

IBM WebSphere Information Integrator
OmniFind Edition



管理企业搜索

版本 8.3

IBM WebSphere Information Integrator
OmniFind Edition



管理企业搜索

版本 8.3

在使用本资料及其支持的产品之前，请务必阅读『声明』中的一般信息。

本文档包含 IBM 的专利信息。它是根据许可协议提供的，并受版权法保护。本出版物包含的信息不包括任何产品保证，且本手册提供的任何声明不应作如此解释。

可以在线方式或通过您当地的 IBM 代表订购 IBM 出版物：

- 要以在线方式订购出版物，可访问“IBM 出版物中心”（IBM Publications Center），网址为 www.ibm.com/shop/publications/order。
- 要查找您当地的 IBM 代表，可访问“IBM 全球联系人目录”（IBM Directory of Worldwide Contacts），网址为 www.ibm.com/planetwide。

当您发送信息给 IBM 后，即授予 IBM 非专有权，IBM 可以它认为合适的任何方式使用或分发此信息，而无须对您承担任何责任。

© Copyright International Business Machines Corporation 2004, 2005. All rights reserved.

目录

关于这些主题	vii
这些主题的读者	vii
什么是企业搜索?	1
企业搜索支持的数据源类型	2
企业搜索组件概述	2
企业搜索搜寻器	3
企业搜索解析器	4
企业搜索索引	5
企业搜索的搜索服务器	6
企业搜索管理控制台	7
监视企业搜索系统	8
企业搜索日志文件	8
定制企业搜索	8
企业搜索的样本搜索应用程序	9
企业搜索数据流	10
企业搜索管理概述	13
登录到管理控制台	16
更改单服务器配置中的企业搜索管理员密码	17
更改多服务器配置中的企业搜索管理员密码	18
更改企业搜索系统的端口号	21
企业搜索集合	25
使用集合向导来创建集合	25
使用“集合”视图来创建集合	26
编辑集合	27
删除集合	28
企业搜索搜寻器管理	31
创建搜寻器	33
编辑搜寻器属性	34
编辑搜寻空间	34
删除搜寻器	35
Content Edition 搜寻器	35
以服务器方式访问 WebSphere II Content Edition 存储库	36
以直接方式访问 WebSphere II Content Edition 存 储库	36
在 UNIX 上为 WebSphere II Content Edition 配置 搜寻器服务器	37
在 Windows 上为 WebSphere II Content Edition 配置搜寻器服务器	38
DB2 搜寻器	39
为 DB2 搜寻器配置 WebSphere Information Integrator Event Publisher Edition	40
配置 WebSphere MQ for DB2 搜寻器	42
将 UNIX 上的搜寻器服务器配置为使用事件发布功 能	44
将 Windows 上的搜寻器服务器配置为使用事件发 布功能	45

DB2 Content Manager 搜寻器	45
在 UNIX 上为 DB2 Content Manager 配置搜寻器 服务器	46
在 Windows 上为 DB2 Content Manager 配置搜寻 器服务器	48
Domino Document Manager 搜寻器	49
Exchange Server 搜寻器	50
验证对安全 Exchange Server 文档的访问权	51
NNTP 搜寻器	51
Notes 搜寻器	52
Lotus Domino 数据库搜寻技巧	53
在 UNIX 上配置搜寻器服务器以搜寻 Lotus Domino 源	54
配置 Windows 上的搜寻器服务器以搜寻 Lotus Domino 源	56
配置使用 DIIOP 协议的服务器	57
在 AIX 上配置 I/O 完成端口以搜寻 Lotus Domino 源	58
QuickPlace 搜寻器	60
配置 QuickPlace 服务器以使用“本地用户”安全性 在 QuickPlace 服务器上配置“目录辅助”	61
UNIX 文件系统搜寻器	62
Web 搜寻器	63
用户代理配置	64
支持 JavaScript	64
限制 Web 搜寻空间的规则	65
使用 Web 搜寻器测试 URL 连接	68
Web 搜寻器中的重新搜寻时间间隔设置	68
有关使用 Web 搜寻器来访问 URL 的选项	69
Web 搜寻器处理软错误页面的方式	69
支持搜寻安全 Web 站点	70
由代理服务器提供服务的 Web 站点	72
Cookie 管理	73
全局 Web 搜寻空间配置	74
No-follow 和 no-index 伪指令	75
覆盖 Web 页面中的 no-follow 和 no-index 伪指令 WebSphere Portal 搜寻器	76
为 WebSphere Portal 搜寻器部署企业应用程序	77
从 WebSphere Portal 复制要搜寻的 URL	78
Windows 文件系统搜寻器	79
配置对数据侦听器应用程序的支持	80
定制搜寻器插件	81
企业搜索索引中的 URI 格式	82
企业搜索解析器管理	89
使用类别	89
基于规则类别	91
基于模型类别	93
类别树	93
选择分类类型	94
配置类别	94

使用 XML 搜索字段	96	基于文本的打分	137
XML 搜索字段	96	静态排名	137
将 XML 元素映射至搜索字段	96	定制升降字典	138
使用 HTML 搜索字段	98	将升降字典添加到系统中	139
HTML 搜索字段	98	使升降字典与集合相关联	139
将 HTML 元数据元素映射至搜索字段	99	基于 URI 模式的文档排名	140
定制文本处理	99	影响与 URI 模式匹配的文档的分数	140
将文本分析引擎添加至系统	100	基于升降类的文档排名	141
使文本分析引擎与集合相关联	101	将字段映射至升降类	143
将 XML 元素映射至公共分析结构	102	配置升降类的升降因子	143
将公共分析结构映射至索引	103	缺省升降类值	144
将公共分析结构映射至 JDBC 表	104	企业搜索的搜索应用程序 147	
配置解析器服务的线程	105	将搜索应用程序与集合相关联	147
启用复合项高级分析功能	105	样本搜索应用程序功能	148
启用对本机 XML 搜索的支持	106	样本搜索应用程序属性	149
中文、日语和韩国语文档的语言分析	106	编辑样本搜索应用程序属性	157
N-gram 分段	107	访问样本搜索应用程序	158
从非 ASCII 字符范围中除去换行符	107	为样本搜索应用程序启用安全性	158
与集合解析器和 Stellent 会话相关联的文档类型	107	企业搜索外部源 161	
将文档类型与集合解析器关联	108	将外部源添加到系统中	161
缺省集合解析器服务规则	109	使搜索应用程序与外部源相关联	162
将文档类型与 Stellent 会话关联	110	企业搜索安全性 165	
Stellent 会话的缺省解析规则	112	管理角色	166
企业搜索索引管理 115		配置管理用户	167
安排索引构建时间表	116	认证与访问控制	167
更改索引时间表	117	对 WebSphere Application Server 中的企业应用程序	
启用和禁用索引时间表	117	禁用安全性	168
配置并发索引构建	118	集合级安全性	169
影响索引的可搜索视图的选项	118	重复文档分析	169
用于搜索文档的建立索引选项	119	锚点文本分析	170
查询中的通配符	121	对指向被禁止文档的链接中的锚点文本建立索引	170
配置通配符查询选项	122	通过搜索应用程序标识实现的安全性	171
作用域	123	文档级安全性	172
配置作用域	123	通过存储安全性令牌进行验证	172
折叠的 URI	124	在查询处理期间验证当前凭证	173
折叠搜索结果中的 URI	125	对 Windows 文件系统文档强制实施文档级安全性	175
从索引中除去 URI	126	对 Lotus Domino 文档强制实施文档级安全性	177
企业搜索的搜索服务器管理 127		禁用文档级安全性	178
搜索高速缓存	127	企业搜索与 WebSphere Portal 的集成 181	
配置搜索高速缓存	128	通过 Portal Search Engine 实现的文档级安全性	182
定制同义词字典	128	部署“搜索”portlet	183
将同义词字典添加到系统中	129	为企业搜索配置 WebSphere Portal“搜索和浏览	
使同义词字典与集合相关联	130	”portlet	184
定制无用词典	130	为搜索中心安装企业搜索适配器	185
将无用词典添加到系统中	131	为搜索中心安装企业搜索注册 portlet	186
使无用词典与集合相关联	131	从 WebSphere Portal 迁移到企业搜索 189	
动态摘要	132	从 WebSphere Portal 迁移基于模型的分类法	189
在管理控制台中定制文档摘要	132	从 WebSphere Portal 迁移集合	190
通过编辑属性定制文档摘要	133	迁移的集合设置	192
使用快速链接	133	迁移向导日志文件	193
快速链接	133		
配置快速链接	134		
企业搜索中的文档排名 137			

启动和停止企业搜索服务器	195	接收关于已记录消息的电子邮件	221
启动企业搜索服务器	195	查看日志文件	223
停止企业搜索服务器	197	备份和复原企业搜索系统	225
监视企业搜索活动	199	备份企业搜索系统	225
估算集中的文档数	199	复原企业搜索系统	226
检查系统资源的可用性	200	将企业搜索系统文件复原至新服务器	227
监视集合	201	I 企业搜索命令、返回码和会话标识	229
查看关于 URI 的详细信息	201	企业搜索文档	255
监视搜寻器	203	WebSphere II OmniFind Edition 辅助	
查看有关 Web 搜寻器活动的详细信息	204	功能选项	257
Web 搜寻器线程详细信息	205	企业搜索术语词汇表	259
Web 搜寻器活动站点	205	访问关于 WebSphere Information	
Web 搜寻器搜寻速率	206	Integration 的信息	269
创建 Web 搜寻器报告	206	提供对文档的意见	271
Web 搜寻器 HTTP 返回码	208	与 IBM 联系	273
监视解析器	210	商标	275
监视集合的索引活动	211	声明	277
监视企业搜索索引队列	212	索引	279
监视搜索服务器	213		
监视数据侦听器	214		
文档跟踪	214		
配置用于进行文档跟踪的日志文件	215		
查看关于废弃的文档的报告	215		
企业搜索日志文件和警报	217		
警报	217		
配置集合级警报	218		
配置系统级警报	219		
配置日志文件	219		
配置 SMTP 服务器信息	221		

关于这些主题

在管理 IBM® WebSphere® Information Integrator OmniFind™ Edition V8.3 系统时，请使用本信息。

WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) 提供了称为企业搜索的技术。企业搜索组件是在您安装 WebSphere II OmniFind Edition 产品时安装的。在 WebSphere II OmniFind Edition 文档中，除了引用安装路径和产品打包标签的情况以外，都使用术语企业搜索。

企业搜索管理文档阐述了下列主题：

- 企业搜索概念、组件和功能的简介
- 有关如何创建、监视和管理集合的指示信息
- 有关如何解析数据以及如何定制解析活动以优化搜索和检索的信息
- 有关如何对解析的数据建立索引以及如何管理建立索引活动的信息
- 有关搜索服务器如何查找搜索结果、对搜索结果进行排名以及返回搜索结果的信息
- 有关使定制搜索应用程序与集合相关联的信息
- 企业搜索提供的不同安全级别的概述
- 有关企业搜索如何与 IBM WebSphere Portal 集成的信息
- 有关如何将 WebSphere Portal 分类法和集合迁移到企业搜索类别树和集合的信息
- 有关备份和复原系统的信息
- 有关创建和查看日志文件的信息
- 有关如何使用企业搜索命令和解释所返回的内容的信息

这些主题的读者

本信息是为负责创建、监视和管理企业搜索集合的系统管理员和系统操作员提供的。

使用本信息来创建集合、选择集合内容以及配置选项以使内容可供搜索。并且，使用本信息来监视集合和系统活动、将用户登记为企业搜索管理员以及使集合和外部可搜索源与搜索应用程序相关联。

为了有效地使用本信息，您必须熟悉 Web 应用程序并有所要搜索的数据源的使用经验。

什么是企业搜索？

企业搜索系统提供了强大的功能来通过单个查询搜索任意数目的结构化和非结构化数据源。企业搜索系统提供了快速查询响应时间和统一的排名结果集，该结果集使您能够方便地找到所需的信息。

随 IBM WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) 一起安装的企业搜索组件从整个企业中收集信息并使该信息可供搜索。通过在 Web 浏览器中输入查询，您可以同时搜索本地和远程数据库、协作系统、内容管理系统、文件系统以及内部和外部 Web 站点。

企业搜索系统设计为能够与现有系统无缝集成，该系统处理从各种来源收集数据所需的繁琐工作并对数据建立索引以便可以进行快速检索。通过对数据应用语言分析和其它类型的分析，企业搜索可以提供高度相关的搜索结果。您不需要了解不同的界面就可以搜索各种存储库类型。

可以添加支持以允许搜索您不想包括在企业搜索索引中的数据源。借助企业搜索的联合搜索功能，可以在搜索已建立了索引的数据源时搜索这些外部源。

搜索质量

为了确保用户能够找到他们所查找的信息，WebSphere II OmniFind Edition 支持 IBM Unstructured Information Management Architecture (UIMA)。UIMA 是一个开放式的框架，它定义了公共的标准文本分析接口。借助扩展文本分析功能，企业搜索可以标识通常隐藏在无结构文本背后的概念、潜在含义、关系、事实以及其它相关数据。可以使用分析期间抽取的信息来提高搜索结果的质量，这些信息也可用来提高其它应用程序（例如商业智能应用程序和数据挖掘应用程序）的质量。

安全性

安全性是企业搜索的组成元素。只有有权管理系统的用户才能管理系统。借助 IBM WebSphere Application Server 提供的安全性机制，您可以配置管理角色并控制哪些用户有权访问各种管理功能。

您还可以指定选项以便在收集数据时使安全性令牌与该数据相关联。如果搜索应用程序启用了安全性，则可以使用这些令牌（它们与文档一起存储在索引中）来强制实施访问控制并确保只有拥有正确凭证的用户才能查询数据和查看搜索结果。

对于某些类型的数据源来说，可以配置选项以便在查询处理期间使用当前访问控制来验证用户的登录凭证。这一层附加的安全措施确保实时地针对本机数据源验证用户的特权。此选项可以保护用户凭证以避免它在文档及其安全性令牌建立索引后发生更改的情况。

相关概念

企业搜索安全性

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

企业搜索支持的数据源类型

提供了用于搜索各种数据源类型的预定义支持。

安装 IBM WebSphere Information Integrator OmniFind Edition 后，就可以开始从下列类型的数据源收集数据了：

- IBM DB2[®] Content Manager 项类型（文档、资源和项）
- IBM DB2 通用数据库[™]数据库
- IBM Domino[®] Document Manager（旧称 Domino.Doc[®]）数据库
- IBM Lotus Notes[®] 数据库
- IBM Lotus[®] QuickPlace[®] 数据库
- Microsoft[®] Exchange Server 公共文件夹
- Microsoft Windows[®] 文件系统
- 网络新闻传输协议（NNTP）新闻组
- UNIX[®] 文件系统
- IBM WebSphere Information Integrator Content Edition 存储库：Documentum、FileNet Panagon Content Services、FileNet P8 Content Manager、Hummingbird[®] Document Management（DM）、OpenText Livelink 和 Portal Document Manager（PDM）项类
- IBM DB2 通用数据库 z/OS[®] 版、IBM Informix[®]、Microsoft SQL Server、Oracle 和 Sybase 数据库的 IBM WebSphere Information Integrator 昵称表
- IBM WebSphere Portal 站点
- 因特网上或内部网中的 Web 站点

还可以添加支持以搜索下列类型的外部源，而不必将这些来源中的文档添加到企业搜索索引中：

- Java[™] 数据库连接（JDBC）数据库表（仅限于 IBM DB2 通用数据库（DB2 UDB）和 Oracle 数据库）。将为 JDBC 数据库中的每个表创建一个独立的外部源。
- 轻量级目录访问协议（LDAP）服务器。将为每台 LDAP 服务器创建一个外部源。

要了解有关受支持数据源类型的最新信息，请参阅 WebSphere Information Integrator OmniFind Edition Web 站点。

相关概念

企业搜索外部源

企业搜索组件概述

企业搜索组件从整个企业中收集数据；对信息进行分析、解析和分类；并且创建用户可以搜索的索引。

企业搜索集合表示用户可以通过单个查询来搜索的一组来源。在创建集合时，您指定要包括哪些来源，并配置有关用户可以如何搜索已建立了索引的数据的选项。

您可以创建多个集合，每个集合都可以包含来自各种数据源的数据。例如，可以创建包含来自 IBM DB2 通用数据库、IBM Lotus Notes，和 IBM DB2 Content Manager 数据库的文档的集合。当用户搜索此集合时，搜索结果有可能包含来自每个数据源的文档。

对联合搜索的支持使用户可以使用一个查询搜索多个集合。搜索结果有可能包含来自企业搜索系统中所有集合和外部源的文档。

创建和管理集合涉及下列活动:

收集数据

搜寻器组件持续不断地或者根据指定的时间表来从数据源收集文档。频繁的搜寻确保用户始终能够访问最新的信息。

分析数据

解析器组件从文档中抽取文本，并且对搜寻器所搜寻的每个文档执行语言分析和其它类型的分析。详细的内容分析能够改进搜索结果的质量。

对数据建立索引

索引组件定期地运行，以将关于新的和已更改的文档的信息添加到索引中。索引组件还可以对集合中的文档执行全局分析以提高搜索结果的质量。

搜索数据

搜索组件搜索索引并与搜索应用程序配合工作以处理搜索请求和返回搜索结果。

其它 WebSphere Information Integrator OmniFind Edition 组件使您能够指定安全首选项、监视系统活动并对发生的问题进行故障诊断。该产品还提供了一个能够发挥作用的样本搜索应用程序，您可以将该搜索应用程序用作模板以创建自己的搜索应用程序。

相关概念

企业搜索搜寻器管理

企业搜索解析器管理

企业搜索索引管理

企业搜索的搜索服务器管理

企业搜索搜寻器

企业搜索搜寻器从数据源收集文档，以便可以对那些文档进行分析、建立索引和进行搜索。

WebSphere Information Integrator OmniFind Edition 附带提供的搜寻器组件具有下列功能:

- 在配置搜寻器时，发现进程查找关于可以搜寻的来源的信息，例如 Lotus Notes 数据库中所有视图和文件夹的名称或者 UNIX 服务器上所有文件系统的名称。
- 在选择要搜寻的来源并启动搜寻器之后，搜寻器组件将从这些来源收集数据，以便可以对那些数据进行分析 and 建立索引。

单个集合可以有多个搜寻器，并且每个搜寻器都被设计成从特定类型的数据源收集数据。例如，可以创建 3 个搜寻器以便将来自文件系统、Notes® 数据库和关系数据库的数据组合到同一个集合中。或者，可以创建若干个具有相同类型的搜寻器，然后根据每个搜寻器所搜寻的数据的更改频率来为那些搜寻器设置不同的搜寻时间表。

用于 Web、WebSphere Portal 和网络新闻传输协议 (NNTP) 来源的搜寻器持续运行。在指定所要搜寻的统一资源定位符 (URL) 或 NNTP 新闻组之后，搜寻器将定期地返回以检查新的和已更改的数据。可以手工启动和停止其它类型的搜寻器，也可以设置搜

寻时间表。如果安排搜寻器的时间表，则指定它第一次运行的时间以及它需要以何频率访问数据源以搜寻新的和已更改的文档。

搜寻器属性是一组规则，这些规则控制特定搜寻器在执行搜寻时的行为。例如，您指定规则以控制搜寻器使用系统资源的方式。可供搜寻的来源集合组成了搜寻器的搜寻空间。在创建搜寻器之后，随时可以编辑搜寻器属性以改变搜寻器收集数据的方式。还可以编辑搜寻空间以更改搜寻器时间表、添加新来源或删除您不想再搜索的来源。

相关概念

企业搜索搜寻器管理

相关任务

监视搜寻器

企业搜索解析器

企业搜索解析器对搜寻器收集到的文档进行分析并对它们进行准备以便建立索引。

WebSphere Information Integrator OmniFind Edition 附带提供的解析器组件对文档内容和文档元数据进行分析。它将分析结果存储在数据商店中，以供索引组件访问。解析器执行下列任务：

- 从文档使用的任何格式中抽取文本。例如，解析器从 XML 和 HTML 文档中的标记中抽取文本。通过使用 Stellent for IBM WebSphere Information Integrator OmniFind Edition Outside In Viewer Technology，解析器还可以从二进制格式（例如 Microsoft Word 和 Adobe Acrobat 可移植文档格式（PDF）文档）中抽取文本。
- 检测每个文档的字符集编码。在执行任何语言分析之前，解析器使用此信息来将所有文本转换为 Unicode。
- 检测每个文档的源语言。
- 应用您对集合指定的解析规则。在配置解析器时，您可以配置：

XML 和 HTML 文档的字段映射规则

此选项使用户能够搜索 XML 和 HTML 文档中的结构化内容和非结构化内容。如果将 XML 元素或 HTML 元数据元素映射到企业搜索索引中的搜索字段，则用户可以在查询中指定字段名并搜索 XML 和 HTML 文档的特定部分。（与搜索所有文档内容的自由文本查询相比，搜索特定字段的查询可以提供更精确的搜索结果。）

类别 此选项使用户能够按文档所属的类别来搜索文档。用户还可以在搜索结果中选择类别并且只浏览属于该类别的文档。

在创建集合时，您选择要使用的类别的类型（如果有的话）。如果使用基于*规则*的类别，则将根据您定义的规则来使文档与类别相关联。可以对您创建的企业搜索集合以及从 IBM WebSphere Portal 迁移的集合配置基于规则的类别。

如果使用基于*模型*的类别，则文档与 WebSphere Portal 系统中存在的基于模型的类别相关联。要使用此选项，必须在企业搜索索引服务器上安装 WebSphere Portal。还必须使用 WebSphere Portal 中的分类工具来管理类别。

定制文本分析

应用程序开发者可以创建定制分析程序以执行要搜索的数据的复杂语言分析。您可以将这些程序插入企业搜索系统并使用它们来注释集合的内容。通过对注释进行索引，可以对集合进行语义搜索。

例如，用户既可以根据查询项相互之间的近似性来搜索出现的查询项也可以搜索同一个句子中的查询项，他们还可以搜索查询项之间的关系（如搜索关于 IBM 的销售人员 Smith 而不是 IBM 工程师 Smith 的文档）。

支持 n-gram 分段

要增强使用中文、日语和韩国语编写的文档的可检索性，您可以指定启用 n-gram 分段方法进行词法分析。这种形式的分析不使用空格来定界字。（在创建集合后不能更改分段方法。）

支持使用本机 XML 搜索来搜索 XML 文档

通过搜索 XML 标记，本机 XML 搜索能够提供更精确的搜索结果。例如，查询可以指定某个单词必须出现在特定 XML 元素中。

用于提高或降低字段相对重要程度分数的类

通过将字段映射到升降类，可以影响文档在搜索结果中的排名。例如，可以提高标题字段的分数，从而确保当查询项出现在标题中时，标题包含该项的文档在搜索结果中的排名较高。

- 抽取文本并添加标记以增强数据的可检索性。在此阶段，解析器执行下列任务：
 - 字符规范化，例如，对大写内容和变音标记（如，德语元音变音）进行规范化。
 - 分析段落、句子、词和空格的结构。通过进行语言分析，解析器将复合词分解并指定允许进行字典和同义词查找的标记。

相关概念

使用类别

XML 搜索字段

HTML 搜索字段

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Text analysis included in enterprise search”

相关任务

监视解析器

企业搜索索引

企业搜索索引组件定期地运行以将关于新的和已更改的文档的信息添加到索引中。

为了确保用户始终能够访问他们所搜索的来源中的最新信息，构建索引的过程包括两个阶段：

重组索引

重组索引时将重新构建整个索引，以便索引结构具有最佳的组织结构。建立索引的过程读取由搜寻器收集并由解析器分析的所有数据。

刷新索引

当刷新索引时，将把自从上次重组索引以来搜寻到的信息添加至索引。

在为集合配置索引选项时，可以指定用于进行索引重组和刷新的时间表。重组和刷新索引的频率依赖于系统资源以及正在建立索引的来源是包含静态内容还是包含动态内容。

为了确保新信息的可用性，请安排时间表以便频繁地刷新索引。请定期地安排索引重组，以便合并所有新信息、分析新内容和优化索引性能。

也可以在不安排索引建立过程时间表的情况下启动建立索引的过程。例如，如果您更改了某些解析规则并要让那些更改对搜索应用程序起作用，则可以在重新搜寻并解析数据后启动索引重组，而不是等待索引重组在它的安排时间启动。

为了控制资源使用量，您控制可以同时共享建立索引进程和提交索引构建请求的集合数。并发地构建索引有助于确保对非常大的索引执行的重组不会阻碍对其它索引执行的刷新。索引构建是非常耗费资源的进程，因此，对于大型系统，您必须监视系统负载以调整索引重组和刷新率。

构建索引时，建立索引的过程将执行全局文档分析。在此阶段，将应用算法以便标识重复的文档，以便对文档的链接结构进行分析以及对 Web 文档中的锚点文本（超文本链接中用于描述目标页面的文本）执行特殊处理。

可以为以下索引活动指定选项：

- 要使用户能够指定通配符，可以构建支持以将查询项扩展到索引中，也可以指定在查询处理期间扩展查询项。您的决定是在资源使用和查询响应时间之间所做的权衡。
- 您可以配置作用域。作用域使您能够限制用户在集合中可以查看的内容。例如，您可以创建一个包含技术支持部门中的文档 URI 的作用域和另一个包含人力资源部门中的文档 URI 的作用域。如果搜索应用程序支持此作用域，则用户可以仅搜索和检索该集合的那些子集中的文档。
- 您可以指定一些选项以将搜索结果中带有相同 URI 前缀的结果文档折叠起来。您也可以指定一个组名，这样就能够将带有不同 URI 前缀的文档在搜索结果中折叠到一起。
- 在构建索引之后，可以除去不想让用户搜索的 URI。

相关概念

企业搜索索引管理

查询中的通配符

作用域

折叠的 URI

基于 URI 模式的文档排名

相关任务

安排索引构建时间表

配置并发索引构建

从索引中除去 URI

监视集合的索引活动

监视企业搜索索引队列

企业搜索的搜索服务器

企业搜索的搜索服务器与搜索应用程序配合工作以处理查询、搜索索引和返回搜索结果。

企业搜索的搜索服务器是在您安装 WebSphere Information Integrator OmniFind Edition 时安装的。在为集合配置搜索服务器时，可以指定选项以指示如何搜索该集合：

- 可以配置搜索高速缓存以存放频繁请求的搜索结果。搜索高速缓存可以改进搜索和检索性能。
- 可以指定用来搜索该集合中的文档的缺省语言。
- 如果应用程序开发者创建了定制字典，则可以使那些字典与集合相关联：
 - 当用户查询使用*同义词*字典的集合时，包含查询项同义词的文档将包括在搜索结果中。
 - 当用户查询使用*无用词*字典的集合时，在处理该查询前将从该查询中除去那些无用词。
 - 当用户查询使用*升降字*字典的集合时，包含升降字的文档的重要程度将提高或降低，这取决于该字典中与升降字相关联的升降因子。
- 如果预先确定某些文档与某些查询相关，则可以配置快速链接。快速链接使特定 URI 与特定关键字和短语相关联。如果查询包含快速链接定义中指定的任何关键字或短语，则将在搜索结果中自动返回相关联的 URI。

在多服务器配置中，在集合级别提供了故障保护，而不是仅仅在服务器级别提供此功能。如果一台搜索服务器上的集合由于任何原因而变得不可用，则将自动地把对该集合执行的查询定向到另一台搜索服务器。

相关概念

企业搜索的搜索应用程序

搜索高速缓存

定制同义词字典

定制无用词字典

定制升降字字典

快速链接

相关任务

监视搜索服务器

企业搜索管理控制台

企业搜索管理控制台在浏览器中运行，这表示管理用户可以在任何时候从任何位置访问它。安全性机制确保只有那些有权访问管理功能的用户才能执行此操作。

当您安装 WebSphere Information Integrator OmniFind Edition 时，将把企业搜索管理控制台安装在搜索服务器上。

管理控制台包括可以帮助您执行若干项主要管理任务的向导。例如，集合向导帮助您创建集合并允许以草稿方式保存工作。不同的数据源类型有不同的搜寻器向导，这些向导帮助您选择要让用户能够搜索的来源。

对于其它管理任务，您可以选择要管理的各个项。例如，在编辑集合时，您可以选择索引页面以更更改索引时间表或选择解析页面以修改 XML 文档的解析规则。

相关概念

企业搜索管理概述

管理角色

相关任务

登录到管理控制台

监视企业搜索系统

可以使用企业搜索管理控制台来监视系统活动并根据需要调整操作。

在安装 WebSphere Information Integrator OmniFind Edition 并创建至少一个集合之后，您可以查看每项主要活动（搜寻、解析、建立索引和搜索）的详细统计信息。此信息包括平均响应时间和进度信息，如在特定搜寻或索引构建会话中已搜寻的或建立索引的文档数。

可以停止和启动大部分活动。例如，可以暂停一个活动、更改其配置或对问题进行故障诊断，并在准备好允许活动继续时重新启动处理。

还可以配置警报，这使您能够在所监视的事件发生时接收关于某些被监视活动的电子邮件。例如，如果搜索响应时间超出指定的阈值，您就会接收到警报。

如果从企业搜索系统中废弃了某个文档，您可以跟踪该文档并确定废弃该文档的时间、位置和原因。例如，解析器可能无法解析某个文档，或者管理员从索引中除去了某个文档。

相关概念

监视企业搜索活动

启动和停止企业搜索服务器

企业搜索日志文件

日志文件是为各个集合和为系统级会话创建的。

当为企业搜索集合或为系统配置日志记录选项时，您指定要记录的消息的类型（如错误消息和警告消息）。您还指定要让系统循环使用旧日志文件以便为最新消息腾出空间的频率。您可以选择选项以便在特定消息（包括警报）或所有错误消息发出时接收关于那些消息的电子邮件。

在查看日志文件时，您选择要查看的日志文件（文件名包含文件创建时间以及哪些组件发出了消息的信息）。您还可以指定查看过滤器。例如，可以选择仅查看错误消息或仅查看来自特定企业搜索会话的消息。

相关概念

企业搜索日志文件和警报

警报

企业搜索消息

相关任务

配置日志文件

配置 SMTP 服务器信息

接收关于已记录消息的电子邮件

查看日志文件

定制企业搜索

企业搜索的应用程序编程接口使您能够创建定制搜索应用程序、用于更新集合内容的定制应用程序、用于进行文本分析的定制程序以及定制同义词字典、定制无用词字典和定制升降字字典。

在安装 WebSphere Information Integrator OmniFind Edition 之后，提供了以下 API 系列用于扩展企业搜索集合：

搜索和索引 API (SI-API)

使用此 API 来构建定制搜索应用程序和定制管理界面。

数据侦听器 API

使用此 API 来接收来自外部搜寻器的数据。外部搜寻器可以连接到企业搜索数据侦听器，然后将数据添加到集合中或从集合中除去数据。

搜寻器插件

使用用于 Web 搜寻器和非 Web 搜寻器的 API 来在文档被搜寻时对那些文档添加元数据，或者使安全性令牌关联以强制实施贵单位的业务规则和安全规则。

可以通过将用于进行语言分析的定制程序与企业搜索集集成，来增强信息的可检索性。在将定制文本分析引擎添加到系统后，就可以使那些引擎与集合相关联。在用户查询集合时，他们将受益于定制程序构建到索引中的单词关联。例如，用户可以搜索词条的概念以及之间的关系，而不仅仅是搜索词条本身。

也可以通过集成反映了特定于行业的词条（例如首字母缩写词、简写和词汇表词条）的定制字典来提高信息的可检索性。在将字典添加到系统后，可以使那些字典与集合相关联。当用户查询集合时，他们就会在下列各方面受益：

- 如果查询包含已定义的同义词，包含查询项同义词的文档就会包括在搜索结果中。
- 如果查询包含无用词，就会从该查询中除去那些无用词，这样就不会在搜索结果中返回不相关的文档。
- 如果查询包含升降字，包含那些升降字的文档在搜索结果中的排名就会提高或降低，这取决于字典中与该升降字相关联的升降值。

