manager サーバーをクローラーできるようにクローラー・サーバーを構成するには、次のようにします。

1. クローラー・サーバー上で、管理者グループのメンバーであるユーザー ID でログインします。このユーザー ID に Lotus Notes をインストールする権限があることを確認してください。
2. Lotus Notes をインストールします。
  - a. Lotus Notes CD リリース 6.0.2 以降を挿入します。(CD がない場合、イメージをダウンロードできます。)
  - b. インストール・プログラム `setup.exe` を開始します。
  - c. プロンプトに回答し、デフォルト値を受け入れるか、希望するインストール設定値 (インストール・ディレクトリーおよびデータ・ディレクトリーのパスなど) を指定します。

不明な点があれば、Lotus Notes 資料を参照してください。

3. クローラー・サーバー上で、WebSphere II OmniFind Edition によって提供されるセットアップ・スクリプトを実行します。
  - a. エンタープライズ・サーチ管理者 ID でログインします (このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです)。
  - b. 以下のスクリプトを開始します。これは、`%ES_INSTALL_ROOT%\bin` ディレクトリーにインストールされています。

```
escrnote.vbs
```

- c. プロンプトに回答します。
  - 次のプロンプトに対して、Lotus Notes がデフォルトのディレクトリーにインストールされている場合は Y、そうでない場合は N と回答します。

```
Lotus Notes ディレクトリー・パス c:\lotus\notes が見つかりました。  
これは正しい Lotus Notes ディレクトリー・パスですか?
```



Windows コンピューターでの標準的なインストール・パスは、  
c:\lotus\notes または c:\lotus\domino です。

- Lotus Notes がクローラー・サーバー上のデフォルトのディレクトリーにインストールされていない場合、次のプロンプトに対して Lotus Notes がインストールされている場所を指定します。

Lotus Notes ディレクトリーのパスを入力してください

- 次のプロンプトに対して、Lotus Notes データ・ディレクトリーがデフォルトの場所にデプロイされている場合は Y、そうでない場合は N と応答します。

Lotus Notes データ・ディレクトリー・パス c:\lotus\notes\data  
が見つかりました。  
これは正しい Lotus Notes データ・ディレクトリー・パスですか?

Windows コンピューターでの標準的なパスは、c:\lotus\notes\data または c:\lotus\domino\data です。

- Lotus Notes データ・ディレクトリーがクローラー・サーバー上のデフォルトの場所にデプロイされていない場合、次のプロンプトに対して、データ・ディレクトリー・パスを指定します。

Lotus Notes データ・ディレクトリーのパスを入力してください。

4. エンタープライズ・サーチ共通通信層 (CCL) のすべてのセッションを含め、エンタープライズ・サーチ・システムを停止して再始動します。

- a. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```

- b. 「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」を選択して、IBM WebSphere Information Integrator OmniFind Edition サービスを再始動します。

- c. コマンド・プロンプトで、次のようにしてエンタープライズ・サーチ・システムを始動します。

```
esadmin start
```

## DIIOP プロトコルを使用するサーバーの構成

Domino Internet Inter-ORB Protocol (DIIOP) を使用するサーバーをクロールするには、Lotus Notes クローラー、QuickPlace クローラー、および Domino Document Manager クローラーがこのプロトコルを使用できるようにサーバーを構成する必要があります。

### 始める前に

クロールしたいサーバーは、DIIOP タスクおよび HTTP タスクを実行中でなければなりません。

### 手順

DIIOP プロトコルを使用するサーバーを構成するには、次のようにします。

1. サーバー文書を構成します。



- a. クロールしたい Lotus Notes サーバー、Lotus QuickPlace サーバー、または Domino Document Manager サーバー上の server 文書を開きます。この文書は、Domino ディレクトリーに格納されています。
- b. 「構成」ページで、「サーバー」セクションを拡張します。
- c. 「セキュリティー」ページの「プログラマビリティ制約事項」領域の以下のフィールドに、ご使用の環境に適したセキュリティー制約事項を指定します。

- 制限された Lotus Script/Java エージェントの実行
- 制限された Java/Javascript/COM の実行
- 制限されない Java/Javascript/COM の実行

例えば、アスタリスク (\*) を指定して Lotus Script/Java エージェントによる無制限のアクセスを許可し、Java/Javascript/COM 制約事項に対しては Domino ディレクトリーに登録済みのユーザー名を指定します。

**重要:** DIOP プロトコルを使用するこのサーバーをクロールするために構成するクローラーは、これらのフィールドに指定するユーザー名を使用できなければなりません。

- d. 「インターネット・プロトコル」ページを開き、次に「HTTP」ページを開き、「HTTP クライアントにデータベースの参照を許可する」オプションを「はい」に設定します。
2. ユーザー文書を構成します。

- a. クロールしたい Lotus Notes サーバー、Lotus QuickPlace サーバー、または Domino Document Manager サーバー上の user 文書を開きます。この文書は、Domino ディレクトリーに格納されています。
- b. 「基本」ページの「インターネット・パスワード」フィールドに、パスワードを指定します。

エンタープライズ・サーチ管理コンソールを使用してこのサーバーのクロールについてのオプションを構成するときに、このユーザー ID とパスワードを、クロールするサーバーを指定するページで指定します。クローラーは、これらの信用証明情報を使用してサーバーにアクセスします。

3. サーバーで DIOP タスクを再始動します。

## Lotus Domino ソースをクロールするための AIX での I/O 完了ポートの構成

Lotus Notes クローラー、QuickPlace クローラー、または Domino Document Manager クローラーを IBM AIX システム上で使用するには、その前に、I/O 完了ポート (I/O completion port (IOCP)) モジュールをインストールし、このモジュールをクローラーが使用できるよう構成する必要があります。

### このタスクについて

IOCP モジュールがないと、クローラーを作成しようとしたときにディスカバリー処理が失敗します。以下のエラー・メッセージが表示されます。

FFQM0105E サーバーからエラーを受け取りました -  
メッセージ: FFQG0024E 予期しない例外がキャッチされました: discover

以下のメッセージ (ENOEXEC エラーを含んでいます) が、  
\$ES\_NODE\_ROOT/logs/system\_yyyyymmdd.log ファイルに書き込まれます (読みやすく  
するために、一部のメッセージ・テキストは複数行に分けて記述してあります)。

```
5/20/05 18:08:52.423 JST [Error] [ES_ERR_EXCEPTION_DEFAULT_MESSAGE] [] [discovery]
ies10.yamato.ibm.com:0:2108088751:control:ComponentDiscoveryW.java:
com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover:86
FFQ00277E 例外がキャッチされました。詳細 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a'
and a stack trace of 'java.lang.UnsatisfiedLinkError:
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a
at java.lang.ClassLoader$NativeLibrary.load(Native Method)
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)
at java.lang.Runtime.loadLibrary0(Runtime.java:824)
at java.lang.System.loadLibrary(System.java:908)
at lotus.domino.NotesThread.load(NotesThread.java:306)
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)
at com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover
(Unknown Source)
at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)
at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover
(ComponentDiscoveryW.java:72)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
at sun.reflect.DelegatingMethodAccessorImpl.invoke
(DelegatingMethodAccessorImpl.java:60)
at java.lang.reflect.Method.invoke(Method.java:391)
at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)
```

## 手順

IOCP モジュールをインストールし、それがクローラー・サーバー上に正しくインス  
トールされたことを確認するには、次のようにします。

1. AIX 製品 CD から IOCP モジュール (bos.iocp.rte) をクローラー・サーバーにイ  
ンストールします。

IOCP モジュールのインストールを完了したら、Lotus Notes クローラー、  
QuickPlace クローラー、または Domino Document Manager クローラーを作成す  
る前に、このモジュールのソフトウェア修正を適用します。以下のリンクで説明  
を参照してください。

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

2. IOCP モジュールがクローラー・サーバーにインストールされたことを確認す  
るため、以下のコマンドを入力します。

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

lspp コマンドからの出力は、次の例に似たものです。

Fileset	Level	State	Description
-----			
Path: /usr/lib/objrepos			
bos.iocp.rte	5.2.0.10	COMMITTED	I/O Completion Ports API
Path: /etc/objrepos			
bos.iocp.rte	5.2.0.10	COMMITTED	I/O Completion Ports API

3. IOCP ポート状況が **Available** であることを確認するため、以下のコマンドを入力します。

```
$ lsdev -Cc iocp
```

lsdev コマンドからの出力は、次の例と一致するはずです。

```
iocp0 Available I/O Completion Ports
```

4. IOCP ポート状況が **Defined** である場合、次のようにして状況を **Available** に変更します。

- a. root としてクローラー・サーバーにログインし、以下のコマンドを発行します。

```
# smit iocp
```

- b. 「I/O 完了ポートの特性の変更/表示」を選択し、「システム再始動時に構成される状態」を「定義済み」から「使用可能」に変更します。

- c. クローラー・サーバーをリブートします。

- d. もう一度 lsdev コマンドを入力し、IOCP ポート状況が **Available** に変わったことを確認します。

---

## QuickPlace クローラー

Lotus QuickPlace のプレースおよびルームをエンタープライズ・サーチ・コレクションに入れる場合は、QuickPlace クローラーを構成する必要があります。

### クローラー・サーバーの構成

クロールする QuickPlace サーバーが Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用している場合は、クローラー・サーバーでスクリプトを実行する必要があります。このスクリプトは、WebSphere Information Integrator OmniFind Edition で提供されており、これによって、QuickPlace クローラーが NRPC を使用するサーバーと通信できるようになります。

クロールする QuickPlace サーバーが Domino Internet Inter-ORB Protocol (DIIOP) を使用している場合、クローラー・サーバーでセットアップ・スクリプトを実行する必要はありません。ただし、QuickPlace クローラーが QuickPlace サーバーにアクセスできるようにサーバーを構成する必要があります。

クロールする QuickPlace サーバーが Lightweight Directory Access Protocol (LDAP) サーバーを使用している場合は、DIIOP プロトコルを使用するように QuickPlace サーバーを構成する必要があります (QuickPlace クローラーは NRPC プロトコルを使用して LDAP データをクロールできません)。また、Directory Assistance データベースを構成し、LDAP サーバーを 2 次 Lotus Domino Server として使用するよう QuickPlace サーバーを構成する必要があります。

WebSphere II OmniFind Edition が IBM AIX システムにインストールされている場合は、I/O 完了ポート・モジュールがクローラー・サーバーにインストールされており、使用可能であることを確認してください。

エンタープライズ・サーチ管理コンソールを使用して、QuickPlace クローラーを構成する前に、ご使用の環境に該当するタスクを実行する必要があります。

- 68 ページの『Lotus Domino ソースをクロールするための UNIX でのクローラー・サーバーの構成』
- 70 ページの『Lotus Domino ソースをクロールするための Windows でのクローラー・サーバーの構成』
- 72 ページの『DIOP プロトコルを使用するサーバーの構成』
- 77 ページの『ローカル・ユーザー・セキュリティーを使用するための QuickPlace サーバーの構成』
- 78 ページの『QuickPlace での Directory Assistance の構成』
- 73 ページの『Lotus Domino ソースをクロールするための AIX での I/O 完了ポートの構成』

## 構成の概要

QuickPlace クローラーを使用して、任意の数の QuickPlace プレースをクロールすることができます。クローラーを作成する際に、単一の QuickPlace サーバーからクロールするプレースを選択します。後でクロール・スペースを編集して、同じクロール・スペースに別の QuickPlace サーバーから文書を追加できます。クローラーを作成または編集する際に、クロール対象として選択したプレース内のすべてのルームをクロールするか、または、特定のルームをクロールするかを指定できます。

QuickPlace クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティーを指定します。クローラーのプロパティーで、クローラーがクロール・スペース内のすべての文書をクロールする方法を制御します。
- QuickPlace サーバーおよび通信プロトコルを識別します。
- サーバーに関連したユーザー・ディレクトリーに関する情報を指定します (クローラーはこの情報を必要とするので、ユーザーがコレクションを検索の際にアクセス制御が行われます)。
- クロールするプレースを選択します。
- プレースをクロールするスケジュールを設定します。
- クロールする文書を選択します。クローラーは、プレース内のすべてのルームをクロールできるか、またはユーザーが選択したルーム内の特定の文書のみをクロールすることができます。
- さまざまなプレースおよびルームのフィールドを検索可能にするオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外したり、添付ファイル検索のオプションを指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケ

ーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

また、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・ソースで維持されている現行のアクセス制御リストと比較します。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

## ローカル・ユーザー・セキュリティーを使用するための QuickPlace サーバーの構成

ローカル・ユーザー・オプションを使用してセキュリティーをインプリメントするように QuickPlace クローラーを構成する予定の場合、クローラーを作成する前に、Lotus QuickPlace サーバー上に Domino Directory を構成する必要があります。

### このタスクについて

QuickPlace クローラーを構成するときに、文書レベル・セキュリティーを施行するために使用するクローラーのセキュリティー・モードを選択します。「ローカル・ユーザー」モードを選択する場合、すべてのローカル・ユーザー ID およびローカル・グループが Domino Directory (Domino Directory 階層は QuickPlace 階層に対応していなければなりません) に登録されていることを確認する必要があります。

また、クローラーが使用するよう指定するユーザー ID とパスワードが Domino Directory に登録されていて、クロール対象データベースの読み取り許可があることも確認する必要があります。

QuickPlace を使用するには、ユーザー名のみが必要です。しかし、QuickPlace ソースをクロールするには、完全に展開されたユーザー ID が必要です。展開されたユーザー ID のフォーマットは次のとおりです。

`username/placename/QP/domainname`

この手順を使用して、完全に展開されたバージョンのユーザー ID を決定し、このユーザー ID が QuickPlace データベースの読み取りを認可されていることを確認し、ユーザー ID を Domino Directory に追加します。Domino Directory には、QuickPlace データベースのクロールに使用されるユーザー ID と、すべての QuickPlace ローカル・ユーザーおよびグループが含まれている必要があります (Domino Directory 階層は QuickPlace 階層に対応していなければなりません)。

### 手順

ローカル・ユーザー・セキュリティーを使用するために QuickPlace サーバーを構成するには、次のようにします。

1. ユーザー ID 許可を確認します。
  - a. QuickPlace サーバーで Server 文書を開きます。

- b. 「ファイル」ページを開き、クローलするデータベースのアクセス制御リスト (ACL) を開きます。
- c. クローラーが使用するよう構成するローカル・ユーザー ID が ACL に存在することと、このユーザー ID がデータベースの読み取り許可を持っていることを確認します。

このユーザー ID は、ステップ 2 に説明がある完全に展開された形式で指定する必要があります。

2. ユーザーを Domino Directory に追加します。
  - a. QuickPlace サーバーで Server 文書を開きます。
  - b. 「ユーザーとグループ」ページのユーザー・ツリー項目に、ステップ 1 で確認した、完全に展開されたユーザー ID を追加します。
  - c. 「インターネット・パスワード」フィールドに、このユーザー ID のパスワードを指定します。

## QuickPlace での Directory Assistance の構成

LDAP ディレクトリーを使用してセキュリティーをインプリメントするように QuickPlace クローラーを構成する予定の場合、クローラーを作成する前に、Lotus QuickPlace サーバー上に Directory Assistance データベースを作成する必要があります。

### 制約事項

クローラ対象にする QuickPlace サーバーは、DIIOP タスクおよび HTTP タスクを実行中でなければなりません。

### 手順

QuickPlace サーバー上に LDAP Directory Assistance を構成するには、次のようにします。

1. Directory Assistance データベースを作成します。
  - a. QuickPlace サーバーで Server 文書を開きます。
  - b. **Directory Assistance(6)** テンプレートを使用してデータベースを作成します。このテンプレートは、サーバー上にあります。
  - c. データベース内に文書を作成するため「**Directory Assistance の追加**」をクリックします。
  - d. 「基本」タブを開き、「**DomainType**」フィールドで「**LDAP**」を選択します。
  - e. 「命名コンテキスト」タブを開き、「**信用証明情報に対してトラステッド**」チェック・ボックスが選択されていることを確認します。
  - f. 「LDAP」タブで LDAP サーバーに関する情報を指定します。
  - g. Server 文書を保管してクローズします。
2. 次のように、Directory Assistance データベースを使用するための QuickPlace サーバーの構成を行います。
  - a. QuickPlace サーバーで Server 文書を開きます。



- b. 「基本」タブを開き、「**Directory Assistance データベース名**」フィールドに、ステップ 1 で作成したデータベースの名前を指定します。
- c. Server 文書を保管してクローズします。

これで、QuickPlace サーバーは、2 次 Domino ディレクトリーとして LDAP サーバーを使用できるようになります。

---

## UNIX ファイル・システム・クローラー

UNIX ファイル・システムに保管されている文書をエンタープライズ・サーチ・コレクションに入れる場合は、UNIX ファイル・システム・クローラーを構成する必要があります。

UNIX ファイル・システム・クローラーを使用して、任意の数の UNIX ファイル・システムをクローラーすることができます。クローラーの構成時に、クローラーするローカルおよびリモートのディレクトリー、サブディレクトリーを選択します。

Windows コンピューターにクローラー・サーバーをインストールした場合、そのサーバーを使用して UNIX ファイル・システム・ソースをクローラーすることはできません (UNIX ファイル・システム・クローラーは、選択可能なクローラー・タイプのリストに表示されません)。

UNIX ファイル・システム・クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクローラー・スペース内のすべてのサブディレクトリーに対するクローラーを制御します。
- ファイル・システムをクローラーするスケジュールを設定します。
- クローラーがクローラーするサブディレクトリー、サブディレクトリーのレベルを選択します。
- サブディレクトリー内の文書を検索可能にするときのオプションを指定します。例えば、特定タイプの文書をクローラー・スペースから除外することができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクローラーに必要な情報をクローラーに指定する方法が表示されます。



---

## Web クローラー

Web サイトのページをエンタープライズ・サーチ・コレクションに入れる場合は、Web クローラーを構成する必要があります。

Web クローラーを使用すれば、任意の数の HTTP サーバーおよびセキュア HTTP (HTTPS) サーバーをクロールすることができます。クローラーが Web サイトにアクセスし、サイト上のデータを読み取ります。その後、文書内のリンクをたどって他の文書をクロールします。Web クローラーは、個々のページあるいはフレームセット (HTML フレームを使用して作成されたページ) のリンクをクロールして、抜き出すことができます。

クロールされたデータは、フォーマットもソースも様々です。一般的に使用されるフォーマットは数多くありますし、イントラネットまたはインターネットの様々なところから取得されます。一般的なフォーマットには、HTML、PDF、Microsoft Word、Lotus WordPro、Extensible Markup Language (XML) などがあります。

Web クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。また、この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクロール・スペース内の全 Web ページに対するクロールを制御します。
- Web サイトへのアクセスを許可または禁止するための規則を指定します。クロール規則を指定する場合、規則をテストして、ユーザーがクロール・スペースに入りたいサイトにクローラーがアクセスできるかどうか検証することができます。
- 特定のタイプのファイルを含めるためのオプション、および特定のファイル拡張子を持つファイルを除外するためのオプションを指定します。
- Web クローラーがソフト・エラーのページを処理するための規則を指定します。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。
- パスワード保護された Web サイトをクロールする場合のオプションを指定します (クロール対象の Web サーバーは HTTP 基本認証、またはパスワード入力のプロンプトを出す HTML フォームを使用する必要があります)。
- プロキシ・サーバーによってサービスされている Web サイトをクロールするためのオプションを指定します。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

## ユーザー・エージェントの構成

Web クローラーは、ロボット排他プロトコルに従います。このプロトコルを使用する Web サイトをクロールするには、その Web サイト上の robots.txt ファイルで、Web クローラー用に設定したユーザー・エージェント名が Web サイトへのアクセスを許可されるようにしてください。

エンタープライズ・サーチ・システムが開始されると、Web クローラーは、構成されたユーザー・エージェント名をロードします。クローラーが、まだ訪れたことのない (あるいは、しばらく訪れていない) Web サイトからページをダウンロードする前に、クローラーは、まず robots.txt というファイルをダウンロードしようとします。このファイルは、Web サイトのルート・ディレクトリーにあります。

robots.txt ファイルがない場合、その Web サイトは、制限なくクロールできるように公開されます。このファイルがある場合、それは、サイト (ディレクトリー) のどの領域がクローラーの立ち入り禁止になっているかを明示します。robots.txt ファイルは、ユーザー・エージェント名を特定してクローラーに許可を出します。

Robots Exclusion プロトコルは、任意のものですが、エンタープライズ・サーチ Web クローラーはこれに従います。

- robots.txt ファイルに、Web クローラー用に設定されたユーザー・エージェント名の項目が含まれている場合、Web クローラーはそのユーザー・エージェント上の制約に従います。
- ユーザー・エージェント名が robots.txt ファイルになく、最後の項目が User-agent: \* (任意のユーザー・エージェントを意味する) と指定されていて、その制約が Disallow: / (この Web サイトのルートを起点とするいかなるクロールも許可しないことを意味する) になっている場合、Web クローラーはそのサイトのクロールができません。
- ユーザー・エージェント名が robots.txt ファイルになく、最後の項目が User-agent: \* と指定されていて、その制約が Allow: / になっている場合は、Web クローラーはそのサイトをクロールすることができます。

Web サイト管理者は、明示的にアクセスを認可されていないすべてのクローラーに対してアクセスを禁止する最後の項目を指定することがよくあります。新規に Web クローラーを構成するとき、クロール対象の Web サイトの一部で Robots Exclusion プロトコルを使用していることがわかっている場合は、Web サイト管理者に依頼して、クローラーの項目をその robots.txt ファイルに追加してもらってください。

Web クローラーのプロパティーと対象とする Web サイトに含まれるすべての robots.txt ファイルに同じユーザー・エージェント名が指定されていることを確認してください。

クロール予定の Web サイトのいずれも Robots Exclusion プロトコルを使用していない場合は、ユーザー・エージェント・プロパティーに指定する値は通常、何でもかまいません。ただし、一部のアプリケーション・サーバー、JSP、およびサーブレットでは、ユーザー・エージェント名に対応するように調整しています。例えば、ブラウザの非互換性に対処するために異なる応答が用意されています。このような場合は、Robots Exclusion プロトコルの使用に関係なく Web クローラーに指定し

たユーザー・エージェント名が重要になります。このタイプのサイトをクロールする必要がある場合は、Web サイト管理者に相談して Web クローラーのアクセスが許可されるようにしてください。

## JavaScript のサポート

エンタープライズ・サーチの Web クローラーが、Web 文書の JavaScript™ 部分に含まれているリンク (URL) を検出する場合があります。

Web クローラーは、相対リンクと絶対リンクの両方を検出できます。HTML 文書に BASE エlementが含まれている場合、クローラーはそのElementを使用して相対リンクを解決します。そうでない場合は、クローラーは、文書固有の URL を使用します。

JavaScript のサポートはリンク抽出に限定されています。クローラーは JavaScript の構文解析、DOM (Document Object Model) の構築、および、JavaScript ステートメントの解釈や実行は行いません。クローラーは JavaScript ステートメントの URL である可能性の高いストリングを文書コンテンツ (JavaScript 部分を含むがそれに限定しない) で検索します。これは 2 つのことを意味しています。

- 一部の URL は検出されても、厳しい HTML パーサーでは無視されます。クローラーは、構文的に有効でない URL はすべて拒否しますが、スキャンのステップで戻される有効な URL の中には、検索にあまり重要でないものがあります。
- 人がブラウザでページを表示させて、そのブラウザが JavaScript を実行するときなどに JavaScript によって生成される文書コンテンツは、Web クローラーでは見ることができず、したがって、索引付けもされません。

## Web クロール・スペースを制限する規則

ユーザーが、検索を許可されている Web サイトのみにアクセスするように、Web クローラーのクロール対象を制限する規則を指定します。

Web クローラーが Web ページをクロールするときに他のページへのリンクをディスカバーすると、それらのリンクを次にクロールされるようにキューに入れます。このクロールとディスカバーは、時間とメモリー・リソースが許す限り繰り返して行われます。Web クローラーを構成する場合、クローラーがクロールを開始する場所を指定します。Web クローラーは、この初期 URL (開始 URL といいます) から始めて、直接または間接的に接続されている Web 上のあらゆる文書に到達できます。

クロール・スペースを制限するには、Web クローラーを構成するときに特定の URL は完全にクロールし、関心がある領域外を示すリンクは無視するように設定します。デフォルトではクローラーは、ディスカバーしたすべての URL を受け付けるようになっているため、コレクションに入れる URL を指定し、それ以外のページは除外する規則を指定する必要があります。

Web クローラーにクロール対象とクロールから除外する対象を指定するには、いくつかの方法があります。次のような指定を行うことができます。

- クローラーがクロールを開始する開始 URL のリスト
- 3 種類のクロール規則 (ドメイン、IP アドレス、URL 接頭部)
- クロール対象の文書を指定する MIME タイプのリスト

- 除外対象の文書を指定するファイル拡張子のリスト
- 1 つの URL パスに許容される最大ディレクトリー数

クロール規則は、次の書式で指定します。

```
action type target
```

`action` は `forbid` または `allow`、`type` はドメイン (domain)、IP アドレス (IP address)、または URL 接頭部 (HTTP または HTTPS) で、`target` は `type` の値によって異なります。パターンに一致するターゲットの指定には、ワイルドカード文字としてアスタリスク (\*) を指定できますが、その使用は限定的になります。

## ドメイン規則

ドメイン規則のターゲットは DNS ドメイン・ネームです。例えば、`www.ibm.com` ドメイン全体をクロール対象にする場合、次のように指定します。

```
allow domain www.ibm.com
```

ターゲットの最初の文字は、アスタリスクでもかまいません。その場合、ホスト名の末尾が後半部分のパターンに一致するあらゆるホスト名がこの規則に該当することになります。例えば、次のパターンに一致するすべてのドメインをクロール対象から除外する場合、次のように指定します。

```
forbid domain *.ibm.com
```

ホスト名の突き合わせでは、具体的なドメイン・ネームを明示して指定した場合でも、ドメイン・ネームのパターンを指定した場合でも大/小文字は区別されます。例えば、「`*.user.ibm.com`」と指定すると、「`joe.user.ibm.com`」および「`mary.smith.user.ibm.com`」は一致となりますが、「`joe.user.IBM.com`」は一致となりません。

ポート番号を指定しないドメイン規則は、そのドメイン上のすべてのポートに適用されます。以下の例では、`sales` ドメインのすべてのポートが許可されます。

```
allow domain sales.ibm.com
```

ドメイン規則でポート番号を指定する場合、その規則はそのポートのみに適用されます。以下の例では、`sales` ドメインのポート 443 のみ許可されます。

```
allow domain sales.ibm.com:443
```

## 接頭部規則

接頭部規則は、指定されたストリングで始まる URL のクロールを制御します。ターゲットは、単独の URL で、通常、1 つ以上アスタリスクを入れてパターンを指定します。例えば、よくある方法として、接頭部ストリングの末尾文字にアスタリスクを指定します。

接頭部規則を使用すれば、Web サイトのすべて、または一部をクロールすることができます。ディレクトリー・パスかパターンを指定し、ディレクトリー・ツリーのそれ以降のすべてを対象として、クロールを許可または禁止します。例えば、次の 2 つの規則では、`sales.ibm.com` の `public` ディレクトリーのすべてのコンテンツをクローラーのクロール対象とし、そのサイトの他のすべてのページについてはアクセスを禁止します。

```
allow prefix http://sales.ibm.com/public/*
forbid prefix http://sales.ibm.com/*
```

接頭部規則を指定する場合、アスタリスクは複数指定が可能で、末尾に限らず接頭部ストリングの任意の位置に指定できます。例えば、次の規則では、sales.ibm.com サイトでディレクトリー名の末尾が fs の最上位ディレクトリーの文書を、クローラーがクロールするのを禁止します。(検索索引で有用な情報がないファイル・システム・マウントなどの場合が考えられます。)

```
forbid http://sales.ibm.com/*fs/*
```

## アドレス規則

アドレス規則は、ターゲットとして IP アドレス (IPv4 のみ) とネットマスクを指定することにより、ホストまたはネットワーク全体のクロールを制御します。例えば、次のように指定します。

```
allow address 9.0.0.0 255.0.0.0
```

ネットマスクを使用するとパターン・マッチングを指定できます。アドレス規則を候補となる IP アドレスに適用するには、規則の IP アドレスと候補 IP アドレスが、ネットマスクでゼロで隠されている部分を除き、同一でなければなりません。アドレス規則はパターンを定義し、ネットマスクはアドレス・パターン中の有意のビット部分を定義します。ネットマスク中のゼロは、ワイルドカードのような働きをし、アドレス内でゼロと同じビット位置にある値はすべて一致すると見なされます。

上記の例の allow 規則は、最初のオクテットが 9 で末尾の 3 つのオクテットが任意の値のすべての IP アドレスに適合します。

規則リストの最後に入れるアドレス規則として、「forbid address 0.0.0.0 0.0.0.0」を入れておくと便利です。この規則は、ネットマスクの全ビットを有意でないビットとしているため、あらゆる IP アドレスが一致となります (この規則では、規則リストでこの前のすべての規則で許可されていないすべてのアドレスを禁止することになります)。

**プロキシ・サーバーに対する制限:** プロキシ・サーバーによってサービスされている Web サイトをクロールする場合は、IP アドレス規則を指定しないでください。プロキシ・サーバーは、通常、ユーザー・エージェント (ブラウザまたはクローラー) が、Web サーバーがあるネットワークに対して直接アクセス権限を持っていない場合に使用されます。例えば、HTTP プロキシ・サーバーは、HTTP 要求をクローラーから Web サーバーに中継し、それに対する応答をクローラーに送信することができます。

Web クローラーがプロキシ・サーバーを使用する場合、プロキシ・サーバーの IP アドレスは、クローラーが別のホスト用に持っている唯一の IP アドレスです。IP アドレス規



則が、クローラーを IP アドレスのサブネットに制約するために使用される場合、制約によってほとんどすべての URL が戻りコード 760 に分類されます (この戻りコードは、URL が Web スペースによって禁止されていることを示します)。

## クローラー規則の順序

クローラーは、URL のディスカバーおよびクローラー処理中、クローラー規則を何度も適用します。規則の順序は重要ですが、それは同タイプの規則に限られます。あるアドレス規則が別のアドレス規則の前に来るか、後ろに来るかでは意味が違いますが、あるアドレス規則が別の接頭部規則の前に来るか、後ろに来るかでは意味の違いはありません。クローラーがこの 2 つの規則を同時に適用することはないからです。

同タイプの規則セット内については、クローラーは候補ドメイン、アドレス、または URL を最初に指定した規則から順次最後の規則まで適合する規則が見つかるまで 1 つずつテストします。適合する最初の規則に指定された処置が使用されます。

順序に依存しているため、多くのクローラー規則は以下のような典型的構成になります。

- ドメイン規則のセットは、通常、クローラー・スペースから単独のドメインを除外する `forbid` 規則を先に置きます。例えば、コレクション管理者は、有用な情報がないドメインをここで除外指定します。
- `forbid` 規則のリストの後には、通常、ワイルドカード文字を入れた一連の `allow` 規則を設定し、これによって、エンタープライズ・イントラネットを定義する上位レベルのドメイン・ネームが末尾に来る任意のドメイン (`*.ibm.com`、`*.lotus.com` など) をクローラーがアクセスできるようにします。

ドメイン規則のセットの最後に次のデフォルト規則を入れ、それより前の規則で許可されなかったドメインをクローラー対象から除外します。

```
forbid domain *
```

この最後の規則は、クローラー・スペースにインターネット全体が入らないようにするため、重要です。

- アドレス規則のセットは、通常、少数の `allow` 規則が先頭に置かれ、これによってクローラーがエンタープライズ・イントラネットにわたる上位レベル (クラス A、クラス B、またはクラス C) のネットワークをクローラーできるようにします。

アドレス規則の最後は、通常、次の規則が置かれ、これによってクローラーが企業ネットワーク外の Web サイトをクローラーしないようにします。

```
forbid 0.0.0.0 0.0.0.0
```

- 通常、接頭部規則のセットは、ツリーまたはサブツリーで許可および禁止指定される領域が詳細にその都度指定されるため、この規則セットが一番大きなセットになります。よい方法としては、まず、より特定化された領域を許可または禁止し、それから逆に一般的なパターンでその他すべての領域を許可または禁止します。

接頭部セクションの最後は、通常の規則とは異なります。ドメイン規則およびアドレス規則について推奨される最終規則を使用すれば、URL の接頭部をテストするよりも効率的に、クローラーがエンタープライズ・ネットワークを超えてクローリングしないようにすることができます。

接頭部規則の場合、処置 (forbid または allow) 別に規則をまとめるほうが、クローラーはより効率的に処理を行うことができます。例えば、allow および forbid 規則を交互に短い指定で並べるよりも、片方の処置を規定する規則を長く並べ、それから他方の処置を規定する規則を同様に長く並べるようにします。allow 規則と forbid 規則を混在させるようにしても目的のクローリング・スペースを作成することはできますが、allow 規則と forbid 規則をそれぞれひとまとめにしたほうが、クローラーのパフォーマンスは良くなります。

## ファイル拡張子、MIME タイプ、およびクローリングの最大の深さ

これらのオプションを使用すれば、クローリング・スペースの内容をさらに詳細に指定することができます。特定タイプの文書を文書のファイル拡張子に基づいて除外したり、文書の MIME タイプを基に特定タイプの文書を組み入れたりすることができます。クローラーにクローリングさせる MIME タイプを指定する場合、Web 文書では MIME タイプが誤って設定されていることが多いので注意してください。

クローリングの最大の深さは、サイト・ルートから数えた URL 内のスラッシュ数で表されます。このオプションを使用すると、クローラーが深さが無限の再帰的なファイル・システム構造にはまり込むのを防ぐことができます。クローリングの深さは、クローラーがある文書から別の文書にリンクをトラバースする時のレベル数とは一致しません。

## 開始 URL

開始 URL は、クローラーがクローリングを開始する URL のことを指し、クローラーが開始されるたびにこの URL がクローリング対象となります。開始 URL がすでにディスカバーされた場合は、その URL がクローリング規則で許可された他の Web サイトよりも早くクローリングまたは再クローリングされることはありません。

開始 URL は Web クローラーが初めて始動され、クローリング・スペースが空の場合に重要です。また、まだディスカバーされていなかった開始 URL をクローリング・スペースの開始 URL のリストに追加するときにも重要となります。

開始 URL にはドメイン・ネームだけでなく完全修飾 URL を指定する必要があります。プロトコル、およびポートが 80 でない場合は、ポート番号も指定してください。

以下に、有効な開始 URL の例を示します。

```
http://w3.ibm.com/  
http://sales.ibm.com:9080/
```

以下の URL は無効な開始 URL です。

```
www.ibm.com
```

また、開始 URL をクローリング規則に入れる必要があります。例えば、クローリング規則で開始 URL をクローリング対象として許可していなければ、クローラーは指定された



開始 URL でクローリングを開始することができなくなります。

## Web クローラーとの URL 接続のテスト

Web クローラーがクローリングする URL を指定したら、クローリング規則の構成をテストできます。

テスト結果により、クローラーのプロパティに指定されているユーザー・エージェント名を使用して、クローラーが URL にアクセスできるかどうかを示されます。また、テスト結果により、URL が排他規則によりクローリングできないかどうかを示されます (例えば、クローリング・スペースから除外されている拡張子に一致するファイル拡張子を持つ文書はクローリングされません)。

サイトが少なくとも 1 度クローリングされた後、URL をテストして、追加情報を入手できます。例えば、テスト・レポートにより、最新の HTTP 戻りコード (URL のクローリングが正常に行われたかどうかを示します)、URL の最終クローリング日時、再クローリングがスケジュールされている日時、ユーザー・エージェントが Web サーバーの現在の robots.txt ファイルを使用しているかどうか、などが示されます。

## Web クローラーにおける再クローリング間隔の設定

Web クローラーが URL に再アクセスする頻度を調整するには、Web クローラーのプロパティのオプションを指定します。

エンタープライズ・サーチ・システムの他のほとんどのクローラー・タイプの場合、管理者が指定したスケジュールに従ってクローリングが実行されます。これに対して、Web クローラーの場合、通常、開始されると連続して実行されます。前にクローリングした URL に再アクセスする頻度を調節するには、最大および最小再クローリング間隔を指定します。

エンタープライズ・サーチ管理コンソールを使用して Web クローラーを作成する、または Web クローラーのプロパティを編集するときに、拡張プロパティを構成するオプションを選択できます。最大および最小再クローリング間隔は、この「Web クローラーの拡張プロパティ」ページで指定します。Web クローラーは、指定したこの値を使用して、データの再クローリング間隔を計算します。

あるページが最初にクローリングされると、クローラーはそのページがクローリングされた日時と指定された最大と最小再クローリング間隔の平均値を使用して、再クローリング日付を設定します。そのページはその日付までは再クローリングされません。その日付以後、そのページが再クローリングされる実際の日時は、クローラー負荷とクローリング・スペース内の新旧の URL のバランスによって異なります。

毎回、そのページが再クローリングされるごとに、クローラーはその内容に変更がないか確認します。内容に変更がある場合、次の再クローリング間隔は前回よりも短くなりますが、指定された最小再クローリング間隔よりも短くなることはありません。内容に変更がない場合、次の再クローリング間隔は前回よりも長くなりますが、指定された最大再クローリング間隔よりも長くなることはありません。

## Web クローラーにおける URL アクセスのオプション

Web クローラーには、特定の URL をできる限り早くアクセスするように強制することができます。

特定の Web サイトの情報でクロール・スペースをリフレッシュする必要がある場合、クローラーをモニターし、「**アクセスまたは再アクセスする URL**」オプションを選択して、クロールまたは再クロールする必要があるページの URL または URL パターンを指定します。

例えば、企業の情報通信部門がイントラネットにある Web ページを追加する、または重要な方針変更があったため、それを反映するようにページを改訂する場合、その新規、または変更済みのページの URL を指定します。クローラーが実行中の場合、クローラーは指定された URL を次にクロールするようにキューに入れ、アクセスを待機しているページがないかその URL を確認します (通常 10 分ごと)。クローラーが実行されていない場合は、クローラーは指定された URL をキューに入れ、クローラーが次回開始されたときにクロールできるようにします。

クロール規則の中に、指定した URL のアクセスをクローラーに許可する規則が必ず含まれるようにしてください。クローラーは、指定した URL には通常より早くアクセスできますが、前提としてすべてのクロール対象の URL は、その URL のクロールを許可するクロール規則がなければなりません。

新規にクロールされたデータは、次回索引が再編成またはリフレッシュされた時点で検索可能になります。

## Web クローラーがソフト・エラーのページを処理する方法

特定ページ要求の返答に標準的なエラー・コードを返したくない場合に、Web サイト管理者が作成したカスタム・ページを処理する Web クローラーを構成することができます。

HTTP サーバーが、クライアントが要求したページを返すことができない場合、サーバーは、通常、ヘッダーから成る応答に戻りコードを付けて返します。 戻りコードで、何が問題であることを示します (例えば、ファイルが見つからなかったことを表すエラー 404 など)。Web サイト管理者によっては、問題をより詳細に説明する特別なページを作成し、代わりにそのページを戻すように HTTP サーバーを構成します。 こうしたカスタム・ページは、ソフト・エラー・ページと呼ばれます。

ソフト・エラー・ページは、Web クローラーの結果をゆがめることがあります。例えば、クローラーは、問題を表すヘッダーを受け取る代わりに、ソフト・エラー・ページと戻りコード 200 (有効な HTML ページの正常なダウンロードを表す) を受け取ることがあります。しかし、ダウンロードされるこのソフト・エラー・ページは要求された URL には関連せず、その内容は、要求されたページの代わりに戻されるときは毎度、ほとんど同じです。関連のない、ほとんど重複しているこれらのページが、索引および検索結果をゆがめます。

この状況に対応するために、Web クローラーの構成時にソフト・エラー・ページの処理オプションを指定することができます。 Web クローラーは、ソフト・エラー・ページを返す各 Web サイトについて以下の情報を必要とします。

- ソフト・エラー・ページを使用するサイトの URL パターン。この URL パターンは、プロトコル (HTTP または HTTPS)、ホスト名、ポート番号 (標準でない場合)、およびパス名で構成されます。パターンには、ワイルドカード文字としてア

アスタリスク (\*) を使用することができ、非ワイルドカード文字が次に現れる位置までの 1 字以上の文字とマッチングさせることができます。指定するパターンは、大/小文字を区別します。

- HTML 文書の <TITLE> タグに対応するテキストのタイトル・パターン。このパターンの指定には、ワイルドカード文字としてアスタリスク (\*) を使用できません。指定するパターンは、大/小文字を区別します。
- HTML 文書のコンテンツに対応するテキストのコンテンツ・パターン。<BODY> タグがある場合、コンテンツは、単に <BODY> タグの内容だけではありません。コンテンツは、ファイルの HTTP ヘッダーの後にあるすべてです。このパターンの指定には、ワイルドカード文字としてアスタリスク (\*) を使用できます。指定するパターンは、大/小文字を区別します。
- 指定した URL、タイトル、およびコンテンツ・パターンと一致する文書に使用する戻りコードを表す整数。

## 例

以下の構成は、Web クローラーに、`http://www.mysite.com/hr/*` Web サイトから戻される、すべての有効な HTML ページ (戻りコード 200) を、指定されたタイトルおよびコンテンツ・パターンと比較するように指示します。ページの <TITLE> タグが、「Sorry, the page」で始まっている、文書のコンテンツに任意の文字 (\*) が含まれている場合には、クローラーはそのページを、戻りコード 404 を返す (ページが見つからなかった) 場合と同じ方法で扱います。

表 2. ソフト・エラー・ページの例

URL パターン	タイトル・パターン	コンテンツ・パターン	HTTP 戻りコード
<code>http://www.mysite.com/hr/*</code>	<code>Sorry, the page*</code>	<code>*</code>	404

異なる戻りコードを扱うために、同一の Web サイトについて複数の項目を作成できます。同一の Web サイトからの戻りコードはそれぞれ、Web クローラーの構成で固有の項目を必要とします。

## ワイルドカード文字の使用

URL、タイトル、および、コンテンツ・パターンは、正規表現ではありません。アスタリスク文字は、非ワイルドカード文字が次に現れる位置までの任意の文字と一致します。例えば、以下ようになります。

\*404 は、*any characters*404 と一致します。

404: \* は、404: *any characters* と一致します。

`http://*.mysite.com/*` は、`http://any host.mysite.com/任意のファイル名` と一致します。

\* は、*any characters* と一致します。

## パフォーマンスへの影響

ソフト・エラー・ページの処理オプションを構成すると、正常にクロールされたページをすべて検査する必要があるため、クローラー・プロセスの時間が長くなります。パターン・マッチングの検査と、ページまたは置き換えた戻りコードを返すか

どうかの判断に、さらに処理時間が必要になります。

## セキュア Web サイトのクロールのサポート

エンタープライズ・サーチ管理コンソールに信用証明情報を指定することで、アクセスにパスワードが必要な文書などの制限付きコンテンツに Web クローラーが、アクセスできるように設定することができます。

Web サーバーが HTTP 基本認証または HTML フォーム・ベースの認証を使用して、Web サイトへのアクセスを制限している場合、Web クローラーの構成で信用証明情報を指定して、パスワード保護された Web サイト上のページをクロールできるようにします。また、Cookie ファイルを手動で構成するオプションも指定できます。

### HTTP 基本認証によって保護された Web サイト

Web サーバーが HTTP 基本認証を使用して Web サイトへのアクセスを制限している場合、Web クローラーがパスワード保護されたページにアクセスできるように認証のための信用証明情報を指定してください。

ユーザー (あるいはクライアント・アプリケーション) に Web サイト上のページへのアクセス権があるかどうかを判別するために、多くの Web サーバーが HTTP 基本認証と呼ばれるクライアント認証方式を採用し、ユーザーの身元を確認しています。通常、この操作は、対話式に行われます。

- HTTP ユーザー・エージェント (Web ブラウザーなど) が HTTP 基本認証で保護されているページを要求すると、Web サーバーは、要求側に要求ページへのアクセス権限がないことを意味する、戻りコード 401 を返します。
- また、Web サーバーは、ユーザーが制限付きコンテンツへのアクセス許可を得ているかどうかを検証するために信用証明情報を提示するよう要求側に求めます。
- Web ブラウザーは、ユーザー名、パスワード、その他のユーザーの信用証明情報に必要な情報を要求するダイアログをユーザーに表示します。
- Web ブラウザーは信用証明情報をエンコードして、保護されたページの要求のたびにそれを組み込みます。
- 信用証明情報が有効であれば、Web サーバーは、戻りコード 200 と要求ページのコンテンツを返します。
- それ以降の同じ Web サーバーからのページ要求には、通常、同じ信用証明情報が組み込まれ、許可ユーザーが要求のたびごとに信用証明情報の入力を求められることなく、他の制限付きコンテンツにアクセスできるようになります。

一度、ユーザーの身元が確認されると、Web サーバーと HTTP ユーザー・エージェントは、通常、Cookie というトークンを交換し、複数の HTTP 要求にわたって維持する必要のあるユーザー・ログイン状況の認識を可能にします。

Web クローラーは対話式で動作しないので、パスワード保護されたページのクロールを可能にするための信用証明情報を、クロールの開始前に指定しておく必要があります。Web クローラーを作成したり、クロール・スペースを編集する場合、クロール対象のセキュア Web サイトそれぞれについて情報を指定してください。

この情報を指定するには、HTTP 基本認証によって保護されている Web サイトまたは Web サーバーの管理者と緊密に作業する必要があります。管理者は、クロー

ル対象の Web サイトのセキュリティー要件を提供してくれるはずですが。これには、Web クローラーの識別の認証や、制限付きページをクロールするアクセス権がクローラーにあるかどうかの判別に使用するあらゆる情報が含まれます。

コレクションの作成時にそのコレクションに対してセキュリティーを使用可能に設定している場合は、文書へのアクセスを制御するために、ユーザー ID、グループ ID、あるいはユーザー役割などのセキュリティー・トークンを、クローラー構成時に指定してください。Web クローラーは、このセキュリティー・トークンを、指定されたルート URL のファイル・システム・ツリーでクロールするあらゆる文書に関連付けます。このトークンは Web クロール・スペース全体に対して設定する文書レベルのセキュリティー・トークンとは別に使用されます。

URL の順序は、重要です。パスワード保護された Web サイトに関する情報を追加したら、それをクローラーに処理させたい順序に配置する必要があります。リスト内では、特定の URL を最初に置き、汎用的な URL が続くように配置してください。Web クローラーは、候補の URL を評価するとき、その候補の URL と一致した、リストの最初にある URL に対して指定された認証データを使用します。

### フォーム・ベースの認証によって保護された Web サイト

Web サーバーが HTML フォームを使用して Web サイトへのアクセスを制限する場合、Web クローラーがパスワード保護されたページにアクセスできるようにする認証のための信用証明情報を指定できます。

ユーザー (あるいはクライアント・アプリケーション) に Web サイト上のページへのアクセス権があるかどうかを判別するために、多くの Web サーバーで HTML フォームを使用して、ユーザーの身元を確認しています。通常、この操作は、対話式に行われます。

- HTTP ユーザー・エージェント (Web ブラウザーなど) が フォーム・ベースの認証で保護されているページを要求すると、Web サーバーは、要求にユーザーの身元を証明する Cookie が含まれているかどうかを確認します。
- Cookie がない場合は、Web サーバーはユーザーにプロンプトを出して、セキュリティー・データをフォームに入力するよう求めます。ユーザーがフォームをサブミットすると、Web サーバーは、必要な Cookie を戻し、それからパスワード保護されたページの要求処理が進められます。
- また、必要な Cookie を組み込んでいれば、その後の要求も処理されます。許可ユーザーは、要求のたびごとにフォームへの入力や信用証明情報の指定を求められることなく、他の制限付きコンテンツにアクセスできます。

Web クローラーは対話式で動作しないので、パスワード保護されたページのクロールを可能にするための信用証明情報を、クロールの開始前に指定しておく必要があります。Web クローラーを作成したり、クロール・スペースを編集する場合、クロール対象のセキュア Web サイトそれぞれについて情報を指定してください。

指定したフィールドは、対話式ユーザーが Web ブラウザーにプロンプトで求められて入力するフィールド、および、ログイン実行に必要な隠しフィールドや静的フィールドと対応します。

この情報を指定するには、フォーム・ベースの認証によって保護されている Web サイトまたは Web サーバーの管理者と緊密に作業する必要があります。管理者



は、クローリング対象の Web サイトのセキュリティー要件を提供してくれるはずで  
す。これには、Web クローラーの識別の認証や、制限付きページをクローリングする  
アクセス権がクローラーにあるかどうかの判別に使用するあらゆる情報が含まれま  
す。

URL パターンの順序は、重要です。パスワード保護された Web サイトに関する情  
報を追加したら、それをクローラーに処理させたい順序に配置する必要があります  
。リスト内では、特定の URL パターンを最初に置き、汎用的な URL パター  
ンが続くように配置してください。Web クローラーは、候補の URL を評価する  
とき、その候補の URL と一致した、リストの最初にある URL パターンに対して指  
定されたフォーム・データを使用します。

## プロキシ・サーバーによってサービスされている Web サイト

Web クローラーがネットワークへの直接アクセスを許可されていない場合、クロー  
ラーが HTTP プロキシ・サーバーを使用してクローリング対象のコンテンツにアクセ  
スするように構成することができます。

Web クローラーが稼働するコンピューター上で TCP/IP ネットワークへのアクセス  
が使用できない場合、あるいは、アクセスが特権のある処理に制限されている場合  
は、Web クローラーで HTTP プロキシ・サーバーを使用するように構成できま  
す。HTTP プロキシは、指定されたホストの指定されたポートで HTTP 要求を  
listen するプロセスです。プロキシ・サーバーは要求を Web サーバーにつなぎ、  
Web サーバーからの応答を要求元のクライアント (Web クローラー) にリレーしま  
す。プロキシ・サーバーは、Web クローラーと同じコンピューター上で稼働する  
ことも、別のコンピューター上で稼働することもできます。

プロキシ以外のクローリングでは、URL に対する要求は直接ホストに送信されます。  
プロキシのクローリングでは、要求はプロキシ・サーバーに送信されます。

Web クローラーの作成や、クローリング・スペースの編集の際は、Web クローラーがプ  
ロキシ・サーバー・ドメインでページのクローリングに使用するプロキシ・サーバ  
ーの情報を指定してください。以下の情報を入手してから、プロキシ・サーバ  
ーをクローリング・スペースに追加してください。

### プロキシ・サーバー・ドメイン

プロキシ・サーバーによってサービスされているドメイン。アスタリスク  
(\*) をワイルドカード文字として使用できます。例えば、\* は、このプロキ  
シー・サーバーによってサービスされているすべてのドメインと一致し、  
\*.resource.com は、末尾に resource.com があるすべてのドメインと一致  
します。

**制約事項:** プロキシ・サーバーのクローリングに IP アドレス規則を指定する  
ことはできません。プロキシ・サーバーの IP アドレスは、ク  
ローラーが別のホスト用に持っている唯一の IP アドレスである  
からです。IP アドレス規則が、クローラーを IP アドレスのサブ  
ネットに制約するために使用される場合、制約によってほとん  
どすべての URL が戻りコード 760 に分類されます (この戻りコ  
ードは、URL が Web スペースによって禁止されていることを示  
します)。



### プロキシ・サーバーのホスト名または IP アドレス

DNS ホスト名あるいはプロキシ・サーバーのドットの付いた IP アドレス。

### プロキシ・サーバーのポート番号

プロキシ・サーバーが HTTP プロキシ要求を listen する TCP/IP ポート番号。

プロキシ・サーバーを追加した後は、それを選択して、クローラーに処理させたい順序で配置する必要があります。リスト内では、特定のドメイン・ネームを最初に置き、汎用的なドメイン・ネームが続くように配置してください。Web クローラーは、候補 URL を評価するときに、候補 URL と一致して、かつ、リストの最初のドメインに指定されたプロキシ・サーバー・データを使用します。(どのプロキシ規則にも一致しない URL は、クローラーに直接アクセス可能であると想定されます。)

## Cookie の管理

通常、Cookie の管理は自動的に行われ、エンタープライズ・サーチ管理者が特別な処置をする必要はありません。必要な場合は、Web クロール・セッションの Cookie を手動で指定することができます。

Cookie は、Web サーバーが HTTP 応答ヘッダーの一部としてユーザー・エージェントに戻す内部が見えないトークンです。これは、発行した Web サーバーにのみ意味があるもので、HTTP 要求の間の状態維持に使用されます。例えば、クライアントの認証の間、Web サーバーは Cookie を戻し、認証されたユーザーがすでにログインされていることをサーバーが判断できるようにします。Cookie があることでユーザーは再度ログインを求められることなく、その Web サーバー上の他のページの要求を出すことができます。

Web クローラーは、Web サーバーから受け取った Cookie を保持し、そのクローラー・インスタンスが継続している間、それを使います。Cookie は、cookies.ini ファイルに保管されます。これは、クローラー・セッションが終了するたびに再度書き込まれます。Web クローラーが停止すると、期限切れになっていない Cookie はすべて保管されて次のセッションの開始時に再ロードされます。

手動で Cookie を指定する場合は、まず別ファイルに保管して、必要なときに cookies.ini ファイルの Cookie とマージしてください。クローラーは、期限切れになっていない Cookie を廃棄しませんが、問題が発生して Cookie コレクション全体の書き込みができない場合には、手動で指定した Cookie は失わずにすみませぬ。クロール・セッションが開始するまでにクローラーが自動的に保守している Cookie とユーザーの Cookie をマージする必要があります。

### Cookie フォーマット

エンタープライズ・サーチの cookies.ini ファイルとマージする予定の Cookie は特定のフォーマットである必要があります。

- 各 Cookie は単一行に入っていないなければなりません。空白行とコメントは許可されていますが、cookies.ini ファイルに保存することはできません。
- 各 Cookie は次のフォーマットでなければなりません。

```
CookieN(cookie_length,URL_length)cookie_text,validation_URL
```

## Cookie

Cookie 項目の始まりを示す必要なキーワード。

Cookie キーワードには、空白を含むことはできず、また、最後に 0 か、1 か、2 のいずれかの 1 桁の数字が付いていなければなりません。この数字は、次の Cookie タイプを表します: バージョン-0 (Netscape)、バージョン-1 (RFC2109)、あるいは バージョン-2 (RFC2965)。ポート・リストは、RFC2965 Cookie ではサポートされていません。

### *cookie\_length*

関連した Cookie テキストの文字数。

### *URL\_length*

関連した検証 URL の文字数。

### *cookie\_text*

発信元の Web サーバーに送信される予定の Cookie のコンテンツ。このストリング (HTTP 応答ヘッダーにある Set-Cookie ディレクティブの右側に相当する部分) には、Cookie の名前と値の組、およびその Cookie と一緒に送信するその他のコンテンツ (パス、セキュリティー設定など) が入ります。このストリングの後ろには、区切り記号のコンマ (,) が続きます。

### *validation\_URL*

この Cookie がディスカバーされた URL。この URL は、Cookie の送信先の決定に使用されます (例えば、ドメイン名とパス名を指定することで決定します)。検証 URL は、発信元の Web サーバーの Cookie に関するセキュリティーとプライバシーの制約事項を満たしている必要があります。

次の例では、読みやすくするために 2 行に分けて示します。Cookie を指定するときは、一行に入れてください:

```
Cookie0(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSINFLSINL;path=/,  
https://www.ibm.com:443/help/solutions/
```

## Web クローラーの Cookie の構成

Web クロール・セッションの Cookie を手動で指定し、Web クローラーが維持する Cookie とマージできます。

### 始める前に

Web クローラーが使用する Cookie を手動で構成するには、エンタープライズ・サーチ管理者でなければなりません。

### 手順

Web クローラーの Cookie を手動で構成するには、次のようにします。

1. エンタープライズ・サーチ管理コンソールから、Cookie を指定したいコレクションをモニターして、Web クローラーを停止します。
2. クローラー・サーバーで、エンタープライズ・サーチ管理者としてログインします。(このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。)

3. 構成したいクローラーのデータ・ディレクトリーに移動します。ここで、`crawler_session_ID` は、エンタープライズ・サーチ・システムによりクローラー・セッションに割り当てられたものです。以下に例を示します。  
`ES_NODE_ROOT/data/col_56092.WEB_88534`
4. `cookies.ini` ファイルを編集し、手動で指定した Cookie 項目を、すでにリストされている項目に追加して、そのファイルを保管して終了します。追加する Cookie が、すでに存在しているものをオーバーライドしないようにします。
5. エンタープライズ・サーチ管理コンソールから、停止した Web クローラーを再始動します。

## グローバル Web クロール・スペースの構成

Web クローラーのためにグローバル・クロール・スペースを構成できます。これにより索引からの URL の除去をより制御できるようになります。

Web クローラーはそれぞれ、クロール対象の URL あるいは、クロール対象外の URL を定義するクロール・スペースで構成されます。クロール・スペースにあるディスカバー済み URL は、後でクロールするときのために (データベースに) 保存されます。クロール・スペースにない URL は、廃棄されます。クローラーが空のデータベースで始動した場合、クローラーが稼働中は、そのクロール・スペースの定義とデータベースは整合しています。

時々、クローラーは停止されて、そのクロール・スペースが、(例えば、ページのクロールを禁止する新規規則によって) 削減されます。クローラーが再始動されるとそのクロール・スペース定義とデータベースは、整合しなくなります。データベースには URL (クロールされるものもあれば、クロールされないものもある) が含まれ、これらの URL は小さくなった新しいクロール・スペースにはありません。

コレクションに Web クローラーが 1 つしかない場合、その Web クローラーがこれらの URL 用の HTTP 戻りコードを 760 (除外される予定であると指定する) に変更し、新規に除外されたページを索引から除去することで、整合性が戻ります。

複数の Web クローラー間でクロール・スペースを分割する場合 (例えば、一部のページは残りのものより頻繁にクロールされるようにするなど)、各 Web クローラーが独立したデータベース表 (初期は空) を保守し、それらはそれぞれ Web クロール・スペースの別の部分をクロールします。元のクローラーのクロール・スペースは削減されて、他のクローラーがクロールする予定の部分が除去された後に残った分のみになります。元のクローラーが、移動済みのページを索引から除去することで整合性を復元しようとするとう問題が発生します。移動済みのページは、他のクローラーによってクロールされているので、そのページは索引に残っている必要があります。

高水準のグローバル・クロール・スペースを構成することで、元のクローラーによってクロールされないが、索引からも削除されない URL を識別することができます。クローラーのクロール・スペースにすでにない URL は、ディスカバリー・プロセスによって除外とマークされたままになり、再度クロールされるときに索引から除去されます。

グローバル・クロール・スペースは、`global.rules` という名前の構成ファイルに定義されます。このファイルは、クローラー構成ディレクトリーに存在しなければなり

ません。(global.rules ファイルがあることで、グローバル・クロール・スペース機能が使用可能になります。) このファイルがある場合は、クローラーの初期化時に読み取られます。このファイルが存在しないと、クローラーは単一レベルのクロール・スペースで作動することになり、クロール・スペースの定義とデータベースの間の整合性を維持するために、必要に応じて索引から文書を除去します。

グローバル・クロール・スペースが存在する場合、クローラーは従来どおり URL が存在するかしないかを判定しますが、索引からの URL の除去を要求するのは URL がどの Web クロール・スペースにも存在しない場合のみです。

global.rules ファイルは、ローカルの crawl.rules ファイルと構文は同じですが、ドメイン名の規則のみを含むことができることが異なります。この制約事項により、クロール・スペースは IP アドレスや HTTP 接頭部パターンではなく DNS ホスト名に基づいてのみ、複数クローラー間でパーティション化されるようになります。URL 接頭部やローカル・クロール・スペースの IP アドレスの規則で (crawl.rules に定義されているように) 除外される URL はグローバル・クロール・スペースによって影響を受けません。そのような URL はやはり除外されます。

グローバル・クロール・スペースは、索引から URL を除去できなくするためにのみ使用されます。この URL はローカルのドメイン規則によってクローラーのクロール・スペースから除去されます。以下の規則が以下の順序で適用されます。

1. ローカルの接頭部規則やアドレス規則によって、クローラーのデータベースから URL が除去される場合、その URL には戻りコード 760 が割り当てられ、索引から除去されます。URL は再度クロールされません。
2. ローカルのドメイン規則によって、クローラーのデータベースから URL が除去されて、さらにグローバル・クロール・スペースがない場合に、URL には戻りコード 760 が割り当てられ、索引から除去されます。URL は再度クロールされません。
3. ローカルのドメイン規則によって、クローラーのデータベースから URL が除去されても、グローバル・クロール・スペースの規則によって明示的に許可されている場合、URL には戻りコード 761 が割り当てられます。クローラーは、その URL を再度クロールしませんが、索引から URL が除去されることはありません。(他のクローラーのローカル・クロール・スペースにあるものと想定されます。)
4. ローカルのドメイン規則によってクローラーのデータベースから URL が除去されており、グローバル・クロール・スペースの規則によって明示的には許可されていない場合、その URL には戻りコード 760 が割り当てられ、索引から除去されます。

グローバル・クロール・スペースは、ローカル・クロール・スペースによりすでに除外されている URL の除去を防ぐためにのみ調べられるので、グローバル・クロール・スペースからのデフォルトの結果は、候補 URL に適用する規則がない場合に、それに対するクロールを禁止することです。

global.rules ファイルは、グローバル・クロール・スペースを共用するあらゆるクローラーの master\_config ディレクトリに存在しなければなりません。global.rules ファイルとローカルの crawl.rules ファイルのすべてのコピーを注意深く編集して、それらが相互に整合性を保っていることを確認してください。

## リンクをたどらない、索引付けしないディレクティブ

ページ上のリンクをたどるか、ページに索引付けをするかを制御する Web クローラー用のディレクティブを指定することで検索の質を上げることができます。

一部の Web ページには、リンクをたどらない、あるいは索引付けをしないディレクティブがあり、これにより Web クローラーなどのロボットがそのページで検出したリンクをたどらないように指示したり、あるいは、そのページのコンテンツを索引に含めないように指示したり、あるいは、そのどちらも行わないように指示します。

これらの設定を制御することで、クロールを向上させることができます。例えば、ディレクトリー・ページの中には、多数のリンクを含んでいるものの、他に有効なコンテンツが無いものがあります。こうしたページはクロールしてリンクをたどる必要はありますが、ディレクトリー・ページそのものを索引付けする利点はありません。

また、クローラーが階層の下方に行かないようにしたい場合がありますが、要求したリーフ・ページにリンクが組み込まれていて、しかもリンクをたどらないディレクティブが含まれていない場合があります。これらのページには自動生成されているものがあり、必要なディレクティブを挿入できる所有者がいません。

そうしたページのクロールに規則を指定するために、`followindex.rules` という名前の構成ファイルを作成、編集します。このファイルに規則を指定するときは、以下のガイドラインに従ってください。

- 設定する規則では、URL 接頭部を指定する必要があります (IP アドレスあるいは DNS ホスト名で Web サイトを識別することはできません)。
- URL 接頭部には、類似した URL を持つ複数のサイトを許可したり禁止するためにワイルドカード文字としてアスタリスク (\*) を含めることができます。
- 順序は重要です (クローラーは、候補 URL と一致する最初の規則を適用します)。
- リンクをたどることや索引付けを明示的に許可および禁止する規則は、ターゲット文書内の設定を含む他の設定に優先します。

## Web ページでのリンクをたどらないディレクティブおよび索引付けをしないディレクティブのオーバーライド

構成ファイルに規則を指定して、Web クローラーがリンクをたどらないディレクティブまたは索引付けをしないディレクティブが含まれているページへのリンクまたは索引ページをたどるかどうかを制御できます。

### 始める前に

Web クローラーについてリンクをたどらないディレクティブおよび索引付けをしないディレクティブを指定するには、エンタープライズ・サーチ管理者でなければなりません。指定されたディレクティブは、クロールされるページに存在するディレクティブをオーバーライドします。

### 手順



リンクをたどらないディレクティブおよび索引付けをしないディレクティブをオーバーライドするには、次のようにします。

1. エンタープライズ・サーチ管理コンソールから、規則を構成したいコレクションをモニターして、Web クローラーを停止します。
2. クローラー・サーバーで、エンタープライズ・サーチ管理者としてログインします。(このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。)
3. 構成したいクローラーの構成ディレクトリーに移動します。ここで、`crawler_session_ID` は、エンタープライズ・サーチ・システムによりクローラー・セッションに割り当てられたものです。以下に例を示します。

```
ES_NODE_ROOT/master_config/col_56092.WEB_88534
```

4. `followindex.rules` という名前のファイルを作成または編集します。
5. 以下の形式でクローラーの規則を入力します。ここで、`URLprefix` は、リンクのたどりまたは索引付けを許可または禁止したい Web サイトの先頭文字です。

```
forbid follow URLprefix
allow follow URLprefix
forbid index URLprefix
allow index URLprefix
```

6. ファイルを保管して、終了します。
7. エンタープライズ・サーチ管理コンソールから、停止した Web クローラーを再始動します。

---

## WebSphere Portal クローラー

IBM WebSphere Portal サイトのページをエンタープライズ・サーチ・コレクションに入れるには、WebSphere Portal クローラーを構成する必要があります。

### WebSphere Portal サーバーの構成

WebSphere Portal クローラーを作成する前に、エンタープライズ・アプリケーション `ESPACServer.ear` を WebSphere Portal にデプロイする必要があります。このエンタープライズ・アプリケーションは、WebSphere Information Integrator OmniFind Edition のインストール時に検索サーバーにインストールされます。このエンタープライズ・アプリケーションをデプロイするには、以下のタスクを完了させます。

- 99 ページの『WebSphere Portal クローラー用エンタープライズ・アプリケーションのデプロイ』。

### 構成の概要

WebSphere Portal クローラーを使用して、単一の WebSphere Portal サイトをクロールすることができます。クローラーを構成する際に、クロールするポータル・サイトの URL を指定します。これにより、クローラーは、指定されたポータルで使用可能なポートレットをダウンロードします。

WebSphere Portal クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。



クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるサイトの全ページのクロールを制御します。
- クロールするポータル・サイトの URL とクローラーがサイトに接続するための情報を指定します。このようなタイプの URL は長く、エンコードされた非 ASCII 文字が含まれているので、URL を WebSphere Portal サーバーからコピーし、エンタープライズ・サーチ管理コンソールに貼り付けると便利です。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

また、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・ソースで維持されている現行のアクセス制御リストと比較します。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

## WebSphere Portal クローラー用エンタープライズ・アプリケーションのデプロイ

WebSphere Portal クローラーを作成する前に、エンタープライズ・アプリケーション ESPACServer.ear を WebSphere Portal にデプロイする必要があります。

### このタスクについて

ESPACServer.ear ファイルは、WebSphere II OmniFind Edition のインストール時に、検索サーバーの ES\_INSTALL\_ROOT/bin ディレクトリーにインストールされます。デフォルトのインストール・パスは、次のとおりです。

#### UNIX システム:

`/opt/IBM/es/bin/ESPACServer.ear`

#### Windows システム:

`C:\Program Files\IBM\es\bin\ESPACServer.ear`

### 手順

WebSphere Portal クローラーで WebSphere Portal サイトをクロールできるようにするエンタープライズ・アプリケーションをデプロイするには、次のようにします。

1. WebSphere\_Portal サーバー・インスタンスを停止します。
2. WebSphere Application Server server1 サーバー・インスタンスがまだ開始されていない場合は、開始します。
3. WebSphere Portal サーバーで、WebSphere Application Server 管理コンソールを開始します。ログインを指示するプロンプトが出されたら、ログインします。

管理コンソールは、以下の方法でオープンできます。

- Windows の「スタート」メニューを使用して、該当するプログラムを選択します。
  - WebSphere Application Server バージョン 5 の場合、Web ブラウザーをオープンし、`http://hostname:port/admin` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9090 です。
  - WebSphere Application Server バージョン 6 の場合、Web ブラウザーをオープンし、`http://hostname:port/ibm/console` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9060 です。
4. 「アプリケーション」をクリックして、「新規アプリケーションのインストール」をクリックします。
  5. 「ブラウズ」をクリックし、ご使用のシステムの ESPACServer.ear ファイルを選択します。
  6. 「次へ」を 2 回クリックします。ポリシー・ファイルに関する警告が出された場合は、「継続」をクリックします。
  7. 「アプリケーション・サーバーへのモジュールのマップ」ページが表示されるまで「次へ」をクリックし、以下を選択して、「適用」をクリックします。
    - a. 「クラスターおよびサーバー」テキスト・ボックスで、**WebSphere:cell=cell\_name、 node=node\_name、 server=WebSphere Portal** を選択します。
    - b. **ESPACServer.ear** モジュールの横のチェック・ボックスを選択します。
  8. 「サマリー」ページが表示されるまで「次へ」をクリックし、「完了」をクリックします。
  9. 「マスター構成に保管」リンクをクリックして、「保管」ボタンをクリックし、WebSphere Application Server 構成に変更を保管します。
  10. WebSphere Portal サーバーを再始動します。

#### 関連概念

233 ページの『エンタープライズ・サーチの WebSphere Portal との統合』  
WebSphere Portal および WebSphere Portal 検索センターにエンタープライズ・サーチ・ポートレットをデプロイすることにより、IBM WebSphere Portal の検索機能を拡張することができます。

## クロールする URL を WebSphere Portal からコピーする

誤った URL の入力をできる限り少なくするために、WebSphere Portal クローラーの構成時に、クロールしたい WebSphere Portal サイトの URL を適切なフィールドにコピーして貼り付けることができます。

#### このタスクについて

WebSphere Portal クローラーの作成時に、クロールしたい WebSphere Portal サーバー上のポータル URL を指定します。URL は長く、通常、エンコードされた非

ASCII 文字が含まれているので、この手順を使用して、URL を WebSphere Portal サーバーからコピーし、エンタープライズ・サーチ管理コンソールに貼り付けると便利です。

### 手順

WebSphere Portal クローラーがクロールする URL を指定するには、次のようにします。

1. エンタープライズ・サーチ管理コンソールで「クロールする WebSphere Portal Server」ページを開き、WebSphere Portal Server が始動していることを確認したら、管理者として WebSphere Portal にログインします。
2. 右上にある「管理」を選択します。
3. 左側のナビゲーション・エリアで「ポータル設定」を選択して、「検索管理」を選択します。
4. 「検索コレクションの管理」ページで、「検索コレクション」エリアにある「PortalCollection」を選択します。(他のコレクションが選択可能な場合、別のコレクションを選択できます。)
5. 「コレクション」エリアの「コンテンツ・ソース」で、「コンテンツ・ソースの追加」をクリックします。
6. 「クロール・ソース・タイプ」で、「ポータル・サイト」を選択します。「この URL でリンクされる文書を収集」フィールドにサイトの URL が表示されず。
7. URL をクリップボードにコピーします。(例えば、URL を強調表示して、Ctrl キーを押しながら Insert (挿入) キーを押します。)
8. エンタープライズ・サーチ管理コンソールに戻って、「WebSphere Portal サイトの URL」フィールドにコピーした URL を貼り付けます。

---

## Windows ファイル・システム・クローラー

Microsoft Windows ファイル・システムに格納された文書をエンタープライズ・サーチ・コレクションに入れる場合は、Windows ファイル・システム・クローラーを構成する必要があります。

Windows ファイル・システム・クローラーを使用して、任意の数の Windows ファイル・システムをクロールすることができます。クローラーの構成時に、クロールするローカルおよびリモートのディレクトリー、サブディレクトリーを選択します。

UNIX コンピューターにクローラー・サーバーをインストールした場合、Windows ファイル・システム・ソースのクロールにそのサーバーを使用することはできません (Windows ファイル・システム・クローラーは、選択可能なクローラー・タイプのリストに表示されません)。

Windows ファイル・システム・クローラーを作成または変更する場合は、エンタープライズ・サーチ管理コンソールにログインします。この場合、操作者はエンタープライズ・サーチ管理者役割の担当者か、またはクローラーを所有するコレクションのコレクション管理者でなければなりません。

クローラーを作成する場合、ウィザードに従えば以下のタスクを実行できます。

- クローラーの動作およびシステム・リソースの使用を制御するプロパティを指定します。クローラーのプロパティは、クローラーによるクロール・スペース内のすべてのサブディレクトリーに対するクロールを制御します。
- ファイル・システムをクロールするスケジュールを設定します。
- クロールするサブディレクトリーを選択します。

クローラーでクロールするサブディレクトリーのレベル数も指定できます。リモート・ファイル・システムをクロールするには、クローラーがデータにアクセスできるようにするためのユーザー ID とパスワードも併せて指定します。