相关概念

企业搜索的搜索应用程序

定制同义词字典

定制无用词字典

定制升降字字典

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API overview”

“Programming Guide and API Reference for Enterprise Search”中的“Data listener”

“Programming Guide and API Reference for Enterprise Search”中的“Crawler plug-ins”

相关任务

配置对数据侦听器应用程序的支持

企业搜索的样本搜索应用程序

可以按原样使用企业搜索的样本搜索应用程序，也可以将其用作模板来开发定制搜索应用程序。

样本搜索应用程序是在安装 WebSphere Information Integrator OmniFind Edition 时安装的。样本搜索应用程序演示了大部分可用于企业搜索的搜索和检索功能。此应用程序也是一个使您能够搜索企业搜索系统中的所有活动集合和外部源的工作示例。在将新集合和外部源提供给用户之前，可以使用样本应用程序来对它们进行测试。

通过使您能够同时搜索一个或多个集合和外部源，样本搜索应用程序演示了对联合搜索的支持。

如果对集合启用了安全性，并且对系统启用了 WebSphere II OmniFind Edition 标识管理，用户在使用搜索应用程序时就可以创建概要文件。用户概要文件用来存储用户在登录到各个域时需要指定的凭证。

在查询处理期间，搜索进程使用存储的凭证（这些凭证以加密格式存储在 WebSphere II OmniFind Edition 管理的安全密钥库中）来确定用户是否有权搜索安全域。如果没有凭证或者凭证对某个域无效，搜索结果就不会包括该域中的文档。

要了解有关使用样本搜索应用程序的信息，请在使用应用程序时单击帮助。要创建定制搜索应用程序，请使用企业搜索的搜索和索引 API。

相关概念

企业搜索的搜索应用程序

样本搜索应用程序功能

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API overview”

相关任务

访问样本搜索应用程序

编辑样本搜索应用程序属性

为样本搜索应用程序启用安全性

企业搜索数据流

随 WebSphere Information Integrator OmniFind Edition 一起安装的企业搜索组件 紧密地进行交互，以确保数据流经系统。

搜寻器从整个企业内的数据源收集文档。解析器从搜寻到的文档中抽取有用的信息并生成标记，例如，这些标记可以使文档与类别相关联并帮助确定那些文档与搜索请求中的项的相关性。索引存储数据以便进行有效的检索。

通过使用 Web 浏览器和搜索应用程序，用户搜索已建立了索引的集合和外部源。搜索应用程序可以显示可以供用户在浏览器中单击的结果列表，搜索应用程序也可以更为复杂并动态地返回根据不同来源中的信息生成的内容。

例如，目录搜索应用程序可以定制满足搜索请求的产品显示。单个查询可以搜索来自不同类型数据源的文档，例如来自 IBM DB2 Content Manager 和 Lotus Notes 存储库的文档的组合。

管理员确定所要收集的数据以及对其进行搜寻、进行解析、建立索引和进行搜索的方式。通过监视系统活动，管理员还可以进行调整以优化数据吞吐量。

下图显示了企业搜索系统中的信息流。

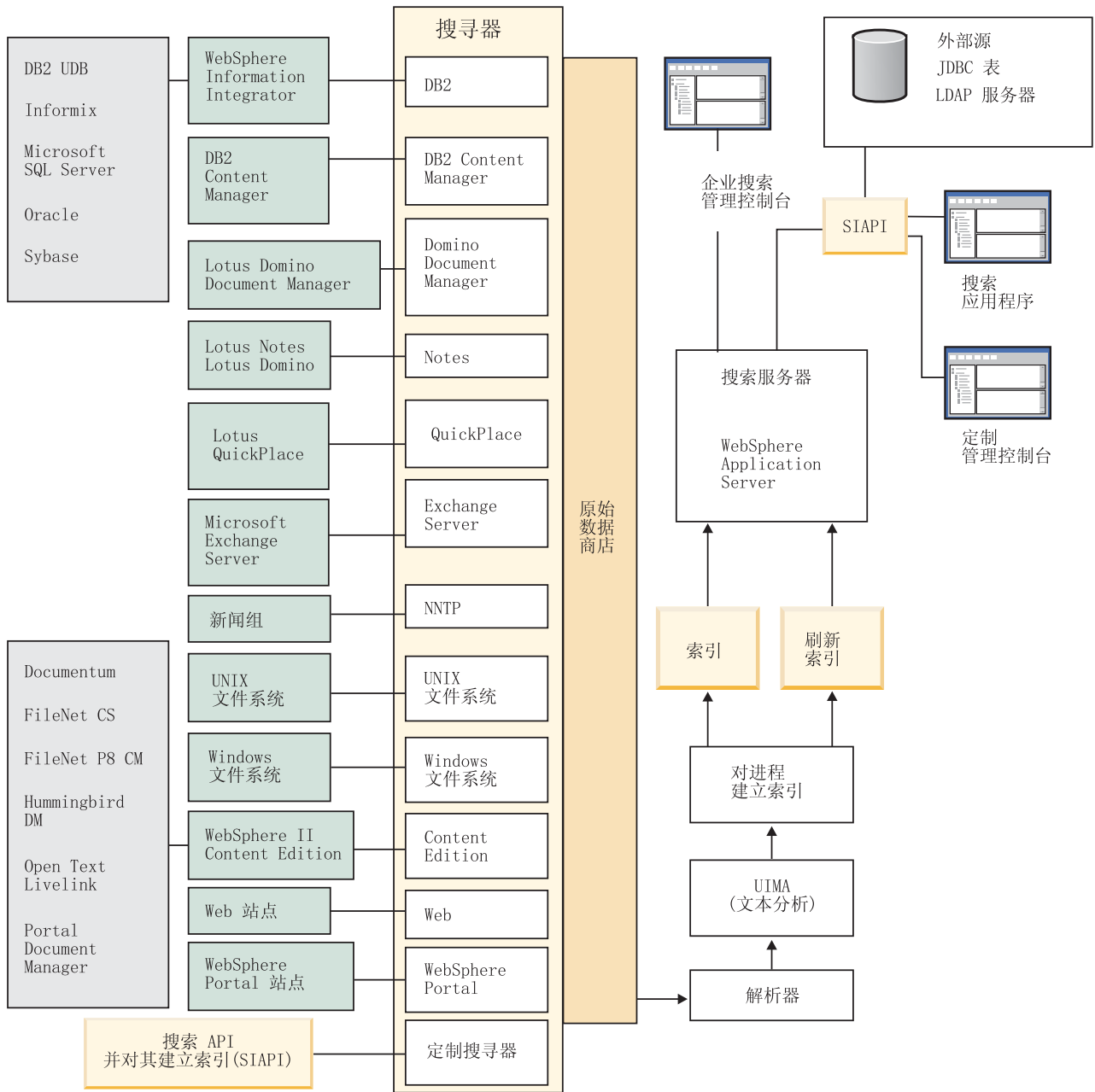


图 1. 数据在企业搜索系统中的流动方式


企业搜索管理概述

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

“集合”视图

依据下列准则来创建第一个集合以及管理系统。

登录 登录到企业搜索管理控制台。集合视图是创建和管理集合的入口点。

提示: 为了获取有关管理控制台的帮助, 随时可以单击工具栏上的**帮助**, 也可以单击**此页的帮助**。如果消息有详细说明和纠正操作, 则可以单击消息末尾的  **更多信息**图标以查看详细信息。

创建集合

选择下列其中一种方法来创建新的空集合:

- 要使用集合向导来创建集合, 请单击**集合向导**并遵循向导提示进行操作。
- 要使用“集合”视图来创建集合, 请单击**创建集合**, 填写创建集合页上的字段, 然后单击**确定**。

配置集合

您必须编辑一个新的空集合以便向它添加内容, 并指定要如何搜寻数据以及使数据可供搜索的选项。

对新集合单击  **编辑**, 然后选择一个页面以指定该集合的选项。

警告: 如果单击 Web 浏览器中的“后退”或“刷新”按钮, 结果就会不一致, 并有可能丢失数据。当配置集合并需要转到先前页面或刷新管理控制台中的信息时, 请单击企业搜索管理控制台中的**上一步按钮**和**刷新按钮**, 而不要单击 Web 浏览器中的“后退”按钮和“刷新”按钮。

- 在常规页上, 可以指定应用于整个集合的选项:
 - 可以编辑常规选项以更改集合的名称或描述, 或更改集合的估算的大小。
 - 可以查看关于集合的不能更改的信息, 例如集合标识或用于对搜索结果中的文档进行排名的静态排名方法。
 - 如果在创建集合时对集合启用了安全性, 则可以启用或禁用文档级安全性控制。
- 在搜寻页上, 配置至少一个搜寻器。

单个集合可以包含来自各种数据源的数据。必须为要包括的每种数据源类型配置至少一个搜寻器。当创建搜寻器时, 特定于所搜寻的数据类型的向导将帮助您配置该搜寻器。

- 在解析页上, 可以配置选项以指示如何对搜寻到的数据进行解析以便可以有效地搜索该数据:

- 可以指定是否将解析 XML 文档以便可使用本机 XML 搜索来搜索这些文档。
- 要把将文档与类别关联，从而使用户可搜索集合的一个子集，或按照搜索结果文档所属的类别来浏览搜索结果文档。
- 可以将 XML 元素和 HTML 元数据元素映射至索引中的搜索字段，从而使用户可以在查询中指定字段名并搜索文档的特定部分。
- 如果向企业搜索系统添加了定制文本分析引擎，则可以选择用于集合的定制文本分析引擎，然后指定文本处理选项以提高信息的可检索性。
- 可以使字段与升降类相关联以影响与查询项匹配的字段在搜索结果中的排名。
- 在索引页上，配置重组索引和刷新索引的时间表。安排索引构建时间表使索引构建频繁地发生，以使用户总是能够访问到最新的信息。也可以执行以下可选活动：
 - 使用户可在查询项中指定通配符。
 - 配置作用域，以使用户可搜索集合中有限的部分，而不必搜索索引中的所有文档。
 - 折叠搜索结果，以便将搜索结果中来自相同来源的所有文档分组并折叠在一起。
 - 指定升降因子以影响与 URI 模式相匹配的文档在搜索结果中的排名。
 - 从索引中除去 URI。例如，在创建某个集合后，您可能需要阻止用户看见某些文档。
- 在搜索页上，可以指定选项以指示如何搜索集合中的文档：
 - 可以为搜索结果设置额外的高速缓存空间，并更改集合的缺省语言。
 - 如果已将定制同义词字典、无用词字典和升降字典添加到企业搜索系统中，则可以选择用户搜索该集合时要使用的字典。
 - 可以指定搜索结果中文档摘要的显示长度。
 - 如果希望只要查询包括特定的关键字或短语，特定的 URI 就自动出现在搜索结果中，可以配置快速链接。
- 在日志页上，可执行以下活动：
 - 指定消息记录类型选项和日志文件回收频率选项。
 - 指定选项，用于接收关于集合活动的警报。例如，当平均搜索响应时间超出指定的限制时，警报可以通知您。
 - 指定选项，以便只要记录某些消息或某些类型的消息时就可接收到电子邮件。
 - 指定选项以记录信息，这样就可以确定从企业搜索系统中废弃文档的时间、位置以及原因。

启动组件

在指定了要搜寻的数据源以及用于收集和搜索数据的选项之后，就可以启动构建集合的过程。启动组件的顺序很关键。搜寻器必须先搜寻数据，这样才能对数据进行解析；解析器必须先对搜寻到的数据进行分析，然后才能对该数据建立索引；必须先刷新索引或重组索引，然后搜索服务器才能开始处理搜索请求。

“外部源”视图

如果要搜索数据源而不对它们进行搜寻或建立索引，则可以单击工具栏上的**外部源**以指定选项，从而使数据源可供搜索。必须指定信息以使 Java 数据库连接 (JDBC) 数据库和轻量级目录访问协议服务器可供企业搜索访问。在使外部源与搜索应用程序相关联之后，用户可以在搜索包含已搜寻、已解析并且已建立索引的数据的集合时搜索这些外部源。

“系统”视图

如果您是**企业搜索管理员**角色的成员，则可以单击工具栏上的**系统**以执行下列活动（**集合管理员**、**操作员**和**监视员**只有在得到**企业搜索管理员**授权后才能访问此视图）：

- 检查系统资源的可用性。
- 配置客户机数据侦听器应用程序以使它们能够更新集合。
- 将定制文本分析引擎添加至系统。
- 将定制的同义词字典、无用词字典和升降字字典添加至系统。
- 指定多少个集合可以并行地构建索引，并指定是否可以并发地运行单个集合的刷新请求和重组请求。
- 为系统级事件配置警报。
- 指定选项以指示如何记录由系统级会话生成的消息。
- 指定邮件服务器信息，以便可以接收关于企业搜索活动的电子邮件。


“安全性”视图

如果您是**企业搜索管理员**角色的成员，可以单击**安全性**以指定安全性选项。**集合管理员**、**操作员**和**监视员**不能访问此视图。

如果在 IBM WebSphere Application Server 中启用了安全性，则可以使用安全性视图来配置管理角色。通过配置管理角色，可以允许更多的用户来管理系统，然而限制每个用户只能访问特定的功能和集合。

在创建自己的搜索应用程序之前，可以使用**样本搜索应用程序**来搜索所有集合以及外部源。在创建定制搜索应用程序之后，使用**安全性视图**来使应用程序与它可以搜索的集合和外部源相关联。

“监视”视图

可以随时单击  **监视**来监视系统或集合组件。如果管理角色允许的话，还可以在监视组件时启动和停止组件进程。

相关概念

第 31 页的『[企业搜索搜寻器管理](#)』

为要包括在集合中的不同类型的数据配置搜寻器。单个集合可以包含任意数目的搜寻器。

第 199 页的『[监视企业搜索活动](#)』

在监视系统和集合活动时，您可以查看各种进程的状态、监视潜在的问题或调整配置以提高性能。

相关任务

第 195 页的『启动企业搜索服务器』

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

第 197 页的『停止企业搜索服务器』

如果对企业搜索服务器的配置作了更改，或者需要对问题进行故障诊断，则可能需要停止并重新启动企业搜索服务器。

使用集合向导来创建集合

如果不熟悉企业搜索的话，向导可以帮助您创建集合。向导将提供关于过程中的每个步骤的详细信息并使您能够在进展的过程中保存设置。

第 26 页的『使用“集合”视图来创建集合』

使用集合视图来创建空集合。然后，可以对该集合进行编辑，以指定一些选项 以指示如何将数据添加到集合中以及使该集合可搜索。

登录到管理控制台

要管理企业搜索系统，在 Web 浏览器中指定一个 URL，然后登录到管理控制台。

在开始之前

必须使用有权访问企业搜索管理控制台的用户标识来登录：

- 如果未在 WebSphere Application Server 中启用全局安全性，则只有在安装 WebSphere II OmniFind Edition 时指定的企业搜索管理员能够访问管理控制台。
- 如果在 WebSphere Application Server 中启用了全局安全性，则可以使用企业搜索管理控制台来配置管理角色。配置的用户标识必须存在于 WebSphere Application Server 用户注册表中。通过配置管理角色就可以允许更多的用户登录到管理控制台，但可以控制每个管理用户可以访问的功能和集合。

过程

要登录到企业搜索管理控制台：

1. 在 Web 浏览器中输入管理控制台的 URL。例如：

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

SearchServer.com 是企业搜索的搜索服务器的主机名。

根据 Web 服务器的配置，可能还需要指定端口号。例如：

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

2. 在欢迎页上，输入用户标识和密码并单击**登录**。

将显示集合视图，此视图是管理系统和集合的入口点。如果使用管理角色，则可以执行的操作以及所看到的集合取决于您的管理角色。

如果您的会话在一段时间内不活动，则系统将自动使您注销。要继续管理系统，请再次登录。

在完成对集合的管理后，可以单击**注销**从控制台注销。然后，可以使用另一标识和密码登录，也可以关闭 Web 浏览器以退出管理控制台。

相关概念

第 166 页的『管理角色』

企业搜索使用角色的概念来控制对管理控制台中各种功能的访问。

相关任务

第 195 页的『启动企业搜索服务器』

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

更改单服务器配置中的企业搜索管理员密码

企业搜索管理员的密码是以加密格式存储的。要更改此密码，可使用 `eschangepw` 脚本。

在开始之前

企业搜索管理员标识和密码在操作系统上必须有效，并且必须有权访问和配置 DB2 通用数据库。

关于本任务

初始企业搜索管理员标识的密码是在安装 WebSphere II OmniFind Edition 时指定的。

要更改该密码，必须运行 `eschangepw` 脚本在整个企业搜索系统中应用此更改。安装程序创建了两个可以用于 `eschangepw` 脚本的环境变量：

ES_INSTALL_ROOT

企业搜索安装目录。

ES_NODE_ROOT

企业搜索数据目录。企业搜索管理员标识的密码存储在此目录中的 `es.cfg` 文件中。

由于 `eschangepw` 脚本安装在 `ES_INSTALL_ROOT/bin` 目录中，因此可以从系统中的任何地方运行该脚本。

过程

要更改单服务器配置中的企业搜索管理员密码：

1. 作为企业搜索管理员登录。
2. 通过输入此命令来停止服务器：`esadmin stop`
3. 打开 WebSphere Application Server 管理控制台，然后停止 `server1` 和 `ESSearchServer` 企业应用程序。
4. 通过使用 UNIX 操作系统命令或 Microsoft Windows 更改密码工具来更改企业搜索管理员用户标识的系统密码。
5. 运行以下脚本，其中 `newValue` 是在第 4 步中指定的密码：

操作系统	命令
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

6. 在 WebSphere Application Server 管理控制台中，启动 `server1` 和 `ESSearchServer` 应用程序。

7. 通过输入下列命令，重新启动 WebSphere II OmniFind Edition 公共通信层（CCL）：

操作系统	命令
UNIX	stopccl.sh, 然后是 startccl.sh -bg
Windows 命令提示符	stopccl, 然后是 startccl
Windows 服务管理工具	<ol style="list-style-type: none">1. 启动 Windows 服务。2. 右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择停止。3. 再次右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择属性。4. 单击登录选项卡。5. 通过指定 <i>newValue</i>, 然后单击确定来更改密码。6. 再次右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择启动。

8. 通过输入此命令来重新启动企业搜索：`esadmin start`。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

更改多服务器配置中的企业搜索管理员密码

企业搜索管理员的密码是以加密格式存储的。要更改该密码，可使用 `eschangePW` 脚本在企业搜索系统中的所有计算机上做出更改。

在开始之前

企业搜索管理员标识和密码在操作系统上必须有效，并且必须有权访问和配置 DB2 通用数据库。

企业搜索管理员的密码在安装此 WebSphere II OmniFind Edition 的所有计算机上必须是相同的。

关于本任务

初始企业搜索管理员标识的密码是在安装 WebSphere II OmniFind Edition 时指定的。

要更改此密码，并在整个企业搜索系统中应用此更改，必须在多服务器配置中的每台计算机上运行 `eschangePW` 脚本。安装程序创建了两个可以用于 `eschangePW` 脚本的环境变量：

ES_INSTALL_ROOT

企业搜索安装目录。

ES_NODE_ROOT

企业搜索数据目录。企业搜索管理员标识的密码存储在此目录中的 `es.cfg` 文件中。

由于 `eschangepw` 脚本安装在 `ES_INSTALL_ROOT/bin` 目录中，因此可以从系统中的任何地方运行该脚本。

过程

要更改多服务器配置中的企业搜索管理员密码：

1. 在安装了索引服务器的计算机上，作为企业搜索管理员登录。
2. 通过输入此命令来启动系统维护方式：`esadmin system maintenance`
3. 通过使用 UNIX 操作系统命令或 Microsoft Windows 更改密码工具来更改企业搜索管理员用户标识的系统密码。
4. 运行以下脚本，其中 `newValue` 是在第 3 步中指定的密码：

操作系统	命令
UNIX	<code>eschangepw.sh newValue</code>
Windows	<code>eschangepw newValue</code>

5. 通过在索引服务器上输入下列命令，重新启动 WebSphere II OmniFind Edition 公共通信层 (CCL)：

操作系统	命令
UNIX	<code>stopccl.sh</code> ，然后是 <code>startccl.sh -bg</code>
Windows 命令提示符	<code>stopccl</code> ，然后是 <code>startccl</code>
Windows 服务管理工具	<ol style="list-style-type: none">1. 启动 Windows 服务。2. 右键单击 WebSphere Information Integrator OmniFind Edition，然后选择停止。3. 再次右键单击 WebSphere Information Integrator OmniFind Edition，然后选择属性。4. 单击登录选项卡。5. 通过指定 <code>newValue</code>，然后单击确定来更改密码。6. 再次右键单击 WebSphere Information Integrator OmniFind Edition，然后选择启动。

6. 在安装了搜寻器服务器的计算机上，作为企业搜索管理员登录。
7. 通过使用 UNIX 操作系统命令或 Windows 更改密码工具来更改企业搜索管理员用户标识的系统密码。此密码必须与在第 3 步中指定的密码相匹配。
8. 运行以下脚本，其中 `newValue` 是在第 3 步中指定的密码：

操作系统	命令
UNIX	<code>eschangepw.sh newValue</code>

操作系统	命令
Windows	eschangepw <i>newValue</i>

9. 通过输入以下命令在搜寻器服务器上重新启动搜寻器:

操作系统	命令
UNIX	stopccl.sh, 然后是 startccl.sh -bg
Windows 命令提示符	stopccl, 然后是 startccl
Windows 服务管理工具	<ol style="list-style-type: none"> 1. 启动 Windows 服务。 2. 右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择停止。 3. 再次右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择属性。 4. 单击登录选项卡。 5. 通过指定 <i>newValue</i>, 然后单击确定来更改密码。 6. 再次右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择启动。

10. 在其中一台安装了搜索服务器的计算机上, 作为企业搜索管理员登录。

11. 停止 IBM HTTP Web 服务器。(此步骤可防止 Network Dispatcher 将查询传递给此服务器。)

12. 通过输入下列命令在搜索服务器上停止 CCL:

操作系统	命令
UNIX	stopccl.sh
Windows 命令提示符	stopccl
Windows 服务管理工具	<ol style="list-style-type: none"> 1. 启动 Windows 服务。 2. 右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择停止。

13. 打开 WebSphere Application Server 管理控制台, 然后停止 server1 和 ESSearchServer 应用程序。

14. 通过使用 UNIX 操作系统命令或 Windows 更改密码工具来更改企业搜索管理员用户标识的系统密码。此密码必须与在第 19 页的 3 步中指定的密码相匹配。

15. 运行以下脚本, 其中 *newValue* 是在第 19 页的 3 步中指定的密码:

操作系统	命令
UNIX	eschangepw.sh <i>newValue</i>
Windows	eschangepw <i>newValue</i>

16. 通过输入下列命令在搜索服务器上重新启动 CCL:

操作系统	命令
UNIX	startccl.sh -bg
Windows 命令提示符	startccl
Windows 服务管理工具	<ol style="list-style-type: none">1. 启动 Windows 服务。2. 右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择属性。3. 单击登录选项卡。4. 通过指定 <i>newValue</i>, 然后单击确定来更改密码。5. 再次右键单击 WebSphere Information Integrator OmniFind Edition, 然后选择启动。

17. 在第二个搜索服务器上, 重复第 20 页的 10 步至 16 步。

18. 启动 IBM HTTP Web 服务器。

19. 在 WebSphere Application Server 管理控制台中, 启动 server1 和 ESSearchServer 企业应用程序。

20. 在安装了索引服务器的计算机上, 作为企业搜索管理员登录, 然后输入此命令以启动企业搜索: `esadmin start`

21. 登录到企业搜索管理控制台, 监视集合, 单击搜索页, 然后单击**停止**和**启动**以重新启动搜索服务器。对企业搜索系统中的每个集合重复此步骤。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

更改企业搜索系统的端口号

如果企业搜索系统进行通信时使用的端口号与另一产品使用的端口号有冲突, 就必须更改企业搜索端口号。

关于本任务

企业搜索系统的端口号是在安装 WebSphere II OmniFind Edition 时指定的。(在多服务器配置中, 在所有服务器上指定同一个端口号。)

如果端口号不可用(例如, 已将该端口号分配给同一服务器上的另一产品), 冲突就会导致在 `CCLServer_date.log` (其中 *date* 指定日志文件的创建日期) 文件中记录以下错误消息:

```
FFQ00273W 内部警告 java.net.BindException: 地址已在使用中: 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)
```


过程

要更改企业搜索使用的端口号:

1. 输入以下命令以停止企业搜索系统:

```
esadmin stop
```

2. 转到需要更改端口号的服务器, 然后通过输入下列命令来停止公共通信层 (CCL):

操作系统	命令
UNIX	stopccl.sh,
Windows 命令提示符	stopccl
Windows 服务管理工具	1. 启动 Windows 服务。 2. 右键单击 WebSphere Information Integrator OmniFind Edition , 然后选择停止。

3. 编辑 `$ES_NODE_ROOT/nodeinfo/es.cfg` 文件 (在 UNIX 上) 或 `%ES_NODE_ROOT%\nodeinfo\es.cfg` 文件 (在 Windows 上), 找到以下属性, 指定新的端口号值, 然后保存并关闭该文件:

```
CCLPort=new_port_number
```

4. 通过输入下列命令重新启动 CCL:

操作系统	命令
UNIX	startccl.sh -bg
Windows 命令提示符	startccl
Windows 服务管理工具	1. 启动 Windows 服务。 2. 右键单击 WebSphere Information Integrator OmniFind Edition , 然后选择启动。

5. 转到索引服务器, 按照步骤 2 中的指示信息来停止 CCL。
6. 编辑 `$ES_NODE_ROOT/nodeinfo/es.cfg` 文件 (在 UNIX 上) 或 `%ES_NODE_ROOT%\nodeinfo\es.cfg` 文件 (在 Windows 上)。
 - a. 找到以下属性, 其中 `computer_name` 是服务器的名称, 在步骤 3 中修改了该服务器的端口号。nodeN 属性中的 N 是用于标识服务器的编号。

```
nodeN.destination=computer_name
```
 - b. 找到以下子属性, 在此处指定在步骤 3 中对服务器指定的端口号, 然后保存并关闭该文件:

```
nodeN.port=new_port_number
```
7. 按照步骤 4 中的指示信息来重新启动 CCL。
8. 输入以下命令以重新启动企业搜索系统:

```
esadmin start
```

此命令完成后, 在所有企业搜索服务器上都将开始使用新端口号。

|

相关参考

| 第 229 页的『企业搜索命令、返回码和会话标识』

| 可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和
| 停止系统。

企业搜索集合

企业搜索集合包含用户可以通过单个查询来搜索的整一组源。通过联合，用户就可以在一个查询中搜索多个集合。

创建集合时，指定应用于整个集合的选项。在将内容添加至集合之前，该集合是空的。

可以通过两种方法来将集合添加至企业搜索系统：

- 如果您不熟悉企业搜索管理控制台，或者如果您仍然正在了解集合组件的协同工作方式，则可能想使用集合向导来创建集合。集合向导帮助您完成任务并允许您在创建集合时将工作保存为草稿集合。
- 当您更加熟悉管理控制台时，可能会通过选择要在集合视图中管理的特定页面来创建集合。

在创建集合之后，使用集合视图中的控件来编辑和监视集合、企业搜索系统和安全性选项。

集合联合

如果搜索应用程序中构建了联合支持，用户可同时搜索多个集合。联合还使您能够超越集合大小的限制（即每个集合 20 000 000 个文档）。例如，用户可以搜索两个集合，每个集合都包含 20 000 000 个文档。

搜索质量取决于各个集合产生的分数，然后将合并这些分数以产生最终的结果集。其结果与提交两个单独的搜索，然后再对这两个结果进行合并和排名一样。

相关任务

第 201 页的『监视集合』

您可以查看关于集合中的每个组件的状态的一般信息，也可以选择选项以查看关于各个组件和 URI 的详细信息。

第 190 页的『从 WebSphere Portal 迁移集合』

要将集合从 WebSphere Portal 迁移至企业搜索，请在 WebSphere Portal 中准备集合，然后使用迁移向导来迁移它们。

使用集合向导来创建集合

如果不熟悉企业搜索的话，向导可以帮助您创建集合。向导将提供关于过程中的每个步骤的详细信息并使您能够在进展的过程中保存设置。

在开始之前

要创建集合，您必须是企业搜索管理员角色的成员。

要将内容添加至集合或者要指定有关如何对集合内容进行解析、建立索引或进行搜索的选项，您必须是企业搜索管理员或该集合的集合管理员。

关于本任务

在创建集合时，可以以草稿状态保存该集合。当集合处于草稿状态时，任何有权管理该集合的管理人员都可以更改该集合。例如，可能想请有 Lotus Notes 数据源经验的集合管理员来配置 Notes 搜寻器。以后，有 UNIX 系统经验的集合管理员可以编辑草稿集合以配置 UNIX 文件系统 搜寻器。


过程

要使用集合向导来创建集合：

1. 单击**集合**以打开集合视图。
2. 单击**集合向导**。
3. 遵循向导中的指示信息来创建一个空集合并向该集合添加内容。

必须配置关于该集合的一般信息并创建至少一个搜寻器。可以接受其余配置选项的缺省值，也可以为新集合指定选项。

4. 要在创建集合完成之前保存该集合，请单击**保存为草稿**。

您的集合将与其他草稿集合一起列在集合视图中。如果对集合启用了安全性，在集合名旁边就会显示  **已启用集合安全性**图标。

5. 要返回到仍在创建的集合，请在集合视图中单击**返回至向导**。
6. 单击**完成**以创建集合。

新集合将与其他集合一起列在集合视图中。

在创建集合之后，必须启动用于搜寻、解析、建立索引和搜索集合的过程。在准备好使该集合与可以搜索该集合的搜索应用程序相关联之前，可以使用样本搜索应用程序（名为 Default）来搜索这个新集合。

使用“集合”视图来创建集合

使用集合视图来创建空集合。然后，可以对该集合进行编辑，以指定一些选项 以指示如何将数据添加到集合中以及使该集合可搜索。

在开始之前

要创建集合，您必须是企业搜索管理员角色的成员。

要将内容添加至集合或者要指定有关如何对集合内容进行解析、建立索引或进行搜索的选项，您必须是企业搜索管理员或该集合的集合管理员。

关于本任务


有关可以对新集合指定的值的信息，请在创建集合时单击**帮助**。

过程

要从集合视图中创建集合：

1. 在集合视图中，单击**创建集合**。
2. 在创建集合页上，在下列字段中提供信息或进行选择：
 - **集合名**指定一个描述性名称来说明该集合的内容或用途。

- **集合安全性。** 指定是否要启用集合安全性。创建集合后，就不能更改此设置。如果启用了集合安全性，以后就可以指定选项以强制实施文档级访问控制。
 - **文档重要性（静态排名模型）。** 指定一种策略，以便指定用来对搜索结果中的文档进行排名的静态排名系数。创建集合后，就不能更改此值。
 - **分类类型。** 指定是否希望能够按文档所属的类别来搜索文档。
 - **要使用的语言。** 指定用来搜索该集中的文档的缺省语言。
3. 接受下列字段的缺省值，或者指定要对此集合使用的选项：
- **描述。** 缺省情况下，不创建任何描述。
 - **估算的文档数。** 集合的缺省估算大小是 1 000 000 个文档。系统使用此值来估算该集合的内存和磁盘资源需求，而不是限制该集合的大小。
 - **集合数据的位置。** 与集合相关的文件的缺省位置是在索引服务器上。创建集合后，就不能更改此值。
 - **集合标识。** 缺省集合标识基于集合名。创建集合后，就不能更改此值。（如果指定定制集合标识，搜索应用程序就会使用此标识来调用该集合，而不是使用系统创建的有可能含义模糊的标识。）
 - **n-gram 分段。** 缺省分段方法是基于 Unicode 的空格分段。仅当集合包含中文、日语或韩国语文档并且想让解析器使用 n-gram 分段来进行字定界时，才应该选择此选项以使用 n-gram 分段。创建集合后，就不能更改此值。
4. 单击**确定**。