- サブディレクトリー内の文書を検索可能にするときのオプションを指定します。例えば、特定タイプの文書をクロール・スペースから除外したり、クローラーが特定サブディレクトリーのファイルにアクセスできるようにユーザー ID とパスワードを指定したりすることができます。
- 文書レベルのセキュリティー・オプションを構成します。コレクションの作成時にセキュリティーが使用可能にされている場合、クローラーは、セキュリティー・データを索引内の文書に関連付けます。このデータによって、検索アプリケーションは、保管されたアクセス制御リストまたはセキュリティー・トークンに基づいたアクセス制御を実行できるようになります。

また、ユーザーが照会をサブミットするときにユーザーの信用証明情報を検証するオプションを選択することもできます。この場合、システムは、ユーザーの信用証明情報を索引付きセキュリティー・データではなく、元のデータ・ソースで維持されている現行のアクセス制御リストと比較します。

文書レベルのセキュリティーを施行するには、ユーザーおよびドメイン・アカウント情報がクローラー・サーバーに正しく構成されていることを確認する必要があります。

クローラーの作成中に「ヘルプ」をクリックすると、ウィザードのフィールドに関する説明や、データのクロールに必要な情報をクローラーに指定する方法が表示されます。

#### 関連概念

225 ページの『Windows ファイル・システム文書に対する文書レベルのセキュリティーの施行』

Windows ファイル・システム・クローラーでクロールされた文書をユーザーが検索するときに現行の信用証明情報を検証するようするには、クローラー・サーバーおよびMicrosoft Windows サーバーの両方でドメイン・アカウント情報を構成する必要があります。

---

## データ・リスナー・アプリケーションのサポートの構成

外部クローラーを作成するデータ・リスナー API を使用することによって、エンタープライズ・サーチを拡張することができます。カスタム・データ・リスナー・アプリケーションは、コレクションへのデータの追加、コレクションからのデータの除去、または Web クローラーに対する URL へのアクセスおよび再アクセスの指示を行えます。

## 始める前に

データ・リスナー・アプリケーションを構成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて


クライアント・データ・リスナー・アプリケーションは、エンタープライズ・サーチのデフォルトのクローラーではクロールできないデータ・ソース・タイプのクローリングを可能にします。データ・リスナー・アプリケーションを使用するためには、アプリケーションがコレクションのアクセスおよび更新を行えるようにする信用証明情報を構成しておく必要があります。

クライアント・データ・リスナー・アプリケーションは、データ・リスナーに接続するときに、クライアント・アプリケーション ID とパスワード、更新されるコレクションの ID を渡す必要があります。この情報は、管理コンソールでアプリケーションについて構成した情報と同じでなければなりません。



データ・リスナーは、エンタープライズ・サーチ・システムが開始されるときに自動的に開始されます。管理コンソールでアプリケーションを構成した後でポート番号を変更した場合は、データ・リスナーを再始動する必要があります。

### 手順

データ・リスナー・アプリケーションを構成するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「 編集」をクリックして、システム編集ビューに移動します。
3. 「データ・リスナー」ページで、「データ・リスナー・アプリケーションの構成」をクリックします。
4. 「データ・リスナー・アプリケーション」ページで、データ・リスナーがクライアント・アプリケーションからの要求の処理用に作成できるスレッドの数、およびデータ・リスナーが要求を listen するポート番号を指定します。
5. クライアント・アプリケーションについての情報を追加するため、「データ・リスナー・アプリケーションの追加」をクリックします。
6. 「データ・リスナー・アプリケーションの追加」ページで、クライアント・データ・リスナー・アプリケーションがエンタープライズ・サーチ・コレクションにアクセスできるようにする認証情報を指定します。データ・リスナーのクライアント ID は、エンタープライズ・サーチ・システム内で固有のものでなければなりません。
7. データ・リスナー・アプリケーションが更新できるコレクションを選択します。
  - すべてのコレクションをアプリケーションが更新できるようにするには、「すべてのコレクション」をクリックします。
  - 指定するコレクションのみをアプリケーションが更新できるようにするには、「特定のコレクション」をクリックします。

このオプションを選択すると、コレクション名のリストが表示されます。アプリケーションが更新できるコレクションごとに、その「選択」チェック・ボックスを選択します。

8. 「OK」をクリックします。
9. データ・リスナーのポート番号を変更した場合は、データ・リスナーを再始動します。
  - a. 「 モニター」をクリックして、システム・モニター・ビューに移動します。
  - b. 「データ・リスナー」ページで、「 再始動」をクリックします。

#### 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『データ・リスナー』

#### 関連タスク

277 ページの『データ・リスナーのモニター』

データ・リスナーをモニターしてその状況を確認し、クライアント・データ・リスナー・アプリケーションのアクティビティーに関する詳細を表示します。

---

## カスタム・クローラー・プラグイン

クローラーのプロパティーを構成するときに、文書レベルのアクセス制御の施行、およびメタデータとエンタープライズ・サーチ索引内の文書との関連付けを行なうために Java クラスを使用するよう指定できます。

プラグインには、クローラーがクロールする各文書ごとに呼び出される Java クラスが含まれています。その Java クラスには、エンタープライズ・サーチ索引からの文書 ID (URI)、セキュリティー・トークン、およびメタデータが渡されます。クラスは、セキュリティー・トークンまたはメタデータの新規セットまたは変更されたセットを戻すか、あるいは、文書がクローラーに無視されることを示すことができます。

クロール・スペース内のすべての文書が 1 回クロールされた後は、プラグインは新規文書または修正された文書についてのみ呼び出されます。エンタープライズ・サーチ索引内にあるが、元のデータ・ソースでは更新されていない文書のセキュリティー・トークンおよびメタデータを変更するには、クロール・スペース内のすべての文書のフル・クロールを開始し、索引を再編成します。

既存のクローラーにプラグインを関連付けることはできません。プラグインのクラス名とクラス・パスは、新規クローラーのプロパティーを構成するときに指定する必要があります。

### プラグインを使用してセキュリティーを施行する

文書レベルのセキュリティーは、クローラーがクロールする各文書に 1 つ以上のセキュリティー・トークン (コンマで区切られた文字列) を関連付けることによって施行されます。通常、セキュリティー・トークンとしてグループ ID が使用されます。

デフォルトで、各文書には、文書を全員に使用可能にするパブリック・トークンが割り当てられています。パブリック・トークンは、管理者が提供する値、またはクロール対象文書中のフィールドから抽出された値で、置き換えることができます。



プラグインを使用すると、独自のビジネス・ルールを適用して、クロールされる文書のセキュリティ・トークンの値を決定できます。各文書に関連付けられたセキュリティ・トークンは、索引に保管されます。セキュリティ・トークンを使用して、それらに一致する文書を選出するようフィルター操作し、ユーザーに表示を許可された文書だけが検索結果に戻されるようになります。

## プラグインを使用してメタデータを文書に追加する

文書の最終変更日時などの文書メタデータは、クロールされるすべての文書について作成されます。クローラー・プラグインを使用すると、独自のビジネス・ルールを適用して、索引付けされるメタデータの値を各文書について決定することができます。

メタデータは、名前と値のペアとして作成されます。ユーザーは、フリー・テキスト検索またはメタデータ・フィールド名を指定した照会で、メタデータを検索できます。

## Web クローラー・プラグイン

Web クローラー用のアプリケーション・プログラミング・インターフェースで、文書がどのようにクロールされ、構文解析用に準備されるのかを制御できます。例えば、クローラーが文書を要求するときに使用されるフィールドを HTTP 要求ヘッダーに追加できます。文書がクロールされた後、構文解析およびトークン化される前に、コンテンツ、セキュリティ・トークン、およびメタデータを変更できます。また、文書がパーサーに送られるのを停止することもできます。

### 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『クローラー・プラグイン』

---

## エンタープライズ・サーチ索引における URI フォーマット

エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

コレクションのカテゴリ、有効範囲、クイック・リンクを構成する際、URI および URI パターンを指定できます。また、索引から文書を除去する必要がある場合、または、特定の URI に関する詳細状況情報を表示する必要があるときにも、URI を指定します。

文書の URI または URI パターンを判別するには、コレクションを検索してください。検索結果内の URI をクリックすれば、興味のある文書を取得できます。検索結果から URI をコピーして、その URI をエンタープライズ・サーチ管理コンソールで使用できます。例えば、URI パターンを指定して、その URI パターンに一致する文書を、エンタープライズ・サーチ・カテゴリに自動的に関連付けることができます。

## Content Edition クローラー

サーバー・アクセス・モードの Content Edition クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
vbr://Server_Name/Repository_System_ID/Repository_Persistent_ID  
/Item_ID/Version_ID  
/Item_Type/[?Page=Page_Number&] JNDI_properties
```

直接アクセス・モードの Content Edition クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
vbr:///Repository_System_ID/Repository_Persistent_ID  
/Item_ID/Version_ID  
/Item_Type/[?Page=Page_Number]
```

#### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

##### *Server\_Name*

WebSphere Information Integrator Content Edition サーバーの名前。

##### *Repository\_System\_ID*

リポジトリのシステム ID。

##### *Repository\_Persistent\_ID*

リポジトリの永続 ID。

##### *Item\_ID*

項目の ID。

##### *Version\_ID*

バージョンの ID。バージョン ID がブランクの場合、この値は文書の最新バージョンを示します。

##### *Item\_Type*

項目のタイプ (CONTENT または FOLDER)。

##### *Page\_Number*

ページ番号。

##### *JNDI\_properties*

J2EE アプリケーション・クライアントの JNDI プロパティ。プロパティには、次の 2 つのタイプがあります。

##### **java.naming.factory.initial**

EJB ハンドルを作成するのに使用されるアプリケーション・サーバーのクラスの名称。

##### **java.naming.provider.url**

EJB ハンドルを要求するのに使用されるアプリケーション・サーバーのネーミング・サービスへの URL。

#### 例 Documentum:

```
vbr://vbrsrv.ibm.com/Documentum/c06b/094e827780000302//CONTENT/?  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

#### FileNet PanagonCS:

```
vbr://vbrsrv.ibm.com/PanagonCS/4a4c/003671066//CONTENT/?Page=1&  
java.naming.provider.url=iiop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

## DB2 クローラー

DB2 クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
db2://Database_Name/Table_Name  
/Unique_Identifier_Column_Name1/Unique_Identifier_Value1  
[/Unique_Identifier_Column_Name2/Unique_Identifier_Value2/...  
/Unique_Identifier_Column_NameN/Unique_Identifier_ValueN]
```

### パラメーター:

すべてのフィールドに URL エンコード方式が適用されます。

*Database\_Name*

データベースの内部名またはデータベースの別名。

*Table\_Name*

スキーマの名前を含む、ターゲット表の名前。

*Unique\_Identifier\_Column\_Name1*

表の先頭の固有 ID 列の名前。

*Unique\_Identifier\_Value1*

先頭の固有 ID 列の値。

*Unique\_Identifier\_Column\_NameN*

表の *n* 番目の固有 ID 列の名前。

*Unique\_Identifier\_ValueN*

*n* 番目の固有 ID 列の値。

**例** ローカル、カタログ・データベース:

```
db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20
```

リモート、アンカタログ・データベース

```
db2://myserver.mycompany.com:50001/REMOTEDB/SCHEMA2.TABLE2/NAME/DAVID
```

## DB2 Content Manager クローラー

DB2 Content Manager クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
cm://Server_Name/Item_Type_Name/PID
```

### パラメーター

*Server\_Name*

IBM DB2 Content Manager ライブラリー・サーバーの名前。

*Item\_Type\_Name*

ターゲット項目タイプの名前。

*PID* DB2 Content Manager 永続 ID

**例**

```
cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icmn1sdb12+ITEMTYPE159+26+A1001001A  
03F27B94411D1831718+A03F27B+94411D183171+14+1018
```

## Domino Document Manager クローラー

Domino Document Manager クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
dominodoc://Server_Name:Port_Number/Database_Replica_ID/Database_Path_and_Name  
/View_Universal_ID/Document_Universal_ID  
/?AttNo=Attachment_Number&AttName=Attachment_File_Name
```

### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

*Server\_Name*

Domino Document Manager サーバーの名前。

*Port\_Number*

オプション: Domino Document Manager サーバー用のポート番号。

*Database\_Replica\_ID*

データベース・レプリカの ID。

*Database\_Path\_and\_Name*

ターゲット Domino Document Manager サーバー上の文書 NSF データベースのパスとファイル名。

*View\_Universal\_ID*

Domino Document Manager 文書をクロールするのに使用される View Universal ID。

*Document\_Universal\_ID*

クロール対象文書で定義される Document Universal ID。

*Attachment\_Number*

オプション: ゼロから始まる、各添付ファイルの連続番号。

*Attachment\_File\_Name*

オプション: 添付ファイルのオリジナル名。

**例** Domino Document Manager 文書:

```
dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf  
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

文書添付ファイル:

```
dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf  
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226  
?AttNo=0&AttName=AttachedFile.doc
```

## Exchange Server クローラー

Exchange Server クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
exchange://OWA_path[?useSSL=true]
```

### パラメーター

*OWA\_Path*

プロトコルを含まない、Outlook Web Access (OWA) パス。

**useSSL=true**

オリジナルの OWA パスのプロトコルが HTTPS である場合に追加される。

**例** 文書本体:

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML

## 文書添付ファイル:

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/  
AttachedFile.doc

## SSL で使用可能:

exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML  
?useSSL=true

**Lotus Notes クローラー**

Lotus Notes クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
domino://Server_Name[:Port_Number]/Database_Replica_ID/Database_Path_and_Name  
/[View_Universal_ID]/Document_Universal_ID  
[?AttNo=Attachment_Number&AttName=Attachment_File_Name]
```

**パラメーター**

すべてのフィールドに URL エンコード方式が適用されます。

*Server\_Name*

Lotus Notes サーバーの名前。

*Port\_Number*

Lotus Notes サーバーのポート番号。ポート番号はオプションです。

*Database\_Replica\_ID*

データベース・レプリカの ID。

*Database\_Path\_and\_Name*

ターゲット Lotus Notes サーバー上の NSF データベースのパスとファイル名。

*View\_Universal\_ID*

ターゲット・データベース上で定義された View Universal ID。この ID は、文書がビューまたはフォルダーから選択されたときのみ指定されます。ビューまたはフォルダーをクロール対象として指定しない場合 (例えば、データベース内の全文書をクロールするよう指定する場合)、View Universal ID は指定されません。

*Document\_Universal\_ID*

クローラーによってクロールされる文書内で定義された Document Universal ID。

*Attachment\_Number*

ゼロから始まる、各添付ファイルの連続番号。添付ファイル番号はオプションです。

### *Attachment\_File\_Name*

添付ファイルのオリジナル名。添付ファイル名はオプションです。

**例** ビューまたはフォルダーでクロールするために選択された文書:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

ビューまたはフォルダーでクロールするために選択されていない文書:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226
```

文書添付ファイル:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf//  
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

## QuickPlace クローラー

QuickPlace クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
quickplace://Server_Name:Port_Number/Database_Replica_ID/Database_Path_and_Name  
/View_Universal_ID/Document_Universal_ID  
/?AttNo=Attachment_Number&AttName=Attachment_File_Name
```

### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

#### *Server\_Name*

Lotus QuickPlace サーバーの名前。

#### *Port\_Number*

オプション: QuickPlace サーバー用のポート番号。

#### *Database\_Replica\_ID*

データベース・レプリカの ID。

#### *Database\_Path\_and\_Name*

ターゲット QuickPlace サーバー上の文書 NSF データベースのパスとファイル名。

#### *View\_Universal\_ID*

QuickPlace 文書をクロールするのに使用される View Universal ID。

#### *Document\_Universal\_ID*

クロール対象文書で定義される Document Universal ID。

#### *Attachment\_Number*

オプション: ゼロから始まる、各添付ファイルの連続番号。

#### *Attachment\_File\_Name*

オプション: 添付ファイルのオリジナル名。

**例** 文書:

```
quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498
```



添付ページ:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498  
?AttNo=0&AttName==QPCons3.ppt
```

## UNIX ファイル・システム クローラー

UNIX ファイル・システム クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

`file:///Directory_Name/File_Name`

### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

*Directory\_Name*

ディレクトリーの絶対パス名。

*File\_Name*

ファイルの名前。

### 例

`file:///home/user/test.doc`

## WebSphere Portal クローラー

WebSphere Portal クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

`wps://WPS_Access_Path?portletDefID=Portlet_Def_ID&pageID=Page_ID&useSSL=SSL`

### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

*WPS\_Access\_Path*

WebSphere Portal サーバー・パス (プロトコルなし)。

*Portlet\_Def\_ID*

WebSphere Portal サーバー用ポータル定義 ID。

*Page\_ID*

WebSphere Portal サーバー用ページ ID。

*SSL* `useSSL=true` は、パスのプロトコルが `HTTPS` であるときに追加されます。そうでない場合は、`useSSL=false` が追加されます。

### 例

文書本体:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+  
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8gHRj/delta  
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!  
?portletDefID=3_0_3S&pageID=6_0_6J&useSSL=false
```

### 例

SSL で使用可能:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+  
LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8gHRj/delta  
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!  
?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

## Windows ファイル・システム クローラー

Windows ファイル・システム クローラーによってクロールされる文書の URI フォーマットは、次のとおりです。

```
file:///Directory_Name/File_Name  
file:///Network_Folder_Name/Directory_Name/File_Name
```

### パラメーター

すべてのフィールドに URL エンコード方式が適用されます。

*Directory\_Name*

ディレクトリーの絶対パス名。

*File\_Name*

ファイルの名前。

*Network\_Folder\_Name*

リモート・サーバー上のみにある文書の場合の、Windows ネットワーク上の共有フォルダーの名前。

**例** ローカル・ファイル・システム:

```
file:///d:/directory/test.doc
```

ネットワーク・ファイル・システム:

```
file:///filesvr.ibm.com/directory/file.doc
```

### 関連概念

37 ページの『エンタープライズ・サーチ・クローラーの管理』

コレクションに入れるデータ・タイプに応じて、さまざまなタイプのクローラーを構成できます。1 つのコレクションに任意の数のクローラーを入れることができます。

### 関連タスク

122 ページの『カテゴリの構成』

1 つのコレクションのカテゴリはいくつも作成することができ、各カテゴリにはいくつもの規則を含めることができます。規則によって、カテゴリと自動的に関連付けられる文書が決まります。

160 ページの『有効範囲の構成』

エンタープライズ・サーチ・コレクションの有効範囲を構成するときに、ユーザーが検索を許可される索引内の文書の範囲に合うように URI または URI パターンを指定します。

163 ページの『索引から URI を除去』

コレクション内の文書をユーザーが検索できないようにするため、それらの文書の URI を索引から除去することができます。

174 ページの『クイック・リンクの構成』

エンタープライズ・サーチ・コレクションのクイック・リンクを作成するため、文書の URI を、検索結果に含ませたものをトリガーするキーワードと関連付けます。

260 ページの『URI についての詳細表示』

URI に関する詳細情報を見ることができます。この URI で表される文書がどのようにクロール、索引付け、検索されるのかについての現在情報および履歴情報を知ることができます。

279 ページの『除去された文書に関するレポートの表示』  
エンタープライズ・サーチ・システムから除去された文書に関する詳細情報を  
表示することができます。この情報は、コレクションに対する文書トラッキング  
を 使用可能にした場合にのみ有効です。



---

## エンタープライズ・サーチ・パーサーの管理

文書の検索性を高めるために、文書およびメタデータが、エンタープライズ・サーチ索引に追加される前に、どのように構文解析、分析、および分類されるのかに関するオプションを指定できます。

文書コンテンツの構文解析および情報の検索性の最適化のために指定できるオプションには、以下のものがあります。

### 中国語、日本語、韓国語の文書に対する構文解析オプションの構成

中国語、日本語、韓国語で書かれた文書を構文解析するために N-gram セグメンテーション使用のオプションを指定することができます。また、中国語と日本語の文書内の空白文字から改行文字を除去することもできます。

### ネイティブ XML 検索の使用可能化

ご使用のコレクションに XML 文書が含まれている場合、XPath および XML フラグメントなどの、ネイティブ XML 照会構文を使用した検索を可能にできます。ネイティブ XML 検索により、ユーザーは、さまざまな XML エlement間の関係に基づいた照会を指定できるようになります。

### カテゴリーの構成

類似 URI パターンを共有する文書、または特定の語句を含む文書をカテゴリーとしてまとめることができます。ユーザーは、コレクションの検索時に、検索結果を特定のカテゴリーに属する文書に限定することができます。

### 検索フィールドの構成

XML 文書内のエレメントを索引の検索フィールドにマップできます。また、HTML 文書のメタデータ・エレメントを検索フィールドにマップすることもできます。エンタープライズ・サーチ索引に検索フィールドを作成することで、ユーザーが XML および HTML 文書の特定部分を照会できるようになり、検索結果の精度を上げることができます。

### テキスト処理オプションの構成

カスタム・テキスト分析エンジンがエンタープライズ・サーチ・システムに追加されている場合、その 1 つを特定のコレクションで使用するよう選択できます。分析エンジンをコレクションに関連付けた後、コンテンツが言語的に分析され注釈が付けられるようにコンテンツをマップするオプションを指定できます。また、分析結果がどのようにエンタープライズ・サーチ索引または JDBC データベース表にマップされるのかを指定することもできます。

### ランキング調整クラスへのフィールドのマッピング

照会用語に一致するフィールドが含まれている文書が、照会用語に一致する他の文書よりも検索結果内で高くランク付けされるよう指定できます。フィールドをランキング調整クラスにマップするときには、どのコンテンツおよびメタデータ・フィールドがランキング調整されるのかを指定します。また、各ランキング調整クラスが文書のランク付けに使用するスコアを構成することもできます。

### 関連概念

- 「テキスト分析機能ガイド」の『セマンティック検索の言語サポート』
- 「テキスト分析機能ガイド」の『エンタープライズ・サーチに組み込まれているテキスト分析』
- 「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』
- 「テキスト分析機能ガイド」の『セマンティック検索アプリケーション』
- 「テキスト分析機能ガイド」の『セマンティック検索照会』

---

## カテゴリーに関する作業

カテゴリーを使用すると、共通の特性を共有する文書をグループにまとめ、そのグループのメンバーに対する基準に合う文書だけを検索、取得することができます。

文書がカテゴリーに関連付けられていて、検索アプリケーションがこの機能をサポートする場合、ユーザーはカテゴリー名を指定してコレクションのサブセットを検索することができます。コレクション全体を検索する場合、ユーザーは検索結果を絞り込んで、同じカテゴリーにある文書だけを結果文書として参照できるようにすることができます。

エンタープライズ・サーチでは、文書をカテゴリーに分類する場合、次の 2 つの方式をサポートしています。

### ルール・ベース

この方式は、指定した規則に従って文書をカテゴリーに関連付ける場合に使用します。例えば、URI パターンを共有する文書をグループ化したり、または文書コンテンツ (特定の語句を含む、あるいは含まない文書など) に基づいて文書をグループ化することができます。

### モデル・ベース

この方法は、IBM WebSphere Portal を使用しており、WebSphere Portal で事前定義されたカテゴリーを使ってエンタープライズ・サーチ・コレクションを検索する場合に使用します。

## カテゴリー化の変更

カテゴリーを作成、管理するには、エンタープライズ・サーチ管理コンソールを使用します。

- コレクションの作成時に、カテゴリー化タイプを選択します。カテゴリーを使用しない、ルール・ベースのカテゴリーを使用する、またはモデル・ベースのカテゴリーを使用する、のいずれかを選択できます。
- コレクションの構文解析規則を構成する場合、必要に応じてカテゴリー化タイプを変更できます。文書のクロール、索引作成の終了後にカテゴリー化タイプを変更する場合、その全文書を再クロールして、再解析し、索引を再編成しないと、検索の品質が低下します。
- ルール・ベースのカテゴリーを使用する場合、管理コンソールを使用して、カテゴリー・ツリー、カテゴリー、およびカテゴリー規則を管理することができます。文書のクロール、索引作成の終了後にカテゴリーおよびカテゴリー規則を変更する場合、その全文書を再クロールして、再解析し、索引を再編成しないと、検索の品質が低下します。



## 文書コンテンツ・フィールド

モデル・ベースのカテゴリあるいは、(URI パターン規則ではなく) 文書コンテンツ規則を使用するルール・ベースのカテゴリでコレクションを作成する場合、カテゴリ化する文書にコンテンツ・フィールドが含まれるようにしてください。

モデル・ベースのカテゴリおよび文書コンテンツに基づいたカテゴリ規則は、文書のコンテンツ・フィールドについて作用しますが、文書メタデータは考慮しません。コンテンツ・フィールドを含む文書、あるいは、クローラーの構成時に、コンテンツ・フィールドに構成可能なフィールドを含む文書のみ、これらのカテゴリ化オプションを使用してカテゴリ化することができます。

モデル・ベースのカテゴリを構成するか、または文書コンテンツに基づくカテゴリ規則を作成する場合、規則は、文書のコンテンツ部分にのみ適用されます。以下のクローラー・タイプでは、規則はファイル (UNIX のコンテンツまたは Microsoft Windows ファイルなど) あるいは Web ページのコンテンツに適用されず。

- Exchange Server
- NNTP
- UNIX ファイル・システム
- Web
- WebSphere Portal
- Windows ファイル・システム

以下のクローラー・タイプでは、個々のデータ・ソースに対してオプションを構成する際に、どのフィールドが文書コンテンツ・フィールドであるかを指定します。このオプションは、クローラー作成時に、クローラー・ウィザードの最終ページで選択できます。クローラーがすでに存在する場合は、クローラー・スペースを編集し、オプションを編集するデータ・ソースを選択してから、どのフィールドがそのデータ・ソースの文書コンテンツ・フィールドであるかを指定します。

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Lotus Notes
- QuickPlace

## ルール・ベースのカテゴリ

エンタープライズ・サーチ・コレクションで文書をカテゴリに関連付ける場合、規則を構成してどの文書を関連付けるかを制御することができます。

カテゴリの作成、およびエンタープライズ・サーチで作成したコレクションに対するカテゴリ規則および、 IBM WebSphere Portal コレクションからマイグレーションしたルール・ベースのカテゴリに対するカテゴリ規則の作成を行うことができます。文書をカテゴリ化するときの規則を構成する場合、コレクション作成時、またはコレクションの解析規則の指定時にルール・ベースのカテゴリを使用することを指定する必要があります。

パーサーは、指定した規則を使用して文書を 1 つまたは複数のカテゴリに関連付けます。

- 文書が特定のカテゴリ内の少なくとも 1 つの規則に該当した場合、パーサーはその文書をその特定カテゴリに関連付けます。
- 文書が複数のカテゴリ内の少なくとも 1 つの規則に該当した場合、パーサーはその文書をそれらすべてのカテゴリに関連付けます。
- 文書がカテゴリのどの規則にも該当しない場合は、パーサーはその文書をカテゴリに関連付けません。コレクションを検索すれば、この文書を検索、取得することができますが、特定のカテゴリを検索してもこの文書が取得されることはありません。

コレクションのカテゴリ・ツリー (分類構造) を管理する場合、新規のカテゴリをカテゴリ階層のどこに追加するかを指定します。また、カテゴリ・ツリーを使用して、編集するカテゴリを選択し、文書をカテゴリ化する場合の規則を追加、削除、および個別規則の内容を変更することもできます。

文書のカテゴリ化規則を構成する場合、エンタープライズ・サーチが、文書がカテゴリに属しているかどうかを判別する際に文書の URI を使用するか、文書内のコンテンツを使用するかを選択します。

#### URI パターン

URI 規則を文書の URI に適用します。URI の一部 (パターン) を指定し、指定されたパターンを URI に持つ文書がその規則に該当することになります。

例えば、規則のテキストに「/hr/」と指定した場合、次に挙げる最初の URI は規則に該当しますが、2 番目の URI は該当しないことになります。

```
file:///corporate/hr/medicalform.doc  
http://company.com/human_resources/medicalform.htm
```

URI はすべてパターンとして扱うため、パターンの最初あるいは最後に指定したワイルドカード文字としてのアスタリスクは、システムによって無視されます。例えば、\*/hr/\* と /hr/ は、同一の URI とマッチングすることになります。

URI のパターン規則は、大/小文字の区別をしません。URI にスペースが含まれる場合、URI パターンは URI エンコード用のエンタープライズ・サーチ規則に従う必要があります。次の例は、Windows ファイル・システム・パスの URI を指定する場合の正しい方法と間違っている方法を示しています。

```
Incorrect URI: file:///c:/program files/  
正しい URI: file:///c:/program+files/
```

#### 文書コンテンツ

文書コンテンツ規則の指定は、照会と同じフォーマットで行います。文書がその照会に対して有効な場合、その文書は規則に該当することになります。規則を構成する場合、文書に含まれる、または含まれてはならない語句を指定し、語のステミング規則を適用する言語を選択します。

例えば、次の規則では、語として「hr」を含むか、または句として「human resources」を含む文書が規則に該当することになります。

hr "human resources"

また別の例として、次の規則では、語として「hr」を含み、かつ「benefits」という句を含まない文書が規則に該当することになります。

+hr -benefits

コンテンツ規則では、検索および索引 API (SIAPI) 照会と同様の言語正規化が行われます。しかし、コンテンツ規則の構文でサポートされるのは、SIAPI 照会構文で使用可能な操作のサブセットです。以下の照会演算子のみを使用できます。

- + 項の前に正符号を付けると、その項が文書に出現しなければならないことを示します。
- 項の前に負符号を付けると、その項が文書に出現してはならないことを示します。
- " 複数の項を引用符で囲むと、その句が文書中に正確に出現する必要があることを示します。

文書コンテンツ規則は、文書のコンテンツ部分にのみ適用されます。以下のクローラー・タイプでは、規則はファイル (UNIX のコンテンツまたは Microsoft Windows ファイルなど) あるいは Web ページのコンテンツに適用されます。

- Exchange Server
- NNTP
- UNIX ファイル・システム
- Web
- WebSphere Portal
- Windows ファイル・システム

以下のクローラー・タイプでは、個々のデータ・ソースに対してオプションを構成する際に、どのフィールドが文書コンテンツ・フィールドであるかを指定します。このオプションは、クローラー作成時に、クローラー・ウィザードの最終ページで選択できます。クローラーがすでに存在する場合は、クロール・スペースを編集し、オプションを編集するデータ・ソースを選択してから、どのフィールドがそのデータ・ソースの文書コンテンツ・フィールドであるかを指定します。

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Lotus Notes
- QuickPlace

#### 関連タスク

245 ページの『WebSphere Portal からのコレクションのマイグレーション』  
WebSphere Portal からエンタープライズ・サーチにコレクションをマイグレーション

ョンするには、WebSphere Portal にコレクションを準備し、マイグレーション・ウィザードを使用して準備したコレクションをマイグレーションします。

## モデル・ベースのカテゴリ

IBM WebSphere Portal システムで、モデル・ベースのカテゴリを使用している場合、エンタープライズ・サーチ・コレクションでもそれらのカテゴリを継続して使用することができます。

WebSphere Portal では、2,300 を超えるサブジェクトで構成される事前定義の分類を提供しています。これらのサブジェクトは、「コンピューター」、「金融」、および「交通機関」などの業種別カテゴリ分野に分類されています。Portal ユーザーは、これらのサブジェクト・エリアに対応する文書を自動的に判別するアプリケーションを作成でき、さらに、ユーザーの業務ニーズに合わせてこれらのカテゴリをカスタマイズできます。

WebSphere Portal カテゴリをエンタープライズ・サーチで使用する場合、次の条件に従う必要があります。

- マイグレーション・ウィザードを使ってモデル・ベース分類構造ファイルをエンタープライズ・サーチにインポートすること。
- コレクションの作成時、またはコレクションの解析規則の構成時に、モデル・ベースのカテゴリの使用を指定すること。
- WebSphere Portal がエンタープライズ・サーチ索引サーバーにインストールされていること。
- カテゴリの管理には WebSphere Portal のカテゴリ化ツールを使用すること。モデル・ベースのカテゴリは、エンタープライズ・サーチ管理コンソールで管理することはできません。

### 関連タスク

243 ページの『WebSphere Portal からのモデル・ベース分類構造のマイグレーション』

WebSphere Portal 分類管理ポートレットを使用すれば、エンタープライズ・サーチ・コレクションで使用するモデル・ベースの分類構造を選択することができます。すでにエンタープライズ・サーチにマイグレーション済みのコレクションは、新たな分類マイグレーションを選択してもその影響を受けることはありません。

## カテゴリ・ツリー

カテゴリ・ツリーを使用すると、コレクション内のルール・ベースのすべてのカテゴリを表示できます。カテゴリ・ツリーを使って、カテゴリの作成、カテゴリの削除、およびカテゴリに文書を関連付けるときの規則を編集することもできます。

カテゴリ・ツリーは、分類構造とも呼ばれ、階層構造になっています。このツリーは、ルート・カテゴリを起点とし、他のカテゴリはすべてこのルート・カテゴリから分岐します。任意の数のカテゴリおよびサブカテゴリをネスト化して、文書の参照および検索のさまざまな選択肢をユーザーに提供することができます。

例えば、ある文書が複数のカテゴリに規則を受け渡す場合、その文書はそれらすべてのカテゴリに関連付けられます。ユーザーが特定のカテゴリを検索する、または、検索結果を参照するときに特定のカテゴリに属する文書を参照する場合、文書が複数のカテゴリに属していればユーザーがその文書を見つける可能性が高まります。

カテゴリ・ツリーを管理する場合、新規のカテゴリを既存のカテゴリ下にネストすることにより、どの文書がどのカテゴリ (1 つ以上) に属するかを調整することができます。カテゴリを作成するときに、そのカテゴリがルート・レベルで作成されるのか、別のカテゴリのサブカテゴリとして作成するのかを指定できます。また、カテゴリ・ツリーを使用して、コレクションからカテゴリを削除したり、文書をカテゴリに関連付けるときの規則を変更したりすることもできます。カテゴリの編集時には、カテゴリの名前変更、カテゴリ化規則の追加または削除、また個別規則の内容を変更することができます。

カテゴリ・ツリーを管理する場合、検索および参照動作のガイドラインとして以下に記載する説明を参考にしてください。

- ユーザーが上位レベルのカテゴリを検索すると、そのカテゴリ、およびそのカテゴリ以下のすべてのサブカテゴリが、検索基準に合致する文書の検索対象となります。ユーザーがサブカテゴリのないカテゴリを検索する場合、そのカテゴリだけが検索されます。
- ユーザーが検索結果を参照しているときに、特定のカテゴリに属する文書だけを参照するオプションを選択すると、そのカテゴリの文書だけが表示されます。サブカテゴリの名前も検索結果に表示されるので、ユーザーは異なるカテゴリに切り替え、文書のサブセットを一度に表示することができます。

#### 関連タスク

245 ページの『WebSphere Portal からのコレクションのマイグレーション』  
WebSphere Portal からエンタープライズ・サーチにコレクションをマイグレーションするには、WebSphere Portal にコレクションを準備し、マイグレーション・ウィザードを使用して準備したコレクションをマイグレーションします。

## カテゴリ化タイプの選択

カテゴリ化タイプを選択するときに、コレクションの中で文書をカテゴリに関連付けるのに使用する方法を指定します。

#### 始める前に

カテゴリ化タイプを変更するには、エンタープライズ・サーチ管理者役割のメンバーであるか、変更するコレクションのコレクション管理者である必要があります。

#### このタスクについて

カテゴリ化タイプは、コレクションが作成されるときに指定されます。必要であれば、コレクションの文書をカテゴリ化する方法を変更できます。コレクションのために特別に構成するルール・ベースのカテゴリを使用するか、IBM WebSphere Portal システムの中にあるモデル・ベースのカテゴリを使用するか、あるいはカテゴリを使用しないこともできます。

**重要:** データをクロールし、コレクションの索引を作成したあとでカテゴリ化タイプを変更すると、索引の整合性がなくなります。検索結果の正確性を確保するには、コレクション内のすべての文書を再クロールし、文書の構文解析が済んでから索引を再編成します。

## 手順

カテゴリ化タイプを選択するには、次のようにします。

1. コレクションを編集して「構文解析」ページを選択し、「**カテゴリ化タイプの選択**」をクリックします。
2. 「カテゴリ化タイプの選択」ページで、次のオプションの 1 つを選択します。

**なし** このオプションは、このコレクションの中の文書をカテゴリ化しない場合に選択します。

### ルール・ベース

このオプションは、このコレクションのために特別に構成したカテゴリ規則を含む分類構造を使用する場合に選択します。

- エンタープライズ・サーチ用に作成したコレクションを構成する場合は、このオプションを選択して、文書のカテゴリ化のためのカテゴリ名と規則を指定します。
- WebSphere Portal から移行したコレクションを構成する場合は、このオプションを選択して、インポートしたルール・ベースのカテゴリを使用するか変更します。

### モデル・ベース

このオプションは、WebSphere Portal システムにあるモデル・ベースのカテゴリに文書を関連付ける場合に選択します。このオプションを使用するには、エンタープライズ・サーチ索引サーバーに WebSphere Portal がインストールされている必要があります。また、WebSphere Portal 内のカテゴリ化ツールを使用して、カテゴリを管理する必要があります。

3. 「OK」をクリックします。

## カテゴリの構成

1 つのコレクションのカテゴリはいくつも作成することができ、各カテゴリにはいくつかの規則を含めることができます。規則によって、カテゴリと自動的に関連付けられる文書が決まります。

### 始める前に

カテゴリを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのカテゴリが所属するコレクションのコレクション管理者である必要があります。

ルール・ベースのカテゴリを使用するオプションが、カテゴリ化タイプとして選択される必要があります。



文書をカテゴリに関連付ける規則をどのように指定するかについての例は、カテゴリを作成または編集しているときに、「ヘルプ」をクリックしてください。

### このタスクについて

検索アプリケーションでカテゴリのサポートが使用可能であれば、ユーザーは、カテゴリ名を指定することによって、コレクションのサブセットを検索できます。ユーザーはまた、検索結果の中でカテゴリを選択して、そのカテゴリに属する文書のみを参照することもできます。

**重要:** データをクロールし、コレクションの索引を作成したあとでカテゴリまたはカテゴリ規則を変更すると、索引の整合性がなくなります。検索結果の正確性を確保するには、コレクション内のすべての文書を再クロールし、文書を再解析してから、索引を再編成します。

### 手順

カテゴリを構成するには、次のようにします。

1. コレクションを編集して「構文解析」ページを選択し、「**カテゴリ・ツリーの構成**」をクリックします。
2. 「カテゴリ・ツリー」ページで、カテゴリを追加するツリー内の場所を選択し、「**カテゴリの作成**」をクリックします。

ルートを選択すると、新規カテゴリがルート・レベルに作成されます。カテゴリ名を選択すると、新規カテゴリは、カテゴリ・ツリーの中の選択されたカテゴリの下にネストされます。

ウィザードが開き、文書を新規カテゴリに関連付ける規則の指定をガイドします。

- a. 「カテゴリの作成」ページで、カテゴリの記述名を入力し、「**次へ**」をクリックします。
- b. 「カテゴリ規則の作成」ページで、「**規則の追加**」をクリックします。
- c. 「カテゴリ規則の作成」ページで、「**規則名**」フィールドに規則の固有の名前を入力します。この名前はコレクションの中の全カテゴリで固有でなければなりません。
- d. 文書をこのカテゴリに関連付けるために使用する規則を指定して、「**OK**」をクリックします。
  - 文書がカテゴリに所属するかどうかを決めるときにエンタープライズ・サーチが文書の URI を使用するようにする場合は、「**URI パターン**」をクリックして、URI パターンを指定します。

指定したテキストが URI に存在すれば、文書はそのカテゴリに関連付けられます。

例: `file:///c:/program+files/finance`

- 文書がカテゴリに所属するかどうかを決めるときにエンタープライズ・サーチが文書内の語を分析するようにする場合は、「**文書コンテンツ**」をクリックして、文書の言語を選択し、文書コンテンツの中に出現しなけれ

ばならない語、または出現してはならない語を指定します。規則の指定は、照会と同じ書式で行います (照会演算子 +、-、" " のみが許可されます)。

指定した語の包含または除外に適合する文書がそのカテゴリに関連付けられます。

例: +finance -accounting +"fiscal year"

- e. 「完了」をクリックします。

新規カテゴリが、このコレクションに所属する他のカテゴリと一緒に、「カテゴリ・ツリー」ページにリストされます。

#### 関連概念

243 ページの『WebSphere Portal からエンタープライズ・サーチへのマイグレーション』

エンタープライズ・サーチでは、IBM WebSphere Portal からエンタープライズ・サーチに分類構造およびコレクションをマイグレーションするときに使用するマイグレーション・ウィザードが用意されています。

#### 関連タスク

243 ページの『WebSphere Portal からのモデル・ベース分類構造のマイグレーション』

WebSphere Portal 分類管理ポートレットを使用すれば、エンタープライズ・サーチ・コレクションで使用するモデル・ベースの分類構造を選択することができます。すでにエンタープライズ・サーチにマイグレーション済みのコレクションは、新たな分類マイグレーションを選択してもその影響を受けることはありません。

#### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

---

## XML 検索フィールドの作業

ユーザーが XML 文書の特定部分を検索できるようにしたい場合、XML エlement を検索フィールドにマップします。

XML エlement を検索フィールドにマップするには、エンタープライズ・サーチ管理コンソールを使用します。Linux

### XML 検索フィールド

XML 検索フィールドは、ユーザーによる XML 文書の特定部分の照会を可能にします。

XML 文書は、半構造化テキストと構造化されていないテキストの両方を含むため、ますます一般的になってきています。XML の構造はカプセル化されていて、テキストを囲む XML エlement によって明示的に定義されたコンテキストを使用します。例えば、作成者の名前が次のように表されます。

```
<author>John Smith</author>
```

このコンテキストでは、テキスト John Smith によって、XML 文書の作成者が示されます。

エンタープライズ・サーチは、XML エLEMENTの内側にあるテキストを検索フィールド名と関連付ける、つまりマップすることができます。コレクションの構文解析オプションを構成するときに、どの XML ELEMENTがどの検索フィールド名にマップされるのかを指定します。XML ELEMENTを検索フィールドにマップすると、ユーザーはマップ・フィールド名を照会で指定することで、それらのELEMENTを検索できます。特定フィールドを検索する照会を使用すると、全文書コンテンツを検索するフリー・テキスト照会よりも精度の高い検索結果が得られます。

例えば、コレクション中に XML 文書が含まれていて、title ELEMENTと author ELEMENTを索引内で検索フィールドとしてマークされるように指定すると、ユーザーがこれらの特定ELEMENTを照会できるようになります。author:Smith の検索では、author ELEMENTが Smith である XML 文書を検索します。

## XML ELEMENTを検索フィールドにマップ

XML ELEMENTを検索フィールドにマップするときには、ユーザーが照会にフィールド名を指定することで検索可能な XML ELEMENTを指定します。

### 始める前に

XML ELEMENTを検索フィールドにマップするには、エンタープライズ・サーチ管理者役割のメンバーであるか、その XML 文書が所属するコレクションのコレクション管理者である必要があります。

### このタスクについて

XML フィールド・マッピングを作成した場合、または、既存の XML フィールド・マッピングでフィールドの追加、変更、削除を行った場合、それらの変更内容が有効になるのは、パーサーを停止して再始動した後です。パーサーの再始動後、新規マッピングおよび変更されたマッピングは、構文解析されるデータに適用されます。新規マッピングおよび変更されたマッピングは、すでに構文解析され索引付けされたデータには影響しません。

このタスクでは、以下の XML 文書をサンプルとして使用することで、従業員 (personnel) レコードをマップし、ユーザーが特定のELEMENTを直接照会できるようにする方法を示します。

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<personnel>
  <personnelrecord>
    <phone>5555</phone>
    <email>joe@us.ibm.com</email>
    <jobroles>Manager, architect
      <jobrole>Managing Search Development Group</jobrole>
      <jobrole>Architecting Search Technology</jobrole>
    </jobroles>
    <location>New York</location>
    <section id="expertise">
```

```
<text>Linguistics</text>
</section>
</personnelrecord>
</personnel>
```

## 手順

この例の XML エlementを検索フィールドにマップするには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「フィールドへの **XML エlementのマッピング**」をクリックします。
2. 「XML フィールドのマッピング」ページで、「**XML マッピングの作成**」をクリックします。「XML フィールド・マッピングの作成」ページが開きます。
3. 「**XML ルート・Element名**」フィールドに、ルート・Element名 `personnel` を入力します。

ここで指定する名前が、検索したい XML 文書内のルート・Elementと完全に一致するようにしてください。エンタープライズ・サーチは、XML 文書の構文解析および索引付けを行う際、ルート・Element名に従って、どのマッピングを使用するのを選択します。

4. 「**XML マッピング名**」フィールドに、この XML フィールド・マッピング規則セットの名前を入力します。

XML マッピング規則セットを作成した後、この名前が「XML フィールド・マッピング」ページに表示されるので、マッピング規則の追加、削除、または変更を行う対象としてこの名前を選択できます。

5. XML Element `jobrole` を、`jobrole` という名前の検索フィールドにマップします。
  - a. 「**フィールド名**」フィールドに `jobrole` と入力します。
  - b. 「**XML Element名**」フィールドに `jobrole` と入力します。
  - c. 「`jobrole`」フィールドをユーザーが照会できるようにし、検索結果に業務を表示できるようにするには、「**フィールド名による検索**」および「**検索結果**」チェック・ボックスを選択します。
6. XML Element `jobroles` を、同じ検索フィールドにマップします。
  - a. 「**フィールドの追加**」をクリックして、フィールド・マッピング規則のリストに空白行を追加します。
  - b. 「**フィールド名**」フィールドに `jobroles` と入力します。
  - c. 「**XML Element名**」フィールドに `jobrole` と入力します。

**ヒント:** XML Element名は、検索フィールド名と一致する必要はなく、複数の XML Elementを同じ検索フィールドにマップできます。

- d. 「`jobrole`」フィールドをユーザーが照会できるようにし、検索結果に業務を表示できるようにするには、「**フィールド名による検索**」および「**検索結果**」チェック・ボックスを選択します。
7. 属性 `expertise` を持つ XML Element `section` を、`expertise` という名前の検索フィールドにマップします。

- a. 「フィールドの追加」をクリックして、フィールド・マッピング規則のリストに空白行を追加します。
  - b. 「フィールド名」フィールドに expertise と入力します。
  - c. 「フィールド名」フィールドに section と入力します。
  - d. 「XML 属性名」フィールドに id と入力します。
  - e. 「XML 属性値」フィールドに expertise と入力します。
  - f. 「expertise」フィールドをユーザーが照会できるようにし、検索結果に専門的知識を表示できるようにするには、「フィールド名による検索」および「検索結果」チェック・ボックスを選択します。
8. 「OK」をクリックします。

例:

検索 (search) 製品を業務とする部門の全員を検出するには、次のような照会を指定します。

```
jobrole:search
```

言語 (linguistics) を専門にする部門の全員を検出するには、次のような照会を指定します。

```
expertise:linguistics
```

---

## HTML 検索フィールドの作業

ユーザーが HTML 文書の特定メタデータ部分を検索できるようにする場合、HTML メタデータ・エレメントを索引の検索フィールドにマップします。

HTML メタデータ・エレメントを検索フィールドにマップするには、エンタープライズ・サーチ管理コンソールを使用します。

## HTML 検索フィールド

HTML 検索フィールドにより、ユーザーは HTML 文書の属性を照会できます。

HTML 文書のメタデータ・エレメントは、文書に関する情報を提供するという点、そのフォーマット方法、および Web 上のアクセス許可の方法の点において、文書の属性と類似しています。以下に例を示します。

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" /> ]
<meta name="copyright" content="(C) Copyright IBM Corporation 2005" />
<meta name="content.owner" content="(C) Copyright IBM Corporation 2005" />
<meta name="security" content="public" />
<meta name="abstract" content="This topic describes an IBM product." />
<meta name="format" content="XHTML" />
```

エンタープライズ・サーチは、HTML メタデータ・エレメントの名前を検索フィールド名に関連付ける、すなわちマップすることができます。コレクションの構文解析オプションを設定するときに、どの HTML メタデータ・エレメントをどの検索フィールド名にマップするのかを指定します。HTML メタデータ・エレメントを検索フィールドにマップすると、ユーザーは照会に検索フィールド名を指定するこ

とで、それらのエレメントを含む文書を検索できます。特定フィールドを検索する照会を使用すると、全文書コンテンツを検索するフリー・テキスト照会よりも精度の高い検索結果が得られます。

例えば、コレクションに HTML 文書が含まれていて、copyright メタデータ・エレメントと abstract メタデータ・エレメントを検索フィールドとして索引付けすると指定すると、ユーザーはこれらの特定エレメントを照会できるようになります。

「copyright:IBM」で検索すると、copyright メタデータに IBM という語のある HTML 文書を検出できます。

## 検索フィールドへの HTML メタデータ・エレメントのマッピング

HTML メタデータ・エレメントを検索フィールドにマップするときには、ユーザーが照会にフィールド名を指定することで検索可能な HTML エレメントを指定します。

### 始める前に

HTML メタデータ・エレメントを検索フィールドにマップするには、エンタープライズ・サーチ管理者役割のメンバーであるか、その HTML 文書が所属するコレクションのコレクション管理者である必要があります。

### このタスクについて

HTML フィールド・マッピングを作成したり、既存の HTML フィールド・マッピングでフィールドの追加、変更、削除を行った場合、それらの変更内容が有効になるのは、パーサーを停止して再始動した後です。変更は、パーサーを再始動した後で構文解析されたデータにのみ適用されます。すでに構文解析され、索引付けられている文書を更新するには、目的の文書を再度クロールして構文解析してから、索引を再編成する必要があります。

### 手順

検索フィールドに HTML メタデータ・エレメントをマップするには、次のようにします。

1. コレクションを編集して「構文解析」ページを選択し、「HTML メタデータをフィールドにマップ」をクリックします。
2. 「HTML フィールド・マッピング」ページで、「フィールドの追加」をクリックして、フィールド・マッピング規則のリストに空白行を追加します。
3. 「フィールド名」フィールドに、マップしようとする HTML メタデータ・エレメントと関連付けたい名前を入力します。ユーザーは、このコレクション内の HTML 文書を照会するときに、このフィールド名を指定できます。
4. 「HTML メタデータ・エレメント名」フィールドに、マップしようとする HTML メタデータ・エレメントの名前を入力します。
5. ユーザーがこのフィールドを照会し、検索結果にそれを表示できるようにするために、「フィールド検索」チェック・ボックスと「検索結果」チェック・ボックスを選択します。
6. このフィールドのデータ・タイプが、DECIMAL、DOUBLE、INTEGER、SHORT、TIME、または TIMESTAMP で、



かつ、このフィールドの検索時にユーザーがパラメトリック照会を指定できるようにしたい場合は、「パラメトリック検索」チェック・ボックスを選択します。

7. 「OK」をクリックします。

例:

これで、ユーザーは、マップ・フィールド名を照会して、特定のメタデータをもつ HTML 文書を検索できるようになりました。例えば、description という名前の HTML メタデータ・エレメントを、abstract という名前の検索フィールドにマップしてある場合、ユーザーは、以下のような照会を入力すると、Thinkpad コンピューターに関する文書を検索できます。

```
abstract:thinkpad
```

---

## カスタム・テキスト処理

カスタム・テキスト処理アルゴリズムをエンタープライズ・サーチ・コレクションと統合することにより、検索結果の品質と精度を高めることができます。

WebSphere Information Integrator OmniFind Edition は、IBM Unstructured Information Management Architecture (UIMA) をサポートしています。これは、テキスト分析機能の作成、ディスカバリー、合成、デプロイを行うためのフレームワークです。アプリケーション開発者は、コンテンツが検索される分析アルゴリズムを作成し、テストしてから、エンタープライズ・サーチにアーカイブを使用するのに必要なすべてのリソースが含まれている処理エンジン・アーカイブ (.pear ファイル) を作成します。カスタム・テキスト分析アルゴリズムを使ってコレクションを検索できるようにするために、アーカイブ (分析エンジンが含まれています) をエンタープライズ・サーチ・システムに追加する必要があります。

テキスト分析エンジン内の分析論理コンポーネントをアノテーター といいます。各アノテーターは、特定の言語分析タスクを実行します。テキスト処理エンジンは、任意の数のアノテーターが含まれているものであっても、また、それぞれに固有のカスタム・アノテーターが含まれている幾つかの分析エンジンの複合であっても構いません。

アノテーターによって作成された情報を、分析結果 といいます。分析結果は、検索したい情報に相当するもので、共通分析構造 と呼ばれるデータ構造に書き込まれます。

コレクションについてテキスト処理オプションを構成する際には、以下の操作を行ってください。

- コレクション内の文書に注釈を付けるのに使用するテキスト分析エンジンを選択します。
- コレクションに分かりやすいマークアップをもつ XML 文書が含まれており、このマークアップをカスタム・テキスト分析で使いたい場合は、XML マッピング・ファイルをコレクションと関連付けて、XML マッピングの出力を共通分析構造にマップすることができます。

例えば、共通分析構造内で、<addressee> エレメントおよび <customer> エレメントの内容を Person 注釈にマップできます。こうすると、これらの注釈にカスタ

ム・アノテーターからアクセスできるようになり、その他の情報を検出できます (例えば、Person の性別を検出できます)。Person 注釈をエンタープライズ・サーチ索引にマップすることもできます。こうすると、ユーザーは、元の XML エレメントを知らなくても、Person を検索することができます。

ユーザーが照会で元の XML エレメントを指定できるようにしたい場合、XML マッピングを定義する必要はありません。その代わりに、構文解析オプションを構成し、コレクションについてネイティブ XML 検索を使用可能に設定できます。

- エンタープライズ・サーチ索引に共通分析構造にあるデータ構造をマップします。これにより、注釈が付けられた文書をセマンティック検索で検索することができるようになります。

例えば、アノテーターによって検出されるエンティティと関係に応じて、ユーザーは、同じ文で発生する概念や、キーワードと概念 (例えば、用語 Alex と電話番号) を検索できます。

- 共通分析構造にあるデータ構造を、Java Database Connectivity (JDBC) 対応であるデータベース表にマップします。データを IBM DB2 Universal Database (DB2 UDB) または Oracle の表にマップできます。このタイプのマッピングによって、分析の結果をデータ・マイニングのようなデータベース・アプリケーションで使用することができるようになります。また、SQL 照会を使用して、エンタープライズ・サーチの外部にあるデータを検索することができるようになります。

#### 関連概念

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』

「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』

「テキスト分析機能ガイド」の『テキスト分析アルゴリズム』

「テキスト分析機能ガイド」の『セマンティック検索アプリケーション』

「テキスト分析機能ガイド」の『セマンティック検索照会』

## システムへのテキスト分析エンジンの追加

カスタム・テキスト分析エンジンを作成する場合は、システムにそれを追加してからでないと、エンタープライズ・サーチ用に使用できません。コレクションは、エンジンを使用して、文書の分析と注釈付けを行い、検索結果の精度を高めることができます。

### 始める前に

システムにテキスト分析エンジンを追加するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて

アプリケーション開発者は、テキスト分析用の UIMA フレームワークに従う処理エンジン・アーカイブ (.pear) を作成できます。アーカイブには、エンタープライズ・サーチ・コレクションを検索するのに必要なリソースのすべてが含まれます。カス


タム・テキスト分析アルゴリズムを使ってコレクションを検索できるようにするために、アーカイブ (分析エンジンが含まれています) をエンタープライズ・サーチ・システムに追加する必要があります。

テキスト分析エンジンをシステムに追加後、その表示名を変更して、XML ソースを表示するオプションを選択できます。(XML ソースは、どの情報がこのエンジンで作成されたのかを示します。)

テキスト分析エンジンが既にこのコレクションと関連付けられている場合、テキスト分析エンジンをシステムから除去することはできません。

### 手順

カスタム・テキスト分析エンジンをエンタープライズ・サーチ・システムに追加するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2.  「編集」をクリックして、システム編集ビューに移動します。
3. 「構文解析」ページで、「テキスト分析エンジンの構成」をクリックします。
4. 「テキスト分析エンジン」ページで、「テキスト分析エンジンの追加」をクリックします。
5. 「テキスト分析エンジンの追加」ページで、新規エンジンの記述名を入力します。システムは、この表示名を使用して、管理コンソール全体でテキスト分析エンジンを識別します。
6. .pear ファイルのロケーションを指定します。ファイルが 8 MB より小さい場合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より大きい場合、そのファイルは索引サーバーに置かなければなりません。
7. 「OK」をクリックします。テキスト分析エンジンが「テキスト分析エンジン」ページにリストされます。

### 関連概念

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』

「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

「テキスト分析機能ガイド」の『分析および検索における XML マークアップの使用方法』

### 関連タスク

「テキスト分析機能ガイド」の『UIMA タイプへの XML マッピング構成ファイルの作成』

## テキスト分析エンジンとコレクションとの関連付け

カスタム・テキスト分析エンジンがエンタープライズ・サーチ・システムと関連付けられている場合、そのうちの 1 つを、このコレクションで使用するよう選択できます。そうすると、ユーザーは、コレクションを検索するときに意味的照会を指定し、検索結果の品質と精度を高めることができます。

### 始める前に

テキスト分析エンジンをコレクションと関連付けるには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

テキスト分析エンジンがすでにこのコレクションと関連付けられている場合、別のエンジンを関連付けるときに、以下のアクションが発生します。

- 「**カスタム分析なし**」を選択した場合、コレクションについて以前に定義したすべてのテキスト分析マッピングはリセットされます。コレクションはシステム・デフォルト値を使用し始めます。
- 別のカスタム・テキスト分析エンジンを選択した場合、コレクションについて以前に定義したすべてのテキスト分析マッピングは保存されます。例えば、engine\_1 から engine\_2 に変更した場合、engine\_2 は、engine\_1 について構成された XML マッピング・ファイルを継承します。

### 手順

テキスト分析エンジンをコレクションと関連付けるには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「**テキスト処理オプションの構成**」をクリックします。
2. 「**テキスト分析エンジンの選択**」をクリックします。エンタープライズ・サーチ・システムにカスタム・テキスト分析エンジンが追加されていない場合、またはコレクションがデフォルトの分析アルゴリズムを使用している場合、エンジン名は **Default** です。
3. 「コレクションのテキスト分析エンジンの選択」ページで、このコレクションで使用したいエンジンの名前を選択します。選択可能なテキスト分析エンジンがない場合、または「**カスタム分析なし**」を選択した場合、パーサーは、文書に注釈を付け、索引の準備をするときに、デフォルトのテキスト分析規則を適用します。
4. 「**OK**」をクリックします。

#### 関連概念

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』

「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

## 共通分析構造への XML エLEMENTのマッピング

コレクションに分かりやすいマークアップをもつ XML 文書が含まれており、このマークアップを使用して、ユーザーがエンタープライズ・サーチ索引を、またはセマンティック検索で JDBC 表を検索できるようにしたい場合、XML エLEMENTを共通分析構造にマップすることができます。

### 始める前に

XML エlementを共通分析構造にマップするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

カスタム・テキスト分析プロセスが XML 文書内の特定のElementにアクセスできるようにしたり、セマンティック検索で使用するために共通の Type にいくつかの XML Elementをマップするために、カスタム XML マッピング・ファイルを作成できます。XML マッピング・ファイルは、テキスト分析のための UIMA フレームワークに従うものでなければなりません。

カスタム・テキスト分析エンジンを使用するコレクションに XML マッピング・ファイルを追加すると、ソース文書内の XML Elementを共通分析構造内の注釈にマップすることができます。こうすると、これらの注釈をカスタム・テキスト分析エンジンで 사용할ことができます。それにより、(共通分析構造内の) 分析結果を索引にマップすることができ、ユーザーは、セマンティック検索を使用してコレクションを検索するときに注釈を照会することができます。

例えば、共通分析構造内で、addressee Elementおよび customer Elementの内容を Person 注釈にマップできます。こうすると、これらの注釈にカスタム・アナレーターからアクセスできるようになり、その他の情報を検出できます (例えば、Person の性別を検出できます)。Person 注釈をエンタープライズ・サーチ索引にマップすることもできます。こうすると、ユーザーは、元の XML Elementを知らなくても、Person を検索することができます。

ユーザーが照会で元の XML Elementを指定できるようにしたい場合、XML マッピングを定義する必要はありません。その代わりに、構文解析オプションを構成し、コレクションについてネイティブ XML 検索を使用可能に設定できます。

## 手順

XML Elementを共通分析構造にマップするには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「**テキスト処理オプションの構成**」をクリックします。
2. 「**XML Elementを共通分析構造にマップ**」領域で、「**マッピングの追加**」をクリックします。
3. 「XML Elementを共通分析構造にマップ」ページで、XML マッピング・ファイルの記述表示名を入力します。
4. ファイルのロケーションを指定します。XML マッピング・ファイルが 8 MB より小さい場合、そのファイルのパスを入力するか、またはブラウズすることができます。XML マッピング・ファイルが 8 MB より大きい場合、そのファイルが索引サーバー上にあることを確認し、そのファイルの完全修飾パスを指定します。
5. 「**OK**」をクリックします。新規 XML マッピング・ファイルは、「テキスト処理オプション」ページの XML マッピング・ファイルのリストに追加されます。

## 関連概念

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』

「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』



「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

「テキスト分析機能ガイド」の『分析および検索における XML マークアップの使用方法』

### 関連タスク

「テキスト分析機能ガイド」の『UIMA タイプへの XML マッピング構成ファイルの作成』

## 索引への共通分析構造のマッピング

ユーザーがセマンティック検索でコレクションを照会するとき、必要とする共通分析構造を指定できます。

### 始める前に

共通分析構造を索引にマップするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

共通分析構造をエンタープライズ・サーチ索引にマップすることにより、ユーザーが意味的に的確な照会を指定して、検索結果の品質を高めることができます。

例えば、アノテーターによって検出されるエンティティと関係に応じて、ユーザーは、同じ文で発生する概念（例えば、特定の人物と任意の競争者名）、あるいはキーワードと概念（例えば、用語 Alex と電話番号）を検索できます。

### 手順

共通分析構造を索引にマップするには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「テキスト処理オプションの構成」をクリックします。
2. 「1 つの共通分析構造を索引にマップ」領域で、「共通分析構造の選択」をクリックします。
3. 「コレクションの共通分析構造の選択」ページで、エンタープライズ・サーチ索引で使用したいマッピングを選択します。
  - 共通分析構造にあるデータ構造をエンタープライズ・サーチ索引にマッピングするためのデフォルトのシステム規則を使用するには、「デフォルト」を選択します。
  - 特定の共通分析構造を索引にマップするには、共通分析構造ファイルのロケーションを指定します。ファイルが 8 MB より小さい場合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より大きい場合、そのファイルは索引サーバーに置かなければなりません。
4. 「OK」をクリックします。指定した共通分析構造は、「テキスト処理オプション」ページに表示されます。

### 関連概念

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』



「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』  
「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』  
「テキスト分析機能ガイド」の『カスタム分析結果の索引付けの方法』

#### 関連タスク

「テキスト分析機能ガイド」の『索引作成構成ファイルの作成』

## JDBC 表への共通分析構造のマッピング

データベース・アプリケーションで使用するために、JDBC 表にマップしたい共通分析構造を指定できます。

### 始める前に

共通分析構造を Java Database Connectivity (JDBC) にマップするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

共通分析構造を JDBC 表にマップすることにより、データがデータベース・アプリケーションで使用できるようになります。例えば、ユーザーは、エンタープライズ・サーチの外部の SQL 照会を指定して、共通分析構造によって追加された注釈を検索できます。また、データ・マイニング・アプリケーションの情報を使用するなど、一層のテキスト処理の情報を使用できます。

### 手順

共通分析構造を JDBC 表にマップするには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「テキスト処理オプションの構成」をクリックします。
2. 「JDBC 表への共通分析構造のマッピング」領域で、「マッピングの追加」をクリックします。
3. 「JDBC 表への共通分析構造のマッピング」ページで、JDBC 表に情報をマップするのに使用したい共通分析構造の記述表示名を入力します。
4. 共通分析構造ファイルのロケーションを指定します。ファイルが 8 MB より小さい場合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より大きい場合、そのファイルは索引サーバーに置かなければなりません。
5. 「OK」をクリックします。新規ファイルは、「テキスト処理オプション」ページのファイルのリストに追加されます。ファイルをシステムに追加した日時も表示されます。

### 関連概念

「テキスト分析機能ガイド」の『カスタム分析の組み込みのワークフロー』  
「テキスト分析機能ガイド」の『カスタム・テキスト分析の組み込み』  
「テキスト分析機能ガイド」の『Unstructured Information Management Architecture (UIMA)』

「テキスト分析機能ガイド」の『分析結果のデータベースへのマッピングの方法』

### 関連タスク

「テキスト分析機能ガイド」の『XML マッピング構成ファイルの作成』

---

## パーサー・サービス用スレッドの構成

十分なメモリー・リソースがある場合、パーサーが文書の構文解析に使用できるスレッドの数を増やすことができます。

### 始める前に

コレクションが多数ある場合などに、パーサー・スレッド数を増やすことができます。ただし、追加スレッドをサポートするのに十分なメモリーがシステムにあることを確認してください。スレッドが 1 つのパーサーは 200 MB のメモリーを必要とします。スレッドを 1 つ増やすごとに 50 MB のメモリーが追加が必要になります。

パーサー用に開始されるスレッドの数を構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

パーサー・スレッドの数を構成するには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「**構文解析オプションの構成**」をクリックします。
2. パーサー開始時に開始されるパーサー・スレッドの最大数を指定し、「**OK**」をクリックします。
3. パーサーを再始動します。

---

## 複合語の拡張分析の使用可能化

パーサーが複合語に対して拡張分析を使用することを使用可能化することにより、検索の品質を高めることができます。拡張分析を使用すると、複合語は分解され、各パーツを 1 つの語であるように扱うことができるようになります。

### 始める前に

複合語の構文解析に関するオプションを指定するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

言語によっては、複数の語からなる単語を空白を入れずに 1 つの語にまとめるものがあります (複合語)。複合語の拡張分析と分解は、ドイツ語のような言語の検索で有用であり、韓国語のような言語の検索では特に重要です。

複合語の拡張分析を使用可能にしておくと、ユーザーは、ワイルドカード文字を使用して複数の照会用語の組み合わせを表す必要なく、語句の検索を行うことができます。例えば、Organ (臓器) を検索すると、Organspender (臓器ドナー) が含まれている文書が戻されますが、Organisation (組織) が含まれている文書は戻されません。ワイルドカード照会 Organ\* では Organ の後にどのようなストリングが続いていてもすべて戻されますが、この検索はそれとは異なり、複合語を構成する言語的に可能なサブワード全体を突き合わせます。

同義語やランキング調整ワードなどのユーザー定義の語も、照会で個々の語として使用される複合パーツに当てはまります。

### 手順

複合語の拡張分析を使用可能化するには、次のようにします。

1. コレクションを編集して「構文解析」を選択し、「**構文解析オプションの構成**」をクリックします。
2. 「**複合語の拡張分析を使用可能にする**」チェック・ボックスを選択し、「**OK**」をクリックします。

### 関連概念

「テキスト分析機能ガイド」の『セマンティック検索の言語サポート』

「テキスト分析機能ガイド」の『エンタープライズ・サーチに組み込まれているテキスト分析』

---

## ネイティブ XML 検索のサポートの使用可能化

コレクションに XML 文書が含まれている場合、ユーザーがコレクションのネイティブ XML 検索を使用可能にすることにより文書を検索するときに XML マークアップを使用できるようにすることができます。

### 始める前に

XML 文書をネイティブ XML 検索で 検索するオプションを使用可能にするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

XPath または XML フラグメントなど、ネイティブ XML 検索は、照会内で文書の XML マークアップを活用して検索結果の精度を高めることができます。例えば、ユーザーは、特定の XML エlement または属性で照会用語が発生するよう指定できます。

例えば、XML 形式のコンピューター小売業者からの送り状に、<company> エlement と <computertype> エlement が組み込まれている <order> 項目が含まれていることがあります。IBM ノートブックの注文が含まれている送り状を取得するために、IBM とノートブックのキーワード検索により、Dell ノートブック・コンピューターおよび IBM デスクトップ・モデルが含まれている文書が取り出されます。XML 検索を使用すると、<company> エlement に IBM が出現すること、<computertype> エlement に notebook が出現すること、さらに、両方のエlemen

トと同じ <order> エレメントの下にあることを指定できます。この方法で、IBM ノートブックを明確に論じている送り状を取得します。

### 手順

ユーザーがネイティブ XML 検索でコレクションを検索できるようにするには、次のようにします。

1. コレクションを編集、「構文解析」ページを選択、そして「**構文解析オプションの構成**」をクリックします。
2. 「**XML 文書の検索でネイティブ XML 検索を可能にする**」を選択します。
3. 「**OK**」をクリックします。

### 関連概念

「テキスト分析機能ガイド」の『セマンティック検索の言語サポート』

「テキスト分析機能ガイド」の『エンタープライズ・サーチに組み込まれているテキスト分析』

「テキスト分析機能ガイド」の『セマンティック検索アプリケーション』

「テキスト分析機能ガイド」の『セマンティック検索照会』

---

## 中国語、日本語、韓国語の文書の言語分析

中国語、日本語、韓国語で書かれた文書の検索性を高めるには、言語分析オプションを指定します。

中国語、日本語、韓国語の文書の場合、パーサーで字句解析に N-gram セグメンテーション方式を使用することを指定できます。また、中国語および日本語の文書では、パーサーが空白から改行文字を除去するように設定できます。

### N-gram セグメンテーション

コレクションの作成時に、ワードの区切り文字として空白を使用しない言語で書かれた文書の構文解析に使用する字句解析のタイプを選択します。

Unicode ベースの空白文字セグメンテーションでは、ワードとワードの間の区切り文字としてブランク・スペースを使用します。N-gram セグメンテーションは、任意の数の文字が並ぶ連続を単一ワードとみなします。ワードの区切り文字としてブランクを使用しない、中国語、日本語、韓国語のような言語では、N-gram セグメンテーションを使用すれば、Unicode ベースの空白文字セグメンテーションよりも質のよい検索結果を戻すことができます。

コレクションの作成時に文書の構文解析に使用するセグメンテーション方式を選択します。コレクション作成後は、構文解析オプションを表示して設定を確認することはできますが、変更することはできません。

### 非 ASCII 文字範囲からの改行文字の除去

中国語や日本語など、ワード境界を定めるのに空白文字が使用されない言語では、改行の原因となる特定の空白文字を除去するようパーサーを構成できます。

### 始める前に

この操作を完了するには、エンタープライズ・サーチ管理者としてログインする必要があります。

### このタスクについて

このオプションがコレクションについて使用可能になっており、文書が中国語または日本語で作成されている場合、パーサーは、2つの文字を分離する空白文字のシーケンスを除去します。その文字は、非 ASCII ユニコード文字範囲からのものでなければなりません。以下の文字が除去されます。

- タブ (0x09)
- LF または改行 (0x0A)
- CR または復帰 (0x0D)

変更を有効にするために、パーサーを停止し、再開します。すでに構文解析済みで、索引に入れられている文書に変更を適用するには、目的の文書をもう一度クロールして構文解析してから、索引を再編成する必要があります。

### 手順

空白文字から改行文字を除去するには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. テキスト・エディターを使用して以下のファイルを編集します。ここで、`collection_ID` は、コレクションの作成時にそのコレクションについて指定された ID です (または、システムによって割り当てられたもの)。  
`collection_ID.parserdriver/collection.properties`
3. `removeCjNewlineChars` プロパティの値を、`false` から `true` に変更します。

---

## コレクション・パーサーと Stellent セッションに関連付けられた文書タイプ

クロール・スペース内の文書が正確にかつ効率的に構文解析されるように、どのタイプの文書をコレクション・パーサーで構文解析し、どの文書を Stellent 文書フィルター処理セッションで構文解析するかを指定する構成ファイルを作成できます。

エンタープライズ・サーチ・コレクションでは、ほとんどの文書フォーマットが標準装備の HTML パーサーまたは XML パーサーで処理されます。一定タイプの文書 (ポストスクリプト文書など) が、通常、構文解析されず、他に Stellent 解析機能で処理される文書タイプ (Microsoft Word、Microsoft Excel、Microsoft PowerPoint、Lotus Freelance、Lotus 123、PDF、RT、および Ichitaro 文書タイプなど) があります。

メタデータは誤りを招きやすいため、プレーン・テキストおよび HTML 文書が Stellent セッションに間違えて送信されて、それから標準装備のパーサーの 1 つに戻される可能性があり、パフォーマンスに影響を与えかねない状態になります。この状況を避けるために、構成ファイルを作成して種々のタイプの文書をどこでどのように構文解析するかを制御することができます。