集合视图将列出新集合以及企业搜索系统中的其他集合。如果对集合启用了安全性，在集合名旁边就会显示  **已启用集合安全性**图标。

在将内容添加至集合之前，该集合是空的。要将内容添加到新集合中，请在集合视图中选择该集合，编辑它，创建至少一个搜寻器，然后指定选项以指示如何对数据进行解析、建立索引和进行搜索。

然后，必须启动用于对该集合进行搜寻、进行解析、建立索引和进行搜索的过程。在准备好使用户能够通过定制搜索应用程序搜索新集合之前，可以使用样本搜索应用程序来搜索新集合。

相关概念

第 13 页的『企业搜索管理概述』

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

编辑集合

编辑集合以指定关于要包括在集合中的文档的信息。

在开始之前

要编辑集合，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。


关于本任务

编辑集合时，指定一些选项，以便搜寻数据源、解析文档、重组索引和刷新索引、搜索已建立了索引的内容以及记录错误消息。创建集合时，必须编辑该集合以将内容添

加到其中。以后，可以编辑该集合以更新内容，也可以更改搜寻信息、解析信息、对信息建立索引、搜索信息或记录信息的方式。

过程

要编辑集合：

1. 单击工具栏中的**集合**以打开集合视图。
2. 在集合列表中找到要编辑的集合，然后单击 **编辑**。
3. 对下列任何一页进行更改：

常规 指定有关集合的一般信息并查看您不能更改的设置。如果创建集合时为其启用了安全性，则可以配置文档级安全性选项。

搜寻 指定要搜寻的数据源，并指定选项以指示如何搜寻内容。每个集合都必须至少包含一个搜寻器，单个集合可以包含来自多类数据源的数据。必须为要包括在集合中的每种数据源类型配置至少一个搜寻器。

解析 指定选项以指示如何对搜寻到的文档进行解析和分析。可以配置类别（它们使用户可以搜索集合的子集），并可以配置一些规则来使用户能够搜索 XML 和 HTML 文档的特定部分。如果将定制文本分析引擎添加到企业搜索系统中，则可以选择用于对此集合中的内容进行分析和添加注释的定制文本分析引擎。还可以使字段与升降类相关联以影响文档排名。

索引 指定时间表以重组整个索引和使用新的或已更改的内容来刷新索引。也可以配置选项以便在查询中使用通配符、将集合视图限制为某个范围的 URI、折叠来自同一 Web 站点的搜索结果以及从索引中除去 URI。

搜索 指定搜索进程选项，例如配置搜索高速缓存和选择搜索语言。还可以配置快速链接，快速链接是一项功能，它确保每当用户在查询中包括了特定的单词或短语时都返回预先确定的 URI。如果已经在企业搜索系统中添加了定制字典，则可以选择用于搜索此集合的字典。

日志 指定要记录的消息类型，并指定选项以指示如何创建和循环使用日志文件。可以配置警报以便可以在某些事件发生时接收到通知，并可以指定一些选项以便每当记录某些消息或某些类型的消息时就可以接收到电子邮件。也可以指定选项以便记录关于从企业搜索系统中废弃的文档的信息。

删除集合

彻底删除集合时，将从企业搜索系统中除去与该集合相关的所有信息。

在开始之前

要删除集合，您必须是企业搜索管理员角色的成员。

在可以删除集合之前，必须停止与该集合相关联的所有进程。


关于本任务

删除集合可能是一个相当耗时的过程。在确认要删除集合之后，系统将删除系统中与该集合相关的所有数据。

提示：可能会看到一条消息指示请求的操作已超时，即使该进程仍在后台运行亦如此。要确定该任务是否已完成，请单击管理控制台中的**刷新**（不要单击 Web 浏览器中的**刷新**）。当集合名不再出现在集合列表中时，表示删除过程已完成。

过程

要删除集合：

1. 单击**集合**以打开集合视图。
2. 在集合列表中，找到要删除的集合并单击 **删除**。

企业搜索搜寻器管理

为要包括在集合中的不同类型的数据配置搜寻器。单个集合可以包含任意数目的搜寻器。

配置搜寻器

使用企业搜索管理控制台来创建、编辑和删除搜寻器。通常，所搜寻的数据类型方面的专家负责配置搜寻器。例如，要设置一个搜寻器以搜寻 Lotus Notes 数据源，则集合管理员应该是 Notes 管理员，或者集合管理员应该与了解要搜寻的数据库的人员密切配合。

当创建搜寻器时，用于所搜寻的数据类型的向导可以帮助您指定控制搜寻器如何使用系统资源的属性。该向导还帮助您选择要搜索的源。

可以随时对现有搜寻器进行更改。可以根据需要编辑搜寻器属性或搜寻空间的各部分。搜寻器向导还可以帮助您进行这些更改。

用基本值来填充新的搜寻器

可以使用系统缺省值来创建搜寻器，也可以通过复制为另一个同类型搜寻器指定的值来创建搜寻器。如果使用现有搜寻器来作为新搜寻器的基础，则可以快速创建多个具有类似属性的搜寻器，然后对它们进行配置，例如，以便搜寻不同的源或者根据不同的搜寻时间表来进行操作。

通过复制搜寻器，可以在多个使用相同搜寻规则的搜寻器之间分配搜寻工作负载。例如，可能会因为要对另一台 Lotus Notes 服务器使用相同的属性和字段搜寻规则而复制 Notes 搜寻器。唯一的差别可能是每个搜寻器所搜寻的数据库以及文档级安全性设置。

将搜寻器类型组合到集合中

企业搜索搜寻器被设计为收集来自特定类型的数据源的信息。当为集合配置搜寻器时，必须决定如何对这些不同的数据源类型进行组合，以便用户可以方便地搜索企业数据。例如，如果要让用户能够通过单个查询来搜索 Microsoft Windows 文件系统和 Microsoft Exchange Server 公共文件夹，则创建一个包含 Windows 文件系统 搜寻器和 Exchange Server 搜寻器的集合。

当将多种类型的搜寻器组合在单个集合中时，请确保所有搜寻器都可以使用相同的静态排名方法。（在创建集合时指定静态排名方法。）例如，如果组合 Web 源（这些源使用文档链接数作为排名系数）和 NNTP 源（这些源通常使用文档日期作为排名系数），则搜索结果的质量可能会下降。

文档级安全性

如果在创建集合时对集合启用了安全性，则可以配置文档级安全性选项。每个搜寻器都可以使安全性令牌与它搜寻的文档相关联。如果在配置搜寻器时指定想要使用文档级安全性，则搜寻器会将指定的安全性令牌与每个文档关联，并且这些令牌将与文档一起被添加到索引中。

如果在定制搜索应用程序中启用安全性，则应用程序可以使用搜寻器设置的与文档相关联的安全性令牌来认证用户。此功能使您能够将访问限制为只对集合中的某些文档进行并允许其他文档被所有用户搜索。例如，在一个集合中，可以允许所有用户访问 Microsoft Exchange Server 公共文件夹中的所有文档，但是只允许使用特定用户标识的用户访问 Lotus Notes 数据库中的文档。

通过在 Java 类中对规则进行编码，可以应用定制业务规则以确定安全性令牌值。配置搜寻器属性时，请指定搜寻器搜寻文档时要使用的插件名。插件添加的安全性令牌存储在索引中，这些安全性令牌可用于控制对文档的访问。

在配置某些类型的搜寻器时，可以指定其他安全性控制。例如，可以指定要在查询处理期间验证用户。如果启用此选项，就会将用户凭证与所要搜索的数据源维护的当前访问控制表进行比较。对当前凭证执行的这种验证可以取代基于企业搜索索引中的安全性令牌的验证，也可以同时进行这两种验证。

安排搜寻器的时间表

为 Web、NNTP 和 WebSphere Portal 源创建的搜寻器持续运行。在启动这样的搜寻器之后，通常不需要停止它们，除非更改该搜寻器的配置。

对于所有其他搜寻器类型，在配置搜寻器时指定搜寻时间表。对于某些数据源类型，单个时间表控制搜寻器访问搜寻空间中的所有数据源的时间。对于其他数据源类型，可以为特定的数据源指定不同的时间表。例如，可以指定不同的时间表来搜寻该搜寻器所搜寻的每个 Lotus Notes 数据库。

在配置时间表时，应该指定所要执行的搜寻的类型。可以安排完整搜寻时间表以在搜寻空间中搜寻所有文档，可以安排搜寻时间表以搜寻对搜寻空间所作的所有更新（新文档、修改过的文档和已删除的文档），也可以安排搜寻时间表以便只搜寻新文档和修改过的文档。完整搜寻的耗时最长。除去已删除文档的搜寻比忽略已删除文档的搜寻更耗时。

在编辑搜寻器的搜寻空间时，可以指定第二个搜寻时间表。例如，可以配置一个时间表以便在每个星期六晚上在搜寻空间中搜寻所有文档，然后配置运行频率更高的第二个时间表来搜寻新文档和修改过的文档。

通过创建多个搜寻器时间表，可以更好地控制该搜寻器访问目标源的时间。例如，要在不同的时区搜寻数据库，可以将搜寻器的时间安排在用户很可能已完成当天工作的时间。

相关概念

第 13 页的『企业搜索管理概述』

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

第 172 页的『文档级安全性』

如果创建集合时对集合启用了安全性，则可以配置文档级安全性控制。文档级安全性确保搜索集合的用户只能够访问他们被允许查看的文档。

相关任务

第 203 页的『监视搜寻器』

您可以查看关于集合中每个搜寻器的状态的一般信息，也可以选择选项以查看关于搜寻器活动的详细信息。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识（URI）都指示了将文档添加到该集合的搜寻器的类型。

创建搜寻器

在创建搜寻器时，指定要创建的搜寻器的类型。一个向导将帮助您指定关于要包括在集合中的数据的信息。

在开始之前

要创建搜寻器，您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

关于本任务

必须为集合至少创建一个搜寻器。创建的搜寻器的类型取决于要包括在集合中的数据的类型。对于创建的搜寻器类型，将有一个向导帮助您指定该搜寻器的选项。例如，该向导将帮助您指定选项以指示搜寻器使用系统资源的方式。该向导还将帮助您选择要包括在集合中的数据源。

过程

要创建搜寻器：

1. 编辑集合，选择搜寻页，然后单击**创建搜寻器**。
2. 选择搜寻器的搜寻器类型和基本值：
 - a. 选择支持所要搜寻的数据的搜寻器类型，如 Web 站点、Lotus Notes 数据库或 UNIX 文件系统。

在选择搜寻器类型之后，将显示有关搜寻器创建方式的选项。

- b. 选择搜寻器的基本值：

将系统缺省值用于新搜寻器

用安装缺省值来填充初始搜寻器设置。

如果选择此选项，则单击**下一步**以开始配置新的搜寻器。

为新搜寻器克隆现有搜寻器的值

使用为另一个此类搜寻器配置的值来填充初始搜寻器设置。

如果选择此选项，则将显示与此搜寻器类型相匹配的搜寻器的列表。请选择要用于新搜寻器的搜寻器，然后单击**下一步**以开始配置新的搜寻器。

将打开用于所创建的搜寻器类型的向导。遵循该向导的提示来创建搜寻器。单击该向导中的任何页上的**帮助**以了解有关可以对该类型的搜寻器指定的选项的更多信息。

新的搜寻器将与其他属于该集合的搜寻器一起列在搜寻页上。任何时候当需要更改搜寻器时，都可以单击选项以编辑搜寻器属性和搜寻空间。

编辑搜寻器属性

可以更改关于搜寻器以及它的数据搜寻方式的信息。例如，可以更改搜寻器使用系统资源的方式。

在开始之前


要编辑搜寻器属性，您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

关于本任务

在编辑搜寻器属性时，单击**帮助**以了解可以进行的更改的类型。可以编辑的属性取决于搜寻器类型。

过程

要编辑搜寻器的属性：

1. 编辑集合，选择搜寻页，找到要编辑的搜寻器，然后单击  **搜寻器属性**。
2. 更改搜寻器属性，然后单击**确定**。
3. 要使更改生效，请停止并重新启动搜寻器。（如果只更改了搜寻器描述，就不需要重新启动搜寻器。）

编辑搜寻空间

可以更改关于搜寻器所搜寻的数据源的信息。例如，可以添加数据源、除去数据源、更改搜寻时间表以及更改在特定数据源中搜寻文档的规则。

在开始之前


要编辑搜寻空间，您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

关于本任务

要了解可以对所管理的搜寻器类型进行的更改的信息，请在编辑搜寻空间时单击**帮助**。

过程

要编辑搜寻空间：

1. 编辑集合，选择搜寻页，找到要编辑的搜寻器，然后单击  **搜寻空间**。
2. 通过选择要更改的选项来更改搜寻空间。

可以选择的选项取决于搜寻器类型。对于某些选项，如将数据源添加至集合，将打开用于该搜寻器类型的向导以帮助您更改搜寻空间。

3. 要使更改生效，请停止并重新启动搜寻器。

删除搜寻器

删除搜寻器时，将从企业搜索系统中除去与该搜寻器相关的所有信息。但是，在重组索引之前，该搜寻器先前搜寻到的信息将继续存在于索引中。

在开始之前

要删除搜寻器，您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。


关于本任务

删除搜寻器可能是一个相当耗时的过程。在确认要删除搜寻器之后，系统将删除系统中与该搜寻器相关的所有数据。

提示： 由于此任务需要一些时间才能完成，所以可能会看到一条消息指示请求的操作已超时，即使该进程仍在后台运行亦如此。要确定该任务是否已完成，请间歇地单击管理控制台中的**刷新**（不要单击 Web 浏览器中的**刷新**）。当搜寻器名不再出现在搜寻器列表中时，表示删除过程已完成。

过程

要删除搜寻器：

1. 编辑集合，然后选择搜寻页。
2. 找到要删除的搜寻器，然后单击  **删除**。

Content Edition 搜寻器

要在企业搜索集合中包括 IBM WebSphere Information Integrator Content Edition 存储库，必须配置 Content Edition 搜寻器。

可以使用 Content Edition 搜寻器搜寻 Documentum、FileNet Panagon Content Services、FileNet P8 Content Manager、Hummingbird Document Management (DM)、OpenText Livelink 和 Portal Document Manager (PDM) 存储库。

配置搜寻器时，指定关于搜寻器如何搜寻空间中搜寻所有存储库的选项。还需要选择要在每个存储库中搜寻的项类。

要创建或更改 Content Edition 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有存储库。
- 指定搜寻器是使用直接方式还是服务器方式来访问存储库。对于服务器方式，还必须指定一些信息以使搜寻器能够访问 Web 应用程序服务器。
- 选择要搜寻的存储库。
- 指定使搜寻器能够访问所选存储库中的内容的用户标识和密码。
- 设置用于搜寻存储库的时间表。
- 选择在每个存储库中要搜寻的项类。

- 指定用于使项类的属性可供搜索的选项。例如，可以将某些类型的文档排除在搜寻空间之外，也可以指定想要搜寻特定版本的存储库。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据访问控制表或安全性令牌来强制实施访问控制。

对于 Documentum、FileNet Panagon Content Services 和 Portal Document Manager 项类来说，还可以选择一个选项来在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。（此类当前凭证验证功能不适用于其他存储库类型。）

创建搜寻器时，可以单击[帮助](#)以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

以服务器方式访问 WebSphere II Content Edition 存储库

可以将 Content Edition 搜寻器配置成以服务器方式访问存储库。

在服务器方式下，搜寻器访问数据时使用的 WebSphere Information Integrator Content Edition 连接器作为企业应用程序安装在 WebSphere Application Server 上，搜寻器通过服务器访问存储库。这种方法使您能够利用 J2EE 应用程序服务器环境。

在将搜寻器配置成以服务器方式访问 WebSphere Information Integrator Content Edition 存储库前，必须在搜寻器服务器上运行一个脚本。此脚本是与 WebSphere Information Integrator OmniFind Edition 一起提供的，它使 Content Edition 搜寻器能够访问服务器上的存储库。

在使用企业搜索管理控制台来将 Content Edition 搜寻器配置成使用服务器方式之前，根据适合于环境与否来完成下列任务：

- 第 37 页的『在 UNIX 上为 WebSphere II Content Edition 配置搜寻器服务器』。
- 第 38 页的『在 Windows 上为 WebSphere II Content Edition 配置搜寻器服务器』。

以直接方式访问 WebSphere II Content Edition 存储库

可以将 Content Edition 搜寻器配置成以直接方式访问存储库。

在直接方式下，搜寻器使用 WebSphere Information Integrator Content Edition 连接器，该连接器是在安装 WebSphere II OmniFind Edition 时安装在搜寻器服务器上的。

过程

要对系统进行配置以便能够以直接方式访问存储库：

1. 确认 `iice_install_root/bin/config.sh` 文件（在 UNIX 上）或 `iice_install_root/bin/config.bat` 文件（在 Microsoft Windows 上）中的 `VBR_HOME` 和 `JAVA_HOME` 环境变量指定正确的目录。
2. 要将 WebSphere Information Integrator Content Edition 管理控制台配置成以直接方式运行，请在 `iice_install_root/bin/Admin.bat` 文件（在 UNIX 上）或 `iice_install_root\bin\Admin.bat` 文件（在 Windows 上）中添加 `vbr.as.operationMode=direct` Java 系统属性。

3. 以直接方式启动 WebSphere Information Integrator Content Edition 管理控制台并为 WebSphere II OmniFind Edition 搜寻器服务器配置连接器。（请参阅 WebSphere Information Integrator Content Edition 文档以获取指示信息。）
4. 在使用 WebSphere II OmniFind Edition 管理控制台来配置 Content Edition 搜寻器时，请选择直接方式选项。

在 UNIX 上为 WebSphere II Content Edition 配置搜寻器服务器

如果将 WebSphere II OmniFind Edition 安装在运行 IBM AIX®、Linux® 或 Solaris 操作环境的计算机上，并且将 Content Edition 搜寻器配置为以服务器方式访问存储库，则必须运行一个脚本来配置搜寻器服务器。该脚本使 Content Edition 搜寻器能够访问 WebSphere Information Integrator Content Edition 存储库。

关于本任务

Content Edition 搜寻器使用 WebSphere Information Integrator Content Edition 的 Java 库作为 Java 客户机。在服务器方式下，这些 Java 库需要 WebSphere Application Server 的与 EJB 相关的 Java 库。为了确保 Content Edition 搜寻器可以使用 Java 库，在安装 WebSphere Application Server 之后，必须在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

WebSphere Information Integrator Content Edition 是在安装 WebSphere II OmniFind Edition 时安装在搜寻器服务器上的。为了能够以服务器方式使用 Content Edition 搜寻器，必须将 `vbr_access_services.jar` 文件从 WebSphere Information Integrator Content Edition 服务器复制到搜寻器服务器。

过程

要配置搜寻器服务器以使其能够搜寻 WebSphere Information Integrator Content Edition 存储库：

1. 如果 WebSphere II OmniFind Edition 安装在多服务器配置中，请安装并绑定 WebSphere Application Server Java 库。
2. 在搜寻器服务器上运行 Content Edition 搜寻器的设置脚本：
 - a. 作为企业搜索管理员登录。
 - b. 启动安装在 `$ES_INSTALL_ROOT/bin` 目录中的以下脚本，然后回答提示：

```
escrvbr.sh
```

3. 停止并重新启动企业搜索系统，包括企业搜索公共通信层（CCL）上的所有会话：

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

4. 将 `vbr_access_services.jar` 文件从 WebSphere Information Integrator Content Edition 服务器复制到搜寻器服务器。

复制自：

`vbr_access_services.jar` 文件位于以下缺省位置中：

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

`was_install_root` 是 WebSphere Application Server 安装目录, `server_name` 是对服务器指定的名称, `application_name` 是对 WebSphere Application Server 中的 WebSphere Information Integrator Content Edition 应用程序指定的名称。

复制到:

搜寻器服务器上的目标目录是 `iice_install_root/lib`, 其中 `iice_install_root` 是搜寻器服务器上的 WebSphere Information Integrator Content Edition 安装目录。

在 Windows 上为 WebSphere II Content Edition 配置搜寻器服务器

如果将 WebSphere II OmniFind Edition 安装在 Microsoft Windows 计算机上, 并且将 Content Edition 搜寻器配置为以服务器方式访问存储库, 则必须运行一个脚本来配置搜寻器服务器。该脚本使 Content Edition 搜寻器能够访问 WebSphere Information Integrator Content Edition 存储库。

关于本任务

Content Edition 搜寻器使用 WebSphere Information Integrator Content Edition 的 Java 库作为 Java 客户机。在服务器方式下, 这些 Java 库需要 WebSphere Application Server 的与 EJB 相关的 Java 库。为了确保 Content Edition 搜寻器可以使用 Java 库, 在安装 WebSphere Application Server 之后, 必须在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

WebSphere Information Integrator Content Edition 是在安装 WebSphere II OmniFind Edition 时安装在搜寻器服务器上的。为了能够以服务器方式使用 Content Edition 搜寻器, 必须将 `vbr_access_services.jar` 文件从 WebSphere Information Integrator Content Edition 服务器复制到搜寻器服务器。

过程

要配置搜寻器服务器以使其能够搜寻 WebSphere Information Integrator Content Edition 存储库:

1. 如果 WebSphere II OmniFind Edition 安装在多服务器配置中, 请安装并绑定 WebSphere Application Server Java 库。
2. 在搜寻器服务器上运行 Content Edition 搜寻器的设置脚本:
 - a. 使用企业搜索管理员标识登录 (此用户标识是在安装 WebSphere II OmniFind Edition 时指定的)。
 - b. 启动安装在 `%ES_INSTALL_ROOT%\bin` 目录中的以下脚本, 然后回答提示:

```
escrvbr.vbs
```

3. 停止并重新启动企业搜索系统, 包括企业搜索公共通信层 (CCL) 上的所有会话:
 - a. 在命令提示符处, 停止企业搜索系统:

```
esadmin stop
```

- b. 选择开始 → 程序 → 管理工具 → 服务, 然后重新启动 IBM WebSphere Information Integrator OmniFind Edition 服务。

c. 在命令提示符处，启动企业搜索系统：

```
esadmin start
```

4. 将 `vbr_access_services.jar` 文件从 WebSphere Information Integrator Content Edition 服务器复制到搜寻器服务器。

复制自：

`vbr_access_services.jar` 文件位于以下缺省位置中：

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

was_install_root 是 WebSphere Application Server 安装目录，*server_name* 是对服务器指定的名称，*application_name* 是对 WebSphere Application Server 中的 WebSphere Information Integrator Content Edition 应用程序指定的名称。

复制到：

搜寻器服务器上的目标目录是 `iice_install_root\lib`，其中 *iice_install_root* 是搜寻器服务器上的 WebSphere Information Integrator Content Edition 安装目录。

DB2 搜寻器

使用 DB2 搜寻器来将 IBM DB2 通用数据库数据库包括在集合中。也可以使用 DB2 搜寻器来包括为 IBM DB2 通用数据库 z/OS 版、IBM Informix、Oracle 和 Microsoft SQL Server 数据库创建的呢称表。

必须为所要搜寻的每台数据库服务器配置独立的搜寻器。配置搜寻器时，指定关于搜寻器如何搜寻同一服务器上的所有数据库的选项。还需要选择要在每个数据库中搜寻的特定表。

选择搜寻的表应该是数据库表、呢称表或视图。DB2 搜寻器不支持连接表。

事件发布

如果使用 WebSphere Information Integrator Event Publisher Edition，并且已使所要搜寻的数据库与发布队列映射相关联，DB2 搜寻器就可以使用那些映射来搜寻数据库表更新。

发布队列映射标识一个 WebSphere MQ 队列，该队列在数据库表更新被发布时接收 XML 消息。搜寻器侦听该队列以获取关于这些已发布的事件的信息，并且在那些表被更新时更新搜寻空间（搜寻器第一次搜寻一个表时，搜寻器会搜寻所有文档）。

与搜寻器根据搜寻器时间表搜寻的文档相比，事件发布功能使新文档和更改过的文档能够更快地可供搜索。

如果已将某些表或全部表配置为使用事件发布功能，则可以指定一些信息以使搜寻器能够在配置该搜寻器时访问 WebSphere MQ 和发布队列映射。

还必须确保在所要搜寻的服务器上配置了 WebSphere MQ 和 WebSphere Information Integrator Event Publisher Edition，并且在搜寻器服务器上配置了 WebSphere MQ 客户机模块。完成下列任务以对 DB2 搜寻器使用事件发布功能：

- 第 42 页的『配置 WebSphere MQ for DB2 搜寻器』。

- 『为 DB2 搜寻器配置 WebSphere Information Integrator Event Publisher Edition』.
- 第 44 页的『将 UNIX 上的搜寻器服务器配置为使用事件发布功能』.
- 第 45 页的『将 Windows 上的搜寻器服务器配置为使用事件发布功能』.

配置概述

要创建或更改 DB2 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何搜寻特定数据库服务器上的所有数据库。
- 指定关于要搜寻的数据库类型的信息。

如果计划搜寻未在本地数据库服务器上编目的远程数据库，则在可以使用 DB2 搜寻器搜寻那些数据库之前，必须先远程服务器上启动 DB2 管理服务器。在配置该搜寻器时，还必须指定远程数据库服务器的主机名和端口。

- 选择要搜寻的数据库。
- 指定用户标识和密码，它们使搜寻器能够访问那些使用访问控制的数据库。
- 设置用于搜寻数据库的时间表。
- 选择要在每个数据库中搜寻的表。

警告： 为了提高发现进程的性能（以及为了防止搜寻器配置进程超时），仅当数据库包含数目不多的表或者表包含数目不多的列时，才应该选择搜寻所有的表。如果现在选择了一些要搜寻的表，则以后可以编辑搜寻空间并将更多的表添加到集合中。

- 选择表（当在事件发布队列中发布对这些表所作的更新时，将要搜寻这些表），然后指定一些信息以使搜寻器能够访问事件发布队列。
- 指定用于使特定表中的列可供搜索的选项。例如，可以使某些列能在参数查询中使用，也可以指定将在搜索结果中返回哪些列。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

创建搜寻器时，可以单击**帮助**以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

为 DB2 搜寻器配置 WebSphere Information Integrator Event Publisher Edition

在将 DB2 搜寻器配置为使用事件发布功能前，请确保在要搜寻的服务器上配置了 IBM WebSphere Information Integrator Event Publisher Edition。

关于本任务

在配置 WebSphere Information Integrator Event Publisher Edition 以便与 DB2 搜寻器配合使用时，请使用下列准则：

- 必须选择源表中要发布的已更改的列和未更改的列。

- 必须选择源表中要发布的已删除的行。
- 不能在多个数据库之间共享事件发布队列。
- 一个数据库可以有多个队列映射和队列。
- 表应该有一个与一个发布队列映射相关联的 XML 发布。（表不应该有多个与单个发布队列映射相关联的 XML 发布。如果每个 XML 发布都与不同的发布队列映射相关联，表就可以有多个 XML 发布。）

过程

请完成下列数据库服务器配置步骤以使 DB2 搜寻器能够访问事件发布队列中发布的表更新。（要获取有关这些步骤的帮助，请参阅 WebSphere Information Integrator Publisher Edition 文档。）

1. 在要搜寻的数据库服务器上安装 WebSphere Information Integrator Event Publisher Edition。
2. 启动复制中心启动板：

操作系统	命令
UNIX	db2rc
Windows	选择开始 → IBM DB2 复制中心

3. 创建 Q Capture 控制表：
 - a. 选择**事件发布**作为启动板视图，选择**创建 Q Capture 控制表**，然后单击**下一步**。
 - b. 在 **Q Capture 服务器**字段中，从可用数据库服务器列表中选择要用作 Q Capture 服务器的服务器，然后单击**确定**。
 - c. 指定有权访问所选 Q Capture 服务器的用户标识和密码。更改 Q Capture 模式或接受缺省模式名，然后单击**下一步**。
 - d. 指定在此数据库服务器上配置 WebSphere MQ 时指定的队列管理器、管理队列和重新启动队列的名称，然后单击**下一步**。
 - e. 单击**完成**。在带有消息和 SQL 脚本的页显示后，单击**关闭**。
 - f. 对于处理选项，选择**立即运行**，然后单击**确定**。在指示 SQL 脚本已完成的消息显示之后，单击**关闭**。
4. 创建 XML 发布：
 - a. 在复制中心启动板中，选择**事件发布**作为启动板视图，选择**创建 XML 发布**，然后单击**下一步**。
 - b. 在“开始”页中，单击**下一步**。
 - c. 在“服务器和队列映射”页中，确认 Q Capture 服务器和 Q Capture 模式都正确，然后单击**发布队列映射**字段旁边的按钮，接着单击**新建**以创建发布队列映射。
 - d. 在“常规”页中，输入队列映射的名称。
 - e. 在“属性”页中，指定发送队列的名称（例如在此服务器上配置 WebSphere MQ 时指定的数据队列名），选择**行操作**或**事务**作为消息内容类型，清除用于发送脉冲信号消息和添加 JMS 消息头的复选框，然后单击**确定**。
 - f. 在带有消息和 SQL 脚本的页显示后，单击**关闭**。
 - g. 对于处理选项，选择**立即运行**，然后单击**确定**。在指示 SQL 脚本已完成的消息显示之后，单击**关闭**。

- h. 在“选择发布队列映射”页上，选择已创建的队列映射，然后单击**确定**。
 - i. 在“服务器和队列映射”页上，确认队列映射名正确，然后单击**下一步**。
 - j. 在“源表”页上，单击**添加**，单击**检索全部**，选择要支持事件发布的表，单击**确定**，然后单击**下一步**。
 - k. 在“列和行”页上，选择要让 DB2 搜寻器搜寻的列（或所有列）并选择键列。在选择所要搜寻的行（或所有行）的页上，选择用于发布源表删除内容的选项。在完成配置这些选项之后，单击**下一步**。
 - l. 在“消息内容”页上，选择“既包括已更改的列也包括未更改的列来作为列数据”选项，并选择“仅新数据值”选项。确保选中了用于自动启动 XML 发布的复选框，然后单击**下一步**。
 - m. 在“查看并完成 XML 发布”页上，单击**下一步**。
 - n. 在“总结”页上，单击**完成**。在带有消息和 SQL 脚本的页显示后，单击**关闭**。
 - o. 对于处理选项，选择**立即运行**，然后单击**确定**。在指示 SQL 脚本已完成的消息显示之后，单击**关闭**。
5. 启动 Q Capture 服务器：
- a. 关闭“复制中心启动板”，然后启动“复制中心”。
 - b. 在对象树中，单击 **Q Replication** → **定义** → **Q Capture 服务器**。
 - c. 右键单击已配置的 Q Capture 服务器的图标，然后选择**启用 Q Replication 的数据库**。
 - d. 在警告消息显示后，单击**确定**。
 - e. 在带有 DB2 消息的页显示后，单击**关闭**。
 - f. 在对象树中，右键单击 Q Capture 服务器的图标，然后选择**启动 Q Capture 程序**。
 - g. 对于处理选项，选择**立即运行**，指定系统名称、DB2 用户的用户标识和密码、日志存储目录的路径以及 DB2 实例名，然后单击**确定**。
 - h. 在指示已提交请求的消息显示后，单击**关闭**。
 - i. 在对象树中，右键单击 Q Capture 服务器的图标，然后选择**检查状态**。

这就显示了 Q Capture 服务器状态。如果发生了错误，就会有状态消息指示服务器可能已关闭。要查看日志并确定任何错误的原因，请在命令行上输入以下命令：

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

配置 WebSphere MQ for DB2 搜寻器

在将 DB2 搜寻器配置为使用事件发布功能之前，请确保在搜寻器将要侦听的服务器上配置了 IBM WebSphere MQ。

在开始之前

确保在目标数据库服务器上安装了 DB2 UDB、WebSphere Information Integrator Event Publisher Edition 和 WebSphere MQ。

限制

如果目标数据库服务器安装在 Linux 计算机上，则所有 DB2 通用数据库 (DB2 UDB) 用户、WebSphere MQ 用户和 WebSphere II OmniFind Edition 用户都必须设置以下环境变量：

```
export LD_ASSUME_KERNEL=2.4.19
```

此环境变量允许从任何执行安装、发出 WebSphere MQ 控制命令或运行 WebSphere MQ 应用程序的 shell 中导出 LinuxThread 线程技术实施。WebSphere MQ 要求导出此环境变量。

关于本任务

DB2 搜寻器支持以客户机方式连接至 WebSphere MQ 服务器。搜寻器侦听事件发布队列中发布的 XML 消息。搜寻器无法侦听通过多个队列传输的 XML 消息。

配置 WebSphere MQ 后，DB2 搜寻器将使用队列管理器名、队列名、服务器主机名、服务器端口号和服务器通道名来从发布队列中获取 XML 消息。搜寻器将对消息进行解析并使用有关已更新的表的信息来更新搜寻空间。

过程

请完成下列数据库服务器配置步骤以使 DB2 搜寻器能够侦听事件发布队列。（要获取有关这些步骤的帮助，请参阅 WebSphere MQ 文档。）

1. 作为 WebSphere MQ 管理员角色登录，然后输入下列命令以创建队列管理器和队列。

a. 在命令行上，输入以下命令：

```
crtmqm QM1
```

b. 在设置已完成消息显示后，输入以下命令：

```
strmqm QM1
```

c. 在已启动"QM1"消息显示后，输入以下命令：

```
runmqsc QM1
```

d. 在正在启动队列管理器 QM1 的 MQSC 消息显示后，输入以下命令以创建管理队列：

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

e. 在已创建 WebSphere MQ 队列消息显示后，输入以下命令以创建重新启动队列：

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

f. 在已创建 WebSphere MQ 队列消息再次显示后，输入以下命令以创建数据队列：

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

g. 在已创建 WebSphere MQ 队列消息再次显示后，输入以下命令以退出：

```
end
```

2. 输入以下命令以便在数据库服务器上启动 MQ 侦听器（在创建使用事件发布功能的 DB2 搜寻器时，MQ 侦听器必须正在运行）。在本示例中，1414 是服务器端口号，并且使用了缺省通道 SYSTEM.DEF.SVRCONN：

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

3. 输入下列命令以授权 DB2 UDB 用户通过用于发布事件的消息队列接口 (MQI) 来访问队列管理器和队列 (在本示例中, 用户标识是 db2inst1):

```
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. 对用来创建和运行使用事件发布功能的 DB2 搜寻器的用户标识输入下列命令。这些命令将授权该用户标识通过用于发布事件的消息队列接口 (MQI) 来访问队列管理器和队列。在本示例中, 用户标识是 esuser:

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

将 UNIX 上的搜寻器服务器配置为使用事件发布功能

如果将 WebSphere II OmniFind Edition 安装在运行 IBM AIX、Linux 或 Solaris 操作环境的计算机上, 并且将 DB2 搜寻器配置为使用事件发布功能, 则必须运行一个脚本来配置搜寻器服务器。该脚本使该搜寻器能够访问 WebSphere MQ 队列管理器和队列。

关于本任务

DB2 搜寻器使用 WebSphere MQ 5.3 modules for Java Messaging 来访问 WebSphere MQ 队列管理器和队列。必须在搜寻器服务器上安装这些模块。

要确保 DB2 搜寻器能够使用事件发布功能, 在安装 WebSphere MQ 模块后, 还必须在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

过程

要将搜寻器服务器配置为使用事件发布功能:

1. 在搜寻器服务器上安装 WebSphere MQ 5.3 modules for Java Messaging:
 - a. 作为 root 用户登录并输入以下命令:

```
export LD_ASSUME_KERNEL=2.4.19
```

- b. 插入 WebSphere MQ CD。
- c. 切换到 MQ modules for Java Messaging 所在的目录。
- d. 输入以下命令以安装模块:

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

2. 在搜寻器服务器上运行 DB2 搜寻器的设置脚本:
 - a. 作为企业搜索管理员登录 (此用户标识是在安装 WebSphere II OmniFind Edition 时指定的)。
 - b. 启动安装在 \$ES_INSTALL_ROOT/bin 目录中的以下脚本, 然后回答提示:

```
escrdb2.sh
```

3. 停止并重新启动企业搜索系统, 包括企业搜索公共通信层 (CCL) 上的所有会话:


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

将 Windows 上的搜寻器服务器配置为使用事件发布功能

如果将 WebSphere II OmniFind Edition 安装在运行 Microsoft Windows 的计算机上，并且将 DB2 搜寻器配置为使用事件发布功能，则必须运行一个脚本以配置搜寻器服务器。该脚本使该搜寻器能够访问 WebSphere MQ 队列管理器和队列。

关于本任务

DB2 搜寻器使用 WebSphere MQ 5.3 modules for Java Messaging 来访问 WebSphere MQ 队列管理器和队列。必须在搜寻器服务器上安装这些模块。

要确保 DB2 搜寻器能够使用事件发布功能，在安装 WebSphere MQ 模块后，还必须在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

过程

要将搜寻器服务器配置为使用事件发布功能：

1. 在搜寻器服务器上安装 WebSphere MQ 5.3 modules for Java Messaging:
 - a. 插入 WebSphere MQ CD。
 - b. 启动 WebSphere MQ 安装程序。
 - c. 在“选择产品功能部件”窗口中，选择 **Java 消息传递** 安装选项。
2. 在搜寻器服务器上运行 DB2 搜寻器的设置脚本：
 - a. 使用企业搜索管理员标识登录（此用户标识是在安装 WebSphere II OmniFind Edition 时指定的）。
 - b. 启动安装在 %ES_INSTALL_ROOT%\bin 目录中的以下脚本，然后回答提示：

```
escrdb2.vbs
```
3. 停止并重新启动企业搜索系统，包括企业搜索公共通信层（CCL）上的所有会话：
 - a. 在命令提示符处，停止企业搜索系统：

```
esadmin stop
```
 - b. 选择 **开始** → **程序** → **管理工具** → **服务**，然后重新启动 IBM WebSphere Information Integrator OmniFind Edition 服务。
 - c. 在命令提示符处，启动企业搜索系统：

```
esadmin start
```

DB2 Content Manager 搜寻器

要在企业搜索集合中包括 IBM DB2 Content Manager 项类型，必须配置 DB2 Content Manager 搜寻器。

搜寻器服务器配置

在可以搜寻 DB2 Content Manager 服务器之前，必须在搜寻器服务器上运行一个脚本。此脚本是与 WebSphere Information Integrator OmniFind Edition 一起提供的，它使 DB2 Content Manager 搜寻器能与 DB2 Content Manager 服务器通信。

在使用企业搜索管理控制台来配置 DB2 Content Manager 搜寻器之前，根据适合于环境与否来完成下列任务：

- 『在 UNIX 上为 DB2 Content Manager 配置搜寻器服务器』。
- 第 48 页的『在 Windows 上为 DB2 Content Manager 配置搜寻器服务器』。

配置概述

可以使用 DB2 Content Manager 搜寻器搜寻任意数目的 DB2 Content Manager 服务器。配置该搜寻器时，指定关于搜寻器如何在搜寻空间中搜寻所有 DB2 Content Manager 服务器的选项。还需要选择要在每台服务器上搜寻的特定项类型。

要创建或更改 DB2 Content Manager 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有 DB2 Content Manager 服务器上的所有项类型。
- 选择要搜寻的 DB2 Content Manager 服务器。
- 指定使搜寻器能够访问 DB2 Content Manager 服务器上的内容的用户标识和密码。
- 设置用于搜寻服务器的时间表。
- 选择要在每台 DB2 Content Manager 服务器上搜寻的项类型。
- 指定用于使某些项类型中的属性可供搜索的选项。例如，可以将某些类型的文档排除在搜寻空间之外，并指定在搜索结果中返回哪些属性。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

创建搜寻器时，可以单击帮助以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

在 UNIX 上为 DB2 Content Manager 配置搜寻器服务器

如果将 WebSphere II OmniFind Edition 安装在运行 IBM AIX、Linux 或 Solaris 操作环境的计算机上，则必须运行一个脚本来配置搜寻器服务器。该脚本使 DB2 Content Manager 搜寻器能够与 IBM DB2 Content Manager 服务器通信。

关于本任务

DB2 Content Manager 搜寻器使用 Java connector for DB2 Content Manager V8 来访问 DB2 Content Manager 服务器。通过在搜寻器服务器上安装 IBM DB2 Information Integrator for Content V8.2 或更新版本来安装此连接器。要确保 DB2 Content Manager 搜寻器能够与 DB2 Content Manager 配合工作，应在安装连接器之后，在搜寻器服务器上运行 WebSphere II OmniFind Edition 提供的一个设置脚本。

过程

要配置搜寻器服务器以使其能够搜寻 DB2 Content Manager 服务器:

1. 在搜寻器服务器上安装 Java connector for DB2 Content Manager V8:
 - a. 在搜寻器服务器上, 作为 root 用户登录:

```
su - root
```
 - b. 运行 db2profile 文件。例如:

```
./home/db2inst/sqlllib/db2profile
```
 - c. 导出 JAVAHOME 环境变量。例如:

```
export JAVAHOME=/usr/IBMJava2-141
```
 - d. 将 Java 目录添加到 PATH 环境变量:

```
export PATH=$PATH:$JAVAHOME/bin
```
 - e. 插入 DB2 Information Integrator for Content 安装 CD 并运行安装向导。
 - f. 在“组件选择”窗口中, 执行下列操作。(如果使用的是 Information Integrator for Content V8.3, 则可以查看带有“定制安装”选项的“组件选择”窗口。)
 - 1) 从组件列表中选择本地连接器, 然后从子组件列表中选择 **Content Manager V8 连接器**。
 - 2) 从组件列表中选择连接器工具箱和样本, 然后从子组件列表中选择 **Content Manager V8 连接器**。
 - g. 指定 DB2 Content Manager 库的数据库名称、用户名和密码, 并接受余下窗口的缺省设置。
2. 在搜寻器服务器上, 使用属于 DB2 管理组中的用户标识进行登录。
3. 编目远程 DB2 Content Manager 库服务器数据库, 然后验证搜寻器服务器是否能够连接至 DB2 Content Manager 服务器:

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

4. 可选: 作为 root 用户登录并测试数据库连接:

```
./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. 在搜寻器服务器上运行 DB2 Content Manager 搜寻器的设置脚本:

- a. 切换到 ES_INSTALL_ROOT/bin 目录:

```
cd $ES_INSTALL_ROOT/bin
```

- b. 启动以下脚本并回答提示:

```
escrcm.sh
```

6. 停止并重新启动企业搜索系统, 包括企业搜索公共通信层 (CCL) 上的所有会话:

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

在 Windows 上为 DB2 Content Manager 配置搜寻器服务器

如果将 WebSphere II OmniFind Edition 安装在 Microsoft Windows 计算机上，则必须运行一个脚本来配置搜寻器服务器。该脚本使 DB2 Content Manager 搜寻器能够与 IBM DB2 Content Manager 服务器通信。

关于本任务

DB2 Content Manager 搜寻器使用 Java connector for DB2 Content Manager V8 来访问 DB2 Content Manager 服务器。通过在搜寻器服务器上安装 IBM DB2 Information Integrator for Content V8.2 或更新版本来安装此连接器。要确保 DB2 Content Manager 搜寻器能够与 DB2 Content Manager 配合工作，应在安装连接器之后，在搜寻器服务器上运行 WebSphere II OmniFind Edition 提供的一个设置脚本。

过程

要配置搜寻器服务器以使其能够搜寻 DB2 Content Manager 服务器：

1. 在搜寻器服务器上安装 Java connector for DB2 Content Manager V8:
 - a. 插入 DB2 Information Integrator for Content 安装 CD。 安装程序自动启动。

DB2 Content Manager Enterprise Information Portal 安装向导将打开。
 - b. 在“组件选择”窗口中，执行下列操作。（如果使用的是 Information Integrator for Content V8.3，则可以查看带有“定制安装”选项的“组件选择”窗口。）
 - 1) 从组件列表中选择本地连接器，然后从子组件列表中选择 **Content Manager V8 连接器**。
 - 2) 从组件列表中选择连接器工具箱和样本，然后从子组件列表中选择 **Content Manager V8 连接器**。
 - c. 指定 DB2 Content Manager 库的数据库名称、用户名和密码，并接受余下窗口的缺省设置。
2. 编目远程 DB2 Content Manager 库服务器数据库，然后验证搜寻器服务器是否能够连接至 DB2 Content Manager 服务器。 在搜寻器服务器上，在命令提示符处输入下列命令：

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

3. 可选：通过打开命令提示符并输入下列命令来测试数据库连接：

```
cmbenv81.bat
cd Information_Integrator_for_Content_install_directory\samples\java\icm
javac *.java
java SConnectDisconnect ICMdatabase_name CMadmin_ID CMadmin_password
```

4. 在搜寻器服务器上运行 DB2 Content Manager 搜寻器的设置脚本：
 - a. 切换到 ES_INSTALL_ROOT\bin 目录：

```
cd %ES_INSTALL_ROOT%\bin
```

- b. 启动以下脚本并回答提示:

```
escrcm.vbs
```

5. 停止并重新启动企业搜索系统，包括企业搜索公共通信层（CCL）上的所有会话:

- a. 在命令提示符处，停止企业搜索系统:

```
esadmin stop
```

- b. 选择开始 → 程序 → 管理工具 → 服务，然后重新启动 IBM WebSphere Information Integrator OmniFind Edition 服务。

- c. 在命令提示符处，启动企业搜索系统:

```
esadmin start
```

Domino Document Manager 搜寻器

要在企业搜索集合中包括 Domino Document Manager 库和文件柜，必须配置 Domino Document Manager 搜寻器。

搜寻器服务器配置

如果计划搜寻的 Domino Document Manager 服务器使用 Notes 远程过程调用（NRPC）协议，则必须在搜寻器服务器上运行一个脚本。此脚本是与 WebSphere Information Integrator OmniFind Edition 一起提供的，它使 Domino Document Manager 搜寻器能够与使用 NRPC 的服务器通信。

如果计划搜寻的 Domino Document Manager 服务器使用 Domino 因特网 ORB 间协议（DIIOP），则不需要在搜寻器服务器上运行设置脚本。但是，必须对 Domino Document Manager 服务器进行配置以使 Domino Document Manager 搜寻器能够访问该服务器。

如果 WebSphere II OmniFind Edition 安装在 IBM AIX 系统上，则必须确保已在搜寻器服务器上安装 I/O 完成端口模块并且该模块可用。

在使用企业搜索管理控制台来配置 Domino Document Manager 搜寻器之前，根据适合于环境与否来完成下列任务:

- 第 54 页的『在 UNIX 上配置搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 56 页的『配置 Windows 上的搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 57 页的『配置使用 DIIOP 协议的服务器』。
- 第 58 页的『在 AIX 上配置 I/O 完成端口以搜寻 Lotus Domino 源』。

配置概述

可以使用 Domino Document Manager 搜寻器搜寻任意数目的 Domino Document Manager 库。在创建搜寻器时，应该只从一台 Domino Document Manager 服务器中选择要搜寻的库。此后，在编辑搜寻空间时，可以添加另一台要包括在同一搜寻空间中的 Domino Document Manager 服务器上的文档。在创建或编辑搜寻器时，可以指定是搜寻所要搜寻的库中的所有文件柜还是要搜寻特定的文件柜。

要创建或更改 Domino Document Manager 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有文档。
- 标识 Domino Document Manager 服务器和通信协议。
- 选择要搜寻的库。
- 设置用于搜寻库的时间表。
- 选择要搜寻的文档。搜寻器可以搜寻库中的所有文件夹，也可以只搜寻所选文件夹中的文档。
- 指定用于使各种库和文件夹中的字段可供搜索的选项。例如，可以从搜寻空间中排除某些字段，也可以指定用于搜索附件的选项。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

也可以选择一个选项以便在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。

创建搜寻器时，可以单击帮助以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

Exchange Server 搜寻器

要在企业搜索集合中包括 Microsoft Exchange Server 公共文件夹，必须配置 Exchange Server 搜寻器。

可以使用 Exchange Server 搜寻器搜寻 Exchange Server 公共文件夹服务器上的任意数目的文件夹和子文件夹。创建搜寻器时，选择要在公共文件夹服务器上搜寻的内容。以后，可以编辑搜寻空间以从另一台公共文件夹服务器添加内容。

要创建或更改 Exchange Server 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有服务器上的子文件夹。
- 指定有关要搜寻的 Exchange Server 公共文件夹服务器的信息。

必须指定用户标识和密码，以便搜寻器可以访问服务器上的内容。如果服务器使用安全套接字层（SSL）协议，则可以指定使搜寻器能够访问搜寻器服务器上的密钥库文件的选项。

- 设置用于搜寻公共文件夹服务器的时间表。
- 选择要搜寻的文件夹和子文件夹。
- 指定用于使子文件夹中的文档可供搜索的选项。例如，可以将某些类型的文档排除在搜寻空间之外。

- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

创建搜寻器时，可以单击**帮助**以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

验证对安全 Exchange Server 文档的访问权

要使用 Exchange Server 搜寻器搜寻受防火墙保护的文档，您必须验证搜寻器服务器是否能够访问 Microsoft Exchange Server 公共文件夹服务器。

关于本任务

如果搜寻器服务器无法访问安全 Exchange Server 服务器，将接收到该服务器发出的 HTTP 代码 501（未实施）。还可能会看到指示接收到意外 HTTP 响应的消息。

过程

要确保搜寻器服务器能够访问防火墙背后的文档：

1. 在搜寻器服务器上启动 Web 浏览器。
2. 访问所要搜寻的 Exchange Server 公共文件夹服务器的 URL。例如：
`http://exchange.yourCompany.com/public/`
3. 验证是否能够打开 Exchange Server。

如果无法访问 Exchange Server 服务器，请与贵单位的服务器管理员联系。

NNTP 搜寻器

要在企业搜索集合中包括来自 NNTP 新闻组的文章，必须配置 NNTP 搜寻器。

可以使用 NNTP 搜寻器搜寻任意数目的 NNTP 服务器。配置该搜寻器时，选择要在每台服务器上搜寻的新闻组。还可以指定要排除的新闻组的模式。通过这种设计，可以很方便地使搜寻器能搜索服务器上的大部分新闻组，并禁止搜寻器搜寻不想让用户搜索的少量新闻组。

例如，可以指定规则以包括特定 NNTP 服务器上的所有新闻组，然后指定要排除该服务器上名称包含 `private` 字符串的新闻组。

要创建或更改 NNTP 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有的新闻组。
- 指定模式以包括新闻组，并指定模式以从搜寻空间中排除某些新闻组。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

创建搜寻器时，可以单击**帮助**以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

Notes 搜寻器

要在企业搜索集合中包括 IBM Lotus Notes 数据库，必须配置 Notes 搜寻器。

搜寻器服务器配置

如果计划搜寻的 Lotus Notes 服务器使用 Notes 远程过程调用 (NRPC) 协议，则必须在搜寻器服务器上运行一个脚本。此脚本是与 WebSphere Information Integrator OmniFind Edition 一起提供的，它使 Notes 搜寻器能够与使用 NRPC 的服务器通信。

如果计划搜寻的 Lotus Notes 服务器使用 Domino 因特网 ORB 间协议 (DIOP)，则不需要在搜寻器服务器上运行设置脚本。但是，必须配置 Lotus Notes 服务器以使 Notes 搜寻器能够访问服务器。

如果 WebSphere II OmniFind Edition 安装在 IBM AIX 系统上，则必须确保已在搜寻器服务器上安装 I/O 完成端口模块并且该模块可用。

在使用企业搜索管理控制台来配置 Notes 搜寻器之前，根据适合于环境与否来完成下列任务：

- 第 54 页的『在 UNIX 上配置搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 56 页的『配置 Windows 上的搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 57 页的『配置使用 DIOP 协议的服务器』。
- 第 58 页的『在 AIX 上配置 I/O 完成端口以搜寻 Lotus Domino 源』。

文档级安全性

如果已启用集合安全性，并且计划搜寻的 Lotus Notes 服务器使用 NRPC 协议，则必须在搜寻器服务器上配置 Lotus Domino Trusted Server。Trusted Server 用来强制实施文档级访问控制。在使集合可供用户搜索之前，请完成下列任务：

- 第 177 页的『配置 Lotus Domino Trusted Server 以验证用户凭证』。
- 在 WebSphere Application Server 中启用全局安全性，并将搜索应用程序配置为使用安全性。此步骤确保当用户尝试使用搜索应用程序时将被提示指定凭证。然后，搜索服务器可以使用这些凭证来验证每个用户是否有权访问 Lotus Notes 文档。

配置概述

可以使用 Notes 搜寻器搜寻任意数目的标准 Lotus Notes 数据库 (.nsf 文件)。在创建搜寻器时，应只从一台 Lotus Notes 服务器中选择要搜寻的数据库或目录。此后，在编辑搜寻空间时，可从另一想要包括在同一搜寻空间中的 Lotus Notes 服务器上添加文档。在创建或编辑搜寻器时，可指定是想搜寻服务器上的所有数据库或目录，还是想搜寻特定的数据库、视图和文件夹。

要创建或更改 Notes 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有文档。
- 标识 Lotus Notes 服务器主机名、端口和通信协议。
- 选择想要搜寻的数据库或目录。
- 设置用于搜寻数据库或目录的时间表。
- 选择要搜寻的文档。可以搜寻目录中的所有文档、数据库中的所有文档或来自数据库的所选视图和文件夹的文档。
- 指定用于使各种数据库、视图和文件夹中的字段可供搜索的选项。例如，可以从搜寻空间中排除某些字段，也可以指定用于搜索附件的选项。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

也可以选择一个选项以便在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。

创建搜寻器时，可以单击帮助以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

相关概念

第 177 页的『对 Lotus Domino 文档强制实施文档级安全性』

如果要搜寻的 Lotus Notes 服务器使用 Notes 远程过程调用 (NRPC) 协议，则必须对搜寻器服务器进行配置以便可以强制实施文档级访问控制。

相关任务

第 177 页的『配置 Lotus Domino Trusted Server 以验证用户凭证』

要对使用 Notes 远程过程调用 (NRPC) 协议的 Notes 搜寻器搜寻到的文档强制实施安全性，必须将所要搜寻的 Domino 服务器配置成 Lotus Domino Trusted Server。

Lotus Domino 数据库搜寻技巧

在配置 Notes 搜寻器前，请查看 Lotus Domino 数据库的搜寻准则。

- 基于标准模板的 Notes 数据库（例如讨论数据库）是搜寻效果最佳的数据库类型。
- Notes 搜寻器将应用下列字段映射规则：
 - 首先注册 Domino 标准模板中的主要字段名。
 - 将映射规则表中指定的 Notes 字段中的值用作搜索结果中的文档摘要。
 - 在文档摘要中不使用映射规则表中未指定的 Notes 字段中的值。
 - 将映射到 Title 字段的 Notes 字段中的值用作搜索结果中的文档标题。
 - 缺省情况下，下表中的字段映射到搜索字段名：

表 1. 缺省字段映射规则

Notes 数据库字段名	搜索字段名
Title	Title
EventTitle	Title
Subject	Title
Body	Body

表 1. 缺省字段映射规则 (续)

Notes 数据库字段名	搜索字段名
Mission	Body
From	Creator
Author	Creator
Keywords	Categories
Categories	Categories
TeamRoomName	Organization
TeamName	Organization
Department	Organization

- Notes 搜寻器可以搜寻除为显示计算的字段以外的所有字段类型。
- 不搜寻 Notes 表单上的静态文本和图像。
- 在配置搜寻器时，请选中**搜寻全部**复选框以搜寻所有字段并最大程度地增加所要搜寻的字段数据（可以使用**搜寻所有字段**，除了字段来限制要搜寻的字段）。

为了最大程度地减少对非必需字段的搜寻工作，可以对除映射到搜索字段的字段外的所有字段清除**搜寻**复选框。

在 UNIX 上配置搜寻器服务器以搜寻 Lotus Domino 源

如果将 WebSphere II OmniFind Edition 安装在运行 IBM AIX、Linux 或 Solaris 操作环境的计算机上，并且计划搜寻使用 Notes 远程过程调用（NRPC）协议的服务器，则必须运行一个脚本来配置搜寻器服务器。该脚本使 Notes、QuickPlace 和 Domino Document Manager 搜寻器能够与数据库服务器通信。

限制

在同一台计算机上，Domino Server 不能与配置为使用 NRPC 协议的 Notes、QuickPlace 或 Domino Document Manager 搜寻器同时运行。如果在 Domino Server 正在运行时尝试启动这些搜寻器中的任何一个搜寻器，就会发生错误，搜寻器将停止。

关于本任务

使用 NRPC 协议的搜寻器将 Domino 库用作客户机。通过在搜寻器服务器上安装 Lotus Domino Server V6.0.2 或更新版本来安装这些库。要确保搜寻器可以使用 Domino 库，在安装 Domino 库之后，应该在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

过程

要配置搜寻器服务器以使它能够搜寻 Lotus Notes、Lotus QuickPlace 和 Domino Document Manager 服务器：

1. 在搜寻器服务器上创建用户 notes 和组 notes:
 - a. 作为 root 用户登录：

```
su - root
```

- b. 添加用户：


```
useradd notes
```

- c. 为此用户添加密码:

```
passwd notes
```

将提示您更改密码。

2. 在搜寻器服务器上安装 Lotus Domino Server:

- a. 插入 Domino Server CD (V6.0.2 或更新版本), 并挂载它。(如果没有 CD, 可以下载映像。)
- b. 切换到操作系统的文件夹。

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

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

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

- c. 启动安装程序:

```
./install
```

- d. 回答提示并接受缺省值或指定首选安装设置 (如安装目录和数据目录的路径)。

如果需要有关安装 Domino Server 的辅助, 请查阅 Domino 文档。

3. 在搜寻器服务器上运行 WebSphere II OmniFind Edition 提供的设置脚本:

- a. 作为企业搜索管理员登录 (此用户标识是在安装 WebSphere II OmniFind Edition 时指定的)。
- b. 启动安装在 \$ES_INSTALL_ROOT/bin 目录中的以下脚本:

```
escrnote.sh
```

- c. 回答提示:

- 对于以下提示, 如果 Domino Server 安装在缺省目录中, 则回答 Y, 否则回答 N:

```
已找到 Lotus Notes 目录路径 /opt/lotus/notes/latest/linux。  
这是正确的 Lotus Notes 目录路径吗?
```

```
在 AIX 上的缺省路径是
```

```
/opt/lotus/notes/latest/ibmpow。在
```

```
Linux 上的缺省路径是
```

```
/opt/lotus/notes/latest/linux。在
```

```
Solaris 上的缺省路径是
```

```
/opt/lotus/notes/latest/sunspa。
```

- 如果 Domino Server 未安装在搜寻器服务器上的缺省目录中, 则指定 Domino 的安装位置来响应以下提示:

```
输入 Lotus Notes 目录的路径
```

```
例如, 在 Linux 计算机上, 可以指定 /opt/lotus/notes/latest/linux。
```

- 对于以下提示, 如果 Domino Server 数据目录安装在缺省目录中, 则回答 Y, 否则回答 N:

```
已找到 Lotus Notes 数据目录路径 /local/notesdata。  
这是正确的 Lotus Notes 数据目录路径吗?
```

缺省路径是 /local/notesdata。

- 如果 Domino Server 数据目录未部署在搜寻器服务器上的缺省位置中，则指定 Domino 数据路径以响应以下提示：

请输入 Lotus Notes 数据目录的路径。

4. 停止并重新启动企业搜索系统，包括企业搜索公共通信层（CCL）上的所有会话：

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

配置 Windows 上的搜寻器服务器以搜寻 Lotus Domino 源

如果将 WebSphere II OmniFind Edition 安装在 Microsoft Windows 计算机上，并且计划搜寻使用 Notes 远程过程调用（NRPC）协议的服务器，则必须运行一个脚本来配置搜寻器服务器。该脚本使 Notes、QuickPlace 和 Domino Document Manager 搜寻器能够与数据库服务器通信。

限制

在同一台计算机上，Lotus Domino Server 和 Lotus Notes 客户机不能与配置为使用 NRPC 协议的 Notes、QuickPlace 或 Domino Document Manager 搜寻器同时运行。如果在 Domino Server 正在运行时尝试启动这些搜寻器中的任何一个搜寻器，就会发生错误，搜寻器将停止。

关于本任务

使用 NRPC 协议的搜寻器使用 Lotus Notes 客户机库。通过在搜寻器服务器上安装 Lotus Notes R6.0.2 或更新版本来安装这些库。要确保搜寻器可以使用 Lotus Notes 客户机库，在安装 Lotus Notes 客户机库之后，请在搜寻器服务器上运行一个由 WebSphere II OmniFind Edition 提供的设置脚本。

过程

要配置搜寻器服务器以使它能够搜寻 Lotus Notes、Lotus QuickPlace 和 Domino Document Manager 服务器：

1. 在搜寻器服务器上，使用作为 Administrators 组的成员的用户标识登录。确保该用户标识具有安装 Lotus Notes 的权限。
2. 安装 Lotus Notes:
 - a. 插入 Lotus Notes R6.0.2 或更新版本的 CD。（如果没有 CD，可以下载映像。）
 - b. 启动安装程序：setup.exe
 - c. 回答提示并接受缺省值或指定首选安装设置（如安装目录和数据目录的路径）。

如果需要帮助，请查阅 Lotus Notes 文档。

3. 在搜寻器服务器上运行 WebSphere II OmniFind Edition 提供的设置脚本：
 - a. 使用企业搜索管理员标识登录（此用户标识是在安装 WebSphere II OmniFind Edition 时指定的）。
 - b. 启动安装在 %ES_INSTALL_ROOT%\bin 目录中的以下脚本：

escrnote.vbs

c. 回答提示:

- 对于以下提示, 如果 Lotus Notes 安装在缺省目录中, 则回答 Y, 否则回答 N:
已找到 Lotus Notes 目录路径 c:\lotus\notes。
这是正确的 Lotus Notes 目录路径吗?

在 Windows 计算机上, 典型的安装路径是 c:\lotus\notes 或 c:\lotus\domino。

- 如果 Lotus Notes 未安装在搜寻器服务器上的缺省目录中, 则指定 Lotus Notes 的安装位置来响应以下提示:

输入 Lotus Notes 目录的路径

- 对于以下提示, 如果 Lotus Notes 数据目录部署在缺省位置中, 则回答 Y, 否则回答 N:

已找到 Lotus Notes 数据目录路径 c:\lotus\notes\data。
这是正确的 Lotus Notes 数据目录路径吗?