文書タイプをコレクション・パーサーと Stellent セッションに関連付けるには、以下の作業が必要です。

1. コレクション・パーサー用の文書タイプを構成する。このステップでは、文書タイプをコレクションが使用するパーサーにマップする構成ファイルを作成します。コレクションごとに、こうした構成ファイルを 1 つ作成することができます。
2. Stellent セッション用の文書タイプを構成する。このステップでは、文書タイプをコレクションが使用する Stellent 文書フィルターにマップする構成ファイルを作成します。コレクションごとに、こうした構成ファイルを 1 つ作成することができます。
3. パーサーの停止と再始動。変更を有効にするには、エンタープライズ・サーチ管理コンソールを使用して、文書タイプを構成したコレクションをモニターしてから、パーサーを停止して再始動します。

## 文書タイプとコレクション・パーサーとの関連付け

特定の文書タイプをコレクション・パーサーと関連付けるには、構成ファイルを作成します。エンタープライズ・サーチ管理コンソールでは、このタスクのサポートはありません。

### 始める前に

この操作を完了するには、エンタープライズ・サーチ管理者としてログインする必要があります。

### このタスクについて

構成ファイルが存在しない場合、コレクション・パーサーはデフォルトのパーサー・サービス規則を使用します。構成ファイルが存在する場合、ファイル内の規則は次のものを指定します。

- ファイル拡張子またはコンテンツ・タイプに応じて、Stellent セッションに送信したい文書。
- メタデータが不完全であるためにタイプが不明な文書の構文解析方法。

ファイルのフォーマットは、一連の行で、各行が以下の規則の 1 つです。

#### **EXTENSION** *extension parser*

指定の拡張子で終わる URL をもつすべての文書が、指定のパーサーによって処理されます。拡張子にピリオドを含めないでください。比較は大/小文字の区別はありません。

#### **CONTENTTYPE** *type/subtype parser*

指定のタイプ/サブタイプに一致するコンテンツ・タイプをもつすべての文書が、指定のパーサーによって処理されます。文書のコンテンツ・タイプ t/s を指定した場合、一致は、t が type に等しく、かつ、s が subtype に等しいか、またはサブタイプがワイルドカード文字 (アスタリスク \*) である場合に発生します。



### UNKNOWN parser

不明な拡張子とコンテンツ・タイプ (すなわち、クローラーにより使用可能にされていない) をもつすべての文書が、指定のパarserによって処理されます。

### DEFAULT parser

他のどの規則の対象にもならないすべての文書が、指定のパarserによって処理されます。

いかなる場合でも、*parser* は、html、xml、stellent、または none を指定する必要があります。ここで、none は、そのタイプの文書が構文解析されないことを意味します。

1 つの文書に複数の規則が一致する場合は、規則の出現順は関係なく、より明確な特定の規則が優先されます。

- EXTENSION 規則は、CONTENTTYPE 規則よりも明確です。
- サブタイプが含まれている CONTENTTYPE 規則は、ワイルドカード文字をもつ規則よりも明確です。例えば、コンテンツ・タイプ application/postscript の規則は、application/\* の規則よりも優先されます。
- 同じ拡張子またはコンテンツ・タイプに規則が 2 つあってはなりません。規則が 2 つある場合、どちらの規則が優先されるかはインプリメンテーション次第です。

### 手順

文書タイプをコレクション・parserと関連付けるには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. 以下のように、構成ファイルを作成します。ここで、*collection\_ID* は、構成したいコレクションを示します。

```
ES_NODE_ROOT/master_config/collection_ID.parserdriver/parserTypes.cfg
```
3. テキスト・エディターを使用して構成ファイルを編集し、parser・サービス規則を指定したら、そのファイルを保管して終了します。
4. 変更を有効にするために、エンタープライズ・サーチ管理コンソールを使用してコレクションのparserをモニターし、parserを停止して、再始動します。

### 例

この例では、組み込み HTML parserは、拡張子 txt、htm、または html をもつか、text/ で始まるコンテンツ・タイプをもつ、あるいは不明な拡張子とコンテンツ・タイプをもつ、すべての文書を処理します。組み込み XML parserは、拡張子 xml をもつか、またはコンテンツ・タイプ text/xml をもつ、すべての文書を処理します。application/ で始まるコンテンツ・タイプをもつ文書を含め、他の文書はすべて、Stellent セッションに送信されます。

```
EXTENSION doc stellent
EXTENSION txt html
EXTENSION htm html
EXTENSION html html
EXTENSION xml xml
```

```
EXTENSION ps none
CONTENTTYPE text/xml xml
CONTENTTYPE text/* html
CONTENTTYPE application/* stellent
UNKNOWN html
DEFAULT stellent
```

## デフォルトのコレクション・パーサー・サービス規則

ファイル・タイプとコンテンツ・タイプをコレクションのパーサーにマッピングする構成ファイルを作成しない場合、文書の構文解析にはデフォルトの規則が使用されます。

コレクション・パーサーで使用されるデフォルトの規則は以下のようになります。

```
EXTENSION pdf stellent
EXTENSION ppt stellent
EXTENSION prz stellent
EXTENSION lwp stellent
EXTENSION doc stellent
EXTENSION rtf stellent
EXTENSION xls stellent
EXTENSION 123 stellent
EXTENSION vsd stellent
EXTENSION vdx stellent
EXTENSION jxw stellent
EXTENSION jsw stellent
EXTENSION jtw stellent
EXTENSION jaw stellent
EXTENSION juw stellent
EXTENSION jbw stellent
EXTENSION jvw stellent
EXTENSION jfw stellent
EXTENSION jtt stellent
EXTENSION jtd stellent
EXTENSION jttd stellent
EXTENSION jtdd stellent
EXTENSION jtddx stellent
EXTENSION ps none
EXTENSION xml xml
EXTENSION txt text
EXTENSION htm html
EXTENSION html html
EXTENSION shtml html
EXTENSION xhtml html
EXTENSION asp html

CONTENTTYPE application/postscript none
CONTENTTYPE application/* stellent
CONTENTTYPE text/rtf stellent
CONTENTTYPE text/richtext stellent
CONTENTTYPE text/xml xml
CONTENTTYPE text/html html
CONTENTTYPE text/plain text

UNKNOWN html
DEFAULT html
```

## 文書タイプと Stellent セッションとの関連付け

Stellent 文書フィルターで構文解析する文書のタイプを指定するために、構成ファイルを作成します。エンタープライズ・サーチ管理コンソールでは、このタスクのサポートはありません。

## 始める前に

この操作を完了するには、エンタープライズ・サーチ管理者としてログインする必要があります。

## このタスクについて

構成ファイルは、次のものを指定します。

- **Stellent** セッションによって構文解析されるファイル・タイプ。ファイル・タイプは、**Stellent** ライブラリーが認識するファイル・タイプの 1 つに一致します。
- コレクション・パーサーに送り返して、組み込みパーサーの 1 つで処理するファイル・タイプ。(このアクションが必要なのは、紛らわしいメタデータのために、コレクション・パーサーが誤って文書を **Stellent** セッションに送信する場合があります。)
- エンタープライズ・サーチでサポートされていないためにリジェクトされるファイル・タイプ。

構成ファイルが指定されたが、存在しない場合、パーサーは開始できません。  
`stellent.properties` ファイル内の `OutsideInSupportedTypes` プロパティーに構成ファイルが指定されていない場合、**Stellent** セッションのデフォルトの構文解析規則が使用されます。

構成ファイルには、文書タイプと、それらの取り扱い方法がリストされています。ファイルのフォーマットは、一連の行で、各行が以下の規則の 1 つです。

```
accept DEFAULT
accept ALL doctype
accept type doctype
native DEFAULT
native type doctype
reject type
```

ここで、それぞれ次のことを表します。

### *doctype*

*doctype* 照会トークンに使用される値です。文書を文書タイプで検索できます。例えば、ユーザーは、PDF 文書を検索するために `$doctype::pdf` を指定できます。

*type* **Stellent** ライブラリー内の `FI_` 値の 1 つです。規則が適用された場合、*doctype* は *doctype* トークンに使用される値です。

### **DEFAULT**

規則のタイプに応じて、受け入れられたタイプまたはネイティブ・タイプのリストにデフォルト・リストが含まれることを意味します。このオプションを選択すると、デフォルトの構成を置き換える代わりに、展開することができます。

**A11** 明示的にリストされないすべてのタイプが指定の *doctype* トークンを使用して受け入れられることを意味します。

構成ファイル内の規則は、以下のように処理されます。

- `accept DEFAULT` が指定された場合のデフォルト・リストを含め、*type* について `accept` 規則がある場合、その規則は受け入れられます。

- また、*type* について reject 規則がある場合、その規則は受け入れられません。
- さらに、accept ALL が指定されている場合、その規則は受け入れられます。
- 上記以外の場合、その規則は受け入れられません。

文書タイプが受け入れられる場合、適用された規則に指定されている *doctype* 値が使用されます。この値は、構文解析されたコンテンツに加えて、コレクション・パーサーに戻されます。文書タイプが受け入れられない場合に起こる動作は、以下の通りです。

- *type* のネイティブ規則 (native DEFAULT が指定されている場合のデフォルトの構文解析規則を含む) がある場合、文書は、この規則で指定された *doctype* トークンの値に加えて組み込みパーサーに戻されます。*doctype* の値は、txt、htm、または xml のいずれかでなければなりません。それらは、それぞれ、プレーン・テキスト、HTML、または XML を示します。
- 上記以外の場合、文書はリジェクトされ、構文解析されません。

## 手順

文書タイプを Stellent セッションと関連付けるには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. ES\_NODE\_ROOT/master\_config/collection\_ID.stellent/stellent.properties ファイルを編集します。ここで、*collection\_ID* は、構成したいコレクションを示します。
3. OutsideInSupportedTypes プロパティに、作成している構成ファイルの絶対パスを指定します。

例えば、1 つのコレクションについて以下の構成ファイルを作成し、それを他のコレクション固有ファイルと一緒に保管できます。

```
ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```

また、以下の構成ファイルを作成すると、同じ設定をすべてのコレクションで使用し、このファイルを他のシステム・レベルのファイルと一緒に保管することもできます。(この方法の場合、ステップ 2 に指定されているとおり、各コレクションの *stellent.properties* ファイルに、必ず、このパスを指定してください。)

```
ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

4. テキスト・エディターを使用して構成ファイルを作成し、Stellent 構文解析規則を指定したら、そのファイルを保管して終了します。
5. 変更を有効にするために、エンタープライズ・サーチ管理コンソールを使用してコレクションのパーサーをモニターし、パーサーを停止して、再始動します。

## 例

以下の構成ファイルでは、Stellent セッションは、デフォルトのリストのほかに、Microsoft Visio フォーマットを受け入れます。

```
accept DEFAULT
accept FI_VISIO3 visio
accept FI_VISIO4 visio
accept FI_VISIO5 visio
accept FI_VISIO6 visio
```

以下の構成ファイルでは、Postscript 文書がサポートされるため、ps という文書タイプで検索が可能です。X ピクスマップ・フォーマット (XPM) が組み込みテキスト・パーサーに戻され、PNG イメージ・フォーマットはリジェクトされます。他のファイル・タイプはすべて受け入れられ、other という文書タイプで検索が可能になります。

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
accept ALL other
reject FI_PNG
```

## Stellent セッション用デフォルト構文解析規則

ファイル・タイプを Stellant セッション文書フィルターにマッピングする構成ファイルを作成しない場合、文書の構文解析にはデフォルトの規則が使用されます。

Stellent セッションで使用されるデフォルトの規則は以下のようになります。

```
ACCEPT FI_WORD4          doc
ACCEPT FI_WORD5          doc
ACCEPT FI_RTF            rtf
ACCEPT FI_WINWORD1       doc
ACCEPT FI_WINWORD1COMPLEX doc
ACCEPT FI_WINWORD2       doc
ACCEPT FI_WORD6          doc
ACCEPT FI_WINWORD6       doc
ACCEPT FI_ICHITAR03      jxw
ACCEPT FI_ICHITAR04      jsw
ACCEPT FI_WINWORD1J      doc
ACCEPT FI_WINWORD5J      doc
ACCEPT FI_RTFJ           rtf
ACCEPT FI_WINWORD7       doc
ACCEPT FI_WORDPRO        lwp
ACCEPT FI_WINWORD97      doc
ACCEPT FI_ICHITAR08      jtd
ACCEPT FI_WORDPRO97      lwp
ACCEPT FI_WINWORD2000    doc
ACCEPT FI_WINWORD2002    doc
ACCEPT FI_WINWORD2003    doc
ACCEPT FI_123R1          123
ACCEPT FI_123R2          123
ACCEPT FI_123R3          123
ACCEPT FI_EXCEL          xls
ACCEPT FI_EXCEL3         xls
ACCEPT FI_EXCEL4         xls
ACCEPT FI_123R4          123
ACCEPT FI_EXCEL5         xls
ACCEPT FI_123R6          123
ACCEPT FI_EXCEL97        xls
ACCEPT FI_123R9          123
ACCEPT FI_EXCEL2000      xls
ACCEPT FI_EXCEL2002      xls
ACCEPT FI_EXCEL2003      xls
ACCEPT FI_FREELANCE      prz
ACCEPT FI_POWERPOINT4    ppt
ACCEPT FI_POWERPOINT3    ppt
ACCEPT FI_POWERPOINT7    ppt
```

ACCEPT FI_FREELANCE3	prz
ACCEPT FI_POWERPOINTMAC3	ppt
ACCEPT FI_POWERPOINTMAC4	ppt
ACCEPT FI_PDF	pdf
ACCEPT FI_EXTPOWERPOINT4	ppt
ACCEPT FI_EXTPOWERPOINTMAC4	ppt
ACCEPT FI_POWERPOINTMACB3	ppt
ACCEPT FI_POWERPOINTMACB4	ppt
ACCEPT FI_POWERPOINT97	ppt
ACCEPT FI_PDFMACBIN	pdf
ACCEPT FI_POWERPOINT9597	ppt
ACCEPT FI_POWERPOINT2000	ppt
ACCEPT FI_POWERPOINT2	ppt
NATIVE FI_HTML	htm
NATIVE FI_HTML_LATIN2	htm
NATIVE FI_HTML_JAPANESESJIS	htm
NATIVE FI_HTML_JAPANESEEUC	htm
NATIVE FI_HTML_CHINESEBIG5	htm
NATIVE FI_HTML_CHINESEEUC	htm
NATIVE FI_HTML_CHINESEGB	htm
NATIVE FI_HTML_KOREANHANGUL	htm
NATIVE FI_HTML_CYRILLIC1251	htm
NATIVE FI_HTML_CYRILLICKO18	htm
NATIVE FI_CYRILLIC1251	txt
NATIVE FI_CYRILLICKO18	txt
NATIVE FI_W2KHTML	htm
NATIVE FI_XL2KHTML	htm
NATIVE FI_PP2KHTML	htm
NATIVE FI_XML	xml
NATIVE FI_WML	xml
NATIVE FI_HTML_JAPANESEJIS	htm
NATIVE FI_WML_CHINESEBIG5	xml
NATIVE FI_WML_CHINESEEUC	xml
NATIVE FI_WML_CHINESEGB	xml
NATIVE FI_WML_CYRILLIC1251	xml
NATIVE FI_WML_CYRILLICKO18	xml
NATIVE FI_WML_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESESJIS	xml
NATIVE FI_WML_JAPANESEEUC	xml
NATIVE FI_WML_KOREANHANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_HTMLUNICODE	htm
NATIVE FI_XML_DOCTYPE_HTML	htm
NATIVE FI_XHTML	htm
NATIVE FI_ASCII	txt
NATIVE FI_ANSI	txt
NATIVE FI_UNICODE	txt
NATIVE FI_ASCII8	txt
NATIVE FI_ANSI8	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt
NATIVE FI_SHIFTJIS	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_LATIN2	txt
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_HEBREW_OLDCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_7BITTEXT	txt
NATIVE FI_JAPANESE_JIS	txt



NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_UTF8	txt
NATIVE FI_EBCDIC_37	txt
NATIVE FI_EBCDIC_273	txt
NATIVE FI_EBCDIC_277	txt
NATIVE FI_EBCDIC_278	txt
NATIVE FI_EBCDIC_280	txt
NATIVE FI_EBCDIC_284	txt
NATIVE FI_EBCDIC_285	txt
NATIVE FI_EBCDIC_297	txt
NATIVE FI_EBCDIC_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_EBCDIC_1026	txt



---

## エンタープライズ・サーチ索引の管理

ユーザーが常に最新の情報にアクセスできるようにするため、エンタープライズ・サーチでは、各コレクションに索引を作成し、定期的にその内容をリフレッシュ、再編成して保守します。

クローラーが収集したデータを検索可能にするためには、索引を作成する必要があります。コレクションを最初に作成すると、エンタープライズ・サーチによって、当初クロールされたすべてのデータの索引が作成されます。クローラーが新規または変更されたデータ・ソースをクロールすると、エンタープライズ・サーチはその新規のコンテンツで索引をリフレッシュします。そして、リフレッシュされたコンテンツは、最終的にベースとなる索引にマージされる必要があります。このマージ・プロセスを再編成と言います。索引がリフレッシュまたは再編成されると、その新規のコンテンツが検索サーバーにコピーされ、検索可能にされます。

クローラーは、データ収集を連続的に、または定期的なスケジュールに基づいて行います。索引を頻繁にリフレッシュすれば、ユーザーが最新のデータを検索できるようになります。継続的にリフレッシュされている索引は、最終的に再編成する必要があります。リフレッシュされた索引が大きくなるにつれて、システム・リソースの消費が増大します。したがって、最適なパフォーマンスを維持するには、定期的に索引を再編成する必要があります。

再編成する頻度は、以下のような条件によって異なります。

- システム・リソース (ファイル・システム・スペース、プロセッサの速度、およびメモリ)
- クロールおよび再クロールの必要がある文書数
- クロール対象のデータ・タイプ
- カテゴリ規則の変更回数 (変更は、索引を再編成しないと有効になりません)
- クローラーをスケジュールされていた時間に実行するのではなく強制始動した回数
- 外部クローラーが URI を除去または追加した回数 (外部クローラーはデータ・リッスナー API によってエンタープライズ・サーチと通信します)

コレクションの文書が数百万あり、ほとんどが Web 文書で作成されている場合、索引の再編成を 1 日に 1 回、リフレッシュを 1、2 時間おきに行うのがその目安となります。

現行の、検索可能な索引を保守するには、次の作業を行います。

- 索引のリフレッシュおよび再編成のスケジュールを指定する
- 索引スケジュールを変更する
- 索引スケジュールを使用可能/使用不可に設定する
- 索引の並行作成を構成する

索引のユーザーのビューに影響するオプションを指定するには、次の作業を行います。

- 照会でのワイルドカード文字のサポートを構成する
- ユーザーが検索できる文書の範囲を制限する有効範囲を構成する
- 検索結果で同じソースからの文書を縮小表示する
- 索引から URI を除去する

#### 関連タスク

273 ページの『コレクションの索引アクティビティのモニター』

作成中の索引の進行を知る必要がある場合、索引スケジュールを使用可能/使用不可にする必要がある場合、または、索引付けアクティビティの開始/停止を行なう必要がある場合に、コレクションの索引をモニターします。

274 ページの『エンタープライズ・サーチの索引キューのモニター』

索引キュー内のすべての索引作成の状況の表示、作成中の索引の停止、またはキューからの索引の削除を行うことができます。

---

## 索引作成のスケジューリング

索引を再編成するスケジュールと、索引を新しいコンテンツでリフレッシュするスケジュールを指定できます。

### 始める前に

索引作成をスケジュールするには、エンタープライズ・サーチ管理者役割のメンバーであるか、コレクションのコレクション管理者である必要があります。

### このタスクについて

ユーザーが検索するソースで常に最新情報にアクセスできるようにするため、索引の再編成とリフレッシュが定期的に行われるようにスケジュールしてください。索引が再編成される際には、索引全体が再作成されます。索引作成プロセスは、クローラーが収集し、パーサーが分析したすべてのデータを読み込みます。索引がリフレッシュされる際には、索引が最後に再編成されて以降にクロールされた情報が検索可能にされます。

索引作成をスケジュールするオプションは、デフォルトで選択されます。このオプションが選択されていると、スケジューラー・プロセスは、エンタープライズ・サーチ・システムが開始されるたびに索引の再編成とリフレッシュを実行するよう作業をスケジュールします。必要であればいつでも「システム始動時に有効化」チェック・ボックスをクリアして、スケジュール済みの索引作成が実行されないようにできます。例えば、問題をトラブルシューティングするときなどに、スケジュールを無効にしたいことがあります。

### 手順

索引作成をスケジュールするには、次のようにします。

1. コレクションを編集して「索引」ページを選択し、「索引作成のスケジュール」をクリックします。
2. 索引が新しいコンテンツでリフレッシュされる頻度を指定するには、「索引のリフレッシュ・スケジュールの指定」領域の「索引作成のスケジュール」ページで以下のオプションを指定します。

- a. 「開始日」領域の、「年」、「月」、「日」、「時間」「分」フィールドで、索引が最初にリフレッシュされる日時を指定します。
- b. 「更新間隔」領域の、「日」、「時間」、および「分」フィールドで、索引がリフレッシュされる頻度を指定します。

通常、索引は、1 時間または 2 時間ごとのように、頻繁にリフレッシュするべきです。ソース・コンテンツの変更頻度に応じて、間隔の長短を調整してください。例えば、毎時間 (0 日と 1 時間)、あるいは 12 時間ごと (0 日と 12 時間) などと指定します。

3. 索引が完全に再作成される頻度を指定するには、「索引の再編成スケジュールの指定」領域の以下のオプションを指定します。
  - a. 「開始日」領域の、「年」、「月」、「日」、「時間」「分」フィールドで、索引が最初に再編成される日時を指定します。
  - b. 「更新間隔」領域の、「日」、「時間」、および「分」フィールドで、索引が再編成される頻度を指定します。

通常、索引は、24 時間ごとのように、定期的に再編成するべきです。ソース・コンテンツの変更頻度に応じて、間隔の長短を調整してください。例えば、12 時間ごと (0 日と 12 時間)、あるいは 2.5 日間ごと (2 日と 12 時間) などと指定します。

4. 「OK」をクリックします。

## 索引スケジュールの変更

索引の再編成またはリフレッシュのスケジュールを変更することができます。

### 始める前に

索引スケジュールを変更するには、操作者がエンタープライズ・サーチ管理者の役割のメンバーか、またはそのコレクションのコレクション管理者である必要があります。

### 手順

索引スケジュールを変更するには、次のようにします。

1. コレクションを編集して「索引」ページを選択し、「月」、「日」、「年」、および「時」フィールドで該当する値を変更します。索引が新しいコンテンツでリフレッシュされる頻度と、索引が再編成される頻度を指定します。
2. 「適用」をクリックします。

## 索引スケジュールの使用可能および使用不可設定

索引のリフレッシュおよび再編成のスケジュールを使用可能または使用不可に設定できます。

### 始める前に

索引スケジュールを使用可能または使用不可にするには、操作者がエンタープライズ・サーチ管理者の役割のメンバーか、またはそのコレクションのコレクション管理者である必要があります。

## このタスクについて

スケジュール済みの索引作成を実行しないようにする必要がある場合、索引スケジュールを使用不可にすることができます。例えば、問題が発生した場合に、トラブルシューティングを行うために、スケジュールされた日時に索引作成が行われないようにそのスケジュールを使用不可にすることができます。

スケジュールの使用可能および使用不可設定は、コレクションの編集時、およびコレクションのモニター時に行うことができます。

### 手順

1. コレクションの編集で索引スケジュールを使用可能または使用不可にするには、次のステップに従います。
  - a. 変更するコレクションを編集します。
  - b. 「索引」 ページで、「**システム始動時に有効化**」チェック・ボックスを選択、または選択解除すると、索引のリフレッシュのスケジュールが使用可能または使用不可になります。
  - c. 索引の再編成のスケジュールを使用可能または使用不可にする場合も、同様に「**システム始動時に有効化**」チェック・ボックスを選択、または選択解除します。
  - d. 「**適用**」をクリックします。
2. コレクションのモニターで索引スケジュールを使用可能または使用不可にするには、次のステップに従います。
  - a. 変更するコレクションをモニターします。
  - b. 「索引」 ページで、索引がスケジュール済みの場合にスケジュールされた日時に索引が作成されないようにするには、「 **スケジュールを使用不可にする**」をクリックします。これで、索引は、スケジュールを使用可能にするか、「 **開始**」をクリックして索引作成プロセスを開始するまで作成されなくなります。
  - c. 索引がスケジュール済みであるが、索引作成のスケジュールが使用不可にされている場合、「 **スケジュールを使用可能にする**」をクリックします。

索引は、索引スケジュールで指定した日時に作成されるようにキューに入れます。

## 索引の並行作成の構成

索引作成要求の同時処理が可能なコレクションの数を指定することによって、索引作成リソースの使用を制御します。十分なシステム・リソースがある場合、コレクションの索引の再編成とリフレッシュが同時に実行されるように設定することにより、検索の質を高めることができます。

### 始める前に

システムの索引作成オプションを指定するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

## このタスクについて




エンタープライズ・サーチは、複数のコレクションでリソースを共有することによって複数の索引を同時に作成でき、それによって、複数のコレクションの索引作成要求が並行して処理されることが可能になります。プロセスを共有すると、大規模な索引の再編成で、キューで作成を待っている他の索引のオペラビリティが妨害されません。

索引作成が要求またはスケジュールされると、索引キューに入れられて処理される順番を待ちます。各コレクションには独自の索引があるので、さまざまなコレクションからの複数の索引作成要求が同時に索引キューにある場合があります。システムの索引オプションを構成するとき、いくつかのコレクションが索引作成用リソースを共有して要求を同時に処理させることができるのかを指定します。

また、コレクションの索引をリフレッシュする要求が、そのコレクションの索引が再編成されると同時に処理されるかどうかも指定できます。このオプションを使用可能に設定すると、実行に時間のかかる索引再編成が処理されている間に、検索サーバーは最新文書でリフレッシュされます (索引リフレッシュから)。ただし、索引作成は、リソース集約的なプロセスです。索引作成中には、大量のシステム・メモリーとディスク・スペースが消費されます。ディスク・スペースまたはメモリーが不十分な場合にこのオプションを使用可能に設定すると、システムの全体的パフォーマンスが低下することがあります。

## 手順

システムの索引作成オプションを指定するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「 編集」をクリックして、システム編集ビューに移動します。
3. 「索引」ページで、「索引作成オプションの構成」をクリックします。
4. 「システム・レベルの索引付けオプション」ページで、システム・リソースを共有し、同時に索引作成要求を処理することが可能なコレクションの数を入力します。

索引作成用リソースを共有するコレクションの数は、エンタープライズ・サーチ・システム内のコレクション数より大きくすることはできません。例えば、コレクションが 5 個ある場合、5 以下の数を入力する必要があります。

5. 個々のコレクションの索引作成を同時に複数サポートできる十分なシステム・リソースがある場合、索引のリフレッシュと再編成が同時に実行されるようにするオプションを選択できます。
6. 「OK」をクリックします。

---

## 索引の検索可能ビューに影響するオプション

文書の索引作成が終わった後、ユーザーが文書を検索したり検索結果内の文書を表示する方法を制御するオプションを指定できます。

索引に関するユーザーのビューを制御するオプションを指定するために、以下の作業を実行できます。

- 照会用語内でのワイルドカード文字のサポートを構成する。ワイルドカード照会のサポートを索引に構築するか、照会処理の間に照会用語を展開するオプションを指定できます。
- ユーザーが検索できる文書の範囲を制限する有効範囲を構成する。ユーザーがコレクションを検索するとき、ユーザーは索引全体ではなく、有効範囲にある文書のみを検索します。
- 検索結果で同じソースからの文書を縮小表示する。索引の URI あるいは URI パターンと一致する文書をグループ化して、検索結果に先頭の結果文書のみを表示することができます (ユーザーは、結果文書を縮小表示するオプションを指定できます)。
- 索引から URI を除去する。一時的に、ユーザーが索引の特定文書を検索できないようにする必要がある場合があります。

## 文書の検索に対する索引付きオプション

クロールされたデータまたは外部ソースを検索するオプションを構成するとき、または、XML および HTML メタデータ・エレメントを検索フィールドにマップするときに、文書がどのように検索でき、検索結果に表示されるのかを指定します。

フィールドを含んでいるクローラー・オプションを編集するとき、フィールドが検索可能かどうか、どのように検索できるのか、検索結果に戻されるかどうかを、以下のオプションを指定することによって制御できます。

- フリー・テキスト検索
- フィールド検索
- パラメトリック検索
- 検索結果
- 文書コンテンツ

外部ソースに関するオプションを編集するときには、フィールドを文書コンテンツ・フィールドであるとマーク付けするオプションは選択できません。

パーサーを構成し、XML エレメントおよび HTML メタデータ・エレメントを索引中のフィールドにマップすることを指定する場合、それらのフィールドがフィールド検索、パラメトリック検索、または検索結果で使用可能かどうかを指定します。

指定するオプションは、文書と共に索引に保管されます。それらのオプションを使用して、検索結果を照会できるユーザーおよび検索結果を表示できるユーザーを制限することができます。

### フリー・テキスト検索

エンタープライズ・サーチ索引は、さまざまなデータ・ソースからのコンテンツで構成されるフルテキスト索引です。単純な照会を自然言語で指定することによって、コンテンツを検索できます。検索プロセスは、フィールドおよび文書コンテンツを検索し、照会に関連する文書を検出します。

**例:**

フリー・テキスト検索は、次の例に示すように単純な照会です。

bicycle chain

文書に含まれるべき語、含まれてはならない語を示したい場合、特殊な表記を入れればそのような指定が可能です。例えば、正符号 (+) を単語の前に置くと、文書にその語と一致する語がなければならないことを指定します。また、負符号 (-) を単語の前に置くと、その語を含む文書を検索結果から除外します。複数の語を引用符 (") で囲むと、その句全体に正確に一致するものを対象として検索が行われます。

**例:**

次のフリー・テキスト照会では、検索一致となるのは文書に「science fiction」に完全一致する句が含まれ、「robot」の語を含まない場合に限られます。

```
+ "science fiction" -robot
```

## フィールド検索

フィールド検索を使用すると、照会対象を文書の特定のフィールドおよびメタデータに制限することができます。例えば、文書のタイトルに特定の語がなければならないといった指定をすることができます。

エンタープライズ・サーチでフィールド検索を指定するには、照会にフィールド名とそのフィールドになければならない語句とを指定します。

**例:**

次の照会では、タイトル・フィールドに「ibm」という語と「enterprise search」という句を含む文書を検索します。

```
title:ibm title:"enterprise search"
```

フィールドをフィールド名で検索するには、クローラーの構成時にそのフィールドをフィールド検索使用可能に設定する必要があります。

## パラメトリック検索

パラメトリック検索は、フィールド検索の一種で、数値フィールド、日付フィールド、およびメタデータに対して比較、評価を行う照会を指定できます。例えば、一定サイズの文書、または一定の日付以降に書かれた文書などを検索することができます。また、属性値が指定した値より大きい、小さい、または等しいなどの条件に当てはまる文書を検索することもできます。

**例 1:**

次の照会では、価格がちょうど 50 ドル (ないし price フィールドに指標付けられている任意の通貨単位) の品目を検索します。

```
#price::=50
```

**例 2:**

次の照会では、ファイル・サイズが 1024 より大きく 2048 以下の文書を検索します。

```
#filesize::>1024<=2048
```

フィールドをパラメトリック照会で検索するには、クローラーの構成時にそのフィールドをパラメトリック検索使用可能に設定する必要があります。

## 検索結果

検索対象にはするが検索結果には表示させたくないフィールドや、照会を行わないが検索結果には表示させたいフィールドがある場合があります。例として、財務データを照会して意味のあるレポートを取得する必要があるが、結果には従業員の給与と名前を表示する必要がある場合などが考えられます。

## 文書コンテンツ

文書コンテンツ・フィールドであるとマークされたフィールドの内容は、文書をモデル・ベースのカテゴリ、および文書コンテンツに基づく規則を指定するカテゴリと関連付けるのに使用できます。

### 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『照会構文』

189 ページの『エンタープライズ・サーチの検索アプリケーション』

検索アプリケーションは、エンタープライズ・サーチ・システム内のコレクションおよび外部ソースの検索を可能にします。検索アプリケーションはいくつでも作成できます。また、1 つの検索アプリケーションで、任意の数のコレクションおよび外部ソースを検索できます。

## 照会内のワイルドカード文字

ユーザーが照会用語にワイルドカード文字を含め、指定のパターンと一致するワードを検索できるようにすることができます。

ワイルドカード照会用語は、アスタリスク (\*) が含まれている用語です。ワイルドカード文字が含まれている照会をユーザーが実行依頼した場合、検索結果には、その照会用語に一致する索引内のすべての文書のほか、ワイルドカード文字で表されるパターンに一致する索引内のすべての文書が含まれます。例えば、照会用語 `sea*` の末尾ワイルドカードは、`search`、`season`、および `seals` と一致が可能です。

索引に対してワイルドカード文字オプションを構成する場合、ユーザーが照会でワイルドカード文字を指定できるようにするかどうか、また、その場合、どのようにこのサポートが提供されるかを選択します。

- 文書のすべての部分に対して、ワイルドカード文字パターンを突き合わせた、ワードでの検索を使用可能にできたり、あるいはフィールドに対するパターン・マッチングを制限できたりします。
- すべてのフィールドがワイルドカード文字を含む照会をサポートすることができるようにしたり、あるいは指定したフィールドに対するパターン・マッチングを制限できたりします。

- 照会用語の最終文字にワイルドカード文字を制限できたり (末尾ワイルドカード文字)、あるいはワイルドカード文字を照会用語のいずれの場所にも出現させることができたりします。(ワイルドカード文字は、フィールド名に出現させることはできません。)
- ワイルドカード文字が出現することを、どこに許可するかによって、どのように照会用語を展開するかを選択できます (ワイルドカード文字を含む照会用語は、マッチングする索引の用語のすべてに対して展開されます)。索引は、すべての可能な用語の展開を保管できたり、あるいは検索プロセスは、照会処理中に用語を展開できたりします。

ワイルドカード文字設定に行う変更はいずれも、索引が再編成される次回に有効になります。

## 索引展開

索引内の用語の展開を含めるには、出現するマッチングに対する照会用語のワイルドカード文字パターンに、ワード内のいくつかの先行文字がマッチングするかを指定します。少なくともこの数の文字をもつ照会用語のみ (\* を除く)、結果を返します。例えば、4 を指定すると、照会用語は、マッチングとして出現するために、最小で 4 文字を指定する必要があります。

4 を指定すると、ワード `technology` は、照会用語の `tech*` および照会用語 `techno*` とマッチングしますが、照会用語 `te*` とはマッチングしません。

索引が更新または再編成されると、文書内の各用語の考えられるすべての展開に、元の用語に加えて索引付けされます。この方法は、照会処理中に用語を展開するのに追加の時間は必要でないという利点があります。ただし、この方法の場合、索引のサイズが大きくなります。つまり、大きい索引に対応できるような十分なシステム・リソースが使用可能でなければなりません。

これは、コレクションが相対的に小さい場合、または索引を作成するためのスペースと時間よりも照会応答時間を重視する場合に、とても役に立つ方法です。例えば、カタログまたは従業員ディレクトリーを検索するのに、この方法を選択できます。

このアプローチは、末尾ワイルドカード文字のサポートを使用可能にした場合にのみ有効です。照会用語にいずれの場所にも出現するワイルドカード文字のサポートを使用可能にした場合、索引内に用語の展開を含めるオプションを選択することはできません。

## 照会展開

ワイルドカード文字が含まれている照会をユーザーが実行依頼したときに照会を展開し、パターン・マッチング規則を適用するには、一致を構成する照会用語のバリエーションの数を指定します。例えば、50 を指定した場合、照会用語の最大 50 のバリエーションを、その照会用語の一致とみなすことができます。

この例を示すと、照会用語 `tech*` はワード `technical`、`technique`、`technology` と、さらに、文字 `tech` で始まる最大 50 個の異なるワードに一致します。



照会展開により、索引のサイズはあまり影響を受けませんが、照会パフォーマンスは低下する可能性があります。検索プロセスは、ワイルドカード文字設定に指定した限度まで、ワイルドカード照会用語の考えられるすべての展開に対して繰り返す必要があります。

この方法は、コレクションが相対的に大きく、しかも、索引を作成するためのスペースと時間が最小でなければならない場合に、とても役立ちます。例えば、E メール・リポジトリ用にこの方法を選択します。その場合、索引は、急速に変化する文書に応じて増加する必要がありますが、照会応答時間はあまり重要ではありません。

このアプローチは、末尾ワイルドカードのサポートを使用可能にしているか、または照会用語のいずれの場所にも出現するワイルドカード文字のサポートを使用可能にしているかどうかにかかわらず有効です。

## 照会でのワイルドカード文字のサポート

ワイルドカード照会用語の展開のセットには、ワイルドカード文字を任意の文字のシーケンスと置換することによって取得できる索引内のすべての用語が含まれます。そのセットは、次のように判断されます。

- コレクションが、照会用語のいずれの場所でも出現できるワイルドカード文字をサポートする場合、アスタリスクを含むいずれの照会用語も、ワイルドカード用語と解釈されます。
- そのセットには、最大限、エンタープライズ・サーチ管理者が構成する展開の最大数が含まれます。索引に、この数を超過が含まれている場合、無視されます。(検索結果は、ワイルドカード展開が無視されたかどうかを示します。)
- ワイルドカード文字サポートが、あるセットのフィールドに制限される場合、そのセットには、指定されたフィールドの 1 つに表示される用語のみが含まれます。用語は、少なくとも、索引内の 1 つの文書の 1 つのフィールドにのみ表示される必要があります。
- 照会用語が「フィールド用語」である場合、ワイルドカード文字は、フィールド指定子の後に表示される必要があります (例えば、`fieldname:*sphere`)。フィールド名にコロン (:) を含めることはできません。
- ワイルドカード文字サポートが、あるセットのフィールドに制限される場合、ワイルドカード照会用語のフィールド名は、エンタープライズ・サーチ管理者コンソールで指定されたフィールドの 1 つである必要があります。そうでないと、用語に対する展開を 1 つも検出しません。
- ワイルドカード文字は、プレーン・テキスト用語でのみサポートされ、XML エレメント名、属性名、または属性値ではサポートされません。ワイルドカード文字でのみ構成される用語はサポートされません。

## 照会でのワイルドカード文字のオプションの構成

エンタープライズ・サーチ・コレクションの索引作成オプションを構成するときに、ユーザーが照会用語にワイルドカード文字を含められるようにしたいかどうかを指定できます。

始める前に



ワイルドカード文字のオプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、その索引が所属するコレクションのコレクション管理者である必要があります。

### このタスクについて

ワイルドカード文字のオプションを指定した場合、変更は、索引が再編成される次回に有効になります。

### 手順

照会でのワイルドカード文字のサポートを構成するには、次のようにします。

1. コレクションを編集、「索引」ページを選択、そして「**ワイルドカード文字のオプションの構成**」をクリックします。
2. 「ワイルドカード文字のオプションの構成」ページで、「**照会でのワイルドカード文字のサポート**」チェック・ボックスを選択します。
3. 「**フィールドを検索しない照会でのワイルドカード文字のサポート**」チェック・ボックスを使用して、フィールドにないテキストを検索する照会でワイルドカード文字がサポートされるかどうかを指定します。例えば、名前付きフィールドを検索しないフリー・テキスト照会 `tech*` は、このチェック・ボックスが選択されている場合のみ、展開された結果（「`technology`」または「`technique`」など）を戻します。
4. ワイルドカード文字をサポートするフィールドの指定:
  - ワイルドカード文字を含む照会をサポートする文書のフィールドすべてを使用可能にするには、「**すべてのフィールド**」を選択します。
  - ワイルドカード文字のサポートを幾つかのフィールドに制限するには、「**特定のフィールド**」を選択して、フィールド名を入力します。展開された結果は、指定したフィールドに対してのみ戻されます。例えば、照会 `author:john*` は、「`author`」フィールドがワイルドカード文字をサポートするように指定した場合にのみ、展開された結果を戻します。
5. ワイルドカード文字が照会用語の最終位置に出現しなければならないかどうか（末尾ワイルドカード）、またはワイルドカード文字が非限定であり、照会用語のいずれの場所にも出現できるかどうかを指定します。

ワイルドカードの位置とタイプを選択する場合、どのようにワイルドカード文字のサポートを使用可能にしたいかをもまた指定する必要があります。詳細については、ワイルドカード文字のオプションを構成しているときに、「ヘルプ」をクリックしてください。

6. 「**OK**」をクリックします。

## 有効範囲

ユーザーにコレクションの限定ビューを提供するには、有効範囲を構成します。

有効範囲とは、索引内の関連した URI のグループのことです。有効範囲を構成することで、コレクションの中のユーザーが見ることができる文書を制限します。ユーザーがコレクションを検索するとき、ユーザーは索引全体ではなく、有効範囲内の文書のみを検索します。この機能を使用するには、検索アプリケーションで有効範囲の検索をサポートしている必要があります。

有効範囲を作成するとき、ユーザーが検索できる索引中の URI の範囲を指定します。ユーザーが検索できる文書を制限することで、検索結果内の文書をユーザーが探している情報に特定できることとなります。

例えば、技術サポート部門の URI を含んだ有効範囲と、人事管理部門の URI を含む別の有効範囲を作成することができます。検索アプリケーションが有効範囲をサポートする場合、技術サポート部門のユーザーは、技術サポート部門有効範囲から文書を取得し、人事管理部門のユーザーは、人事管理部門有効範囲から文書を取得します。

有効範囲は必要なだけ作成できますが、多すぎるとパフォーマンスに影響を及ぼす場合があります。ほとんどの検索要求が、1 つまたは 2 つの有効範囲のみをフィルタリングする必要があるように有効範囲を構成してください。有効範囲には URI 全体または URI パターンを含めることができるので、同じ文書が複数の有効範囲に属することがあります。

有効範囲の構成時、変更を有効にするために、索引の再編成を 2 回行わなければならない場合があります。コレクションの索引の再編成を一度もしないうちに有効範囲を構成すると、ユーザーはコレクションを検索できても、検索結果に有効範囲データは表示されません。検索結果が URI の有効範囲を確実に反映するようにするため、索引をもう一度再編成してください。

有効範囲を構成する前に、少なくとも 1 度は索引を再編成していれば、次の索引の再編成後に、その変更内容が有効になります。

## 有効範囲の構成

エンタープライズ・サーチ・コレクションの有効範囲を構成するときに、ユーザーが検索を許可される索引内の文書の範囲に合うように URI または URI パターンを指定します。

### 始める前に

有効範囲を構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、その有効範囲が所属するコレクションのコレクション管理者である必要があります。

### このタスクについて

検索アプリケーションで有効範囲のサポートが使用可能に設定されている場合、ユーザーはコレクションを検索する際に、有効範囲の境界を定義する URI に一致する文書のみを検索できます。

有効範囲の構成時、変更を有効にするために、索引の再編成を 2 回行わなければならない場合があります。コレクションの索引の再編成を一度もしないうちに有効範囲を構成すると、ユーザーはコレクションを検索できても、検索結果に有効範囲データは表示されません。検索結果が URI の有効範囲を確実に反映するようにするため、索引をもう一度再編成してください。

有効範囲を構成する前に、少なくとも 1 度は索引を再編成していれば、次の索引の再編成後に、その変更内容が有効になります。

## 手順

有効範囲を構成するには、次のようにします。

1. コレクションを編集、「索引」ページを選択、そして「有効範囲の構成」をクリックします。
2. 「有効範囲」ページで、「有効範囲の作成」をクリックします。
3. 有効範囲の名前と、有効範囲の境界を定義する URI および URI パターンを指定します。有効範囲から除外したい URI および URI パターンを指定することもできます。
4. 「OK」をクリックします。

新規有効範囲が、このコレクションに属している他の有効範囲と共に「有効範囲」ページにリストされます。

### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

## 縮小表示した URI

エンタープライズ・サーチは、Web ソースおよび NNTP ソースからの文書のうち URI 接頭部が同じものがグループ化され、検索結果に縮小表示されるように、検索結果を編成できます。

結果が縮小表示される場合、先頭の結果は、通常、左揃えで表示されます。それより下のランキングの結果は、グループ化され、先頭の結果より字下げされて表示されます。

URI 接頭部が異なる結果文書を単一のグループとして縮小表示するには、作成する 1 つのグループ名に、それらの接頭部を関連付けます。例えば、財務データを管理するサーバーが 3 つある場合、3 つのサーバーすべてからの文書を検索結果においてグループ化し、先頭の結果文書の下に低いランキングの結果を縮小表示できます。

検索アプリケーションは、URI 接頭部あるいはグループ名を使用して、検索結果で文書を縮小表示できます。エンタープライズ・サーチのサンプル検索アプリケーションでは、先頭の 2 つの検索結果文書が表示されます。同じ URI 接頭部を持つ (あるいは、同じ URI グループに属す) 3 つ以上の結果文書が返された場合、結果を縮小表示させるオプションを選択できます。

ユーザーは、エンタープライズ・サーチ照会構文 (`samegroupas:URI prefix`) を使用して、指定された URI 接頭部と同じグループ内のすべての文書を検索できます。

### URI 接頭部とグループ名の編成方法

管理コンソールを使用して検索結果の縮小表示規則を構成するときに、縮小表示する文書の URI 接頭部を指定します。また、オプションで、それらの URI 接頭部を 1 つのグループ名に関連付けることができます。

設定する URI 接頭部の順序は重要です。索引サーバーは、コレクションの中の各 URI の値を計算するときに、URI 接頭部の順序を使用します。それぞれの URI ごとに以下ようになります。

1. 索引サーバーは、検索結果の縮小表示規則にある URI 接頭部を順番にスキャンします。
2. 索引サーバーは、索引内の文書と接頭部が一致する最初の URI 接頭部を検出すると、グループ名 (あるいは、規則でグループ名を指定していない場合は、URI 接頭部) をその文書の追加の検索語として関連付けます。
3. 文書が URI 接頭部と一致しない場合には、次のようにします。
  - Web URI の場合、索引サーバーは、URL のホスト名を URI 接頭部として使用します。
  - NNTP URI の場合、索引サーバーは、参照ヘッダーの値にある最初のメッセージ ID を URI 接頭部として使用します。

検索結果で縮小表示する URI 接頭部のリストに URI 接頭部を追加した後、索引サーバーがその URI 接頭部をスキャンして、追加の検索語として索引内の文書に関連付ける際の望ましい順序にそれを配置してください。

- URI 接頭部を追加しても、それをグループ名に関連付けないときは、個別の URI 接頭部を選択して、リスト内でそれを上下に移動させます。
- URI 接頭部を追加してそれをグループ名に関連付ける場合に、リスト内で URI 接頭部を上下に移動させるときはいつでも、同じグループに属している URI 接頭部のグループ全体を移動させます。(グループ内の URI 接頭部の順序は問題ではありません。個々の URI 接頭部を選択すれば、自動的にグループ全体が選択されます。)

## 検索結果内の URI の縮小表示

同じ URI 接頭部をもつ Web および NNTP ソースからの検索結果文書をグループ化および縮小表示するオプションが指定できます。異なる URI 接頭部をもつ結果文書をまとめて縮小表示できるようにするグループ名も作成できます。

### 始める前に

検索結果を縮小表示するオプションを指定するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて



検索結果を縮小表示するために行った変更は、次に索引が再編成されるまで、有効になりません。

### 手順

検索結果を縮小表示するためのオプションを指定するには、次のようにします。

1. コレクションを編集、「索引」ページを選択、そして「縮小表示した検索結果」をクリックします。
2. 「縮小表示した検索結果」ページで、「URI 接頭部の追加」をクリックします。

3. 「結果を縮小するための URI 接頭部の追加」 ページで、検索結果で縮小表示したい文書の URI接頭部を入力します。 例:  

```
http://finance/ROI/  
http://server1.com/finance/
```
4. この URI 接頭部と関連付けたい説明的なグループ名を入力できます。 いくつかのソースからの結果文書を単一グループとして縮小表示するために、各 URI 接頭部を追加するときに同じグループ名を入力してください。
5. 「OK」をクリックします。
6. 「縮小表示した検索結果」 ページで、新しい規則を、索引サーバーにスキャンさせる順序で並べます。
  - URI 接頭部を追加し、それをグループ名と関連付けていない場合、新しい URI 接頭部は、リストの一番下に示されます。矢印キーを使用して、新しい接頭部を正しい位置に移動します。
  - 新しい URI 接頭部をグループ名と関連付けた場合、その URI 接頭部は、同じグループに所属する URI 接頭部のセットの一番下に示されます。矢印キーを使用して、URI 接頭部のグループ全体を正しい位置に移動します。
7. URI 接頭部またはグループ名を変更するには、目的の URI 接頭部を選択して、 「編集」をクリックします。
8. リストから URI 接頭部を除去するには、目的の URI 接頭部を選択して、 「除去」をクリックします。

## 索引から URI を除去

コレクション内の文書をユーザーが検索できないようにするため、それらの文書の URI を索引から除去することができます。

### 始める前に

索引から URI を除去するには、エンタープライズ・サーチ管理者役割のメンバーであるか、コレクションのコレクション管理者である必要があります。

### このタスクについて

完全修飾 URI を指定すると、その URI はユーザーの検索結果に現れなくなります。ただし、ユーザーが同じ照会をサブミットし、その照会の結果文書が検索キャッシュに残っている場合には、除去した URI のキャッシュされた結果ページが検索結果に戻され続けます。索引が次にリフレッシュまたは再編成されるまで、検索キャッシュは更新されず、URI は索引から除去されません。

複数の URI を除去するようパターンを指定した場合、索引が次にリフレッシュまたは再編成されるまでは、そのパターンに一致する URI がユーザーの検索結果に戻され続けます。

索引から URI を除去しても、その URI はクロール・スペースから除去されるわけではありません。クローラーが次回にその文書をクロールすると、その URI が索引内に作成され、再び検索に使用可能になります。URI をクロール・スペースから除去するには、該当する文書を除外するようにクロール規則を更新し、クローラーをいったん停止して、再始動する必要があります。

## 手順

特定の文書の URI を索引から除去するには、次のようにします。

1. コレクションを編集、「索引」ページを選択、そして「索引から URI を除去」をクリックします。
2. 「索引から URI を除去」ページで、索引から除去したい URI (または URI パターン) を入力します。

例:

```
http://domain.org/hr/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

3. 「OK」をクリックします。

### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI)  
は、その文書をコレクションに追加したクローラーのタイプを示します。



---

## エンタープライズ・サーチの検索サーバー管理

検索サーバーに指定できるオプションとして、検索結果を戻すキャッシュ・スペースを使用する、検索結果に表示する文書サマリーの最大長を制御する、検索品質向上のためにカスタム辞書を関連付ける、照会に特定の用語が使用されたときに事前定義しておいた URI を検索結果に戻す、などがあります。

ユーザーが照会をサブミットすると、検索サーバーは索引を使用して、該当する文書を素早く探し出します。検索サーバーは、構文解析され、トークン化されたデータが保管されているエンタープライズ・サーチ・データ・ストアを使用して、該当する文書のメタデータを取得します。メタデータには、文書の URI、タイトル、説明、日付、データ・タイプなど（さらに多くのデータを格納することもできます）を入れることができます。

コレクションに対して検索サーバーを設定するときに、照会の処理方法に影響するオプションを指定することができます。これには、照会のパフォーマンスを左右する可能性のあるオプションも含まれます。

### 検索キャッシュの構成

照会のパフォーマンスを最適化するには、結果の文書をキャッシュに保管するように指定することができます。また、キャッシュに入れる文書用に割り当てるスペース量を設定することができます。

### 文書サマリーの表示最大長の設定

結果文書のほとんどで、ユーザーにとってその文書が取得したいものかどうかを見極められるように文書コンテンツのサマリーを表示します。この要約情報を表示するために、検索結果上で使用するスペースを指定することができます。

### 異なるデフォルト言語の指定

コレクションの文書を検索するためのデフォルト言語は、コレクション作成時に指定されていますが、必要に応じて別の言語を指定することができます。

### カスタム辞書の関連付け

アプリケーション開発者によって、同義語、ストップワード、ランキング調整ワードに関するカスタム辞書が作成されている場合、コレクションの検索時にその辞書を指定することができます。

### クイック・リンクの構成

一定のキーワードおよび句に対して戻す URI を事前設定できます。ユーザーが照会に該当するキーワードまたは句を指定すると、事前定義された URI が検索結果と共に戻されます。クイック・リンク URI は、検索サーバーが索引検索によって戻す URI に追加される形で戻されます。

### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』

ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。

178 ページの『カスタム・ランキング調整ワード辞書』  
検索結果の質を向上させるために、カスタム・ランキング調整ワード辞書を作成して、検索結果における文書のランクを調整することができます。

---

## 検索キャッシュ

検索サーバーの負荷が比較的高い場合、検索結果をキャッシュに入れることでパフォーマンスが向上します。

検索サーバーが検索要求を処理する場合、まず、同じ照会の結果がキャッシュにすでに存在していないかチェックされます。検索サーバーが該当する結果文書を検出できれば、検索結果をそのまますぐにユーザーに戻すことができます。該当の結果文書が見つからないと、検索サーバーは索引を検索します。

検索キャッシュが満杯になると、最も古い結果文書および頻度の低い照会の結果文書が、新たな検索結果用の場所を空けるためにキャッシュから出されます。

エンタープライズ・サーチ管理コンソールから、検索キャッシュを使用可能に設定し、また、キャッシュの容量 (結果を同時にキャッシュに入れられる照会数) を指定することができます。

検索キャッシュの設定を変更した場合は、検索サーバーを再始動して、その変更を有効にする必要があります。

---

## 検索キャッシュの構成

コレクションの検索キャッシュを使用可能または使用不可にできます。また、検索キャッシュのサイズを制御するオプションも指定できます。

### 始める前に

コレクションの検索キャッシュを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

検索キャッシュを構成するには、次のようにします。

1. コレクションを編集して「検索」ページを選択し、「**検索サーバー・オプションの構成**」をクリックします。
2. 「検索サーバー・オプション」ページで、「**検索キャッシュを使用**」チェックボックスを選択します。
3. 「**キャッシュ項目の最大数**」フィールドに、検索キャッシュが保持できる検索結果セットの最大数を入力します。
4. 「**OK**」をクリックします。
5. 変更を有効にするために、検索サーバーをモニターし、サーバー・プロセスを再始動します。

## カスタム同義語辞書

検索結果の質を高めるために、ユーザーがコレクションを検索するときに、照会用語の同義語を検索することができます。

同義語辞書を作成し、それをエンタープライズ・サーチ・システムに追加して、コレクションに関連付けると、ユーザーは、そのコレクションを検索するときに、照会用語の同義語を含む文書を検索できます。このように照会を拡張することで、ユーザーは、正確に照会用語と一致する文書だけでなく、興味のある文書をすべて検出できる可能性が高くなります。同義語辞書を作成するときに、どのワードとどのワードが互いに同義語であるかを定義するので、ユーザーがあらゆるバリエーションを考慮して照会用語を指定しなくても、関係のある文書を見つけられるようになります。

例えば、ユーザーの組織で、部門や設備などを頭字語や省略語を使用して参照している場合もありますし、ユーザーのコレクション内の文書に、業界固有の用語が含まれている場合もあります。同義語辞書を作成すれば、頭字語 (例えば ACL など) を含む照会で、その頭字語を拡張したもの (例えば、ACLs、アクセス制御リスト (access control lists)、アクセス制御 (access controls) など) を扱う文書を返すことができます。

エンタープライズ・サーチ照会言語では、照会用語の前にチルド演算子を付加して同義語を指定することをサポートしています。例えば、`~WAS` という照会では、WebSphere Application Server を扱う文書を返すことができます。また、アプリケーション開発者は、照会のプロパティによって同義語サポートを使用可能にでき、これには特別な構文は必要ありません。

同義語辞書には、ワードのさまざまな変形を含み、以下のような特性があります。

- ワードは、言語に固有なものではなく、異なる言語で使用されることがある。コレクションごとに 1 つだけ同義語辞書があります。
- ワードは、語尾変化をしない。あらゆるすべての語尾変化は、同義語リストに追加する必要があります。例えば、ワードの単数形と複数形の語尾変化の場合もあります (ACL と ACLs など)。

同義語辞書に追加する用語のほとんどは、厳密な意味で同等です。つまり、用語 A が用語 B の同義語である場合、すなわち B は A の同義語です。A が照会で使用されるたびに、B が使用され、その反対もまた真です。

ただし、ある用語の異なる用途に対応する用語を追加することもでき、これにはその用語の汎用的変形、あるいはより特定のな変形も含まれます。例えば、1 つの同義語グループは building と house を含み、もう 1 つのグループは bank、shore、と credit union を含みます。

用語と用語の関係が厳密でないほど、検索結果は大きくなり、検索結果によっては照会に関係ないものになる場合があります。検索および索引 API では、ユーザーが検索要求をサブミットしたときに、適切な同義語を選択できるようにする方法や、ユーザーに照会用語がどのような同義語に拡張されたかを示す方法もあります。

同義語辞書を作成するには、そのコレクションの主題についてのエキスパートが XML フォーマットで同義語リストを作成する (あるいは、アプリケーション開発者

と一緒に作業して XML ファイルを作成する) 必要があります。 WebSphere II OmniFind Edition で提供されるツールを使用して、その XML ファイルをバイナリー (.dic ファイル) に変換する必要があります。

エンタープライズ・サーチ管理者は、このバイナリー・ファイルをシステムにアップロードし、表示名を割り当てます。コレクション管理者は、コレクションに検索サーバー・オプションを設定するときに、コレクション内の文書の検索に使用する同義語辞書を選択できます。

**制約事項:** カスタム同義語辞書をシステムに追加後は、それを編集することができません。コレクションに使用可能な同義語を修正するには、以下のようになります。

1. ソース XML ファイルを更新します。
2. XML ソースを新規ディクショナリー・ファイルに変換します。
3. 古い同義語辞書を使用するコレクションからその辞書を削除します。
4. 古い同義語辞書をシステムから削除します。
5. 新しい同義語辞書をシステムに追加します。
6. 新しい同義語辞書を使用するコレクションにその辞書を関連付けます。

#### 関連概念

「テキスト分析機能ガイド」の『検索アプリケーションの同義語サポート』

#### 関連タスク

「テキスト分析機能ガイド」の『同義語に使用できる XML ファイルの作成』

「テキスト分析機能ガイド」の『同義語辞書の作成』

## システムへの同義語辞書の追加


コレクション内で文書を検索するためにカスタム同義語辞書を作成する場合、辞書をエンタープライズ・サーチ・システムと関連付ける必要があります。コレクションの検索に使用したい同義語辞書を、後で選択できます。

#### 始める前に

エンタープライズ・サーチ照会を使用するためにカスタム同義語辞書を追加するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

#### 手順

同義語をエンタープライズ・サーチ・システムと関連付けるには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2.  「編集」をクリックして、システム編集ビューに移動します。
3. 「検索」ページで、「同義語辞書の構成」をクリックします。
4. 「同義語辞書の構成」ページで、「同義語辞書の追加」をクリックします。
5. 「同義語辞書の追加」ページで、同義語辞書の固有の表示名を入力し、オプションで、説明を入力します。

- .dic ファイルのロケーションを指定します。ファイルが 8 MB より小さい場合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より大きい場合、そのファイルは索引サーバーに置かなければなりません。
- 「OK」をクリックします。カスタム同義語辞書が、エンタープライズ・サーチ・システムに追加され、コレクションの検索に使用できるようになります。

## 同義語辞書とコレクションとの関連付け

同義語辞書がエンタープライズ・サーチ・システムと関連付けられている場合、そのうちの 1 つを選択して、コレクションの検索時に使用できます。照会用語が辞書内の用語に一致した場合、その用語の同義語が含まれている結果文書も、検索結果に入れて戻されます。

### 始める前に

コレクションの同義語辞書を選択するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

同義語辞書をコレクションと関連付けるには、次のようにします。

- コレクションを編集して「検索」ページを選択し、「**検索サーバー・オプションの構成**」をクリックします。
- 「検索サーバー・オプション」ページの「**同義語辞書名**」フィールドで、ユーザーがこのコレクションを照会するときを使用したい同義語辞書を選択します。

使用できる同義語辞書のリストに、エンタープライズ・サーチ・システムに追加されたすべての同義語辞書が含まれます。

- 「OK」をクリックします。

---

## カスタムのストップワード辞書

検索結果の質を高めるために、照会処理中に一定の語が自動的に照会用語から除去されるよう指定できます。

ストップワード辞書には、頻繁に使用されるために照会用語としては実用的でない、エンタープライズ特有の語句が含まれます。これらの語を照会から除外することによって、わずかに関連するだけのおびただしい数の結果文書をユーザーが受け取るのを防止できます (照会内の他の用語に一致する文書のみが戻されます)。照会処理中に、検索サーバーが照会からストップワードを除去します。除去される語には、カスタム辞書中のストップワードと、エンタープライズ・サーチに事前定義されたストップワード (一般的な前置詞や冠詞など) があります。

WebSphere II OmniFind Edition は、デフォルトで、言語特有のストップワード認識を実行します。このプロセスによって、頻繁に使用される一般的な語 (例えば、a や the) が照会から除去されます。カスタムのストップワード辞書を定義する必要があるのは、エンタープライズ特有またはドメイン特有のストップワードのみについてです。



ストップワード辞書を作成するには、そのコレクションの主題についてのエキスパートが XML フォーマットでストップワード・リストを作成する (あるいは、アプリケーション開発者と一緒に作業して XML ファイルを作成する) 必要があります。WebSphere II OmniFind Edition で提供されるツールを使用して、XML ファイルをバイナリー (.dic) ファイルに変換する必要があります。

エンタープライズ・サーチ管理者は、このバイナリー・ファイルをシステムにアップロードし、表示名を割り当てます。コレクション管理者は、コレクションに検索サーバー・オプションを設定するときに、コレクション内の文書の検索に使用するストップワード辞書を選択できます。

**制約事項:** カスタムのストップワード辞書をシステムに追加した後は、それを編集することができません。照会処理に使用可能なストップワードを訂正するには、以下の手順を実行する必要があります。

1. ソース XML ファイルを更新します。
2. XML ソースを新規ディクショナリー・ファイルに変換します。
3. 古いストップワード辞書を使用するコレクションからその辞書を削除します。
4. 古いストップワード辞書をシステムから削除します。
5. 新しいストップワード辞書をシステムに追加します。
6. 新しいストップワード辞書を使用するコレクションにその辞書を関連付けます。

#### 関連概念

「テキスト分析機能ガイド」の『カスタムのストップワード辞書』

#### 関連タスク

「テキスト分析機能ガイド」の『ストップワードに使用できる XML ファイルの作成』

「テキスト分析機能ガイド」の『ストップワード辞書の作成』

## システムへのストップワード辞書の追加


照会から語を除去するためのカスタムのストップワード辞書を作成する場合、それらの辞書をエンタープライズ・サーチ・システムと関連付ける必要があります。どのストップワード辞書をコレクションの検索に使用するのかは、後で選択できます。

#### 始める前に

カスタムのストップワード辞書をシステムに追加するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

#### 手順

カスタムのストップワード辞書をエンタープライズ・サーチ・システムと関連付けるには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「 編集」をクリックして、システム編集ビューに移動します。



3. 「検索」 ページで、「ストップワード辞書の構成」 をクリックします。
4. 「ストップワード辞書の構成」 ページで、「ストップワード辞書の追加」 をクリックします。
5. 「ストップワード辞書の追加」 ページで、辞書の固有の表示名を入力します。
6. .dic ファイルのロケーションを指定します。 ファイルが 8 MB より小さい場合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より大きい場合、そのファイルは索引サーバーに置かなければなりません。
7. 「OK」 をクリックします。 カスタムのストップワード辞書が、エンタープライズ・サーチ・システムに追加され、コレクションの検索に使用できるようになります。

## ストップワード辞書とコレクションとの関連付け

ストップワード辞書がエンタープライズ・サーチ・システムと関連付けられている場合、そのうちの 1 つをコレクションの検索時に使用するよう選択できます。照会用語に一致する語が辞書内にある場合、その語は処理される前に照会から除去されます。

### 始める前に

コレクションに対してストップワード辞書を選択するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

ストップワード辞書をコレクションと関連付けるには、次のようにします。

1. コレクションを編集して「検索」 ページを選択し、「検索サーバー・オプションの構成」 をクリックします。
2. 「検索サーバー・オプション」 ページの「ストップワード辞書名」 フィールドに、ユーザーがこのコレクションを照会するときを使用するようにしたいストップワード辞書の名前を入力します。

使用できるストップワード辞書のリストに、エンタープライズ・サーチ・システムに追加されたすべてのストップワード辞書が含まれます。

3. 「OK」 をクリックします。

---

## 動的要約

動的要約は、結果文書中のどの語句がユーザーの検索対象概念を最もよく表しているかを判別する技法です。

エンタープライズ・サーチで、動的要約は、文書中のさまざまな検索語を多く含む文を捕捉しようとする機能です。検索結果には、いくつかの文、または各文の一部が選択されて表示されます。対象となる検索語は、検索結果の中で、HTML レンダリングによって強調表示されます。

コレクションの検索サーバー・オプションを構成するときに、検索結果内での文書サマリーの最大表示長を指定できます。サマリーには強調表示文字が含まれるの

で、検索アプリケーションに戻されるバッファは指定の最大値より大きくなる場合があります。ただし、表示長が指定の最大値を超えることはありませんが、(ソース文書から抽出されるサマリー・データに従って) サマリーはより短くなる可能性があります。

## 管理コンソールの文書サマリーのカスタマイズ

エンタープライズ・サーチ管理コンソールで検索サーバーのオプションを指定することによって、文書サマリーに表示される情報量をカスタマイズすることができます。

### 始める前に

コレクションのサマリー表示長を制御するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

文書サマリーの最大表示長に指定する値は、各サマリーに入れることのできるセンテンスの数に指定する値と連動して機能します。文書サマリーが最も短くなる値が優先されます。

例えば、センテンス数を 4 に限定すると指定した場合、表示長では 4 つのセンテンスに含まれる文字総数より多い文字が許可されていても、文書サマリーに含まれるセンテンスの数は 4 になります。別の例として、センテンス数が 10 に制限され、表示長が 500 文字に制限されているという組み合わせを想定すると、結果のセンテンス数は 10 より少なくなることもあります。

### 手順

文書サマリーの表示長を構成するには、次のようにします。

1. コレクションを編集して「検索」ページを選択し、「**検索サーバー・オプションの構成**」をクリックします。
2. 「検索サーバー・オプション」ページで、文書サマリーの最大表示長を指定します。ユーザーが検索結果を表示するときの文書サマリーは、ここで指定する値を超えなくなります。
3. 各文書が含むことのできるセンテンスの数を指定します (サマリーは 10 個までのセンテンスを含むことができます)。
4. 「**OK**」をクリックします。
5. 変更を有効にするために、検索サーバーをモニターし、サーバー・プロセスを再始動します。

## プロパティの編集による文書サマリーのカスタマイズ

エンタープライズ・サーチ照会の各結果文書にサマリーが含まれています。プロパティ・ファイルの編集によって、各サマリーに含まれる情報をカスタマイズすることができます。

### このタスクについて

ES\_NODE\_ROOT/master\_config/collection\_ID.runtime.node1/runtime-generic.properties ファイル内の以下のプロパティの値を変更することにより、検索結果の説明をカスタマイズできます。

#### **MinWordsPerSentence**

説明文に含めることができるワードの最小数。デフォルト値は 4 です。

#### **MaxWordsPerSentence**

説明文に含めることができるワードの最大数。デフォルト値は 20 です。

#### **NumberOfReturnedSentences**

文書の説明を構成する文の数。デフォルト値は 5 です。

#### **MaxSentencesPerDocument**

説明を作成するプロセスで候補と見なされる文書内の文の最大数。デフォルト値は 1000 です。

### 手順

検索結果内の文書サマリーをカスタマイズするには、次のようにします。

1. 検索サーバーに、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. テキスト・エディターを使用して以下のファイルを編集します。ここで、`coll_ID` は、コレクションの作成時にそのコレクションについて指定された ID です (または、システムによって割り当てられたもの)。

```
ES_NODE_ROOT/master_config/coll_ID.runtime.node1/runtime-generic.properties
```

**ヒント:** コレクション名とその ID との間でのマッピングを確認するためには、`ES_NODE_ROOT/master_config/collections.ini` ファイルを参照してください。

3. カスタマイズしたいプロパティを変更してから、ファイルを保管して終了します。
4. 検索サーバーを停止して再始動して、変更を適用します。

---

## クイック・リンクに関する作業

クイック・リンクは、ユーザーが特定の語句を含む照会をサブミットしたときに検索結果に戻される文書を指しています。

コレクションのクイック・リンクを構成するには、エンタープライズ・サーチ管理コンソールを使用します。

### クイック・リンク

クイック・リンクによって、照会用語に関連すると事前指定された文書へのリンクをユーザーに提供できます。

クイック・リンクは、照会に特定の語句が含まれる場合に、エンタープライズ・サーチが自動的に検索結果に入れるようにする URI です。通常、クイック・リンク URI は結果リストの先頭に表示され、照会に適合するとあらかじめ設定した文書をユーザーが確実に参照できるようになっています。

クイック・リンクは、他の検索結果に追加する形で戻されます。検索プロセスでは、照会用語に一致する文書がないか索引を検索し、それらの文書の URI をクイック・リンクの URI に追加して戻します。

クイック・リンクを構成する場合、その URI を示すタイトルと要約を指定して、ユーザーがその文書の内容を把握し、それがそのユーザーが取得したい対象の文書かどうかをすぐに判断できるようにすることができます。

例えば、URI 「<http://www.ibm.com/education/us/>」ならば、「米国 IBM 教育関係サイト」というタイトルをつけ、要約として「米国内の教育関係者、専門家、および学生向けソリューション、製品、およびリソースの紹介」などと設定します。

エンタープライズ・サーチ・コレクションでクイック・リンクを使用するには、クイック・リンクを表示するオプションが検索アプリケーションで使用可能になっていなければなりません。検索アプリケーションによっては、ユーザーがコレクションの検索時にクイック・リンクを戻す機能を使用可能/使用不可に設定できる場合があります。

## クイック・リンクの構成

エンタープライズ・サーチ・コレクションのクイック・リンクを作成するため、文書の URI を、検索結果に含ませたものをトリガーするキーワードと関連付けます。

### 始める前に

クイック・リンクを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのクイック・リンクが所属するコレクションのコレクション管理者である必要があります。

### このタスクについて

クイック・リンクのキーワードおよび URI の指定方法の例を参照するには、クイック・リンクを作成または編集しているときに「ヘルプ」をクリックしてください。

変更を有効にするのに、検索サーバーを再始動する必要はありません。

### 手順

クイック・リンクを構成するには、次のようにします。

1. コレクションを編集、「検索」ページを選択、そして「クイック・リンクの構成」をクリックします。
2. 「クイック・リンク」ページで、「クイック・リンクの作成」をクリックします。
3. このクイック・リンクが検索結果に戻されるようにするためのキーワードと句、この照会に関連すると事前判定した文書の URI、このクイック・リンクのその他のオプションを指定します。

1 行につき、1 つまたは複数のキーワードを指定するか、1 つの句 (2 つ以上の語を引用符で囲んだもの) を指定します。キーワードはスペースで区切ります (コンマを使用してキーワードを区切ることはできません)。Enter キーを押すと改行されます。

4. 「OK」をクリックします。

「クイック・リンク」ページに、このコレクションに属している他のクイック・リンクと共に新しいクイック・リンクがリストされます。

#### **関連資料**

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI)  
は、その文書をコレクションに追加したクローラーのタイプを示します。





---

## エンタープライズ・サーチの文書ランキング

ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。

検索サーバーは、最も関連性の高い検索結果を生成するために、豊富な照会構文をサポートし、テキスト・ベースのスコアリングや静的ランキングなどの幾つかの技法を使用します。次のように、検索結果での文書の重要度を操作するオプションを構成することによって、ランキングのデフォルト動作を拡張することができます。

- カスタムのランキング調整ワード辞書を作成して、指定されたランキング調整ワードが含まれている文書が検索結果内でどのようにランク付けされるのかを操作できます。
- 指定された URI パターンに一致する文書のスコアを操作できます。
- ランキング調整クラスにマップされたフィールドが含まれている文書のスコアを操作できます。

### 関連概念

183 ページの『ランキング調整クラスに基づいた文書ランキング』  
フィールドをランキング調整クラスにマッピングすることにより、検索結果における文書のランクを調整することができます。

181 ページの『URI パターンに基づいた文書ランキング』  
URI パターンにランキング調整要因を割り当てることによって、文書の重要度を増減できます。

178 ページの『カスタム・ランキング調整ワード辞書』  
検索結果の質を向上させるために、カスタム・ランキング調整ワード辞書を作成して、検索結果における文書のランクを調整することができます。