在 Windows 计算机上, 典型的路径是 c:\lotus\notes\data 或 c:\lotus\domino\data。

- 如果 Lotus Notes 数据目录未部署在搜寻器服务器上的缺省位置中, 则指定数据目录路径以响应以下提示:

请输入 Lotus Notes 数据目录的路径。

4. 停止并重新启动企业搜索系统, 包括企业搜索公共通信层 (CCL) 上的所有会话:

- a. 在命令提示符处, 停止企业搜索系统:

```
esadmin stop
```

- b. 选择开始 → 程序 → 管理工具 → 服务, 然后重新启动 IBM WebSphere Information Integrator OmniFind Edition 服务。

- c. 在命令提示符处, 启动企业搜索系统:

```
esadmin start
```

配置使用 DIIOIP 协议的服务器

要搜寻使用 Domino 因特网 ORB 间协议 (DIIOIP) 的服务器, 必须对该服务器进行配置以使 Notes、QuickPlace 和 Domino Document Manager 搜寻器都能够使用该协议。

在开始之前

要搜寻的服务器必须正在运行 DIIOIP 和 HTTP 任务。

过程

要对使用 DIIOIP 协议的服务器进行配置:

1. 配置 server 文档:

- a. 打开所要搜寻的 Lotus Notes、Lotus QuickPlace 或 Domino Document Manager 服务器上的 server 文档。此文档存储在 Domino 目录中。

- b. 在配置页上, 展开**服务器**部分。

- c. 在安全性页上的**可编程性限制**区域中，在下列字段中为环境指定适当的安全性限制：

- 运行受限的 **Lotus 脚本 / Java 代理程序**
- 运行受限的 **Java/Javascript/COM**
- 运行不受限的 **Java/Javascript/COM**

例如，可以指定一个星号 (*) 以允许 Lotus 脚本 / Java 代理程序进行不受限的访问，并指定在 Domino 目录中注册的用户名以便进行 Java/Javascript/COM 限制。

要点：配置的要通过 DIOP 协议来搜寻此服务器的搜寻器必须能够使用这些字段中指定的用户名。

- d. 打开因特网协议页，然后打开 HTTP 页并将**允许 HTTP 客户机浏览数据库**选项设置为**是**。

2. 配置 user 文档：

- a. 打开所要搜寻的 Lotus Notes、Lotus QuickPlace 或 Domino Document Manager 服务器上的 user 文档。此文档存储在 Domino 目录中。
- b. 在基本页中，在**因特网密码**字段中指定密码。

当使用企业搜索管理控制台来配置此服务器的搜寻选项时，请在标识所要搜寻的服务器的那一页上指定此用户标识和密码。搜寻器将使用这些凭证来访问该服务器。

3. 在服务器上重新启动 DIOP 任务。

在 AIX 上配置 I/O 完成端口以搜寻 Lotus Domino 源

在 IBM AIX 系统上，在可以使用 Notes、QuickPlace 或 Domino Document Manager 搜寻器之前，必须安装 I/O 完成端口 (IOCP) 模块并对其进行配置以供搜寻器使用。

关于本任务

如果没有 IOCP 模块，当尝试创建搜寻器时，发现进程就会失败。将显示以下错误消息：

```
FFQM0105E 从服务器接收到错误 -  
消息: FFQG0024E 捕获到意外异常: discover
```

将把以下消息（它包含 `ENOEXEC` 错误）写入 `$ES_NODE_ROOT/logs/system_yyyymmdd.log` 文件。（某些消息文本将换行以提高可读性。）

```
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 捕获到异常，详细信息为"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"  
堆栈跟踪为"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)"

```

过程

要安装 IOCP 模块并确保它正确地安装在搜寻器服务器上:

必须

1. 在搜寻器服务器上从 AIX 产品 CD 安装 IOCP 模块 (bos.iocp.rte)。

安装 IOCP 模块后, 在创建 Notes、QuickPlace 或 Domino Document Manager 搜寻器之前, 请对该模块应用软件修订。要获取指示信息, 请参阅以下链接提供的信息:

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

2. 输入以下命令以确保在搜寻器服务器上安装了 IOCP 模块:

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

lspp 命令的输出应该类似于以下示例:

文件集	级别	状态	描述

路径: /usr/lib/objrepos			
bos.iocp.rte	5.2.0.10	已落实	I/O 完成端口 API

路径: /etc/objrepos			
bos.iocp.rte	5.2.0.10	已落实	I/O 完成端口 API

3. 输入以下命令以确保 IOCP 端口状态为可用:

```
$ lsdev -Cc iocp
```

lsdev 命令的输出应该与以下示例相同:

```
iocp0 可用      I/O 完成端口
```

4. 如果 IOCP 端口状态是已定义, 请将状态更改为可用:

- a. 作为 root 用户登录到搜寻器服务器, 然后发出以下命令:

```
# smit iocp
```

- b. 选择更改 / 显示 I/O 完成端口特征并将在系统重新启动时要配置的状态由已定义更改为可用。
- c. 重新引导搜寻器服务器。
- d. 再次输入 lsdev 命令并确认 IOCP 端口状态是否已更改为可用。

QuickPlace 搜寻器

要在企业搜索集合中包括 Lotus QuickPlace 位置和空间，必须配置 QuickPlace 搜寻器。

搜寻器服务器配置

如果计划搜寻的 QuickPlace 服务器使用 Notes 远程过程调用 (NRPC) 协议，则必须在搜寻器服务器上运行一个脚本。此脚本是与 WebSphere Information Integrator OmniFind Edition 一起提供的，它使 QuickPlace 搜寻器能够与使用 NRPC 的服务器通信。

如果计划搜寻的 QuickPlace 服务器使用 Domino 因特网 ORB 间协议 (DIOP)，则不需要在搜寻器服务器上运行设置脚本。但是，必须对 QuickPlace 服务器进行配置以使 QuickPlace 搜寻器能够访问该服务器。

如果计划搜寻的 QuickPlace 服务器使用轻量级目录访问协议 (LDAP) 服务器，则必须将该 QuickPlace 服务器配置为使用 DIOP 协议 (QuickPlace 搜寻器无法使用 NRPC 协议来搜寻 LDAP 数据)。还必须配置“目录辅助”数据库并将 QuickPlace 服务器配置成将该 LDAP 服务器用作辅助 Domino 服务器。

如果 WebSphere II OmniFind Edition 安装在 IBM AIX 系统上，则必须确保已在搜寻器服务器上安装 I/O 完成端口模块并且该模块可用。

在使用企业搜索管理控制台来配置 QuickPlace 搜寻器之前，根据适合于环境与否来完成下列任务：

- 第 54 页的『在 UNIX 上配置搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 56 页的『配置 Windows 上的搜寻器服务器以搜寻 Lotus Domino 源』。
- 第 57 页的『配置使用 DIOP 协议的服务器』。
- 第 61 页的『配置 QuickPlace 服务器以使用“本地用户”安全性』。
- 第 62 页的『在 QuickPlace 服务器上配置“目录辅助”』。
- 第 58 页的『在 AIX 上配置 I/O 完成端口以搜寻 Lotus Domino 源』。

配置概述

可以使用 QuickPlace 搜寻器搜寻任意数目的 QuickPlace 位置。在创建搜寻器时，应该只从一台 QuickPlace 服务器中选择要搜寻的位置。此后，在编辑搜寻空间时，可以添加另一台要包括在同一搜寻空间中的 QuickPlace 服务器上的文档。在创建或编辑搜寻器时，可以指定是搜寻所要搜寻的位置中的所有空间还是要搜寻特定的空间。

要创建或更改 QuickPlace 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有文档。
- 标识 QuickPlace 服务器和通信协议。
- 指定关于与该服务器相关联的用户目录的信息（搜寻器需要此信息以便可以在用户搜索该集合时强制实施访问控制）。
- 选择要搜寻的位置。
- 设置用于搜寻位置的时间表。

- 选择要搜寻的文档。搜寻器可以搜寻某个位置中的所有空间，也可以只搜寻所选空间中的文档。
- 指定用于使各种位置和空间中的字段可供搜索的选项。例如，可以从搜寻空间中排除某些字段，也可以指定用于搜索附件的选项。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

也可以选择一个选项以便在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。

创建搜寻器时，可以单击帮助以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

配置 QuickPlace 服务器以使用“本地用户”安全性

如果计划将 QuickPlace 搜寻器配置为使用“本地用户”选项来实施安全性，则必须在创建该搜寻器之前在 Lotus QuickPlace 服务器上配置 Domino 目录。

关于本任务

在配置 QuickPlace 搜寻器时为搜寻器选择用于强制实施文档级安全性的安全方式。如果选择“本地用户”方式，则必须确保所有本地用户标识和本地组都已在 Domino 目录中注册（Domino 目录层次结构必须与 QuickPlace 层次结构相对应）。

还必须确保指定让搜寻器使用的用户标识和密码已在 Domino 目录中注册并且有权读取所要搜寻的数据库。

要使用 QuickPlace，只有用户名是必需的。但是，要搜寻 QuickPlace 源，需要完全展开的用户标识。展开的用户标识具有以下格式：

`username/placename/QP/domainname`

使用此过程来确定用户标识的完全展开版本、确保此用户标识有权读取 QuickPlace 数据库以及将该用户标识添加到 Domino 目录中。Domino 目录必须包含用来搜寻 QuickPlace 数据库以及所有 QuickPlace 本地用户和本地组的用户标识（Domino 目录层次结构必须与 QuickPlace 层次结构相对应）。

过程

要配置 QuickPlace 服务器以使用“本地用户”安全性：

1. 确认用户标识许可权：
 - a. 打开 QuickPlace 服务器上的 Server 文档。
 - b. 打开“文件”页，然后打开所要搜寻的数据库的访问控制表（ACL）。
 - c. 确认搜寻器要使用的本地用户标识包含在 ACL 中并且此用户标识有权读取数据库。

必须在步骤 2 中指定此用户标识的完全展开格式。

2. 将用户添加到 Domino 目录中：
 - a. 打开 QuickPlace 服务器上的 Server 文档。

- b. 在“人员和组”页上，在人员树项中，添加步骤 1 中确认的完全展开用户标识。
- c. 在因特网密码字段中，指定此用户标识的密码。

在 QuickPlace 服务器上配置“目录辅助”

如果计划将 QuickPlace 搜寻器配置为使用 LDAP 目录来实施安全性，则必须先在 Lotus QuickPlace 服务器上创建“目录辅助”数据库，然后才能配置搜寻器。

限制

要搜寻的 QuickPlace 服务器必须正在运行 DIIOP 和 HTTP 任务。

过程

要在 QuickPlace 服务器上配置“LDAP 目录辅助”：

1. 创建“目录辅助”数据库：
 - a. 打开 QuickPlace 服务器上的 Server 文档。
 - b. 通过使用目录辅助（6）模板来创建数据库。此模板在服务器上。
 - c. 单击添加目录辅助以在该数据库中创建文档。
 - d. 打开“基本”选项卡，然后在域类型字段中选择 LDAP。
 - e. 打开“命名上下文”选项卡，确保选中了信任凭证复选框。
 - f. 打开 LDAP 选项卡，指定有关 LDAP 服务器的信息。
 - g. 保存 Server 文档，然后将其关闭。
2. 配置 QuickPlace 服务器以使用“目录辅助”数据库：
 - a. 打开 QuickPlace 服务器上的 Server 文档。
 - b. 打开“基本”选项卡，在目录辅助数据库名称字段中，指定在步骤 1 中创建的数据库的名称。
 - c. 保存 Server 文档，然后将其关闭。

现在，QuickPlace 服务器就可以将该 LDAP 服务器用作辅助 Domino 目录了。

UNIX 文件系统搜寻器

要在企业搜索集合中包括 UNIX 文件系统中存储的文档，必须配置 UNIX 文件系统搜寻器。

可以使用 UNIX 文件系统搜寻器搜寻任意数目的 UNIX 文件系统。配置该搜寻器时，选择要搜寻的本地目录和子目录以及远程目录和子目录。

如果将搜寻器服务器安装在 Windows 计算机上，则不能使用该服务器来搜寻 UNIX 文件系统源（UNIX 文件系统搜寻器不会出现在可用搜寻器类型的列表中）。

要创建或更改 UNIX 文件系统搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有子目录。

- 设置用于搜寻文件系统的时间表。
- 选择要让搜寻器搜寻的子目录以及子目录层。
- 指定用于使子目录中的文档可供搜索的选项。例如，可以将某些类型的文档排除在搜寻空间之外。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

创建搜寻器时，可以单击**帮助**以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

Web 搜寻器

要在企业搜索集合中包括来自 Web 站点的页面，必须配置 Web 搜寻器。

可以使用 Web 搜寻器来搜寻任意数目的超文本传输协议（HTTP）服务器和安全 HTTP（HTTPS）服务器。该搜寻器访问 Web 站点并读取该站点上的数据。然后，可以根据文档中的链接来搜寻其他文档。Web 搜寻器可以从各个页面或框架集（使用 HTML 框架创建的页面）搜寻并抽取链接。

搜寻的数据可以具有许多常用格式中的其中一种，并且来自内部网或因特网中的各种来源。常用格式包括 HTML、PDF、Microsoft Word、Lotus WordPro 以及可扩展标记语言（XML）等等。

要创建或更改 Web 搜寻器，请登录到企业搜索管理控制台。您还必须是企业搜索管理角色成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有 Web 页面。
- 指定规则以允许和禁止访问 Web 站点。在指定搜寻规则时，可以对规则进行测试并验证搜寻器是否能够访问要包括在搜寻空间中的站点。
- 指定选项以包括某些类型的文件和排除具有某些文件扩展名的文件。
- 指定 Web 搜寻器处理软错误页面的规则。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。
- 指定用于搜寻受密码保护的 Web 站点（要搜寻的 Web 服务器必须使用 HTTP 基本认证或 HTML 表单来提示输入密码）的选项。
- 指定用于搜寻由代理服务器提供服务的 Web 站点的选项。

创建搜寻器时，可以单击**帮助**以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

用户代理配置

Web 搜寻器遵守 Robots Exclusion Protocol 协议。要搜寻使用此协议的 Web 站点，应确保 Web 站点上的 robots.txt 文件允许对 Web 搜寻器配置的用户代理名称访问该 Web 站点。

当企业搜索系统启动时，Web 搜寻器将装入对它配置的用户代理名称。在搜寻器从一个先前从未访问过的 Web 站点（或已有一段时间未访问的 Web 站点）下载页面时，搜寻器会首先尝试下载一个名为 robots.txt 的文件。此文件在 Web 站点的根目录中。

如果 robots.txt 文件不存在，则该 Web 站点是完全开放不限制搜寻的。如果该文件存在，它将指定该站点（目录）的哪些区域是对搜寻器没有限制的。robots.txt 文件通过标识搜寻器的用户代理名称来指定搜寻器的许可权。

虽然 Robots Exclusion Protocol 协议是自愿性质的，但企业搜索 Web 搜寻器遵守该协议：

- 如果 robots.txt 文件中包含了对 Web 搜寻器配置的用户代理名称的一个条目，则该 Web 搜寻器符合对那个用户代理的限制。
- 如果用户代理名称未出现在 robots.txt 文件中，但最后一个条目指定 User-agent: *（表示任何用户代理），且限制被指定为 Disallow: /（表示从 Web 站点的根目录开始不允许任何搜寻），则将禁止该 Web 搜寻器搜寻那个站点。
- 如果用户代理名称未出现在 robots.txt 文件中，但最后一个条目指定 User-agent: * 且限制被指定为 Allow: /，则将允许该 Web 搜寻器搜寻那个站点。

Web 站点管理员通常会指定最后一个条目以禁止未被显式授予访问权的所有搜寻器访问该站点。如果正在配置新的 Web 搜寻器并且知道想要搜寻的某些 Web 站点使用 Robots Exclusion Protocol 协议，则应要求那些 Web 站点的管理员在他们的 robots.txt 文件中为搜寻器添加一个条目。

应确保在 Web 搜寻器的属性中与您感兴趣的 Web 站点的所有 robots.txt 文件中指定的用户代理名称相同。

如果要搜寻的 Web 站点中没有任何一个使用 Robots Exclusion Protocol 协议，则对用户代理属性指定的值通常不会产生影响。但某些应用程序服务器、JSP 和 servlet 会使它们的响应适应用户代理名称。例如，针对浏览器的不兼容性，会有不同的响应。在这些情况下，无论是否使用 Robots Exclusion Protocol 协议，对 Web 搜寻器指定的用户代理名称都可能会有影响。如果需要搜寻这些类型的站点，请咨询 Web 站点管理员以确保允许 Web 搜寻器进行访问。

支持 JavaScript

企业搜索的 Web 搜寻器可查找包含在 Web 文档的 JavaScript™ 部分的一些链接（URL）。

Web 搜寻器既可查找相对链接也可查找绝对链接。如果 HTML 文档中包含了 BASE 元素，则搜寻器使用该元素来解析相对链接。否则，搜寻器将使用该文档自己的 URL。

对 JavaScript 的支持仅限于抽取链接。搜寻器不解析 JavaScript、不构建 DOM（文档对象模型）也不解释或执行 JavaScript 语句。搜寻器在文档内容（包括但不限于 JavaScript 部分）中查找有可能是 JavaScript 语句中的 URL 的字符串。这有两层含义：

- 将找到一些被更严格的 HTML 解析器所忽略的 URL。搜寻器将拒绝任何语法无效的 URL，但扫描步骤返回的某些有效 URL 可能不值得搜索。
- 由 JavaScript 产生的文档内容（例如，当某个用户通过浏览器查看一个页面，且该浏览器执行某些 JavaScript 时）将不会被 Web 搜寻器发现，因此将不会对它们建立索引。

限制 Web 搜寻空间的规则

要确保用户仅访问您让他们搜索的 Web 站点，请指定规则以限制 Web 搜寻器可以搜寻的内容。

当 Web 搜寻器搜寻 Web 页面时，它会发现指向其他页面的链接并将那些链接放入接着要搜寻的队列中。只要时间和内存资源允许，就可以重复进行搜寻和发现。在配置 Web 搜寻器时，请指定该搜寻器的起始搜寻位置。从这些初始 URL（它们被称为起始 URL）开始，Web 搜寻器能够到达 Web 上与这些初始 URL 直接或间接链接的任何文档。

要限制搜寻空间，请将 Web 搜寻器配置为彻底搜寻某些 URL 并忽略指向您不感兴趣的区域的链接。缺省情况下，因为搜寻器接受它所发现的任何 URL，所以必须指定一些规则，这些规则标识要包括在集合中的 URL 并除去余下的页面。

可以通过几种方法来指定要让 Web 搜寻器搜寻和不搜寻的内容。可以指定：

- 作为搜寻器的起始搜寻位置的起始 URL 列表
- 三种类型的搜寻规则：域、因特网协议（IP）地址和 URL 前缀
- 要包括的文档的 MIME 类型列表
- 要排除的文档的文件扩展名列表
- URL 路径中的最大目录数

搜寻规则具有以下格式：

```
action type target
```

action 是 forbid 或 allow；type 是域、IP 地址或 URL 前缀（HTTP 或 HTTPS）；target 视 type 的值而定。可以通过几种有限的方法来指定星号（*）作为通配符，以指定与模式相匹配的目标。

域规则

域规则的目标是 DNS 域名。例如，可以指定要搜寻整个 www.ibm.com 域：

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

目标中的第一个字符可以是一个星号，这将导致该规则应用于任何以该模式的余下部分结束的主机名。例如，可以指定不搜寻与以下模式相匹配的域：

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

无论指定的是显式域名还是域名模式，主机名匹配都是区分大小写的。例如，*.user.ibm.com 与 joe.user.ibm.com 和 mary.smith.user.ibm.com 都匹配，但与 joe.user.IBM.com 不匹配。

未指定端口号的域规则适用于该域上的所有端口。在以下示例中，sales 域上的所有端口都是允许的：

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

如果域规则指定了端口号，则该规则仅适用于该端口。在以下示例中，只允许 sales 域上的端口 443:

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

前缀规则

前缀规则控制以指定字符串开头的 URL 的搜寻。目标是单个 URL，它通常包含一个或多个星号以表示一个模式。例如，通常将星号指定为前缀字符串中的最后一个字符。

前缀规则使您能够搜寻整个 Web 站点或 Web 站点的一部分。可以指定目录路径或模式，然后允许或禁止目录树中从该点开始的所有内容。例如，下列规则共同起作用以允许搜寻器搜寻 sales.ibm.com 上 public 目录中的所有内容，但禁止搜寻器访问该站点上的任何其他页面:

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

当指定前缀规则时，可以指定多个星号，并且可以在前缀字符串中的任何位置指定这些星号，而不仅仅是在最后一个位置指定它们。例如，以下规则禁止搜寻器搜寻 sales.ibm.com 站点的顶级目录（如果目录名以 fs 结尾的话）中的任何文档。（例如，可能有一些文件系统安装未包含对于搜索索引有用的信息。）

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

地址规则

地址规则使您能够通过将 IP 地址（仅 IPv4）和网络掩码指定为目标来控制整个主机或网络的搜寻。例如:

```
allow address 9.0.0.0 255.0.0.0
```

此网络掩码使您能够指定模式匹配。为了将地址规则应用于候选 IP 地址，规则中的 IP 地址与候选 IP 地址必须完全相同（除了网络掩码中被零屏蔽掉的部分以外）。地址规则定义一个模式，网络掩码定义地址模式中的有效位。网络掩码中的零充当通配符，它表示在地址中的同一个位的位置指定的任何值都匹配。

在上一个示例中，allow 规则应用于第一个八位元为 9 且后三个八位元为任意值的所有 IP 地址。

可以将一个很有用的规则作为最终地址包括在规则列表中: forbid address 0.0.0.0 0.0.0.0。此规则与任何 IP 地址都匹配，这是因为网络掩码使所有位都无效（此规则禁止规则列表中的先前规则所不允许的所有地址）。

代理服务器限制: 如果计划搜寻由代理服务器提供服务的 Web 站点，请不要指定 IP 地址规则。代理服务器通常是在用户代理（浏览器或搜寻器）无法直接访问 Web 服务器所在的网络时使用的。例如，HTTP 代理服务器可以将 HTTP 请求从搜寻器传递至 Web 服务器并将响应回传至搜寻器。

当 Web 搜寻器使用代理服务器时，该代理服务器的 IP 地址是搜寻器所了解的唯一其他主机 IP 地址。如果使用 IP 地址规则来将搜寻

器约束到 IP 地址子网，该约束将导致使用返回码 760 对几乎所有 URL 进行分类（该返回码表示 Web 空间禁止那些 URL）。

搜寻规则顺序

在发现和搜寻 URL 的过程中的各个时间，搜寻器都可能应用搜寻规则。规则的顺序十分重要，但仅在每种类型的规则中才如此。一个地址规则是出现在另一个地址规则之前还是之后是有区别的，但是一个地址规则是出现在一个前缀规则之前还是之后是没有区别的，因为搜寻器不会同时应用这两种规则。

在单一类型的规则集中，搜寻器针对每条规则测试候选的域、地址或 URL，顺序是从第一条指定的规则开始到最后一条指定的规则结束，直到找到适用的规则为止。将使用对第一条适用规则所指定的操作。

对顺序的依赖性决定了大部分搜寻规则的典型结构：

- 域规则集合通常以从搜寻空间中排除单个域的 `forbit` 规则开始。例如，集合管理员可能会确定某些域不包含有用的信息。
- 在 `forbit` 规则列表的后面通常跟着一系列 `allow` 规则（带有通配符），这些 `allow` 规则使搜寻器能够访问任何以其中一个高级域名结束的域，那些域名定义了企业内部网（如 `*.ibm.com` 和 `*.lotus.com`）。

域规则集合以下列缺省规则结束，此规则排除了先前规则所不允许的域：

```
forbit domain *
```

这个最终规则很重要，因为它可防止搜寻空间包括整个因特网。

- 地址规则集合通常以少数 `allow` 规则开始，这些 `allow` 规则使搜寻器能够搜寻跨企业内部网的高级（A 类、B 类或 C 类）网络。

地址规则通常以下列规则结束，该规则防止搜寻器搜寻位于企业网络外部的 Web 站点。

```
forbit 0.0.0.0 0.0.0.0
```

- 因为前缀规则集包含对指定为树和子树的允许和禁止区域的非常详细的规范说明，所以前缀规则集通常是最大的。一种较好的方法是，首先允许或禁止更紧密集中的区域，然后采用更一般的模式来指定相反的规则，以允许或禁止其余内容。

前缀部分通常不以典型规则结束。建议的最终域和地址规则可以确保搜寻器不会搜寻到企业网络范围之外，这比测试 URL 前缀更有效率。

如果按操作（`forbit` 或 `allow`）将规则分组，则搜寻器可以更有效地应用前缀规则。例如，不要指定相互交替的短的 `allow` 和 `forbit` 规则序列，而是指定规定一项操作的一长串规则，然后指定规定另一项操作的一长串规则。可以交错使用 `allow` 和 `forbit` 规则以达到搜寻空间的目标。但是，将 `allow` 规则组合到一起并将 `forbit` 规则组合到一起可以提高搜寻器性能。

文件扩展名、MIME 类型和最大搜寻深度

这些选项提供了指定搜寻空间内容的其他方法。可以根据文档的文件扩展名来排除某些类型的文档，并且可以根据文档的 MIME 类型来包括某些类型的文档。当指定要让搜寻器搜寻的 MIME 类型时，请注意，在 Web 文档中通常未正确地设置 MIME 类型。

最大搜寻深度是 URL 中从其站点根开始的斜杠数。此选项使您能够防止搜寻器陷入具有无限深度的递归文件系统结构中。当搜寻器沿着从一个文档到另一个文档的链接进行搜寻时，搜寻深度不对应于搜寻器经过的层数。

起始 URL

起始 URL 是搜寻器首先搜寻的 URL，每当搜寻器启动时，这些 URL 就会被插入到搜寻中。如果已经发现了起始 URL，则它们不会比在搜寻规则中允许的其他 Web 站点更早地被搜寻或重新搜寻。

当第一次启动 Web 搜寻器且搜寻空间为空时，起始 URL 特别重要。在将先前未发现的 URL 添加到搜寻空间中的起始 URL 列表时，起始 URL 也很重要。

起始 URL 必须是标准 URL，而不能仅仅是域名。必须指定协议和端口号（如果端口不是 80 的话）。

下列 URL 是有效的起始 URL:

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

下列 URL 不是有效的起始 URL:

```
www.ibm.com
```

在搜寻规则中必须包括起始 URL。例如，如果搜寻规则不允许搜寻指定的起始 URL，搜寻器就无法从该 URL 开始搜寻。

使用 Web 搜寻器测试 URL 连接

在指定要让 Web 搜寻器搜寻的 URL 之后，可以测试搜寻规则的配置。

测试结果显示了搜寻器是否能够使用搜寻器属性中指定的用户代理名来访问那些 URL。测试结果还显示了 URL 是否由于排除规则而无法被搜寻（例如，文档可能由于其文件扩展名与搜寻空间所排除的扩展名匹配而无法被搜寻）。

在对站点至少执行了一次搜寻后，可以测试 URL 以获取其他信息。例如，测试报告可以提供最新的 HTTP 返回码（此返回码指示对该 URL 执行的搜寻是否已成功），可以显示该 URL 的上次搜寻时间以及已安排的再次搜寻时间，并且可以显示用户代理是否正在使用 Web 服务器的当前 robots.txt 文件。

Web 搜寻器中的重新搜寻时间间隔设置

要影响 Web 搜寻器重新访问 URL 的频率，请在 Web 搜寻器属性中指定选项。

企业搜索系统中的大多数其他搜寻器类型是根据管理员指定的时间表运行的。相反，在启动 Web 搜寻器之后，它通常是持续运行的。要控制它重新访问它先前已搜寻的 URL 的频率，请指定最小重新搜寻时间间隔和最大重新搜寻时间间隔。

在使用企业搜索管理控制台来创建 Web 搜寻器或编辑 Web 搜寻器属性时，可以选择一个选项以配置高级属性。在高级 Web 搜寻器属性页上，指定最小重新搜寻时间间隔和最大重新搜寻时间间隔选项。Web 搜寻器使用指定的值来计算重新搜寻数据的时间间隔。

第一次搜寻页面时，搜寻器使用搜寻该页面的日期和时间以及指定的最小重新搜寻时间间隔与最大重新搜寻时间间隔的平均值来设置重新搜寻日期。在该日期以前，不会重新搜寻该页面。在该日期以后重新搜寻该页面的时间取决于搜寻器负载以及搜寻空间中的新旧 URL 之间的均衡。

每次重新搜寻该页面时，搜寻器都检查内容是否已更改。如果内容已更改，则下一个重新搜寻时间间隔将比上一个短，但是决不会比指定的最小重新搜寻时间间隔短。如果内容未更改，则下一个重新搜寻时间间隔将比上一个长，但是决不会比指定的最大重新搜寻时间间隔长。

有关使用 Web 搜寻器来访问 URL 的选项

可以强制 Web 搜寻器尽可能快地访问特定 URL。

如果需要使用来自某个 Web 站点的信息来刷新搜寻空间，则可以监视搜寻器，选中**要访问或重新访问的 URL** 选项，然后指定需要搜寻或重新搜寻的页面的 URL 或 URL 模式。

例如，如果“通信”部门将一个 Web 页面添加到内部网中或者修订了一个页面以反映重要的政策更改，则可以指定新页面或已更改的页面的 URL。如果搜寻器正在运行，则搜寻器将把指定的 URL 排队，以便在它下次检查正在等待被访问的页面时搜寻那些 URL（通常是每 10 分钟搜寻一次）。如果搜寻器未在运行中，则将指定的 URL 排队，以便搜寻器下次启动时可以搜寻该 URL。

确保搜寻规则包括允许搜寻器访问所指定的 URL 的规则。搜寻器可以比在正常情况下更快地访问指定的 URL。但是，要真正地搜寻该 URL，必须存在允许搜寻该 URL 的搜寻规则。

新搜寻到的数据将在下次重组或刷新索引时可供搜索。

Web 搜寻器处理软错误页面的方式

可以配置 Web 搜寻器处理定制页面，Web 站点管理员在对特定页面的请求做出响应时如果不想返回标准错误代码，可创建定制页面。

如果 HTTP 服务器无法返回客户机请求的页面，则服务器通常会返回一个由响应头和返回码组成的响应。返回码将指示问题的所在（例如，错误 404 表示无法找到文件）。有些 Web 站点管理员会创建特殊页面用来更详细的对问题进行解释，并配置 HTTP 服务器返回这些页面。这些定制页面称为软错误页面。

软错误页面会干扰 Web 搜寻器的结果。例如，搜寻器本应接收到一个指示出现了问题的响应头，但是却接收到一个软错误页面和返回码 200，表示成功下载了有效的 HTML 页面。但此下载的软错误页面与请求的 URL 无关，而在每次应返回请求的页面却返回了此软错误页面时，它的内容都是非常相似的。这些不相关并且几乎重复的页面会干扰索引和搜索结果。

要处理这种情况，可以在配置 Web 搜寻器时指定用来处理软错误页面的选项。Web 搜寻器需要关于返回软错误页面的每个 Web 站点的以下信息：

- 使用软错误页面的站点的 URL 模式。此 URL 模式由协议（HTTP 或 HTTPS）、主机名、端口号（如果非标准）和路径名组成。可以使用星号（*）作为通配符以便与一个或多个字符匹配，直到遇到模式中下一个出现的非通配符。指定的模式区分大小写。
- 对应于 HTML 文档 <TITLE> 标记的文本的标题模式。指定此模式时，可以使用星号（*）作为通配符。指定的此模式区分大小写。
- 对应于 HTML 文档内容的文本的内容模式。如果存在 <BODY> 标记，内容不仅是 <BODY> 标记的内容。内容是文件中 HTTP 头后面的所有内容。指定此模式时，可以使用星号（*）作为通配符。指定的此模式区分大小写。
- 代表返回码的一个整数，用于与指定的 URL、标题和内容模式相匹配的文档。

示例

以下配置告诉 Web 搜寻器将从 `http://www.mysite.com/hr/*` Web 站点返回的所有有效 HTML 页面（返回码 200）与指定的标题和内容模式进行比较。如果某个页面的 <TITLE> 标记以“Sorry, the page”开始，并且文档的内容包含了任何内容（*），则搜寻器将以接收到返回码 404（未找到页面）时相同的方法处理该页面。

表 2. 软错误页面示例

URL 模式	标题模式	内容模式	HTTP 返回码
<code>http://www.mysite.com/hr/*</code>	<code>Sorry, the page*</code>	<code>*</code>	404

可以对同一 Web 站点创建多个条目，以处理不同的返回码。来自相同 Web 站点的每个返回码都需要在 Web 搜寻器的配置中有自己的条目。

使用通配符

URL、标题和内容模式不是正则表达式。星号与任何字符匹配，直到遇到下一个出现的任何非通配符。例如：

`*404` 与 *any characters*404 匹配

`404: *` 与 `404:` 任何字符匹配

`http://*.mysite.com/*` 与 `http://any host.mysite.com/any file` 匹配

`*` 与 *any characters* 匹配

性能影响

在配置处理软错误页面的选项时，由于必须检查所有成功搜寻的页面，所以增加了搜寻器的处理时间量。还需要更多的处理时间来检查模式匹配并确定应返回一个页面还是返回一个返回码。

支持搜寻安全 Web 站点

通过在企业搜索管理控制台中指定凭证，可以使 Web 搜寻器能访问受限内容，例如，需要密码才能访问的文档。

如果 Web 服务器使用 HTTP 基本认证或基于 HTML 表单的认证来限制对 Web 站点的访问，可以在 Web 搜寻器的配置中指定使受密码保护的 Web 站点中的页面能被搜寻的凭证。也可以指定选项用于手工配置 cookie 文件。