---

## テキスト・ベースのスコアリング

エンタープライズ・サーチでは、照会中の用語に一致した各文書について、動的にスコアを計算します。

照会に一致する各文書のテキストのスコアを計算する場合、エンタープライズ・サーチでは次のようなさまざまな要素を計算に入れます。

- コレクション全体で各照会用語が出現する度数。一般的に、多くの文書に出現する照会用語は、より少ない文書にしか出現しない照会用語に比べて、文書のスコアを高くする程度は小さくなります。
- 照会用語と一致する文書内で各照会用語が出現する回数。一般的に、特定文書の中でその照会用語が多く出現するほどスコアは高くなります。
- 照会用語と一致する各文書内で照会用語が出現する近さ。一般的に、特定の文書中で照会用語が互いに近い場所に出現する場合、それらと同じ照会用語が互いに離れている場所に出現する場合よりも高いスコアがその文書に与えられます。

- 照会用語と一致する各文書内で照会用語が出現するコンテキスト。例えば、文書のタイトルに照会用語が出現する場合、それと同じ照会用語が文書のプレーン・テキスト中に出現する場合よりも高いスコアがその文書に与えられます。

各文書の長さ、および語彙の多さもスコアを決定する際の要因となります。

---

## 静的ランキング

文書のタイプによっては、検索結果における文書の重要度を高くする静的ランキング要素を関連付けることができます。

コレクションを作成するときに、コレクション内の文書に静的ランキング要素を関連付けるかどうかを指定します。Web コンテンツの場合、他の文書から特定の文書に対するリンク数、およびそれらのリンク元によって、検索結果におけるその文書の重要度を高くすることができます。

日付フィールドまたは日付メタデータを含む文書の場合、文書の日付を使用して適合度を高くするように設定することができます。例えば、NNTP ニュースグループではより最近の記事を古い記事よりも適合度を高くする設定ができます。データ・ソースに複数の日付値がある場合は、そのデータ・ソース内の文書の適合度を判定する上で、どの日付を最も重視するかを選択することができます。

静的ランキングをコレクションに使用する場合、異なるランキング・タイプを使用するデータ・ソースを同一のコレクションに混在させないように注意してください。例えば、ある文書に対して静的ランキング要素としてリンク数を使用したい場合、コレクション内の文書が Web 文書だけになっていることを確認してください。異なるランキング・モデルのソースが同じコレクションにまとめられていると、検索品質が低くなることがあります。

また、静的ランキングを適用できるフィールドと値がその文書に含まれていることを確認してください。例えば、コレクションの文書のランキング要素として文書日付を使用する指定をした場合に文書に日付フィールドまたは日付属性がないと、検索品質は低下することがあります。

### 関連タスク

245 ページの『WebSphere Portal からのコレクションのマイグレーション』  
WebSphere Portal からエンタープライズ・サーチにコレクションをマイグレーションするには、WebSphere Portal にコレクションを準備し、マイグレーション・ウィザードを使用して準備したコレクションをマイグレーションします。

---

## カスタム・ランキング調整ワード辞書

検索結果の質を向上させるために、カスタム・ランキング調整ワード辞書を作成して、検索結果における文書のランクを調整することができます。

照会でランキング調整ワード辞書にある用語が指定されると、その用語が含まれている文書の重要度が、辞書の実語に構成されているランキング調整要因に応じて高くなったり、低くなったりします。

ランキング調整要因の範囲は -10 から 10 までです。照会処理時に、検索サーバーは、正の数のランキング調整要因の用語が含まれている文書の重要度を高くし、負の数のランキング調整要因の用語が含まれている文書の重要度を低くします。

例えば、ランキング調整要因が高い照会用語が含まれている文書は、ランキング調整要因が適用されない場合に比べて、高くランクされます。(ランキング調整されている照会用語のみが検索結果における文書のランクに影響します。ランキング調整要因は、文書のスコアに影響する唯一の係数です。)

辞書を作成する場合、同じランキング調整要因を複数の用語に割り当てることができます。辞書には、単一の用語または複数の用語を含めることができます (複数の用語は句として一致します)。

ランキング調整値によって重み付けされた用語が OR 演算子を使用する照会 (例えば this | that) で指定された場合、その照会用語に対して重み付けされた平均が計算されます。計算結果の集約スコアが OR 照会オペランドのすべてのオカレンスに使用されます (個々の OR 照会オペランドについて、個々のスコアが計算されることはありません)。

ランキング調整ワード辞書に基づくランキング調整は、フィールド照会用語ではサポートされません。照会用語の解析時に、フィールド名ではなく照会テキストのみが、文書スコアの計算に使用されます。(フィールド内の照会用語にランキング調整要因を適用するためには、フィールド名をランキング調整クラスにマップすることができます。)

ランキング調整ワード辞書を作成するには、そのコレクションの主題についてのエキスパートが XML フォーマットでランキング調整ワード・リストを作成する (あるいは、アプリケーション開発者と一緒に作業して XML ファイルを作成する) 必要があります。WebSphere II OmniFind Edition で提供されるツールを使用して、XML ファイルをバイナリー (.dic) ファイルに変換する必要があります。

エンタープライズ・サーチ管理者は、このバイナリー・ファイルをシステムにアップロードし、表示名を割り当てます。コレクション管理者は、コレクションの検索サーバー・オプションを構成するときに、コレクション内の文書の検索に使用するランキング調整ワード辞書を選択します。

**制約事項:** カスタム・ランキング調整ワード辞書は、システムに追加した後で編集することはできません。照会処理に使用可能なランキング調整ワードを変更するには、以下を行う必要があります。

1. ソース XML ファイルを更新します。
2. XML ソースを新規ディクショナリー・ファイルに変換します。
3. ランキング調整ワード辞書を使用するコレクションから古い辞書を削除します。
4. システムからランキング調整ワード辞書を削除します。
5. 新規のランキング調整ワード辞書をシステムに追加します。
6. 新規ランキング調整ワード辞書を、その辞書を使用するコレクションに関連付けます。

## 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』  
ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語  
および条件に照らして最も関連する結果を戻します。

「テキスト分析機能ガイド」の『カスタム・ランキング調整ワード辞書』

#### 関連タスク

「テキスト分析機能ガイド」の『ランキング調整ワードに使用できる XML ファ  
イルの作成』

「テキスト分析機能ガイド」の『ランキング調整ワード辞書の作成』

## ランキング調整ワード辞書のシステムへの追加


カスタム・ランキング調整ワード辞書を作成する場合、辞書をエンタープライズ・  
サーチ・システムに関連付ける必要があります。そうすることにより、コレクショ  
ンの検索に使用したいランキング調整ワード辞書を後で選択することができます。

### 始める前に

カスタム・ランキング調整ワード辞書をシステムに追加するには、エンタープライ  
ズ・サーチ管理者役割の担当者である必要があります。

### 手順

カスタム・ランキング調整ワードをエンタープライズ・サーチ・システムに関連付  
けるには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2.  「編集」をクリックして、システム編集ビューに移動します。
3. 「検索」ページで、「ランキング調整ワード辞書の構成」をクリックします。
4. 「ランキング調整ワード辞書の構成」ページで、「ランキング調整ワード辞書の追加」をクリックします。
5. 「ランキング調整ワード辞書の追加」ページで、辞書の固有の表示名を入力し、オプションで、説明を入力します。
6. .dic ファイルのロケーションを指定します。ファイルが 8 MB より小さい場  
合、そのファイルをシステム上に置くことができます。ファイルが 8 MB より  
大きい場合、そのファイルは索引サーバーに置かなければなりません。
7. 「OK」をクリックします。カスタム・ランキング調整ワード辞書が、エンター  
プライズ・サーチ・システムに追加され、コレクションの検索に使用できるよ  
うになります。

## ランキング調整ワード辞書とコレクションとの関連付け

ランキング調整ワード辞書がエンタープライズ・サーチ・システムと関連付けられ  
ている場合、そのうちの 1 つを選択して、コレクションの検索時に使用できます。  
照会用語が辞書内の用語に一致した場合、その用語が含まれている文書の重要度  
が、辞書の用語に割り当てられているランキング調整要因に応じて高くなったり、  
低くなったりします。

### 始める前に

コレクションのランキング調整ワード辞書を選択するには、エンタープライズ・サーチ管理者役割の担当者であるか、そのコレクションのコレクション管理者である必要があります。

### 手順

ランキング調整ワード辞書をコレクションと関連付けるには、次のようにします。

1. コレクションを編集して「検索」ページを選択し、「**検索サーバー・オプションの構成**」をクリックします。
2. 「検索サーバー・オプション」ページの「**ランキング調整ワード辞書名**」フィールドで、ユーザーがこのコレクションを照会するとき使用するランキング調整ワード辞書の名前を入力します。

使用できる辞書のリストに、エンタープライズ・サーチ・システムに追加されたすべてのランキング調整ワード辞書が含まれます。

3. 「**OK**」をクリックします。

---

## URI パターンに基づいた文書ランキング

URI パターンにランキング調整要因を割り当てることによって、文書の重要度を増減できます。

すべての文書には、索引に追加されるときにデフォルトの静的ランキング・スコアが割り当てられます。デフォルトのスコアは、コレクションに対して静的ランキングが使用可能化されたかどうかによって変わり、使用可能化された場合には静的ランキング・タイプ (文書日付、または、Web 文書の場合はリンクする他の文書の数) によって変わります。

URI パターンにランキング調整要因を割り当てることによって、文書の相対的重要度を操作できます。ランキング調整要因は、デフォルトの静的ランキング・スコアおよび他の要因と共に、文書の最終的な静的スコアを決定するのに使用されます。

構成する URI パターンの順序は重要です。索引サーバーは、コレクション内の各文書の値を計算する際、URI パターンをリストされている順に評価します。それぞれの URI ごとに以下ようになります。

1. 索引サーバーは、URI パターンを順番にスキャンします。
2. 索引サーバーは、索引内に一致する文書がある最初の URI を検出すると、その URI パターンに構成されたランキング調整要因をその文書に適用します。
3. どの URI パターンにも一致しない文書には、デフォルトの静的ランキング・スコアが使用されます。

URI パターンにランキング調整要因を構成する場合は、その後、索引サーバーにスキャンさせたい順序でその URI パターンを配置する必要があります。

### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』

ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。



## URI パターンに一致する文書のスコアの操作

デフォルトの静的ランキング・スコアにランキング調整要因を適用することによって、URI パターンに一致する文書の重要度を増減することができます。

### 始める前に

URI パターンに一致する文書の重要度を操作するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

構成するランキング調整要因は、デフォルトの静的ランキング・スコアと共に使用されて、指定された URI パターンに一致するすべての文書について新しい静的スコアが計算されます。

ランキング調整要因は、静的スコアのみを調整するものであり、文書の最終的なランクを決定する計算に影響する 1 つの要素にすぎません。例えば、多数からリンクされる文書 (この場合は初期スコアが高くなります) がある場合、リンクがまったくない文書は常にそれよりも低くランク付けされます。

### 手順

URI パターンに一致する文書のスコアを操作するには、次のようにします。

1. コレクションを編集して「索引」ページを選択し、「**URI パターン・マッチングによるスコアの操作**」をクリックします。
2. 「**URI パターン・マッチングによるスコアの操作**」ページで、「**URI パターンの追加**」をクリックします。
3. 検索結果内での重要度を増減したい文書の URI パターンを指定します。以下に例を示します。



```
http://domain.org/hr/*  
db2://*ROI*  
*/afs/*
```

4. ランキング調整要因に -10 と 10 の間の値を入力します。このランキング調整要因に基づいて、指定した URI パターンに一致するすべての文書の最終的な静的スコアが計算されます。
5. 「**OK**」をクリックします。
6. 「**URI パターン・マッチングによるスコアの操作**」ページで、新しい URI パターンを、索引サーバーにスキャンさせたい順序で配置します。

索引サーバーは、URI がリストされている順序で、静的ランキング・スコアを計算します。最良の結果を得るには、より特定度の高い URI を先にリストします。以下の例で、/forms サブディレクトリーは `http://www.ibm.com/hr/*` URI パターンに一致します。/forms サブディレクトリー内の文書のスコアが確実に正しく計算されるようにするには、/forms サブディレクトリーの URI パターンを先にリストします。

```
http://www.ibm.com/hr/forms/* 8  
http://www.ibm.com/hr/* -2
```



- URI パターンまたはランキング調整要因を変更するには、目的の URI パターンを選択して、「 編集」をクリックします。
- リストから URI パターンを除去するには、目的の URI パターンを選択して、「 除去」をクリックします。
- 以前に索引作成された文書にランキング調整要因を適用するには、索引を再編成します。

#### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』  
ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。

#### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

---

## ランキング調整クラスに基づいた文書ランキング

フィールドをランキング調整クラスにマッピングすることにより、検索結果における文書のランクを調整することができます。

文書の構文解析時に、パーサーは、トークンが属しているフィールドに応じて **ランキング調整クラス** を文書トークンに割り当てます。これらのランキング調整クラスは、索引に組み込まれ、照会の評価時に、結果文書のランクに影響するスコアの計算に使用されます。

スコアの計算方法に影響を与えるには、ランキング調整クラスのランキング調整要因を構成します。照会用語が、ランキング調整クラスにマップされているフィールドのトークンに一致すると、トークンのこのオカレンスに課せられている係数が文書の合計スコアに影響します。スコアは、ランキング調整クラスに構成されているランキング調整要因を適用して計算されます。

例えば、タイトル・フィールドのスコアのランキングを上げると、照会用語がタイトルにある場合、オカレンスによって文書スコアが上昇するため、検索結果における文書のランクが上昇します。

文書ランキングを調整するには、エンタープライズ・サーチ管理コンソールを使用して、ランキング調整クラスのランキング調整要因を指定し、フィールドをランキング調整クラスにマップします。エンタープライズ・サーチ用に 16 のランキング調整クラスが事前構成されています。そのうちの 8 つのランキング調整クラスはコンテンツ・フィールドで使用するように設計されており、それ以外の 8 つのランキング調整クラスはメタデータ・フィールド用に設計されています。デフォルトのランキング調整クラスに関連付けられているスコアを編集したり、ランキング調整クラスに別のフィールドや追加フィールドを関連付けたりすることができます。

フィールドのマッピングを変更する場合には、以前に索引付けされていた文書に変更が適用されるように、文書を再度クロールおよび構文解析する必要があります。

ランキング調整クラスに指定されている係数を変更する場合には、検索サーバーをモニターし、変更が有効になるように、検索サーバー処理を停止して再始動してください。

## 重複文書の検出

フィールドをランキング調整クラスにマップする際には、重複文書の検出にフィールドが使用されるかどうかを指定する必要があります。

- フィールドが重複文書の検出に使用される場合、フィールドはコンテンツ・フィールドであるとみなされ、コンテンツ・フィールド用に設計されているランキング調整クラスのみが選択対象となります。
- フィールドが重複文書の検出に使用されない場合、フィールドはメタデータ・フィールドであるとみなされ、メタデータ・フィールド用に設計されているランキング調整クラスのみが選択対象となります。この場合、指定されているフィールド以外すべての点において同じである 2 つの文書は、互いに重複しているとみなされます。

## 高い再呼び出し値と低い再呼び出し値

照会を評価する際に、検索プロセスは、戻される結果文書数を見積もります。しきい値によって、照会が低い再呼び出し値であるか、または高い再呼び出し値であるかが決定されます。

### 低い再呼び出し

結果文書の見積もり数が下限しきい値を下回る場合、照会は低い再呼び出し照会であるとみなされます。

### 高い再呼び出し

結果文書の見積もり数が上限しきい値を上回る場合、照会は高い再呼び出し照会であるとみなされます。

**混合** 文書の見積もり数が 2 つのしきい値の間である場合、照会の再呼び出し値は 2 つのしきい値の混合になります。

各ランキング調整クラスは、照会処理時に、低い再呼び出し照会および高い再呼び出し照会に関連付けられます。低いランキング調整要因は低い再呼び出し照会の相対重要度に影響し、高いランキング調整要因は高い再呼び出し照会の相対重要度に影響します。2 つのランキング調整要因の混合は、混合の再呼び出し値を持つ照会の相対重要度に影響します。

ランキング調整要因の値は、文書内の照会用語の各オカレンスの相対重要度を制御します。文書内の照会用語の各オカレンスは、対応するランキング調整要因に応じてカウントされます。

コレクションにランキング調整クラスを構成する場合は、デフォルトのランキング調整要因を編集できます。例えば、タイトル・フィールドにある照会用語が、通常のテキストにある照会用語に比べて 5 倍にカウントされるようにランキング調整要因を指定することができます。

### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』  
ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語  
および条件に照らして最も関連する結果を戻します。

## ランキング調整クラスへのフィールドのマッピング

フィールド名をランキング調整クラスにマッピングすることにより、フィールドの  
相対重要度に影響を与えることができます。

### 始める前に

フィールドをランキング調整クラスにマップするには、エンタープライズ・サーチ  
管理者役割の担当者であるか、そのコレクションのコレクション管理者である必要  
があります。

### このタスクについて

システムはランキング調整要因を使用して、ランキング調整クラスにマップされて  
いるフィールド内に照会用語が含まれている文書のランキングを調整します。

エンタープライズ・サーチでは、いくつかのフィールド名が予約されているので、  
それ以外の定義特性を持たないフィールドおよびテキスト (例えば、タイトル・フ  
ィールドでないフィールドや強調されない通常のテキストなど) のスコアを計算で  
きます。予約フィールド以外のフィールドを予約フィールドが使用するランキング  
調整クラスにマップすることはできますが、予約フィールドを編集したり削除する  
ことはできません。

### 手順



フィールドをランキング調整クラスにマップするには、次のようにします。

1. コレクションを編集し、「構文解析」ページを選択して、「**フィールドをランキ  
ング調整クラスにマップ**」をクリックします。
2. 「**フィールドをランキング調整クラスにマップ**」ページで、「**フィールドの追  
加**」をクリックします。
3. 「**フィールドをランキング調整クラスに追加**」ページで、ランキング調整クラス  
にマップしたいフィールドの名前を入力します。

クロールされたソースや外部ソースにあるフィールド名、XML エlementによ  
りマップされるフィールド名、HTML メタデータ・Elementによりマップさ  
れるフィールド名、事前定義されているフィールド名を指定できます。

4. フィールドが重複文書の検出に使用されるかどうかを指定します。チェック・  
ボックスを選択すると、選択可能なランキング調整クラスのリストに、コンテン  
ツ・フィールドに適用されるクラスが表示されます。チェック・ボックスをクリ  
アすると、選択可能なランキング調整クラスのリストに、メタデータ・フィール  
ドに適用されるクラスが表示されます。
5. ランキング調整クラスを選択して、「**OK**」をクリックします。

追加したフィールドが「フィールドをランキング調整クラスにマップ」ページに表示されます。オプションを選択して、ランキング調整クラスを編集し、このフィールドが含まれている文書のスコアの決定に使用されるさまざまなランキング調整要因を構成します。

6. フィールドが重複文書の検出に使用されるかどうかを変更する場合、またはフィールドを別のランキング調整クラスにマップする場合は、「 **編集**」をクリックします。(エンタープライズ・サーチによって予約されているフィールドは編集できません。)
7. ランキング調整クラスからフィールドを除去する場合は、「 **除去**」をクリックします。(エンタープライズ・サーチによって予約されているフィールドは除去できません。)
8. 以前に索引付けされている文書に変更を適用するには、文書を再度クロールし、構文解析し、索引付けする必要があります。

#### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』  
ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。

## ランキング調整クラスのランキング調整要因の構成

ランキング調整クラスに対して構成するランキング調整要因は、結果文書内の特定のフィールドの存在が照会にどのように関連するかについてのユーザーの見積もりを表します。ランキング調整クラスに最も高いランキング調整要因が指定されている場合、そのランキング調整クラスにマップされているフィールドが含まれている結果文書の重要度が高くなります。

#### 始める前に


ランキング調整クラスのランキング調整要因を構成するには、エンタープライズ・サーチ管理者役割の担当者であるか、そのコレクションのコレクション管理者である必要があります。

#### このタスクについて

システムは、ランキング調整クラスに構成されているランキング調整要因、デフォルトの静的ランキング・スコア、およびその他の係数を使用して、ランキング調整クラスにマップされているフィールドが含まれている結果文書の新規スコアを計算します。

#### 手順

ランキング調整クラスのランキング調整要因を構成するには、次のようにします。

1. コレクションを編集し、「構文解析」ページを選択して、「**フィールドをランキング調整クラスにマップ**」をクリックします。
2. 「**フィールドをランキング調整クラスにマップ**」ページで、「**ランキング調整クラスの編集**」をクリックします。
3. 「**ランキング調整クラス**」ページで、変更したいランキング調整クラスを指定し、「 **編集**」をクリックします。

4. 「ランキング調整クラスの編集」ページで、ランキング調整要因の最大値と最小値に新規の値を指定します。両方の係数に同じ値を指定することもできます。
5. 「OK」をクリックします。
6. 変更を有効にするために、検索サーバーをモニターして、検索処理の停止および再始動のアイコンを選択します。ユーザーが照会を実行依頼すると、このランキング調整クラスにマップされているフィールドが含まれている結果文書の相対重要度が、新規ランキング調整要因によって判別されます。

#### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』

ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。

## ランキング調整クラスのデフォルト値

エンタープライズ・サーチでは、検索結果における文書のランク付けに影響を与える 16 のランキング調整クラスが用意されています。

エンタープライズ・サーチでは、他の定義特性を持たないフィールドおよびテキストのスコアを計算するために、以下のフィールドが予約されています。

```
es_special_field.regular_text
es_special_field.default_field
```

予約フィールド以外のフィールドを予約フィールドが使用するランキング調整クラスにマップすることはできますが、予約フィールドを編集したり削除することはできません。

予約フィールド以外のすべてのフィールドについては、システムが文書のランク計算に使用するランキング調整要因を編集することができます。任意の数のフィールドを任意のランキング調整クラス (予約フィールドによって使用されるランキング調整クラスも含む) にマップすることができます。

以下の表に、ランキング調整クラス名、低い再呼び出し値を持つ照会用のデフォルト・ランキング調整要因、高い再呼び出し値を持つ照会用のデフォルト・ランキング調整要因、デフォルト構成でランキング調整クラスにマップされている事前定義フィールド名を示します。

デフォルトのランキング調整要因は、コレクション作成時にコレクションに対して選択された静的ランキング・メソッドによって異なります。オプションには、静的でないランキング、文書に対するリンク数によって決まるランク (Web ソースの場合)、文書日付によって決まるランクがあります。

表 3. ランキング調整クラスのデフォルト値

ランキング調整クラス名	デフォルト時の低いランキング調整要因および高いランキング調整要因			
	静的でないランキング	文書リンク	文書日付	事前定義フィールド・マッピング
Content class A	低: 4 高: 2	低: 6 高: 1	低: 4 高: 2	es_special_field.regular_text

表3. ランキング調整クラスのデフォルト値 (続き)

デフォルト時の低いランキング調整要因および高いランキング調整要因				
ランキング調整クラス名	静的でないランキング	文書リンク	文書日付	事前定義フィールド・マッピング
Content class B	低: 5 高: 4	低: 7 高: 3	低: 5 高: 4	es_special_field.html_emphasized_text  次の HTML エlementを含む: b、big、caption、dfn、em、h4、h5、h6、strong
Content class C	低: 7 高: 4	低: 9 高: 3	低: 7 高: 4	es_special_field.html_headers  次の HTML エlementを含む : h1、h2、h3
Content class D	低: 2 高: 5	低: 1 高: 5	低: 2 高: 5	title
Content class E	低: 1 高: 1	低: 5 高: 10	低: 1 高: 1	es_special_field.anchor
Content class F	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.anchor_same_dir
Content class G	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.anchor_same_host
Content class H	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
Metadata class A	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.default_field
Metadata class B	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
Metadata class C	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
Metadata class D	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
Metadata class E	低: 1 高: 1	低: 5 高: 1	低: 1 高: 1	keywords
Metadata class F	低: 1 高: 1	低: 3 高: 1	低: 1 高: 1	es_special_field.urlhost
Metadata class G	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.urlpath
Metadata class H	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	description

### 関連概念

177 ページの『エンタープライズ・サーチの文書ランキング』  
ユーザーがコレクションを検索すると、検索プロセスは照会に指定された用語および条件に照らして最も関連する結果を戻します。



---

## エンタープライズ・サーチの検索アプリケーション

検索アプリケーションは、エンタープライズ・サーチ・システム内のコレクションおよび外部ソースの検索を可能にします。検索アプリケーションはいくつでも作成できます。また、1つの検索アプリケーションで、任意の数のコレクションおよび外部ソースを検索できます。

### サンプル検索アプリケーション

サンプル検索アプリケーションによって、エンタープライズ・サーチで使用可能な多くの検索、取り出し機能を実際に試すことができます。サンプル・アプリケーションは、企業の目標を反映する、対話式のカスタム検索アプリケーションを作成するために、IBM 検索および索引 API (SI-API) の使用方法を示す実施例でもあります。

デフォルトの構成ファイルのプロパティを変更しないかぎり、サンプル検索アプリケーションを使用して、エンタープライズ・サーチ・システム内のすべてのアクティブなコレクションおよび外部ソースを検索することができます。新しいコレクションおよび外部ソースをユーザーに使用可能にする前に、サンプル検索アプリケーションを使用してそれらをテストできます。

サンプル検索アプリケーションは、自動的にすべてのコレクションおよび外部ソースと関連付けられます。実稼働環境では、エンタープライズ・サーチ管理者が、さまざまなコレクションをどの検索アプリケーションで検索できるかを制御します。

### カスタム検索アプリケーション

作成する検索アプリケーションは、IBM WebSphere Application Server 環境で独立型 Web アプリケーションとして実行することも、IBM WebSphere Portal 環境でポータルレットとして起動することもできます。検索および索引 API を使用することによって、サンプル検索アプリケーションのように、両方の環境でシームレスに機能する検索アプリケーションを設計できます。

#### 関連概念

154 ページの『文書の検索に対する索引付きオプション』

クロールされたデータまたは外部ソースを検索するオプションを構成するとき、または、XML および HTML メタデータ・エレメントを検索フィールドにマップするときに、文書がどのように検索でき、検索結果に表示されるのかを指定します。

220 ページの『検索アプリケーション ID を使用したセキュリティー』

コレクション・レベルのセキュリティーを提供するために、各コレクションおよび外部ソースを検索できる検索アプリケーションを指定します。

221 ページの『文書レベル・セキュリティー』

コレクションが作成されたときに、そのコレクションに対するセキュリティーを使用可能にすると、文書レベルのセキュリティー制御を構成することができます。文書レベルのセキュリティーによって、コレクションを検索するユーザーは、見ることが許される文書にしかアクセスできないようになります。

235 ページの『ポータル検索エンジンでの文書レベル・セキュリティー』  
IBM WebSphere ポータル検索エンジンを使用して、ユーザーがエンタープライズ・サーチ・コレクションを検索するときに文書レベルのセキュリティーを施行することができます。

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API の概要』

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『照会構文』

---

## コレクションとの検索アプリケーションの関連付け

新規の検索アプリケーションを使用するには、その検索アプリケーションを、それが検索できるコレクションに関連付けておく必要があります。

### 始める前に

検索アプリケーションを、それが検索できるコレクションに関連付けるには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### 手順

検索アプリケーションを 1 つ以上のコレクションに関連付けるには、次のようにします。

1. 管理コンソールのツールバーで「**セキュリティー**」をクリックします。
2. 「検索アプリケーション」ページで、「**検索アプリケーションの構成**」をクリックします。
3. 「検索アプリケーションの構成」ページで、「**検索アプリケーションの追加**」をクリックします。
4. 検索アプリケーションの名前を入力します。
5. アプリケーションが検索できるコレクションを選択します。

・ システムに追加するすべてのコレクションを検索アプリケーションがアクセスできるようにするには、「**すべてのコレクションと外部ソース**」をクリックします。

・ 指定するコレクションのみを検索アプリケーションがアクセスできるようにするには、「**特定のコレクションと外部ソース**」をクリックします。

このオプションを選択すると、コレクション名および外部ソース名のリストが表示されます。アプリケーションが検索できるコレクションごとに、その「**選択**」チェック・ボックスを選択します。

6. 「**OK**」をクリックします。

---

## サンプル検索アプリケーションの機能

エンタープライズ・サーチのサンプル検索アプリケーションを使用して、カスタム検索アプリケーションに構築できる検索機能のほとんどを実際に試してみることができます。

サンプル検索アプリケーションを使用して、一度に 1 つ、一度に複数、あるいは一度にすべてのコレクションおよび外部ソースの検索を行うことができます。アプリケーション・プロパティのデフォルトが変更されていない限り、このアプリケーションを使用して、エンタープライズ・サーチ・システム内のすべてのコレクションおよび外部ソースを検索できます。

## 照会機能

これらの機能を使用して、次のことができます。

- 単純なフリー・テキストの照会を指定する。
- 検索結果の精度を上げるために、より複雑な照会を指定する。例えば、文書内の特定のフィールドを検索したり、特定の語句を含ままたは除外する照会構文を使用して文書を検索できます。
- 検索する対象のコレクションおよび外部ソースを指定する。
- 特定のソース・タイプまたはすべてのソース・タイプを検索する。
- 特定のタイプの文書を検索する。例えば、Microsoft Word 文書のみを検索したり、PDF 形式の文書のみを検索できます。
- 使用する照会用語の言語を指定する。検索対象にする文書の言語も指定できます。
- コレクションの特定のサブセットを検索する。例えば、検索アプリケーションは、事前に定義された文書の範囲 (有効範囲) にビューを制限できます。あるいは、指定されたカテゴリーに属する文書のみを検索する照会をサブミットできます。
- 照会を拡張して、照会用語の同義語を含める。同義語辞書がコレクションに関連付けられている場合、照会用語の同義語を含む文書が検索結果に戻されます。

## 検索結果機能

これらの機能を使用して、次のことができます。

- 照会用語に一致する検索結果を表示する。
- 各ページに表示する結果文書数を制御し、結果セットの間で前方または後方を参照する。
- 結果文書についての詳細を表示、または非表示にする。例えば、文書の要旨を表示したり、あるいは、各結果文書内のフィールドの名前などの詳細を表示できます。
- 同じソースからの文書を縮小表示する。例えば、1 つのソースが 100 件の文書を返す場合、結果セットの中では、もっとも関連する 2 文書がグループ化されて表示されます。残りの 98 文書は、同じソースからもっと多くの文書を表示するオプションを選択すると表示されます。
- 関連性で文書をソートしたり、文書の日付の昇順にあるいは降順に文書をソートする。
- 照会ストリングの中にミススペルの可能性のある語があれば、スペルの修正候補のプロンプトを出す。
- 結果の文書が属するカテゴリーについての情報を表示し (コレクションがカテゴリーを使用する場合)、特定のカテゴリーに属する文書のみを参照する。
- 照会用語を追加で指定し、検索結果内でさらに検索を行う。

## 文書検索機能

これらの機能を使用して、次のことができます。

- 文書 URI をクリックして、文書を取得する。クライアント・アプリケーションが使用可能な場合、ネイティブ・ビューアーで結果文書を表示することもできます。例えば、Notes Client アプリケーションがインストールされている場合、そのアプリケーションを使用して Lotus Notes データベースからの文書を表示できます。

文書レベルのセキュリティがクローラーに対して構成されている場合、セキュア・コンテンツへのアクセスを認可されているユーザーのみが文書を取得できます。

- クイック・リンクをクリックして、文書を取得する。クイック・リンクは、キーワードを URI に関連付けます。指定されたキーワードが照会の中に含まれていれば、関連した URI (これは、これらのキーワードに高い関連があると事前に決められている) が検索結果の先頭に表示されます。

---

## サンプル検索アプリケーションのプロパティ

config.properties ファイルをサンプル検索アプリケーションについて編集することによって、環境に合わせたオプションを指定し、アプリケーションの外観を変更し、ユーザーが検索アプリケーションを開始した後に使用できるオプションを制御できます。

### 環境パラメーター

検索アプリケーションの操作を制御するオプションを指定できます。

#### applicationName

有効な検索アプリケーションの名前を指定します。デフォルト値は Default です。

別のアプリケーションをデフォルト・アプリケーションとして使用する場合、このデフォルト値を変更してください。

**ヒント:** アプリケーション名が Default の場合、このサンプル検索アプリケーションを使用して、すべてのコレクションおよび外部ソースを 1 回の照会で検索できます。

#### hostname

ご使用の WebSphere Application Server インスタンスをサポートするように構成された Web サーバーの、完全修飾ホスト名を指定します。デフォルト値は localhost です。

検索アプリケーションが正しく機能するようにするため、デフォルト値を、WebSphere Application Server が使用すると構成されている完全修飾ホスト名に変更してください。例えば、コンピューターのホスト名が MyMachine で、Web サーバーがそのホスト名として www.ibm.com を使用するように構成されている場合、www.ibm.com を指定します。

**port** ご使用の WebSphere Application Server インスタンスをサポートするように

構成された Web サーバーのポート番号を指定します。デフォルト値は 80 です。Web サーバー・インスタンスのポートを変更した場合にのみ、デフォルト値を変更してください。

#### **timeout**

検索サーバーからの応答をどれだけ待機すれば検索要求がタイムアウトになるのかを秒数で指定します。この数値は整数でなければなりません (60.5 や sixty ではなく、60)。タイムアウト値を指定しない場合、デフォルト値は 30 秒です。

#### **username**

ユーザーによって検索要求が実行依頼されたときにエンタープライズ・サーチがユーザー認証を実行し、WebSphere Application Server でユーザーを認証することができるようにするユーザー名を指定します。このフィールドは、WebSphere Application Server でグローバル・セキュリティーを使用可能に設定した場合にのみ使用されます。

#### **password**

指定されたユーザー名のパスワードを指定します。

#### **filter**

検索結果内にリストされた文書を取得するのに使用するクラスを指定します。デフォルトのクラスは、`com.ibm.es.api.filters.SetDocumentURIFilterFetch` です。代わりにカスタムのクラスを使用して文書を取得する場合のみ、この値を変更してください。

#### **logging.level**

ログに出力する詳細の量を以下のように指定します。

**OFF** どのメッセージもログに記録されません。

**INFO** 通知メッセージがログに記録されます。

**FINE** エラー・メッセージがログに記録されます。

**ALL** すべてのメッセージがログに記録されます。

## **ソース・タイプ用アイコン**

結果文書が属しているデータ・ソースのタイプを表すイメージをカスタマイズできます。ソース・タイプ用の以下のアイコンが `config.properties` ファイルに事前定義されていて、WebSphere Information Integrator OmniFind Edition のインストール時にサポートされたクローラーおよび外部ソースを示します。



#### **documentSource.vbr.icon**

文書が Content Edition クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは `/images/sourceVBR.gif` です。



#### **documentSource.db2.icon**

文書が DB2 クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは `/images/sourceDB2.gif` です。



#### **documentSource.cm.icon**

文書が DB2 Content Manager クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceCM.gif です。



#### **documentSource.dominodoc.icon**

文書が Domino Document Manager クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceDominoDoc.gif です。



#### **documentSource.exchange.icon**

文書が Exchange Server クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceExchange.gif です。



#### **documentSource.nntp.icon**

文書が NNTP クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceNNTP.gif です。



#### **documentSource.notes.icon**

文書が Lotus Notes クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceNotes.gif です。



#### **documentSource.quickplace.icon**

文書が QuickPlace クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceWorkplace.gif です。



#### **documentSource.unixfs.icon**

文書が UNIX ファイル・システム・クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceUnixFS.gif です。



#### **documentSource.web.icon**

文書が Web クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceWeb.gif です。



#### **documentSource.wps.icon**

文書が WebSphere Portal クローラーによってクロールされたことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceWPS.gif です。



#### **documentSource.winfs.icon**

文書が Windows ファイル・システム・クローラーによってクロールされた



ことを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceWindowsFS.gif です。



#### **documentSource.ldap.icon**

文書が LDAP サーバー用に作成された外部ソースに属していることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceLDAP.gif です。



#### **documentSource.jdbc.icon**

文書が Java Database Connectivity (JDBC) データベース表用に作成された外部ソースに属していることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/sourceJDBC.gif です。

## クライアント・ビューアー用アイコン

結果文書は Web ブラウザーで表示できます。結果文書タイプ固有のクライアント・アプリケーションが使用可能な場合、クライアント・ビューアーでも文書を表示できます。文書を表示するのに使用されるクライアント・ビューアーのタイプを表すイメージをカスタマイズできます。以下の例では、文書を Lotus Notes クライアント・アプリケーションで表示できることを示すために Lotus Notes アイコンが使用されています。

```
client.notes.icon=/images/notes.gif
```

検索結果には、アイコンとクライアント・アプリケーションへのリンクが、以下のように表示されます。



### クライアント・ビューアー

## 文書フィールド

フィールドがあるデータ・ソースのタイプに対して、どのフィールドが結果文書に表示されるのかを制御できます。

**fields.URI prefix=**スペースで区切ったフィールド名リスト

URI 接頭部にあるコロン文字 (:) の前に円記号 (¥) を付けて、コロン文字をエスケープする必要があります。フィールド名リストを次の行に継続する場合、前の行が円記号 (¥) で終わるようにしてください。以下に例を示します。

```
fields.db2¥://=databasename tablename  
fields.domino¥://=servername databasename databasetitle filename extension ¥  
createddate modifieddate  
fields.cm¥://=servername itemtypename createddate modifieddate mimetype  
fields.file¥://=directory filename extension modifieddate filesize title
```

## フィールド用アイコン

フィールドがあるデータ・ソースのタイプおよび文書に対して、フィールドを表すイメージをカスタマイズできます。フィールド用の以下のアイコンが config.properties ファイルに事前定義されています。



#### **field.icon.databasetitle**

文書タイトルが入っているフィールドであることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/notesdb.gif です。



#### **field.icon.databasename**

文書が属しているデータベースの名前が入っているフィールドであることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/db2.gif です。



#### **field.icon.tablename**

文書が属している表の名前が入っているフィールドであることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/table.gif です。



#### **field.icon.directory**

文書が属しているディレクトリーの名前が入っているフィールドであることを示すイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/closedFolder.gif です。

## デフォルトのフィールド用アイコン

検索結果に表示されるフィールド用にアイコンが何も構成されていない場合に使用するイメージを指定できます。以下のデフォルトのフィールド用アイコンが config.properties ファイルに事前定義されています。



#### **field.defaultIcon**

検索結果内のフィールド用のデフォルトのアイコンであるイメージ・ファイルのパスと名前を指定します。デフォルトのアイコンは /images/database.gif です。

## 日付フィールド

どのフィールドが日付フィールドであるかを指定できます。ここで指定するフィールド名は、検索結果内で日付データのようにフォーマットされます。日付のフォーマットは、Web ブラウザーでのロケール設定に一致します。

**date.fields=**スペースで区切ったフィールド名リスト

フィールド名リストを次の行に継続する場合、前の行が円記号 (¥) で終わるようにしてください。

例:

```
date.fields=modifieddate createddate
```

## 文書タイトル

文書タイトルのテキストをより分かりやすいデータで置換することにより、文書の代替タイトルを指定できます (つまり、タイトルを簡潔化 (clean) ) できます。例えば、「Slide 1」のような内容を表していないラベルが含まれた一連の文書タイトルがある場合、タイトルをそのまま表示する代わりに、検索結果で「Slide 1」を抑制

するよう指定できます (結果文書を識別できるように、より意味のあるフィールド (ファイル名など) を代わりに使用することが考えられます)。

また、文書タイトルから無意味な語を除去することにより、文書の代替タイトルを指定できます (つまり、タイトルを切り捨て (*truncate*) できます)。例えば、多数の結果文書が「Microsoft Word -」で始まっている場合、繰り返されるこのテキストを抑制することによって、検索結果を読みやすくなります。

**titles.clean=**コンマで区切ったタイトル・リスト

**titles.truncatePrefix=**コンマで区切った接頭部リスト

コンマで区切られたこれらのリストには、コンマを除き、スペースおよびその他の文字を入れることができます。リストを次の行に継続する場合、前の行が円記号 (¥) で終わるようにしてください。

以下に例を示します。

```
titles.clean=Slide 1, Layout 1, IBM Software Group Presentation Template, ¥
untitled, Untitled Document, PowerPoint Presentation, ¥
(no title for this page)
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

## ユーザー・プリファレンスのデフォルト値

検索アプリケーションの「プリファレンス」ページのデフォルト値を指定できます。ユーザーがプリファレンスを変更した場合、そのユーザーの現行セッションでのみ新しい値が有効になります。以下のプリファレンスが `config.properties` ファイルに事前定義されています。

**preferences.resultsRange=10**

検索結果の各ページに 10 個の結果文書をリストできることを指定します。

**preferences.siteCollapsing=Yes**

エンタープライズ・サーチ管理コンソールにおけるサイト縮小表示規則が構成されている場合に、URI 接頭部規則に一致する URI は検索結果で縮小表示されることを指定します。

**preferences.spellCorrections=Yes**

ミススペルの可能性がある語を含んでいる照会をユーザーが実行した場合にスペル修正候補が表示されることを指定します。

**preferences.extendedHighlighting=No**

文書サマリーのフィールドに加えて、追加のフィールド (文書タイトルなど) で、照会用語が強調表示されないことを指定します。

## デフォルトのコレクションおよび外部ソース

「プリファレンス」ページと「拡張検索」ページで事前選択されるコレクションおよび外部ソースを指定できます。ユーザーはこのデフォルトのセットを編集して、デフォルトで使用可能にされているより少ないコレクションおよび外部ソースを検索対象にできます。コレクションおよび外部ソースのセットをここで制限すると、ユーザーはプリファレンスまたは拡張検索オプションを変更する際に、検索アプリケーションで使用可能になっているコレクションまたは外部ソースのうち任意のものを選択できます。

**preferences.defaultCollections=\***

**preferences.defaultCollections=**スペースで区切ったコレクション ID リスト

すべてのコレクションおよび外部ソースを検索対象として使用可能にするには、アスタリスク (\*) を指定します (コレクションおよび外部ソースは、エンタープライズ・サーチ管理コンソールで、検索アプリケーションと関連付けられていなければなりません)。これは、`config.properties` ファイルでのデフォルト設定です。

プリファレンスまたは拡張検索オプションを変更していないユーザーが何を検索対象とするのかを制限するには、デフォルトでユーザーが検索するようにしたいコレクションおよび外部ソースのコレクション ID を指定します。

以下に例を示します。

```
preferences.defaultCollections=*
preferences.defaultCollections=coll_id1 coll_id2
```

## 検索結果の追加情報

検索結果で提供される情報量をカスタマイズすることができ、また、ユーザーが検索結果をフィルターできるかどうかを制御できます。`config.properties` ファイル中のデフォルト設定は、次のとおりです。

**refreshButton.show=false**

「リフレッシュ」ボタンが表示されるかどうかと、検索アプリケーションをリフレッシュすることができるかどうかを制御します。このオプションを `true` に設定すると、ユーザーは検索に使用可能なコレクションおよび外部ソースのリストをリフレッシュできます (例えば、検索アプリケーションがエンタープライズ・サーチ管理コンソールで追加のコレクションまたは外部ソースと関連付けられた場合)。

「リフレッシュ」ボタンが表示されるようにする場合として、`config.properties` ファイルに加えた変更をテストする場合などが考えられます。変更を保管した後、「リフレッシュ」をクリックすると、その変更が検索アプリケーションにどのように影響するかを確認できます。「リフレッシュ」ボタンがない場合、WebSphere Application Server で `ESSearchApplication` エンタープライズ・アプリケーションを再始動しないと、変更は有効になりません。

検索に使用可能なコレクションまたは外部ソースがない場合 (例えば、誤ったホスト名が指定されている場合や、検索サーバーが開始されていない場合、`ESSearchServer` エンタープライズ・アプリケーションが WebSphere Application Server において開始されなかった場合)、問題のトラブルシューティングを支援するため、「リフレッシュ」ボタンは自動的に表示されません。

**extraMessages.show=false**

検索結果の下部にある、警告メッセージと通知メッセージが表示される領域の表示を制御します。エラー・メッセージは、常にページの上部に表示されます。追加のメッセージを表示するには、このオプションを `true` に設定してください。

### **builtQueryString.show=false**

結果文書リストの前にある領域での、完全に展開された照会構文の表示を制御します。処理された実際の照会が表示されるようにするには、このオプションを `true` に設定してください。

### **refineResults.show=true**

ユーザーが追加の照会用語を指定して検索結果を絞り込むことができるかどうかを制御します。このオプションを `true` に設定すると、「**結果内の検索**」というラベルが付いた照会ボックスが、検索結果ページの下部に表示されます。

### **filter.showOnTwoLines=true**

ソース・タイプによる結果のフィルター処理とファイル・タイプによる結果のフィルター処理に関するオプションが検索結果に 1 行または 2 行で表示されるかどうかを制御します (ユーザーは、検索結果の表示中に、ソース・タイプの選択とファイル・タイプの選択を行って、選択したフィルターに一致する結果文書のみを表示できます)。

検索結果の表示に使用できるスペース量を最大化するには、このプロパティを `false` に設定します。使用可能なフィルターが 1 行を超える場合は特に、このプロパティを `true` に設定して各フィルターが別の行に表示されるようにすると、フィルターの読みやすさが増します。

## **カスタム・バナーおよびロゴ**

検索アプリケーションの上部にあるバナー領域に表示されるイメージをカスタマイズできます。例えば、WebSphere II OmniFind Edition のデフォルトのイメージを、ご使用のエンタープライズ・ブランドを反映するイメージで置き換えることができます。バナーが表示されないようにするには、以下の一方または両方の行をコメント行にしてください。 `banner.icon` プロパティは、バナー領域の左側に表示されるグラフィックを示します。 `banner2.icon` プロパティは、バナー領域の右側に表示されるグラフィックを示します。

```
banner.icon=/images/WS_II_OFEdition.gif  
banner2.icon=/images/WS_II_mosaic.gif
```

## **カスタム背景イメージ**

検索アプリケーションのページ背景に表示されるイメージをカスタマイズできます。例えば、WebSphere II OmniFind Edition のデフォルトのイメージを、ご使用のエンタープライズ・ブランドを反映するイメージで置き換えることができます。背景イメージがページに表示されないようにするには、以下の行の 1 つ以上をコメント行にしてください。

```
search.backgroundImage=/images/IIOF_search.gif  
preferences.backgroundImage=/images/IIOF_options.gif  
advanced.backgroundImage=/images/IIOF_advanced.gif  
browse.backgroundImage=/images/IIOF_tree.gif  
myProfile.backgroundImage=/images/IIOF_profile.gif  
logout.backgroundImage=/images/IIOF_logout.gif
```

## **リンク**

`config.properties` ファイルの `Links` 領域にあるプロパティは、検索アプリケーション・ページの名前が、ツールバーおよびタブ付きページに表示される代わりに、各

ページにリンクとして表示されるようにします。リンクの表示は、検索アプリケーションをポートレットとして実行し、ポータル・ページで検索アプリケーションの表示に使用されるスペース量を最小化する場合に役立ちます。

ツールバーおよびタブ付きページでオプションを選択することによって検索アプリケーションをナビゲートするほうが望ましい場合は、これらの行をコメント化してください。

## 検索タブ

`config.properties` ファイルの `Search tabs` 領域にあるプロパティは、検索アプリケーションの「検索」ビュー内のタブ付きページ（「検索」、「拡張検索」、および「カテゴリー・ツリー」）に使用される `Java Server Page (JSP)` の名前を指定します。Java プログラミングおよび `JSP` に精通している場合以外は、これらのページを編集しないでください。

この領域のカスタマイズ例には、以下のものがあります。

- 検索アプリケーションを、タブ付きページに別の外観を提供するカスタム `JSP` に誘導する。
- カテゴリー・ツリーに関するエントリをコメント化する。例えば、コレクションに対してカテゴリーを構成しない場合、検索アプリケーションで「カテゴリー・ツリー」ページを表示する必要はありません。
- タブ付きページに関するエントリを `config.properties` ファイルの `Toolbars` 領域にコピーし、これらの行をコメント化する。例えば、ツールバーのみが表示されればよく、タブ付きページは全く表示する必要がない場合があります。

## ツールバー

`config.properties` ファイルの `Toolbars` 領域にあるプロパティは、検索アプリケーションのツールバーに使用される `Java Server Page (JSP)` の名前を指定します。Java プログラミングおよび `JSP` に精通している場合以外は、これらのページを編集しないでください。

この領域のカスタマイズ例には、以下のものがあります。

- 検索アプリケーションを、ツールバーに別の外観を提供するカスタム `JSP` に誘導する。
- 表示されないようにする項目について、ツールバーに関するエントリをコメント化する。例えば、「製品情報」ページへのリンクをツールバーに組み込む必要がない場合があります。
- 「拡張検索」ページを表示する機能を `config.properties` ファイルの `Search tabs` 領域から移動し、このオプションをツールバーでのみ使用可能にする。

## 分かりやすい文書タイプのラベル

実際の文書タイプ名を、より正確で分かりやすい語にマップすることによって、文書タイプ・フィルターの読みやすさを向上させることができます。検索アプリケーションで使用可能な文書タイプは、検索および索引 `API (SI-API)` の `AvailableDocumentTypes` クラスで定義されています。利便性のため、使用可能な文書タイプは、`config.properties` ファイルの最後にもリストされています。



**documentType.**ラベル=スペースで区切った文書タイプのリスト

検索結果で文書タイプ・フィルター行に表示される名前と、ユーザーがそのフィルターを選択したときに表示される実際の文書タイプのリストを指定します。

例えば、ラベル `html` を指定し、さまざまな Web 文書のファイル拡張子および MIME タイプをその名前にマップします。ユーザーが検索結果をフィルターに掛けるために `html` をクリックすると、指定された拡張子および MIME タイプの文書のみが表示されます。

文書タイプの以下のマッピングが `config.properties` ファイルに事前定義されています。

```
documentType.html=shtml text/html html xhtml htm
documentType.doc=doc application/msword
documentType.ppt=application/mspowerpoint ppt
documentType.xls=xls application/x-excel application/msexcel ¥
application/x-msexcel application/excel application/vnd.ms-excel
documentType.xml=xml text/xml
documentType.txt=txt text/plain
documentType.pdf=pdf application/pdf
```

## カスタム・フィルター

検索結果の表示をフィルター操作するためのカスタム照会を指定できます。

**filterCustom.**ラベル =照会用語

検索結果でカスタム・フィルター行に表示される名前と、ユーザーがそのフィルターを選択したときに検索結果を絞り込むための照会を指定します (ユーザーは、検索結果の表示中に、カスタム・フィルターを選択して、事前定義済み照会に一致する結果文書のみを表示できます)。

以下の例では、人的資源 (`hr`) データベースに属している文書のみが表示されるよう、検索結果がフィルターに掛けられます。

```
filterCustom.HR_database_only=databasename::hr
```

ユーザーが検索結果をフィルターに掛けるために `HR_database_only` をクリックすると、照会 `databasename::hr` が処理されます。検索結果が表示されるとき、`hr` データベースからの文書のみがリストされます。

`config.properties` ファイルには、例として幾つかのカスタム・フィルターがコメント化されて含まれています。

---

## サンプル検索アプリケーション・プロパティの編集

エンタープライズ・サーチのサンプル検索アプリケーションは、システムにあるすべてのアクティブなコレクションおよび外部ソースを検索できます。プロパティ・ファイルを編集して、ご使用の Web サーバー環境に合うようにオプションを指定したり、デフォルト・アプリケーションとして別の検索アプリケーションを使用したり、検索アプリケーションの開始時に表示されるオプションを制御したりできます。

このタスクについて

インストール・プログラムは、エンタープライズ・サーチ用の検索サーバー上の IBM WebSphere Application Server に、エンタープライズ・サーチのサンプル検索アプリケーションをデプロイします。この検索アプリケーションを構成するには、そのアプリケーションと一緒にデプロイされるプロパティー・ファイル `config.properties` を編集します。

変更を有効にするには、WebSphere Application Server 内の `ESSearchApplication` エンタープライズ・アプリケーションを停止し、再始動する必要があります。

## 手順

サンプル検索アプリケーションのプロパティーを編集するには、次のようにします。

1. 標準のテキスト・エディターで、`config.properties` ファイルを編集します。

`config.properties` ファイルは、以下の場所にインストールされています。ここで、`ES_INSTALL_ROOT` は、検索サーバー上の WebSphere II OmniFind Edition インストール・ディレクトリーです。

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/  
ESSearchApplication.war/WEB-INF/config.properties
```

2. ご使用の Web サーバー環境に関する情報および検索プリファレンスを指定するプロパティーを編集し、ファイルを保管してクローズします。(ファイル中のポンド記号 (#) は、コメント行を示します。)
3. `ESSearchApplication` アプリケーションを停止し、再始動します。
  - a. 検索サーバー上で、WebSphere Application Server 管理コンソールを開始します。

管理コンソールは、以下の方法でオープンできます。

- Windows の「スタート」メニューを使用して、該当するプログラムを選択します。
  - WebSphere Application Server バージョン 5 の場合、Web ブラウザーをオープンし、`http://hostname:port/admin` にアクセスします。ここで、`hostname` は検索サーバーのホスト名であり、`port` は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9090 です。
  - WebSphere Application Server バージョン 6 の場合、Web ブラウザーをオープンし、`http://hostname:port/ibm/console` にアクセスします。ここで、`hostname` は検索サーバーのホスト名であり、`port` は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9060 です。
- b. ユーザー ID とパスワードの入力を求めるプロンプトが出たら、WebSphere Application Server ユーザー・レジストリーに登録されているユーザー名とパスワードを入力します。ユーザー・レジストリーとしてローカル・オペレーティング・システムを使用している場合は、エンタープライズ・サーチ管理者 ID とパスワードを指定できます。
  - c. 管理コンソールにログインしたら、「アプリケーション」をクリックし、さらに「エンタープライズ・アプリケーション」をクリックします。

- d. ESSearchApplication のチェック・ボックスを選択して、「**停止**」をクリックします。
- e. アプリケーションが停止したら、そのアプリケーションのチェック・ボックスをもう一度選択して、「**開始**」をクリックします。

---

## サンプル検索アプリケーションへのアクセス

Web ブラウザーで URL を指定することにより、サンプル検索アプリケーションにアクセスします。

### 始める前に

サンプル検索アプリケーションは、ご使用の Web サーバー環境に合わせて構成する必要があります。

### このタスクについて

サンプル検索アプリケーションは、エンタープライズ・サーチ用の検索サーバーにインストールされています。このアプリケーションをそのまま使用してコレクションおよび外部ソースをテストし、テストが終わった後で、それらをユーザーに対して使用可能にすることができます。また、このアプリケーションをモデルとして使用して、独自の検索アプリケーションを作成することもできます。

### 手順

サンプル検索アプリケーションを開始するには、次のようにします。

1. Web ブラウザーで検索アプリケーションの URL を入力します。以下に例を示します。

```
http://SearchServer.com/ESSearchApplication/
```

SearchServer.com は、検索サーバーのホスト名です。

Web サーバーがポート 80 を使用するように構成されていない場合は、正しいポート番号を指定することも必要です。例えば、次のように指定します。

```
http://SearchServer.com:9080/ESSearchApplication/
```

2. WebSphere Application Server でセキュリティーが使用可能になっている場合は、有効なユーザー ID とパスワードを使ってアプリケーションにログインします。

検索アプリケーションで使用可能なコレクションのいずれかでセキュリティーが使用可能に設定されている場合や、照会処理中にユーザー信用証明情報を検証するよう構成されたクローラーがセキュア・コレクションに含まれている場合は、ユーザー・プロファイルを構成できます。「マイ・プロファイル」ページで、セキュア・ドメインにアクセスするための信用証明情報を指定してください。そうすると、ログインするよう指示されることなく、それらのドメインを検索できます。

3. 「検索」ページで、検索したいコレクションおよび外部ソースを選択し、照会を実行依頼します。

## サンプル検索アプリケーションに対するセキュリティの使用可能化

WebSphere Application Server でグローバル・セキュリティを使用可能に設定し、サンプル検索アプリケーションを使用してセキュア・コレクションを検索しようとする場合は、サンプル検索アプリケーションおよび WebSphere Application Server で構成設定を変更する必要があります。

### 始める前に

- エンタープライズ・サーチ管理者役割のメンバーである必要があります。
- WebSphere Application Server でグローバル・セキュリティを使用可能に設定する必要があります。 WebSphere II OmniFind Edition をマルチサーバー構成としてインストールした場合は、各検索サーバーでグローバル・セキュリティを使用可能に設定します。グローバル・セキュリティを使用可能に設定する方法については、WebSphere Application Server 資料を参照してください。
- WebSphere Application Server 内のユーザー・レジストリーに対して Lightweight Directory Access Protocol (LDAP) を選択する場合、エンタープライズ・サーチ管理者 ID とパスワードを LDAP レジストリーに追加する必要があります (この ID とパスワードは、WebSphere II OmniFind Edition のインストール時に指定されたものです)。

例えば、エンタープライズ・サーチ管理者 ID が adminUser の場合、LDAP レジストリーのユーザー項目は uid=adminUser,ou=Employees,o=IBM,c=US です。手順については、LDAP サーバーの資料を参照してください。

### 手順

エンタープライズ・サーチのサンプル・アプリケーションに対してセキュリティを使用可能にするには、次のようにします。

1. 管理コンソールで検索アプリケーションのプロパティを更新します。
  - a. エンタープライズ・サーチ管理コンソールに、エンタープライズ・サーチ管理者特権をもつユーザーとしてログインします。
  - b. ツールバーで「**セキュリティ**」をクリックします。
  - c. 「検索アプリケーション」ページで、「**検索アプリケーションの構成**」をクリックします。
  - d. 「**検索アプリケーションの追加**」をクリックし、「**検索アプリケーション名**」フィールドに、WebSphere II OmniFind Edition のインストール時に指定されたエンタープライズ・サーチ管理者 ID を入力します。
  - e. 「**すべてのコレクションと外部ソース**」が選択されていることを確認し、「**OK**」をクリックします。
2. 次のようにして、config.properties ファイルを編集します。
  - a. UNIX を使用している場合は、コンソール・ウィンドウを開きます。Microsoft Windows を使用している場合は、コマンド・プロンプト・ウィンドウを開きます。
  - b. サンプル検索アプリケーションの WEB-INF ディレクトリーに移動します。以下の例は、読みやすいように 2 行に示してありますが、コマンドは、1 行に指定してください。

## UNIX:

```
cd $ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/  
ESSearchApplication.war/WEB-INF
```

## Windows:

```
cd %ES_INSTALL_ROOT%\installedApps\ESSearchApplication.ear\  
ESSearchApplication.war\WEB-INF
```

- c. テキスト・エディターを使用して config.properties ファイルを編集します。
  - d. username プロパティを有効な WebSphere Application Server ユーザーの名前に変更します。
  - e. password プロパティを、指定されたユーザー ID のパスワードに変更します。
  - f. ファイルを保管して、終了します。
3. WebSphere Application Server で ESSearchApplication アプリケーションを再始動します。
- a. 検索サーバー上で、WebSphere Application Server 管理コンソールを開始します。

管理コンソールは、以下の方法でオープンできます。

- Windows の「スタート」メニューを使用して、該当するプログラムを選択します。
  - WebSphere Application Server バージョン 5 の場合、Web ブラウザーをオープンし、`http://hostname:port/admin` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9090 です。
  - WebSphere Application Server バージョン 6 の場合、Web ブラウザーをオープンし、`http://hostname:port/ibm/console` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9060 です。
- b. ユーザー ID とパスワードの入力を求めるプロンプトが出たら、WebSphere Application Server でグローバル・セキュリティーを使用可能にしたときに指定した管理者 ID とパスワードを入力します。
  - c. 管理コンソールにログインしたら、「アプリケーション」をクリックし、さらに「エンタープライズ・アプリケーション」をクリックします。
  - d. ESSearchApplication のチェック・ボックスを選択して、「停止」をクリックします。
  - e. アプリケーションが停止したら、ESSearchApplication のチェック・ボックスをもう一度選択して、「開始」をクリックします。





---

## エンタープライズ・サーチ外部ソース

外部ソースとは、そのデータ・ソース内の文書のクロール、構文解析、索引作成を必要とせずにエンタープライズ検索アプリケーションでの検索が可能なデータ・ソースです。

以下のタイプのデータ・ソースを外部ソースとして検索できます。

- Java Database Connectivity (JDBC) データベース表 (IBM DB2 Universal Database (DB2 UDB) および Oracle データベースのみ)。JDBC データベース内の各表ごとに、個別の外部ソースが作成されます。
- Lightweight Directory Access Protocol (LDAP) サーバー。各 LDAP サーバーごとに、1 つの外部ソースが作成されます。

外部ソースに関する情報を構成した後、その外部ソースを少なくとも 1 つの検索アプリケーションと関連付ける必要があります。そうすると、ユーザーは、データのクロール、構文解析、および索引作成によってエンタープライズ・サーチ用に作成されたコレクションを照会すると同時に、その外部ソースを検索できるようになります。

### 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API フェデレーター』

---

## システムへの外部ソースの追加

外部ソースをエンタープライズ・サーチ・システムに追加するときには、追加するソースのタイプを指定します。ウィザードのガイドに従って、データ・ソースに関する情報、および、データ・ソースがどのように検索できるのかについての情報を指定します。

### 始める前に

システムに外部ソースを追加するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### 制約事項

Oracle データベースを外部ソースとして検索するには、エンタープライズ・サーチ用の検索サーバーに Oracle クライアント・プログラムがインストールされていなければなりません。

### このタスクについて

外部ソースに関する情報をシステムに追加するときに、ユーザーがエンタープライズ検索アプリケーションでそのソースを照会できるようにします。Lightweight Directory Access Protocol (LDAP) サーバーおよび Java Database Connectivity (JDBC) データベース表を、検索対象として使用可能にできます。

LDAP サーバーを構成するときには、ウィザードのガイドに従って、そのサーバーにシステムが接続するための情報を指定し、そのサーバーがどのように検索されるのかに関するオプションを指定します。

JDBC サーバーを構成するときには、ウィザードのガイドに従って、そのサーバーにシステムが接続するための情報を指定し、検索に使用可能にする表を選択し、それらの表に含まれるデータがどのように検索されるのかに関するオプションを指定します。システムに追加する各表ごとに、個別に検索可能な外部ソースが作成されます。

## 手順

外部ソースをシステムに追加するには、次のようにします。

1. 「外部ソース」をクリックして「外部ソース」ビューを開きます。
2. 「外部ソースの追加」をクリックします。
3. 追加する外部ソースのタイプとして、LDAP サーバーまたは JDBC データベースのいずれかを選択します。
4. 外部ソースの構成を始めるため、「次へ」をクリックします。

作成するソースのタイプに対応するウィザードが開きます。ウィザードのプロンプトに従って、外部ソースを構成してください。ウィザードのどのページでも、「ヘルプ」をクリックすると、指定可能なオプションの説明が表示されます。

DB2 Universal Database (DB2 UDB) および Oracle データベースの接続情報を構成する際、以下に示すデフォルトの JDBC ドライバー名とロケーションを参考にしてください。

### DB2: レガシー JDBC ドライバー

ドライバ名: `COM.ibm.db2.jdbc.app.DB2Driver`

ロケーション: `db2_install_root/java/db2java.zip`

### DB2: ユニバーサル JDBC ドライバー

ドライバ名: `com.ibm.db2.jcc.DB2Driver`

ロケーション:

`db2_install_root/java/db2jcc.jar`

`db2_install_root/java/db2jcc_license_cu.jar`

### Oracle ドライバ名: `oracle.jdbc.driver.OracleDriver`

ロケーション:

`oracle_home/jdbc/lib/classes12.zip`

`oracle_home/jdbc/lib/nls_charset12.zip`

5. 外部ソース検索に関するオプションの指定が終わったら、「完了」をクリックします。

新しい外部ソースが、システムに以前に追加された他の外部ソースと共に、「外部ソース」ビューにリストされます。

## 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API フェデレーター』

## 外部ソースと検索アプリケーションの関連付け

外部ソースは、少なくとも 1 つの検索アプリケーションを関連付けないと検索できません。

### 始める前に

検索アプリケーションを、検索できる外部ソースと関連付けるには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### 手順

検索アプリケーションを 1 つ以上の外部ソースに関連付けるには、次のようにします。

1. 管理コンソールのツールバーで「**セキュリティ**」をクリックします。
2. 「検索アプリケーション」ページで、「**検索アプリケーションの構成**」をクリックします。
3. 「検索アプリケーションの構成」ページで、「**検索アプリケーションの追加**」をクリックします。
4. 検索アプリケーションの名前を入力します。
5. そのアプリケーションが検索できる外部ソースを次のように選択します。
  - システムに追加するすべての外部ソースを検索アプリケーションがアクセスできるようにするには、「**すべてのコレクションと外部ソース**」をクリックします。
  - 指定する外部ソースのみを検索アプリケーションがアクセスできるようにするには、「**特定のコレクションと外部ソース**」をクリックします。

このオプションを選択すると、コレクション名および外部ソース名のリストが表示されます。アプリケーションが検索できる外部ソースの「**選択**」チェック・ボックスを選択します。

6. 「**OK**」をクリックします。

### 関連概念

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『**検索および索引 API フェデレーター**』



---

## エンタープライズ・サーチ・セキュリティ

エンタープライズ・サーチのセキュリティ機構によって、無許可でソースが検索されないようにし、管理機能の使用を特定のユーザーに制限することができます。

エンタープライズ・サーチによって、ユーザーは広い範囲のデータ・ソースを検索できます。コンテンツに対するアクセス許可を持つユーザーのみがアクセスすることを確認し、また許可されたユーザーのみが管理コンソールにアクセスできることを確実にするために、エンタープライズ・サーチは複数のレベルにおいてセキュリティの調整および実行を行います。

### Web サーバー

セキュリティの第 1 レベルは Web サーバーです。WebSphere Application Server でグローバル・セキュリティを使用可能にすると、ユーザーに管理役割を割り当てたり、システムを管理しようとするユーザーを認証することができます。ユーザーが管理コンソールにログインすると、そのユーザーが管理することを許可された機能とコレクションのみが使用可能になります。

また、検索アプリケーションは WebSphere Application Server の認証サポートを使用して、検索アプリケーションでコレクションを検索するユーザーを認証できます。

### コレクション・レベルのセキュリティ

コレクションを作成するときに、コレクション・レベルのセキュリティを使用可能にできます。コレクションの作成後にこの設定を変更することはできません。コレクション・レベルのセキュリティを使用可能にしていない場合は、あとで文書レベルのセキュリティ制御を指定することはできません。

コレクション・レベルのセキュリティが使用可能な場合に、グローバル分析処理で以下の特別な規則を適用します。

- 文書ごとにセキュリティ制御が評価されるように、重複 (または重複に近い) のコンテンツの文書は、そのコンテンツが正規表現で関連付けられて索引が付けられるのではなく、別個に索引付けされます。
- グローバル分析のアンカー・テキスト処理フェーズは、通常、ある文書 (ソース文書) に現れるテキストを、そのテキストが必ずしも現れるとは限らない別の文書 (ターゲット文書) に関連付けます。これにより、ソース文書に現れるテキストを指定した照会で、ターゲット文書が検索されるようになります。このタイプのアンカー・テキスト処理は、ユーザーがソース文書は表示できないがターゲット文書の表示は許可されている場合に、セキュリティのリスクが生じます。コレクション・セキュリティが使用可能である場合、禁止文書へのリンクのアンカー・テキストは索引から除外されます。文書固有のコンテンツまたはメタデータが照会に一致している場合のみ、文書が検索結果に戻されます。

コレクション・セキュリティを使用可能にすることと、検索の質とでトレードオフが生じます。コレクション・セキュリティを使用可能にすると、

文書ごとに索引付けされる情報量が減ります。この副次作用は、照会によっては結果の数が少なくなることです。

コレクション・レベルのセキュリティーは、アプリケーション ID を通して検索アプリケーションでも使用できます。コレクションを検索するためにエンタープライズ・サーチ管理者は、検索アプリケーションを、それが検索できる特定のコレクションに関連付ける必要があります。そして、標準アクセス制御機構を使用して、ユーザーに対し、検索アプリケーションへのアクセスを許可したり拒否したりできます。

### 文書レベル・セキュリティー

コレクションのクローラーを構成するときに、文書レベルのセキュリティーを使用可能にできます。このオプションを選択すると、クローラーは、クローラーする各文書にセキュリティー・トークンに関連付けることができます。セキュリティー・トークンは文書と一緒に索引の中に保管されます。一部のクローラーでは、照会処理中に、ユーザーの信用証明情報を現行のアクセス制御データで (ネイティブ・データ・ソースで構成されたのと同様に) 検証するように指定できます。

検索アプリケーションではセキュリティー・トークンとユーザーの信用証明情報を使用して、アクセス制御を施行することができます。ユーザーが、アクセス許可を持つ文書のみを検索して取得するために、検索アプリケーションが、検索サーバーに渡す照会にログイン・ユーザーの信用証明情報を組み込むことができます。

コレクションに対するセキュリティーは、索引付けされたコンテンツを保護するためにエンタープライズ・サーチが使用できる認証とアクセス制御機構を越えて拡張されます。転送中に、悪意のある、あるいは無許可のユーザーにデータをアクセスさせないための安全機能も存在します。例えば、検索サーバーは、Secure Sockets Layer (SSL)、Secure Shell (SSH)、および Secure Hypertext Transfer Protocol (HTTPS) といったプロトコルを使用して、索引サーバーおよび検索アプリケーションと通信します。

暗号化によってさらにセキュリティーが強化されます。例えば、エンタープライズ・サーチ管理者のパスワードが製品のインストール時に指定されますが、これは、暗号化されたフォーマットで保管されます。ユーザーがユーザー・プロフィールで指定するパスワードも、暗号化されたフォーマットで保管されます。

セキュリティーを向上させるには、サーバー・ハードウェアが適切に隔離されており、無許可の侵入から保護されている必要があります。ファイアウォールをインストールすると、使用しているネットワークの別の部分を通じた侵入から、エンタープライズ・サーチ・サーバーを保護できます。また、エンタープライズ・サーチ・サーバーに、余計なポートが開いていないことも確認してください。システムが、エンタープライズ・サーチのアクティビティーとアプリケーション用に明示的に割り当てられたポートでのみ要求を listen するように、システムを構成してください。

---

## 管理役割

エンタープライズ・サーチは管理コンソールのさまざまな機能へのアクセスを制御するために、役割の概念を使用します。



WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) のインストール中に、インストーラーは、エンタープライズ・サーチ管理者のユーザー ID とパスワードを構成します。管理コンソールに最初にアクセスするときは、このユーザーとしてログインしなければなりません。WebSphere Application Server 内のグローバル・セキュリティーを使用可能に設定していない場合、このユーザー ID がエンタープライズ・サーチ管理コンソールのアクセスに使用できる唯一のユーザー ID です。

WebSphere Application Server でグローバル・セキュリティーを使用可能にしていれば、追加のユーザーをエンタープライズ・サーチ管理ユーザーとして登録できます。ユーザーを役割に割り当てることによって、アクセスを特定のコレクションに制限し、それぞれの管理ユーザーが実行できる機能を制御できます。エンタープライズ・サーチ内の管理役割に割り当てるユーザー ID は、WebSphere Application Server ユーザー登録の中になければなりません。

管理ユーザーがログインするときに、エンタープライズ・サーチはユーザー ID を認証します。ユーザーが管理できるコレクションと機能のみが、コンソールの中で選択可能になっています。

次の管理役割にユーザーを登録できます。

#### **エンタープライズ・サーチ管理者**

これらのユーザーは、コレクションを作成し、エンタープライズ・サーチ・システムのすべての側面を管理する権限を持ちます。WebSphere II OmniFind Edition をインストールする際に、最初のエンタープライズ・サーチ管理ユーザーのユーザー ID とパスワードを指定します。最初にログインした後に、このユーザーは、他のユーザーをエンタープライズ・サーチ管理者の役割に割り当てることができます。

#### **コレクション管理者**

これらのユーザーは、特定のコレクションまたはすべてのコレクションの操作を、編集、モニター、および制御できます。これらのユーザーは、コレクションを作成することはできません。エンタープライズ・サーチ管理者がコレクション管理者を認可した場合のみ、コレクション管理者は、システム・レベルのアクティビティーをモニターおよび操作することができます。

#### **オペレーター**

これらのユーザーは、特定のコレクションまたはすべてのコレクションの動きをモニターし、制御することができます。これらのユーザーは、コレクション・アクティビティーの開始と停止はできますが、コレクションの作成や編集はできません。エンタープライズ・サーチ管理者がオペレーターを認可した場合のみ、オペレーターは、システム・レベルのアクティビティーをモニターおよび操作することができます。

#### **モニター担当者**

これらのユーザーは、特定のコレクションまたはすべてのコレクションをモニターすることができます。操作の制御 (サーバーの開始や停止など)、コレクションの作成、またはコレクションの編集はできません。エンタープライズ・サーチ管理者がモニター担当者を認可した場合のみ、モニター担当者は、システム・レベルのアクティビティーを監視することができます (しかし、操作はできません)。