由 HTTP 基本认证保护的 Web 站点

如果 Web 服务器使用 HTTP 基本认证来限制对 Web 站点的访问，可以指定使 Web 搜寻器能访问受密码保护的页面的认证凭证。

要确定用户（或客户机应用程序）是否有权访问 Web 站点上的页面，很多 Web 服务器都使用一种客户机认证方案（称作 HTTP 基本认证）来建立用户的身份。通常，这种交互作用是交互式的。

- 当 HTTP 用户代理（例如，Web 浏览器）请求一个受 HTTP 基本认证保护的页面时，Web 服务器会以返回码 401 作出响应，该返回码表示未授权请求方访问所请求的页面。
- Web 服务器也会要求请求方出示可用于验证是否允许该用户访问受限内容的凭证。
- Web 浏览器将为用户显示一个对话框，要求用户提供构成用户凭证所需的用户名、密码和任何其他信息。
- Web 浏览器对凭证进行编码，然后在以后再对受保护的页面发出请求时包括这些凭证。
- 如果凭证有效，Web 服务器将作出响应，返回返回码 200 和请求的页面内容。
- 以后再对来自相同 Web 服务器的页面发出请求时通常会包括相同的凭证，从而使经过授权的用户能访问其他受限内容，而不会被要求对每个请求都指定凭证。

一旦建立了用户身份，Web 服务器和 HTTP 用户代理通常会交换令牌（称为 *cookie*），它使用户登录状态信息在 HTTP 请求之间保留下来。

由于 Web 搜寻器不是交互运行的，所以必须在搜寻器开始搜寻之前指定使它能搜寻受密码保护的页面的凭证。在创建 Web 搜寻器或编辑搜寻空间时，指定关于需要搜寻的每个安全 Web 站点的信息。

要指定此信息，必须与受 HTTP 基本认证保护的 Web 站点或 Web 服务器的管理员密切合作。他们必须向您提供将要搜寻的 Web 站点的安全需求，包括用于认证 Web 搜寻器身份以及确定搜寻器有搜寻受限页面许可权的所有信息。

如果在创建集合时已对集合启用了安全性，可以在配置搜寻器时指定安全性令牌（例如，用户标识、组标识或用户角色）来控制对文档的访问。Web 搜寻器使这些安全性令牌与它在文件系统树中为指定根 URL 搜寻的每个文档相关联。这些令牌是在整个 Web 搜寻空间配置的任何文档级安全性令牌以外额外使用的。

URL 的顺序很重要。在添加关于受密码保护的 Web 站点的信息后，必须按照想要搜寻器处理它的顺序来放置。首先列出较具体的 URL，然后再把较为笼统的 URL 放在列表的较低位置中。当 Web 搜寻器对候选 URL 进行评估时，它将使用列表中第一个与候选 URL 匹配的 URL 指定的认证数据。

由基于表单的认证保护的 Web 站点

如果 Web 服务器使用 HTML 表单来限制对 Web 站点的访问，可以指定使 Web 搜寻器能访问受密码保护的页面的认证凭证。

要确定用户（或客户机应用程序）是否有权访问 Web 站点上的页面，很多 Web 服务器都使用 HTML 表单来建立用户的身份。通常，这种交互作用是交互式的。

- 当 HTTP 用户代理（例如，Web 浏览器）请求一个受基于表单的认证保护的页面时，Web 服务器会检查该请求是否包括了建立用户身份的 *cookie*。

- 如果未出示该 cookie，则 Web 服务器会提示用户将安全性数据输入表单。在用户提交表单时，Web 服务器会返回需要的 cookie 并允许处理对受密码保护的页面的请求。
- 同样也允许处理此后包括了所需 cookie 的请求。授权用户可以访问其他受限内容，而不会被要求对每个请求都填写表单并指定凭证。

由于 Web 搜寻器不是交互运行的，所以必须在搜寻器开始搜寻之前指定使它能搜寻受密码保护的页面的凭证。在创建 Web 搜寻器或编辑搜寻空间时，指定关于需要搜寻的每个安全 Web 站点的信息。

指定的字段对应于交互式用户在受到 Web 浏览器提示时填写的字段，以及成功登录所需的任何隐藏字段或静态字段。

要指定此信息，必须与由基于表单的认证保护的 Web 站点或 Web 服务器的管理员密切合作。他们必须向您提供将要搜寻的 Web 站点的安全需求，包括用于认证 Web 搜寻器身份以及确定搜寻器有搜寻受限页面许可的所有信息。

URL 模式的顺序很重要。在添加关于受密码保护的 Web 站点的信息后，必须按照想要搜寻器处理它的顺序来放置。首先列出特定程度较高的 URL 模式，然后将通用程度较高的 URL 模式放在列表中的较后位置。当 Web 搜寻器对候选 URL 进行评估时，它将使用列表中第一个与候选 URL 匹配的 URL 模式指定的表单数据。

由代理服务器提供服务的 Web 站点

如果不允许 Web 搜寻器直接访问网络，可以配置搜寻器使用 HTTP 代理服务器来访问想要搜寻的内容。

如果在运行 Web 搜寻器的计算机上无法访问 TCP/IP 网络，或者只有拥有特权的进程才能进行访问，则可以将 Web 搜寻器配置为使用 HTTP 代理服务器。HTTP 代理是一个进程，它在指定主机上的指定端口上侦听 HTTP 请求。代理服务器将请求传递给 Web 服务器，并将 Web 服务器的响应传递给发出请求的客户机（Web 搜寻器）。代理服务器可以与 Web 搜寻器在同一台计算机上运行，也可以在不同的计算机上运行。

在非代理搜寻中，对 URL 的请求直接发送给主机。该请求将随着代理搜寻被发送给代理服务器。

在创建 Web 搜寻器或编辑搜寻空间时，可以指定关于代理服务器的信息，Web 搜寻器在搜寻该代理服务器的域中的页面时将使用此信息。在向搜寻空间添加代理服务器之前应先获取以下信息：

代理服务器域

由代理服务器提供服务的域。可以使用星号 (*) 作为通配符。例如，* 与此代理服务器提供服务的域相匹配，*.resource.com 与所有以 resource.com 结尾的域相匹配。

限制： 由于代理服务器的 IP 地址是搜寻器所了解的唯一其他主机 IP 地址，所以，不能指定 IP 地址规则来搜寻代理服务器。如果使用 IP 地址规则来将搜寻器约束到 IP 地址子网，该约束将导致使用返回码 760 对几乎所有 URL 进行分类（该返回码表示 Web 空间禁止那些 URL）。

代理服务器主机名或 IP 地址

代理服务器的 DNS 主机名或点分 IP 地址。

代理服务器端口号

代理服务器侦听 HTTP 代理请求的 TCP/IP 端口号。

在添加代理服务器后，必须选择它并按照搜寻器处理它时依据的顺序来安排它的位置。首先列出特定程度较高的域名，然后将通用程度较高的域名放在列表中的较后位置。当 Web 搜寻器对候选 URL 进行评估时，它将使用列表中第一个与候选 URL 匹配的域指定的代理服务器数据。（与任何代理规则都不匹配的 URL 被假定为直接可供搜寻器访问。）

Cookie 管理

通常，cookie 管理会自动进行，而不需企业搜索管理员执行任何操作。如果需要，也可以为 Web 搜寻会话手工指定 cookie。

Cookie 是 Web 服务器返回给用户代理作为 HTTP 响应头的一部分的不透明令牌。Cookie 只对发出它们的 Web 服务器有意义，用于保持 HTTP 请求之间的状态。例如，在客户机认证期间，Web 服务器可能会返回一个 cookie，使服务器能确定某个经认证的用户已登录。出示该 cookie 使该用户能对那个 Web 服务器上的页面发出更多请求，而不会被提示再次登录。

Web 搜寻器保留从 Web 服务器接收到的 cookie，并在搜寻器实例运行期间使用它们。它将 cookie 存储在 cookies.ini 文件中，该文件由搜寻器在每个搜寻器会话结束时重写。在 Web 搜寻器停止时，它将保存所有未过期的 cookie，并在下个会话开始时重新装入。

如果手工指定 cookie，应将它们存储在单独的文件中，然后在需要时将它们与 cookies.ini 文件中的 cookie 合并。搜寻器不会废弃未到期的 cookie，但是，当由于发生问题而导致无法写整个 cookie 集合时，您不会希望丢失手工指定的 cookie。必须在搜寻会话开始之前，将 cookie 与搜寻器自动维护的 cookie 合并。

Cookie 格式

计划要与企业搜索 cookies.ini 文件合并的 cookie 必须采用特定的格式。

- 每个 cookie 都必须在一行上。允许使用空白行和注释，但不会将它们保留在 cookies.ini 文件中。
- 每个 cookie 都必须有以下格式：

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

Cookie

表示 cookie 条目开始的必需关键字。

Cookie 关键字不能包含空白，并且必须在末尾附加一个数字，可以是 0、1 或 2。该数字表示 cookie 类型：version-0（Netscape）、version-1（RFC2109）或 version-2（RFC2965）。在 RFC2965 cookie 中不支持端口列表。

cookie_length

相关联的 cookie 文本的长度，以字符计。

URL_length

相关联的验证 URL 的长度，以字符计。

cookie_text

将发送到始发 Web 服务器的 cookie 的内容。此字符串（表示 HTTP 响应头中

Set-Cookie 伪指令的右边) 指定 cookie 的名称 / 值对, 以及要随该 cookie 一起发送的任何其他内容 (例如, 路径、安全性设置等)。此字符串后面跟着逗号 (,) 分隔符。

validation_URL

发现此 cookie 的 URL。此 URL 用于确定将 cookie 发送到何处 (例如, 通过提供域名和路径名)。验证 URL 必须满足始发 Web 服务器对 cookie 的安全限制和隐私限制。

以下示例显示为两行是为了更易阅读; 指定的 cookie 必须在一行上:

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

为 Web 搜寻器配置 cookie

可以为 Web 搜寻会话手工指定 cookie, 并将它们与 Web 搜寻器维护的 cookie 合并。

在开始之前

要手工配置 cookie 以供 Web 搜寻器使用, 您必须是企业搜索管理员。

过程

要为 Web 搜寻器手工配置 cookie:

1. 从企业搜索管理控制台中, 监视想要为其指定 cookie 的集合, 然后停止 Web 搜寻器。
2. 作为企业搜索管理员登录到搜寻器服务器上 (此用户标识是在安装 WebSphere II OmniFind Edition 时指定的)。
3. 切换到想要配置的搜寻器的数据目录, 其中 *crawler_session_ID* 是由企业搜索系统分配给搜寻器会话的标识。 例如:

```
ES_NODE_ROOT/data/col_56092.WEB_88534
```
4. 编辑 *cookies.ini* 文件, 将手工指定的那些 cookie 条目附加到已列出的条目后, 然后保存并退出该文件。 确保您的 cookie 不会覆盖任何已存在的 cookie。
5. 从企业搜索管理控制台中, 重新启动停止的 Web 搜寻器。

全局 Web 搜寻空间配置

可以为 Web 搜寻器配置全局搜寻空间, 从而使您能在从索引中除去 URL 时更好地进行控制。

每个 Web 搜寻器都配置有搜寻空间, 搜寻空间定义要搜寻的或不要搜寻的 URL。搜寻空间中已被发现的 URL 将被保留下来 (保留在数据库中) 以用于将来的搜寻; 那些不在搜寻空间中的 URL 将被废弃。如果搜寻器在开始时使用的是空数据库, 则搜寻空间定义与数据库在搜寻器运行时将保持一致。

有时, 搜寻器会停止并且它的搜寻空间会缩小 (例如, 新规则禁止页面被搜寻)。当搜寻器重新启动后, 它的搜寻空间定义和数据库将变得不一致。数据库包含了一些新的较小的搜寻空间中没有的 URL (有些已搜寻, 有些未搜寻)。

如果集合只有一个 Web 搜寻器, 则该 Web 搜寻器可通过将这些 URL 的 HTTP 返回码更改为 760 (以指定将排除这些 URL) 并请求从索引中除去现在已排除的页面, 来恢复一致性。

如果将搜寻空间划分给两个或多个 Web 搜寻器（例如，为确保某些页面比其他页面更频繁地被搜寻），则每个 Web 搜寻器都将维护独立的数据库表（最初是空的），并搜寻 Web 搜寻空间的一个不同的部分。于是，原始搜寻器的搜寻空间即缩小为在除去由其他搜寻器搜寻的那些部分之后剩下的空间。当原始搜寻器尝试通过从索引中除去已移动的页面来恢复一致性时，会出现问题。因为这些已移动的页面现在正由其他搜寻器在搜寻，它们应保留在索引中。

通过配置更高级别的全局搜寻空间，可以标识原始搜寻器将不搜寻，但也不会从索引中除去的 URL。那些不再存在于任何搜寻器的搜寻空间中的 URL 被发现进程继续标记为要排除，并且将在被重新搜寻时从索引中除去。

全局搜寻空间由名为 `global.rules` 的配置文件定义，它必须存在于搜寻器配置目录中（`global.rules` 文件的存在使全局搜寻空间功能可正常发挥作用）。如果此文件存在，将在搜寻器初始化期间读它。如果此文件不存在，搜寻器将以单级搜寻空间运行，并在需要时从索引中除去文档以保持它的搜寻空间定义和数据库之间的一致性。

如果全局搜寻空间存在，搜寻器将和以前一样控制 URL 是否存在于搜寻空间中，但只有在 URL 不在任何 Web 搜寻空间中时才请求从索引中除去该 URL。

`global.rules` 文件的语法与局部 `crawl.rules` 文件相同，但它只包含域名规则。此限制使搜寻空间只根据 DNS 主机名，而不是 IP 地址或 HTTP 前缀模式在搜寻器之间进行分区。局部搜寻空间中被（`crawl.rules` 文件中定义的）URL 前缀规则或 IP 地址规则排除的 URL 不受全局搜寻空间影响；这样的 URL 仍将被排除。

全局搜寻空间仅用于防止从索引中除去被局部域规则从一个搜寻器的搜寻空间中排除的 URL。下列规则按下列顺序应用：

1. 如果局部前缀规则或地址规则排除了搜寻器数据库中的一个 URL，则对该 URL 指定返回码 760 并将其从索引中除去。该 URL 将不再被搜寻。
2. 如果局部域规则排除了搜寻器数据库中的一个 URL，并且没有全局搜寻空间，则对该 URL 指定返回码 760 并将其从索引中除去。该 URL 将不再被搜寻。
3. 如果局部域规则排除了搜寻器数据库中的一个 URL，但全局搜寻空间中的规则明确允许该 URL，则对该 URL 指定返回码 761。搜寻器不会再次搜寻该 URL，但也不会从索引中除去它（假定它存在于另外某个搜寻器的局部搜寻空间中）。
4. 如果局部域规则排除了搜寻器数据库中的一个 URL，并且全局搜寻空间中的规则未明确允许该 URL，则对该 URL 指定返回码 760，并从索引中除去该 URL。

由于查询全局搜寻空间的目的仅仅是为了防止删除已被局部搜寻空间排除的 URL，因此，当没有规则适用于候选 URL 时，全局搜寻空间的缺省结果是禁止它被搜寻。

`global.rules` 文件必须存在于共享全局搜寻空间的每个搜寻器的 `master_config` 目录中。必须仔细地编辑 `global.rules` 文件的所有副本以及各个局部 `crawl.rules` 文件，以确保它们是一致的。

No-follow 和 no-index 伪指令

通过对 Web 搜寻器指定伪指令来控制是否按照页面上的链接进行搜寻以及是否对页面建立索引，可提高搜索质量。

某些 Web 页面中有 no-follow 或 no-index 伪指令，这些伪指令指示机器人（如 Web 搜寻器）不要按照在那些页面中找到的链接进行搜寻，不要包括索引中那些页面的内容，或不要执行这两项操作中的任何一项。

控制这些设置可提高搜寻质量。例如，某些目录页面可能包含了数以千计的链接，但没有其他有用的内容；应该搜寻那些页面并按照它们的链接进行搜寻，但对目录页面本身建立索引则没有必要。

有时候，可能不想让搜寻器搜寻到层次结构中更低的层次，但所需的叶子页中包含了链接但未包含 no-follow 伪指令。由于这些页面中的某些页面是自动生成的，所以它们没有所有者可插入所需的伪指令。

要指定搜寻这些页面的规则，可创建或编辑一个配置文件，文件名是 followindex.rules。在此文件中指定规则时，请使用以下准则：

- 配置的规则必须指定 URL 前缀（不能通过 IP 地址或 DNS 主机名来标识 Web 站点）。
- URL 前缀可包括星号（*）作为通配符，以允许或禁止带有类似 URL 的多个站点。
- 顺序很重要（搜寻器将应用与候选 URL 匹配的第一个规则）。
- 那些明确允许和禁止按照链接进行搜寻或建立索引的规则将覆盖其他设置，包括目标文档中的那些设置。

覆盖 Web 页面中的 no-follow 和 no-index 伪指令

可以在配置文件中指定规则，以便控制 Web 搜寻器是否按照指向那些包含了 no-follow 或 no-index 伪指令的页面的链接进行搜寻或对那些页面建立索引。

在开始之前

要对 Web 搜寻器指定 no-follow 和 no-index 伪指令，您必须是企业搜索管理员。您指定的伪指令将覆盖将要搜寻的页面中现有的伪指令。

过程

要覆盖 no-follow 和 no-index 伪指令：

1. 从企业搜索管理控制台中，监视想要为其配置规则的集合，然后停止 Web 搜寻器。
2. 作为企业搜索管理员登录到搜寻器服务器上（此用户标识是在安装 WebSphere II OmniFind Edition 时指定的）。
3. 切换到想要配置的搜寻器的配置目录，其中 *crawler_session_ID* 是由企业搜索系统分配给搜寻器会话的标识。例如：

```
ES_NODE_ROOT/master_config/col_56092.WEB_88534
```

4. 创建或编辑名为 followindex.rules 的文件。
5. 使用以下格式输入搜寻器的规则，其中 *URLprefix* 是一些 Web 站点的开始字符，想要允许或禁止按照这些站点进行搜寻或对它们建立索引：

```
forbid follow URLprefix  
allow follow URLprefix  
forbid index URLprefix  
allow index URLprefix
```

6. 保存然后退出该文件。
7. 从企业搜索管理控制台中，重新启动停止的 Web 搜寻器。

WebSphere Portal 搜寻器

要在企业搜索集合中包括来自 IBM WebSphere Portal 站点的页面，必须配置 WebSphere Portal 搜寻器。

WebSphere Portal Server 配置

在创建 WebSphere Portal 搜寻器之前，必须在 WebSphere Portal 中部署企业应用程序 ESPACServer.ear。这个企业应用程序是在安装 WebSphere Information Integrator OmniFind Edition 时安装在搜索服务器上的。要部署这个企业应用程序，请完成下列任务：

- 『为 WebSphere Portal 搜寻器部署企业应用程序』。

配置概述

可以使用 WebSphere Portal 搜寻器搜寻单个 WebSphere Portal 站点。配置搜寻器时，指定所要搜寻的门户网站站点的 URL。于是，搜寻器将下载指定门户网站上的可用 portlet。

要创建或更改 WebSphere Portal 搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务：

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何搜寻站点上的所有页面。
- 指定所要搜寻的门户网站站点的 URL，并指定使搜寻器能够连接到该站点的信息。由于这些类型的 URL 可能很长并且包含已编码的非 ASCII 字符，所以，您可能希望从 WebSphere Portal Server 复制 URL 并将其粘贴到企业搜索管理控制台中。
- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

也可以选择一个选项以便在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。

创建搜寻器时，可以单击帮助以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

为 WebSphere Portal 搜寻器部署企业应用程序

在创建 WebSphere Portal 搜寻器之前，必须在 WebSphere Portal 中部署企业应用程序 ESPACServer.ear。

关于本任务

ESPACServer.ear 文件是在安装 WebSphere II OmniFind Edition 时安装到搜索服务器上的 ES_INSTALL_ROOT/bin 目录中的。缺省安装路径如下所示：

UNIX 系统：

/opt/IBM/es/bin/ESPACServer.ear

Windows 系统:

C:\Program Files\IBM\es\bin\ESPACServer.ear

过程

要部署企业应用程序以使 WebSphere Portal 搜寻器能够搜寻 WebSphere Portal 站点:

1. 停止 WebSphere_Portal 服务器实例。
2. 如果 WebSphere Application Server server1 服务器实例尚未启动, 则启动它。
3. 在 WebSphere Portal Server 上, 启动 WebSphere Application Server 管理控制台。
如果提示您登录, 则进行登录。

以下列方式打开管理控制台:

- 使用 Windows 的开始菜单来选择程序。
 - 对于 WebSphere Application Server V5, 打开 Web 浏览器并访问 `http://hostname:port/admin`, 其中 `hostname` 是搜索服务器的主机名, `port` 是 WebSphere Application Server 管理控制台的端口号。通常, 管理控制台端口是 9090。
 - 对于 WebSphere Application Server V6, 打开 Web 浏览器并访问 `http://hostname:port/ibm/console`, 其中 `hostname` 是搜索服务器的主机名, `port` 是 WebSphere Application Server 管理控制台的端口号。通常, 管理控制台端口是 9060。
4. 单击应用程序, 然后单击安装新应用程序。
 5. 单击浏览并从系统中选择 ESPACServer.ear 文件。
 6. 单击两次下一步。如果接收到关于策略文件的警告, 则单击继续。
 7. 单击下一步直到看到“将模块映射至应用程序服务器”页时, 进行下列选择, 然后单击应用:
 - a. 在集群和服务器文本框中, 选择 **WebSphere:cell=cell_name, node=node_name, server=WebSphere Portal**。
 - b. 选中 **ESPACServer.ear** 模块旁边的复选框。
 8. 单击下一步直到看到“总结”页为止, 然后单击完成。
 9. 单击保存至主配置链接, 然后单击保存按钮以保存对 WebSphere Application Server 配置所作的更改。
 10. 重新启动 WebSphere Portal Server。

相关概念

第 181 页的『企业搜索与 WebSphere Portal 的集成』

可以通过在 WebSphere Portal 和 WebSphere Portal 搜索中心中部署企业搜索 portlet 来扩展 IBM WebSphere Portal 的搜索能力。

从 WebSphere Portal 复制要搜寻的 URL

为了降低输入错误 URL 的可能性, 在配置 WebSphere Portal 搜寻器时, 可以将所要搜寻的 WebSphere Portal 站点的 URL 复制并粘贴到相应的字段中。

关于本任务

在创建 WebSphere Portal 搜寻器时，需要指定所要搜寻的 WebSphere Portal Server 上的门户网站 URL。由于 URL 会很长并且通常包含已编码的非 ASCII 字符，所以，可以使用此过程来从 WebSphere Portal Server 复制 URL 并将其粘贴到企业搜索管理控制台中。

过程

要指定让 WebSphere Portal 搜寻器搜寻的 URL:

1. 当在企业搜索管理控制台中看到要搜寻的 WebSphere Portal Server 页时，请确保 WebSphere Portal Server 已启动，然后作为管理员登录到 WebSphere Portal。
2. 从右上角选择**管理**。
3. 从左边的导航区中选择**门户网站设置**，然后选择**搜索管理**。
4. 在“管理搜索集合”页上，从“搜索集合”区域中选择 **PortalCollection**。（如果有其他集合可用，也可以选择别的集合。）
5. 在“集合中的内容源”区域中，单击**添加内容源**。
6. 对于**搜寻源类型**，选择 **Portal 站点**。站点 URL 将显示在收集从此 **URL 链接的文档**字段中。
7. 将该 URL 复制到剪贴板。（例如，选中该 URL，然后按住 Ctrl 键不放并按 Insert 键。）
8. 返回企业搜索管理控制台，将复制的 URL 粘贴到 **WebSphere Portal 站点 URL** 字段中。

Windows 文件系统搜寻器

要在企业搜索集合中包括 Microsoft Windows 文件系统中存储的文档，必须配置 Windows 文件系统 搜寻器。

可以使用 Windows 文件系统 搜寻器搜寻任意数目的 Windows 文件系统。配置该搜寻器时，选择要搜寻的本地目录和子目录以及远程目录和子目录。

如果将搜寻器服务器安装在 UNIX 计算机上，则不能使用该服务器来搜寻 Windows 文件系统源（Windows 文件系统搜寻器不会出现在可用搜寻器类型的列表中）。

要创建或更改 Windows 文件系统搜寻器，请登录到企业搜索管理控制台。您必须是企业搜索管理员角色的成员或者是拥有该搜寻器的集合的集合管理员。

创建搜寻器时，将有一个向导帮助您执行下列任务:

- 指定控制搜寻器如何操作和使用系统资源的属性。搜寻器属性控制搜寻器如何在搜寻空间中搜寻所有子目录。
- 设置用于搜寻文件系统的时间表。
- 选择要搜寻的子目录。

可以指定要让搜寻器搜寻的子目录层数。要搜寻远程文件系统，还需指定使搜寻器能够访问数据的用户标识和密码。

- 指定用于使子目录中的文档可供搜索的选项。例如，可以将某些类型的文档排除在搜寻空间之外，也可以指定一个用户标识和密码以使搜寻器访问特定子目录中的文件。

- 配置文档级安全性选项。如果创建集合时启用了安全性，则搜寻器可以使安全性数据与索引中的文档相关联。此数据使搜索应用程序能够根据存储的访问控制表或安全性令牌来强制实施访问控制。

也可以选择一个选项以便在用户提交查询时验证用户凭证。在这种情况下，系统将用户凭证与原始数据源维护的当前访问控制表进行比较，而不是将用户凭证与建立了索引的安全性数据进行比较。

要强制实施文档级安全性，必须确保在搜寻器服务器上正确地配置了用户和域帐户信息。

创建搜寻器时，可以单击[帮助](#)以了解有关向导中的字段的信息以及如何向搜寻器提供它搜寻数据时所需的信息。

相关概念

第 175 页的『对 Windows 文件系统文档强制实施文档级安全性』

为了在用户搜索 Windows 文件系统 搜寻器搜寻的文档时验证当前凭证，必须在搜寻器服务器和 Microsoft Windows 服务器上配置域帐户信息。

配置对数据侦听器应用程序的支持

通过使用数据侦听器 API 来创建外部搜寻器，可以扩展企业搜索。定制数据侦听器应用程序可以对集合添加数据、从集合中除去数据或者指示 Web 搜寻器访问和重新访问 URL。

在开始之前

要配置数据侦听器应用程序，您必须是企业搜索管理员角色的成员。

关于本任务


客户机数据侦听器应用程序使您能够搜寻缺省企业搜索搜寻器无法搜寻的数据源类型。在可以使用数据侦听器应用程序之前，必须配置凭证以使该应用程序能够访问和更新集合。

当客户机数据侦听器应用程序连接至数据侦听器时，它必须传递客户机应用程序标识和密码以及所要更新的集合的标识。此信息必须与在管理控制台中为应用程序配置的信息匹配。

数据侦听器是在企业搜索系统启动时自动启动的。如果在管理控制台中配置应用程序后更改了端口号，就必须重新启动数据侦听器。



过程

要配置数据侦听器应用程序：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在数据侦听器页上，单击**配置数据侦听器应用程序**。
4. 在数据侦听器应用程序页上，指定数据侦听器为了处理来自客户机应用程序的请求而可以创建的线程数以及数据侦听器侦听请求时使用的端口号。
5. 单击**添加数据侦听器应用程序**以添加关于客户机应用程序的信息。

6. 在添加数据侦听器应用程序页上，指定认证信息以使客户机数据侦听器应用程序能够访问企业搜索集合。数据侦听器客户机标识在企业搜索系统中必须是唯一的。
7. 选择数据侦听器应用程序能够更新的集合：
 - 如果要想让该应用程序更新所有集合，则单击**所有集合**。
 - 如果要想让该应用程序仅更新指定的集合，则单击**特定集合**。

当选择此选项时，将显示一组集合名。对于该应用程序能够更新的每个集合，选中**选择复选框**。

8. 单击**确定**。
9. 如果更改了数据侦听器端口号，则重新启动数据侦听器：
 - a. 单击  **监视**以切换到系统监视视图。
 - b. 在数据侦听器页上，单击  **重新启动**。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Data listener”

相关任务

第 214 页的『监视数据侦听器』

监视数据侦听器以查看其状态和查看有关客户机数据侦听器应用程序活动的详细信息。

定制搜寻器插件

在配置搜寻器属性时，可以指定一个 Java 类，这个 Java 类用来强制实施文档级访问控制并使元数据与企业搜索索引中的文档相关联。

插件包含一个 Java 类，将对搜寻器搜寻的每个文档调用该 Java 类。将把企业搜索索引、安全性令牌和元数据中的文档标识（URI）传递给该 Java 类。该类返回一组新的或修改过的安全性令牌和元数据，该类也可以指示某个文档将被搜寻器忽略。

在搜寻空间中的所有文档都被搜寻一次之后，将仅对新文档的修改过的文档调用该插件。要为已在企业搜索索引中但在原始数据源中尚未更新的文档更改安全性令牌和元数据，请启动对搜寻空间中的所有文档的完整搜寻，然后重组索引。

不能使插件与现有的搜寻器相关联。在配置新搜寻器的属性时，必须指定插件类名和类路径。

使用插件来强制实施安全性

文档级安全性是通过使一个或多个安全性令牌（逗号定界的字符串）与搜寻器搜寻的每个文档相关联来强制实施的。通常将组标识用作安全性令牌。

缺省情况下，对每个文档指定一个公共令牌，这使得每个人都可以使用该文档。可以将公共令牌替换为管理员提供的值或者从搜寻到的文档的字段中抽取的值。

该插件允许应用自己的业务规则以确定搜寻到的文档的安全性令牌值。与每个文档相关联的安全性令牌存储在索引中。这些安全性令牌用来过滤与安全性令牌匹配的文档并确保在搜索结果中只返回用户有权查看的文档。

使用插件来将元数据添加到文档中

将为搜寻到的所有文档创建文档元数据，例如文档的上次修改日期。搜寻器插件允许应用自己的业务规则以确定要对每个文档的哪些元数据值建立索引。

元数据是以“名称 - 值”对形式创建的。用户可以使用自由文本查询或指定了元数据字段名的查询来搜索元数据。

Web 搜寻器插件

使用 Web 搜寻器的应用程序编程接口，可以控制文档搜寻方式以及准备文档以便进行解析的方式。例如，可以将字段添加到搜寻器请求文档时将要使用的 HTTP 请求头中。搜寻文档后，并且在对其进行解析和标记化之前，可以更改内容、安全性令牌和元数据。还可以停止将文档发送至解析器。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Crawler plug-ins”

企业搜索索引中的 URI 格式

企业搜索索引中的每个文档的统一资源标识 (URI) 都指示了将文档添加到该集合的搜寻器的类型。

可以在为集合配置类别、作用域和快速链接时指定 URI 或 URI 模式。也可以在需要从索引中除去文档或查看有关特定 URI 的详细状态信息时指定 URI。

搜索集合以确定文档的 URI 或 URI 模式。可以单击搜索结果中的 URI 来检索您感兴趣的文档。可以从搜索结果中复制 URI 以便在企业搜索管理控制台中使用该 URI。例如，可以指定 URI 模式以便自动地使与该 URI 模式匹配的文档与企业搜索类别相关联。

Content Edition 搜寻器

Content Edition 搜寻器在服务器访问方式下搜寻的文档的 URI 格式为:

```
vbr://Server_Name/Repository_System_ID/Repository_Persistent_ID  
/Item_ID/Version_ID  
/Item_Type/[?Page=Page_Number&] JNDI_properties
```

Content Edition 搜寻器在直接访问方式下搜寻的文档的 URI 格式为:

```
vbr:///Repository_System_ID/Repository_Persistent_ID  
/Item_ID/Version_ID  
/Item_Type/[?Page=Page_Number]
```