## 関連タスク

19 ページの『管理コンソールへのログイン』

エンタープライズ・サーチ・システムを管理するには、Web ブラウザーで URL を 指定した後で、管理コンソールにログインします。

251 ページの『エンタープライズ・サーチ・サーバーの開始』

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

---

## 管理ユーザーの構成

管理役割を構成することで、コレクションへのアクセスを制限でき、各管理ユーザーが実行できる機能を制御できます。

### 始める前に

管理役割にユーザーを割り当てる前に、WebSphere Application Server でセキュリティーが使用可能にされていることを確認してください。また、WebSphere Application Server ユーザー・レジストリー内にユーザー ID が存在することを確認してください。

管理ユーザーを構成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### 手順

ユーザーを管理役割に割り当てるには、次のようにします。

1. 「**セキュリティー**」をクリックして「セキュリティー」ビューを開きます。
2. 「管理役割」ページで、「**ユーザーの追加**」をクリックします。
3. 登録したいユーザーのユーザー ID を入力し、適切な管理役割を選択します。
4. このユーザーをエンタープライズ・サーチ管理者として登録しない場合は、このユーザーが「**システム**」ツールバーからページにアクセスできるかどうかを指定します。

例えば、一部のオペレーターまたはコレクション管理者に、システム・レベルのログ・ファイルのモニターを許可したい場合があります。

5. このユーザーをエンタープライズ・サーチ管理者として登録しない場合は、このユーザーが管理できるコレクションおよび外部ソースを選択します。

個々のコレクションおよび外部ソースのチェック・ボックスを選択するか、あるいはこのユーザーがすべてのコレクションおよび外部ソースを管理できるようにします。

---

## 認証とアクセス制御

無許可のユーザーからコンテンツを保護し、管理機能へのアクセスを制御するために、エンタープライズ・サーチは、ユーザー認証とアクセス制御をサポートします。

## 認証

認証は、ユーザーが誰であるか、つまり、ユーザーが宣言するものをシステムが検証するプロセスです。アクセスは一般にリソースを要求するユーザーの ID を基にしているため、認証は、セキュリティーを実効あるものにするために不可欠です。

管理コンソールにアクセスしようとするユーザーを認証するため、エンタープライズ・サーチは、WebSphere Application Server で提供される認証サポートを利用します。

エンタープライズ・サーチ・コレクションを検索するユーザーを認証するために、検索アプリケーションは WebSphere Application Server 内のセキュリティーを利用し、ユーザーの信用証明情報の認証のための優先メソッドをインプリメントできます。通常、ユーザーの信用証明情報は、ユーザーがログインするとき、または検索アプリケーションにアクセスするときに検索アプリケーションに渡されるユーザー ID とパスワードから成ります。

ユーザー認証は、ユーザーのエンタープライズで使用可能なリソースおよびプロトコルに応じて、他の方法でインプリメントされることも可能です。例えば、ユーザーの識別をスマート・カードで行うようにしたり、デジタル証明書と Public Key Infrastructure で行ったり、あるいは、ユーザーがログインするときにチケットを割り当ててユーザーの認証状態をトラッキングするようにしたりできます。

## アクセス制御

アクセス制御とは、ユーザーが身元を明らかにし、それが認証されたあとで、ユーザーが行えることを制限することです。アクセス制御リスト (ACL) は、リソースへのアクセスを制限するための最も一般的な方法です。ACL はユーザー識別 (ユーザー名、グループ名、ユーザー役割など) のリストです。各ユーザー識別は、ユーザーの権限と特権を定義する許可のセットに関連付けられています。

例えば、アクセス制御によって、ファイル・サーバー上のファイルへのアクセスを許可したり拒否したりできます。また、アクセスを許可されたユーザーが、ファイル・サーバー上のファイルの読み取り、作成、編集、または削除ができるかどうかを制御できます。

エンタープライズ・サーチの中で、すべてのアクセス制御は、ユーザーが索引の中のデータを読み取る権限があるかどうかによって異なります。検索アプリケーションの中でどのようにアクセス制御を有効にするかに応じて、また、エンタープライズ・サーチを管理するときにコレクションとクローラーに指定した規則に応じて、次のことができます。

- すべてのユーザーがコレクションの中のすべての文書を検索できる。
- すべてのユーザーが、特定のクローラーによってクロールされたすべての文書を検索できる。また、他のクローラーによってクロールされた文書へのアクセスは制限される。
- 特定のユーザーが、特定の文書を検索できる。例えば、Lotus Notes クローラーでクロールするデータベースを指定するときに、数人のユーザーが特定のビューおよびフォルダー内の文書を検索できるというオプションを指定し、他のユーザーがこれらの文書を検索できないようにします。

## WebSphere Application Server 内でのエンタープライズ・アプリケーションのセキュリティーの使用不可化

ユーザー認証を必要とする WebSphere II OmniFind Edition アクティビティーを制御するために、WebSphere Application Server 内の個々のエンタープライズ・アプリケーションについてグローバル・セキュリティーを使用不可にすることができます。

### このタスクについて

WebSphere II OmniFind Edition インストール・プログラムは、次の 3 つのエンタープライズ・アプリケーションを WebSphere Application Server にデプロイします。

- ESAdmin アプリケーション。エンタープライズ・サーチ管理コンソールのインターフェースが含まれています。
- ESSearchApplication アプリケーション。サンプル検索アプリケーションのインターフェースが含まれています。
- ESSearchServer アプリケーション。WebSphere II OmniFind Edition SIAPI インプリメンテーションのすべての遠隔通信を提供し、SIAPI インターフェースが検索サーバーと通信できるようにします。

デフォルトで、3 つのエンタープライズ・アプリケーションはすべて、WebSphere Application Server グローバル・セキュリティーをサポートします。これらのアプリケーションは、グローバル・セキュリティーが使用可能であることを検出すると、受信したすべての要求の認証を始めます。

組織によっては、特定の WebSphere II OmniFind Edition エンタープライズ・アプリケーションについてセキュリティーを使用可能または使用不可にしたい場合があります。例えば、エンタープライズ・サーチ管理コンソールにアクセスするすべてのユーザーを認証したいが、WebSphere II OmniFind Edition SIAPI インターフェースまたはサンプル検索アプリケーションを使用するユーザーは認証したくない場合があります。

### 手順

特定のエンタープライズ・アプリケーションについてセキュリティーを使用不可にするには、次のようにします。

1. 検索サーバー上で、WebSphere Application Server 管理コンソールを開始します。

管理コンソールは、以下の方法でオープンできます。

- Windows の「スタート」メニューを使用して、該当するプログラムを選択します。
- WebSphere Application Server バージョン 5 の場合、Web ブラウザーをオープンし、`http://hostname:port/admin` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9090 です。

- WebSphere Application Server バージョン 6 の場合、Web ブラウザーをオープンし、 `http://hostname:port/ibm/console` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9060 です。
- 2. ユーザー ID とパスワードの入力を求めるプロンプトが出たら、WebSphere Application Server でグローバル・セキュリティーを使用可能にしたときに指定した管理者 ID とパスワードを入力します。
- 3. 管理コンソールにログインしたら、「アプリケーション」をクリックし、さらに「エンタープライズ・アプリケーション」をクリックします。
- 4. セキュリティーを使用不可にしたいエンタープライズ・アプリケーションの名前の横にあるチェック・ボックスを選択します。
- 5. スクロールダウンして、「ユーザー/グループへのセキュリティー役割のマッピング」リンクをクリックします。
- 6. **AllAuthenticated** 役割を見つけて、「全利用者?」欄の下のチェック・ボックスを選択します。
- 7. 「OK」をクリックします。
- 8. 「保管」リンクをクリックして、変更を保管します。
- 9. WebSphere Network Deployment を使用している場合は、「ノードとの変更の同期化」チェック・ボックスを選択します。
- 10. 「保管」をクリックします。
- 11. もう一度「アプリケーション」をクリックしてから、「エンタープライズ・アプリケーション」をクリックします。
- 12. 今変更したアプリケーションのチェック・ボックスを選択して、「停止」をクリックします。
- 13. アプリケーションが停止したら、そのアプリケーションのチェック・ボックスをもう一度選択して、「開始」をクリックします。

---

## コレクション・レベルのセキュリティー

コレクション・レベルのセキュリティーを提供するために、コンテンツの索引付けのオプションと、検索アプリケーションが特定のコレクションを検索できるようにするオプションを構成します。

コレクションを作成するときに、コレクション・セキュリティーを使用可能にするオプションを選択できます。このオプションを選択すれば、あとで文書レベルのセキュリティー制御を構成することができます。コレクション・セキュリティーが使用可能である場合、エンタープライズ・サーチのグローバル分析処理により、重複文書や文書内のアンカー・テキストを索引付けする別の規則も適用できます。

検索アプリケーションを作成したあと、検索アプリケーション ID によって、検索アプリケーションが検索できるコレクションと、検索アプリケーションにアクセスできるユーザーを指定できます。



## 重複文書分析

コレクション・セキュリティーを使用可能にすると、グローバル分析プロセスは、コレクションの中の重複文書の識別は行いません。

グローバル分析の間、索引付けプロセスは、互いに重複している、またはほとんど重複している文書を識別します。そして、これらの文書をすべて、コンテンツの 1 つの正規表現に関連付けます。重複文書を識別することで、検索結果に、同じ (あるいはほとんど同じ) コンテンツの複数の文書が入ることがなくなります。

コレクションを作成するときにコレクション・セキュリティーを使用可能にすると、重複文書は識別されず、それらの文書は共通の正規表現に関連付けられません。代わりに、各文書は、別個に索引付けされます。これによって、ユーザーは、ユーザーの信用証明情報とマッチするセキュリティー・トークンの文書のみを検索することになります。例えば、2 つの文書のコンテンツがほとんど同じかもしれませんが、異なるアクセス制御リストを使用して、セキュリティーを施行します。

重複文書分析を使用不可にすればコレクションの中の文書のセキュリティーは強化されますが、検索結果の中に同じ文書の複数のコピーが返されると、検索の質は低下するかもしれません。

## アンカー・テキスト分析

コレクション・セキュリティーを使用可能にすると、グローバル分析プロセスは、Web クローラーでクロールされる文書の中のアンカー・テキストの索引作成に、特別な規則を適用します。コレクション・セキュリティーを使用可能にしない場合は、個別の Web クローラーを構成するときに禁止文書へのリンクのアンカー・テキストの索引を作成するかどうか指定することができます。

アンカー・テキストは、リンクの接続先のページを説明した、ハイパーテキスト・リンク内の情報です。例えば、次のリンクで、テキスト「Query Syntax」は、Web サイトの `syntax.htm` ページに接続するリンクの中の、アンカー・テキストです。

```
<a href="../doc/syntax.htm">Query Syntax</a>
```

一般に、Web クローラーは文書の中のリンクをたどって、追加の文書をクロールし、これらのリンクされたページを索引の中に組み込みます。グローバル分析の間、索引は、アンカー・テキストが組み込まれる文書 (ソース文書) だけでなく、ターゲット文書にもアンカー・テキストに関連付ける処理をします。上記の例では、アンカー・テキスト「Query Syntax」は、ターゲット・ページ `syntax.htm` とアンカー構成を含むページに関連付けられます。

コレクションを作成するときにコレクション・セキュリティーを使用可能にすると、アンカー・テキスト処理は使用不可になります。実際にそのテキストが文書または文書のメタデータに出現しない限り、そのアンカー・テキストはもはや文書に索引付けされません。このセキュリティー制御により、禁止文書へのアンカー・テキストがユーザーのアクセスできる文書に関連付けられることはないため、ユーザーがアクセスできない文書の情報を見ることはありません。

コレクション・セキュリティーを使用可能にすることで、ユーザーが、ユーザーの信用証明情報とマッチするセキュリティー・トークンの文書のみを検索できるようにし、Web 文書のセキュリティーを向上させることができます。しかし、アンカ



ー・テキストを処理しないことで、照会に関連する可能性のある文書が、一部検索結果に含まれなくなるという可能性があります。

コレクション・セキュリティーを使用可能にしない場合は、個別の Web クローラーを構成するときに禁止文書へのリンクのアンカー・テキストの索引付けをすることができます。アンカー・テキストの索引付けを指定すると、分析と索引付けのプロセスは、Web クローラーによって検索されたすべてのページのアンカー・テキストの索引付けを処理します。アンカー・テキストの索引付けをしないと指定すると、禁止文書へのリンクのアンカー・テキストは、この索引から削除されます。

## 禁止文書へのリンク内のアンカー・テキストの索引作成

Web クローラーでのクロールが禁止されている文書へのリンクが文書に含まれている場合、それらのリンク用のアンカー・テキストを索引に保存するかどうかを、Web クローラーの構成時に指定できます。

### 始める前に

アンカー・テキストの索引作成のためのオプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、構成したい Web Crawler のコレクション管理者でなければなりません。


### このタスクについて

robots.txt ファイルまたは Web 文書のメタデータ内のディレクティブにより、Web Crawler が Web サイトで文書にアクセスできないようにすることができます。Web Crawler がクロールを許可されている文書に禁止文書へのリンクが組み込まれている場合、それらのリンクのアンカー・テキストをどのように取り扱うかを指定できます。

Web クローラーを構成するときに、禁止文書へのアンカー・テキストの索引を作成するかどうかを指定することができます。最も厳しいセキュリティーを実現するためには、禁止文書へのリンク内のアンカー・テキストに索引を付けないよう指定してください。アンカー・テキストの索引を作成しないことで、照会に関連する可能性のある文書が、一部検索結果に含まれなくなるという可能性があります。

### 手順

禁止文書へのリンク内のアンカー・テキストの索引作成を使用可能または使用不可に設定するには、以下のようになります。

1. コレクションを編集します。「クロール」ページを選択し、構成したい Web クローラーを見つけて、 「クローラーのプロパティ」をクリックします。
2. 「**拡張 Web クローラー・プロパティの編集**」をクリックします。
3. このクローラーがクロールするすべての文書でアンカー・テキストに索引を付けるには、「**禁止文書へのリンク内のアンカー・テキストの索引作成**」チェック・ボックスを選択します。ユーザーは、それらのページをポイントするリンクのアンカー・テキストに入っているテキストを検索すると、Web Crawler がクロールを許可されていないページを知ることができます。

禁止文書へのアンカー・テキストを索引から除外するには、このチェック・ボックスを外します。ユーザーは、Web Crawler がクロールを許可されていないページを知ることができません。アンカー・テキストは、禁止文書と共に、索引から除外されます。

4. 「OK」をクリックし、「Web Crawler のプロパティ」ページで、もう一度、「OK」をクリックします。
5. 変更を有効にするために、クローラーを停止し、再開します。

以前に索引作成された文書の変更を適用するには、文書を再クロールして、索引をもう一度作成できるようにする必要があります。以前のクロールで禁止文書に関する情報が索引に追加されている場合、その情報は索引から除去されます。

## 検索アプリケーション ID を使用したセキュリティ

コレクション・レベルのセキュリティを提供するために、各コレクションおよび外部ソースを検索できる検索アプリケーションを指定します。

すべての検索アプリケーションは、アプリケーション ID をエンタープライズ・サーチ API に渡す必要があります。エンタープライズ・サーチ管理者と検索アプリケーションは、この ID を使用して、コレクションおよび外部ソースのセキュリティを施行できます。

検索アプリケーションがコレクションまたは外部ソースにアクセスして検索を開始する前に、エンタープライズ・サーチ管理者は、検索アプリケーションを、それが検索できる特定のコレクションおよびソースに関連付ける必要があります。検索アプリケーションには、エンタープライズ・サーチ・システム内のすべてのコレクションおよび外部ソースを検索させることも、あるいは指定したコレクションおよび外部ソースのみを検索させることもできます。

アクセス制御を施行するには、セキュリティ・トークン(例えば、ユーザー ID、グループ ID、またはユーザー役割など)を検索アプリケーションに関連付けて、それらのユーザーのみがアプリケーションにアクセスできるようにします。例えば、ユーザーの検索アプリケーションを起動する URL にアクセスを制限することができます。

検索アプリケーション ID の詳細と、ユーザーのカスタム検索アプリケーションにセキュリティ制御を組み込む方法については、エンタープライズ・サーチの Search and Index API (検索および索引 API) の説明を参照してください。

### 関連概念

189 ページの『エンタープライズ・サーチの検索アプリケーション』

検索アプリケーションは、エンタープライズ・サーチ・システム内のコレクションおよび外部ソースの検索を可能にします。検索アプリケーションはいくつでも作成できます。また、1 つの検索アプリケーションで、任意の数のコレクションおよび外部ソースを検索できます。

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API の概要』

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API のセキュリティ』

---

## 文書レベル・セキュリティー

コレクションが作成されたときに、そのコレクションに対するセキュリティーを使用可能にすると、文書レベルのセキュリティー制御を構成することができます。文書レベルのセキュリティーによって、コレクションを検索するユーザーは、見ることが許される文書にしかアクセスできないようになります。

コレクションの文書へのアクセスを制御するために、クローラーが索引用のセキュリティーを収集します。データ・ソースのうちのいくつかのタイプでは、照会をサブミットするときに、ユーザーの現行の信用証明情報も検証することができます。ユーザーの現行の信用証明情報を検証するには、カスタム検索アプリケーションにユーザー・プロファイルのサポートを組み込んで作成します。ユーザー・プロファイルを保管することで、ユーザーが信用証明情報の入力を何度も求められることなく文書にアクセスできるようにします。

### 関連概念

189 ページの『エンタープライズ・サーチの検索アプリケーション』  
検索アプリケーションは、エンタープライズ・サーチ・システム内のコレクションおよび外部ソースの検索を可能にします。検索アプリケーションはいくつでも作成できます。また、1 つの検索アプリケーションで、任意の数のコレクションおよび外部ソースを検索できます。

235 ページの『ポータル検索エンジンでの文書レベル・セキュリティー』  
IBM WebSphere ポータル検索エンジンを使用して、ユーザーがエンタープライズ・サーチ・コレクションを検索するときに文書レベルのセキュリティーを施行することができます。

「エンタープライズ・サーチ プログラミング・ガイドおよび API リファレンス」の『検索および索引 API のセキュリティー』

## 保管されたセキュリティー・トークンによる検証

コレクションが作成されたときに、そのコレクションに対するセキュリティーを使用可能にした場合、索引にセキュリティー・データを保管することで、文書レベルのセキュリティー制御を構成することができます。

デフォルトで、それぞれの文書は共用文書であるとみなされます。つまり、すべてのユーザーが検索できることを意味します。ほとんどの文書タイプで、1 つ以上のセキュリティー・トークンを文書に関連付けて、その文書付きトークンを索引に保管することで、文書レベルのセキュリティーを実現することができます。クローラーを構成するときに、そのクローラーがクロールした文書にアクセスできるユーザーを制限するために、セキュリティー・トークンを使用することを指定します。

データ・ソース・タイプにフィールドが含まれている場合、アクセス制御を行うために、そのフィールドの中の 1 つの値を使用することを指定できます。データ・ソースにフィールドがない場合、特定のフィールドの値をセキュリティー目的で使用しない場合、あるいは、指定したフィールドに、アクセス制御の実施を可能にする値が含まれない場合は、クローラーに、文書に関連付けるためのセキュリティー・トークンを定義できます。

各コレクションの管理者は、クローラーが文書に関連付けるセキュリティー・トークンを決定します。例えば、セキュリティー・トークンは、ユーザー ID、グループ

ID、ユーザー役割、または管理者が決定した、データ・ソースに有効なその他の値を示すこともあります。データ・ソース管理者がネイティブのアクセス制御リストを更新した場合、その更新済みセキュリティ制御は、次に索引が更新されるか、あるいは再編成されるまで有効になりません。

セキュリティ・トークンは、文書が、構文解析、分析、および索引付けのステージに移動するときに文書に付随します。検索アプリケーションでセキュリティを使用可能にしている場合、セキュリティ・トークンを使用して文書へのアクセスを制御できます。コレクションを検索するユーザーは、ユーザーの信用証明情報によって表示が許可される文書のみを検索できます。ユーザーの信用証明情報がセキュリティの規則をパスしないと、ユーザーはセキュリティ・トークンによって保護される文書を検索できません。

規則を Java クラスでエンコードすることにより、セキュリティ・トークンの値の判別にカスタム・ビジネス・ルールを適用することができます。クローラー・プロパティを構成するときに、文書をクロールする際にクローラーに使用させたいプラグインの名前を指定します。プラグインが追加するセキュリティ・トークンは、索引に保管され、文書へのアクセス制御に使用されます。

## 照会処理中の現行の信用証明情報の検証

コレクションの作成時にそのコレクションのセキュリティを使用可能にした場合、特定のタイプのドメインで、ユーザーが照会をサブミットしたときにユーザーの現行の信用証明情報を検証することができます。

次のタイプのクローラーを構成するときに、ユーザーの信用証明情報をネイティブのリポジトリが管理する現行のアクセス制御と比較して検証するオプションを選択できます。

- Content Edition クローラー (Documentum、FileNet Panagon Content Services、および Portal Document Manager リポジトリ・タイプのみ)
- Domino Document Manager クローラー
- Lotus Notes クローラー
- QuickPlace クローラー
- WebSphere Portal クローラー
- Windows ファイル・システム クローラー

照会の応答を返す前に、検索サーバーはネイティブ・リポジトリとやりとりをして、ユーザーの現行のアクセス権を検証してから、ユーザーがアクセス権を持っていない文書はすべて検索結果に表示しないように除去します。

この文書レベルのセキュリティ実施方法により、ユーザーの信用証明情報が、索引に保管されているセキュリティ・データではなく、現行のセキュリティ・データと比較されるので、高水準のセキュリティが提供されます。また、この方法により、複雑さに関係なくネイティブ・リポジトリのセキュリティ機構のアクセス制御が確実に行われます。文書のフィルター操作がリアルタイムに行われるため、検索結果は、検索基準に一致したそれぞれの文書ごとの最新のアクセス制御設定を反映します。

この方法のもう一つの利点は、索引のサイズに影響を与えない (セキュリティー・トークンの索引付けに追加スペースを必要としない) ということです。ただし、検証にネイティブ・リポジトリとの接続が必要なので、この方法は照会のパフォーマンスに影響を与える可能性があります。

最も厳しいセキュリティーを確保し、照会パフォーマンスへの影響を最小にするには、セキュリティー・トークンを索引に保管するオプションと現行アクセス制御を検証するオプションを組み合わせてください。ユーザーが照会をサブミットしたとき、2段階の検証が行われます。

- まず最初に、検索サーバーが索引付きセキュリティー・データを使用して、文書がクロールされたサーバーとデータベースへのアクセス権をユーザーが持っているかどうかを即時に判断します (索引は速さにおいて最適化されて、サブ秒の応答時間を実現します)。
- 次に、検索サーバーは、ユーザーがアクセスを許可されているサーバー上のドメインにある文書のみを含む一時的なリストを作成します。検索サーバーは、そのリストを使用してネイティブ・リポジトリに接続し、ユーザーがその要求文書の参照を許可されているかどうかを判断します。

あるユーザーにサーバーおよびドメインへのアクセス権がある場合、そのユーザーには文書へのアクセス権がある可能性が高いです。しかし、この最後のフィルター操作段階により、ユーザーの現行アクセス権設定に合った文書のみが検索結果に戻されるようになります。

#### 関連概念

225 ページの『Windows ファイル・システム文書に対する文書レベルのセキュリティーの施行』

Windows ファイル・システム・クローラーでクロールされた文書をユーザーが検索するときに現行の信用証明情報を検証するようにするには、クローラー・サーバーおよびMicrosoft Windows サーバーの両方でドメイン・アカウント情報を構成する必要があります。

228 ページの『Lotus Domino 文書に対する文書レベルのセキュリティーの施行』

クロールされる Lotus Notes サーバーが、Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用する場合、文書レベルのアクセス制御を施行するように、クローラー・サーバーを構成する必要があります。

#### 関連タスク

228 ページの『ユーザーの信用証明情報を検証するための Lotus Domino トラストド・サーバーの構成』

Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用する Lotus Notes クローラーによってクロールされたドキュメントのためにセキュリティーを施行するには、クロールされる Domino サーバーを、Lotus Domino トラストド・サーバーとして構成する必要があります。

## ユーザー・プロファイルおよび ID 管理

エンタープライズ・サーチのためのユーザー・プロファイルを作成することによって、ユーザーは、セキュア・ドメインの検索を可能にする信用証明情報を保管することができます。



照会がサブミットされたときにユーザーの信用証明情報の検証が必要なドメインを検索する場合、ユーザーは、ドメインへのログインで使用する信用証明情報を検索アプリケーションに提供しなければなりません。WebSphere Information Integrator OmniFind Edition ID 管理 の機能により、ユーザーは、ユーザー・プロファイル内のドメイン (いくつでも可) のために信用証明情報を保管することができます。信用証明情報は暗号化されて、エンタープライズ・サーチ・システム内に安全に保管されます。

ユーザーは、ユーザー・プロファイルを作成して、検索アプリケーションの使用中に自分の信用証明情報を登録することができます。エンタープライズ・サーチのサンプル検索アプリケーションでは、この機能を「**個人用プロファイル**」オプションで提供しています (ご使用のカスタム検索アプリケーションで、この機能を異なる方法でインプリメントしている可能性もあります)。

ユーザー・プロファイルには、検索するドメインにログインするためにユーザーが指定しなければならないさまざまな信用証明情報が保管されます。ユーザーは、以下の条件がすべて当てはまる場合に、ユーザー・プロファイルを作成することができます。

- WebSphere Application Server でグローバル・セキュリティーが使用可能になっていること。
- WebSphere II OmniFind Edition エンタープライズ・サーチ管理コンソールで ID 管理が使用可能になっていること。
- 検索アプリケーションが検索できるコレクションの少なくとも 1 つで、セキュリティーが使用可能になっていること。
- 照会がサブミットされたときにユーザーの現行の信用証明情報の検証を行うことでアクセス制御を施行するクローラーによってクロールされた文書が、少なくとも 1 つのセキュア・コレクションに含まれていること。
- 次のクローラー・タイプのうち少なくとも 1 つに対して文書レベルのセキュリティーが構成された場合、照会処理中に現行の信用証明情報を検証するためのオプションが選択されていること。
  - Content Edition (特定のタイプのリポジトリ用)
  - Domino Document Manager
  - Lotus Notes
  - QuickPlace
  - WebSphere Portal
  - Windows ファイル・システム

プロファイルには、照会処理中にユーザーの信用証明情報を検証する必要がある検索アプリケーションで使用可能なすべてのドメインがリストされます。ユーザーは、どのドメインに対して信用証明情報を保管するかを選択できます。ドメインの信用証明情報を指定しないと、そのドメイン内のデータ・ソースからの文書は検索結果から除外されます。

WebSphere II OmniFind Edition ID 管理を使用しない場合、現行の信用証明情報の検証を必要とするドメインをユーザーが照会するときに、検索アプリケーションがユーザーのセキュリティー・コンテキスト (USC) ストリングを供給する必要があります。



## ID 管理の構成

WebSphere II OmniFind Edition ID 管理を使用して、ユーザー・プロフィールを保管することができます。このプロフィールにより、ユーザーは、照会処理中にユーザーの信用証明情報を検証する必要があるドメインを検索できます。

### 始める前に

ID 管理オプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて

クローラーについて文書レベルのセキュリティー・オプションを構成する際に、照会処理中にユーザーの信用証明情報を検証するように指定した場合、ユーザーはその検証が必要なドメインを検索するときに、自分の信用証明情報を提供する必要があります。 WebSphere II OmniFind Edition ID 管理を使用すると、ユーザーは、ユーザー・プロフィールを作成して、セキュア・ドメインへのログインで使用する信用証明情報を登録することができます。信用証明情報は、 WebSphere II OmniFind Edition で管理されているセキュア・データベース内で暗号化されています。

ユーザーがセキュア・ドメインを検索するとき、検索サーバーが、保管された信用証明情報を使用してそのユーザーを認証します。信用証明情報が無効であった場合、そのソース・ドメインの文書は検索結果から除外されます。

### 手順

ID 管理を構成するには、次のようにします。

1. 「**セキュリティー**」をクリックして「セキュリティー」ビューを開きます。
2. 「検索アプリケーション」ページで、「**ID 管理の構成**」をクリックします。
3. 「ID 管理の構成」ページで、ユーザー・プロフィール内のユーザーの信用証明情報を管理するために WebSphere II OmniFind Edition を使用可能にするチェック・ボックスを選択します。
4. 「**OK**」をクリックします。

## Windows ファイル・システム文書に対する文書レベルのセキュリティーの施行

Windows ファイル・システム・クローラーでクロールされた文書をユーザーが検索するときに現行の信用証明情報を検証するようにするには、クローラー・サーバーおよびMicrosoft Windows サーバーの両方でドメイン・アカウント情報を構成する必要があります。

Windows ファイル・システム クローラーを構成するとき、ローカル・コンピューター上のサブディレクトリーまたはリモート・コンピューター上のサブディレクトリーのどちらをクロールするかを指定します。コレクションでセキュリティーが使用可能になっている場合、クロールされたサブディレクトリー内の文書へのアクセスを制御するためのオプションも指定できます。

ユーザーが照会をサブミットしたときにそのユーザーの現行の信用証明情報を検証することでアクセス制御を施行することを選択した場合、ドメイン・アカウントが

正しく構成されていることを確認してください。ローカル・コンピューター上でクローラされたファイルのドメイン・アカウントをセットアップするための要件は、リモート Windows サーバー上でクローラされたファイルのための要件とは異なります。

## ローカル・アクセス制御データでの検証

現行のユーザー信用証明情報を検証するために、システムはローカル・ユーザー・アカウント情報とドメイン・アカウント情報の両方を使用します (コンピューターが Windows ドメインに属している場合)。照会処理中に信用証明情報を検証するには、検索対象の文書のセキュリティー情報に両方のユーザー名をリストしておく必要があります。

### ローカル・アカウント

ローカル・アカウントの場合、ユーザー名は次のフォーマットです。

*COMPUTER NAME\USERNAME*

ログインするには、ユーザーはユーザー名のみを指定します。しかし、正しく指定された Windows ユーザー権限割り当てではフルネームが使用されません。例えば、ローカル・アカウントのユーザー名が *abcuser* である場合、アカウントのフルネームとしては *WINSERVER1\abcuser* などが考えられません。

ユーザーが検索アプリケーションを使用して、ローカル・システム上のセキュア文書の検索用プロファイルを構成するときは、Windows にログインする際に使用するユーザー名 (例えば、*abcuser*) を指定しなければなりません。

### ドメイン・アカウント

ドメイン・アカウントの場合、ユーザー名は次のフォーマットです。

*DOMAIN NAME\USERNAME*

ログインするために、ユーザーはこの情報を次のフォーマットで指定します。

*USERNAME@DOMAIN NAME*

例えば、ファイルのユーザー権限割り当てを構成して、ドメイン *WIN1\abcuser* を選択する場合、アカウントは *abcuser@win1.company.com* として表示されます。

ユーザーが検索アプリケーションを使用して、セキュア・ドメイン内の文書を検索できるようにするプロファイルを構成するときは、Windows にログインする際に使用するユーザー名 (例えば、*abcuser@win1.company.com*) を指定しなければなりません。

ローカル・コンピューター上で現行の信用証明情報の検証を施行するためには、クローラー・サーバーによって使用されるユーザー・アカウントに次の Windows ユーザー権限が必要です。(ユーザー権限を割り当てるには、Windows 管理ツールを使用します。「管理ツール」 → 「ローカル・セキュリティー・ポリシー」 → 「ローカル・ポリシー」 → 「ローカル・ユーザー権限の割り当て」。)

- クローラー・サーバーを実行するユーザー ID には、「オペレーティング・システムの一部として動作 (Act as part of the operating system)」の権限が必要です。(この権限は、WebSphere Information Integrator OmniFind Edition がインストールされるときに、クローラー・サーバー上のエンタープライズ・サーチ管理ユーザー用に構成されます。)
- ユーザーには、「ローカルでログオン」のユーザー権限が必要です。

## リモート・ドメイン・アクセス制御データでの検証

Windows オペレーティング・システムでは、\\servername で始まるディレクトリがリモート・ディレクトリと見なされます。例えば、次のとおりです。

```
\\software\utilities\IBM
```

リモート・ディレクトリにアクセスするには、ユーザーは、ユーザー名を次のフォーマットで指定します。

```
USERNAME@DOMAIN NAME
```

ユーザーが検索アプリケーションを使用して、リモート・システム上のセキュア文書を検索できるようにするプロファイルを構成するときは、リモート Windows システムにアクセスする際に使用するユーザー名 (例えば、abcuser@win1.company.com) を指定しなければなりません。

リモート・コンピューター上で現行の信用証明情報の検証を施行するためには、ユーザー・アカウントに次の Windows ユーザー権限が必要です。(ユーザー権限を割り当てるには、Windows 管理ツールを使用します。「管理ツール」→「ドメイン・セキュリティ・ポリシー」。)

- 検索対象のクローラー・サーバーおよび Windows サーバーは、同じドメインのメンバーでなければなりません。
- クローラー・サーバーを実行するユーザー ID には、「オペレーティング・システムの一部として動作 (Act as part of the operating system)」の権限が必要です。(この権限は、WebSphere Information Integrator OmniFind Edition がインストールされるときに、クローラー・サーバー上のエンタープライズ・サーチ管理ユーザー用に構成されます。)
- ユーザーには、「バッチ・ジョブとしてログオン」のユーザー権限が必要です。

### 関連概念

222 ページの『照会処理中の現行の信用証明情報の検証』

コレクションの作成時にそのコレクションのセキュリティを使用可能にした場合、特定のタイプのドメインで、ユーザーが照会をサブミットしたときにユーザーの現行の信用証明情報を検証することができます。

101 ページの『Windows ファイル・システム・クローラー』

Microsoft Windows ファイル・システムに格納された文書をエンタープライズ・サーチ・コレクションに入れる場合は、Windows ファイル・システム・クローラーを構成する必要があります。

## Lotus Domino 文書に対する文書レベルのセキュリティーの施行

クロールされる Lotus Notes サーバーが、Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用する場合、文書レベルのアクセス制御を施行するように、クローラー・サーバーを構成する必要があります。

NRPC プロトコルを使用する Lotus Notes サーバー上でクロールされた文書で文書レベル・セキュリティーを施行するには、クローラー・サーバーに Domino サーバーをインストールする必要があります。この Domino サーバーは、ご使用の Domino ドメインのメンバーでなければなりません。Domino サーバーをインストールおよび構成するには、Lotus Domino 資料の中の指示に従ってください。

セキュア・コレクションを検索するユーザーが、検索条件に合った Lotus Notes 文書の参照を許可されているかどうかを検索サーバーで検査できるように、以下のタスクも実行する必要があります。ユーザーが参照を許可されていない文書があると、検索結果がユーザーに戻される前に、その文書が検索結果から除去されます。

- 『ユーザーの信用証明情報を検証するための Lotus Domino トラステッド・サーバーの構成』。
- セキュリティーを使用するために、WebSphere Application Server でグローバル・セキュリティーを使用可能にし、検索アプリケーションを構成します。このステップにより、ユーザーが検索アプリケーションを使用する際に、信用証明情報を指定するように指示するプロンプトが出されます。その後、検索サーバーはこれらの信用証明情報を使用して、Lotus Notes 文書に対する各ユーザーのアクセス権を検査することができます。

### 関連概念

222 ページの『照会処理中の現行の信用証明情報の検証』

コレクションの作成時にそのコレクションのセキュリティーを使用可能にした場合、特定のタイプのドメインで、ユーザーが照会をサブミットしたときにユーザーの現行の信用証明情報を検証することができます。

65 ページの『Notes クローラー』

IBM Lotus Notes データベースをエンタープライズ・サーチ・コレクションに入れる場合は、Lotus Notes クローラーを構成する必要があります。

## ユーザーの信用証明情報を検証するための Lotus Domino トラステッド・サーバーの構成

Notes リモート・プロシージャー・コール (NRPC) プロトコルを使用する Lotus Notes クローラーによってクロールされたドキュメントのためにセキュリティーを施行するには、クロールされる Domino サーバーを、Lotus Domino トラステッド・サーバーとして構成する必要があります。

### 始める前に

リモート・データベースを検索する際に文書レベルのセキュリティーを施行したい場合にのみ、この手順を実行する必要があります。クローラー・サーバーのローカル・データベースを検索するには、この手順は必要ありません。

トラステッド・サーバーを構成するためには、クローラーに Domino サーバーをインストールしなければなりません。この Domino サーバーは、ご使用の Domino ドメインのメンバーでなければなりません。

## このタスクについて

Lotus Notes クローラーに対して文書レベルのセキュリティー・オプションを構成するときは、ユーザーが照会をサブミットしたときにそのユーザーの信用証明情報を検証することでアクセス制御を施行したいかどうかを指定します。このタイプのセキュリティーを施行したい場合は、クロールされる Domino サーバーを Lotus Domino トラステッド・サーバーにしなければなりません。

ユーザーの現行の信用証明情報を検証する必要があるドメインをユーザーが検索するとき、トラステッド・サーバーが Domino サーバー ID を使用してコンテキストを現行のユーザー ID に切り替えることができます。Domino データベースは、現行のユーザーがそのデータベースをオープンしたかのようにオープンされ、そのユーザーに関するすべてのデータベース・アクセス制御リスト情報が施行されます。

この方法でコンテキストを切り替える機能は、一般的には、ローカル Domino サーバーの data ディレクトリーに保管されているデータベースでのみ使用可能です。Lotus Domino バージョン 6.5.1 から、この機能はトラステッド・サーバーによって提供されています。トラステッド・サーバーを構成するには、機密の操作 (例えば、データベースがリモート・コンピューターからアクセスされたときに別のユーザーとして動作するなど) を行う際にどの Domino サーバーを信用すべきかを、Domino 管理者が指定します。

## 手順

トラステッド・サーバーを構成するには、Lotus Notes クローラーによってクロールされるすべての Domino サーバー上で次のステップを実行します。

1. Domino サーバーで、Domino ドメイン管理者 ID ファイルを使用して、Lotus Domino Administrator クライアントを開きます。
2. 「ファイル」 → 「サーバーのオープン」を選択します。
3. トラステッド・サーバー機能を使用可能にしたい Domino サーバーの名前を入力します。
4. 「構成」タブを選択します。
5. 「サーバー」オブジェクトを展開し、「現行サーバー」文書を選択して、「サーバーの編集」をクリックします。
6. 「セキュリティー」タブを選択し、文書の最後までスクロールし、「トラステッド・サーバー」項目を見つけて、下矢印ボタンをクリックします。
7. 次のオプションのいずれかを指定します。

### LocalDomainServers

Domino ドメイン内のすべてのサーバーをトラステッド・サーバーと見なす場合は、このオプションを選択します。

### server\_name

トラステッド・サーバーとしてクロールおよび検索可能な Domino サーバーの名前を指定します。

クロールしたい Domino サーバーが別の Domino ドメインにある場合、サーバー名を指定するか、あるいは **OtherDomainServers** グループを選択する必要があります。WebSphere II OmniFind Edition Domino サーバー ID ファイルを他の Domino ドメインとクロス認証するためには、さ



らに Domino 用の手順を実行する必要があります。これらの手順について詳しくは、Domino サーバーの資料を参照してください。

8. 「保管してクローズ」をクリックして、変更内容を保管します。
9. トラストッド・サーバーとして作動するようにしたりリモート Domino サーバーを停止および再始動します。

#### 関連概念

222 ページの『照会処理中の現行の信用証明情報の検証』  
コレクションの作成時にそのコレクションのセキュリティーを使用可能にした場合、特定のタイプのドメインで、ユーザーが照会をサブミットしたときにユーザーの現行の信用証明情報を検証することができます。

65 ページの『Notes クローラー』

IBM Lotus Notes データベースをエンタープライズ・サーチ・コレクションに入れる場合は、Lotus Notes クローラーを構成する必要があります。

---

## 文書レベル・セキュリティーを使用不可にする

アクセス制御が索引内の文書と関連付けられているかどうかに関係なく、ユーザーがコレクションを検索できるようにすることができます。現行の信用証明情報の検証をサポートするクローラーでは、照会処理中に現行のアクセス制御を検証することなく、ユーザーにコレクションを検索させることができます。

### 始める前に

文書レベルのセキュリティー・オプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 制約事項

文書レベルのセキュリティー・オプションを指定できるのは、コレクションが作成されたときにセキュリティーがコレクションに対して使用可能にされている場合だけです。

### このタスクについて

クローラーを構成する際に、クローラー対象の文書にセキュリティー・トークンを関連付けることができます。検索アプリケーションは、索引に格納されているこれらのトークンを使用して、ユーザーがコレクションを検索する際にアクセス制御を施行できます。クローラーの中には、ユーザーが照会をサブミットしたときに、ネイティブ・リポジトリ内の、文書に関連付けられた現行のアクセス制御を検証するように指定できるものもあります。

これらのセキュリティー制限を除去するために、照会を使って渡されるセキュリティー・トークンを検索サーバーが無視するように指定できます。信用証明情報を現行のアクセス制御と比較することなく、ユーザーが文書を照会できるようにすることもできます。



新規コレクションをテストする場合や、検索アプリケーションの問題をトラブルシューティングする必要がある場合に、文書レベルのセキュリティーを一時的に使用不可にしたいことがあります。

## 手順

文書レベルのアクセス制御を使用不可にするには、次のようにします。

1. コレクションを編集します。「一般」ページを選択して、「**文書レベルのセキュリティーの構成**」をクリックします。
2. クローラーが文書と関連付けたセキュリティー・トークンを、ユーザーがコレクションを照会するときを使用したくない場合、「文書レベルのセキュリティー」ページで、「**索引内の文書レベルのアクセス制御を無視する**」チェック・ボックスを選択します。

クローラーは文書にセキュリティー・トークンを追加し続けますが、検索サーバーはトークンを無視し、ユーザーが以前に保護された文書を検索できるようにします。

3. ユーザーが照会をサブミットしたときに、ネイティブ・リポジトリ内の、文書に関連付けられた現行のアクセス制御を検証したくない場合、「**照会処理中に現行の信用証明情報の検証を行わない (Do not validate current credentials during query processing)**」チェック・ボックスを選択します。このチェック・ボックスは、この機能をサポートするクローラーによってクローラされた文書に対してのみ選択可能です。

このチェック・ボックスを選択しても、その他の文書レベルのセキュリティー・オプションは有効なままです。例えば、クローラーの構成時にアクセス制御を索引内に保管するオプションを指定した場合、「**索引内の文書レベルのアクセス制御を無視する**」チェック・ボックスも同様に選択しないと、それらのセキュリティー制御が引き続き適用されます。

4. 「**OK**」をクリックします。



---

## エンタープライズ・サーチの WebSphere Portal との統合

WebSphere Portal および WebSphere Portal 検索センターにエンタープライズ・サーチ・ポートレットをデプロイすることにより、IBM WebSphere Portal の検索機能を拡張することができます。

### 統合の設定ポイント

エンタープライズ・サーチ・ポートレットを WebSphere Portal に統合するには、次のようにいくつかの方法があります。

#### WebSphere Portal

WebSphere Portal には、ユーザーがアプリケーション、コンテンツ、プロセス、および人物と相互通信を行う単一のアクセス・ポイントがあります。この WebSphere Portal フレームワークにより、ポートレットと呼ばれる新規のアプリケーションを、ポータル内の他のアプリケーションに影響を与えることなく統合し、デプロイすることができます。

エンタープライズ・サーチ・ポートレットを WebSphere Portal にデプロイすれば、WebSphere Portal インターフェースからエンタープライズ・サーチ・コレクションを検索することができます。WebSphere Portal の構成設定により、エンタープライズ・サーチ・ポートレットのルック・アンド・フィールは WebSphere Portal 環境の他のポートレットと同じ状態に保たれます。

#### ポータル検索エンジン

WebSphere ポータル検索エンジンは、Web サイトをクロールします。管理ポートレットを使用すれば、管理者は索引付けされたコレクションを作成することができ、さらに検索ポートレットにより、作成されたコレクションをユーザーが検索できるようになります。

WebSphere Portal バージョン 5.0.2 またはそれ以降のバージョンを使用している場合、ポータル検索エンジン・コレクション用の構成情報および分類をエンタープライズ・サーチにマイグレーションすることができます。

WebSphere Portal バージョン 5.1 またはそれ以降のバージョンを使用している場合、ポータル検索エンジン文書検索ポートレットを使用して、エンタープライズ・サーチ・コレクションを検索することができます。構成プロパティを利用すれば、管理者が必要に応じて 2 つの検索機能を簡単に切り替えられるようにすることができます。

#### WebSphere Portal 検索センター

WebSphere Portal 検索センターは、WebSphere Portal で検索が可能になっているすべてのソースを検索するときを中心となる起点です。この検索センターおよびユニバーサル検索ポートレットを使用すれば、WebSphere Portal コンテンツ、および管理者が検索センターに登録したその他すべてのコレクションを検索することができます。

検索センターにはページ式のインターフェースがあります。利用できるすべてのコレクションの検索は、1 つの共通ページから行うこともできますし、

個別のコレクションを検索するページを選択して行うこともできます。例えば、ポータル検索エンジン索引のページ、 Portal Document Management ライブラリーのページなどがあります。

エンタープライズ・サーチ・コレクションを検索センターから検索できるようにするため、WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) では、アダプターおよび登録ポートレットが提供されています。アダプターによってエンタープライズ・サーチ用のタブが検索センター・インターフェースに追加され、登録ポートレットによってエンタープライズ・サーチ・ポートレットが検索センターに登録されます。

アダプターおよび登録ポートレットをインストールしないと、検索センターを使用してエンタープライズ・サーチ・コレクションを検索できるようにはなりません。

### WebSphere Portal クローラー

WebSphere II OmniFind Edition のインストール時に、エンタープライズ・アプリケーション ESPACServer.ear が検索サーバーにインストールされます。このエンタープライズ・アプリケーションを WebSphere Portal にデプロイした後、エンタープライズ・サーチ管理コンソールを使用して、WebSphere Portal クローラーを構成し、WebSphere Portal サイトをエンタープライズ・サーチ・コレクションに追加することができます。

### 統合による利点

エンタープライズ・サーチによって、より広いデータ・ソース・タイプの検索がサポートされ、WebSphere Portal 検索環境が拡張されます。エンタープライズ・サーチ・ポートレットを使用することにより、Web サイトに加え、WebSphere II OmniFind Edition でサポートされる他のすべてのデータ・タイプも検索することができます。

また、エンタープライズ・サーチには、スケーラビリティの面でも利点があります。ポータル検索エンジンは、シングル・サーバーで十分に検索負荷をサポートできる中小規模の企業に便利です。これに対し、エンタープライズ・サーチでは、大規模な企業レベル能力をサポートするため、4 つのサーバー (データのクロールに 1 基、データの構文解析と索引作成に 1 基、および検索処理サポートに 2 基) に負荷を分散します。

### ポートレットのデプロイメント概要

エンタープライズ・サーチに使用できるポートレットは、使用する WebSphere Portal のバージョンによって異なります。

#### WebSphere Portal バージョン 5.0.2

エンタープライズ・サーチ・ポートレットをデプロイし、これをエンタープライズ・サーチ・コレクションの検索に使用することができます。このポートレットは、ポータル検索エンジン・ポートレットと共存できます。

#### WebSphere Portal バージョン 5.1 以降

- エンタープライズ・サーチ・ポートレットをデプロイし、これをエンタープライズ・サーチ・コレクションの検索に使用することができます。このポートレットは、ポータル検索エンジン・ポートレットと共存できます。
- エンタープライズ・サーチ・ポートレットをデプロイ後、文書検索ポートレットをポータル検索エンジンに構成すれば、エンタープライズ・サーチ・コレクションを検索できるようになります。
- エンタープライズ・サーチ・ポートレットを WebSphere Portal 検索センターに登録すると、検索センターを使用してエンタープライズ・サーチ・コレクションを検索できるようになります。この場合、エンタープライズ・サーチ・コレクションだけを検索するページを選択することもできますし、エンタープライズ・サーチ・コレクションおよび検索センターで利用できる他のすべてのコレクションを検索する照会を入力することもできます。

#### 関連概念

243 ページの『WebSphere Portal からエンタープライズ・サーチへのマイグレーション』

エンタープライズ・サーチでは、IBM WebSphere Portal からエンタープライズ・サーチに分類構造およびコレクションをマイグレーションするときを使用するマイグレーション・ウィザードが用意されています。

#### 関連タスク

99 ページの『WebSphere Portal クローラー用エンタープライズ・アプリケーションのデプロイ』

WebSphere Portal クローラーを作成する前に、エンタープライズ・アプリケーション ESPACServer.ear を WebSphere Portal にデプロイする必要があります。

---

## ポータル検索エンジンでの文書レベル・セキュリティ

IBM WebSphere ポータル検索エンジンを使用して、ユーザーがエンタープライズ・サーチ・コレクションを検索するときに文書レベルのセキュリティを施行することができます。

エンタープライズ・サーチ・クローラーが、クロール対象の文書にグループ ID セキュリティー・トークンを関連付けている場合で、ポータル検索エンジンの文書検索ポートレットを、エンタープライズ・サーチ・コレクションを検索するように構成する場合、ポータル検索エンジンは、ログインしたユーザーのグループ ID を取り出し、そのグループ ID のセキュリティ・トークンを照会と共にエンタープライズ・サーチに渡すことができます。セキュリティ・トークンにより、ユーザーに表示を許可された文書だけが検索結果に戻されるようになります。

クローラーが、ユーザー ID またはユーザー役割など、別のタイプのセキュリティ・トークンを文書に関連付けており、かつ、エンタープライズ・サーチ・コレクションを検索するときに文書レベルのセキュリティを施行したい場合は、カスタムの検索ポートレットを作成する必要があります。ポータル検索エンジンは、グループ ID のセキュリティ・トークンしか取り出しません。

#### 関連概念

189 ページの『エンタープライズ・サーチの検索アプリケーション』

検索アプリケーションは、エンタープライズ・サーチ・システム内のコレクショ

ンおよび外部ソースの検索を可能にします。検索アプリケーションはいくつでも作成できます。また、1つの検索アプリケーションで、任意の数のコレクションおよび外部ソースを検索できます。

221 ページの『文書レベル・セキュリティー』

コレクションが作成されたときに、そのコレクションに対するセキュリティーを使用可能にすると、文書レベルのセキュリティー制御を構成することができます。文書レベルのセキュリティーによって、コレクションを検索するユーザーは、見ることが許される文書にしかアクセスできないようになります。

---

## 検索ポータルレットのデプロイ

検索ポータルレット ESSearchPortlet.war によって、エンタープライズ・サーチ・コレクションは WebSphere ポータル検索エンジンで検索できるようになります。

### 始める前に

WebSphere Portal にエンタープライズ・サーチ用の検索ポータルレットをデプロイしてからでないと、ポータルレットを使用するようポータル検索エンジンを構成することはできません。

### このタスクについて

検索ポータルレットをデプロイした後も、引き続きポータル検索エンジンを使用して、WebSphere Portal 内の索引付けされたデータを検索できます。ユーザーがエンタープライズ・サーチ・コレクションを検索できるようにするには、WebSphere Portal 管理者がポータル検索エンジン構成のプロパティを編集します。これら 2つの検索機能を切り替え、ニーズに合った最適な検索ソリューションを選択して使用できます。

ESSearchPortlet.war ファイルは、WebSphere II OmniFind Edition がインストールされるときに、検索サーバー上の ES\_INSTALL\_ROOT/bin ディレクトリーにインストールされます。デフォルトのインストール・パスは、次のとおりです。

- UNIX システム: /opt/IBM/es/bin/ESSearchPortlet.war
- Windows システム: C:\Program Files\IBM\es\bin\ESSearchPortlet.war

### 手順

検索ポータルレットを WebSphere Portal にデプロイするには、次のようにします。

1. ポータルレットをインストールします。
  - a. WebSphere Portal サーバー管理者 ID を使用して WebSphere Portal サーバーにログインします。
  - b. ツールバーから「管理」を選択します。
  - c. 左側のナビゲーション領域で「ポータルレット」を選択し、「ポータルレット」メニューから「インストール」を選択します。
  - d. 「ブラウズ」をクリックし、ご使用のシステムから ESSearchPortlet.war ファイルを選択し、「次へ」をクリックします。



- e. 次のページで、「インストール」をクリックします。ポートレットがインストールされると、「APIN0005I: ポートレットは正常にインストールされました」というメッセージが表示されます。
2. ポートレットのパラメーターを変更します。
    - a. 左側のナビゲーション領域で「**ポートレットの管理**」を選択し、「**IBM Enterprise Search**」を選択します。
    - b. 選択ボックスの右に新しいアイコンが表示されたら、「**パラメーターの変更**」を選択します。
    - c. ポートレット・パラメーターのリストで、以下のパラメーターを変更します。

**port** この値を、WebSphere II OmniFind Edition 検索サーバー用の Web サーバーのポート番号に設定します。デフォルト値は 80 です。

**applicationName** この値を、エンタープライズ・サーチ用の有効な検索アプリケーションの名前に設定します (エンタープライズ・サーチ管理コンソールの「検索アプリケーション」ページに、選択可能な検索アプリケーションの名前がリストされます)。デフォルト値は、検索サーバー上に常に構成されている Default です。

**hostname** この値を、WebSphere II OmniFind Edition 検索サーバーの完全修飾ホスト名に設定します。
    - d. 「**保管**」をクリックして変更内容を保管し、次に「**キャンセル**」をクリックして「パラメーターの変更」ページを終了します。
  3. ポートレットのポータル・ページを作成します。
    - a. 左側のナビゲーション領域で「**ポータル・ユーザー・インターフェース**」を選択し、「**ページの管理**」を選択します。
    - b. 「**マイ・ポータル**」リンクを選択します。
    - c. 「**新規ページ**」をクリックし、ページのタイトルを指定し、「**OK**」をクリックします。ページが作成されると、「APPR0010I: ページ・タイトル は正常に作成されました」というメッセージが表示されます。
    - d. 「**OK**」をクリックします。作成したページが、使用可能なページのリストに表示されます。
  4. ページに検索ポートレットを追加します。
    - a. 作成したページに対して「**編集**」アイコン (鉛筆) を選択します。
    - b. 「**ポートレットの追加**」をクリックします。
    - c. 「**検索対象**」テキスト・ボックスに、Enterprise を入力し、「**検索**」をクリックします。
    - d. 「**IBM Enterprise Search**」の横のチェック・ボックスを選択し、「**OK**」をクリックします。
    - e. 「**完了**」をクリックします。このステップは、検索ポートレットのインストールと WebSphere ポータル・ページへのポートレットの追加が正常に実行されたことを示します。
  5. ポートレットにアクセスします。

- a. ページ上部のナビゲーション・バーで「マイ・ポータル」を選択します。
- b. システムにポートレットを追加したときに指定したページ・タイトルに一致するページを選択します。

このポートレット・ページに初めてアクセスする場合、このポートレット用の Java Server Page (JSP ファイル) をシステムはコンパイルする必要があるため、ページの表示には少し時間がかかります。

---

## WebSphere Portal 検索ブラウザ・ポートレットのエンタープライズ・サーチのための構成

WebSphere Portal には、ネイティブな WebSphere Portal コレクションの検索とブラウズに使用できるポートレットがあります。このポートレットを、エンタープライズ・サーチ・コレクションを検索するように構成することができます。

### 手順

エンタープライズ・サーチ・コレクションを検索するように検索ブラウザ・ポートレットを構成するには、次のようにします。

1. WebSphere\_Portal サーバー・インスタンスを停止します。
2. 以下のファイルを WebSphere II OmniFind Edition ES\_INSTALL\_ROOT/lib ディレクトリーから、WebSphere Portal サーバー上の /WebSphere/PortalServer/shared/app ディレクトリーにコピーします。  
esapi.jar  
siapi.jar
3. WebSphere\_Portal サーバー・インスタンスを開始し、WebSphere Portal 管理者 ID を使用して WebSphere Portal サーバーにログインします。
4. ポートレットをコピーします。
  - a. 右上隅の「管理」をクリックし、「ポートレットの管理」オブジェクトを展開して「ポートレット」をクリックします。
  - b. search という語を検索します。
  - c. 検索ブラウザ・ポートレットに対して「ポートレットのコピー」をクリックし、このポートレットの固有の名前を入力し、「OK」をクリックします。
5. ポートレットを構成します。
  - a. 指定したポートレット名の横の「ポートレットの構成」アイコンをクリックします。
  - b. 右矢印をクリックして、パラメーターの 2 ページ目に進みます。
  - c. 「新規パラメーター」に ApplicationInfoId を入力し、有効な WebSphere II OmniFind Edition 検索アプリケーションの名前を「新しい値」フィールドに入力し、この新規パラメーターを追加するため「追加」をクリックします。

エンタープライズ・サーチ管理コンソールの「検索アプリケーション」ページに、選択可能な検索アプリケーションの名前がリストされます。デフォルト値は、検索サーバー上に常に構成されている Default です。

- d. 以下のパラメーターを削除します。

```
ApplicationInfoId_EXAMPLE
QueryFactoryImp
BrowseFactoryImp
SearchFactoryImp
IndexName
EJB_Example
IIOP_URL_Example
SOAP_URL_Example
EJB
SOAP_URL
IIOP_URL
```

- e. 以下のパラメーターを追加し、下記の値を指定します。各パラメーターの名前と値を追加するため、「追加」をクリックします。パラメーター名は大/小文字を区別します。

```
IMPLEMENT = ES
ApplicationInfoId = 検索アプリケーション名 (例: Default)
SearchFactoryImp = com.ibm.es.api.search.RemoteSearchFactory
BrowseFactoryImp = com.ibm.es.api.browse.RemoteBrowseFactory
hostname = 検索サーバーのホスト名 (例: omnifind.ibm.com)
port = 検索サーバーのポート番号 (例: 80)
```

6. 検索ブラウザ・ポートレットの新規コピーをポータル・ページに割り当てます。詳しくは、WebSphere Portal 管理の資料を参照してください。
7. ポートレットにアクセスします。
  - a. ポータル・ページ上部のナビゲーション・バーで「マイ・ポータル」を選択します。
  - b. ステップ 4 (238 ページ) でポートレットをコピーしたときに指定した名前に一致するページを選択します。

---

## 検索センターを使用するためのエンタープライズ・サーチ・アダプターのインストール

エンタープライズ・サーチ・コレクションの検索に WebSphere Portal 検索センターを使用するには、エンタープライズ・サーチ用のアダプターをインストールする必要があります。

### このタスクについて

アダプター ESSearchAdapter.ear は、エンタープライズ・サーチのページを検索センター・インターフェースに追加します。このアプリケーションと登録ポートレットをインストールすると、それ以降は、エンタープライズ・サーチ・コレクションを検索センターに追加することができ、それらのコレクションをユニバーサル検索ポートレットで検索できます。

ESSearchAdapter.ear ファイルは、WebSphere II OmniFind Edition がインストールされるときに、検索サーバー上の ES\_INSTALL\_ROOT/bin ディレクトリーにインストールされます。デフォルトのインストール・パスは、次のとおりです。

- UNIX システム: /opt/IBM/es/bin/ESSearchAdapter.ear
- Windows システム: C:\Program Files\IBM\es\bin\ESSearchAdapter.ear

### 手順

エンタープライズ・サーチ用のアダプターを検索センターにインストールするには、次のようにします。

1. WebSphere\_Portal サーバー・インスタンスを停止します。
2. WebSphere Application Server server1 サーバー・インスタンスがまだ開始されていない場合は、開始します。
3. 以下のファイルを WebSphere II OmniFind Edition ES\_INSTALL\_ROOT/lib ディレクトリーから、WebSphere Portal サーバー上の /WebSphere/PortalServer/shared/app ディレクトリーにコピーします。

```
esapi.jar  
siapi.jar
```

siapi.jar ファイルを上書きするかどうかを確認するプロンプトが出されたら、「はい」を指定します。

4. WebSphere Portal サーバー上で、WebSphere Application Server 管理コンソールを開始します。ログインを指示するプロンプトが出されたら、ログインします。

管理コンソールは、以下の方法でオープンできます。

- Windows の「スタート」メニューを使用して、該当するプログラムを選択します。
  - WebSphere Application Server バージョン 5 の場合、Web ブラウザーをオープンし、`http://hostname:port/admin` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9090 です。
  - WebSphere Application Server バージョン 6 の場合、Web ブラウザーをオープンし、`http://hostname:port/ibm/console` にアクセスします。ここで、*hostname* は検索サーバーのホスト名であり、*port* は WebSphere Application Server 管理コンソールのポート番号です。通常、管理コンソールのポートは 9060 です。
5. 「アプリケーション」をクリックして、「新規アプリケーションのインストール」をクリックします。
  6. 「ブラウズ」をクリックし、システムから ESSearchAdapter.ear を選択し、「次へ」を 2 回クリックします。ポリシー・ファイルに関する警告が出された場合は、「継続」をクリックします。
  7. 「アプリケーション・サーバーへのモジュールのマップ」ページが表示されるまで「次へ」をクリックし、以下の選択を行ない、「適用」をクリックします。
    - 「クラスターおよびサーバー」フィールドで、**WebSphere:cell=cell\_name、node=node\_name、server=server1** を選択します。
    - **ESSearchAdapterEJB** モジュールの横のチェック・ボックスを選択します。
  8. 「サマリー」ページが表示されるまで「次へ」をクリックし、「完了」をクリックします。
  9. 「マスター構成に保管」リンクをクリックして、「保管」ボタンをクリックし、WebSphere Application Server 構成に変更を保管します。
  10. WebSphere Portal サーバーを再始動します。

## 検索センターを使用するためのエンタープライズ・サーチ登録ポートレットのインストール

WebSphere Portal 検索センターを使用してエンタープライズ・サーチ・コレクションを検索するには、エンタープライズ・サーチ用の登録ポートレットをインストールする必要があります。

### このタスクについて

登録ポートレット `ESSearchAdapterPortlet.war` は、エンタープライズ・サーチ・ポートレットを検索センターに登録します。このポートレット (およびエンタープライズ・サーチ用のアダプター・アプリケーション) をインストールすると、それ以降は、エンタープライズ・サーチ・コレクションを検索センターに追加することができ、それらのコレクションをユニバーサル・サーチ・ポートレットで検索できます。

`ESSearchAdapterPortlet.war` ファイルは、WebSphere II OmniFind Edition がインストールされるときに、検索サーバー上の `ES_INSTALL_ROOT/bin` ディレクトリーにインストールされます。デフォルトのインストール・パスは、次のとおりです。

- UNIX システム: `/opt/IBM/es/bin/ESSearchAdapterPortlet.war`
- Windows システム: `C:\Program Files\IBM\es\bin\ESSearchAdapterPortlet.war`

### 手順

エンタープライズ・サーチ用の登録ポートレットをインストールするには、次のようにします。

1. WebSphere Portal 管理者 ID を使用して WebSphere Portal サーバーにログインします。
2. ツールバーから「管理」を選択します。
3. 左側のナビゲーション領域で「ポートレット」を選択し、「ポートレット」メニューから「インストール」を選択します。
4. 「ブラウズ」をクリックし、ご使用のシステムから `ESSearchAdapterPortlet.war` ファイルを選択し、「次へ」をクリックします。
5. 次のページで、「インストール」をクリックします。ポートレットがインストールされると、「APIN00051: ポートレットは正常にインストールされました」というメッセージが表示されます。
6. このポートレット (IBM WebSphere II OmniFind Edition registration portlet for enterprise search という名前が付けられています) をポータル・ページに割り当てます。詳しくは、WebSphere Portal 管理の資料を参照してください。





---

## WebSphere Portal からエンタープライズ・サーチへのマイグレーション

エンタープライズ・サーチでは、IBM WebSphere Portal からエンタープライズ・サーチに分類構造およびコレクションをマイグレーションするときに使用するマイグレーション・ウィザードが用意されています。

分類構造およびコレクションをマイグレーションする場合、エンタープライズ・サーチ索引サーバーでマイグレーション・ウィザードを実行します。分類構造をマイグレーションすると、その分類をエンタープライズ・サーチ・コレクションで使用することができます。また、エンタープライズ・サーチを使用して、WebSphere Portal からマイグレーションしたコレクションを管理、検索することもできます。

エンタープライズ・サーチでは、分類構造はカテゴリ・ツリーと言います。ルール・ベースの分類構造をマイグレーションした場合、それをカテゴリ・ツリーとしてエンタープライズ・サーチ管理コンソールで管理することができます。モデル・ベースの分類構造をエンタープライズ・サーチで使用するには、索引サーバーに WebSphere Portal がインストールされていなければなりません。

分類構造およびコレクションをマイグレーションしたい場合は、必ずモデル・ベースの分類構造ファイルを先にマイグレーションしてからコレクションをマイグレーションしてください。このようにしないと、モデル・ベースのカテゴリ化が WebSphere Portal からマイグレーションするコレクションについて適切に機能しなくなります。

### 関連概念

233 ページの『エンタープライズ・サーチの WebSphere Portal との統合』  
WebSphere Portal および WebSphere Portal 検索センターにエンタープライズ・サーチ・ポートレットをデプロイすることにより、IBM WebSphere Portal の検索機能を拡張することができます。

### 関連タスク

122 ページの『カテゴリの構成』  
1 つのコレクションのカテゴリはいくつも作成することができ、各カテゴリにはいくつもの規則を含めることができます。規則によって、カテゴリと自動的に関連付けられる 文書が決まります。

---

## WebSphere Portal からのモデル・ベース分類構造のマイグレーション

WebSphere Portal 分類管理ポートレットを使用すれば、エンタープライズ・サーチ・コレクションで使用するモデル・ベースの分類構造を選択することができます。すでにエンタープライズ・サーチにマイグレーション済みのコレクションは、新たな分類マイグレーションを選択してもその影響を受けることはありません。

### 始める前に

初めてマイグレーション・ウィザードを実行する前には、エンタープライズ・サーチ・システムを停止します。これで、構成ファイルへの変更が可能になります。最初にマイグレーション・ウィザードを実行するとき、WebSphere Application Server および WebSphere Portal がインストールされているパスを指定します。2 回目からは、マイグレーション・ウィザードを実行するときは、エンタープライズ・サーチ・システムをアクティブにすることができます。

### このタスクについて

モデル・ベースの分類構造をマイグレーションする場合、WebSphere Portal の分類構造を選択し、エクスポートする必要があります。それから、エンタープライズ・サーチのマイグレーション・ウィザードを使ってその分類構造をエンタープライズ・サーチにマイグレーションします。

### 手順

モデル・ベースの分類構造を WebSphere Portal からエンタープライズ・サーチにマイグレーションするには、次のようにします。

1. 現行のモデル・ベースの分類構造を WebSphere Portal 分類管理ポートレットからエクスポートします。分類構造は、次の XML ファイルで構成されています。

```
synonyms.xml  
titles.xml  
treenodes.xml
```

2. これらのファイルをエンタープライズ・サーチ索引サーバーにコピーします。
3. エンタープライズ・サーチ索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
4. エンタープライズ・サーチのインストール・ディレクトリーに移動します。

```
UNIX: cd $ES_INSTALL_ROOT/bin  
Windows: cd %ES_INSTALL_ROOT%\bin
```

5. マイグレーション・ウィザードを初めて開始する場合は、次のコマンドを入力して、エンタープライズ・サーチ・システムを停止します。

```
esadmin stop
```

6. 次のコマンドを入力してマイグレーション・ウィザードを開始し、「次へ」をクリックします。

```
UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat
```

初めてマイグレーション・ウィザードを実行するのではなく、エンタープライズ・サーチ・システムがアクティブになっている場合、マイグレーション・ウィザードの実行中にエンタープライズ・サーチ・コレクションを作成しないでください。

7. 「WebSphere Portal からのモデル・ベース分類構造ファイルのインポート」を選択し、「次へ」をクリックします。

8. マイグレーション・ウィザードを初めて使用する場合は、WebSphere Application Server および WebSphere Portal がインストールされているパスをそれぞれ指定します。「次へ」をクリックします。
9. モデル・ベースの分類構造ファイルのあるディレクトリーを参照して探し、モデル・ベースの分類構造を使用するためにマイグレーションする必要がある XML ファイル (synonyms.xml、 titles.xml、 および treenodes.xml) を選択して、「次へ」をクリックします。
10. これがマイグレーション・ウィザードを実行する第 1 回目であった場合は、マイグレーションの終了後、次のコマンドを入力して、エンタープライズ・サーチ・システムを再始動してください。

```
esadmin start
```

エラーが発生した場合、マイグレーション・ウィザードがインストールされているディレクトリーにある MigrationWizard.log ファイルを参照してください。

#### 関連概念

120 ページの『モデル・ベースのカテゴリー』

IBM WebSphere Portal システムで、モデル・ベースのカテゴリーを使用している場合、エンタープライズ・サーチ・コレクションでもそれらのカテゴリーを継続して使用することができます。

#### 関連タスク

122 ページの『カテゴリーの構成』

1 つのコレクションのカテゴリーはいくつも作成することができ、各カテゴリーにはいくつもの規則を含めることができます。規則によって、カテゴリーと自動的に関連付けられる 文書が決まります。

---

## WebSphere Portal からのコレクションのマイグレーション

WebSphere Portal からエンタープライズ・サーチにコレクションをマイグレーションするには、WebSphere Portal にコレクションを準備し、マイグレーション・ウィザードを使用して準備したコレクションをマイグレーションします。

### 始める前に

モデル・ベースの分類構造およびコレクションをマイグレーションする場合、この手順を使用して、コレクションをマイグレーションします。このようにしないと、モデル・ベースのカテゴリー化が WebSphere Portal からマイグレーションするコレクションについて適切に機能しなくなります。

### 手順

コレクションを WebSphere Portal からエンタープライズ・サーチにマイグレーションするには、次のようにします。

1. WebSphere ポータル検索エンジンで、マイグレーションするコレクションのすべてのクローラー・プロセスを停止し、保留中のすべての文書を承認または拒否します。(エンタープライズ・サーチでは、保留中の文書という概念はサポートしません。)

2. マイグレーションする各コレクションについて、ポータル検索エンジン・ポートレットを使用してその設定を XML ファイルにエクスポートします。
3. エンタープライズ・サーチ索引サーバーが別のサーバーにインストールされている場合は、エクスポートした XML ファイルを索引サーバーにコピーします。
4. エンタープライズ・サーチ索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
5. エンタープライズ・サーチのインストール・ディレクトリーに移動します。

UNIX: cd \$ES\_INSTALL\_ROOT/bin

Windows: cd %ES\_INSTALL\_ROOT%\bin

6. 次のコマンドを入力してマイグレーション・ウィザードを開始し、「次へ」をクリックします。

UNIX: ./eswpsmigrate.sh

Windows: eswpsmigrate.bat

7. 「**WebSphere Portal のポータル検索エンジンから検索設定をマイグレーションする**」を選択し、「次へ」をクリックします。
8. エクスポートされたポータル検索エンジン構成ファイルがあるディレクトリーを参照し、マイグレーションしたいファイルを選択して「次へ」をクリックします。選択された構成ファイルが分析、検証されます。
9. コレクションごとに次の情報を入力し、「次へ」をクリックすると、コレクションのエンタープライズ・サーチへのマイグレーションが開始されます。
  - エンタープライズ・サーチで使用するコレクションの名前。
  - コレクションで文書の重要度を判断するための基準。この静的ランキング要素は、なし、文書日付基準、他の Web 文書からの Web 文書へのリンク基準のいずれかにすることができます。
  - このコレクションに使用するカテゴリーのタイプ。カテゴリーなし、ルール・ベースのカテゴリー、またはモデル・ベースのカテゴリーのいずれかを選択できます。

ルール・ベースのカテゴリーを選択した場合、WebSphere Portal コレクションの分類構造と規則がエンタープライズ・サーチにマイグレーションされます。

モデル・ベースのカテゴリーを選択できるのは、あらかじめモデル・ベースの分類構造を WebSphere Portal からエンタープライズ・サーチにマイグレーションしてある場合に限られます。

マイグレーション時にエラーが発生した場合、マイグレーション・ウィザードがインストールされているディレクトリーにある MigrationWizard.log ファイルを参照してください。

マイグレーションされたコレクションに追加設定を構成するには、エンタープライズ・サーチ管理コンソールを使用することができます。

**要件:** マイグレーションしたコレクションの Web クローラー・プロパティを構成するときに、クローラーに関するコメントを受信する E メール・アドレスとユーザー・エージェント名を指定する必要があります (Web クローラー・プロパティの構成中に不明な点があったら、「ヘルプ」をクリックしてください)。