参数 URL 编码被应用于所有字段。

Server_Name

WebSphere Information Integrator Content Edition 服务器的名称。

Repository_System_ID

存储库的系统标识。

Repository_Persistent_ID

存储库的持久标识。

Item_ID

项的标识。

Version_ID

版本的标识。如果版本标识为空白，则此值指示文档的最新版本。

Item_Type

项的类型 (CONTENT 或 FOLDER)。

Page_Number

页号。

JNDI_properties

J2EE 应用程序客户机的 JNDI 属性。有两种类型的属性：

java.naming.factory.initial

用来创建 EJB 句柄的应用程序服务器的类名。

java.naming.provider.url

用来请求 EJB 句柄的应用程序服务器的命名服务的 URL。

示例 Documentum:

```
vbr://vbrsrv.ibm.com/Documentum/c06b/094e827780000302//CONTENT/?  
java.naming.provider.url=iop%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=iop%3A%2F%2Fmyvbr.ibm.com%3A2809&  
java.naming.factory.initial=com.ibm.websphere.naming.WsnInitContextFactory
```

DB2 搜寻器

由 DB2 搜寻器搜寻的文档的 URI 格式为:

```
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]
```

参数: URL 编码被应用于所有字段。

Database_Name

数据库的内部名称或数据库的别名。

Table_Name

目标表的名称, 包括模式的名称。

Unique_Identifier_Column_Name1

表中的第一个唯一标识列的名称。

Unique_Identifier_Value1

第一个唯一标识列的值。

Unique_Identifier_Column_NameN

表中的第 *n* 个唯一标识列的名称。

Unique_Identifier_ValueN

第 *n* 个唯一标识列的值。

示例 本地已编目的数据库:

db2://LOCALDB/SCHEMA1.TABLE1/MODEL/ThinkPadA20

远程未编目的数据库:

db2://myserver.mycompany.com:50001/REMOTEDB/SCHEMA2.TABLE2/NAME/DAVID

DB2 Content Manager 搜寻器

由 DB2 Content Manager 搜寻器搜寻的文档的 URI 格式为:

cm://Server_Name/Item_Type_Name/PID

参数

Server_Name

IBM DB2 Content Manager 库服务器的名称。

Item_Type_Name

目标项类型的名称。

PID DB2 Content Manager 持久标识。

示例

cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icmnlbdb12+ITEMTYPE159+26+A1001001A
03F27B94411D1831718+A03F27B+94411D183171+14+1018

Domino Document Manager 搜寻器

由 Domino Document Manager 搜寻器搜寻的文档的 URI 格式为:

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

参数 URL 编码被应用于所有字段。

Server_Name

Domino Document Manager 服务器的名称。

Port_Number

可选: Domino Document Manager 服务器的端口号。

Database_Replica_ID

数据库副本的标识。

Database_Path_and_Name

目标 Domino Document Manager 服务器上的文档 NSF 数据库的路径和文件名。

View_Universal_ID

用来搜寻 Domino Document Manager 文档的视图通用标识。

Document_Universal_ID

所搜寻的文档中定义的文档通用标识。

Attachment_Number

可选: 每个附件的连续编号 (从零开始)。

Attachment_File_Name

可选: 附件文件的原始名称。

示例 Domino Document Manager 文档:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226

文档附件:

dominodoc://dominodocsvr.ibm.com/49256D3A000A20DE/domdoc%2FADMN-6FAJXL.nsf
/8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
?AttNo=0&AttName=AttachedFile.doc

Exchange Server 搜寻器

由 Exchange Server 搜寻器搜寻的文档的 URI 格式为:

exchange://*OWA_path*[?useSSL=true]

参数

OWA_Path

Outlook Web Access (OWA) 路径, 不带协议。

useSSL=true

当原始 OWA 路径的协议是 HTTPS 时添加此内容。

示例 文档主体

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML

文档附件:

exchange://exchangesvr.ibm.com/public/RootFolder1/Folder1/Document.EML/
AttachedFile.doc

已启用 SSL:

exchange://exchangesvr.ibm.com/public/TeamRoom/Folder1/Document.EML
?useSSL=true

Notes 搜寻器

由 Notes 搜寻器搜寻的文档的 URI 格式为:

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*]

参数 URL 编码被应用于所有字段。

Server_Name

Lotus Notes 服务器的名称。

Port_Number

Lotus Notes 服务器的端口号。端口号是可选的。

Database_Replica_ID

数据库副本的标识。

Database_Path_and_Name

目标 Lotus Notes 服务器上的 NSF 数据库的路径和文件名。

View_Universal_ID

对目标数据库定义的视图通用标识。仅当从视图或文件夹中选择文档时才指定此标识。如果未指定要搜寻的视图或文件夹 (例如, 如果指定想要搜寻数据库中的所有文档), 则未指定视图通用标识。

Document_Universal_ID

在由搜寻器搜寻的文档中定义的文档通用标识。

Attachment_Number

每个附件的连续编号（从零开始）。附件号是可选的。

Attachment_File_Name

附件的原始名称。附件名是可选的。

示例 已选择为要按视图或文件夹进行搜寻的文档:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
8178B1C14B1E9B6B8525624F0062FE9F/0205F44FA3F45A9049256DB20042D226
```

未选择为要按视图或文件夹进行搜寻的文档:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
0205F44FA3F45A9049256DB20042D226
```

文档附件:

```
domino://dominosvr.ibm.com/49256D3A000A20DE/Database.nsf/  
0205F44FA3F45A9049256DB20042D226?AttNo=0&AttName=AttachedFile.doc
```

QuickPlace 搜寻器

由 QuickPlace 搜寻器搜寻的文档的 URI 格式为:

```
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
```

参数 URL 编码被应用于所有字段。

Server_Name

Lotus QuickPlace 服务器的名称。

Port_Number

可选: QuickPlace 服务器的端口号。

Database_Replica_ID

数据库副本的标识。

Database_Path_and_Name

目标 QuickPlace 服务器上的文档 NSF 数据库的路径和文件名。

View_Universal_ID

用来搜寻 QuickPlace 文档的视图通用标识。

Document_Universal_ID

所搜寻的文档中定义的文档通用标识。

Attachment_Number

可选: 每个附件的连续编号（从零开始）。

Attachment_File_Name

可选: 附件文件的原始名称。

示例 文档:

```
quickplace://ltsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498
```

页面附件:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csampleplace%5CPageLibrary4925704300021490.nsf/A7986FD2A9CD47090525670800167225/2B02B1DE3A82B2CE49257043001C2498?AttNo=0&AttName==QPCons3.ppt
```

UNIX 文件系统搜寻器

由 UNIX 文件系统搜寻器搜寻的文档的 URI 格式为:

```
file:///Directory_Name/File_Name
```

参数 URL 编码被应用于所有字段。

Directory_Name

目录的绝对路径名。

File_Name

文件名。

示例

```
file:///home/user/test.doc
```

WebSphere Portal 搜寻器

由 WebSphere Portal 搜寻器搜寻的文档的 URI 格式为:

```
wps://WPS_Access_Path?portletDefID=Portlet_Def_ID&pageID=Page_ID&useSSL=SSL
```

参数 URL 编码被应用于所有字段。

WPS_Access_Path

WebSphere Portal 服务器路径 (不带协议)。

Portlet_Def_ID

WebSphere Portal 服务器的 portlet 定义标识。

Page_ID

WebSphere Portal 服务器的页面标识。

SSL 当路径的协议是 HTTPS 时, 将添加 `useSSL=true`。否则, 将添加 `useSSL=false`。

示例 文档主体

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!?portletDefID=3_0_3S&pageID=6_0_6J&useSSL=false
```

示例 已启用 SSL:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+LKnPy1vM0Y_QjzKCN4g3cQbJgQi0-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

Windows 文件系统 搜寻器

由 Windows 文件系统 搜寻器搜寻的文档的 URI 格式为:

```
file:///Directory_Name/File_Name
```

```
file:///Network_Folder_Name/Directory_Name/File_Name
```

参数 URL 编码被应用于所有字段。

Directory_Name

目录的绝对路径名。

File_Name

文件名。

Network_Folder_Name

(仅限于远程服务器上的文档) Windows 网络上的共享文件夹的名称。

示例 本地文件系统:

file:///d:/directory/test.doc

网络文件系统:

file:///filesvr.ibm.com/directory/file.doc

相关概念

第 31 页的『企业搜索搜寻器管理』

要包括在集合中的不同类型的数据配置搜寻器。单个集合可以包含任意数目的搜寻器。

相关任务

第 94 页的『配置类别』

可以为集合创建任意数目的类别，并且每个类别可以包含任意数目的规则。规则确定哪些文档自动地与该类别相关联。

第 123 页的『配置作用域』

在企业搜索集合配置作用域时，指定索引中允许用户搜索的文档范围的 URI 或 URI 模式。

第 126 页的『从索引中除去 URI』

要防止用户搜索集合中的文档，可以从索引中除去那些文档的 URI。

第 134 页的『配置快速链接』

为企业搜索集合创建快速链接，使文档的 URI 与一些关键字相关联，这些关键字将导致该文档包括在搜索结果中。

第 201 页的『查看关于 URI 的详细信息』

您可以查看有关 URI 的详细信息。您可以查看关于由此 URI 表示的文档是如何被搜寻、被建立索引和被搜索的当前信息和历史信息。

第 215 页的『查看关于废弃的文档的报告』

您可以查看关于已从企业搜索系统中废弃的文档的详细信息。仅当对集合启用了文档跟踪功能时，才会提供此信息。

企业搜索解析器管理

要增强文档的可检索性，可以指定选项控制在将文档及元数据添加到企业搜索索引之前如何对它们进行解析、分析和分类。

可以指定用于解析文档内容和优化信息的可检索性的选项包括以下各项：

配置解析中文、日语和韩国语文档的选项

可以对使用 n-gram 分段来解析以中文、日语和韩国语编写的文档指定选项。也可以从中文和日语文档的空格中除去换行符。

启用本机 XML 搜索

如果您的集合包括 XML 文档，可以使它们能通过本机 XML 查询语法（例如，XPath 和 XML 片段）来进行搜索。本机 XML 搜索使用户能根据各种 XML 元素之间的关系来指定查询。

配置类别

可以将共享类似 URI 模式或包含特定单词的文档分组到类别中。当用户搜索集合时，他们可以将搜索结果限制为那些属于特定类别的文档。

配置搜索字段

可以将 XML 文档中的元素映射至索引中的搜索字段。也可以将 HTML 文档中的元数据元素映射至搜索字段。通过在企业搜索索引中创建搜索字段，可以使用户能查询 XML 和 HTML 文档中的特定部分，并提高搜索结果的精确度。

配置文本处理选项

如果已向企业搜索系统添加了定制文本分析引擎，可以选择一个分析引擎用于集合。在将分析引擎与集合关联后，可以指定映射内容选项，以便可进行语言分析和注释。还可以指定分析结果如何映射至企业搜索索引或 JDBC 数据库表。

将字段映射至升降类

对于包含与查询项匹配的字段的文档，可以指定它们在搜索结果中的排名高于其他与查询项匹配的文档。在将字段映射至升降类时，可以指定要对哪些内容和元数据字段进行升降处理。还可以配置每个升降类在影响文档排名时使用的分数。

相关概念

“Text Analysis Integration”中的“Linguistic support for semantic search”

“Text Analysis Integration”中的“Text analysis included in enterprise search”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Semantic search applications”

“Text Analysis Integration”中的“Semantic search query”

使用类别

类别使您能够对共享公共特征的文档进行分组以及仅搜索和检索符合作为该组成员这一条件的文档。

如果将文档与类别相关联，并且搜索应用程序支持此功能，则用户可以通过指定类别名来搜索集合的子集。如果用户搜索整个集合，则他们可以优化搜索结果并且仅浏览与其中一个结果文档位于同一类别中的文档。

企业搜索支持两种文档分类方法：

基于规则

如果要根据指定的规则来使文档与类别相关联，则使用此方法。例如，可以对共享 URI 模式的文档进行分组或根据文档内容（例如，包含或不包含特定词和短语）对文档进行分组。

基于模型

如果使用 IBM WebSphere Portal 并且要使用 WebSphere Portal 中的预定义类别来搜索企业搜索集合，则使用此方法。

应用分类更改

要创建和管理类别，请使用企业搜索管理控制台：

- 在创建集合时选择分类类型。可以选择不使用类别、使用基于规则的类别或使用基于模型的类别。
- 在配置集合的解析规则时，可以更改分类类型（如果有必要的话）。如果在搜寻文档并对文档建立索引之后更改分类类型，搜索质量将会下降，直到重新搜寻所有文档，重新解析文档并重组索引为止。
- 如果选择使用基于规则的类别，则使用管理控制台来管理类别树、类别和分类规则。如果在搜寻文档并对文档建立索引之后更改类别或分类规则，搜索质量将会下降，直到重新搜寻所有文档，重新解析文档并重组索引为止。

文档内容字段

要创建集合并且要让该集合包含基于模型的类别，或者要让该集合包含使用文档内容规则（而不是使用 URI 模式规则）的基于规则的类别，应确保要分类的文档包含内容字段。

基于模型的类别和以文档内容为基础的分类规则对文档的内容字段进行操作，而不考虑文档的元数据。只有包含内容字段的文档，或包含可在配置搜寻器时配置为内容字段的字段的文档，可使用这些分类选项进行分类。

如果配置基于模型的类别或者创建基于文档内容的分类规则，那些规则将仅应用于文档的内容部分。对于下列搜寻器类型来说，这些规则将应用于文件（例如 UNIX 或 Microsoft Windows 文件内容）或 Web 页面内容：

- Exchange Server
- NNTP
- UNIX 文件系统
- Web
- WebSphere Portal
- Windows 文件系统

对于下列搜寻器类型来说，在为各个数据源配置选项时指定哪些字段是文档内容字段。在创建搜寻器时，此选项是在搜寻器向导的最后一页上提供的。如果搜寻器已存在，则编辑搜寻空间，选择数据源以编辑其选项，然后指定哪些字段是该数据源中的文档内容字段。

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

基于规则的分类

可以配置规则以控制哪些文档与企业搜索集中的类别相关联。

可以为在企业搜索中创建的集合以及从 IBM WebSphere Portal 集合迁移的基于规则的分类创建类别和分类规则。要配置用于对文档进行分类的规则，必须在创建集合时或者在指定集合的解析规则时指定要使用基于规则的类别。

解析器使用指定的规则来使文档与一个或多个类别相关联：

- 如果文档符合类别中的至少一个规则，则解析器就使该文档与该类别相关联。
- 如果文档符合数个类别中的至少一个规则，则解析器就使该文档与所有类别相关联。
- 如果文档不符合类别的任何规则，则解析器不会使该文档与类别相关联。当用户搜索集合时，他们可以搜索此文档并检索它，但是他们不能搜索类别并期望检索该文档。

当管理集合的类别树（即分类法）时，决定要在类别层次结构添加新类别的位置。还可以使用类别树来选择要编辑的类别，然后添加用于对文档进行分类的规则、删除规则或更改各个规则的内容。

在配置用于对文档进行分类的规则时，选择企业搜索是使用文档的 URI 还是文档中的内容来确定文档是否属于该类别：

URI 模式

URI 规则应用于文档的 URI。指定部分 URI（一个模式），那些在 URI 中具有指定模式的文档将符合该规则。

例如，如果指定规则文本为 `/hr/`，则下面第一个 URI 将符合该规则，第二个 URI 却不符合该规则：

```
file:///corporate/hr/medicalform.doc
http://company.com/human_resources/medicalform.htm
```

由于所有 URI 都被视为模式，所以系统将忽略模式开头或末尾指定的任何星号通配符。例如，`*/hr/*` 和 `/hr/` 将与同一组 URI 匹配。

URI 模式规则不区分大小写。如果 URI 包含空格，则 URI 模式必须符合用于编码 URI 的企业搜索规则。以下示例显示了为 Windows 文件系统路径指定 URI 的正确方法和不正确方法：

不正确的 URI: file:///c:/program files/

正确的 URI: file:///c:/program+files/

文档内容

按照与查询相同的格式来表达文档内容规则。如果该文档对于查询来说是有效的，则将符合此规则。在配置规则时，指定文档必须包含或不包含的词和短语，并且选择一种语言以应用单词衍生形态查询规则。

例如，以下规则指定文档如果包含 `hr` 一词或短语 `human resources`，该文档就符合该规则：

```
hr "human resources"
```

再例如，以下规则指定如果文档包含 `hr` 一词但不包含 `benefits` 一词，该文档就符合该规则：

```
+hr -benefits
```

内容规则必须进行与搜索和索引 API (SI-API) 查询相同的语言规范化。但是，内容规则的语法支持一小部分在 SI-API 查询语法中可用的操作。只允许下列查询运算符：

- + 在查询项前加一个加号表示该查询项必须出现在文档中。
- 在查询项前加一个减号表示该查询项不能出现在文档中。
- " 将两个或多个查询项包括在引号中，表示引号中的短语项必须精确出现在文档中。

文档内容规则仅应用于文档的内容部分。对于下列搜寻器类型来说，这些规则将应用于文件（例如 UNIX 或 Microsoft Windows 文件内容）或 Web 页面内容：

- Exchange Server
- NNTP
- UNIX 文件系统
- Web
- WebSphere Portal
- Windows 文件系统

对于下列搜寻器类型来说，在为各个数据源配置选项时指定哪些字段是文档内容字段。在创建搜寻器时，此选项是在搜寻器向导的最后一页上提供的。如果搜寻器已存在，则编辑搜寻空间，选择数据源以编辑其选项，然后指定哪些字段是该数据源中的文档内容字段。

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

相关任务

第 190 页的『从 WebSphere Portal 迁移集合』

要将集合从 WebSphere Portal 迁移至企业搜索，请在 WebSphere Portal 中准备集合，然后使用迁移向导来迁移它们。

基于模型类别

如果在 IBM WebSphere Portal 系统中使用了基于模型的类别，则可以对企业搜索集合继续使用那些类别。

WebSphere Portal 提供包含超过 2,300 个主题的预定义分类法。这些主题分组为主要的商业类别领域，如计算机、金融和运输。Portal 用户可以创建自动确定哪些文档与这些主题领域相匹配的应用程序，并且他们可以为他们自己的企业需求定制这些类别。

如果要对企业搜索使用 WebSphere Portal 类别，则必须：

- 使用迁移向导来将基于模型的分法文件导入到企业搜索中。
- 在创建集合时或者在为集合配置解析规则时指定要使用的基于模型的类别。
- 确保在企业搜索索引服务器中安装了 WebSphere Portal。
- 使用 WebSphere Portal 中的分类工具来管理类别。不能使用企业搜索管理控制台来管理基于模型的类别。

相关任务

第 189 页的『从 WebSphere Portal 迁移基于模型的分法』

通过使用 WebSphere Portal 分法管理 Portlet，可以选择要与企业搜索集合配合使用的基于模型的分法。已迁移到企业搜索的集合不受新的分法迁移所影响。

类别树

类别树使您能够查看集合中所有基于规则的类别。使用类别树来创建类别、删除类别和编辑使文档与类别相关联的规则。

类别树（也称为分法）是以层次结构格式安排的。此树以根类别开始，所有其他类别从根类别衍生。可以嵌套任意数目的类别和子类别以向用户提供用于浏览和检索文档的不同选项。

例如，如果某个文档符合数个类别中的规则，则该文档与所有那些类型相关联。当用户搜索类别或者在浏览搜索结果时浏览属于某个类别的文档时，文档属于多个类别这一事实能够提高用户能够找到该文档的可能性。

当管理类别树时，可以通过在现有类别下面嵌套新类别来控制哪些文档属于一个或多个类别。创建类别时，可以指定是要在根级别创建该类别还是作为另一个类别的子类别来创建它。还可以使用类别树来从集合中删除类别以及更改用于使文档与类别相关联的规则。编辑类别时，可以将类别重命名、添加或删除分类规则或者修改各个规则的内容。

在管理类别树时，请使用有关搜索和浏览行为的下列描述作为准则：

- 如果用户搜索高级别的类别，则将在该类别及其所有子类别中搜索与搜索条件相匹配的文档。如果用户搜索某个没有更多子类别的类别，将仅搜索该类别。
- 如果用户正在浏览搜索结果并选择用来浏览属于特定类别的文档的选项，则将仅显示该类别中的文档。任何子类别的名称也将显示在搜索结果中，因此用户能够在类别之间进行导航以及同时查看文档的多个子集。

相关任务

第 190 页的『从 WebSphere Portal 迁移集合』

要将集合从 WebSphere Portal 迁移至企业搜索，请在 WebSphere Portal 中准备集合，然后使用迁移向导来迁移它们。

选择分类类型

在选择分类类型时，指定用来使文档与集合中的类别相关联的方法。

在开始之前

要更改分类类型，您必须是企业搜索管理员角色的成员或者是所更改的集合的集合管理员。

关于本任务

分类类型是在创建集合时指定的。如果有必要的话，可以更改集合的文档分类方式。可以使用专门为集合配置的基于规则类别，可以使用 IBM WebSphere Portal 系统中现有的基于模型的类别，也可以不使用任何类别。

要点：如果在搜寻数据并为集合创建索引之后更改了分类类型，则索引将变为不一致。要确保搜索结果的准确性，请重新搜寻集合中的所有文档，并在解析文档之后重组索引。

过程

要选择分类类型：

1. 编辑集合，选择解析页，然后单击**选择分类类型**。
2. 在选择分类类型页上，选择下列其中一个选项：

无 如果不想对此集合中的文档进行分类，则选择此选项。

基于规则

如果要使用包含专门为此集合配置的分类规则的分类法，则选择此选项。

- 如果正在配置为企业搜索创建的集合，则选择此选项以指定用于对文档进行分类的类别名和规则。
- 如果正在配置从 WebSphere Portal 迁移的集合，则选择此选项以使用或更改已导入的基于规则的类别。

基于模型

如果要使文档与存在于 WebSphere Portal 系统中的基于模型的类别相关联，则选择此选项。要使用此选项，必须在企业搜索索引服务器上安装 WebSphere Portal。还必须使用 WebSphere Portal 中的分类工具来管理类别。

3. 单击**确定**。

配置类别

可以为集合创建任意数目的类别，并且每个类别可以包含任意数目的规则。规则确定哪些文档自动地与该类别相关联。

在开始之前

要配置类别，您必须是企业搜索管理员角色的成员或者是类别所属的集合的集合管理员。

必须选择使用基于规则的类别的选项来作为分类类型。

要获取有关如何指定使文档与类别相关联的规则示例，请在创建或编辑类别时单击帮助。

关于本任务

如果搜索应用程序启用了类别的支持，则用户可以通过指定类别名来搜索集合的子集。用户还可以在搜索结果中选择类别并且仅浏览属于所选类别的文档。

要点：如果在搜寻数据并为集合创建索引之后更改了类别或分类规则，则索引将变为不一致。要确保搜索结果的准确性，请重新搜寻集合中的所有文档、重新解析文档并重组索引。

过程

要配置类别：

1. 编辑集合，选择解析页，然后单击**配置类别树**。
2. 在类别树页中，在树中选择要添加类别的位置，然后单击**创建类别**。

如果选择根，则将在根级别创建新类别。如果选择类别名，则新类别将嵌套在类别树中所选类别下方。

将打开一个向导以帮助指定用于使文档与新类别相关联的规则：

- a. 在创建类别页上，输入类别的描述性名称，然后单击**下一步**。
- b. 在创建分类规则页上，单击**添加规则**。
- c. 在创建分类规则页上，在**规则名**字段中为该规则输入唯一的名称。此名称在集合中的所有类别中必须是唯一的。
- d. 指定要用来使文档与此类别相关联的规则，然后单击**确定**。
 - 如果要让企业搜索在确定文档是否属于该类别时使用文档的 URI，则单击 **URI 模式**，然后指定 URI 模式。

如果指定的文本存在于 URI 中，则该文档与类别相关联。

例如：`file:///c:/program+files/finance`

- 如果要让企业搜索在确定文档是否属于该类别时分析文档内容字段中的词，则单击**文档内容**，选择文档的语言，然后指定必须出现或不能出现在文档内容中的词。按照查询格式来表达规则（只允许使用 +、- 和 " " 查询运算符）。

如果文档包含或不包含指定的单词，则该文档与此类别相关联。

例如：`+finance -accounting +"fiscal year"`

- e. 单击**完成**。

新类别将与属于此集合的其他类别一起列在类别树页中。

相关概念

第 189 页的『从 WebSphere Portal 迁移到企业搜索』

企业搜索提供了一个迁移向导，您可以使用该向导来将分类法和集合从 IBM WebSphere Portal 迁移到企业搜索。

相关任务

第 189 页的『从 WebSphere Portal 迁移基于模型的分类法』

通过使用 WebSphere Portal 分类法管理 Portlet，可以选择要与企业搜索集合配合使用的基于模型的分类法。已迁移到企业搜索的集合不受新的分类法迁移所影响。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识（URI）都指示了将文档添加到该集合的搜寻器的类型。

使用 XML 搜索字段

如果要使用户能够搜索 XML 文档的特定部分，则将 XML 元素映射至搜索字段。

可以使用企业搜索管理控制台，将 XML 元素映射至搜索字段。Linux

XML 搜索字段

XML 搜索字段使用户能够查询 XML 文档的特定部分。

因为 XML 文档包含半结构化的和非结构化的文本，所以它们变得很常用。XML 的结构是封装的，该结构使用由围绕文本的 XML 元素明确定义的上下文。例如，作者的姓名可能以如下形式出现：

```
<author>John Smith</author>
```

在此上下文，文本 John Smith 标识 XML 文档的作者。

企业搜索可以使 XML 元素中的文本与搜索字段名相关联（即将文本映射至字段名）。在为集合配置解析选项时，指定要将哪些 XML 元素映射至哪些搜索字段名。通过将 XML 元素映射至搜索字段，可以使用户能够通过指定映射的字段名来搜索那些元素。与搜索所有文档内容的自由文本查询相比，搜索特定字段的查询可以提供更精确的搜索结果。

例如，如果集合包括 XML 文档，并且您指定要将 title 和 author 元素标记为索引中的搜索字段，则用户可以查询这些特定的元素。搜索 author:Smith 将找到在 author 元素中包含 Smith 的 XML 文档。

将 XML 元素映射至搜索字段

当将 XML 元素映射至搜索字段时，指定用户可以通过在查询中指定字段名来搜索哪些 XML 元素。

在开始之前

要将 XML 元素映射至搜索字段，您必须是企业搜索管理员角色的成员或者是 XML 文档所属的集合的集合管理员。

关于本任务

当创建 XML 字段映射或者添加、更改或删除现有 XML 字段映射中的字段时，该更改将在停止并重新启动解析器之后生效。新的和经过更改的映射将应用于在重新启动解析器之后解析的数据。新的和经过更改的映射不会影响已解析并建立索引的数据。

此任务使用以下样本 XML 文档来显示可以如何映射人员记录并使用户能够直接查询某些元素。

```
<?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>
```

过程

要将此示例中的 XML 元素映射至搜索字段：

1. 编辑集合，选择解析页，然后单击将 **XML 元素映射至字段**。
2. 在 XML 字段映射页上，单击**创建 XML 映射**。这就打开了创建 XML 字段映射页。
3. 在 **XML 根元素名字段**中，输入根元素名：personnel。

确保在这里指定的名称与想要搜索的 XML 文档中的根元素精确匹配。当对 XML 文档进行解析和建立索引时，企业搜索根据根元素名来选择要使用的映射。

4. 在 **XML 映射名字段**中，输入这组 XML 字段映射规则的名称。

在创建一组 XML 映射规则后，此名称将显示在 XML 字段映射页上，可以选择此名称以添加、删除或更改映射规则。

5. 将 XML 元素 jobrole 映射至名为 jobrole 的搜索字段：
 - a. 在**字段名字段**中，输入 jobrole。
 - b. 在 **XML 元素名字段**中，输入 jobrole。
 - c. 要使用户能够查询 jobrole 字段并在搜索结果中查看作业角色，请选中**字段搜索**和**搜索结果**复选框。
6. 将 XML 元素 jobroles 映射至同一个搜索字段：
 - a. 单击**添加字段**以将一个空行添加到字段映射规则列表。
 - b. 在**字段名字段**中，输入 jobroles。
 - c. 在 **XML 元素名字段**中，输入 jobrole。

提示：XML 元素名不需要与搜索字段名相匹配，并且可以将多个 XML 元素映射至同一个搜索字段。

- d. 要使用户能够查询 jobrole 字段并在搜索结果中查看作业角色，请选中**字段搜索**和**搜索结果**复选框。
7. 将具有属性 expertise 的 XML 元素 section 映射至名为 expertise 的搜索字段：

- a. 单击**添加字段**以将一个空行添加到字段映射规则列表。
 - b. 在**字段名字段**中，输入 expertise。
 - c. 在**字段名字段**中，输入 section。
 - d. 在 **XML 属性名字段**中，输入 id。
 - e. 在 **XML 属性值字段**中，输入 expertise。
 - f. 要使用户能够查询 expertise 字段和在搜索结果中查看专业值，请选中**字段搜索**和**搜索结果**复选框。
8. 单击**确定**。

示例:

要查找组织中的每个开发搜索产品的人员，请指定以下查询:

```
jobrole:search
```

要查找组织中的每个在语言学方面有专长的人，请指定以下查询:

```
expertise:linguistics
```

使用 HTML 搜索字段

如果要使用户能够搜索 HTML 文档的特定元数据部分，请将 HTML 元数据元素映射至索引中的搜索字段。

可以使用企业搜索管理控制台来将 HTML 元数据元素映射至搜索字段。

HTML 搜索字段

HTML 搜索字段使用户能够查询 HTML 文档的属性。

HTML 文档中的元数据元素与文档属性在以下方面类似，它们都提供关于文档的信息，格式以及在 Web 上访问该文档的方式。例如:

```
<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" />
```

企业搜索可以使 HTML 元数据元素名与搜索字段名相关联（即将 HTML 元数据元素名映射至搜索字段名）。在为集合配置解析选项时，指定要将哪些 HTML 元数据元素映射至哪些搜索字段名。通过将 HTML 元数据元素映射至搜索字段，可以使用户能够通过查询中指定搜索字段名来找到带有那些元素的文档。与搜索所有文档内容的自由文本查询相比，搜索特定字段的查询可以提供更精确的搜索结果。

例如，如果集合包括 HTML 文档，并且指定要将 copyright 和 abstract 元数据元素作为搜索字段进行索引，则用户可以查询这些特定的元素。搜索 copyright:IBM 即可找到在 copyright 元数据中包含 IBM 的 HTML 文档。

将 HTML 元数据元素映射至搜索字段

当将 HTML 元数据元素映射至搜索字段时，指定用户可以通过在查询中指定字段名来搜索哪些 HTML 元数据元素。

在开始之前

要将 HTML 元数据元素映射至搜索字段，您必须是企业搜索管理员角色的成员或者是 HTML 文档所属的集合的集合管理员。

关于本任务

当创建 HTML 字段映射或者添加、更改或删除现有 HTML 字段映射中的字段时，所作的更改将在停止并重新启动解析器之后生效。那些更改将仅应用于重新启动解析器后解析的数据。要更新已解析并建立了索引的文档，必须再次搜寻和解析该文档，然后重组索引。

过程

要将 HTML 元数据元素映射至搜索字段：

1. 编辑集合，选择解析页，然后单击**将 HTML 元数据映射至字段**。
2. 在 HTML 字段映射页上，单击**添加字段**以在字段映射规则列表中添加一空行。
3. 在**字段名字段**中，输入要与所映射的 HTML 元数据元素相关联的名称。当用户查询此集合中的 HTML 文档时，他们可以指定此字段名。
4. 在**HTML 元数据元素名字段**中，输入要映射的 HTML 元数据元素名。
5. 要使用户能够查询此字段并在搜索结果中查看此字段，请选中**字段搜索**和**搜索结果复选框**。
6. 如果此字段的数据类型是 DECIMAL、DOUBLE、INTEGER、SHORT、TIME 或 TIMESTAMP，并且想使用户搜索此字段时能够指定参数查询，请选中**参数搜索复选框**。
7. 单击**确定**。

示例：

用户现在可以对映射的字段名进行查询以查找具有特定元数据的 HTML 文档了。例如，如果将名为 description 的 HTML 元数据元素映射至名为 abstract 的搜索字段，用户就可以输入类似于以下的查询来查找有关 Thinkpad 计算机的文档：

```
abstract:thinkpad
```