10. エンタープライズ・サーチ管理コンソールから、マイグレーション済みコレクションに対してクローリング、構文解析、および索引作成プロセスを開始します。
11. マイグレーション済みコレクションがエンタープライズ・サーチで検索できるとわかったら、ポータル検索エンジンにある元のコレクションを削除します。
12. オプション: WebSphere Portal 管理者が、WebSphere Portal のポータルからマイグレーションしたコレクションをユーザーが検索できるようにしたい場合は、以下のステップを実行します。
  - a. エンタープライズ・サーチ・ポートレットを WebSphere Portal システムにデプロイします。

WebSphere Portal サーバー・クラスターでは、この作業は WebSphere Application Server Deployment Manager がインストールされているサーバーに対して行う必要があります。WebSphere Portal サーバー・クラスターの他のサーバーには、この Deployment Manager によってエンタープライズ・サーチ・ポートレットが配布されます。

- b. エンタープライズ・サーチ・ポートレットを該当するポータル・ページに追加します。

WebSphere Portal では、検索ポートレットのアクセス制御が、特定ページおよびポートレットに対するアクセス可能度によってモデル化されています。コレクション設定はマイグレーションされますが、ポートレットは WebSphere Portal サーバー管理者によって手動で配置する必要があります。

#### 関連概念

29 ページの『エンタープライズ・サーチ・コレクション』

エンタープライズ・サーチ・コレクションには、ユーザーが 1 回の照会で検索できる全ソースが格納されています。フェデレーションによって、ユーザーは複数のコレクションを 1 回の照会で検索することができます。

117 ページの『ルール・ベースのカテゴリー』

エンタープライズ・サーチ・コレクションで文書をカテゴリーに関連付ける場合、規則を構成してどの文書を関連付けるか制御することができます。

120 ページの『カテゴリー・ツリー』

カテゴリー・ツリーを使用すると、コレクション内のルール・ベースのすべてのカテゴリーを表示できます。カテゴリー・ツリーを使って、カテゴリーの作成、カテゴリーの削除、およびカテゴリーに文書を関連付けるときの規則を編集することもできます。

178 ページの『静的ランキング』

文書のタイプによっては、検索結果における文書の重要度を高くする静的ランキング要素を関連付けることができます。

---

## マイグレーションされたコレクションの設定

IBM WebSphere Portal からコレクションをマイグレーションすると、マイグレーション・ウィザードにより、コレクションおよびクローラーのデフォルト設定が作成されます。

ポータル検索エンジン・コレクションとエンタープライズ・サーチ・コレクションに同じ設定がある場合、ウィザードでは、コレクションのエンタープライズ・サーチへのマイグレーション時にポータル検索エンジンの設定を使用します。エンタープライズ・サーチだけにある設定については、コレクションのマイグレーション時に指定した設定、または、エンタープライズ・サーチのコレクションに対するデフォルト設定がウィザードによって使用されます。

### ポータル検索エンジンとエンタープライズ・サーチの両方にある設定

マイグレーション・ウィザードは、マイグレーション対象の各コレクションについて、以下の設定をマイグレーションします。

- ポータル検索エンジン・コレクション内のポータル検索エンジン・サイト
- コレクション言語
- 分類構造 (カテゴリー・ツリー)、およびエンタープライズ・サーチ・コレクションでルール・ベースのカテゴリーを使用する場合はルール・ベース・カテゴリーの規則

コレクション内の各ポータル検索エンジン・サイトは、エンタープライズ・サーチ Web クローラーに統合されます。マイグレーション・ウィザードは、以下のクローラー設定をマイグレーションします。

- 開始 URL
- 並行クロール・プロセスの数
- クロールの深さ
- 文書検索のタイムアウト値 (秒)
- デフォルトの文字セット
- クロール対象への組み入れ規則および対象からの除外規則

### エンタープライズ・サーチだけにある設定

コレクションをマイグレーションする場合、そのコレクションに関する情報を指定します。マイグレーション・ウィザードでは、それらの設定をマイグレーションし、コレクションに対するエンタープライズ・サーチのデフォルト設定を使用して、マイグレーションした各コレクションを構成します。

エンタープライズ・サーチ管理コンソールを使用すれば、このコレクションおよび Web クローラー構成を変更することができます。括弧内に示された値は、マイグレーションされたデータに対するデフォルト設定です。

- コレクション名
- 文書の静的ランキング計画
- 使用されるカテゴリー化のタイプ: ルール・ベース、カテゴリー化なし、など



- 検索キャッシュを使用するかどうか、および検索キャッシュが保持できる照会および検索結果数 (使用する、5,000)
- 検索応答時間をモニターし、限度を超えたらアラートを発行するかどうか (モニターする、限度 5 秒)
- アクセス制御を使用するかどうか (しない)
- 索引のリフレッシュ・スケジュール
- 索引の再編成スケジュール
- ログの詳細レベル (すべてのメッセージ)

また、マイグレーション・ウィザードでは、各 Web クローラーに対して以下の設定を作成します。

- クローラー名
- クローラーの説明
- 最大ページ長
- 文書セキュリティ設定
- データ・ソース・タイプで該当する場合、クロールする必要がある文書の MIME タイプ

新規にマイグレーションされた Web クローラーを開始する前に、クローラー・プロパティとクロール・スペース設定を全部よく検討し、必要な値がすべて指定されていることを確認してください (必要フィールドは、赤いアスタリスクでマークされています)。特に、クローラーに関するコメントを受信する E メール・アドレスとクローラーのユーザー・エージェント名を指定していることを確認してください。Web クローラー・プロパティの構成中に不明な点があったら、「ヘルプ」をクリックしてください。

---

## マイグレーション・ウィザードのログ・ファイル

マイグレーション・ウィザードでは、すべてのメッセージを、マイグレーション・ウィザードがインストールされているディレクトリーにある WpsMigratorLog.log ファイルに書き込みます。

マイグレーション済みの各コレクションについて、WpsMigratorLog.log ログ・ファイルには、WebSphere ポータル検索エンジンから読み込まれたすべての設定値、およびそれらの設定値がエンタープライズ・サーチ・コレクションにインポートされた場所が格納されています。



---

## エンタープライズ・サーチ・サーバーの開始と停止

コレクションを作成後、データのクロール、構文解析、索引作成を行うためのサーバーを開始する必要があります。コレクションに変更を加えた後も、サーバーを停止して再始動してください。

ほとんどのエンタープライズ・サーチ・サーバーは、連続して、または指定されたスケジュールに従って稼働できます。例えば、索引の再編成またはリフレッシュのスケジュールを指定できます。データ解析用および索引の検索用のサーバーを開始した後、構成設定を変更する (カテゴリーの変更や検索キャッシュのサイズ増加など) 場合以外、通常、それらのサーバーを停止し再始動する必要はありません。

コレクションの内容を変更した場合、あるいは、エンタープライズのソースからクローラーがデータを収集する方法の規則を変更した場合は、それらの変更が有効になるように、通常、クローラーを停止し、再始動する必要があります。クロール規則を変更していなければ、クローラーは、連続して実行されるか (Web および Nntp クローラーの場合)、指定されたスケジュールに従って実行されます。

---

## エンタープライズ・サーチ・サーバーの開始

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

### 始める前に

クロールしたいデータ・ソースを構成し、そのデータがどのように構文解析、索引付け、検索されるのかに関するオプションを指定します。例えば、ユーザーが検索結果にカテゴリー詳細を表示できるようにしたい場合は、パーサーを開始する前にカテゴリーを構成します。

エンタープライズ・サーチ・サーバーを開始するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者、またはそのコレクションを管理する権限をもつオペレーターである必要があります。

エンタープライズ・サーチ・サーバーの開始は、正しい順序で行う必要があります。例えば、クローラーを開始してデータのクロールを行った後でないと、文書の構文解析および索引付けはできません。

### 手順

エンタープライズ・サーチ・サーバーを開始するには、次のようにします。

1. エンタープライズ・サーチ管理コンソールまたは検索アプリケーションを使用するには、次のようにします。
  - a. IBM HTTP Server がまだ稼働していない場合は、開始します。

- b. ESAdmin およびESSearchApplication エンタープライズ・アプリケーションがまだ稼働していない場合は、WebSphere Application Server 管理コンソールを使用して開始します。

2. エンタープライズ・サーチの共通通信層 (CCL) が稼働していない場合は、システムを始動します。

- a. 索引サーバーで、WebSphere II OmniFind Edition のインストール時に指定されたエンタープライズ・サーチ管理者のユーザー ID でログインします。

#### UNIX システム

次のコマンドを入力します。

```
startccl.sh -bg
```

#### Windows システム

「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」を選択して、IBM WebSphere Information Integrator OmniFind Edition サービスを再始動します。


3. エンタープライズ・サーチ・システムが稼働していない場合は、システムを始動します。


- a. 索引サーバーで、WebSphere II OmniFind Edition のインストール時に指定されたエンタープライズ・サーチ管理者のユーザー ID でログインします。

- b. 次のコマンドを入力します。

```
esadmin start
```

4. エンタープライズ・サーチ管理コンソールを始動して、エンタープライズ・サーチ管理者としてログインします。管理役割を使用した場合、開始したいコレクションに対する権限をもつコレクション管理者またはオペレーターとしてログインできます。


5. 「コレクション」ビューで、管理したいコレクションを指して、 「モニター」をクリックします。

6. 「クロール」ページで、開始したい各クローラーごとに、 「開始」をクリックします。


- Web クローラーまたは NNTP クローラーを開始すると、クローラーは即時にデータのクロールを始めます。これらのタイプのクローラーは、連続して稼働し、Web サイトおよび NNTP ニュースグループの文書のクロールと再クロールを行ないます。

- 他のクローラー・タイプのいずれかを開始すると、クローラー・セッションが開始します。クローラーは、スケジュールされた日時にクロールを開始します。クローラーをスケジュールしていない場合、またはクローラーを早く開始したい場合は、クローラーをモニターして、クロールしたい各データ・ソースの開始アイコンをクリックします。

クローラーが開始した後は、それを続けて実行させておくことができます。クローラーをスケジュールしていれば、スケジュール日時に再び実行されます。

7. データのクロールが終わったら、「構文解析」ページを開き、 「開始」をクリックしてパーサーを開始します。

パーサーは続けて実行させておくことができます。通常、データの構文解析方法 (カテゴリーまたは XML フィールド・マッピングの更新) を変更しない限り (例えば、カテゴリーまたは XML フィールド・マッピングの更新)、パーサーを停止する必要はありません。

- オプション: 索引付けがスケジュールされた日時に開始するのを待たずに、索引付けプロセスを強制始動するには、索引」ページを開き、「再編成」領域で  「開始」をクリックします。

索引付けプロセスは続けて実行させておくことができます。索引は、スケジュール日時にリフレッシュおよび再編成されます。

- 「検索」ページで、 「開始」をクリックします。

検索サーバーは続けて実行させておくことができます。通常、検索キャッシュまたは文書サマリーの設定を変更しない限り、検索サーバーを停止する必要はありません。

#### 関連概念

15 ページの『エンタープライズ・サーチ管理の概要』

コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

212 ページの『管理役割』

エンタープライズ・サーチは管理コンソールのさまざまな機能へのアクセスを制御するために、役割の概念を使用します。

#### 関連タスク

19 ページの『管理コンソールへのログイン』

エンタープライズ・サーチ・システムを管理するには、Web ブラウザーで URL を指定した後で、管理コンソールにログインします。

#### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## エンタープライズ・サーチ・サーバーの停止

エンタープライズ・サーチ・コンポーネントの構成に変更を加えた場合や、問題のトラブルシューティングが必要な場合などには、エンタープライズ・サーチ・サーバーを停止して再始動する必要があります。

#### 始める前に

エンタープライズ・サーチ・サーバーを停止するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者、またはそのコレクションを管理する権限をもつオペレーターである必要があります。

## このタスクについて


エンタープライズ・サーチ・サーバーは、他のサーバーとは無関係に停止することができます。例えば、クローラーの構成を変更し、それを組み込むためにクローラーを停止して再始動する場合、パーサーを停止して再始動する必要はありません。

個々のサーバーを停止する代わりにエンタープライズ・サーチ・システムを停止したい場合、エンタープライズ・サーチ管理者 ID で索引サーバーにログインできます (この ID は WebSphere II OmniFind Edition のインストール時に指定されたものです)。次に、以下のコマンドを入力します。

```
esadmin stop
```

## 手順

エンタープライズ・サーチ・サーバーを停止するには、次のようにします。


1. 「コレクション」ビューで、管理したいコレクションを指して、 「モニター」をクリックします。
2. 「クロール」ページで、管理したいクローラーを見つけ、停止または休止します。


クロール・スペースまたはクローラーのプロパティを変更する場合は、変更を組み込むために、クローラーを停止して再始動してください。クロール・スペースを変更し、すでにコレクションに文書に変更を適用したい場合は、文書を再クロールすることも必要です。


**ヒント:** プロセスがバックグラウンドでまだ実行されているにもかかわらず、要求された操作がタイムアウトになったというメッセージが表示される場合があります。タスクが完了したかどうかを判断するには、管理コンソールで「リフレッシュ」をクリックします (Web ブラウザーの「リフレッシュ」はクリックしないでください)。プロセスは、クローラーの状況アイコンが停止したことを示したら完了です。

3. 「構文解析」ページで、 「停止」をクリックして、パーサーを停止します。

データを構文解析するための規則を変更する場合は、変更を組み込むために、パーサーを停止して再始動してください。変更は、新たにクロールされた文書のみ適用されます。すでに索引に入っている文書に変更を適用したい場合は、フル・クロールを開始して、すべての文書を再クロールする必要があります。これにより、すべての文書が構文解析され、再度、索引付けられます。

4. 「索引」ページで、 「停止」をクリックして、リフレッシュまたは再編成中の索引を停止します。

また、索引キューをモニター中にも、索引作成を停止できます。これを行うには、ツールバーで「システム」を選択し、「索引」ページを開いて、作成を停止したい索引について 「停止」をクリックします。

5. 「検索」ページで、 「停止」をクリックして、検索サーバーを停止します。通常、検索サーバーを停止して再始動する必要があるのは、検索キャッシュまたは文書サマリーの設定を変更した場合だけです。

## 関連概念



15 ページの『エンタープライズ・サーチ管理の概要』

コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

#### **関連資料**

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。



---

## エンタープライズ・サーチ・アクティビティのモニター

システム・アクティビティおよびコレクション・アクティビティをモニターすると、さまざまなプロセスの状況の表示、問題発生の可能性の監視、または、パフォーマンス改善のための構成設定の調整を行うことができます。

エンタープライズ・サーチ管理コンソールを使用すれば、システムをモニターし、必要に応じて操作を調整することができます。主要アクティビティ（クロール、構文解析、索引作成、検索）については、詳細な統計データを参照することができます。統計データには、平均応答時間や、クロールまたは索引作成セッションでいくつの文書がクロールまたは索引作成されたかなどの進行状況データがあります。

ほとんどのアクティビティは、アイコンをクリックするだけで、開始および停止することができます。この開始/停止操作により、アクティビティを一時停止して、構成の変更または問題のトラブルシューティングを行い、アクティビティを進められる状況になった時点で処理を再開することができます。

### 関連概念

15 ページの『エンタープライズ・サーチ管理の概要』

コレクションおよび外部ソースの作成と管理、コンポーネントの開始と停止、システム・アクティビティおよびログ・ファイルのモニター、管理ユーザーの構成、検索アプリケーションのコレクションおよび外部ソースへの関連付け、セキュリティ施行情報の指定には、エンタープライズ・サーチ管理コンソールを使用します。

### 関連タスク

251 ページの『エンタープライズ・サーチ・サーバーの開始』

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

253 ページの『エンタープライズ・サーチ・サーバーの停止』

エンタープライズ・サーチ・コンポーネントの構成に変更を加えた場合や、問題のトラブルシューティングが必要な場合などには、エンタープライズ・サーチ・サーバーを停止して再始動する必要があります。

### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## 単一コレクション内の文書数の見積もり

エンタープライズ・サーチ・コレクションを作成または編集するときは、保持しておきたいコレクションの文書数の見積もりを用意します。リソース・マネージャーは、この数値を使用して、コレクションに必要なメモリーとディスク・リソースを見積もりますが、コレクションのサイズを制限することはありません。

### 始める前に

コレクションの見積もりサイズを変更するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

コレクションが大きくなって、見積もったサイズに達しても、システムは索引への文書追加を停止しません。リソース・マネージャーは、メモリーおよびディスク・リソースの所要量見積もりが、システムで現在使用可能になっている量より多くなると、警告を出します。この警告によって、将来のリソース不足を防止することができます。

コレクションについてアラートを構成し、索引内の文書の数が増えすぎたときオプションを通知できるようにした場合、デフォルトの限度は、コレクション内の見積もり文書数に指定した値と一致します。モニターは、指定されたアラートしきい値パーセントと一緒にこの数値を使用し、コレクションについて構成された最大文書数に近づいてくると、Eメールを送信します。

### 手順

コレクションの可能なサイズを見積もるには、次のようにします。

1. コレクションを編集、「一般」ページを選択、そして「**一般オプションの構成**」をクリックします。
2. 「**見積もり文書数**」フィールドに、コレクションの成長が見込まれる大きさを表す数値を入力します。デフォルト値は 1,000,000 文書です。

---

## システム・リソースの可用性のチェック

コレクションまたはクローラーを作成してから文書のクロールを開始するまでにシステム・リソースをチェックすると、使用可能なシステム・リソースが、現在の構成設定に基づき、WebSphere II OmniFind Edition を全能力で実行するのに十分であるかどうかを確認できます。

### 始める前に

システム・リソースをチェックするには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて

コレクションまたはクローラーを作成するときに、システムは、リソースの可用性を自動的にチェックします。オプションを選択して、いつでもシステム・リソースの可用性をチェックすることもできます。

システムは、コレクションおよびクローラーのプロパティを構成する際に使用すると見積もったスペースを、使用可能なシステム・スペースと比較して、リソースの可用性を通知するメッセージを表示します。メッセージは、リソースが十分でないサーバー、コレクションの見積もりサイズに必要なスペースの量、使用可能なフリー・スペースの量を示します。

リソースが十分でない場合は、以下の修正処置のいずれかを試みてください。

- メッセージ・テキストに示されているファイル・システムのサイズを増やす。
- コレクションの一般オプションを編集して、見積もり文書数にもっと小さい数値を指定する。
- クローラー・プロパティを編集して、クロールする文書の最大数、最大ページ・サイズ、スレッドの最大数にもっと小さい数値を指定する。
- Web クローラーについて、クローラー・プロパティを編集して、アクティブ・ホストの最大数、新規文書の最大数、一時ストレージ内の文書の最大数にもっと小さい数値を指定する。
- システムの索引プロパティを編集して、並行実行を許可される索引作成の回数を減らす。
- 複数のコレクション内で、いつでも同時にアクティブなプロセスの数が多すぎないようにする。例えば、複数のコレクションからのクローラーとパーサーが同時に実行すると、リソースが影響を受ける可能性があります。
- システムからクローラーを削除する。
- システムからコレクションを削除する。

リソースが少ない可能性について警告が検出されなかった場合は、クロールされて索引作成されるデータに対応できるだけの十分なリソースが存在すると考えられます。

### 手順

エンタープライズ・サーチ・システムの現在構成に大して十分なリソースが使用可能であるかどうかを判断するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「一般」ページで、「システム・リソースのチェック」をクリックします。

---

## コレクションのモニター

コレクション内の各コンポーネントの状態に関する一般情報を表示するか、または、個々のコンポーネントおよび URI に関する詳細情報を表示するオプションを選択することができます。


### 始める前に


すべてのエンタープライズ・サーチ管理ユーザーは、コレクションをモニターできます。コンポーネントの開始または停止を行うか、スケジュールを使用可能/使用不


可にするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

## 手順

コレクションをモニターするには、次のようにします。

1. 「コレクション」ビューで、モニター対象にしたいコレクションを見つけて、 「モニター」をクリックします。各コレクション・コンポーネントの現在の状況が表示されます。

**ヒント:** コレクションを編集集中であり、すでに「一般」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

2. URI に関する詳細情報を表示したい場合は、 「URI 詳細」をクリックします。

例えば、特定の URI が索引中にあるかどうかや、その URI が含まれる索引が検索サーバーにコピーされたかどうかを見ることができます。

3. 個々のコンポーネントをモニターし、そのコンポーネントのアクティビティに関する詳細な統計を表示するには、「状況」アイコンをクリックします。

## 関連概念

29 ページの『エンタープライズ・サーチ・コレクション』

エンタープライズ・サーチ・コレクションには、ユーザーが 1 回の照会で検索できる全ソースが格納されています。フェデレーションによって、ユーザーは複数のコレクションを 1 回の照会で検索することができます。

---

## URI についての詳細表示

URI に関する詳細情報を見ることができます。この URI で表される文書がどのようにクロール、索引付け、検索されるのかについての現在情報および履歴情報を知ることができます。

### 始める前に

URI レポートを表示する要求、またはレポートを E メール・アドレスへ送信する要求をサブミットする前に、情報の発信元になるコンポーネントがアクティブであることを確認してください。例えば、文書がどのようにクロール、索引付け、検索されるのかについての詳細情報を表示する場合には、Web クローラー、索引サーバー、検索サーバーが稼働中であることを確認してください。除去された文書をトラッキングするには、文書トラッキングのログイン・オプションが構成されていることを確認してください。

### このタスクについて

URI に関する情報の収集は、時間のかかるプロセスです。情報を要求した後、情報が表示されるのを待つというオプションを選択できます。より効率的なオプションは、指定した E メール・アドレスにレポートを送信するようにするオプションです。





レポートを受信できるようにするには、その前に、エンタープライズ・サーチに対してメール・サーバーの情報が構成されていることを確認する必要があります。この情報は、「システム」ビューの「ログ」ページで E メール・オプションを構成するときに指定します。


索引サーバーと検索サーバーは、すべての URI に関する情報 (例えば、ある URI が索引にあるかどうか、それが検索サーバーにコピーされたかどうか) を提供できます。文書がどのようにクロールされたかに関する情報を表示するには、Web クローラーによってクロールされた文書の URI を指定する必要があります。

## 手順

URI に関する詳細を表示するには、次のようにします。

1. 「コレクション」ビューで、モニター対象にしたいコレクションを見つけて、 「モニター」をクリックします。

**ヒント:** コレクションを編集中であり、すでに「一般」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

2.  「URI の詳細」をクリックします。
3. 「URI の詳細」ページで、情報を見たい URI を入力します。
4. 見たい情報のタイプに合わせてチェック・ボックスを選択します。

### クローラー詳細 (Web クローラーにのみ使用可能)

このチェック・ボックスを選択すると、Web クローラーによって文書がどのようにクロールされたかに関する情報と、クロール・スペース内の現在の状況に関する情報が示されます。

### 索引の詳細

このチェック・ボックスを選択すると、文書が索引付けされたかどうかと、文書が検索サーバーにコピーされたかどうかを示されます。

### 検索の詳細

このチェック・ボックスを選択すると、文書がどのように検索できるかに関する情報と、文書が検索に使用可能かどうかを示されます。

### パーサーによって除去された文書

その文書が構文解析中にエンタープライズ・サーチ・システムから除去されたのかどうか、またその場合の除去された理由を参照する場合は、このチェック・ボックスを選択します。

### 索引から除去された文書

その文書が索引作成中またはアナライズ中にエンタープライズ・サーチ・システムから除去されたのかどうか、またその場合の除去された理由を参照する場合は、このチェック・ボックスを選択します。

5. レポートが表示されるのを待つ場合は、「レポート表示」をクリックします。
6. 後でレポートを表示できるように E メール・アドレスにレポートを送信するには、「レポートの送信」をクリックします。
  - a. 「詳細 URI レポートの送信」ページで、「通知を受け取る E メール・アドレス」フィールドに、レポートを受け取る E メール・アドレスを入力します。

- b. 「レポートの送信」をクリックします。

#### 関連概念

278 ページの『文書トラッキング』

文書は、処理の様々な段階でシステムから除去することができます。文書が除去された時、および除去の原因となった問題の内容を確認する場合に、このオプションを指定できます。

#### 関連タスク

279 ページの『除去された文書に関するレポートの表示』

エンタープライズ・サーチ・システムから除去された文書に関する詳細情報を表示することができます。この情報は、コレクションに対する文書トラッキングを使用可能にした場合にのみ有効です。

#### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』

エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

---

## クローラーのモニター


コレクション内の各クローラーの状態に関する一般情報を表示するか、または、クローラーのアクティビティーに関する詳細情報を表示するオプションを選択することができます。


### 始める前に


コレクションのモニター限定の管理役割を使用している場合、クローラー統計を表示することはできませんが、クローラーの動作の変更 (クローラーの開始や停止など) はできません。

### 手順


クローラーをモニターするには、次のようにします。

1. 「コレクション」ビューで、モニターしたいコレクションを見つけて、 「モニター」をクリックします。
2. 「クロール」ページを開きます。

**ヒント:** コレクションを編集集中であり、すでに「クロール」ページが表示されている場合は、 「モニター」をクリックすれば、クローラーをモニターするためのビューに移動できます。

3. クローラーが稼働中または休止されていて、クローラーに関する詳細な状況情報を表示したい場合、 「詳細」をクリックします。表示される統計のタイプは、クローラー・タイプによって異なります。

コレクションのプロセス管理を許可する管理役割を持っている場合、クローラー・アクティビティーに関する詳細を表示中に、クローラーを開始、停止、休止することができます。クローラーがスケジュール可能である場合、クロールのスケジュールを使用可能または使用不可にすることもできます。

4. クローラーが停止または休止されていて、クローラー・セッションを開始したい場合、 「開始」または「再開」をクリックします。

#### Web クローラーの場合:

クローラーが停止されていた場合、クローラーは再びクロールを開始し、クロール・スペース全体をクロールします。クローラーが休止されていた場合、クローラーは、休止されたときのターゲットの先頭からクロールを再開します。

即時にクローラーにフル・クロールを開始させたい場合、「詳細」アイコンをクリックして、その後、「再フル・クロールの開始」アイコンをクリックします。クローラーは、最後にクロールしてから変更が行われなかったページも含めて、クロール・スペース全体のクロールを開始します。すべての文書を再クロールしたい場合があります。例えば、文書の構文解析の規則を変更して、以前に索引作成した文書にその規則を適用したい場合などです。

#### NNTP クローラーの場合:

クローラーが停止されていた場合、クローラーは再びクロールを開始し、クロール・スペース全体をクロールします。クローラーが休止されていた場合、クローラーは、休止されたときのターゲットの先頭からクロールを再開します。

#### 他のすべてのクローラー・タイプの場合:

クローラーが停止されていた場合、クローラーは、スケジュールされた日時にクロールを開始します。クローラーがデータ・ソースをクロールする最初の時、クローラーはフル・クロールを実行します。クロールが繰り返しスケジュールされている場合、クローラーはデータ・ソースに対するすべての更新 (文書の追加、削除、および変更) か、または文書の追加および変更のみのどちらかをクロールします。クローラー・スケジュールでクロールのタイプを構成します。

クローラー・スケジュールがまだ行われていない場合、またはクローラーを早く開始したい場合は、「詳細」アイコンをクリックしてください。その後、クロール・スペースの詳細領域で、開始したいクロールのタイプのアイコン「フル・クロール」、「すべて更新」、または「新規および変更文書のみ」をクリックします。クロールしたい各データ・ソース (サーバー、データベース、サブフォルダーなど) の該当する開始アイコンをクリックする必要があります。

- 稼働中のクローラーを停止したい場合、 「停止」または  「休止」をクリックします。クローラーは、再始動または再開されるまで、データのクロールを停止します。

休止されたクローラーを再開する場合、クローラーは、休止されたターゲットの先頭でクロールを開始します。例えば、DB2 クローラーは、停止されたときにクロール中だった表の先頭行からクロールを再開します。

#### 関連概念

37 ページの『エンタープライズ・サーチ・クローラーの管理』

コレクションに入れるデータ・タイプに応じて、さまざまなタイプのクローラーを構成できます。1 つのコレクションに任意の数のクローラーを入れることができます。

#### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』  
コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## Web クローラーのアクティビティ詳細の表示


Web クローラーのアクティビティに関する詳細を表示することによって、全体的なパフォーマンスを査定し、必要に応じて、Web クローラーのプロパティおよびクロール・スペース定義を調整できます。


### 始める前に


すべてのエンタープライズ・サーチ管理ユーザーは、クローラーのアクティビティをモニターできます。クローラーの開始または停止を行うには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

Web クローラーのアクティビティに関する詳細を表示するには、次のようにします。

1. 「コレクション」ビューで、モニター対象にしたい Web クローラーを所有するコレクションを見つけ、 「モニター」をクリックします。
2. 「クロール」ページを開きます。

**ヒント:** コレクションを編集集中であり、すでに「クロール」ページが表示されている場合は、 「モニター」をクリックすれば、クローラーをモニターするためのビューに移動できます。

3. モニターしたい Web クローラーが稼働中であるか休止している場合、 「詳細」をクリックします。
4. Web クローラーの詳細ページで、クローラーの現在と過去のアクティビティに関する詳細な統計を表示するため、以下のオプションを確認または選択します。
  - a. 「スレッドの詳細」をクリックすると、アクティブに Web サイトをクロールしているスレッドの数と、非アクティブ状態のスレッドの数が示されます。
  - b. 「アクティブ・サイト」をクリックすると、クローラーがアクティブにクロール中の Web サイトに関する情報が示されます。
  - c. 「最近クロールされた URL」をクリックします。この情報は、クローラーが最近クロールした対象を示します。ビューをリフレッシュしてもリスト中の項目が変わらない場合、発生しているクロールはありません。
  - d. 「クローラー履歴」をクリックすると、クローラーの過去のアクティビティに関するレポートが表示されます。
  - e. 「URL 状況」領域で、情報を表示したい Web サイトの URL を入力し、「表示」をクリックします。

例えば、このオプションを使用して、ある URL がクロール・スペース内にあるかどうか、それがクロールされたのか、または、単にディスカバーされただけなのか、再度クロールされるべきなのはいつかを見たり、Web サイトに対する最新クロール試行に関する情報を見ることができます。

URL に関する詳細が表示された後、「**サイト履歴**」をクリックすると、その URL でのクローラーのアクティビティーに関する追加の統計情報を見ることができます。

## Web クローラーのスレッドの詳細

Web クローラーをモニターすることによって、アクティブに Web サイトをクロールしているスレッドの数と、非アクティブ状態のスレッドの数を見ることができます。

コレクションのモニター中に Web クローラーに関する詳細を表示すると、クローラー・スレッドの状況を見ることができます。一般的な状態には次のものがあります。

**待機中** スレッドにはクロールすべき URL がないことを示します。この状態は、スレッドがクロールを完了し、クローラーがそれ以上のクロールすべき URL を見つける速さが不十分な場合に発生します。例えば、クローラーのプロパティーの設定で、クローラーが同じサイトから別のページを取得できるまでに待たなければならない時間が長すぎる場合には、URL が供給される速度が十分ではないことがあります。

### フェッチ中

スレッドが Web サイトからページをダウンロード中であることを示します。

**完了** スレッドが、クロールしたページをクローラーの残りに送信中であるが、別の URL をクロールする準備はまだできていないことを示します。

**中断** クローラーが休止していることを示します。

理想的なのは、すべてのスレッドが常にページをフェッチ中であることです。スレッドが頻繁に完了状態になる場合は、データベースにスループットの問題があると考えられます。

スレッドが頻繁に待機状態になる場合は、クローラーのプロパティーの「**アクティブ・ホストの最大数**」フィールドに指定された値を検討してください。この値が小さい場合、スレッドを継続的にビジーにするにはクロール・スペース内のサイト数が十分でないか、クロールに適格な URL の数が十分でないことが考えられます。低アクティビティーを引き起こす条件として、DNS ルックアップ失敗とロボット・ルックアップ失敗があります。

## Web クローラー・アクティブ・サイト

Web クローラーをモニターして、クローラーがアクティブにクロール中の Web サイトに関する情報を見ることができます。

コレクションのモニター中に Web クローラーに関する詳細を表示すると、アクティブ・サイトに関する統計を見ることができます。統計には、以下の情報が示されます。

- 現時点でクローリングのためにクローラーが内部データベースからメモリーへ移した URL の数。
- これまでにクローラーがクローリングを試行した URL の数。
- サイトが非アクティブにされてクローラーのこの回の反復用のメモリーから除去されるまでの残り時間。
- これまでにサイトがメモリー内にあった時間。

この情報は、クローラーに対して構成されたクローリング規則をクローラーが進んでいくのに従って刻々と変化します。アクティブにされた URL 数が、クローラーのプロパティの「**アクティブ・ホストの最大数**」フィールドに構成された値と近いのが理想的です。

アクティブにされた URL 数がゼロに近い場合、クローラーは適格 URL を検出していません。そのような低いアクティビティを引き起こす可能性のある条件には、DNS ルックアップの失敗、ネットワーク接続性の問題、データベース・エラー、クローリング・スペース定義の問題があります。例えば、次のようになります。

- 長時間にわたってメモリー内に多数のサイトがあり、クローリングされた URL が少数である場合、ネットワーク接続性の問題がないか調べてください。
- 十分な数のサイトがリストにある場合、クローリング・スペース定義の問題、または DNS ルックアップの問題がないか調べてください。
- 適切な速度でサイトがクローリングされているが、多数の URL がクローリングされないままメモリーからサイトが出ている場合、クローラーのプロパティを編集し、「**メモリー内 URL の最大残存時間**」フィールド内のタイムアウト値を調整して、サイトをメモリーに保持する時間を長くしてください。

## Web クローラーのクローリング速度

Web クローラーをモニターして、クローラーが Web サイトからページをダウンロードしている速度に関する情報を見ることができます。

コレクションのモニター中に Web クローラーに関する詳細を表示すると、クローラーがデータをクローリングする速さ (クローリング速度) に関する統計を見ることができます。また、現行セッションの開始以降にクローラーがクローリングした URL の数に関する統計も見ることができます。

クローリング速度は、1 秒あたりにクローリングされているページ数です。この数は、Web クローラーに対して構成できるいくつかのプロパティと関連しています。

- クローラー・スレッドの数
- アクティブ・サイトの数
- クローラーが同じ Web サーバーから別のページを取得できるまでに待たなければならない時間。

クローラーがクローラー・スレッドごとに 1 つのアクティブ・サイトを持っている場合で、クローラーが 2 秒間待った後でないと同じ Web サーバーから別のページを取得できない場合、クローラーは 2 秒あたりに、スレッド当たり 1 ページより



速くクローリングすることはできません。例えば、クローラーがデフォルト数 (200) のスレッドを使用する場合、クローラーは 200 スレッドで 1 秒当たり 100 ページをクローリングできます。

クローラー・スレッド数の倍のアクティブ・サイトがある場合で、クローラーが 2 秒間待った後でないと同じ Web サーバーから別のページを取得できないとすると、クローラーは 1 秒にスレッド当たり 1 ページの速度に達すると考えられます。しかし、そうすると、ネットワークのダウンロード速度およびデータベース・スループットが速度を限定する要因になってきます。良好なクローラー・パフォーマンスが示されるのは、クローリング速度が、クローラー・スレッド数、アクティブ・サイト数、およびクローラー待ち時間と調和している場合です。

Web クローラーのパフォーマンスをモニターするときに検討が必要なもう 1 つの因子は、現行クローラー・セッションの開始以降にクローラーがクローリングした URL の数です。この数を、クローラーが稼働した合計時間で除算した値が、長期間の平均スループットです。この数値が増加しない場合、クローラーは完了しているか、進行できないかのいずれかです。例えば、ネットワーク接続エラー、データベース・エラー、DNS ルックアップ失敗などがあると、クローラーの進行が妨げられます。

## Web クローラー・レポートの作成

Web クローラーの過去のアクティビティに関するレポートを表示することによって、全体的なパフォーマンスを査定し、必要に応じて、Web クローラーのプロパティおよびクローリング・スペース定義を調整できます。

### 始める前に

コレクションのモニター限定の管理役割を使用している場合、クローラー統計を表示することができ、クローラーのアクティビティに関するレポートを作成することもできますが、クローラーの動作の変更 (クローラーの開始や停止など) はできません。


### このタスクについて


さまざまなタイプのレポートがあり、Web クローラーのアクティビティに関する情報が提供されます。レポートのタイプによっては、クローラーの内部データベースから収集できる限りの速度で情報が戻されるものもあります。サイト・レポートおよび HTTP 戻りコード・レポートは、作成に時間を要します。これらのタイプのレポートを作成する場合、結果がエンタープライズ・サーチ管理コンソールに戻されるのを待つ代わりに、レポートを受け取る E メール・アドレスを指定できます。


レポートでの統計の解釈方法については、Web クローラーのモニター中およびレポートの作成中に「ヘルプ」をクリックしてください。

### 手順

Web クローラー・レポートを作成するには、次のようにします。

1. 「コレクション」ビューで、モニター対象にしたい Web クローラーを所有するコレクションを見つけ、 「モニター」をクリックします。
2. 「クローリング」ページを開きます。

**ヒント:** コレクションを編集集中であり、すでに「クローラ」ページが表示されている場合は、 「モニター」をクリックすれば、クローラーをモニターするためのビューに移動できます。

3. レポートを作成したい Web クローラーが稼働中であるか休止している場合、 「詳細」をクリックします。
4. Web クローラーの詳細ページで、作成したいレポートのタイプに応じてオプションを選択します。
  - クローラーと、クローラーがディスカバーまたはクロールするすべてのサイトに関するレポートを作成するには、「クローラー状況サマリー」領域で「クローラー履歴」をクリックします。
  - 「URL 状況」領域で、レポートを作成したい特定のサイトの URL を指定し、「表示」をクリックし、「サイト履歴」をクリックします。
5. クローラー履歴とサイト履歴の両方のレポートについて、レポートに入れたい統計のチェック・ボックスを選択し、「レポート表示」をクリックします。

これらのタイプの統計について、クローラーは、内部データベースから情報を取得できる限りの速度で、レポートを管理コンソールに戻します。

6. クローラー履歴レポートを作成する場合、サイト・レポートを作成するためのオプションを指定し、「レポートの実行」をクリックします。

このレポートは、選択した統計が組み込まれて作成され、指定するファイルに保存されます (このファイルの名前は絶対名でなければなりません)。レポート作成後に E メールを受け取ることを指定できます。

7. クローラー履歴レポートを作成する場合、HTTP 戻りコード・レポートを作成するためのオプションを指定し、「レポートの実行」をクリックします。

このレポートには、サイトごとに分布している HTTP 戻りコードの数についての情報が示されます。このレポートは、指定するファイルに保存されます (このファイルの名前は絶対名でなければなりません)。レポート作成後に E メールを受け取ることを指定できます。

このレポートを使用すると、4xx 戻りコード (ページが見つからなかったことを示す)、5xx 戻りコード (サーバー問題を示す)、6xx 戻りコード (接続での問題を示す) 等を多く戻しているサイトを確認できます。

このレポートは、クローラーがある程度の時間アクティブであった場合に最も利用価値があります (例えば、数週間にわたってアクティブであったクローラーの場合など)。このレポートは、消失したサイト、新着サイト、大量の URL が含まれるサイト (Lotus Notes データベースの重複クロールを表している可能性があります)、HTTP サーバーによる再帰的ファイル・システムのサービスを受けているサイトを識別するのに役立ちます。多数の HTTP 戻りコードがあるサイトが索引に寄与していない場合、それらのサイトをクロール・スペースから除去することによって、クローラーのパフォーマンスを向上させることができます。

## Web クローラーの HTTP 戻りコード

Web クローラーをモニターするとき、クロールするページからクローラーが受け取る HTTP 戻りコードに関する情報を見ることができます。

## 表のサマリー

Web クローラー履歴をモニターするか、または特定の URL の状況をモニターするとき、クローラーに戻された HTTP 戻りコードを見ることができます。この情報を使用して、クローラー・スペースの管理およびクローラー・パフォーマンスの最適化を行えます。例えば、ある URL に対して多数の HTTP 戻りコードをクローラーが受け取り、それらの戻りコードが、その場所にあるページがクローリングできないことを示すものである場合、その URL をクローラー・スペースから除去することによってパフォーマンスを改善できます。

以下の表に、HTTP 戻りコードと、Web クローラーがそれらをどのように解釈するのかを示します。100 から 505 までの値は標準 HTTP 戻りコードです (詳しくは、<http://www.w3.org/Protocols/rfc2616/rfc2616.html> を参照してください)。それ以外の HTTP 戻りコードは、エンタープライズ・サーチおよび Web クローラー専用のコードです。

表 4. Web クローラーからの HTTP 戻りコード

コード	説明	コード	説明	コード	説明	コード	説明
NULL	未クローリング	401	無許可	500	内部サーバー・エラー	700	構文解析エラー (ヘッダー・エンドなし)
100	継続	402	支払要	501	未実装	710	構文解析エラー (ヘッダー)
200	成功	403	禁止	502	不正なゲートウェイ	720	構文解析エラー (HTTP コードなし)
201	作成されました	404	検出されません	503	サービス利用不可	730	構文解析エラー (本文)
202	受け入れられました	405	許可されないメソッド	504	ゲートウェイのタイムアウト	740	robots.txt ファイルにより除外
203	信頼できない情報	406	受け入れ不能	505	サポートされない HTTPバージョン	741	ロボットが一時的に利用不可
204	コンテンツなし	407	プロキシ認証が必要	611	読み取りエラー	760	クローラー・スペース定義により除外
205	コンテンツのリセット	408	要求タイムアウト	612	接続エラー	770	不正なプロトコルまたは非標準システム・ポート
206	部分的コンテンツ	409	競合	613	読み取りタイムアウト	780	ファイル・タイプ排他により除外
300	多肢選択	410	もう存在しない	614	SSL ハンドシェイク失敗	2004	索引 META タグなし
301	永久に移動済み	411	長さが必要	615	他の読み取りエラー	3020	ソフト・リダイレクト

表 4. Web クローラーからの HTTP 戻りコード (続き)

コード	説明	コード	説明	コード	説明	コード	説明
302	検出されました	412	前提条件失敗	616	FBA 異常	4044	robots.txt ファイルにより除外
303	他を参照	413	要求エンティティが大きすぎる	617	エンコード・エラー		
304	未変更	414	要求 URI が長すぎる	680	DNS ルックアップ失敗		
305	プロキシ使用	415	サポートされないメディア・タイプ				
306	(未使用)	417	予想失敗				
307	一時的リダイレクト						

## 表の注

### 4xx 戻りコード

400 (不正な要求) コードは、めったにありません。HTTP 戻りコード標準によると、4xx コードは、クライアント (クローラー) が失敗したことを示します。しかし、サーバーに問題があるか、クローラーがリンクとして受け取った URL に問題があるのが普通です。例えば、一部の Web サーバーは、サイト・ルート (例えば、<http://xyz.ibm.com/././sales>) からナビゲートしようとする URL を許容しません。他の Web サーバーでは、このような上方ナビゲーションでの問題はなく、クローラーがすでにルートにある場合は親ディレクトリー記号 (..) を無視します。

一部のサーバーはサイト・ルートに対する要求をエラーとして扱い、古くなったリンクが、もう認識または実装されていない操作を要求する可能性があります。もうサービスされていないページに対する要求があると、その要求はもう有効とは見なされないの、アプリケーション・サーバーは例外をスローし、それが原因となって Web サーバーは HTTP 戻りコード 400 を戻します。

**615** Web サイトからデータをダウンロードするクローラー・サーバーで予期しない例外が検出されたことを示します。このタイプの戻りコードが多数ある場合、クローラーでの問題があると考えられます。

### 6xx 戻りコード

615 を除く 6xx 戻りコードは、タイムアウトなど、クローラーにおいて予想される問題を示します。以下の戻りコードについては、修正処置が必要な場合があります。

#### 611、612、613

サイトが低速であるか、ネットワーク・パフォーマンスが不十分であることを示します。

**614** クローラーがセキュア (HTTPS) サイトをクローラーできないことを示します。これらのサイトがアクセス可能であるという確信がある場合、証明書がクローラー・サーバーおよびターゲット Web サー

バー上で正しくセットアップされていることを確認してください。例えば、認識された認証局 (CA) によって認証済みのサイトの場合、クローラーが使用するトラスト・ストアに新しい CA を追加することができます。

また、クロールしようとしているサイトで自己署名証明書がどのように構成されているかも確認してください。クローラーは、自己署名証明書を受け入れるように構成されています。サイトによっては、ルート URL (例えば、<http://sales.ibm.com/>) に対して自己署名証明書を作成し、その後、サブドメイン (例えば、<http://internal.sales.ibm.com/>) でその証明書の使用を試みます。クローラーは、このような方法で使用される証明書を受け入れられません。クローラーが自己署名証明書を受け入れるのは、サブジェクトのドメイン名 (sales.ibm.com) および証明書の署名者が、要求されているページのドメイン名と一致する場合のみです。

- 616** ログイン・フォームが、再認証後にダウンロードにも出現することを表します。
- 617** エンコード・ストリング (文字セット) が正しくないか、または文書に無効なバイトが含まれているために、文書のバイト・コンテンツからストリングを作成できないことを表します。
- 680** おそらくネットワーク・アクセスの問題のため、クローラーが、クロール・スペース内でホストの IP アドレスを入手できなかったことを示します。このタイプのエラーは、クローラーが、いくつかの URL をクロールできなかったのではなく、サイト全体をクロールできないことを示します。このタイプの戻りコードが多数ある場合、スループットが大幅に減少します。

#### 7xx 戻りコード

7xx コードは、ほとんどの場合、クロール・スペースにおける規則が原因です。

#### 710 - 730

問題があるためにクローラーがダウンロードを完全にできないこと、または、クローラーがサイトで無効な HTML データを検出したことを示します。これらのタイプの戻りコードが多数ある場合、エンタープライズ・サーチのサポート担当者に支援を依頼してください。

#### 740 または 4044

サイトの robots.txt ファイル内の制限によって文書が除外されているため、ファイルのコンテンツを索引付けできないことを示します。

**740** 除外された文書をポイントするアンカー・リンクを索引に含められることを示します。

**4044** 除外された文書をポイントする文書内のアンカー・リンクが索引でも除外されることを示します。

- 741** サイトにはクロールを許可する robots.txt ファイルがあるが、ダウンロードが失敗したことを示します。URL をクロールできない状態

が繰り返される場合、その URL はクローल・スペースから除去されます。このタイプの戻りコードが多数ある場合、ターゲット・サイトが永久に利用不可なのか一時的に利用不可なのかをチェックしてください。ターゲット・サイトがもう使用可能でない場合、それをクローल・スペースから除去してください。

残りの 7xx 戻りコードは、クローラーがしばらく稼働した後で、クロール・スペースに変更を加えた場合に発生することがほとんどです。通常、これらの戻りコードが発生しても、解決が必要な問題はありません。

**3020** 戻りコード 200 をもつ文書に、別の URL へのユーザー・エージェントを参照するロケーション・ヘッダーが含まれることを表します。

---

## パーサーのモニター

パーサーによって分析される文書に関する情報を表示する必要がある場合に、それらがエンタープライズ・サーチ索引に追加される前にパーサーをモニターします。オプションを使用して、統計を検討し、パーサーのアクティビティを管理できます。

### 始める前に

管理役割がコレクションのモニターに限定されている場合、パーサーの状況を表示することはできますが、パーサーを開始したり、停止したりすることはできません。

### このタスクについて

パーサーの詳細をモニターしているとき、特定の瞬間に合わせて構文解析アクティビティに関する統計を提供するパーサー・アクティビティのスナップショットが表示されます。統計により、クロールが終了し、現在、構文解析中または構文解析を待っている文書の数や、構文解析が終了し、索引への格納を待っている文書の数が示されます。


パーサーがアクティブであれば、メッセージにより、パーサーの状態に関するその他の情報が提供されます。例えば、次のとおりです。


- パーサーは、アクティブに文書を構文解析しています。
- パーサーはアイドル中です。パーサーは、さらに多くの文書が構文解析可能になるまでスリープ状態です。エラーが発生すると、パーサーは再始動を待機します。使用可能なパーサー・サービスがなくなると、パーサーは再始動します (例えば、自動再始動は、パーサー・サービスへの接続が確立できない時、またはパーサー Java 仮想マシンのすべてが他のコレクションのためビジーである場合に発生します)。
- パーサーは休止しています (例えば、パーサーは、索引再編成が完了するまで休止します)。


### 手順

コレクションのパーサーをモニターするには、次のようにします。




1. 「コレクション」ビューで、モニターしたいコレクションを見つけて、 「モニター」をクリックします。
2. 「構文解析」ページを開きます。


**ヒント:** コレクションを編集集中であり、すでに「構文解析」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

3. パーサーが稼働中で、構文解析アクティビティに関する詳細な状況情報を表示したい場合は、 「詳細」をクリックします。

コレクションのプロセス管理を許可する管理役割を持っている場合、構文解析アクティビティに関する詳細を表示中にパーサーを開始および停止することができます。

4. パーサーが停止されていて、それを開始したい場合、 「開始」をクリックします。

最初にコレクションを作成するときには、クローラーがデータのクローリングを開始した後で、パーサーを開始してください。そうすれば、パーサーが分析とカテゴリ化を行うデータが必ずあることになります。構文解析規則を変更しない限り、パーサーは続けて実行させておくことができます。

5. パーサーが実行中であり、それを停止したい場合は、 「停止」をクリックします。

構文解析規則を変更する場合、パーサーを停止して再始動する必要があります。例えば、パーサー構成を変更した場合、パーサーを停止して、再始動した後でないと、変更が有効になりません。

---

## コレクションの索引アクティビティのモニター


作成中の索引の進行を知る必要がある場合、索引スケジュールを使用可能/使用不可にする必要がある場合、または、索引付けアクティビティの開始/停止を行なう必要がある場合に、コレクションの索引をモニターします。


### 始める前に




すべてのエンタープライズ・サーチ管理ユーザーは、索引アクティビティをモニターできます。索引作成の開始または停止を行うか、索引スケジュールを使用可能/使用不可にするには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順


コレクションの索引をモニターするには、次のようにします。

1. 「コレクション」ビューで、モニターしたいコレクションを見つけて、 「モニター」をクリックします。
2. 「索引」ページを開きます。

**ヒント:** コレクションを編集中であり、すでに「索引」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

3. 索引がスケジュール済みであり、スケジュールされた日時にその索引が作成されることを望まない場合、 「スケジュールを使用不可にする」をクリックします。その索引は、スケジュールを使用可能にするか、索引作成プロセスを開始するまで、作成されません。
4. 索引がスケジュール済みであるが、索引作成のスケジュールが使用不可にされている場合、 「スケジュールを使用可能にする」をクリックします。索引は、索引付けスケジュールに指定した日時に、作成のためにキューに入られます。
5. 索引が停止されていて、それを開始したい場合、 「開始」をクリックします。

通常、索引付けは定期的なスケジュールに従って行われます。索引が作成されているときに停止した場合、または、索引のスケジュールを使用不可にした場合、「開始」をクリックして索引作成を強制開始できます。

6. 索引作成がアクティブであり、それを停止したい場合、 「停止」をクリックします。

例えば、索引作成を停止する必要がある場合として、コレクション内で使用されるカテゴリ化のタイプを変更した後に索引の再編成を強制する場合などが考えられます。

7. 索引作成中にエラーが発生する場合、 「エラー」をクリックします。

「ログ・ファイルのコンテンツ」ページが表示されて、索引作成エラーに関する追加情報を表示できます。そのページで、個々のエラー・メッセージを選択して、問題に関する詳細を参照することができます。

### 関連概念

149 ページの『エンタープライズ・サーチ索引の管理』

ユーザーが常に最新の情報にアクセスできるようにするため、エンタープライズ・サーチでは、各コレクションに索引を作成し、定期的にその内容をリフレッシュ、再編成して保守します。

---

## エンタープライズ・サーチの索引キューのモニター

索引キュー内のすべての索引作成の状況の表示、作成中の索引の停止、またはキューからの索引の削除を行うことができます。

### 始める前に

索引キューを管理するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて

複数の索引を同時に作成することができますが、一度にキューに入れられるコレクションは 1 つだけです。システムの索引オプションを構成するとき、いくつの索引がキューおよび索引作成用リソースを同時に共用できるのかを指定します。

## 手順


索引キューをモニターするには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「索引」ページを選択します。


索引キュー内に索引があるコレクションのリストが表示されます。それぞれの索引ごとに、作成中の索引のタイプ (リフレッシュまたは再編成)、索引が索引キューに入れられた時刻、索引作成が始まった時刻 (作成が進行中である場合) が表示されます。


3. 個々の索引を管理するには、「状況」アイコンをクリックします。

例えば、ある索引がどの程度まで完了しているかを見たり、索引内の文書数を見たり、索引スケジュールを使用不可にすることができます。

4. 作成されている索引を停止するには、 「停止」をクリックします。

例えば、カテゴリ規則を変更した場合などに、索引の再編成を強制開始できるように索引のリフレッシュを停止することができます。

索引作成を停止した後に開始するには、スケジュールされている次の開始時刻に索引が索引キューに入るのを待つか、あるいは、「状況」アイコンをクリックして索引をモニターし、次に 「開始」をクリックして索引のリフレッシュまたは再編成を行います。

5. 索引キューから索引を削除するには、 「除去」をクリックします。

### 関連概念

149 ページの『エンタープライズ・サーチ索引の管理』

ユーザーが常に最新の情報にアクセスできるようにするため、エンタープライズ・サーチでは、各コレクションに索引を作成し、定期的にその内容をリフレッシュ、再編成して保守します。

---

## 検索サーバーのモニター

ある特定のコレクションについて検索サーバーのアクティビティーに関する詳細な状況情報を表示するか、または、エンタープライズ・サーチ・システム全体について検索サーバーの詳細な状況情報を表示することができます。


### 始める前に


すべてのエンタープライズ・サーチ管理ユーザーは、管理を認可されているコレクションの検索サーバーをモニターできます。エンタープライズ・サーチ・システムのすべての検索サーバーをモニターするには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

検索サーバーの開始または停止を行うには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。



## 手順

1. 単一のコレクションの検索サーバーをモニターするには、次のようにします。

- a. 「コレクション」ビューで、モニターしたいコレクションを見つけて、 「モニター」をクリックします。
- b. 「検索」ページを開きます。

**ヒント:** コレクションを編集中であり、すでに「検索」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

2. エンタープライズ・サーチ・システムのすべての検索サーバーをモニターするには、次のようにします。
  - a. 「システム」をクリックして「システム」ビューを開きます。
  - b. 「検索」ページを選択します。

3. 検索サーバーが停止されていて、それを開始したい場合、 「開始」をクリックします。
4. 検索サーバーが実行中であり、それを停止したい場合は、 「停止」をクリックします。

検索キャッシュを使用可能または使用不可にした、検索キャッシュのサイズを変更した、またはクイック・リンクを変更した場合は、変更を有効にするために、検索サーバーを停止して、再始動する必要があります。

5. 検索サーバーが検索要求の処理に費やす時間のサマリーを表示するには、「**応答時間履歴**」をクリックします。

このレポートには、ある特定の日に検索要求への応答に検索サーバーが費やした平均時間がミリ秒で示されます。

平均応答時間は、システムのパフォーマンスの指標であり、サービス品質に対応します。応答時間の増加は、システムに負荷がかかっていることを示します。例えば、検索対象のコレクションの数およびコレクション・サイズがシステムの負担になっていることが考えられます。

6. 最も頻繁にサブミットされている照会のリストを表示するには、「**高頻度の照会**」をクリックします。

このレポートでは、最も頻繁にサブミットされた上位 50 の照会におけるキーワードと、ある特定の照会をユーザーがサブミットした回数が見られます。

高頻度の照会を検討することによって、クイック・リンクの候補を識別できます。クイック・リンクを作成すれば、多くのユーザーのために検索品質に良い影響を与えることができます。関連の高い文書が常に検索結果に戻されることを保証できます。

また、これらの照会に答えるようなリソースへのエンタープライズ・ポータルからのリンクを作成するといった場合もあります。例えば、交際費に関する情報をユーザーが頻繁に検索している場合に、イントラネット・ホーム・ページに、交際費手続きを議論するページへのリンクを組み込みます。

7. 最近サブミットされた照会のリストを表示するには、「**最新の照会**」をクリックします。

このレポートでは、最近サブミットされた 50 個の照会におけるキーワードが示されます。

最新の照会を検討することによって、組織における現在の傾向と緊急事態を識別できます。例えば、何らかのトピックに対する関心が急に高まっていることなどが分かります。そのような関心の高まりは、そのトピック用のクイック・リンクの必要性を示していたり、そのトピックを他の方法 (エンタープライズ・ポータルでのリンクの提供など) でユーザーが使用できるようにする必要があることを示している可能性があります。

---

## データ・リスナーのモニター


データ・リスナーをモニターしてその状況を確認し、クライアント・データ・リスナー・アプリケーションのアクティビティーに関する詳細を表示します。

### 始める前に


データ・リスナーをモニターするには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### 手順

データ・リスナーをモニターするには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2. 「データ・リスナー」ページで、状況アイコンを表示して、データ・リスナーがアクティブであるのか、停止しているのかを確認します。
3. データ・リスナーが稼働中で、クライアント・アプリケーションのアクティビティーに関する詳細な状況情報を表示したい場合、 「詳細」をクリックします。

「データ・リスナー詳細」ページの状況アイコンは、データ・リスナーが稼働中か、停止しているかを示します。統計により、処理を待っている要求の数、クライアント・アプリケーション要求に対して作動している各スレッドの現在の状態、指定のスレッド状態についてアクティブなスレッドの数が示されます。

4. データ・リスナーのポート番号を変更する場合は、 「再始動」をクリックします。

データ・リスナーは、エンタープライズ・サーチ・システムが開始されたときに開始されます。データ・リスナーのポート番号を変更した場合を除き、データ・リスナーを再始動する必要はありません。

### 関連タスク

102 ページの『データ・リスナー・アプリケーションのサポートの構成』外部クローラーを作成するデータ・リスナー API を使用することによって、エンタープライズ・サーチを拡張することができます。カスタム・データ・リスナー・アプリケーションは、コレクションへのデータの追加、コレクションからのデータの除去、または Web クローラーに対する URL へのアクセスおよび再アクセスの指示を行えます。

## 文書トラッキング

文書は、処理の様々な段階でシステムから除去することができます。文書が除去された時、および除去の原因となった問題の内容を確認する場合に、このオプションを指定できます。

文書の構文解析を妨げるエラーにパーサーが遭遇した場合、除去された文書に関して、理由コードの付いたメッセージがログに記録されます。(このタイプのエラーによって、旧バージョンの文書が索引から除去されることはありません。)

索引作成の段階で、文書が除去されることがあり、この情報もまたログに記録されます。例えば、URI および URI パターンは、明示的に削除できます。ある文書をクロールし、クロールしたそのクローラーを後で削除してしまうということがあります。ソース文書がもはや存在しないか (その文書にはネガティブ HTTP コードが関連付けられます)、またはその文書に関連付けられた HTTP コードが認識されないこととなります。また、グローバル分析を要求する文書にランク情報が欠如していると、文書が除去されるということもあります。

クロールされた文書がわかっているが、その文書が索引に表示されない場合、エンタープライズ・サーチ管理コンソールを使用して、システム全体にわたる文書のフローをトラックすることができます。いつ、どこで、なぜ、その文書が除去されたのかを示す詳細レポートを表示することができます。例えば、そのレポートは、グローバル分析中に文書が予期せずに除去されたことを示したり、または管理者によって URI が索引から除去されたことを示したりすることがあります。

### 関連タスク

260 ページの『URI についての詳細表示』

URI に関する詳細情報を見ることができます。この URI で表される文書がどのようにクロール、索引付け、検索されるのかについての現在情報および履歴情報を知ることができます。

## 文書トラッキングに対するログ・ファイルの構成

いつ、どこで、およびなぜ、文書がシステムから除去されたかを判別する場合、除去された文書に関する情報をトラッキングするログ・ファイルを構成することができます。

### 始める前に

コレクション・レベル・メッセージの 除去された文書をトラッキングするためのオプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### このタスクについて

ログ・ファイルが消費するディスク・スペースが大きくなりすぎるのを防止するために、システムはログ・ファイルを循環させ、現在日付が変わるたびに新しいログ・ファイルを開始します。1 つのログ・ファイルが許容最大サイズに達し、日付が変わっていない場合、システムは新しいログ・ファイルを作成します。ログ・ファイルの最大数に達した場合、新しいログ・ファイルを作成できるように、最も古いログ・ファイルが廃棄されます。



## 手順

文書トラッキングのログ・ファイルを構成するには、次のようにします。

1. コレクションを編集、「ログ」ページを選択、そして「文書トラッキングの構成」をクリックします。
2. 「文書トラッキング」ページで、文書トラッキングのチェック・ボックスが選択されていることを確認します。
3. システムから除去された文書に関する情報をログに記録するのに使用するログ・ファイルの数を指定します。これらのログ・ファイルは、文書が除去される可能性のあるすべてのセッションで共用されます。
4. 「OK」をクリックします。

## 除去された文書に関するレポートの表示

エンタープライズ・サーチ・システムから除去された文書に関する詳細情報を表示することができます。この情報は、コレクションに対する文書トラッキングを使用可能にした場合にのみ有効です。

### 始める前に

除去された文書に関する要求をサブミットする、または E メール・アドレスにレポートを送信する前に、情報の発信元になるセッションがアクティブであることを確認してください。例えば、構文解析中または索引作成中に除去された文書に関して確認するには、コレクションに対するパーサーまたは索引セッションが開始していることを確認してください。

レポートを受信できるようにするには、その前に、メール・サーバーがエンタープライズ・サーチに対して確実に構成されているようにしてください。この情報は、「システム」ビューの「ログ」ページで E メール・オプションを構成するときに指定します。


### このタスクについて


削除された文書に関する情報の収集は、時間のかかるプロセスです。情報を表示する場合に、その情報が表示されるのを待つというオプションを選択できます。より効率的なオプションは、指定した E メール・アドレスにレポートを送信するようにするオプションです。


文書が除去された場合、文書が除去された日時、エラーの重大度レベル、問題が発生したコンポーネントとセッション、およびエラー・メッセージが、そのレポートによって表示されます。

## 手順

除去された文書に関する詳細を表示するには、次のようにします。

1. 「コレクション」ビューで、モニター対象にしたいコレクションを見つけて、 「モニター」をクリックします。

ヒント: コレクションを編集中であり、すでに「一般」ページが表示されている場合は、 「モニター」をクリックすれば、目的のコレクションをモニターするためのビューに移動できます。

2.  「URI の詳細」をクリックします。
3. 「URI の詳細」ページで、情報を見たい URI を入力します。
4. 見たい情報のタイプに合わせてチェック・ボックスを選択します。

#### パーサーによって除去された文書

その文書が構文解析中に除去されたのかどうか、またその場合の除去された理由を参照する場合は、このチェック・ボックスを選択します。

#### 索引から除去された文書

その文書が索引作成中またはアナライズ中に除去されたのかどうか、またその場合の除去された理由を参照する場合は、このチェック・ボックスを選択します。

5. どのようにレポートを表示するかを指定します。
  - レポートが表示されるのを待つ場合は、「レポート表示」をクリックします。
  - 後でレポートを表示できるように E メール・アドレスにレポートを送信するには、「レポートの送信」をクリックします。

「詳細 URI レポートの送信」ページで、「通知を受け取る E メール・アドレス」フィールドに、レポートを受け取る E メール・アドレスを入力して、「レポートの送信」をクリックします。

#### 関連タスク

260 ページの『URI についての詳細表示』

URI に関する詳細情報を見ることができます。この URI で表される文書がどのようにクロール、索引付け、検索されるのかについての現在情報および履歴情報を知ることができます。

#### 関連資料

105 ページの『エンタープライズ・サーチ索引における URI フォーマット』  
エンタープライズ・サーチ索引内の各文書の Uniform Resource Identifier (URI) は、その文書をコレクションに追加したクローラーのタイプを示します。

---

## エンタープライズ・サーチのログ・ファイルおよびアラート

コレクションおよびシステムに対してログを記録するときのメッセージのタイプを選択し、ログ・ファイルの作成と表示のオプション、アラート受信のオプション、およびメッセージに関する E メール受信のオプションを指定することができます。

通常運用では、エンタープライズ・サーチ・コンポーネントは、メッセージを共通のログ・ファイルに書き込みます。このログ・ファイルは、索引サーバーの `ES_NODE_ROOT/logs` ディレクトリにあります。管理コンソールを使用して、この共通ログ・データを表示することができます。

ネットワーク通信障害などの問題が発生すると、当該コンポーネントは、そのコンポーネントがインストールされているサーバーの `logs` ディレクトリにログ・メッセージを書き込みます。これらのローカル・ログ・ファイルを表示するには、UNIX システムの `tail` ユーティリティなど、当該コンピューターのファイル・ビューアーを使用してください。このタイプのログ・ファイルの表示に管理コンソールを使用することはできません。

ログ・ファイルの構成時には、ログに記録するメッセージのタイプ (エラー・メッセージ、警告メッセージなど) の選択、古くなったログ・ファイルを廃棄して新規のログ・ファイル用のスペースを確保する場合の廃棄頻度の指定、ログ・ファイルの最大サイズの指定、およびメッセージ言語の選択を行うことができます。また、特定の事象が発生した場合や、特定のメッセージ、あるいは特定のタイプのメッセージがログに記録された場合に、必ず E メールを受信するようにするオプションを指定することもできます。

ログ・ファイルをモニターする場合、開きたいログ・ファイルを選択できます。ログ・ファイルの内容をフィルターに掛けて、特定の重大度レベル (エラー・メッセージのみなど) のメッセージ、または特定のエンタープライズ・サーチ・セッションが生成したメッセージだけを表示するように設定することができます。ログ・ファイルを表示する場合、個別のメッセージの詳細を表示することができます。例えば、メッセージを生成した機能名、その他問題を修正する際に役立つ情報を必要に応じて確認することができます。

### 関連概念

「メッセージ・リファレンス」の『エンタープライズ・サーチのメッセージ』

---

## アラート

エンタープライズ・サーチでは、特定のイベントの発生が検出された場合に、メッセージをログ・ファイルに書き込むように構成することができます。

イベントによって起動され、作成されるメッセージ (アラートと呼ばれます) は、特定のリソースでフリー・スペースがなくなりかけているなど、対処が必要と思われる状態の発生を管理者に通知します。エンタープライズ・サーチでアラートを構成

する場合、システムでモニターする条件を指定します。その条件に該当する状態が発生した場合、システムによって自動的にログ・ファイルにメッセージが書き込まれます。

特定の状態について、直接管理者に通知されるようにしたい場合は、モニターによる特定のメッセージがログに記録されたときに E メールを送信するオプションを指定できます。

アラートは、コレクション・レベルのイベント、およびシステム・レベルで発生したイベントに対して構成できます。コレクション・レベルでは、システムは、次のような構成が可能です。

- 各クローラーがクロールする文書数をモニターし、指定された最大文書数に近くなった場合にアラート・メッセージを発行する。
- コレクションの索引に追加される文書数をモニターし、指定された最大文書数に近くなった場合にアラート・メッセージを発行する。
- 検索要求に対する応答時間が指定した限度を超える場合に通知する。

システム・レベルでは、エンタープライズ・サーチ・サーバーのディスク・スペースをモニターし、フリー・スペースの量が少なくなった場合にアラート・メッセージを発行する構成を設定することができます。

## コレクション・レベル・アラートの構成

アラートを構成することによって、特定のコレクション・レベルのイベントが発生すると必ずログ・ファイルにメッセージが書き込まれるようにすることができます。また、これらのイベントに関するメッセージがログに記録されると、E メールを受信できます。

### 始める前に

コレクションのアラートを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

### 手順

コレクション・レベルのアラートを構成するには、次のようにします。

1. コレクションを編集します。「ログ」ページを選択し、「アラートの構成」をクリックします。
2. システムが、それぞれのクローラーがクロールしている文書の数モニターするようにしたければ、次のステップを行います。
  - a. 「クローラーがクロールした文書数が最大許容値の一定比率に達した場合」チェック・ボックスを選択します。
  - b. 「比率 (%)」フィールドに、メッセージをログに記録する時点のパーセンテージを指定します。この数値は、クローラーがクロールできる最大文書数に対するパーセンテージで指定します (クローラー・プロパティを構成するときに、「クロールする最大文書数」を指定します)。デフォルト値は 90 % です。

それぞれのクローラーに異なる制限を構成できるので、クローラーごとに別個のメッセージがログに記録されます。例えば、アラートしきい値の比率はデフォルト値を使用し、DB2 クローラーのクロール可能文書数を 200 万 (2 000 000) 文書に設定し、Lotus Notes クローラーのクロール可能文書数を 100 万 (1 000 000) 文書に設定した場合、DB2 クローラーが 180 万 (1 800 000) 文書をクロールした時点でメッセージがログに記録され、Lotus Notes クローラーが 90 万 (900 000) 文書をクロールした時点で別のメッセージがログに記録されます。

3. システムが、索引に追加されている文書の数モニターするようにしたければ、次のステップを行います。
  - a. 「**コレクション内の文書数が見積もり数の一定比率に達した場合**」チェック・ボックスを選択します。
  - b. 「**比率 (%)**」フィールドに、メッセージをログに記録する時点のパーセンテージを指定します。この数値は、コレクションが保持すると想定される見積もり文書数のパーセンテージで指定します。デフォルト値は 85 % です。

「**限度**」フィールドに、コレクションの現在の見積もり数が表示されます。この値を変更するには、コレクションの「**一般**」ページを開き、一般オプションを構成するオプションを選択して、「**見積もり文書数**」フィールドに新しい値を指定します。

**重要:** この限度、およびあるコレクションのために構成する文書の見積もり数は、そのコレクションの拡大をモニターするためにのみ使用されます。それらの数値は、索引が拡大できる大きさの絶対的な制限を押しつけるものではありません。

4. システムが、検索要求の応答に必要な時間が限界を超えた場合に通知するようにしたければ、次のステップを行います。
  - a. 「**検索応答時間が限度を超えた場合**」チェック・ボックスを選択します。
  - b. 「**限度**」フィールドに、最長の検索応答時間として許容できる秒数を入力します。

この数を超えると、システムは、このイベントについてのログ・メッセージを書き込みます。例えば、デフォルト値のままにした場合、システムは、検索サーバーの検索要求に応答する時間が平均 5 秒以上かかると、ログ・メッセージを作成します。

標準的な応答時間は、0.5 秒以内です。平均が 1 秒を超える場合、オペレーティング・システムにパフォーマンスを改善する調整が必要であるか、あるいは、検索サーバーの構成設定に問題があることを示しています。例えば、検索キャッシュに割り振るスペースの量を増やすことも考えられます。

5. 「**OK**」をクリックします。

これらのイベントについてのメッセージをシステムがログに記録したときに、E メールを受信するようにしたければ、「ログ」ページを開き、「**メッセージに対する E メール・オプションの構成**」をクリックして、E メール・アドレスを指定します。使用可能にしたアラートのメッセージ ID は、E メールが送信されるメッセージ ID のリストに自動的に追加されます。

E メールを受信できるようにするには、その前に、ユーザーのメール・サーバーの情報が構成されていることを確認する必要があります。これを行うには、エンタープライズ・サーチ管理者は、ツールバーの「システム」を選択し、「ログ」ページを開いて、「メッセージに対する E メール・オプションの構成」をクリックする必要があります。

#### 関連タスク

287 ページの『ログ・メッセージについての E メール受信』

特定のメッセージまたはメッセージ・タイプがログに記録されたら E メールを受信するというオプションを指定できます。

## システム・レベル・アラートの構成


アラートを構成することによって、特定のシステム・レベルのイベントが発生すると必ずログ・ファイルにメッセージが書き込まれるようにすることができます。また、これらのイベントに関するメッセージがログに記録されると、E メールを受信できます。

#### 始める前に

システム・レベルのアラートを構成するには、エンタープライズ・サーチ管理者でなければなりません。

#### 手順

システム・レベルのアラートを構成するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2.  「編集」をクリックして、システム編集ビューに移動します。
3. 「ログ」ページで、「アラートの構成」をクリックします。
4. システムが、エンタープライズ・サーチ・サーバーで使用可能なフリー・スペースの量をモニターするようにしたい場合は、「ファイル・システムの使用可能スペース量が全体スペースの一定比率に達した場合」チェック・ボックスを選択します。
5. 「比率 (%)」フィールドに、システムが、サーバーのフリー・スペースの量がなくなったことを通知する時点のパーセンテージを指定します。この数値は、ファイル・システムの合計スペースに対するパーセンテージで指定します。デフォルト値は 80 % です。

使用しているエンタープライズ・サーチ・システムが複数のサーバー上にセットアップされている場合、システムは、サーバーごとに別個のログ・メッセージを作成します。例えば、あるメッセージはクローラー・サーバー上のスペースがなくなったことを通知し、また別のメッセージは、索引および検索サーバーのスペース制約について通知します。

6. 「OK」をクリックします。

このイベントについてのメッセージをシステムがログに記録したときに、E メールを受信するようにしたければ、「ログ」ページを開き、「メッセージに対する E メール・オプションの構成」をクリックして、E メール・アドレスと、メール・サーバーについての情報を指定します。



## 関連タスク

287 ページの『ログ・メッセージについての E メール受信』

特定のメッセージまたはメッセージ・タイプがログに記録されたら E メールを受信するというオプションを指定できます。

---

## ログ・ファイルの構成

ログに記録したいメッセージのタイプを指定し、ログ・ファイル作成のオプションを指定できます。

### 始める前に

コレクション・レベルのログ・ファイルを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。システム・レベルのログ・ファイルを構成するには、エンタープライズ・サーチ管理者でなければなりません。



### このタスクについて

ログ・ファイルが消費するディスク・スペースが大きくなりすぎるのを防止するために、システムはログ・ファイルを循環させ、現在日付が変わるたびに新しいログ・ファイルを開始します。1 つのログ・ファイルが許容最大サイズに達し、日付が変わっていない場合、システムは新しいログ・ファイルを作成します。ログ・ファイルの最大数に達した場合、新しいログ・ファイルを作成できるように、最も古いログ・ファイルが廃棄されます。

ログに記録されたメッセージについての E メールを受信するには、まず E メールが配信される方法に関する情報を指定します。次に、どのメッセージについて E メールを受信したいのかを指定します。

### 手順

エンタープライズ・サーチ・ログ・ファイルを構成するには、次のようにします。

1. システム・レベルのログ・ファイルの作成と循環についてのオプションを構成したい場合、次のようにします。
  - a. 「システム」をクリックして「システム」ビューを開きます。
  - b.  「編集」をクリックして、システム編集ビューに移動します。
  - c. 「ログ」ページで、「ログ・ファイル・オプションの構成」をクリックします。「システム・レベルのログ・ファイル・オプション」ページが表示されます。
2. コレクション・レベルのログ・ファイルの作成と循環についてのオプションを構成したい場合、次のようにします。
  - a. 「コレクション」ビューで、オプションを指定したいコレクションを見つけ、 「編集」をクリックします。
  - b. 「ログ」ページで、「ログ・ファイル・オプションの構成」をクリックします。「コレクション・レベルのログ・ファイル・オプション」ページが表示されます。

3. 「**ログ対象の情報タイプ**」フィールドで、ログに記録したいメッセージのタイプを選択します。

#### エラー・メッセージのみ

エラー・メッセージは、望ましくない状態または予期しない動作が発生したこと、処理が継続できないことを示します。何らかの処置を実行して問題を解決する必要があります。

#### エラーおよび警告メッセージ

警告メッセージは、起こりうる競合または不整合を示しますが、そのために処理が停止するわけではありません。このオプションがデフォルトです。

#### すべてのメッセージ

情報メッセージは、システムまたは現行タスクの一般情報を示し、修正処置は何も必要ありません。

4. 「**各ログ・ファイルの最大サイズ**」フィールドに、各ログ・ファイルの最大サイズを M バイトで入力します。デフォルト値は 5 MB です。

ログ・ファイルがこのサイズに達すると、ログ・ファイルの許容最大数に達するまで新しいログ・ファイルが作成されます。ログ・ファイルを比較的小さいサイズにしておく、表示して調べるときの効率がよくなります。

5. 「**ログ・ファイルの最大数**」フィールドに、作成したいログ・ファイルの最大数を入力します。デフォルト値は 10 です。

古いログ・メッセージも確実に参照できるようにしたい場合は、この数値を大きくしてください。関心があるのは最近のメッセージであり、長期間のアクティビティ履歴を保持する必要がない場合は、この値を減らしてください。

6. 「**デフォルト・ロケール**」フィールドで、ログ・メッセージに使用したい言語を選択します。デフォルト値は英語です。
7. 「**OK**」をクリックします。

---

## SMTP サーバー情報の構成

エンタープライズ・サーチのアクティビティに関する E メールを受信できるようにするには、その前に、SMTP (Simple Mail Transfer Protocol) サーバーの情報を構成しておく必要があります。

### 始める前に

SMTP サーバーの情報を構成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。

### このタスクについて

いくつかのエンタープライズ・サーチ管理機能を使用して、E メールを受信できます。これらの機能のいずれかから E メールを受信するには、その前に、SMTP サーバーの情報を指定する必要があります。


- コレクション・レベルのアラートまたはシステム・レベルのアラートを構成すると、これらのメッセージがログに記録されるかどうかに関係なく、E メールを受

信できます。また、モニター対象のイベントに起因するメッセージだけでなく、他のメッセージがログに記録されたときに E メールを受信することもできます。

- 索引内の URI またはエンタープライズ・サーチ・システムからドロップされた文書に関して詳しい情報を表示したい場合、レポートを E メールで受け取ることができます。
- Web クローラーをモニターし、Web クローラーの履歴レポートを作成するよう指定した場合、レポートの作成後に E メールで通知されます。

### 手順

SMTP サーバーの情報を構成するには、次のようにします。

1. 「システム」をクリックして「システム」ビューを開きます。
2.  「編集」をクリックして、システム編集ビューに移動します。
3. 「ログ」ページで、「メッセージに対する E メール・オプションの構成」をクリックします。
4. 「システム・メッセージに対する E メール・オプション」ページで、「E メール配信に使用する SMTP メール・サーバー」フィールドに、使用したい SMTP サーバーの完全修飾ホスト名または IP アドレスを入力します。

システムは、このサーバーを使用して、指定したアドレスに E メールを送信します。

5. 「Eメールの確認頻度」フィールドに、適格メッセージのチェックとそれらに関する Eメールの送信をシステムが行う頻度を指定します。

システムは、1 つの特定の E メール・アドレス向けのメッセージはすべて結合して 1 つのメッセージにし、指定した頻度でそのメッセージを送信します。

6. 「OK」をクリックします。

---

## ログ・メッセージについての Eメールの受信

特定のメッセージまたはメッセージ・タイプがログに記録されたら Eメールを受信するというオプションを指定できます。

### 始める前に

システム・レベル・メッセージの Eメール・オプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーである必要があります。コレクション・レベル・メッセージの Eメール・オプションを構成するには、エンタープライズ・サーチ管理者役割のメンバーであるか、そのコレクションのコレクション管理者である必要があります。

Eメールを受信できるようにするには、最初に、Eメールが送達できるように SMTP (Simple Mail Transfer Protocol) サーバーの情報を構成しておく必要があります。


### このタスクについて

アラートの構成では、特定のイベントが発生したらメッセージをログに記録するというオプションを選択できます。それらのオプションを使用可能にした場合、それ

らのメッセージがログに記録されたら自動的に E メールを受信するというオプションを構成できます。また、イベントに起因するメッセージだけでなく、他のメッセージがログに記録されたときに E メールを受信するというオプションを指定することもできます。

## 手順

メッセージについての E メール・オプションを構成するには、次のようにします。

1. システム・メッセージに関して E メールを受け取りたい場合は、次のようにします。
  - a. 「システム」をクリックして「システム」ビューを開きます。
  - b.  「編集」をクリックして、システム編集ビューに移動します。
  - c. 「ログ」ページで、「メッセージに対する E メール・オプションの構成」をクリックします。
  - d. 「システム・メッセージの E メール・オプション」ページで、「システム・レベル・メッセージに関する E メールを送信する」チェック・ボックスを選択します。
  - e. 「E メール受信用 E メール・アドレス」フィールドに、1 つ以上の E メール・アドレスを入力します。通常、エンタープライズ・サーチ管理者は、システム・メッセージに関する情報を受信する必要があります。


各アドレスはコンマで区切ってください。例えば、次のとおりです。

steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com

- f. ログに記録されるすべてのエラー・メッセージについて E メールを受信したい場合は、「全エラー・メッセージに関する E メールを送信する」チェック・ボックスを選択します。
- g. 特定のシステム・レベル・メッセージがログに記録されたときにのみ E メールを受信したい場合は、「特定のメッセージに関する E メールを送信する」領域でそれらのメッセージのメッセージ ID を入力します。1 行につき 1 つのメッセージ ID を入力してください。例えば、次のとおりです。

FFQC4819E  
FFQ00005E

デフォルトでは、いくつかのメッセージ ID がリストされます (これらのメッセージの説明については、「ヘルプ」をクリックしてください)。

- h. 「OK」をクリックします。
2. コレクションのメッセージに関して E メールを受け取りたい場合は、次のようにします。
  - a. 「コレクション」をクリックして「コレクション」ビューを開きます。
  - b. コレクションのリストで、構成したいコレクションを見つけ、 「編集」をクリックします。
  - c. 「ログ」ページで、「メッセージに対する E メール・オプションの構成」をクリックします。
  - d. 「コレクション・メッセージの E メール・オプション」ページで、「コレクション・レベル・メッセージに関する E メールを送信する」チェック・ボックスを選択します。

- e. 「E メール受信用 E メール・アドレス」フィールドに、1 つ以上の E メール・アドレスを入力します。通常、コレクション管理者は、コレクション・レベルのメッセージに関する情報を受信する必要があります。

各アドレスはコンマで区切ってください。例えば、次のとおりです。

steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com

- f. ログに記録されるすべてのエラー・メッセージについて E メールを受信したい場合は、「**全エラー・メッセージに関する E メールを送信する**」チェック・ボックスを選択します。
- g. 特定のコレクション・レベル・メッセージがログに記録されたときのみ E メールを受信したい場合は、「**特定のメッセージに関する E メールを送信する**」領域でそれらのメッセージのメッセージ ID を入力します。1 行につき 1 つのメッセージ ID を入力してください。例えば、次のとおりです。

FFQC4819E  
FFQ00005E

デフォルトでは、いくつかのメッセージ ID がリストされます (これらのメッセージの説明については、「ヘルプ」をクリックしてください)。

- h. 「OK」をクリックします。

#### 関連概念

「メッセージ・リファレンス」の『エンタープライズ・サーチのメッセージ』

#### 関連タスク

282 ページの『コレクション・レベル・アラートの構成』

アラートを構成することによって、特定のコレクション・レベルのイベントが発生すると必ずログ・ファイルにメッセージが書き込まれるようにすることができます。また、これらの イベントに関するメッセージがログに記録されると、E メールを受信できます。

284 ページの『システム・レベル・アラートの構成』

アラートを構成することによって、特定のシステム・レベルのイベントが発生すると必ずログ・ファイルにメッセージが書き込まれるようにすることができます。また、これらの イベントに関するメッセージがログに記録されると、E メールを受信できます。

---


## ログ・ファイルの表示


システムおよびコレクション・コンポーネントが共通ログ・ファイルに書き込むログ・メッセージを表示できます。また、フィルターを指定して、特定の重大度レベルのメッセージ、および特定のエンタープライズ・サーチ・セッションのメッセージのみを表示することもできます。

#### 始める前に

すべてのエンタープライズ・サーチ管理ユーザーは、管理を認可されているコレクションのログ・ファイルを表示できます。システム・レベルのログ・ファイルを表示するには、エンタープライズ・サーチ管理者役割のメンバーであるか、または「システム」ツールバーへのアクセス許可を持っている必要があります。

## 手順

1. 単一のコレクションのログ・ファイルを表示するには、次のようにします。
  - a. 「**コレクション**」をクリックして「コレクション」ビューを開きます。
  - b. コレクションのリストで、表示したいコレクションを指して、 「**モニター**」をクリックして、「ログ」ページを開きます。

**ヒント:** コレクションを編集集中であり、すでに「ログ」ページが表示されている場合は、 「**モニター**」をクリックすれば、コレクションをモニターするためのビューに移動できます。

2. システム・レベルのログ・ファイルを表示するには、次のようにします。
  - a. 「**システム**」をクリックして「システム」ビューを開きます。
  - b. 「ログ」ページを選択します。
3. 「**ログ・ファイル**」フィールドで、表示したいログ・ファイルを選択します。


各ログ・ファイルの名前には、ログ・ファイルのタイプ (システムまたはコレクション ID など)、作成日、その日の作成順を示す数値が含まれています。例えば、次のとおりです。

```
log_file_type_2005-05-26_1.log  
log_file_type_2005-05-26_2.log  
log_file_type_2005-05-25_1.log  
log_file_type_2005-05-25_2.log  
log_file_type_2005-05-25_3.log
```

4. 特定の重大度レベルのメッセージのみを表示するには、「**重大度**」フィールドで適切なチェック・ボックスを選択します。
5. 特定のセッションからのメッセージのみを表示するには、「**セッション**」フィールドで適切なチェック・ボックスを選択します。
6. 「**ログの表示**」をクリックします。

「ログ・ファイルのコンテンツ」ページの各メッセージごとに、メッセージが発行された日時、メッセージ重大度レベル、メッセージを発行したセッションの名前、およびメッセージ ID とエラー・テキストが表示されます。

ボタンをクリックすることで、ログ・ファイルの先頭ページ、最終ページ、直前のページ、または次のページに移動することができます。ページ番号を指定することで、そのページに直接移動することもできます。

7. メッセージについて詳しい情報を表示したい場合は、 「**詳細**」をクリックします。

「ログ・メッセージの詳細」ページに、メッセージを出したエンタープライズ・サーチ・サーバーのホスト名、エラーを引き起こしたファイルの名前、エラーが発生した関数名および行番号、プロセス ID、スレッド ID が表示されます。



---

## エンタープライズ・サーチ・システムのバックアップとリストア

バックアップ・スクリプトおよびリストア・スクリプトを使用して、エンタープライズ・サーチ・システムをバックアップおよびリストアすることができます。

回復不能エラーによるシステム障害が発生した場合は、WebSphere Information Integrator OmniFind Edition を再インストールし、リストア・スクリプトを実行する必要があります。また、これらのスクリプトは、1 つまたは複数の新規サーバーに対して基本的な必須システム・ファイルをリストアする場合に使用することもできます。

これらのスクリプトは、次のファイルをバックアップおよびリストアします。

- ES\_NODE\_ROOT/master\_config ディレクトリーの構成ファイル。
- クローラーのデータベース・ファイル。
- 索引ファイル。エンタープライズ・サーチ索引ファイルが ES\_NODE\_ROOT/data ディレクトリーにない場合は、エンタープライズ・サーチ・スクリプトを使用して索引ファイルをバックアップおよびリストアすることはできません。

バックアップ・スクリプトは、スクリプトの実行時に指定したディレクトリーの下に以下のサブディレクトリーを作成します。(エンタープライズ・サーチ管理者 ID は、指定したディレクトリーに対して書き込み許可を持っている必要があります。)

### **master\_config**

ES\_NODE\_ROOT/master\_config ディレクトリーの構成ファイルが保管されます。

### **database**

クローラー・サーバーのデータベース・ファイルが保管されます。

**data** 索引サーバーの索引ファイルが保管されます。

エンタープライズ・サーチ・システム・ファイルを別のディレクトリーにバックアップする場合、十分なディスク・スペースを確保する必要があります。バックアップ・スクリプトおよびリストア・スクリプトでは、ファイルのチェックは行いません。また、バックアップを開始すると、ほとんどのシステム・セッションは一時的に使用不可になりますが、検索プロセスは継続して実行されます。索引を再編成してからバックアップを開始してください。そうすることにより、最新の索引を使用することができます。

マルチサーバー・インストールの場合、システムのバックアップおよびリストアは、エンタープライズ・サーチ索引サーバーから行ってください。索引サーバーは、データベース・カタログを作成するため、クローラー・サーバーの製品データベース表にアクセスし、バックアップすることができます。

### **関連資料**

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## エンタープライズ・サーチ・システムのバックアップ

エンタープライズ・サーチ・システムのバックアップは、UNIX の場合は `esbackup.sh` スクリプトを、Microsoft Windows の場合は `esbackup.bat` スクリプトを使用して行います。

### 制約事項

バックアップ・スクリプトおよびリストア・スクリプトの実行中は、すべてのシステム・セッションが停止されます。誤ったシステム情報または矛盾するシステム情報が表示されるのを回避するため、スクリプトの実行中は、エンタープライズ・サーチ管理コンソールを使用しないでください。

エンタープライズ・サーチ管理者 ID には、バックアップ・スクリプトを実行するときに指定するディレクトリーへの書き込み権限が必要です。

### 手順

エンタープライズ・サーチ・システムをバックアップするには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. 以下のコマンドを入力します。`backup_directory` は、ファイルをバックアップするディレクトリーです。

UNIX: `esbackup.sh -c -d -i backup_directory`

Windows: `esbackup.bat -c -d -i backup_directory`

指定できるオプションは、次のとおりです。

- c 構成ファイルをバックアップします。これはデフォルト・オプションです。
- d クローラー・データ・ストア内のクローラされた文書をバックアップします (これらの文書はまだ構文解析または索引付けされていません)。
- i 索引ファイルをバックアップします。索引ファイルが `ES_NODE_ROOT/data` ディレクトリーにない場合は、`esbackup` スクリプトを使用して索引ファイルをバックアップすることはできません。

### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

## エンタープライズ・サーチ・システムのリストア

WebSphere II OmniFind Edition を再インストールした後、`esrestore.sh` スクリプト (UNIX) または `esrestore.bat` スクリプト (Microsoft Windows) を使用して、システム構成ファイルをリストアできます。

### 制約事項

バックアップ・スクリプトおよびリストア・スクリプトの実行中は、すべてのシステム・セッションが停止されます。誤ったシステム情報または矛盾するシステム情報が表示されるのを回避するため、スクリプトの実行中は、エンタープライズ・サーチ管理コンソールを使用しないでください。

### 手順

エンタープライズ・サーチ・システムをリストアするには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. 共通通信層 (CCL) が開始されていない場合、これを開始します。

UNIX: `startccl.sh -bg`

Windows: `startccl.bat`

Windows システムで CCL をバックグラウンドで開始するには、「スタート」 → 「プログラム」 → 「管理ツール」 → 「サービス」の順にクリックし、WebSphere Information Integrator OmniFind Edition サービスを再始動します。

3. コントローラーがまだ停止されていない場合は、次のコマンドによって停止します。

`esadmin stop`

4. 以下のコマンドを入力します。ここで、`backup_directory` は、ファイルをバックアップしたディレクトリーです。

UNIX: `esrestore.sh -c -d -i backup_directory`

Windows: `esrestore.bat -c -d -i backup_directory`

指定できるオプションは、次のとおりです。

- c 構成ファイルをリストアします。これはデフォルト・オプションです。
- d クローラー・データ・ストア内のクローラ済み文書をリストアします。(これらの文書は、まだ構文解析も索引作成もされていません。)
- i 索引ファイルをリストアします。索引ファイルが `ES_NODE_ROOT/data` ディレクトリーにない場合、`esrestore` スクリプトを使用して索引ファイルをリストアすることはできません。

### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

---

## エンタープライズ・サーチ・システム・ファイルを新規のサーバーにリストアする

システム・ファイルをあるエンタープライズ・サーチ・サーバーからバックアップして、そのシステム・ファイルを現行サーバーではなく、1 基または複数の新規のサーバーにリストアすることもできます。

### 始める前に

リストア・スクリプトを実行する前に、WebSphere II OmniFind Edition を対象となる新規のサーバーにインストールしておく必要があります。

### 制約事項

バックアップ・スクリプトおよびリストア・スクリプトの実行中は、すべてのシステム・セッションが停止されます。誤ったシステム情報または矛盾するシステム情報が表示されるのを回避するため、スクリプトの実行中は、エンタープライズ・サーチ管理コンソールを使用しないでください。

エンタープライズ・サーチ管理者 ID には、バックアップ・スクリプトを実行するときに指定するディレクトリーへの書き込み権限が必要です。

### このタスクについて

ES\_NODE\_ROOT/master\_config/nodes.ini ファイルに格納されているサーバー情報は、バックアップ・ファイルには含まれていません。

### 手順

1 基または複数の新規のサーバーにエンタープライズ・サーチのシステム・ファイルをリストアするには、次のようにします。

1. 索引サーバーで、エンタープライズ・サーチ管理者としてログインします。このユーザー ID は、WebSphere II OmniFind Edition のインストール時に指定されたものです。
2. バックアップ・スクリプトを現行 (リストア元) 索引サーバーで実行します。ここで、*backup\_directory* は、ファイルをバックアップするディレクトリーです。

UNIX: `esbackup.sh backup_directory`

Windows: `esbackup.bat backup_directory`

3. FTP プログラムを使用して、*backup\_directory* 内のすべてのファイルを新規の索引サーバーに送信します。
4. 新規の索引サーバーでリストア・スクリプトを実行します。

UNIX: `esrestore.sh backup_directory`

Windows: `esrestore.bat backup_directory`

### 関連資料

297 ページの『エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID』

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。





## エンタープライズ・サーチ・コマンド、戻りコード、およびセッション ID

コマンドを使用して、問題の診断、システムの別の部分の状況の判別、セッションの開始および停止、あるいはシステムの始動および停止を行うことができます。

複数のサーバーがインストールされている環境では、システム内のどのサーバーからでもコマンドを実行することができます。しかしながら、索引サーバーからコマンドを実行することをお勧めします。索引サーバー (コントローラー・サーバー) は、システム内にある他のすべてのシステムの情報にアクセスすることができます。

大部分のコマンドは、以下のようなフォーマットです。

```
esadmin command_name arguments  
esadmin session_ID action -option
```

すべてのコマンドについて詳しく知りたい場合は、`esadmin help` と入力してください。特定のコマンドについて詳しく知りたい場合は、`esadmin action help` と入力してください。

### エンタープライズ・サーチ `esadmin` コマンド

次のコマンドを 1 行に入力します。

表 5. エンタープライズ・サーチ `esadmin` コマンド

コマンド	説明
<code>esadmin <i>crawler_session_id</i> start</code>	クローラー・セッションを開始します。このコマンドでは、クローラー・アクティビティは開始されません。  コマンドの例: <code>esadmin col1.WEB1.esadmin start</code>  メッセージおよび戻りコードの例: FFQC5310I WEBCrawler1 (sid: col1.WEB1.esadmin) is not running. FFQC5314I Result: 0
<code>esadmin <i>crawler_session_id</i> startCrawl</code>	クローラーを開始します。  コマンドの例: <code>esadmin col3.DB21.esadmin startCrawl</code>  メッセージおよび戻りコードの例: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is already running. PID: 23650 FFQC5314I Result: 0

| 表 5. エンタープライズ・サーチ *esadmin* コマンド (続き)

コマンド	説明
<p>  <i>esadmin crawler_session_id</i> pause</p>	<p data-bbox="714 260 1427 291">クローラを一時停止します。</p> <p data-bbox="714 317 1427 348"><b>コマンドの例:</b></p> <p data-bbox="714 359 1427 390"><i>esadmin col3.DB21.esadmin</i> pause</p> <p data-bbox="714 415 1427 447"><b>メッセージおよび戻りコードの例:</b></p> <p data-bbox="714 457 1427 489">FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)</p> <p data-bbox="714 489 1427 520">is already running. PID: 23650</p> <p data-bbox="714 520 1427 552">FFQC5314I Result: 0</p>
<p>  <i>esadmin crawler_session_id</i> resume</p>	<p data-bbox="714 562 1427 594">クローラを再開します。</p> <p data-bbox="714 619 1427 651"><b>コマンドの例:</b></p> <p data-bbox="714 661 1427 693"><i>esadmin col3.DB21.esadmin</i> resume</p> <p data-bbox="714 718 1427 749"><b>メッセージおよび戻りコードの例:</b></p> <p data-bbox="714 760 1427 791">FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)</p> <p data-bbox="714 791 1427 823">is already running. PID: 23650</p> <p data-bbox="714 823 1427 854">FFQC5314I Result: 0</p>
<p>  <i>esadmin crawler_session_id</i> stopCrawl</p>	<p data-bbox="714 854 1427 886">クローラを停止します。</p> <p data-bbox="714 911 1427 942"><b>コマンドの例:</b></p> <p data-bbox="714 953 1427 984"><i>esadmin col3.DB21.esadmin</i> stopCrawl</p> <p data-bbox="714 1010 1427 1041"><b>メッセージおよび戻りコードの例:</b></p> <p data-bbox="714 1052 1427 1083">FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)</p> <p data-bbox="714 1083 1427 1115">is already running. PID: 23650</p> <p data-bbox="714 1115 1427 1146">FFQC5314I Result: 0</p>
<p>  <i>esadmin crawler_session_id</i> stop</p>	<p data-bbox="714 1146 1427 1178">クローラ・セッションを停止します。</p> <p data-bbox="714 1203 1427 1234"><b>コマンドの例:</b></p> <p data-bbox="714 1245 1427 1276"><i>esadmin col3.DB21.esadmin</i> stop</p> <p data-bbox="714 1302 1427 1333"><b>メッセージおよび戻りコードの例:</b></p> <p data-bbox="714 1344 1427 1375">FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin)</p> <p data-bbox="714 1375 1427 1407">is already running. PID: 23650</p> <p data-bbox="714 1407 1427 1438">FFQC5314I Result: 0</p>

表 5. エンタープライズ・サーチ *esadmin* コマンド (続き)

コマンド	説明
<code>esadmin crawler_session_id getCrawlerStatus</code>	<p>クローラーの状況を取得します。戻される情報は、そのクローラーが Web クローラーであるか、あるいはその他すべてのデータ・ソース用のクローラーであるかによって決まります。</p>
	<p><b>Web クローラーの場合の例:</b></p>
	<pre>esadmin col1.WEB1.esadmin getCrawlerStatus</pre>
	<p><b>Web クローラーの場合の戻りコードおよびメッセージの例:</b></p>
	<pre>FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) is already running. PID: 23650</pre>
	<p><b>Web 以外のクローラーの場合の例:</b></p>
	<pre>esadmin col3.DB21.esadmin getCrawlerStatus</pre>
	<p><b>Web 以外のクローラーの場合の戻りコードおよびメッセージの例:</b></p>
	<pre>FFQC5303I db2crawler (sid: db2col.DB2_96945) is already running. PID: 5936</pre>
	<p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p>
<code>esadmin dscrawler_session_id getCrawlSpaceStatus</code>	<p>Web 以外のクローラーのクロール・スペース状況概要を取得します。</p>
<code>esadmin web_crawler_session_id getCrawlStatus -selections value</code>	<p><b>コマンドの例:</b></p>
	<pre>esadmin col3.DB21.esadmin getCrawlSpaceStatus</pre>
	<p><b>メッセージおよび戻りコードの例:</b></p>
	<pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is already running. PID: 23650</pre>
	<p>Web クローラーのクロール・スペース状況概要を取得します。</p>
	<p><b>コマンドの例:</b></p>
	<pre>esadmin col1.WEB1.esadmin getCrawlStatus</pre>
	<p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p>

表 5. エンタープライズ・サーチ *esadmin* コマンド (続き)

コマンド	説明
<pre>esadmin dscrawler_session_id getCrawlSpaceStatusDetail -ts target_server_id  esadmin webcrawler_session_id getCrawlDetailsPerSite -url string selections num -threshold num</pre>	<p>Web 以外のクローラーのクロール・スペース状況詳細を取得します。ターゲット・サーバーのオプションを指定しないと、すべてのターゲット・サーバーのデータが戻されます。例えば、DB2 クローラーが FOUNTAIN および SAMPLE データベースをクロールする場合、ターゲット・サーバー・オプションを指定しないと、FOUNTAIN および SAMPLE データベースのすべての表の状況が戻されます。</p>
	<p><b>コマンドの例:</b></p>
	<pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre>
	<p><b>メッセージおよび戻りコードの例:</b></p>
	<pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) is already running. PID: 23650</pre>
	<p>Web クローラーのクロール・スペース状況詳細を取得します。</p>
	<p><b>コマンドの例:</b></p>
	<pre>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</pre>
	<p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p>
<pre>esadmin startParser -cid collection_ID</pre>	<p>パーサーを開始します。</p>
	<p><b>コマンドの例:</b></p>
	<pre>esadmin startParser -cid col1</pre>
	<p><b>メッセージおよび戻りコードの例:</b></p>
	<pre>FFQC5303I Controller (server1) (sid: controller) is already running. PID: 25917 FFQC5314I Result: 0</pre>
<pre>esadmin stopParser collection_id</pre>	<p>パーサーを停止します。</p>
	<p><b>コマンドの例:</b></p>
	<pre>esadmin stopParser -cid col1</pre>
	<p><b>メッセージおよび戻りコードの例:</b></p>
	<pre>FFQC5303I Controller (node1) (sid: controller) is already running. PID: 15292 FFQC5312E Error executing action: [stopParser] FFQC4823E Session Parser Driver - Collection col1 (node1) [col1.parserdriver] is not running.</pre>

表 5. エンタープライズ・サーチ *esadmin* コマンド (続き)

コマンド	説明
<pre>esadmin monitor getCollectionParserMonitorStatus -cid <i>collection_ID</i></pre>	<p>パーサーの状況を取得します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b></p> <pre>FFQC5303I Monitor (node1) (sid: monitor) is already running. PID: 12543</pre> <p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p>
<pre>esadmin startMain -cid <i>collection_id</i></pre>	<p>索引の再編成を開始します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin startMain -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b></p> <pre>FFQC5303I Controller (node1) (sid: controller) is already running. PID: 25917 FFQC5314I Result: 1117671147056</pre>
<pre>esadmin startDelta -cid <i>collection_id</i></pre>	<p>索引のリフレッシュを開始します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin startDelta -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b></p> <pre>FFQC5303I Controller (node1) (sid: controller) is already running. PID: 4548 FFQC5314I Result: 1117670603408</pre>
<pre>esadmin monitor getCollectionIndexMonitorStatus -cid <i>collection_id</i> -buildType [main   delta] -numrecords <i>lastNrecords</i></pre>	<p>索引のリフレッシュまたは再編成の状況を取得します。(main 索引は、再編成された索引を参照します。delta 索引は、リフレッシュされた索引を参照します。オプション <i>numrecords</i> を指定すると、最後の N 個の索引作成状況レポートが表示されます。<i>numrecords</i> を省略した場合は、最後の 20 個の索引作成状況が戻されます。</p> <p><b>コマンドの例:</b></p> <pre>esadmin monitor getCollectionIndexMonitorStatus -cid coll -buildType main -numrecords 4</pre> <p><b>メッセージおよび戻りコードの例:</b></p> <pre>FFQC5303I Monitor (node1) (sid: monitor) is already running. PID: 12649</pre> <p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p>

表 5. エンタープライズ・サーチ *esadmin* コマンド (続き)

コマンド	説明
<code>esadmin startSearch -cid collection_id</code>	<p>検索サーバー・プロセスを開始します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin startSearch -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b> <pre>FFQC5303I Controller (node1) (sid: controller) is already running. PID: 25917 FFQC5314I Result: 0</pre> </p>
<code>esadmin stopSearch -cid collection_id</code>	<p>検索サーバー・プロセスを停止します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin stopSearch -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b> <pre>FFQC5303I Controller (node1) (sid: controller) is already running. PID: 15292 FFQC5314I Result: 0</pre> </p>
<code>esadmin monitor</code> <code>getCollectionSearchMonitorStatus</code> <code>-cid collection_id</code>  <code>esadmin searchmanager_session_id</code> <code>getStatus -cid collection_id</code>	<p>検索サーバーの状況を取得します。</p> <p><b>コマンドの例:</b></p> <pre>esadmin monitor getCollectionSearchMonitorStatus -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b> <pre>FFQC5303I Monitor (node1) (sid: monitor) is already running. PID: 12649</pre> <p>所定の検索サーバー上のコレクションに関する詳細な検索索引状況情報を戻します。検索サーバーごとに、検索マネージャー・セッションが 1 つあります。特定の検索サーバー上の検索索引をモニターおよび操作するのは、各検索マネージャー・セッションの役目です。</p> <p><b>コマンドの例:</b></p> <pre>esadmin searchmanager.node1 getStatus -cid coll</pre> <p><b>メッセージおよび戻りコードの例:</b> <pre>FFQC5303I Search Manager (node1) (sid: searchmanager.node1) is already running. PID: 15711 FFQC5314I Result: PID=18390 CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1</pre> <p>戻される状況メッセージについて詳しくは、303 ページの『状況コマンドの詳細情報』を参照してください。</p> </p></p>



## 状況コマンドの詳細情報

コマンドの中には、広範囲にわたる情報を戻せるものがあります。このセクションでは、クローラー状況およびクロール・スペース状況に関して戻ることができる情報について説明します。297 ページの『エンタープライズ・サーチ esadmin コマンド』の表で、各 esadmin コマンドから戻される情報を示しています。このセクションでは、以下のコマンドから戻される情報について説明します。

- Web クローラー状況
- Web 以外のクローラー状況
- Web クローラーのクロール・スペース状況
- Web 以外のクローラーのクロール・スペース状況
- Web クローラーの詳細なクロール・スペース状況
- Web 以外のクローラーの詳細なクロール・スペース状況
- パーサー状況
- 索引作成状況
- 検索サーバー状況
- 詳細な検索サーバー状況

**Web クローラー状況:** Web クローラー状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。Web クローラー状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<CrawlerStatus>
<CrawlerRunLevel Value="Running"/>
<CrawlerThreadStateDist Count="4" Total="200">
<CrawlerThreadState State="FETCHING" Count="100"/>
. . .
</CrawlerThreadState State="FETCHING" Count=100>
<ActiveBucketList Count="500">
<ActiveBucket URL="http://w3.ibm.com/"
NumActURLs="355"
NumProcURLs="350"
TimeRem="5" Duration="1195"/>
. . .
</ActiveBucketList>
<CrawlRate Value="75"/>
<RecentlyCrawledURLList Count="40">
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<RecentlyCrawledURL URL="http://w3.ibm.com/foo.html"/>
<NumURLsThisSession Value="160000"/>
</CrawlerStatus>
```

以下の表で、Web クローラー状況コマンドによって戻される各 XML エレメントおよびその属性について説明します。

表 6. Web クローラー状況情報

エレメント	属性	説明
CrawlerStatus	<ul style="list-style-type: none"> <li>CrawlerThreadStateDist</li> <li>ActiveBucketList</li> <li>CrawlRate</li> <li>RecentlyCrawledURLList</li> <li>NumURLsThisSession</li> </ul>	クローラー状況。
CrawlerRunLevel Value	<ul style="list-style-type: none"> <li>『Not started』: クローラー・セッションが存在します。しかし、文書処理するための開始メッセージはまだ受け取っていません。</li> <li>『Started』: クローラーは開始しています。</li> <li>『Running』: クローラーは初期化および開始処理を終えて、アクティブな状態でクロールしています。</li> <li>『Paused』: クローラーは、アクティブ・クロールを中断し、しかし終了はしないように指示されました。</li> <li>『Stopping』: クローラーはストップ信号を受け取り、停止しようとしています。</li> <li>『Error』: クローラーはリカバリー不能な状態に陥っており、クロールを再開するにはクローラーを停止および再開する必要があります。</li> </ul>	クローラーが何をしているかという情報。
CrawlerThreadState State	ストリング (英語)	クローラー・スレッドのアクティビティ。このフィールドは、スレッド (1 つまたは複数) が何をしているかということを示します。
ActiveBucket	<ul style="list-style-type: none"> <li>URL: ストリング (URL 仕様) URL をクロールしているプロトコル、ホスト、およびポート。</li> <li>NumActURLs: 整数 (正数) クロール用にバケットが使用可能になった (アクティブになった) ときの、そのバケット内の URL の数。</li> <li>NumProcURLs: 整数 (負数以外) それまでに処理された (クロールまたは拒否された) バケットの URL の数。</li> <li>TimeRem: 正数 バケットがタイムアウトになるまでの残りの秒数。</li> <li>Duration: 整数 (負数以外) バケットがアクティブになってからの秒数。</li> </ul>	指定された Web サイトの現在のアクティビティ。

表 6. Web クローラー状況情報 (続き)

エレメント	属性	説明
CrawlRate	値: 整数 (負数以外)  1 秒あたりにクロールされたページ (全バケットを結合したもの)。	クローラーのスループット測定。
RecentlyCrawledURL	URL: ストリング (URL 仕様)  クロールされたプロトコル、ホスト、ポート、およびファイルを示すストリング。	最近クロールされたページ。
NumURLsThisSession	値: 整数 (負数以外)	クローラー (プロセス) のこのインスタンスがクロールを開始してからクロールされた URL の数。

**Web 以外のクローラー状況:** Web 以外のクローラー状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。 Web 以外のクローラー状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<GeneralStatus>
<Status>0</Status>
<StatusMessage>Idle</StatusMessage>
<NumberOfServers>1</NumberOfServers>
<NumberOfCompletedServers>1</NumberOfCompletedServers>
<NumberOfTargets>3</NumberOfTargets>
<NumberOfCompletedTargets>3</NumberOfCompletedTargets>
<NumberOfCrawledRecords>115</NumberOfCrawledRecords>
<RunningThreads>0</RunningThreads>
</GeneralStatus>
```

以下の表で、Web クローラー以外の各エンタープライズ・サーチ・クローラーでの、XML エレメントおよび属性について説明します。この情報は、クローラー状況コマンドで戻されるものです。

表 7. クローラー状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーのエレメントおよび属性

エレメントおよび属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Status	状況 (0、1、2、-1)	状況 (0、1、2、-1)	状況 (0、1、2、-1)
StatusMessage	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー
NumberOfServers	クロール・スペース内の NNTP サーバーの数。	クロール・スペース内のデータベースの数。	クロール・スペース内のデータベースの数。
NumberOfCompletedServers	クロールされた NNTP サーバーの数。	クロールされたデータベースの数。	クロールされたデータベースの数。
NumberOfTargets	クロール・スペース内の新規グループの数。	クロール・スペース内のデータベースの数。	クロール・スペース内のビューおよびフォルダーの数。
NumberOfCompletedTargets	クロールされた新規グループの数。	クロールされた表の数。	クロールされたビューおよびフォルダーの数。
NumberOfCompletedRecords	クロールされた項目の数。	クロールされたレコードの数。	クロールされた文書の数。

表7. クローラー状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーの要素および属性 (続き)

要素および属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
RunningThreads	クローラー・スレッドの数。	クローラー・スレッドの数。	クローラー・スレッドの数。

表8. クローラー状況コマンドで戻される、Exchange Server クローラー、DB2 Content Manager クローラー、および Content Edition クローラーの要素および属性

要素および属性の名前	Exchange Server クローラー	DB2 Content Manager クローラー	Content Edition クローラー
Status	状況 (0、1、2、-1)	状況 (0、1、2、-1)	状況 (0、1、2、-1)
StatusMessage	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー
NumberOfServers	クロール・スペース内の Exchange Server サーバーの数。	クロール・スペース内の Content Manager サーバーの数。	クロール・スペース内のリポジトリの数。
NumberOfCompletedServers	クロールされた Exchange Server サーバーの数。	クロールされた Content Manager サーバーの数。	クロールされたりリポジトリの数。
NumberOfTargets	クロール・スペース内のサブフォルダーの数。	クロール・スペース内の項目タイプの数。	クロール・スペース内のクラスの数。
NumberOfCompletedTargets	クロールされたサブフォルダーの数。	クロールされた項目タイプの数。	クロールされた項目クラスの数。
NumberOfCompletedRecords	クロールされた文書の数。	クロールされた文書の数。	クロールされた文書の数。
RunningThreads	クローラー・スレッドの数。	クローラー・スレッドの数。	クローラー・スレッドの数。

表9. クローラー状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーの要素および属性

要素および属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
Status	状況 (0、1、2、-1)	状況 (0、1、2、-1)	状況 (0、1、2、-1)
StatusMessage	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー	状況: 0 - アイドル、1 - 実行中、2 - 休止、-1 - エラー
NumberOfServers	クロール・スペース内のプレースの数。	クロール・スペース内のライブラリーの数。	固定値 1。
NumberOfCompletedServers	クロールされたプレースの数。	クロールされたライブラリーの数。	すべてのサブディレクトリーがクロールされた場合、0 または 1。
NumberOfTargets	クロール・スペース内のプレース・データベースおよびルーム・データベースの数。	クロール・スペース内のキャビネットの数。	クロール・スペース内のサブディレクトリーの数。

表 9. クローラー状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーの要素および属性 (続き)

要素および属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
NumberOfCompletedTargets	クローリングされたプレース・データベースおよびループ・データベースの数。	クローリングされたキャビネットの数。	クローリングされたサブディレクトリ数。
NumberOfCompletedRecords	クローリングされた文書の数。	クローリングされた文書の数。	クローリングされたファイルの数。
RunningThreads	クローラー・スレッドの数。	クローラー・スレッドの数。	クローラー・スレッドの数。

**Web クローラーのクローリング・スペース状況:** Web クローラーのクローリング・スペース状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。Web クローリング・スペース状況コマンドによって、以下の情報を戻すことができます。

表 10. Web クローラーのクローリング・スペース状況コマンドの選択マスク値

マスク・ビット	選択
1	ロー・データ・ストア内のページの数。
2	ディスカバーされたサイトの数。
4	DNS のあるサイトの数。
8	DNS のないサイトの数。
16	ディスカバーされた URL の数。
32	保管された固有ページの数。
64	クローリングされた URL の数。
128	クローリングされていない URL の数。
256	期限切れの URL の数。
512	HTTP 戻りコードの配布。

すべての値が、現行の内部データベースを使用するすべてのセッションでの累積合計を表します。

```
<CrawlStatus>
  <NumPagesInRDS Value="5422386"/>
  <NumSitesDiscovered Value="15332"/>
  <NumSitesWithDNS Value="14832"/>
  <NumSitesWithoutDNS Value="500"/>
  <NumURLsDiscovered Value="15222999"/>
  <NumUniquePagesSaved Value="6234789"/>
  <NumURLsCrawled Value="7800422"/>
  <NumURLsUncrawled Value="7422577"/>
  <NumURLsOverdue Value="14000"/>
  <HTTPCodeDist Count="4" Total="1031000"/>
  <HTTPCode Code="200" Count="1000000"/>
  <HTTPCode Code="301" Count="1000"/>
  <HTTPCode Code="404" Count="10000"/>
  <HTTPCode Code="780" Count="20000"/>
  </HTTPCode Code="780" Count="20000"/>
</CrawlStatus>
```

戻りデータには、以下のエレメントのいくつかまたはすべてが含まれます (無いこともあります)。

表 11. Web クローラーのクロール・スペース状況コマンドから戻される情報

エレメント	属性	説明
CrawlerStatus	<ul style="list-style-type: none"> <li>• NumPagesInRDS</li> <li>• NumSitesDiscovered</li> <li>• NumSitesWithDNS</li> <li>• NumSitesWithoutDNS</li> <li>• NumURLsDiscovered</li> <li>• NumUniquePagesSaved</li> <li>• NumURLsCrawled</li> <li>• NumURLsUncrawled</li> <li>• NumURLsOverdue</li> <li>• HTTPCodeDist</li> </ul>	クロールの累積状態に関して迅速に取得できる情報 (すべてのセッション)。
NumPagesInRDS	<p>値: 負でない整数</p> <p>ロー・データ・ストア (RDS) のステージング・エリア内に現在入っているページの数 (このクローラーのみ)。</p>	ロー・データ・ストア (RDS) がどのくらいフルになっているか (このクローラーのコントリビューションのみ)。
NumSitesDiscovered	<p>値: 負でない整数</p> <p>クロールによってディスカバーされたホストの数 (または、シード)。</p>	クロールすべきドメインの、クローラーの適用範囲の指標 (ホスト・カウント)。
NumSitesWithDNS	<p>値: 負でない整数</p> <p>IP アドレスに関連付けられたホストの数 (バックグラウンドのクローラーによって解決されたもの)。</p>	URL 内で DNS 名によってディスカバーされるホストの IP アドレスを、クローラーがどのくらい効率的に取得できるかの指標。
NumSitesWithoutDNS	<p>値: 負でない整数</p> <p>IP アドレスに関連付けられていないホストの数 (バックグラウンドのクローラーによって解決されたもの)。</p>	URL 内で DNS 名によってディスカバーされるホストの IP アドレスを、クローラーがどのくらい効率的に取得できるかの指標。
NumURLsDiscovered	<p>値: 負でない整数</p> <p>クローラーによって訪問された固有の URL の数。</p>	クロールすべきドメインの、クローラーの適用範囲の指標 (URL カウント)。
NumUniquePagesSaved	<p>値: 負でない整数</p> <p>その他のエンタープライズ・サーチ・コンポーネントによってさらに処理するために RDS に書き込まれた固有のページの数。</p>	索引のサイズに対する、このクローラーのコントリビューション。
NumURLsCrawled	<p>値: 負でない整数</p> <p>クローラーによってクロールされた固有の URL の数。</p>	データを処理するための、クローラーの機能の指標 (エンドツーエンド)。この数は、RDS に書き込まれるページの数によって異なります。それは、クロールされたすべてのページが RDS に書き込まれるわけではないからです。



表 11. Web クローラーのクロール・スペース状況コマンドから戻される情報 (続き)

エレメント	属性	説明
NumURLsOverdue	値: 負でない整数  再クロールする対象として適格な固有の URL の数。	Web スペースをトラバースするための、クローラーの機能の指標。

**Web 以外のクローラーのクロール・スペース状況:** Web 以外のクローラーのクロール・スペース状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。 Web 以外のクロール・スペース状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<ServerStatus>
  <Server Name ="FOUNTAIN">
    <Status>5</Status>
    <StatusMessage>Scheduled</StatusMessage>
    <NumberOfTargets>1</NumberOfTargets>
    <NumberOfCompletedTargets>1</NumberOfCompletedTargets>
    <NumberOfErrors>0</NumberOfErrors>
    <StartTime>1118354510512</StartTime>
    <EndTime>1118354514386</EndTime>
    <ScheduleConfigured>2</ScheduleConfigured>
    <ScheduleTime>1118393377000</ScheduleTime>
    <TotalTime>3874</TotalTime>
  </Server>
</ServerStatus>
```

以下の表で、Web クローラー以外の各エンタープライズ・サーチ・クローラーでの、XML エレメントおよび属性について説明します。この情報は、クロール・スペース状況コマンドで戻されるものです。Notes クローラーの場合、集約レベルが 0 のとき、Server@Name は サーバー名 + データベース名 です。集約レベルが 1 のとき、Server@Name は サーバー名 + ディレクトリー名 です。

表 12. クロール・スペース状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーのエレメントおよび属性

エレメントおよび属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Server@Name	ニュース・サーバー名	データベース名	データベース名またはディレクトリー名
Server/Status	状況: (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• -1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>

表 12. クロール・スペース状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーの要素および属性 (続き)

要素および属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>
Server/NumberOfTargets	クロール・スペース内の新規グループの数。	クロール・スペース内のデータベースの数。	クロール・スペース内のビューおよびフォルダーの数。
Server/NumberOfCompletedTargets	クロールされた新規グループの数。	クロールされた表の数。	クロールされたビューおよびフォルダーの数。
Server/NumberOfErrors	適用されない。	エラーの数。	エラーの数。
Server/StartTime	適用される場合、開始時刻。	適用される場合、開始時刻。	適用される場合、開始時刻。
Server/EndTime	適用される場合、終了時刻。	適用される場合、終了時刻。	適用される場合、終了時刻。
Server/ScheduleConfigured	適用されない。	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジュールされるように構成されていません。</li> <li>1: クローラーが、スケジュールされるように構成されています。しかし、スケジュールはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジュールされるように構成されています。そして、スケジュールはそのセッションで使用可能になっています。</li> </ul>	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジュールされるように構成されていません。</li> <li>1: クローラーが、スケジュールされるように構成されています。しかし、スケジュールはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジュールされるように構成されています。そして、スケジュールはそのセッションで使用可能になっています。</li> </ul>
Server/ScheduleTime	適用されない。	適用される場合、スケジュール時刻。	適用される場合、スケジュール時刻。
Server/TotalTime	適用される場合、合計時間。	適用される場合、合計時間。	適用される場合、合計時間。

表 12. クロール・スペース状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーの要素および属性 (続き)

要素および属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Server/AggregationLevel	適用されない。	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0、1: <ul style="list-style-type: none"> <li>0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)</li> <li>Notes クローラーは、ディレクトリー・モードで文書をクロールします。</li> </ul>

表 13. クロール・スペース状況コマンドで戻される、Exchange Server クローラー、DB2 Content Manager クローラー、および Content Edition クローラーの要素および属性

要素および属性の名前	Exchange Server クローラー	DB2 Content Manager クローラー	Content Edition クローラー
Server@Name	Exchange Server サーバー名	DB2 Content Manager サーバー	リポジトリ名
Server/Status	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>5: スケジュールされている</li> <li>-1: エラー</li> </ul>
Server/NumberOfTargets	クロール・スペース内のサブフォルダーの数。	クロール・スペース内の項目タイプの数。	クロール・スペース内の項目クラスの数。
Server/NumberOfCompletedTargets	クロールされたサブフォルダーの数。	クロールされた項目タイプの数。	クロールされた項目クラスの数。
Server/NumberOfErrors	エラーの数。	エラーの数。	エラーの数。

表 13. クロール・スペース状況コマンドで戻される、Exchange Server クローラー、DB2 Content Manager クローラー、および Content Edition クローラーの要素および属性 (続き)

要素および属性の名前	Exchange Server クローラー	DB2 Content Manager クローラー	Content Edition クローラー
Server/StartTime	適用される場合、開始時刻。	適用される場合、開始時刻。	適用される場合、開始時刻。
Server/EndTime	適用される場合、終了時刻。	適用される場合、終了時刻。	適用される場合、終了時刻。
Server/ScheduleConfigured	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>
Server/ScheduleTime	適用される場合、スケジュール時刻。	適用される場合、スケジュール時刻。	適用される場合、スケジュール時刻。
Server/TotalTime	適用される場合、合計時間。	適用される場合、合計時間。	適用される場合、合計時間。
Server/AggregationLevel	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)

表 14. クロール・スペース状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーの要素および属性

要素および属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
Server@Name	ブレース・ディレクトリース	ライブラリー・データベース	ローカル・ホストの固定値

表 14. クロール・スペース状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーのエレメントおよび属性 (続き)

エレメントおよび属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
Server/Status	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• -1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>	状況 (0、1、2、3、4、5、-1) <ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>
Server/StatusMessage	<ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>• 0: クロールされていない</li> <li>• 1: クロール中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• 5: スケジュールされている</li> <li>• -1: エラー</li> </ul>
Server/NumberOfTargets	クロール・スペース内のプレース・データベースおよびルーム・データベースの数。	クロール・スペース内のキャビネットの数。	クロール・スペース内のサブディレクトリーの数。
Server/NumberOfCompletedTargets	クロールされたプレース・データベースおよびルーム・データベースの数。	クロールされたキャビネットの数。	クロール・スペース内のサブディレクトリーの数。
Server/NumberOfErrors	エラーの数。	エラーの数。	エラーの数。
Server/StartTime	適用される場合、開始時刻。	適用される場合、開始時刻。	適用される場合、開始時刻。
Server/EndTime	適用される場合、終了時刻。	適用される場合、終了時刻。	適用される場合、終了時刻。

表 14. クロール・スペース状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーのエレメントおよび属性 (続き)

エレメントおよび属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
Server/ScheduleConfigured	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>	0、1、2 <ul style="list-style-type: none"> <li>0: クローラーが、クローラー構成ファイルに従って、スケジューリングされるように構成されていません。</li> <li>1: クローラーが、スケジューリングされるように構成されています。しかし、スケジューリングはそのセッションで使用不可になっていました。</li> <li>2: クローラーが、スケジューリングされるように構成されています。そして、スケジューリングはそのセッションで使用可能になっています。</li> </ul>
Server/ScheduleTime	適用される場合、スケジュール時刻。	適用される場合、スケジュール時刻。	適用される場合、スケジュール時刻。
Server/TotalTime	適用される場合、合計時間。	適用される場合、合計時間。	適用される場合、合計時間。
Server/AggregationLevel	0、1 <ul style="list-style-type: none"> <li>0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)</li> <li>Notes クローラーは、ディレクトリー・モードで文書をクロールします。</li> </ul>	0、1 <ul style="list-style-type: none"> <li>0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)</li> <li>Notes クローラーは、ディレクトリー・モードで文書をクロールします。</li> </ul>	0、1 <ul style="list-style-type: none"> <li>0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)</li> <li>Notes クローラーは、ディレクトリー・モードで文書をクロールします。</li> </ul>

**Web クローラーの詳細なクロール・スペース状況:** Web クローラーの詳細なクロール・スペース状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。詳細なクロール・スペース状況コマンドによって、以下の情報を戻すことができます。

表 15. Web クローラーの詳細なクロール・スペース状況コマンドの選択マスク値

マスク・ビット	選択
1	ロー・データ・ストア内のページの数。
2	ディスクカバーされたサイトの数。



表 15. Web クローラーの詳細なクロール・スペース状況コマンドの選択マスク値 (続き)

マスク・ビット	選択
4	DNS のあるサイトの数。
8	DNS のないサイトの数。
16	ディスカバーされた URL の数。
32	保管された固有ページの数。
64	クロールされた URL の数。
128	クロールされていない URL の数。
256	期限切れの URL の数。
512	HTTP 戻りコードの配布。

戻り情報の例:

```
<CrawlDetailsPerSite>
  <Site URL=http://w3.ibm.com/">
    <NumURLsDiscovered Value="5422386"/>
    <NumURLsOverdue Value="15332"/>
    <NumURLsCrawled Value="15332"/>
    <NumURLsUncrawled Value="15332"/>
    <NumURLsOverdueBy Threshold="604800" Value="14832"/>
    <NumURLsActivated Value="2200"/>
    <LastActivationTime Value="1076227340"/>
    <LastActivationDuration Value="4300"/>
    <IPAddressList Count="1"/>
      <IPAddress Value="9.205.41.33"/>
    </IPAddressList>
    <RobotsContent>
      robots content. . .
    </RobotsContent>
    <HTTPCodeDist Count="4" Total="1031000"/>
      <HTTPCode Code="200" Count ="1000000"/>
      <HTTPCode Code="301" Count ="1000"/>
      <HTTPCode Code="404" Count ="10000"/>
      <HTTPCode Code="780" Count="20000"/>
    </HTTPCodeDist>
  </CrawlDetailsPerSite>
```

以下の表で、Web クローラーの詳細なクロール・スペース状況として戻される各フィールドについて説明します。

表 16. Web クローラーの詳細なクロール・スペース状況コマンドから戻される情報

エレメント	属性	説明
CrawlDetailsPerSite	<ul style="list-style-type: none"> <li>LastActivationTime:</li> <li>LastActivationDuration:</li> <li>IPAddressList:</li> <li>RobotsContent:</li> <li>HTTPCodeDist:</li> </ul>	あるサイトの詳細な状態に関して迅速に取得できる情報。
Site	URL	サイト・ルート・ページの URL。
NumURLsDiscovered	値	サイトからディスカバーされた URL の数。

表 16. Web クローラーの詳細なクロール・スペース状況コマンドから戻される情報 (続き)

エレメント	属性	説明
NumURLsOverdue	値	サイトから再クロールする対象として適格な URL の数。
NumURLsCrawled	値	サイトに応じてクロールされた URL の数。
NumURLsUncrawled	値	サイトに応じてまだクロールされていない URL の数。
NumURLsOverdueBy	しきい値、値: 整数 (正数または負数)	少なくとも数秒前に (しきい値) 再クロールする対象として適格だった、あるいは、この後さらに長い秒数が経って (しきい値) 再クロールする対象として適格になる予定の URL の数。
	クロールを検出するためのしきい値が、指定された時間で (負数 = 現在までの秒数)、あるいはこの後指定された時間間隔内で (正数 = 現在からの秒数)、期限切れになります。	
NumURLsActivated	値	このサイトの最後のスキャン中にメモリー内に入れられ、クローラー・スレッドから使用できるようになった URL の数。
LastActivationTime	値	このサイトの URL が最後にメモリー内に入れられたエポック以降、経過した秒数。
LastActivationDuration	値	このサイトの URL が最後にメモリーに入れられ、クローラー・スレッドから使用できるようになっていた秒数。
IPAddressList	IPAddress	このサイトのサーバー・ホスト用の、認識されているすべての IP アドレス。
IPAddress	値	サイトのサーバー・ホスト用の、IPv4 ドット表記アドレス。
RobotsContent	テキスト	ロボット・ファイルにテキストが存在している場合、そのテキスト。
HTTPCodeDist	HTTPCode	このサイトの試行ダウンロードからの、HTTP コードの配布。
HTTPCode	コード: 整数 HTTP 戻りコードまたは別の内部コード。	このサイトのクロール中に、特定の HTTP 戻りコードが発生した回数。

**Web 以外のクローラーの詳細なクロール・スペース状況:** Web 以外のクローラーの詳細なクロール・スペース状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。Web 以外のクローラーの詳細なクロール・スペース状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <?xml version='1.0' encoding='UTF-8'?>
<TargetStatus>
  <Target Name ="escmgr.crawlerinstances">
    <Status>2</Status>
    <StatusMessage>Completed</StatusMessage>
    <NumberOfRecords></NumberOfRecords>
```

```

|         <NumberOfCrawledRecords>117</NumberOfCrawledRecords>
|         <NumberOfInsertedRecords>21</NumberOfInsertedRecords>
|         <NumberOfUpdatedRecords>45</NumberOfUpdatedRecords>
|         <StartTime>1118354510727</StartTime>
|         <EndTime>1118354514386</EndTime>
|         <AggregationLevel>0<AggregationLevel>
|     <Target>
| </TargetStatus>

```

表 17. 詳細なクローल・スペース状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーの要素および属性

要素および属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Target@Name	ニュース・グループ名	テーブル名	ビューまたはフォルダー名
Target@CrawlType	適用されない。	0、1: <ul style="list-style-type: none"> <li>0: アクティブ・クロール (通常)</li> <li>1: パッシブ・クロール (DB2 Event Publishing)</li> </ul>	0
Target/Status	状況: (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況: (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況: (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>
Target/NumberOfRecords	サーバー上の最後の項目番号。	クロールされたレコードの数。	クロールされた文書の数。
Target/NumberOfCompletedRecords	クロールされた項目の数。	クロールされたレコードの数。	クロールされた文書の数。
Target/NumberOfInsertedRecords	新しく通知された項目の数。	挿入されたレコードの数。	挿入されたレコードの数。
Target/NumberOfUpdatedRecords	適用されない。	更新されたレコードの数。	更新されたレコードの数。
Target/NumberOfDeletedRecords	適用されない。	削除されたレコードの数。	削除されたレコードの数。
Target/StartTime	クローラーが最後に開始された日時。	クローラーが最後に開始された日時。	クローラーが最後に開始された日時。
Target/EndTime	クロールが完了した日時。	クロールが完了した日時。	クロールが完了した日時。

表 17. 詳細なクローल・スペース状況コマンドで戻される、NNTP クローラー、DB2 クローラー、および Notes クローラーの要素および属性 (続き)

要素および属性の名前	NNTP クローラー	DB2 クローラー	Notes クローラー
Target/TotalTime	クローラーがクロールに費やした時間の合計。	クローラーがクロールに費やした時間の合計。	クローラーがクロールに費やした時間の合計。
Target/AggregationLevel	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0、1: <ul style="list-style-type: none"> <li>0: Notes クローラーは、通常モードで文書をクロールします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)</li> <li>Notes クローラーは、ディレクトリー・モードで文書をクロールします。</li> </ul>
Target/LastUpdateTime	適用されない。	最後に更新された時刻: <ul style="list-style-type: none"> <li>0: アクティブ・クロール (通常)</li> <li>1: パッシブ・クロール (DB2 Event Publishing)</li> </ul>	適用されない。
Target/LastResetTime	適用されない。	最後に統計をリセットした時間: <ul style="list-style-type: none"> <li>0: アクティブ・クロール (通常)</li> <li>1: パッシブ・クロール (DB2 Event Publishing)</li> </ul>	適用されない。

表 18. 詳細なクロール・スペース状況コマンドで戻される、Exchange Server クローラー、DB2 Content Manager クローラー、および Content Edition クローラーの要素および属性

要素および属性の名前	Exchange Server クローラー	DB2 Content Manager クローラー	Content Edition クローラー
Target@Name	サブフォルダー名	項目タイプ名	項目クラス名
Target@CrawlType	0	0	0
Target/Status	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クロールされていない</li> <li>1: クロール中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>

表 18. 詳細なクローल・スペース状況コマンドで戻される、Exchange Server クローラー、DB2 Content Manager クローラー、および Content Edition クローラーの要素および属性 (続き)

要素および属性の名前	Exchange Server クローラー	DB2 Content Manager クローラー	Content Edition クローラー
Target/StatusMessage	<ul style="list-style-type: none"> <li>• 0: クローラされていない</li> <li>• 1: クローラ中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• -1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>• 0: クローラされていない</li> <li>• 1: クローラ中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• -1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>• 0: クローラされていない</li> <li>• 1: クローラ中</li> <li>• 2: 完了 (スケジュールされていない)</li> <li>• 3: 待機中</li> <li>• 4: 休止されている</li> <li>• -1: エラー</li> </ul>
Target/NumberOfRecords	適用されない。	適用されない。	適用されない。
Target/NumberOfCompletedRecords	クローラされた文書の数。	クローラされた文書の数。	クローラされた文書の数。
Target/NumberOfInsertedRecords	挿入されたレコードの数。	挿入されたレコードの数。	挿入されたレコードの数。
Target/NumberOfUpdatedRecords	適用されない。	更新されたレコードの数。	更新されたレコードの数。
Target/NumberOfDeletedRecords	適用されない。	削除されたレコードの数。	削除されたレコードの数。
Target/StartTime	クローラが最後に開始された日時。	クローラが最後に開始された日時。	クローラが最後に開始された日時。
Target/EndTime	クローラが完了した日時。	クローラが完了した日時。	クローラが完了した日時。
Target/TotalTime	クローラがクローラに費やした時間の合計。	クローラがクローラに費やした時間の合計。	クローラがクローラに費やした時間の合計。
Target/AggregationLevel	0: Notes クローラは、通常モードで文書をクローラします。(Notes クローラ以外のクローラでは、常に 0 が戻されます。)	0: Notes クローラは、通常モードで文書をクローラします。(Notes クローラ以外のクローラでは、常に 0 が戻されます。)	0: Notes クローラは、通常モードで文書をクローラします。(Notes クローラ以外のクローラでは、常に 0 が戻されます。)
Target/LastUpdateTime	適用されない。	適用されない。	適用されない。
Target/LastResetTime	適用されない。	適用されない。	適用されない。

表 19. 詳細なクローラ・スペース状況コマンドで戻される、QuickPlace クローラ、Domino Document Manager クローラ、UNIX ファイル・システム・クローラ、および Windows ファイル・システム・クローラの要素および属性

要素および属性の名前	QuickPlace クローラ	Domino Document Manager クローラ	UNIX および Windows ファイル・システム・クローラ
Target@Name	プレース・データベース名またはルーム・データベース名	キャビネット・データベース名	サブディレクトリー名
Target@CrawlType	0	0	0

表 19. 詳細なクローल・スペース状況コマンドで戻される、QuickPlace クローラー、Domino Document Manager クローラー、UNIX ファイル・システム・クローラー、および Windows ファイル・システム・クローラーの要素および属性 (続き)

要素および属性の名前	QuickPlace クローラー	Domino Document Manager クローラー	UNIX および Windows ファイル・システム・クローラー
Target/Status	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	状況 (0、1、2、3、4、-1) <ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>
Target/StatusMessage	<ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>	<ul style="list-style-type: none"> <li>0: クローラされていない</li> <li>1: クローラ中</li> <li>2: 完了 (スケジュールされていない)</li> <li>3: 待機中</li> <li>4: 休止されている</li> <li>-1: エラー</li> </ul>
Target/NumberOfRecords	適用されない。	適用されない。	適用されない。
Target/NumberOfCompletedRecords	クローラされた文書の数。	クローラされた文書の数。	クローラされたファイルの数。
Target/NumberOfInsertedRecords	挿入されたレコードの数。	挿入されたレコードの数。	挿入されたレコードの数。
Target/NumberOfUpdatedRecords	クローラされたブレース・データベースおよびルーム・データベースの数。	クローラされたキャビネットの数。	クローラ・スペース内のサブディレクトリーの数。
Target/NumberOfDeletedRecords	更新されたレコードの数。	更新されたレコードの数。	更新されたレコードの数。
Target/StartTime	クローラーが最後に開始された日時。	クローラーが最後に開始された日時。	クローラーが最後に開始された日時。
Target/EndTime	クローラが完了した日時。	クローラが完了した日時。	クローラが完了した日時。
Target/TotalTime	クローラーがクローラに費やした時間の合計。	クローラーがクローラに費やした時間の合計。	クローラーがクローラに費やした時間の合計。
Target/AggregationLevel	0: Notes クローラーは、通常モードで文書をクローラします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0: Notes クローラーは、通常モードで文書をクローラします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)	0: Notes クローラーは、通常モードで文書をクローラします。(Notes クローラー以外のクローラーでは、常に 0 が戻されます。)
Target/LastUpdatedTime	適用されない。	適用されない。	適用されない。
Target/LastResetTime	適用されない。	適用されない。	適用されない。



**パーサー状況:** パーサー状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。パーサー状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <Monitor Type="Parser">
<ParserStatus>
  <Status>1</Status>
  <State>Parsing</State>
  <SnapshotTimeStamp>1124318637564</SnapshotTimeStamp>
  <NumberOfDocsToBeIndexed>231974</NumberOfDocsToBeIndexed>
  <ParseRate>0</ParseRate>
  <ParseRateMBPerHour>0</ParseRateMBPerHour>
  <NumberOfCpmThreads>3</NumberOfCpmThreads>
  <ParserServiceSession>parserservice.1</ParserServiceSession>
</ParserStatus>
</CrawlerStatus>
<Name>WEBCrawler1</Name>
  <Crawlerid>coll.WEB1.esadmin</Crawlerid>
  <Type>WEB</Type>
  <ParserStatus>1</ParserStatus>
  <NumberOfDocsToBeParsed>15881</NumberOfDocsToBeParsed>
  <NumberOfDocsAlreadyParsed>29</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
<CrawlerStatus>
  <Name>Data Listener (server1)</Name>
  <Crawlerid>datalistener</Crawlerid>
  <Type>datalistener</Type>
  <ParserStatus>0</ParserStatus>
  <NumberOfDocsToBeParsed>0</NumberOfDocsToBeParsed>
  <NumberOfDocsAlreadyParsed>0</NumberOfDocsAlreadyParsed>
</CrawlerStatus>
</Monitor>
```

以下の表で、パーサー状況コマンドによって戻される情報の XML エlement について説明します。

表 20. パーサー状況コマンドのElement

Element	説明
Status	このコレクションのためのパーサー・セッションが実行中である場合は 1、パーサー・セッションが停止している場合は 0 です。

表 20. パーサー状況コマンドの要素 (続き)

要素	説明
State	<p>起こりうる状態は、Idle (アイドル)、Restart (再始動)、Parsing (構文解析中)、Resuming (再開中)、Stopped (停止されている)、Initializing (初期化中)、Pause (休止)、NoParserServiceIsAvailable (使用可能なパーサー・サービスがない) です。</p> <p>Idle (アイドル) という状態は、パーサーが、このコレクション内のクローラーからさらに多くの文書が到着するのを待って、N 分間スリープしていることを示します。デフォルトのスリープ時間は 300 秒です。</p> <p>Restart (再始動) という状態は、パーサーが、構文解析/トークン化 JVM が再始動されるのを待っていることを示します。構文解析/トークン化 JVM は別個のセッションで実行されます。これは、文書が最終的に処理される場所です。</p> <p>Parsing (構文解析中) という状態は、パーサーが文書処理していることを示します。</p> <p>Pause (休止) という状態は、パーサーが、このコレクション用の索引作成セッションによって休止されたことを示します。Initializing (初期化中) という状態は、パーサーが開始され、その状態を初期化していることを示します。</p> <p>Resuming (再開) という状態は、パーサーが、このコレクション用の索引作成セッションによって、Pause (休止) 状態から Parsing (構文解析中) 状態に再開されたことを示します。</p> <p>NoParserServiceIsAvailable (使用可能なパーサー・サービスがない) という状態は、このコレクション用の文書処理するために使用可能な構文解析/トークン化 JVM がないことを示します。これは、すべての構文解析/トークン化 JVM が、他のコレクションで使用されていることを意味します。</p>
SnapshotTimeStamp	次回パーサーがクローラー・テーブルから文書を読み取る時刻で、1970 年以降の秒数。
NumberOfDocsToBeIndexed	このコレクション用のストア内のドキュメントの数。この数には、次回の索引作成で削除するようマークされている文書も含まれます。
ParseRate	構文解析速度 (1 秒あたりの文書数)。
ParseRateMBPerHour	構文解析速度 (1 時間あたりの MB)。
NumberOfCpmThreads	このコレクション用の文書処理のために構文解析/トークン化 JVM が使用する CPM スレッドの数。
ParserServiceSession	このコレクション用の文書処理している構文解析/トークン化 JVM の名前。このフィールドは、パーサーが Parsing (構文解析中) 状態になっている場合にのみ使用可能になります。
Name	クローラーの名前。
Type	クローラーのタイプ (Web、NNTP、DB2、など)。
ParserStatus	<ul style="list-style-type: none"> <li>• 0: このクローラーからのドキュメントは、まだ構文解析されていません。</li> <li>• 1: このクローラーからのドキュメントは、現在構文解析されているところです。</li> <li>• 2: このクローラーからのドキュメントは、構文解析されました。</li> </ul>
NumberOfDocsToBeParsed	このクローラーからのドキュメントで、構文解析されるのを待っているドキュメントの数。

表 20. パーサー状況コマンドのエレメント (続き)

エレメント	説明
NubmerOfDocsAlreadyParsed	このクローラーからのドキュメントで、構文解析されたドキュメントの数。

**索引作成状況:** 索引作成状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。索引作成状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <Monitor Type="Parser">
<Monitor Type="MainIndexHistory" Count="1">
  <IndexStatus Id="1"/>
  <StartTime>1123101789411</StartTime>
  <Progress>0</Progress>
  <CurrentPhase>0</CurrentPhase>
  <TotalPhase>0</TotalPhase>
  <IndexCopyProgress>0</IndexCopyProgress>
  <CurrentServer>0</CurrentServer>
  <TotalServer>0</TotalServer>
  <IndexCopyTime>0</IndexCopyTime>
  <IndexBuildTime>0</IndexBuildTime>
  <Status>0</Status>
  <StopTime>1123101789618</StopTime>
  <NumberOfDocuments>0</NumberOfDocuments>
</IndexStatus>
<CurrentIndexWildcardSupport/>
<NextIndexWildcardSupport/>
<ScheduleStatus>
  <Status>1</Status>
</Monitor>
```

以下の表で、索引作成状況コマンドによって戻される情報の各 XML エレメントについて説明します。

表 21. 索引作成状況コマンドのエレメント

エレメント	説明
IndexStatusId	索引状況 ID。
StartTime	この索引作成が開始された時刻で、1970 年以降の秒数。
Progress	この索引作成の完了したパーセンテージ。
CurrentPhase	<ul style="list-style-type: none"> <li>1: 保管再書き込みフェーズ</li> <li>2: グローバル分析フェーズ</li> <li>3: 索引作成フェーズ</li> </ul>
TotalPhase	この索引作成のフェーズの数。現在この値は 3 です。
IndexCopyProgress	索引コピーの完了したパーセンテージ。索引コピー処理では、作成された索引を索引作成サーバーから検索サーバーにコピーします。
CurrentServer	索引コピーで、索引がコピーされる先の検索サーバー。
TotalServer	索引をコピーする先の検索サーバーの数。
IndexCopyTime	索引作成のすべてのフェーズの合計時間。
Progress	索引作成およびコピーが成功した場合は 0、エラー・メッセージ・コードの場合は 0 以外の数字。
StopTime	索引作成 (すべてのフェーズ) および索引コピーの終了時刻。
TotalTime	開始時刻と終了時刻の間の期間。
NumberOfDocuments	索引内の文書の数。

表 21. 索引作成状況コマンドの要素 (続き)

要素	説明
CurrentIndexWildcardSupport	次の索引作成に使用すべきワイルドカード設定。取りうる値は、None、QueryExpansion、または IndexExpansion です。
ScheduleStatus	<ul style="list-style-type: none"> <li>0: このコレクションおよび索引タイプのためのスケジュールが使用可能になっていない場合。</li> <li>1: このコレクションおよび索引タイプのためのスケジュールが使用可能になっている場合。</li> </ul>
ScheduledTimeEnabled	次回このコレクションおよび索引タイプのために索引作成が実行される時刻で、1970 年以降の秒数。

**検索サーバー状況:** 検索サーバー状況を取得するコマンドを実行すると、XML 文書フォーマットで情報が戻されます。検索サーバー状況コマンドによって、以下の情報を戻すことができます。

```
FFQC5314I Result: <?xml version="1.0"?>
<Monitor Type="Search" Count="1">
<SearchStatus Name="Search Manager (node1)" SearchID=
"searchmanager.node1" HostName="myComputer.svl.ibm.com">
<Status>1</Status>
</SearchStatus>
</Monitor>
```

以下の表で、検索サーバー状況コマンドによって戻される情報の XML エlement について説明します。

表 22. 検索サーバー状況コマンドの要素

要素	説明
SearchStatusName	このコレクション用の検索索引のモニターおよび保守を行っている検索マネージャー・セッションの名前および ID。
HostName	検索索引が実行されているサーバーのホスト名。
Status	<ul style="list-style-type: none"> <li>0: このコレクション用の検索索引が実行されていない場合。</li> <li>1: このコレクション用の検索索引が実行されている場合。</li> </ul>

**詳細な検索サーバー状況:** 検索サーバー状況を戻すコマンドによって、以下の情報を戻すことができます。

```
FFQC5303I Search Manager (node1) (sid: searchmanager.node1)
is already running. PID: 15711
FFQC5314I Result: PID=18390
CacheHits=3
QueryRate=1
Port=44008
SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1
```

以下の表で、詳細な検索サーバー状況コマンドによって戻される情報の項目について説明します。

表 23. 詳細な検索サーバー状況コマンドの項目

項目	説明
CacheHits	検索キャッシュから取得される結果の数。
QueryRate	最後の時間間隔で受信した照会の数。デフォルトでは、時間間隔は 5 分です。
Port	照会を listen または受信するために検索索引が使用するポート番号。
SessionId	このコレクションの検索索引のセッション ID。
CacheHitRate	検索キャッシュから取得される結果の数 (すべての検索結果に対するパーセンテージ)。
ResponseTime	指定した時間間隔 (デフォルトは 5 分) での平均応答時間 (ミリ秒)。
Status	<ul style="list-style-type: none"> <li>0: このコレクション用の検索索引が実行されていない場合。</li> <li>1: このコレクション用の検索索引が実行されている場合。</li> </ul>
SessionName	このコレクションの検索索引のセッション名。

## esadmin コマンドの戻りコード

esadmin コマンドから戻るコードは、以下のとおりです。

表 24. esadmin コマンドの戻りコード

コード	名前	説明
0	CODE_ERROR_NONE	コマンドが正常に完了しました。
102	CODE_ERROR_INSTANTIATION_EXCEPTION	コマンド・ハンドラーをインスタンス化するときにエラーが発生しました。
103	CODE_ERROR_ACCESS_EXCEPTION	コマンド・ハンドラーをインスタンス化するときに、正しくないアクセスによるエラーが発生しました。
104	CODE_ERROR_EXECUTE_EXCEPTION	
105	CODE_ERROR_THROWABLE	
106	CODE_ERROR_NO_SUCH_METHOD	
107	CODE_ERROR_INVALID_SESSION	
108	CODE_ERROR_INVALID_PARAMETER	
109	CODE_ERROR_SESSION_NOT_RUNNING	

## セッション ID の取得

esadmin check コマンドを使用して、エンタープライズ・サーチ・コンポーネントおよびそれらに対応するセッション ID のリストを表示することができます。以下の表で、共通セッション、それらの ID、それらが存在しているサーバー、およびセッションの状態のリストを示します。

表 25. セッション名、起点サーバー、セッション ID、およびセッション状態の例

セッション	セッションが実行されているサーバー	セッション ID	セッションの状態
configmanager	索引サーバー	10433	開始済み
controller	索引サーバー	10464	開始済み
customcommunication	索引サーバー	適用されない	適用されない

表 25. セッション名、起点サーバー、セッション ID、およびセッション状態の例 (続き)

セッション	セッションが実行されているサーバー	セッション ID	セッションの状態
datalistener	索引サーバー	10582	開始済み
discovery	索引サーバー	10649	開始済み
monitor	索引サーバー	10682	開始済み
parserservice	索引サーバー	10718	開始済み
resource.node1	索引サーバー	10759	開始済み
samplecpp	索引サーバー	10827	開始済み
sampletest	索引サーバー	10857	開始済み
scheduler	索引サーバー	10889	開始済み
searchmanager.node1	索引サーバー	10927	開始済み
utilities.node1	索引サーバー	10384	開始済み

### 関連概念

291 ページの『エンタープライズ・サーチ・システムのバックアップとリストア』

バックアップ・スクリプトおよびリストア・スクリプトを使用して、エンタープライズ・サーチ・システムをバックアップおよびリストアすることができます。

257 ページの『エンタープライズ・サーチ・アクティビティのモニター』  
システム・アクティビティおよびコレクション・アクティビティをモニターすると、さまざまな プロセスの状況の表示、問題発生の可能性の監視、または、パフォーマンス改善のための構成設定の調整を行うことができます。

### 関連タスク

262 ページの『クローラーのモニター』

コレクション内の各クローラーの状態に関する一般情報を表示するか、または、クローラーのアクティビティに関する詳細情報を表示するオプションを選択することができます。

251 ページの『エンタープライズ・サーチ・サーバーの開始』

ユーザーがコレクションを検索できるようにするために、システム・プロセスを開始し、その後、コレクションのクロール、構文解析、索引付け、検索を行うサーバーを開始する必要があります。

253 ページの『エンタープライズ・サーチ・サーバーの停止』

エンタープライズ・サーチ・コンポーネントの構成に変更を加えた場合や、問題のトラブルシューティングが必要な場合などには、エンタープライズ・サーチ・サーバーを停止して再始動する必要があります。



## エンタープライズ・サーチの資料

WebSphere Information Integrator OmniFind Edition の資料は、PDF または HTML で読むことができます。

WebSphere Information Integrator OmniFind Edition インストール・プログラムは、「インフォメーション・センター」を自動的にインストールします。インストール・プログラムは検索サーバーに「インフォメーション・センター」をインストールします。マルチサーバー・インストールの場合、「インフォメーション・センター」は、両方の検索サーバーにインストールされます。「インフォメーション・センター」をインストールしていない場合、「ヘルプ」をクリックすると、IBM Web サイトの「インフォメーション・センター」が開きます。エンタープライズ・サーチの「HTML トピック」を参照するには、「インフォメーション・センター」を開始します。

PDF 資料を参照するには、docs/locale/pdf に移動します。例えば、英語で資料を見つけるには、docs/en\_US/pdf に移動します。PDF 資料、ダウンロード、フィックス、技術情報、およびインフォメーション・センターを、WebSphere Information Integrator OmniFind Edition サポート・サイトで表示できます。

以下のテーブルは、使用可能な資料、ファイル名、ロケーションを示します。

表 26. エンタープライズ・サーチ用 PDF 資料

ヘッダー	ヘッダー	ヘッダー
エンタープライズ・サーチのインストール・ガイド (この文書のトピックは、「インフォメーション・センター」でも使用可能です。)	iiysi.pdf	docs/locale/pdf/
エンタープライズ・サーチの管理 (この文書のトピックは、「インフォメーション・センター」でも使用可能です。)	iiysa.pdf	docs/locale/pdf/
エンタープライズ・サーチプログラミング・ガイドおよび API リファレンス (この文書のトピックは、「インフォメーション・センター」でも使用可能です。)	iiysp.pdf	docs/locale/pdf/
メッセージ・リファレンス (この文書のトピックは、「インフォメーション・センター」でも使用可能です。)	iiysm.pdf	docs/locale/pdf/

表 26. エンタープライズ・サーチ用 PDF 資料 (続き)

ヘッダー	ヘッダー	ヘッダー
エンタープライズ・サーチのインストール要件 (この文書のトピックは、「インフォメーション・センター」でも使用可能です。)	iiysr.txt または iiysr.htm	docs/locale/ (このファイルは、「ファースト・ステップ」プログラムから起動することも可能です。)
リリース情報	iiysn.pdf	IBM WebSphere Information Integrator OmniFind Edition 資料 の Web サイトでのみ入手可能です。
テキスト分析機能ガイド	iiyst.pdf	docs/locale/pdf/

---

## WebSphere II OmniFind Edition アクセシビリティ

IBM WebSphere Information Integrator OmniFind Edition のユーザー・インターフェースおよび資料はアクセス可能です。

### インストール・プログラム

キーボード・ショートカットを使用して、WebSphere II OmniFind Edition インストール・プログラム全体を移動することができます。以下のテーブルは、キーボード・ショートカットを説明しています。

表 27. インストール・プログラム用キーボード・ショートカット

アクション	ショートカット
ラジオ・ボタンの強調表示	矢印キー
ラジオ・ボタンの選択	タブ・キー
プッシュボタンの強調表示	タブ・キー
プッシュボタンの選択	Enter キー
次のウィンドウまたは前のウィンドウへ移動、またはキャンセル	タブ・キーを押してプッシュボタンを強調表示し、Enter キーを押す
アクティブ・ウィンドウを非アクティブにする	Ctrl + Alt + Esc

### エンタープライズ・サーチ管理コンソールおよびインフォメーション・センター

管理コンソールおよびインフォメーション・センターは、Microsoft Internet Explorer または Mozilla FireFox で表示することのできるブラウザー・ベースのインターフェースです。ブラウザーのキーボード・ショートカットのリストおよび他のアクセシビリティ機能については、Internet Explorer または FireFox のオンライン・ヘルプを参照してください。

### PDF 資料

エンタープライズ・サーチ資料のすべてを PDF で表示できます。PDF 文書は、Adobe Acrobat Version 6.0 によって利用できます。PDF 文書は、ほとんどのスクリーン・リーダー用に構造化され、それによって読むことができます。



---

## エンタープライズ・サーチの用語集

この用語集では、エンタープライズ・サーチのインターフェースおよび資料で使用される用語を定義します。

### ア

#### アクセス制御リスト (access control list)

関連付けられたオブジェクトにアクセスできるユーザーを識別し、そのオブジェクトへのユーザーのアクセス権限を指定するリスト。

#### アノテーター (annotator)

特定の言語分析タスクを実行して、注釈を生成し、記録するソフトウェア・コンポーネント。アノテーターは、分析エンジンにおける分析論理コンポーネントです。

#### エスケープ文字 (escape character)

後続の 1 つ以上の文字に対して特殊な意味を抑制または設定する文字。

#### エンキュー (enqueue)

キューに項目を入れること。

#### エンタープライズ・サーチ管理者 (enterprise search administrator)

エンタープライズ・サーチ・システム全体を管理できる管理役割。

#### オペレーター (operator)

コレクション・レベルのプロセスを監視、開始、停止する権限を持つエンタープライズ・サーチ・ユーザー。

### カ

#### 改行文字 (newline character)

印刷または表示位置を 1 行下へ移動させる制御文字。システムによっては、複数の文字を要求するものもあります。

#### 概念抽出 (concept extraction)

テキスト文書にある重要な語彙項目 (人、場所、製品など) を識別し、その項目リストを生成する検索機能。『テーマ抽出』も参照。

#### 外部データ・ソース (external data source)

WebSphere Information Integrator OmniFind Edition によってクロール、構文解析、または索引付けされていないフェデレーションに対するデータ・ソース。外部データ・ソースの検索は、それらのデータ・ソースの照会アプリケーション・プログラミング・インターフェースに委任されます。

#### 鍵ストア・ファイル (keystore file)

署名者証明書として保管される公開鍵、および個人証明書に保管される秘密鍵を含むキー・データベース・ファイル。

#### カタカナ (Katakana)

2 つの一般的日本語表音文字の 1 つで使用されるシンボルから構成される文字セット。外国語のワードを表音的に書く場合に主に使用されます。

**カテゴリー (category)**

類似した特性を持つ文書のグループ。

**カテゴリー・ツリー (category tree)**

エンタープライズ・サーチ管理コンソールに表示されるカテゴリー階層。

**管理役割 (administrative role)**

エンタープライズ・サーチ管理コンソールで実行できる機能を決めるユーザー種別。この役割により、ユーザーが管理できるコレクションも決まる。

**共通通信層 (CCL) (Common Communication Layer)**

WebSphere Information Integrator OmniFind Edition のさまざまなコンポーネント (コントローラー、パーサー、クローラー、インデクサー) を結びつける通信インフラストラクチャー。

**共通分析構造 (common analysis structure)**

テキスト分析エンジンによって分析される文書を保管する構造。情報は、注釈の形式の共通分析構造および他の機能構造で保管されます。

**近接検索 (proximity search)**

同一のセンテンス、パラグラフ、または文書にある一定の語を探す検索タイプ。

**クイック・リンク (quick link)**

URI とキーワードと句の間のアソシエーション。

**クローラー (crawler)**

データ・ソースから文書を取り出し、検索索引作成用の情報を収集するソフトウェア・プログラム。

**クロール・スペース (crawl space)**

指定パターン (データベース名、ファイル・システム・パス、ドメイン・ネーム、IP アドレス、URL など) に一致するソースの集合。クローラーはここから読み取って索引用の項目を取り出す。

**言語の識別 (language identification)**

文書の言語を判別するエンタープライズ・サーチ機能。

**言語分析検索 (linguistic search)**

基本形に戻したり (例: *mice* は *mouse* として索引付けされる)、または基本形を使用して拡張したり (複合語語のように) した語句を使用して文書を表示、取得、索引付けする検索タイプ。

**検索アプリケーション (search application)**

エンタープライズ・サーチ・システムで、照会の処理、索引の検索、検索結果の表示、コレクション用のソース文書の取得を行うプログラム。

**検索エンジン (search engine)**

検索要求を受け取り、文書リストをユーザーに戻すプログラム。

**検索キャッシュ (search cache)**

以前の検索要求のデータと結果を保持するバッファー。

**検索結果 (search results)**

検索要求に一致する文書のリスト。

**検索索引ファイル (search index files)**

検索エンジンで索引が保管されているファイルのセット。



### 合字 (ligature)

結びつけることによって 1 つの文字として表示される複数の文字。例: f と i を結びつけて合字 fi を形成。

### 高頻度ランキング (popular ranking)

文書の既存のランキングに、その検索頻度を基に加算するランキング・タイプ。

### 語のステミング (word stemming)

言語学的な正規化のプロセス。1 つのワードの異形を一般形に分解する。例えば、connections、connective、および connected のようなワードは connect に戻されます。

### コレクション (collection)

データ・ソースと、そのクロール、解析、索引作成、検索用のオプションの集合。

## サ

### サーブレット (servlet)

Web サーバー上で稼働し、Web クライアント要求に対する応答として動的コンテンツを生成することにより、サーバーの機能性を拡張する Java プログラム。サーブレットは、一般的に、データベースを Web に接続するのに使用されます。

### 索引 (index)

『フルテキスト索引』を参照。

### 索引キュー (index queue)

索引の再編成要求、または処理される索引のリフレッシュ要求のリスト。

### 索引付けしないディレクティブ (no-index directive)

索引内の Web ページのコンテンツを含まないように、ロボット (Web クローラーなど) を指示する、それらのページ内のディレクティブ。

### 索引の再編成 (index reorganization)

エンタープライズ・サーチ・システムで索引を作成する処理。『索引のリフレッシュ』と対比。

### 索引のリフレッシュ (index refresh)

エンタープライズ・サーチ・システム内の既存の索引に新しい情報を追加する処理。『索引の再編成』と対比。

### シード URL (seed URL)

クロールの開始点。

### 識別名 (distinguished name)

ディレクトリーのエントリーを一意的に識別する名前。識別名は、コンマで分離された「属性:値 (attribute:value)」ペアで構成されます。また、デジタル証明書のエンティティを一意的に識別する名前/値ペアのセット(例: CN=個人の名前および C=国または地域)。

### 字句類縁性 (lexical affinity)

文書内で相互に密接な関係で出現する検索語間の関係。字句類縁性を使用して、結果の適合度を算出する。

**自然言語照会 (natural language query)**

キーワードを単純に並べるのではなく、文語表現 (「Who runs the finance department?」など) を分析する検索のタイプ。

**情報抽出 (information extraction)**

概念抽出のタイプの 1 つで、テキスト文書内の重要な語彙項目 (名前、用語、式など) を自動的に認識するもの。

**証明書 (certificate)**

公開鍵を証明書の所有者の ID に結合するデジタル文書で、それによって証明書の所有者を認証済みにすることができます。証明書は、認証局 (CA) によって発行されます。

**処理エンジン・アーカイブ (processing engine archive)**

UIMA 分析エンジン、およびエンタープライズ・サーチのカスタム分析に使用するために要求されるリソースのすべてを含む .pear zip アーカイブ・ファイル。

**信用証明情報 (credential)**

認証中に獲得される詳細情報で、ユーザー、グループ・アソシエーション、および他のセキュリティー関連識別属性を記述しています。信用証明情報は、許可、監査、および委任など、多数のサービスを実行するのに使用できます。

**ステミング (stemming)**

『語のステミング』を参照。

**ストップワード (stop word)**

共通に使用されるワードで、*the*、*an*、または*and* など、検索アプリケーションが無視するもの。

**ストップワードの除去 (stop word removal)**

共通ワードを無視して、より関連性のある結果を戻すために、照会から索引からストップワードを共通を除去するプロセス。

**静的要約 (static summarization)**

要約タイプの 1 つ。検索結果には、文書の指定および保管された要約が含まれる。『動的要約』と対比。

**静的ランキング (static ranking)**

ランキング・タイプの 1 つ。日付や、その文書を指すリンク数など、ランキングされる文書に関する係数でランキングが上がる。『動的ランキング』と対比。

**セキュリティー・トークン (security token)**

コレクションの文書へのアクセス許可に使用される ID とセキュリティーに関する情報。データ・ソース・タイプによって、サポートするセキュリティー・トークンのタイプは異なる。例えば、ユーザー役割、ユーザー ID、グループ ID や、コンテンツへのアクセス制御用のその他の情報などがある。

**セグメンテーション (segmentation)**

パス制御が基本情報単位を、BIU セグメントと呼ばれる、より小さな単位に分割する処理。これによって、隣接サーバーの、より小さなバッファー・サイズに対応します。

### 接語 (clitic)

構文的には分離して機能するが、音声学的には別のワードに接続するワード。接語は、結合されるワードとは、接続して書かれたり、分離して書かれたりします。接語の一般的な例としては、英語における縮小語の終わりの部分が含まれます(*wouldn't* または *you're*)。

### ソフト・エラー・ページ (soft error page)

クライアントが要求したページを HTTP サーバーが戻すことができなかった場合に、詳細に問題を説明し、問題の内容を示す戻りコード付きのヘッダーのみで構成される応答の代わりに、これらのページを戻すように HTTP サーバーを構成する特別なページ。

## タ

### 注釈 (annotation)

テキストのスパンに関する情報。例えば、注釈は、テキストのスパンが会社名を表すことを指示することもあり得ます。UIMA では、注釈は、特別な種類のフィーチャー構造です。

### データ・ソース (data source)

文書を検索できるデータ・リポジトリ。Web、リレーショナルおよび非リレーショナル・データベース、およびコンテンツ・マネージメント・システムなど。

### データ・ソース・タイプ (data source type)

データ・アクセス用のプロトコルに応じたデータ・ソースのグループ。

### テーマ抽出 (theme extraction)

概念抽出のタイプの 1 つで、テキスト文書内の重要な語彙項目を自動的に認識して、文書のテーマやトピックを抽出するもの。『概念抽出』も参照。

### ディスカバラー (discoverer)

クローラー機能の 1 つで、クローラーが情報検索に使用できるデータソースを判別する機能。

### テキスト・ベースのスコアリング (text-based scoring)

照会内の語に対する文書の適合度を表す整数値を、文書に割り当てるプロセス。整数値が大きいほど、照会への一致が緊密であることを表す。『動的ランキング』も参照。

### テキスト分析 (text analysis)

コレクションのデータの検索性を高めるために、テキストから意味やその他の情報を抽出するプロセス。

### テキスト分析エンジン (text analysis engine)

テキスト内のコンテキストおよびセマンティック・コンテンツを検索および表すことに関与するソフトウェア・コンポーネント。

### デキュー (dequeue)

キューから項目を除去すること。

### トークナイザー (tokenizer)

テキストをスキャンし、一続きの文字をトークンとして認識できる場合に、それを判別するテキスト・セグメンテーション・プログラム。

### トークン (token)

エンタープライズ・サーチによって索引付けされる基本テキスト単位。トークンは、言語内のワードにすることもできますし、索引付けに適切な、他のテキスト単位にすることもできます。

### 同義語辞書 (synonym dictionary)

ユーザーがコレクションを検索するときに、その照会用語の同義語を検索できるようにする辞書。

### 動的要約 (dynamic summarization)

要約タイプの 1 つ。検索語が強調表示され、検索結果には検索している文書の概念を最もよく表す句が含まれる。『静的要約』と対比。

### 動的ランキング (dynamic ranking)

照会の条件を検索中の文書に関して分析し、結果のランクを決定するランキングのタイプ。『テキスト・ベースのスコアリング』も参照。『静的ランキング』と対比。

## ナ

### 認証局 (certificate authority)

証明書を発行する組織。CA は、電子トランザクションに含まれるエンティティ(個人または法人)を認証します。認証局は、情報を交換する二者が、実際に、主張しているとおりの者であることを保証します。

## ハ

### パーサー (parser)

エンタープライズ・サーチ・データ・ストアに追加された文書を解釈するプログラム。パーサーは、文書から情報を抽出し、索引付け、検索、取得の準備を行う。

### ハイブリッド検索 (hybrid search)

ブール検索とフリー・テキスト検索を組み合わせたもの。

### パラメトリック検索 (parametric search)

指定された範囲内の数値または属性 (日付、整数、その他のデータ・タイプなど) を含むオブジェクトを探す検索のタイプ。

### ブール検索 (boolean search)

1 つ以上の検索語が、AND、NOT、OR などの演算子を使って結合された検索。

### ファジー検索 (fuzzy search)

検索語にスペルが似た語を戻す検索。

### フィーチャー・パス (feature path)

UIMA フィーチャー構造内のフィーチャーの値にアクセスするのに使用されるパス。

### フィーチャー構造 (feature structure)

テキスト分析の結果を表す、基礎となるデータ構造。フィーチャー構造は、属性値構造です。各フィーチャー構造は、タイプに属します。すべてのタイプは、Java クラスと非常に類似している、有効なフィーチャーまたは属性の指定されたセットを持ちます。

**フィールド (field)**

レコードの最小識別可能パート。

**フィールド検索 (fielded search)**

特定のフィールドに限定された照会。

**フェデレーション (federation)**

命名システムを結合する処理。それによって、集合システムは、命名システムをスパンする複合名を処理できます。

**フェデレーテッド・サーチ (federated search)**

複数の検索サービスにわたって検索を可能にし、検索結果の統合化されたりリストを戻す検索機能。

**付加記号 (diacritic)**

アクセント記号、またはドイツ語のウムラウトなど、ワードの発音を変更したり、同じワードの間で区別したりする場合に、文字に追加されるマーク。

**フリー・テキスト検索 (free text search)**

フリー・フォーム・テキストで検索語を表現した検索。

**フルテキスト索引 (full text index)**

データ項目を参照し、照会用語を含む文書を検索で迅速に見つけられるようにするデータ構造。

**プレース (place)**

個人やグループが共同作業するために出会うポータルで、可視になる仮想ロケーション。ポータルでは、各ユーザーは、専用作業のための個人用プレースを持ち、個人やグループは、さまざまな共有スペースへのアクセス権を持ちます。そこは、パブリック・プレースにも、制限されたプレースにもなり得ます。『Lotus QuickPlace プレース』をも参照。

**プロキシ・サーバー (proxy server)**

アプリケーションまたは Web サーバーがホストする HTTP Web 要求に対する中継として動作するサーバー。プロキシ・サーバーは、エンタープライズのコンテンツ・サーバーの代理として動作します。

**分析エンジン (analysis engine)**

『テキスト分析エンジン』を参照。

**分析結果 (analysis results)**

アナレーターが生成する情報。分析結果は、検索したい情報に相当するもので、共通分析構造と呼ばれるデータ構造に書き込まれます。

**分類構造 (taxonomy)**

類似性に基づいてオブジェクトをグループに分類したもの。エンタープライズ・サーチでは、分類構造によってデータはカテゴリーとサブカテゴリーに編成される。『カテゴリー・ツリー』も参照。

**マ****マスク文字 (masking character)**

検索語の先頭、中間、および末尾にある任意の文字を表す文字。マスク文字は通常、索引で語の異形を検索するために使用される。『ワイルドカード文字』も参照。

**末尾の文字 (trailing character)**

ワードにおける最後の位置を保持する文字。

**見出し語 (lemma)**

ワードの正規書式。見出し語は、チェコ語など、大きく語形変化する言語では重要です。

**見出し語化 (lemmatization)**

辞書に提示されているワードに代わって見出し語を検索する処理。見出し語化は、ステミングとは異なります。ステミングは、アルゴリズム的であり、ある言語のワードをリストする辞書を操作しません。

**文字の正規化 (character normalization)**

大文字化や発音区別符号など、文字の異体形式が共通形式に合わせられる処理。

**モデル・ベースのカテゴリ (model-based category)**

類似した内容を含む文書で文書を索引付けおよび検索するために、文書の主題の判別に使われる事前定義用語を使用した分類構造。

**モニター担当者 (monitor)**

コレクション・レベルのプロセスを監視する権限を持つエンタープライズ・サーチ・ユーザー。

**ヤ****ユーザー・エージェント (user agent)**

Web をブラウズし、アクセスしたサイトに自身の情報を残すアプリケーション。エンタープライズ・サーチで、Web クローラーはユーザー・エージェント。

**有効範囲 (scope)**

検索要求の範囲を定義するのに使用される関連 URI のグループ。

**用語加重検索 (weighted term search)**

一定の用語が重視される照会。

**要約 (summarization)**

文書の内容を簡潔に記述する文を検索結果に組み込むプロセス。『動的的要約』と『静的要約』も参照。

**ラ****ライブラリー (library)**

他のオブジェクトにディレクトリーとしてサービスを提供するシステム・オブジェクト。『Domino Document Manager ライブラリー』を参照。

**ランキング (ranking)**

照会検索結果の各文書に整数値を割り当てるプロセス。検索結果における文書の順序は、照会への適合度に基づいて決まる。ランクが高いほど、緊密な一致を表す。『動的ランキング』と『静的ランキング』も参照。

**ランキング調整クラス (boost class)**

検索結果内の文書の相対的ランクに影響を与えることのできる仕様。

**ランキング調整ワード (boost word)**

検索結果内の文書の関連ランクに影響を与えることのできるワード。照会処



理中に、ワードに事前定義したスコアに従って、ランキング調整ワードを含む文書の重要度を調整することも可能です。

#### **リモート・フェデレーター (Remote Federator)**

検索可能なオブジェクトのセットをフェデレートするサーバー・フェデレーター。

#### **リンク分析 (link analysis)**

文書間のハイパーリンクの分析に基づき、コレクション内のどのページがユーザーにとって重要なかを判別するための方法。

#### **リンクをたどらないディレクティブ (no-follow directive)**

Web ページで検出されるリンクをたどらないように、ロボット (Web クローラーなど) を指示する、それらのページ内のディレクティブ。

#### **ルーム (room)**

ユーザーが、他の人々が読む文書を作成し、他の人々からのコメントに回答し、プロジェクトの状況と期限を検討することができるようにするプログラム。ユーザーは、同じルームにいる他の人々とチャットすることもできます。『Lotus QuickPlace ルーム』をも参照。

#### **ルール・ベースのカテゴリ (rule-based category)**

どの文書が、どのカテゴリと関連付けられるかを指定する規則によって作成されるカテゴリ。例えば、一定の語を含む、または含まない文書や一定の URI パターンに一致する文書を、特定のカテゴリと関連付ける規則を定義する。

#### **ローカル・フェデレーター (Local Federator)**

検索可能なオブジェクトのセット全体をフェデレートするクライアント・フェデレーター。

#### **ロボット排除プロトコル (Robots Exclusion Protocol)**

サイトのある部分をロボットが訪問しないように、Web サイト管理者が、訪問するロボットに指示できるようにするプロトコル。

## **ワ**

#### **ワイルドカード文字 (wildcard character)**

検索語の先頭、中間、または末尾にある任意の文字を表す文字。

## **D**

#### **Document Object Model (DOM)**

XML ファイルなど、構造化文書を、プログラマチックにアクセスおよび更新できるオブジェクトのツリーとして表示するシステム。

#### **Domino Document Manager ライブラリー (Domino Document Manager library)**

Domino Document Manager に対するエントリー・ポイントである Domino Document Manager データベース。

#### **Domino Document Manager キャビネット (Domino Document Manager cabinet)**

文書を編成するのに使用される Domino Document Manager データベース。キャビネットが Domino データベースを保持します。

#### **Domino Internet Inter-ORB Protocol (DIIOP)**

サーバー上で稼働し、Domino Object Request Broker と連動して、Notes

Javaクラスを使用して作成される Java アプレットと Domino サーバーとの間の通信を可能にするサーバー・タスク。ブラウザ・ユーザーおよび Domino サーバーは、DIIOP を使用して通信し、オブジェクト・データを交換します。

## I

### **ID 管理 (identity management)**

セキュア保管でユーザー信用証明情報を暗号化する能力。

### **IP アドレス (IP address)**

ネットワーク上のホストを識別する固有 32 ビット・アドレス。

## J

### **Java Database Connectivity (JDBC)**

Java プラットフォームと広範なデータベースとの間のデータベース依存接続の業界標準。JDBC インターフェースは、SQL ベースのデータベース・アクセスに対する呼び出しレベルの API を提供します。

### **Java 仮想マシン (JVM) (Java virtual machine (JVM))**

コンパイル済み Java コード (アプレットおよびアプリケーション) を実行するプロセッサのソフトウェア・インプリメンテーション。

### **JavaScript**

ブラウザおよび Web サーバーで使用される Web スクリプト言語。

### **JavaServer Pages (JSP)**

動的コンテンツをクライアントに戻すために、Java コードを動的に Web ページ (HTML ファイル) に組み込むことができるようにし、そのページのサービスが提供されるとき実行するサーバー・スクリプト・テクノロジー。

## L

### **Lightweight Directory Access Protocol (LDAP)**

X.500 モデルをサポートするディレクトリーへのアクセス権を提供する TCP/IP を使用し、より複雑な X.500 ディレクトリー・アクセス・プロトコルのリソース要件の影響を受けないオープン・プロトコル。

### **Lotus QuickPlace プレース (Lotus QuickPlace place)**

地理的に分散した参加者が、構造化されてセキュアなワークスペースにおいて、プロジェクトで共同作業し、オンラインで通信することができるようになる、Lotus QuickPlace によって提供される Web の場。

### **Lotus QuickPlace ルーム (Lotus QuickPlace room)**

共通の興味、および集会的作業の必要を共有する、許可されたメンバーに制限された Lotus QuickPlace プレースのパーティション化領域。

## M

### **MIME タイプ (MIME type)**

インターネット全体にわたって転送されるオブジェクトのタイプを識別するためのインターネット標準。

## N

### **N-gram セグメンテーション (n-gram segmentation)**

Unicode ベースの空白文字セグメンテーションのように、ワードを区切るのにブランク・スペースを使用するよりはむしろ、所与の数の文字の重複シーケンスを単一文字とみなす分析方法。

### **Notes リモート・プロシージャー・コール (NRPC) (Notes remote procedure call (NRPC))**

すべての Notes-to-Notes 通信に使用される Lotus Notes のアーキテクチャ一層。

## S

### **Secure Sockets Layer (SSL)**

通信プライバシーを提供するセキュリティー・プロトコル。

## U

### **Unicode ベースの空白スペース・セグメンテーション (Unicode-based white space segmentation)**

トークンと区切り文字を区別するために Unicode 文字プロパティを使用するトークン化の方式。

### **Universal Resource Name (URN)**

特定の構文に準拠する文字の短ストリングからなるインターネット・プロトコル・エレメント。そのストリングは、リソースを参照するのに使用できる名前またはアドレスから構成されます。

### **URI (Uniform Resource Identifier)**

抽象的または物理的リソースを識別するコンパクトな文字ストリング。

### **URL (Uniform Resource Locator)**

コンピューター上、またはインターネットなどのネットワーク内の情報リソースを表す一続きの文字。この一続きの文字には、その情報リソースへのアクセスに使用されるプロトコルの省略名、およびその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る。

### **Unstructured Information Management Architecture (UIMA)**

非構造化データの分析用システムをインプリメントするフレームワークを定義する IBM アーキテクチャー。

## W

### **Web クローラー (Web crawler)**

Web 文書を検索したり、文書内のリンクをたどったりすることによって、Web を探索するロボット・ソフトウェアのクラス。

## X

### **XML パス言語 (XPath) (XML Path Language (XPath))**

一意的に識別されたり、またはソースの XML 文書の一部にアドレッシングしたりする言語。XPath はまた、ストリング、数値、およびブール演算子の操作のための基本機能を提供します。



---

## WebSphere Information Integration に関する情報の入手

WebSphere Information Integration 製品に関する情報は、Web で入手できます。

WebSphere Information Integration に関する情報は、Web の [www.ibm.com/software/data/integration/db2ii/](http://www.ibm.com/software/data/integration/db2ii/) にあります。このサイトには、次の最新情報が入っています。

- 製品資料
- 製品ダウンロード
- フィックスパック
- リリース情報とその他のサポート文書
- WebSphere Information Integration に関する新情報
- ホワイト・ペーパーや IBM Redbooks™ などの Web リソースへのリンク
- ニュースグループやユーザー・グループへのリンク
- WebSphere Information Integration 製品のオンライン・インフォメーション・センターへのリンク
- 資料の注文方法

製品資料を入手するには、以下のようにします。

1. Web の [www.ibm.com/software/data/integration/db2ii/](http://www.ibm.com/software/data/integration/db2ii/) にアクセスします。
2. ドロップダウン・リストから製品 (例えば、WebSphere Information Integrator OmniFind Edition) を選択します。
3. ページ左側の「Support」リンクをクリックします。
4. 「Learn」セクションで必要なリンクを選択します。選択した製品のインフォメーション・センターがある場合は、インフォメーション・センターのリンクを選択できます。 344 ページの図 2 の例を参照してください。

## Learn

- **Product documentation and manuals** (2 items)
- **Redbooks** (1 item)
- **V8.2 Documentation and release notes**

## Information Center

Provides fast, online centralized access to product information.

- [1.0](#)

図 2. WebSphere Information Integration Support Web サイトにおける製品資料へのリンクの例



---

## IBM と連絡を取る

お客様の国または地域で IBM に連絡する方法については、Web の [www.ibm.com/planetwide](http://www.ibm.com/planetwide) にある「IBM Directory of Worldwide Contacts」にアクセスしてください。



---

## 商標

ここでは、IBM の商標と、特定の IBM 以外の商標をリストします。

IBM の商標に関する情報については、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

以下は、それぞれ各社の商標または登録商標です。

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。

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。

Intel、Intel Inside (ロゴ)、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。

UNIX は、The Open Group の米国およびその他の国における登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における商標です。

他の会社名、製品名およびサービス名等はそれぞれ各社の商標です。



---

## 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、米国以外の国においては本書で述べる製品、サービス、またはプログラムを提供し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。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、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。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、様々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。



それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生した創作物には、次のように、著作権表示を入れていただく必要があります。

Outside In (®) Viewer Technology, ©1992-2004 Stellent, Chicago, IL., Inc. All Rights Reserved.

IBM XSLT Processor Licensed Materials - Property of IBM ©Copyright IBM Corp., 1999-2004. All Rights Reserved.



# 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]

アクセシビリティ 329  
アクセス制御  
  現在の検証 222  
  コレクションに対する使用不可化 230  
  説明 215  
  文書レベル・セキュリティ 221, 235  
  ポータル検索エンジン・サポート 235  
  ユーザー・プロファイル 224  
  Lotus Domino での要件 228  
  Windows ファイル・システムでの要件 225  
アクティブ Web サイトのモニター 264, 266  
アダプター, エンタープライズ・サーチ用の 239  
アナテーター 129  
アプリケーション ID 220  
アラート  
  クロール済み文書数 282  
  検索応答時間 282  
  コレクション・レベル 258, 282  
  サーバー上のフリー・スペース 284  
  索引付けされた文書 282  
  索引の制限 258  
  システム・レベル 284  
  説明 281  
  E メールでの受信 287  
  E メール・オプション 282, 284  
  SMTP サーバー構成 286  
アンカー・テキスト分析  
  グローバル分析 218  
  コレクション・セキュリティ 218  
  説明 211  
  文書の索引作成 219  
イベント・パブリッシング  
  説明 48  
  AIX オペレーティング・システムでのセットアップ 55  
  DB2 クローラー構成 50, 53  
  Linux オペレーティング・システムでのセットアップ 55

イベント・パブリッシング (続き)  
  Solaris オペレーティング環境でのセットアップ 55  
  Windows でのセットアップ 56  
エラー・メッセージ  
  ログ・ファイルの表示 289  
  E メールでの受信 285, 287  
  SMTP サーバー構成 286  
エンタープライズ・サーチ  
  クローラー・サーバー 4  
  索引サーバー 7  
  パーサー 5  
エンタープライズ・アプリケーション  
  ESAdmin アプリケーション 216  
  ESPACServer.ear ファイル 99  
  ESSearchAdapter.ear ファイル 239  
  ESSearchApplication アプリケーション 216  
  ESSearchServer アプリケーション 216  
エンタープライズ・サーチ  
  概要 1  
  管理コンソール 9  
  管理役割 213  
  クローラー・サーバー 37  
  検索アプリケーション 12  
  検索サーバー 8, 165  
  コマンド 297  
  コレクション・レベルのセキュリティ 217  
  コンポーネント 3  
  サーバーの開始 251  
  サーバーの停止 251, 253  
  索引サーバー 149  
  システム・ファイルのリストア 294  
  セキュリティ 211  
  セッション ID 297  
  データ・フロー・ダイアグラム 13  
  パーサー 115  
  バックアップ 292  
  バックアップ・スクリプト 291  
  バックアップ・データからのリストア 293  
  文書レベル・セキュリティ 221  
  ポート番号構成 26  
  戻りコード 297  
  モニター 257  
  リストア・スクリプト 291  
  ログ・ファイル 281  
  API 11  
  URI フォーマット 105  
  WebSphere Portal との統合 233

エンタープライズ・サーチ管理者  
  説明 213  
  変更, シングル・サーバー上のパスワード 20  
  変更, マルチサーバー上のパスワード 22  
  役割の構成 214  
エンタープライズ・サーチ資料の 検索 327  
エンタープライズ・サーチのバックアップ 291, 292  
エンタープライズ・サーチのリストア 291  
応答時間履歴のモニター 275  
オペレーター  
  説明 213  
  役割の構成 214

## [カ行]

開始  
  エンタープライズ・サーチ・サーバー 251  
  クローラー・サーバー 262  
  検索サーバー 275  
  索引作成 273  
  サンプル検索アプリケーション 203  
  データ・リスナー 277  
  パーサー 272  
  マイグレーション・ウィザード 243, 245  
外部クローラー  
  構成 103  
  データ・リスナー API 11  
  データ・リスナー・アプリケーション 103  
外部ソース  
  アプリケーション ID セキュリティ 220  
  検索 154  
  検索アプリケーションとの関連付け 209  
  構成 207  
  説明 207  
カスタム・テキスト分析  
  共通分析構造のマッピング 134, 135  
  索引への分析結果のマッピング 134  
  説明 129  
  テキスト分析エンジン 130, 131  
  JDBC 表への分析結果のマッピング 135

- カスタム・テキスト分析 (続き)
  - XML エLEMENTのマッピング 132
- カテゴリー
  - カテゴリー化タイプ 121
  - カテゴリー・ツリー 120
  - 検索 116
  - 作成 122
  - サブカテゴリーのネスティング 120
    - 説明 116
    - モデル・ベース 120, 121
    - ルール・ベース 117, 121
    - URI フォーマット 105
    - WebSphere Portal からのマイグレーション 243, 245
  - カテゴリー化タイプ
    - 選択 31, 121
    - モデル・ベース 120
    - ルール・ベース 117
  - カテゴリー規則
    - 構成 122
    - 文書コンテンツ 117, 122
    - URI パターン 117, 122
  - カテゴリー・ツリー
    - 説明 120
    - WebSphere Portal からのマイグレーション 243, 245
- 韓国語
  - 複合語分析 136
  - N-gram セグメンテーション 138
- 管理コンソール
  - インターフェース 15
    - 説明 9
    - タスクの概要 15
    - ログイン 19
- 管理者パスワード
  - 変更、シングル・サーバー上 20
  - 変更、マルチサーバー上 22
- 管理役割
  - エンタープライズ・サーチ管理者 213, 214
  - オペレーター 213, 214
  - 構成 214
  - コレクション管理者 213, 214
    - 説明 213
    - モニター担当者 213, 214
- キーワード、クイック・リンク 173, 174
- 共通分析構造
  - 索引へのマッピング 134
    - 説明 129
    - JDBC 表へのマッピング 135
    - XML エLEMENTのマッピング 132
- クイック・リンク
  - URI フォーマット 105
- クイック・リンク
  - 検索 173
  - 作成 174
- クイック・リンク (続き)
  - 説明 173
- グローバル Web クロール・スペース 95
- グローバル分析
  - アンカー・テキスト分析 211, 218
    - 説明 7
    - 重複文書の検出 211, 218
- クローラー
  - 外部サポート 11, 103
  - クローラーのプロパティの編集 41
  - クロール・スペースの編集 41
  - 構成の概要 37
  - 異なるクローラー・タイプの組み合わせ 37
  - 削除 42
  - 作成 40
  - システム状況 262
  - 初期値 40
  - スケジューリング 37
    - 説明 4
  - データ・リスナー・アプリケーション 103
    - デフォルトのマイグレーション設定 248
    - プラグイン 104
    - 文書レベル・セキュリティ 221
    - 文書レベル・セキュリティを使用可能にする 37
    - ベースとなる値 37
    - モニター 262
    - Content Edition 43, 44
    - DB2 48
    - DB2 Content Manager 57
    - Domino Document Manager 61
    - Exchange Server 63, 64
    - NNTP 64
    - Notes 65, 67
    - QuickPlace 75
    - UNIX ファイル・システム 79
    - URI フォーマット 105
    - Web 80
    - WebSphere Portal 98, 99, 100
    - Windows ファイル・システム 101
  - クローラーのプロパティ
    - 説明 4
    - 編集 41
  - クローラー履歴レポート
    - サイト・レポート 267
    - 作成 267
    - 説明 264
    - HTTP 戻りコード・レポート 267
  - クローラー・サーバー
    - 開始 251, 262
    - 停止 253, 262
- クローラー・タイプ
  - 単一コレクション内に組み合わせる 37
    - ベースとなる値 37
- クローラー・プラグイン 104
  - 「クロール」ページ、説明 15
- クロール速度のモニター 266
- クロール・スペース
  - アラート 282
    - 説明 4
    - 編集 41
    - Web クローラーの構成 82
- 言語 サポート
  - カスタム・テキスト分析 129
  - セマンティック検索 129
- 言語サポート
  - ストップワード辞書 169
  - セマンティック検索 137
  - 同義語辞書 167
  - ネイティブ XML 検索 137
  - ランキング調整ワード辞書 178
- 検索
  - 外部ソース 154
  - カテゴリー 116
  - クイック・リンク 173
  - コレクション 154
    - HTML 文書 127, 128
    - XML 文書 124, 125, 132
- 検索 結果
  - ワイルドカード文字 156
- 検索 サーバー
  - 開始 275
  - 「検索」ページ、説明 15
- 検索アプリケーション
  - アプリケーション ID 220
  - 外部ソースとの関連付け 209
  - カスタム 189
  - コレクションとの関連付け 190
  - コレクション・レベルのセキュリティ 220
  - サンプル 189, 191
    - 説明 12
- 検索応答時間
  - アラート 282
  - モニター 275
- 検索オプション
  - 検索結果の 154
    - パラメトリック検索 154
    - フィールド検索 154
    - フリー・テキスト検索 154
    - 文書コンテンツ 154
- 検索および索引 API 11, 189
- 検索キャッシュ
  - 構成 166
  - 説明 166

検索結果  
グループ化 161, 162  
サマリー 172  
サマリーのカスタマイズ 172  
縮小表示 161, 162  
静的ランキング 178  
説明 177  
テキスト・ベースのスコアリング 177  
動的要約 171  
動的ランキング 177  
ランキング 182  
ランキング調整クラスの構成 183,  
185, 186, 187  
ワイルドカード文字展開 158  
URI パターン構成 181  
検索結果の縮小表示  
構成 162  
検索結果のランキング  
静的 178  
説明 177  
テキスト・ベースのスコアリング 177  
動的 177  
ランキング調整クラス 183, 185, 186,  
187  
ランキング調整ワード辞書 180  
URI パターン 181, 182  
検索サーバー  
応答時間履歴 275  
開始 251  
検索キャッシュ 166  
高頻度の照会 275  
最新の照会 275  
システム状況 275  
ストップワード辞書 169  
ストップワード辞書の関連付け 170,  
171  
説明 8, 165  
停止 253, 275  
同義語辞書 167, 168  
同義語辞書の関連付け 169  
モニター 275  
ランキング調整ワード辞書 178  
ランキング調整ワード辞書の関連付け  
180  
検索センター、WebSphere Portal の  
エンタープライズ・サーチ用のアダプ  
ター 233, 239  
説明 233  
登録ポートレット 233, 241  
検索ブラウズ・ポートレットの構成 238  
検索ポートレットのデプロイメント 236  
検証、現行の信用証明情報の 222, 225,  
228  
高頻度の照会のモニター 275  
「構文解析」ページ、説明 15

コマンド、エンタープライズ・サーチ  
297  
コレクション  
アプリケーション ID セキュリティー  
220  
アンカー・テキスト・セキュリティ  
218  
検索 154  
検索アプリケーションとの関連付け  
190  
検索サーバー 165  
構文解析 115  
「コレクション」ビューを使用した作  
成 31  
コレクション・ウィザードを使用した  
作成 30  
サイズの見積もり 31  
削除 34  
作成方法 29  
システム状況 259  
セキュリティ 217  
説明 3  
重複文書セキュリティ 218  
デフォルトのマイグレーション設定  
248  
ドラフト 30  
フェデレーション 29  
文書レベルのアクセス制御の迂回 230  
編集 33  
モニター 259  
リソースの見積もり 258  
MigrationWizard.log ファイル 245  
WebSphere Portal からのマイグレーシ  
ョン 243, 245  
コレクション ID、構文規則 31  
「コレクション」ビュー  
コレクションの作成 31  
説明 15  
コレクション管理者  
説明 213  
役割の構成 214  
コレクション・ウィザード 30  
コレクション・レベルのセキュリティ  
アプリケーション ID 220  
アンカー・テキスト分析 218  
使用可能化 31  
説明 211, 217  
重複文書の検出 218

## [サ行]

サーバー・モード、Content Edition リポ  
ジトリー 44  
最近クローラされた URL のモニター  
264  
最小再クローラ間隔 87

最新の照会のモニター 275  
最大再クローラ間隔 87  
サイト履歴レポート  
作成 267  
説明 264  
索引  
アラート 282  
アンカー・テキスト (anchor text) 219  
キューからの削除 274  
縮小表示した URI 153, 161, 162  
スケジューリング 150  
スケジュールの変更 151  
スケジュールを使用可能にする 151,  
273  
スケジュールを使用不可にする 151,  
273  
説明 7, 149  
同時作成 152  
並行作成 152  
モニター 273, 274  
有効範囲 153, 159  
ワイルドカード文字 153, 156, 158  
URI の除去 153, 163  
URI フォーマット 105  
「索引」ページ、説明 15  
索引キュー 274  
索引サーバー  
開始 251  
停止 253  
索引作成  
開始 273  
システム状況 274  
停止 273, 274  
同時 152  
並行 152  
索引スケジュールを使用可能にする 151  
索引スケジュールを使用不可にする 151  
索引付けしないディレクティブ  
構成 97  
説明 97  
索引の再編成 7  
索引の再編成 149  
スケジューリング 150, 151  
説明 7, 149  
索引の並行作成 152  
索引のリフレッシュ 7, 149  
スケジューリング 150, 151  
説明 7, 149  
削除  
キューから索引を 274  
クローラー 42  
コレクション 34  
作成  
クイック・リンク 174  
クローラー 40  
コレクション 30, 31

作成 (続き)

有効範囲 160  
ルール・ベースのカテゴリ 122  
HTML 検索フィールド 128  
Web クローラー・レポート 267  
XML 検索フィールド 125

サマリー

カスタマイズ 172  
動的 171

サンプル検索アプリケーション

アクセス 203  
開始 203  
検索機能 189, 191  
セキュリティの使用可能化 204  
セキュリティの使用不可化 216  
説明 12, 191  
デフォルトのデプロイメント 201  
config.properties ファイル 192, 201  
WebSphere グローバル・セキュリティ  
ー 204

「システム」ビュー、説明 15

システム状況

クローラー 262  
検索サーバー 275  
コレクション 259  
索引作成 274  
パーサー 272  
Web クローラー 264

システムのリストア 291, 293, 294

システム・バックアップ 291, 292

システム・リソース

チェック 258  
見積もり 258

システム・リソースのチェック 258

システム・リソースの見積もり 258

縮小表示した URI

構成 162  
説明 161

縮小表示した検索結果

説明 161

照会の検証 222

除去、索引から URI を 163

除去された 文書

ログ・ファイルの構成 278

除去された文書

説明 278  
に関するレポート 279

資料 327

スクリプト

esbackup.bat 292, 294  
esbackup.sh 292, 294  
escrem.sh 58  
escrem.vbs 60  
escrib2.sh 55  
escrib2.vbs 56  
escrnote.sh 68

スクリプト (続き)

escrnote.vbs 70  
escrvbr.sh 45  
escrvbr.vbs 46  
esrestore.bat 293, 294  
esrestore.sh 293, 294  
startccl 293

スケジューリング

クローラー 37  
索引作成 150, 151

ストップワード辞書

コレクションとの関連付け 171  
システムへの追加 170  
説明 169

スレッド

パーサー 136  
Web クローラー 265

スレッドの詳細のモニター 264

静的ランキング

コレクションの使用可能化 31  
説明 178  
マイグレーションされたコレクション  
における 245

セキュリティ

アクセス制御 215  
アンカー・テキスト分析 218  
エンタープライズ・アプリケーション  
に対する使用不可化 216  
管理役割 214

クローラー・プラグイン 104

検索アプリケーション ID 220

コレクションの使用可能化 31, 211

コレクション・レベル 217, 230

サンプル検索アプリケーション 204

説明 211

重複文書の検出 218

認証 215, 216

文書レベル 221, 222, 225, 230

文書レベルのアクセス制御の迂回 230

ユーザー・プロファイル 224

ID 管理 225

Lotus Domino 文書 228

WebSphere グローバル・セキュリティ  
ー 216

Windows ドメイン 225

「セキュリティ」ビュー、説明 15

セキュリティ・トークン

クローラー構成 221  
コレクションに対する使用不可化 230  
文書レベル・セキュリティ 221,  
235  
ポータル検索エンジン処理 235

セッション ID、エンタープライズ・サー  
チ 297

セマンティック検索 129, 132, 137

ソフト・エラー・ページ、Web クローラ  
ー 88

## 【タ行】

高い再呼び出し照会

説明 183

デフォルトのランキング調整要因 187

タスクの概要、管理コンソール 15

中国語

改行文字の解析 138

N-gram セグメンテーション 138

重複文書の検出

グローバル分析 218

セキュリティの使用可能化 218

説明 211

ランキング調整クラスの構成 183

直接モード、Content Edition リポジトリ  
ー 44

データ・ソース・タイプ

エンタープライズ・サーチでサポート  
されている 2

外部サポート 2, 11

リレーショナル・データベース 48

Content Edition リポジトリ 43, 44

DB2 Content Manager 項目タイプ 57

DB2 データベース 48

Domino Document Manager データベ  
ース 61

Exchange Server のパブリック・フォ  
ルダー 63

NNTP ニュースグループ 64

Notes データベース 65, 67

QuickPlace データベース 75

UNIX ファイル・システム 79

Web サイト 80

WebSphere Portal サイト 98

Windows ファイル・システム 101

データ・フロー、エンタープライズ・サー  
チ・システム 13

データ・リスナー

構成 103

再始動 103, 277

モニター 277

データ・リスナー API 11

停止

エンタープライズ・サーチ・サー  
バー 251, 253

クローラー・サーバー 262

検索 サーバー 275

索引作成 273, 274

パーサー 272

ディスカバリー 4

テキスト 処理

アノテーター 129

共通分析構造 129



テキスト処理  
 テキスト分析エンジン 129  
テキスト分析 エンジン  
 説明 129  
 分析結果のマッピング 134, 135  
テキスト分析 (text analysis)  
 共通分析構造 134, 135  
 テキスト分析エンジン 130, 131  
 XML エLEMENTのマッピング 132  
テキスト分析エンジン  
 コレクションとの関連付け 131  
 システムへの追加 130  
 XML エLEMENTのマッピング 132  
テキスト・ベースのスコアリング 177  
デフォルトの検索アプリケーション 201  
デプロイメント  
 エンタープライズ・サーチ用のアダプ  
 ター 239  
 検索ブラウザ・ポートレット 238  
 検索ポートレット 236  
 登録ポートレット 241  
 ESPACServer.ear ファイル 99  
 ESSearchAdapterPortlet.war ファイル  
 241  
 ESSearchAdapter.ear ファイル 239  
 ESSearchPortlet.war ファイル 236  
同義語辞書  
 コレクションとの関連付け 169  
 システムへの追加 168  
 説明 167  
同時索引作成 152  
動的要約 171  
動的ランキング 177  
トラステッド・サーバー構成 228

## [ナ行]

日本語  
 改行文字の解析 138  
 N-gram セグメンテーション 138  
認証  
 エンタープライズ・アプリケーション  
 対する使用不可化 216  
 説明 215  
ネイティブ XML 検索 137

## [ハ行]

パーサー  
 改行文字 138  
 開始 272  
 空白文字 138  
 構文解析、文書タイプの 139  
 システム状況 272  
 スレッド 136

パーサー (続き)  
 説明 5, 115  
 データ分析タスク 5  
 停止 272  
 ネイティブ XML 検索 137  
 パーサー・サービスの文書タイプ  
 140, 142  
 複合語分析 136  
 モニター 272  
 N-gram セグメンテーション 138  
 Stellent セッション用の文書タイプ  
 143, 145  
パーサー・サーバー  
 開始 251  
 停止 253  
パーサー・サーバー  
 スレッド構成 136  
パスワード、エンタープライズ・サーチ管  
理者 20, 22  
パスワード保護された Web サイト 90  
 フォーム・ベースの認証 91  
 HTTP 基本認証 90  
バックアップ・スクリプト  
 稼働中 292  
 実行中 294  
 説明 291  
パラメトリック検索 154  
低い再呼び出し照会  
 説明 183  
 デフォルトのランキング調整要因 187  
表示  
 ログ・ファイル 289  
 URI の詳細 260  
ファイアウォール、Exchange Server 文書  
のクローラ 64  
ファイル拡張子  
 コレクション・パーサーによってサポ  
 ートされている 140, 142  
 除外、Web クロール・スペースから  
 82  
 Stellent セッションでサポートされて  
 いる 143, 145  
フィールド、ランキング調整クラスへのマ  
ッピング 185  
フィールド検索 154  
フェデレーテッド・コレクション 29  
フォーム・ベースの認証 90, 91  
複合語、構文解析 136  
プラグイン、クローラーの 104  
フリー・スペースのアラート 284  
フリー・テキスト検索 154  
プロキシー・サーバー 92  
文書 トラッキング  
 ログ・ファイルの構成 278  
文書コンテンツ・オプション 154

文書サマリー  
 カスタマイズ 172  
 プロパティの 編集 172  
文書タイプ  
 構文解析 139  
 パーサー・サービスの 140, 142  
 Stellent セッション用 143, 145  
文書トラッキング  
 使用可能化 278  
 使用不可化 278  
 説明 278  
 レポート 279  
文書の重要度  
 コレクションの使用可能化 31  
 静的 178  
 マイグレーションされたコレクション  
 における 245  
 ランキング調整クラス 183, 187  
 ランキング調整ワード辞書 180  
 URI パターン 181, 182  
文書レベル・セキュリティ  
 クローラー構成 37  
 クローラー・プラグイン 104  
 現行の信用証明情報の検証 222  
 索引付きアクセス制御 221  
 セキュリティー・トークン 221  
 説明 211, 221  
 ポータル検索エンジン・サポート 235  
 ユーザー・プロファイル 224  
 リアルタイムの検証 222  
 ID 管理 225  
 Lotus Domino 文書 228  
 Lotus Domino 文書の 228  
 Windows ファイル・システムでの  
 225  
分類、WebSphere Portal からのマイグレ  
ーション 243, 245  
分類管理ポートレット 243  
編集  
 クローラーのプロパティ 41  
 クロール・スペース 41  
 検索アプリケーションのプロパティ  
 192, 201  
 コレクション 33  
 データ・リスナー・アプリケーション  
 103  
ポータル検索エンジン  
 セキュリティー・トークンを渡す 235  
 説明 233  
 文書レベル・セキュリティ 235  
ポート番号、エンタープライズ・サーチ  
26  
ポートレット  
 エンタープライズ・サーチ 233  
 検索 236  
 検索ブラウザ 238

ポートレット (続き)  
サポートされる WebSphere Portal のバージョン 233  
説明 233  
登録 241  
分類管理ポートレット 243  
ESSearchAdapterPortlet.war ファイル 241

## [マ行]

マイグレーション  
コレクション 245  
モデル・ベースの分類 243  
ルール・ベースの分類 245  
マイグレーション・ウィザード  
開始 243, 245  
コレクション 245  
説明 243  
デフォルトのクローラー設定 248  
デフォルトのコレクション設定 248  
モデル・ベースの分類 243  
ルール・ベースの分類 245  
ログ・ファイル 249  
マッピング  
共通分析構造への XML エLEMENTの 132  
共通分析構造を 索引へ 134  
ランキング調整クラスに対するフィールドの 185  
HTML 検索フィールド 128  
JDBC 表へ 共通分析構造を 135  
XML 検索フィールド 125  
モデル・ベースのカテゴリ  
カテゴリ化タイプの選択 121  
説明 120  
モデル・ベースの分類、WebSphere Portal からのマイグレーション 243  
戻りコード、エンタープライズ・サーチ 297  
モニター  
エンタープライズ・サーチ 257  
応答時間履歴 275  
クローラー 262  
検索サーバー 275  
高頻度の照会 275  
コレクション 259  
最新の照会 275  
除去された文書 279  
データ・リスナー 277  
パーサー 272  
ログ・ファイル 289  
URI の詳細 260  
Web クローラー 264  
Web クローラーのクロール速度 266

モニター (続き)  
Web クローラーのスレッドの詳細 265  
Web クローラー・アクティブ・サイト 266  
「モニター」ビュー、説明 15  
モニター担当者  
説明 213  
役割の構成 214

## [ヤ行]

ユーザー・エージェント 81  
ユーザー・プロファイル  
構成 225  
説明 224  
有効範囲  
検索 159  
作成 160  
説明 159  
URI パターン 159, 160  
URI フォーマット 105

## [ラ行]

ランキング調整クラス  
構成 185, 186  
説明 183  
高い再呼び出し照会 183, 187  
重複文書の検出 183  
デフォルト値 187  
低い再呼び出し照会 183, 187  
フィールドのマップ 185  
ランキング調整要因  
ランキング調整クラスの 185, 187  
ランキング調整クラスの構成 183, 186  
ランキング調整ワード辞書の 178  
URI パターン 181, 182  
ランキング調整ワード辞書  
コレクションとの関連付け 180  
システムへの追加 180  
説明 178  
リストア  
システム・ファイルを新規のサーバーに 294  
バックアップ・データからの 293  
リストア・スクリプト  
実行中 293, 294  
説明 291  
リンクをたどらないディレクティブ  
構成 97  
説明 97  
ルール・ベースのカテゴリ  
カテゴリ化タイプの選択 121

ルール・ベースのカテゴリ (続き)  
作成 122  
説明 117  
ルール・ベースの分類、WebSphere Portal からのマイグレーション 243  
ルール・ベースの分類、WebSphere Portal からのマイグレーション 245  
ローカル・ユーザー・セキュリティ、QuickPlace クローラー 77  
「ログ」ページ、説明 15  
ログイン、管理コンソールへの 19  
ログ・ファイル  
文書トラッキングに対する 278  
ログ・ファイル  
最大サイズ 285  
重大度レベル 285  
循環 285  
説明 281  
デフォルトの場所 281  
表示 289  
フィルター 289  
マイグレーション・ウィザード 249  
モニター 289  
E メール・オプション 287  
SMTP サーバー構成 286

## [ワ行]

ワイルドカード文字  
索引展開 156, 158  
照会展開 156, 158

## A

AIX オペレーティング・システム  
イベント・パブリッシング構成 55  
Content Edition クローラーの 構成 45  
DB2 Content Manager クローラー構成 58  
Domino Document Manager クローラー構成 68  
Notes クローラー構成 68  
QuickPlace クローラー構成 68  
API  
検索および索引 11, 189  
説明 11  
データ・リスナー 11

## B

bos.iocep.rte モジュール 73

## C

CCLServer\_date.log ファイル 26  
config.properties ファイル  
プロパティの説明 192  
編集 201  
Content Edition クローラー  
構成 43  
サーバー・モード 44  
直接モード 44  
AIX オペレーティング・システムでの  
セットアップ 45  
Linux オペレーティング・システムでの  
セットアップ 45  
Solaris オペレーティング環境でのセ  
ットアップ 45  
URI フォーマット 105  
Windows でのセットアップ 46  
cookies.ini ファイル  
構成 94  
説明 93  
フォーマット 93  
Cookie、Web クロールの  
構成 94  
フォーマット 93  
crawl.rules ファイル 95

## D

DB2 Content Manager クローラー  
構成 57  
AIX オペレーティング・システムでの  
セットアップ 58  
Linux オペレーティング・システムで  
のセットアップ 58  
Solaris オペレーティング環境でのセ  
ットアップ 58  
URI フォーマット 105  
Windows でのセットアップ 60  
DB2 Content Manager の Java コネクタ  
58, 60  
DB2 クローラー  
イベント・パブリッシング 48  
イベント・パブリッシング構成 55,  
56  
構成 48  
URI フォーマット 105  
WebSphere II Event Publisher Edition  
構成 50  
WebSphere MQ インストール 55  
WebSphere MQ 構成 53  
Windows での WebSphere MQ イン  
ストール 56  
DIIOP プロトコル、クローラー構成 72  
Directory Assistance 構成 78

Domino Document Manager クローラー  
構成 61  
AIX オペレーティング・システムでの  
セットアップ 68  
DIIOP プロトコル構成 72  
IOCP 構成 73  
Linux オペレーティング・システムで  
のセットアップ 68  
NRPC プロトコル 68, 70  
Solaris オペレーティング環境でのセ  
ットアップ 68  
URI フォーマット 105  
Windows でのセットアップ 70  
Domino ユーザー構成、QuickPlace クロ  
ーラー 77

## E

E メール通知  
アラート 287  
メッセージ 287  
SMTP サーバー構成 286  
EAR ファイル  
ESAdmin アプリケーション 216  
ESPACServer.ear 99  
ESSearchAdapter.ear 239  
ESSearchApplication アプリケーション  
216  
ESSearchServer アプリケーション 216  
ESAdmin アプリケーション  
セキュリティの使用不可化 216  
ログイン 19  
esadmin コマンド 297  
esbackup.bat スクリプト 292, 294  
esbackup.sh スクリプト 292, 294  
eschangepw スクリプト 20, 22  
escrcm.sh スクリプト 58  
escrcm.vbs スクリプト 60  
escrdb2.sh スクリプト 55  
escrdb2.vbs スクリプト 56  
escrnote.sh スクリプト 68  
escrnote.vbs スクリプト 70  
escrvbr.sh スクリプト 45  
escrvbr.vbs スクリプト 46  
ESPACServer.ear ファイル 99  
esrestore.bat スクリプト 293, 294  
esrestore.sh スクリプト 293, 294  
ESSearchAdapterPortlet.war ファイル 241  
ESSearchAdapter.ear ファイル 239  
ESSearchApplication アプリケーション  
再始動 201, 204  
セキュリティの使用可能化 204  
セキュリティの使用不可化 216  
config.properties ファイル 192, 201  
ESSearchPortlet.war ファイル 236

ESSearchServer アプリケーション  
セキュリティの使用不可化 216  
es.cfg ファイル 20, 22  
ES\_INSTALL\_ROOT、説明 20, 22  
ES\_NODE\_ROOT、説明 20, 22  
es\_special\_field.default\_field 187  
es\_special\_field.regular\_text フィールド  
187  
Exchange Server クローラー  
構成 63  
セキュア文書 64  
URI フォーマット 105

## F

followindex.rules ファイル  
構成 97  
説明 97

## G

global.rules ファイル 95

## H

HTML 検索フィールド  
エレメントのマッピング 127, 128  
作成 128  
説明 127  
HTML 文書、検索 127  
HTTP 基本認証 90  
HTTP プロキシ・サーバー 92  
HTTP 戻りコード  
Web クローラーが受け取った 267  
Web クローラー・レポート 267

## I

ID 管理  
構成 225  
ユーザー・プロファイル 224  
IOCP、クローラー構成 73  
I/O 完了ポート・モジュール、クローラー  
構成 73

## J

JavaScript サポート、Web クローラーの  
82  
JDBC 外部ソース  
構成 207  
削除 207  
編集 207  
JDBC ドライバー 207

## L

LDAP 外部ソース  
構成 207  
削除 207  
編集 207

Linux オペレーティング・システム  
Content Edition クローラーの構成 45  
Linux オペレーティング・システム  
イベント・パブリッシング構成 55  
DB2 Content Manager クローラー構成  
58  
Domino Document Manager クローラー  
構成 68  
Notes クローラー構成 68  
QuickPlace クローラー構成 68  
Solaris オペレーティング環境  
イベント・パブリッシング構成 55  
Lotus Domino ドメイン 228  
Lotus Domino トラステッド・サーバー  
228

## M

MIME タイプ、Web クロール・スペース  
に入れる 82

## N

NNTP クローラー、構成 64  
Notes クローラー  
検証、現行の信用証明情報の 228  
構成 65  
使用のヒント 67  
フィールド・マッピング規則 67  
文書レベル・セキュリティ構成 228  
AIX オペレーティング・システムでの  
セットアップ 68  
DIHOP プロトコル構成 72  
IOCP 構成 73  
Linux オペレーティング・システムで  
のセットアップ 68  
Lotus Domino トラステッド・サーバ  
ー 228  
NRPC プロトコル 68, 70  
Solaris オペレーティング環境でのセッ  
トアップ 68  
URI フォーマット 105  
Windows でのセットアップ 70  
NRPC プロトコル、クローラー構成 68,  
70  
N-gram セグメンテーション 138

## P

parserTypes.cfg ファイル 140  
PDF 資料 327

## Q

QuickPlace クローラー  
構成 75  
ローカル・ユーザー・セキュリティ  
77  
AIX オペレーティング・システムでの  
セットアップ 68  
DIHOP プロトコル構成 72  
Directory Assistance 構成 78  
Domino ユーザー構成 77  
IOCP 構成 73  
Linux オペレーティング・システムで  
のセットアップ 68  
NRPC プロトコル 68, 70  
Solaris オペレーティング環境でのセッ  
トアップ 68  
URI フォーマット 105  
Windows でのセットアップ 70

## R

Robots Exclusion プロトコル 81  
robots.txt ファイル 81  
runtime-generic.properties ファイル 172

## S

SIAPI (検索および索引 API) 11, 189  
SMTP サーバー構成 286  
Solaris オペレーティング環境  
Content Edition クローラーの 構成  
45  
DB2 Content Manager クローラー構成  
58  
Domino Document Manager クローラー  
構成 68  
Notes クローラー構成 68  
QuickPlace クローラー構成 68  
startccl スクリプト 293  
Stellent セッション  
構文解析、文書タイプの 139  
デフォルトの文書タイプ 145  
文書タイプの関連付け 143  
stellentypes.cfg ファイル 143  
stellent.properties ファイル 143  
synonyms.xml ファイル 243

## T

titles.xml ファイル 243  
treenodes.xml ファイル 243

## U

UIMA  
共通分析構造 134, 135  
コレクションとの関連付け 131  
索引への分析結果のマッピング 134  
システムへのテキスト分析エンジンの  
追加 130  
説明 129  
JDBC 表への分析結果の マッピング  
135  
XML エLEMENTのマッピング 132  
UNIX ファイル・システム・クローラー  
構成 79  
URI フォーマット 105  
URI  
エンタープライズ・サーチにおけるフ  
ォーマット 105  
カテゴリ規則 117, 122  
クイック・リンク 173, 174  
検索結果の縮小表示 161, 162  
索引からの除去 163  
詳細情報の表示 260  
静的スコアの操作 181, 182  
有効範囲 159, 160  
URI の詳細  
除去された文書 279  
モニター 260  
URL にできる限り早くアクセスする 88  
URL にできる限り早く再アクセスする  
88  
URL パスの深さ 82

## V

vbr\_access\_services.jar ファイル 45, 46

## W

Web クローラー  
アクティブ・サイト 264, 266  
開始 URL 82, 88  
グローバル・クロール・スペース 95  
クローラー履歴 264  
クロール規則 82  
クロール速度 266  
クロール・スペース の制限 82  
構成 80  
最近クロールされた URL 264  
再クロール間隔 87

- Web クローラー (続き)
  - サイト履歴 264
  - 索引付けしないディレクティブ 97
  - システム状況 264
  - スレッドの詳細 264, 265
  - ソフト・エラー・ページ 88
  - パスワード保護された Web サイト 90, 91
  - プロキシ・サーバー 92
  - モニター 264
  - ユーザー・エージェント 81
  - リンクをたどらないディレクティブ 97
  - レポートの作成 267
  - Cookie 93
  - Cookie 構成 94
  - Cookie フォーマット 93
  - followindex.rules ファイル 97
  - JavaScript サポート 82
  - URL 状況 264
  - URL にできる限り早くアクセスする 88
- Web クローラーの 開始 URL 88
- Web クローラーの IP アドレス規則 82
- Web クローラーのアドレス規則 82
- Web クローラーの開始 URL 82
- Web クローラーの再クロール間隔 87
- Web クローラーの接頭部規則 82
- Web クローラーのドメイン規則 82
- Web クロールの Cookie 説明 93
- Web クロール・スペースの制限 82
- WebSphere II Event Publisher Edition、DB2 クローラー構成 50
- WebSphere II OmniFind Edition 329
  - アクセシビリティ 329
  - 概要 1
  - 管理コンソール 9
  - クローラー・サーバー 4
  - 検索アプリケーション 12
  - 検索サーバー 8
  - コマンド 297
  - コンポーネント 3
  - 索引サーバー 7
  - セッション ID 297
  - データ・フロー・ダイアグラム 13
  - パーサー 5
  - 変更、シングル・サーバー上のパスワード 20
  - 変更、マルチサーバー上のパスワード 22
  - ポート番号構成 26
  - 戻りコード 297
  - API 11
  - ID 管理 225
  - WebSphere Portal との統合 233
- WebSphere MQ、クローラー・サーバー構成 55, 56
- WebSphere MQ、DB2 クローラー構成 53
- WebSphere Portal
  - エンタープライズ・サーチとの統合 233
  - カテゴリ・ツリーのマイグレーション 243
  - 検索ブラウザ・ポートレット 238
  - 検索ポートレットのデプロイメント 236
  - コレクションのマイグレーション 243
  - デフォルトのマイグレーション設定 248
  - 分類管理ポートレット 243
  - 分類のマイグレーション 243
  - モデル・ベースの分類 243
- WebSphere Portal クローラー
  - エンタープライズ・アプリケーションのデプロイメント 99
  - 構成 98
  - サイト URL のコピー 100
  - ESPACServer.ear ファイルのデプロイ 99
  - URI フォーマット 105
- WebSphere Portal 検索センター
  - エンタープライズ・サーチ用のアダプター 233, 239
  - 説明 233
  - 登録ポートレット 233, 241
- WebSphere グローバル・セキュリティ検索アプリケーションのプロパティ 201
  - 使用不可化 216
- Windows オペレーティング・システム
  - イベント・パブリッシング構成 56
  - クローラー構成 70
  - Content Edition クローラーの 構成 46
  - DB2 Content Manager クローラー構成 60
- Windows ドメイン 225
- Windows ファイル・システム・クローラー
  - 構成 101
  - 文書レベル・セキュリティ構成 225
  - URI フォーマット 105
- WpsMigratorLog.log ファイル 249
- X**
- XML エlement
  - 共通分析構造へのマッピング 132
  - 検索 124, 132
  - 検索フィールドへの マッピング 125
- XML 検索フィールド
  - エレメントのマッピング 124, 125, 132
  - 作成 125
  - 説明 124, 132
- XML 照会構文、ネイティブ 137
- XML フラグメント、ネイティブ XML 検索 137
- XML 文書
  - 検索 125
  - ネイティブ XML 検索 137
- XPath、ネイティブ XML 検索 137







**IBM**

Printed in Japan



SD88-6374-02



**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