定制文本处理

通过将定制文本处理算法与企业搜索集合相结合，可以提高搜索结果的质量和精度。

WebSphere Information Integrator OmniFind Edition 支持 IBM Unstructured Information Management Architecture (UIMA)，这是一个用于创建、发现、组成和部署文本分析功能的框架。应用程序开发者对将要搜索的内容创建并测试分析算法，然后创建处理引擎归档文件（.pear 文件），它包含了将该归档文件用于企业搜索所必需的所有资源。为了能够使用定制分析算法来搜索集合，必须将该归档文件（它包含文本分析引擎）添加到企业搜索系统中。

文本分析引擎中的分析逻辑组件称为注释器。每个注释器都执行特定的语言分析任务。一个文本处理引擎可以包含任意数目的注释器，它也可以由数个文本分析引擎组成，每个文本分析引擎都包含各自的定制注释器。

由注释器产生的信息被称为分析结果。分析结果对应于想要搜索的信息，被写到一种称为公共分析结构的数据结构中。

在为集合配置文本处理选项时，需执行以下任务：

- 选择要用于对集合中的文档创建注释的文本分析引擎。
- 如果集合包含带有有意义标记的 XML 文档，并且想在定制文本分析中使用此标记，可以使 XML 映射文件与该集合相关联，并将 XML 映射的输出映射至公共分析结构。

例如，可将 <addressee> 和 <customer> 元素的内容映射至公共分析结构中的“人员”注释。然后这些注释可由您的定制注释器访问，这些注释器可能会检测到其他信息（例如，它们可能会检测到人员的性别）。也可以将“人员”注释映射至企业搜索索引，以允许用户搜索“人员”而无需了解原始的 XML 元素。

如果想要允许用户在查询中指定原始的 XML 元素，则不需要定义任何 XML 映射。而应对集合配置解析选项并启用本机 XML 搜索。

- 将公共分析结构中的数据映射至企业搜索索引，从而能够通过语义搜索来搜索已创建了注释的文档。

例如，根据由注释器检测到的实体和关系，用户可以搜索出现在同一句子中的概念（例如，特定的某个人以及任何竞争者的姓名）或某个关键字和概念（例如，搜索项 Alex 和某个电话号码）。

- 将公共分析结构中的数据映射至支持 Java 数据库连接（JDBC）的数据库表。可以将数据映射至 IBM DB2 通用数据库（DB2 UDB）或 Oracle 表。此类映射使您能够在诸如数据挖掘应用程序之类的数据库应用程序中使用分析结果。它还使您能够使用 SQL 查询来在企业搜索外部搜索数据。

相关概念

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Text analysis algorithms”

“Text Analysis Integration”中的“Semantic search applications”

“Text Analysis Integration”中的“Semantic search query”

将文本分析引擎添加至系统

如果创建了定制文本分析引擎，就必须将其添加到系统中，这样才能将其用于企业搜索。集合可以使用该引擎来分析文档并为文档创建注释，从而提高搜索结果的精度。

在开始之前

要将文本分析引擎添加到系统中，您必须是企业搜索管理员角色的成员。

关于本任务


应用程序开发者可以创建符合 UIMA 文本分析框架的处理引擎归档 (.pear)。该归档包含搜索企业搜索集合时所需的所有资源。为了能够使用定制分析算法来搜索集合，必须将该归档文件（它包含文本分析引擎）添加到企业搜索系统中。

在将文本分析引擎添加到系统后，可以更改其显示名并选择选项以查看 XML 源代码。（XML 源代码将显示此引擎生成的信息。）

如果文本分析引擎已与某个集合相关联，就不能从系统中除去该文本分析引擎。

过程

要将定制文本分析引擎添加到企业搜索系统中：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在解析页上，单击**配置文本分析引擎**。
4. 在文本分析引擎页上，单击**添加文本分析引擎**。
5. 在添加文本分析引擎页上，输入新引擎的描述性名称。系统将使用这个显示名来在管理控制台中标识该文本分析引擎。
6. 指定 .pear 文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。
7. 单击**确定**。该文本分析引擎将列示在文本分析引擎页上。

相关概念

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Approaches for using XML markup in analysis and search”

相关任务

“Text Analysis Integration”中的“Creating an XML to UIMA types mapping configuration file”

使文本分析引擎与集合相关联

如果有定制文本分析引擎与企业搜索系统相关联，就可以选择要与集合配合使用的定制文本分析引擎。于是，用户在搜索集合时就可以指定语义查询，从而提高搜索结果的质量和精确度。

在开始之前

要使文本分析引擎与集合相关联，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

如果已有文本分析引擎与此集合相关联，在关联另一个引擎时就会执行下列操作：

- 如果选择了**不进行定制分析**，则先前为该集合定义的所有文本分析映射都将被复位。该集合将开始使用系统缺省值。

- 如果选择了另一个定制文本分析引擎的名称，则先前为该集合定义的所有文本分析映射都将保留下来。例如，如果将关联的引擎 engine_1 更改为 engine_2，则 engine_2 将继承为 engine_1 配置的 XML 映射文件。

过程

要使文本分析引擎与集合相关联：

1. 编辑集合，选择解析页，然后单击**配置文本处理选项**。
2. 单击**选择文本分析引擎**。如果尚未将任何定制文本分析引擎添加到企业搜索系统中，或者如果集合使用缺省分析算法，则分析引擎名是 **Default**。
3. 在为此集合选择文本分析引擎页上，选择要与此集合配合使用的引擎的名称。如果没有文本分析引擎可用，或者如果选择了**不进行定制分析**，则解析器在为文档创建注释和为索引准备文档时将应用缺省文本分析规则。
4. 单击**确定**。

相关概念

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

将 XML 元素映射至公共分析结构

如果集合包含带有有意义的标记的 XML 文档，并且想使用此标记来使用户能够通过语义搜索功能对企业搜索索引或 JDBC 表进行搜索，则可以将 XML 元素映射至公共分析结构。

在开始之前

要将 XML 元素映射至公共分析结构，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

要使定制文本分析进程能够访问 XML 文档中的特定元素，或者要将几个 XML 元素映射至公共类型以便在语义搜索功能中使用，可以创建定制 XML 映射文件。XML 映射文件必须符合 UIMA 文本分析框架。

将 XML 映射文件添加到使用定制文本分析引擎的集合后，源文档中的 XML 元素就能够映射至公共分析结构中的注释。于是，定制文本分析引擎就可以使用这些注释。然后，可以将（公共分析结构中的）分析结果映射至索引，这样用户在通过语义搜索功能搜索集合时就可以查询注释了。

例如，可将 addressee 和 customer 元素的内容映射至公共分析结构中的“人员”注释。然后这些注释可由您的定制注释器访问，这些注释器可能会检测到其他信息（例如，它们可能会检测到人员的性别）。也可以将“人员”注释映射至企业搜索索引，以允许用户搜索“人员”而无需了解原始的 XML 元素。

如果想要允许用户在查询中指定原始的 XML 元素，则不需要定义任何 XML 映射。而应对集合配置解析选项并启用本机 XML 搜索。

过程

要将 XML 元素映射至公共分析结构:

1. 编辑集合, 选择解析页, 然后单击**配置文本处理选项**。
2. 在将 **XML 元素映射至公共分析结构**区域中, 单击**添加映射**。
3. 在将 XML 元素映射至公共分析结构页上, 为 XML 映射文件输入描述性的显示名。
4. 指定文件的位置。如果 XML 映射文件小于 8 MB, 则可以输入路径或进行浏览以找到该文件。如果 XML 映射文件大于 8 MB, 请确保该文件在索引服务器上, 然后指定该文件的标准路径。
5. 单击**确定**。新的 XML 映射文件将被添加到文本处理选项页上的 XML 映射文件列表中。

相关概念

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Approaches for using XML markup in analysis and search”

相关任务

“Text Analysis Integration”中的“Creating an XML to UIMA types mapping configuration file”

将公共分析结构映射至索引

可以指定当用户通过语义搜索来查询集合时要使用的公共分析结构。

在开始之前

要将公共分析结构映射至索引, 您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

通过将公共分析结构映射至企业搜索索引, 用户就能够使用语义指定精确的查询, 从而提高搜索结果的质量。

例如, 根据由注释器检测到的实体和关系, 用户可以搜索出现在同一句子中的概念 (例如, 特定的某个人以及任何竞争者的姓名) 或某个关键字和概念 (例如, 搜索项 Alex 和某个电话号码)。

过程

要将公共分析结构映射至索引:

1. 编辑集合, 选择解析页, 然后单击**配置文本处理选项**。
2. 在将一个公共分析结构映射至索引区域中, 单击**选择公共分析结构**。
3. 在为此集合选择公共分析结构页上, 选择要用于企业搜索索引的映射:
 - 要使用缺省系统规则来将公共分析结构中的数据结构映射至企业搜索索引, 请选择**缺省**。

- 要将特定公共分析结构映射至索引，请指定公共分析结构文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。

4. 单击**确定**。指定的公共分析结构将显示在文本处理选项页上。

相关概念

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Approaches for indexing custom analysis results”

相关任务

“Text Analysis Integration”中的“Creating the index build configuration file”

将公共分析结构映射至 JDBC 表

可以指定要将哪个公共分析结构映射至 JDBC 表以便在数据库应用程序中使用。

在开始之前

要将公共分析结构映射至 Java 数据库连接 (JDBC) 数据库表，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

通过将公共分析结构映射至 JDBC 表，数据就可以供数据库应用程序使用。例如，用户可以在企业搜索外部指定 SQL 查询来搜索公共分析结构添加的注释。也可以使用信息来进行进一步进行文本处理，例如使用数据挖掘应用程序中的信息。

过程

要将公共分析结构映射至 JDBC 表：

1. 编辑集合，选择解析页，然后单击**配置文本处理选项**。
2. 在**将公共分析结构映射至 JDBC 表**区域中，单击**添加映射**。
3. 在将公共分析结构映射至 JDBC 表页上，输入用来将信息映射至 JDBC 表的公共分析结构的描述性显示名。
4. 指定公共分析结构文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。
5. 单击**确定**。新文件将被添加到文本处理选项页上的文件列表中。还将显示该文件被添加到系统的日期和时间。

相关概念

“Text Analysis Integration”中的“Workflow for custom analysis integration”

“Text Analysis Integration”中的“Custom text analysis integration”

“Text Analysis Integration”中的“Unstructured information management architecture (UIMA)”

“Text Analysis Integration”中的“Approaches for mapping analysis results to a database”

相关任务

配置解析器服务的线程

如果有足够的内存资源，则可以增加解析器解析文档时可以使用的线程数。

在开始之前

如果集合非常多，则可以增加解析器线程数。请确保系统有足够的内存来支持附加的线程。带有一个线程的解析器需要 200 MB 内存。每个附加线程都另外需要 50 MB 内存。

要为解析器配置启动的线程数，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要配置解析器线程数：

1. 编辑集合，选择解析页，然后单击**配置解析选项**。
2. 指定解析器启动时要启动的最大解析器线程数，然后单击**确定**。
3. 重新启动解析器。

启用复合项高级分析功能

可以通过使解析器能够对复合项使用高级分析功能来提高搜索质量。使用高级分析功能，将对复合项进行分解以便将每个部分都看作单个项。

在开始之前

要指定复合项解析选项，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

某些语言会将多词项累积成不带空格的单个词（复合项）。复合项高级分析和分解功能对于搜索诸如德语之类的语言来说很有帮助，对于搜索诸如韩国语之类的语言来说至关重要。

如果启用复合项高级分析功能，用户就可以搜索各个项，而不必使用通配符来查找查询项的复合格式。例如，搜索 `Organ (organ)` 会返回包含 `Organspender (organ donor)` 的文档，但不会返回包含 `Organisation (organization)` 的文档。与通配符查询 `Organ*`（它可以返回 `Organ` 后面的任何字符串）不同，此搜索仅与较大复合项中的完整语言子词匹配。

用户定义的词汇表项（例如同义词和升降字）也会应用于在查询中用作各个词的复合部分。

过程

要对复合项启用高级分析功能：

1. 编辑集合，选择解析，然后单击**配置解析选项**。

2. 选中对复合项启用高级分析复选框，然后单击**确定**。

相关概念

“Text Analysis Integration”中的“Linguistic support for semantic search”

“Text Analysis Integration”中的“Text analysis included in enterprise search”

启用对本机 XML 搜索的支持

如果集合中包括了 XML 文档，通过对集合启用本机 XML 搜索，可以使用户在搜索文档时能使用 XML 标记。

在开始之前

要支持使用本机 XML 搜索来搜索 XML 文档，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

本机 XML 搜索（例如，XPath 或 XML 片段）通过在查询中利用文档的 XML 标记，可以提供更精确的搜索结果。例如，用户可以指定某个查询项必须出现在特定的 XML 元素或属性中。

例如，来自某个计算机零售商的 XML 格式的发票可能有包括 <company> 和 <computertype> 元素的 <order> 条目。要检索包含 IBM notebooks 订单的发票，搜索关键字 IBM 和 notebook 可能会检索出包括 Dell notebook computers 和 IBM desktop models 的文档。通过使用 XML 搜索，可以指定 IBM 必须出现在 <company> 元素中，而 notebook 必须出现在 <computertype> 元素中，并且这两个元素都必须包含在同一 <order> 元素之下。这样，您检索的发票就应该只涉及 IBM notebooks 了。

过程

要使用户能通过本机 XML 搜索来搜索集合：

1. 编辑集合，选择解析页，然后单击**配置解析选项**。
2. 选择使用用户可通过本机 XML 搜索来搜索 XML 文档。
3. 单击**确定**。

相关概念

“Text Analysis Integration”中的“Linguistic support for semantic search”

“Text Analysis Integration”中的“Text analysis included in enterprise search”

“Text Analysis Integration”中的“Semantic search applications”

“Text Analysis Integration”中的“Semantic search query”

中文、日语和韩国语文档的语言分析

要增强使用中文，日语和韩国语编写的文档的可检索性，可以指定语言分析选项。

对于中文，日语和韩国语文档，可以指定解析器将使用 n-gram 分段来进行词法分析。对于中文和日语文档，还可以配置解析器从空格中除去换行符。

N-gram 分段

创建集合时，可以选择在解析不使用空格进行单词定界的语言编写的文档时要使用的词法分析类型。

基于 Unicode 的空格分段使用空格来作为各个单词之间的定界符。N-gram 分段将任意数目的字符的重叠序列视为一个单词。对于诸如中文、日语和韩国语之类的语言（这些语言不将空格用作字定界符），与基于 Unicode 的空格分段相比，n-gram 分段可以返回更好的搜索结果。

创建集合时，选择用于对文档进行解析的分段方法。在创建集合之后，就可以通过查看解析选项来查看设置，但不能更改该信息。

从非 ASCII 字符范围中除去换行符

在不使用空格来对字边界进行定界的语言（如中文和日语）中，可以将解析器配置为除去某些将会导致换行的空格字符。

在开始之前

要完成此任务，您必须作为企业搜索管理员登录。

关于本任务

如果对集合启用了此选项，并且文档是以中文或日语编写的，则解析器将除去分隔两个字母字符的空格字符序列。字母字符必须在非 ASCII Unicode 字符范围内。将除去以下字符：

- 制表符 (0x09)
- LF 或换行符 (0x0A)
- CR 或回车符 (0x0D)

要使更改生效，请停止并重新启动解析器。要对已解析并存储在索引中的文档应用更改，可再次搜寻和解析该文档，然后重组索引。

过程

要从空格中除去换行符：

1. 在索引服务器上，作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 使用文本编辑器来编辑以下文件，其中 *collection_ID* 是在创建集合时对集合指定的标识（或由系统分配的标识）：
`collection_ID.parserdriver/collection.properties`
3. 将 `removeCjNewlineChars` 属性值由 `false` 更改为 `true`。

与集合解析器和 Stellent 会话相关联的文档类型

为了确保准确并高效地对搜寻空间中的文档进行解析，可以创建配置文件来指定哪些类型的文档将由集合解析器解析以及哪些类型的文档将由 Stellent 文档过滤会话解析。

在企业搜索集合中，大多数文档格式都由内置的 HTML 或 XML 解析器处理。通常不解析某些类型的文档（如 Postscript 文档），并且其他类型的文档由 Stellent 解析功能

进行处理（如 Microsoft Word、Microsoft Excel、Microsoft PowerPoint、Lotus Freelance、Lotus 123、PDF、RT 和 Ichitaro 文档类型）。

由于元数据可能有误，所以纯文本和 HTML 文档可能会被错误地发送给 Stellent 会话，然后被发送回给其中一个内置解析器，这种情况会影响性能。要避免这种情况，可以创建配置文件以控制不同类型的文档在哪解析以及如何解析。

使文档类型与集合解析器和 Stellent 会话相关联的工作涉及下列任务：

1. 为集合解析器配置文档类型。此步骤涉及创建一个配置文件，该配置文件将文档类型映射至集合所使用的解析器。可以为每个集合创建这些配置文件中的其中一个。
2. 为 Stellent 会话配置文档类型。此步骤涉及创建一个配置文件，该配置文件将文档类型映射至集合所使用的 Stellent 文档过滤器。可以为每个集合创建这些配置文件中的其中一个。
3. 停止并重新启动解析器。要使更改生效，请使用企业搜索管理控制台来监视已配置了文档类型的集合，然后停止并重新启动解析器。

将文档类型与集合解析器关联

要使特定类型的文档与集合解析器相关联，应创建配置文件。在企业搜索管理控制台中，未提供对此任务的支持。

在开始之前

要完成此任务，您必须作为企业搜索管理员登录。

关于本任务

如果配置文件不存在，集合解析器就会使用缺省解析器服务规则。如果配置文件存在，该文件中的规则指定：

- 要根据文件扩展名或内容类型将哪些文档发送给 Stellent 会话。
- 如何对由于元数据不完整而具有未知类型的文档进行解析。

文件的格式是行序列，每一行都是下列其中一个规则：

EXTENSION *extension parser*

以指定扩展名作为 URL 结尾的所有文档都将由指定的解析器处理。不要在扩展名中包括句点；比较操作不区分大小写。

CONTENTTYPE *type/subtype parser*

内容类型与指定的类型 / 子类型相匹配的所有文档都将由指定的解析器处理。假定文档的内容类型为 t/s，如果 t 等于类型，并且 s 等于子类型或者子类型是通配符（星号 *），则将发生匹配。

UNKNOWN *parser*

所有具有未知扩展名和内容类型的文档（即，不是由搜寻器提供的文档）都将由指定的解析器处理。

DEFAULT *parser*

所有未被任何其他规则包括的文档都将由指定的解析器处理。

在所有情况下，*parser* 必须指定 html、xml、stellent、或 none，其中 none 表示将不解析那种类型的文档。

如果有多个规则与一个文档相匹配，则规则越具体就越优先，而不考虑规则的出现顺序：

- EXTENSION 规则比 CONTENTTYPE 规则更具体。
- 包含子类型的 CONTENTTYPE 规则比包含通配符的 CONTENTTYPE 规则更具体。例如，内容类型 application/postscript 的规则优先于 application/* 的规则。
- 不应该存在两个用于相同扩展名和内容类型的规则。在那种情况下，由实施进程来负责确定哪个规则具有优先权。

过程

要使文档类型与集合解析器相关联：

1. 在索引服务器上，作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 创建如下配置文件，其中 *collection_ID* 标识您想要配置的集合：
ES_NODE_ROOT/master_config/*collection_ID*.parserdriver/parserTypes.cfg
3. 使用文本编辑器编辑文件并指定解析器服务规则，然后保存并退出此文件。
4. 要使更改生效，请使用企业搜索管理控制台来监视集合的解析器，并停止再重新启动该解析器。

示例

在以下示例中，内置的 HTML 解析器处理所有带有扩展名 txt、htm 或 html、内容类型以 text/ 开头或者带有未知扩展名和内容类型的文档。内置的 XML 解析器处理所有带有扩展名 xml 或者内容类型为 text/xml 的文档。所有其他文档（包括内容类型以 application/ 开头的文档）都将被发送给 Stellent 会话。

```
EXTENSION doc stellent
EXTENSION txt htm
EXTENSION htm htm
EXTENSION html html
EXTENSION xml xml
EXTENSION ps none
CONTENTTYPE text/xml xml
CONTENTTYPE text/* htm
CONTENTTYPE application/* stellent
UNKNOWN htm
DEFAULT stellent
```

缺省集合解析器服务规则

如果未创建配置文件来将文件类型和内容类型映射至集合的解析器，则将使用缺省规则来解析文档。

集合解析器使用的缺省规则如下所示：

```
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 jttc stellent
EXTENSION jtdc stellent
EXTENSION jtdx 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
```

将文档类型与 Stellent 会话关联

要指定 Stellent 文档过滤器要解析的文档类型，您应创建配置文件。在企业搜索管理控制台中，未提供对此任务的支持。

在开始之前

要完成此任务，您必须作为企业搜索管理员登录。

关于本任务

配置文件指定了以下内容：

- Stellent 会话将要对哪些文件类型进行解析。文件类型与 Stellent 库所识别的其中一种文件类型相对应。
- 要将哪些文件类型发送回给集合解析器以便使用其中一个内置解析器进行处理。（由于集合解析器可能会由于元数据不正确而错误地将文档发送给 Stellent 会话，所以需要执行此操作。）
- 哪些文件类型由于不受企业搜索支持而将被拒绝。

如果指定了配置文件，但它不存在，则解析器将无法启动。如果未对 `stellent.properties` 文件中的 `OutsideInSupportedTypes` 属性指定配置文件，将使用 Stellent 会话的缺省解析规则。

配置文件列出了文档类型以及它们的处理方式。文件的格式是行序列，每一行都是下列其中一个规则：

```
accept DEFAULT
accept ALL doctype
accept type doctype
native DEFAULT
native type doctype
reject type
```

其中:

doctype

是用于 *doctype* 查询令牌的值。可按文档类型搜索文档。例如, 用户可以指定 `$doctype::pdf` 以搜索 PDF 文档。

type 是 Stellent 库中的其中一个 FI_ 值, *doctype* 是当规则被应用时用于 *doctype* 令牌的值。

DEFAULT

表示接受类型列表或本机类型列表(这取决于规则的类型)包括缺省列表。此选项使您能够扩展缺省配置, 而不是替换缺省配置。

A11 表示指定的 *doctype* 令牌接受所有未明确列出的类型。

配置文件中的规则处理方式如下:

- 如果存在 *type* 的 *accept* 规则, 包括缺省列表(如果指定了 *accept DEFAULT* 的话), 则接受该类型。
- 否则, 如果存在 *type* 的 *reject* 规则, 则不接受该类型。
- 否则, 如果指定了 *accept ALL*, 则接受该类型。
- 否则, 不接受该类型。

如果接受文档类型, 则使用已应用的规则中指定的 *doctype* 值。将把此值与所解析的内容一起发送回给集合解析器。如果不接受文档类型, 则发生下列行为:

- 如果存在 *type* 的本机规则(如果指定了 *native DEFAULT*, 则还包括缺省解析规则), 就会将文档与此规则指定的 *doctype* 标记值一起发送回给内置解析器。*doctype* 值必须是 `txt`、`htm` 或 `xml`, 它们分别指示纯文本、HTML 或 XML。
- 否则, 拒绝该文档, 并且不对该文档进行解析。

过程

要使文档类型与 Stellent 会话关联:

1. 在索引服务器上, 作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 编辑 `ES_NODE_ROOT/master_config/collection_ID.stellent/stellent.properties` 文件, 其中 *collection_ID* 标识您想要配置的集合。
3. 对于 `OutsideInSupportedTypes` 属性, 指定正在创建的配置文件的绝对路径。

例如, 可以为单个集合创建以下配置文件并将其与其他特定于集合的文件存储在一起:

```
ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```

又例如, 可能会创建以下配置文件, 以便可对所有集合使用相同的设置, 并将此文件与其他系统级文件存储在一起。(如果使用此方法, 则务必在每个集合的 `stellent.properties` 文件中指定此路径, 如步骤 2 所述。)

ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg

4. 使用文本编辑器创建配置文件并指定 **Stellent** 解析规则，然后保存并退出此文件。
5. 要使更改生效，请使用企业搜索管理控制台来监视集合的解析器，并停止再重新启动该解析器。

示例

在以下配置文件中，**Stellent** 会话除接受缺省列表以外还接受 Microsoft Visio 格式。

```
accept DEFAULT
accept FI_VISI03 visio
accept FI_VISI04 visio
accept FI_VISI05 visio
accept FI_VISI06 visio
```

在以下配置文件中，将支持 Postscript 文档并可以对它们执行搜索，这些文档的文档类型为 ps；将把 X 像素图格式 (XPM) 发送回内置文本解析器；将拒绝 PNG 图像格式；接受所有其他文件类型并可以对它们执行搜索，它们的文档类型为 other。

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
accept ALL other
reject FI_PNG
```

Stellent 会话的缺省解析规则

如果未创建配置文件来将文件类型映射至 **Stellent** 会话文档过滤器，则将使用缺省规则来解析文档。

Stellent 会话使用的缺省规则如下所示：

```
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_OLDPCODE	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

企业搜索索引管理

要确保用户总是能够访问最新的信息，企业搜索为每个集合创建索引并通过定期地刷新内容和重组内容来维护该索引。

要使搜寻器收集的数据可供搜索，必须创建索引。第一次创建集合时，企业搜索为最初搜寻的所有数据创建索引。搜寻器搜寻新的和已更改的数据源时，企业搜索将使用新内容来刷新索引。最后，需要将经过刷新内容合并到基本索引中。这个合并过程称为重组。每当刷新索引或重组索引时，都将把新内容复制到搜索服务器并使该内容可供搜索。

搜寻器持续不断地或定期地收集数据。如果频繁地刷新索引，用户就能够搜索最新的数据。最后，必须将持续刷新的索引重组。当经过刷新的索引变大时，它将消耗更多的系统资源。因此，为了维护最佳性能，应该定期重组索引。

重组索引的频率取决于：

- 系统资源（文件系统空间、处理器速度和内存）
- 需要搜寻和重新搜寻的文档数目
- 要搜寻的数据类型
- 更改分类规则的频率（在重组索引之前，更改不会生效）
- 强制搜寻器启动而不是在已安排的时间运行的频率
- 外部搜寻器除去或添加 URI 的频率（这些类型的搜寻器通过数据侦听器 API 来与企业搜索进行交互）

对于有儿百万个主要由 Web 文档构建的集合，应该以大约每天一次的频率重组索引并且每一或两个小时将索引刷新一次。

要维护可搜索的当前索引，请执行下列任务：

- 指定用于刷新索引和重组索引的时间表
- 更改索引时间表
- 启用和禁用索引时间表
- 配置并发索引构建

要指定影响用户的索引视图的选项，还可以执行以下任务：

- 配置对通配符查询的支持
- 配置作用域以限制用户可搜索的文档范围
- 折叠搜索结果中来自相同来源的文档。
- 从索引中除去 URI

相关任务

第 211 页的『监视集合的索引活动』

当您需要查看正在构建的索引的进度、启用或禁用索引时间表或启动和停止建立索引的活动时，请监视集合的索引。

第 212 页的『监视企业搜索索引队列』

您可以查看索引队列中的所有索引构建的状态、停止正在构建的索引或从队列中删除索引。

安排索引构建时间表

可以指定重组索引和使用新内容来刷新索引的时间表。

在开始之前

要安排索引构建时间表，您必须是企业搜索管理员角色的成员或该集合的集合管理员。

关于本任务

要确保用户始终能够访问他们所搜索的来源中的最新信息，请安排定期重组并刷新索引的时间表。重组索引时，将重新构建整个索引。建立索引的过程读取由搜寻器收集并由解析器分析的所有数据。刷新索引时，将使自从上次重组索引以来搜寻的信息可搜索。

缺省情况下，已选择了“安排索引构建时间表”选项。此选项指示调度程序进程每当企业搜索系统启动时都调度任务以刷新并重组索引。如果需要避免运行已安排的索引构建，随时可以清除**当系统启动时启用**复选框。例如，可能需要禁用时间表以对问题进行故障诊断。

过程

要安排索引构建时间表：

1. 编辑集合，选择索引页，然后单击**安排索引构建时间表**。
2. 要指定使用新内容刷新索引的频率，请在安排索引构建时间表页上的**指定时间表以刷新索引**区域中指定下列选项：
 - a. 在**开始时间**区域中，在**年、月、日、小时和分钟**字段中，指定第一次刷新索引的时间。
 - b. 在**更新时间间隔**区域中，在**天、小时和分钟**字段中指定索引刷新频率。

通常，应该频繁地刷新索引，如每 1 个小时或每 2 个小时刷新一次。根据源内容更改的频率，可指定较短或较长的时间间隔。例如，可以指定每小时（0 天又 1 小时）或每 12 小时（0 天又 12 小时）。
3. 要指定完全重新构建索引的频率，请在**指定时间表以重组索引**区域中指定下列选项：
 - a. 在**开始时间**区域中，在**年、月、日、小时和分钟**字段中，指定第一次重组索引的时间。
 - b. 在**更新时间间隔**区域中，在**天、小时和分钟**字段中指定索引重组频率。

通常，应该定期地重组索引，如每 24 个小时重组一次。根据源内容更改的频率，可指定较短或较长的时间间隔。例如，可以指定每 12 小时（0 天又 12 小时）或每两天半（2 天又 12 小时）。
4. 单击**确定**。

更改索引时间表

可以更改用于重组或刷新索引的时间表。

在开始之前

要更改索引时间表，您必须是企业搜索管理员角色的成员或该集合的集合管理员。

过程

要更改索引时间表：

1. 编辑集合，选择索引页，然后在**月**、**日**、**年**和**小时**字段中输入适当的值。指定使用新内容刷新索引的频率或者索引重组频率。
2. 单击**应用**。

启用和禁用索引时间表

可以启用和禁用用于刷新索引和重组索引的时间表。

在开始之前

要启用或禁用索引时间表，您必须是企业搜索管理员角色的成员或该集合的集合管理员。

关于本任务

如果需要防止已调度的索引构建运行，则可以禁用索引的时间表。例如，您可能想禁用时间表以防止在安排的日期和时间构建索引以便可以对问题进行故障诊断。

可以在编辑集合时启用和禁用时间表，也可以在监视集合时启用或禁用时间表。

过程

1. 要通过编辑集合来启用或禁用索引的时间表，执行下列步骤：
 - a. 编辑要更改的集合。
 - b. 在索引页上，选中或清除**当系统启动时启用**复选框以启用或禁用用于刷新索引的时间表。
 - c. 要启用或禁用用于重组索引的时间表，请选中或清除**当系统启动时启用**复选框。
 - d. 单击**应用**。
2. 要通过监视集合来启用或禁用索引的时间表，执行下列步骤：
 - a. 监视要更改的集合。
 - b. 在索引页上，如果已经为索引安排了时间表，并且不想在安排的日期和时间构建该索引，则单击 **禁用时间表**。在启用该时间表或单击 **启动**以启动索引构建过程之前，不会构建该索引。
 - c. 如果已经为索引安排了时间表，但是禁用了用于构建该索引的时间表，则单击 **启用时间表**。

该索引将进行排队，以便在索引时间表中指定的日期和时间进行构建。

配置并发索引构建

通过指定可以同时为多少个集合处理索引构建请求来控制建立索引资源的使用。如果有足够的系统资源，就可以通过允许在重组集合索引的同时刷新该索引来提高搜索质量。

在开始之前

要为系统指定索引构建选项，您必须是企业搜索管理员角色的成员。

关于本任务


通过在集合之间共享资源，企业搜索可以同时构建多个索引，这样就可以并行地处理多个集合的索引构建请求。通过共享进程，可以确保对非常大的索引执行的重组不会妨碍队列中其他正在等待构建的索引的可用性。

当请求索引构建或安排它的时间表时，它将进入索引队列并等待轮到它被处理。由于每个集合都有自己的索引，所以在索引队列中可以同时有各种集合的数个索引构建请求。在为系统配置建立索引选项时，指定有多少个集合可以共享建立索引资源和并行地处理它们的请求。

还可以指定在重组集合索引的同时处理集合索引刷新请求。如果启用此选项，则将在处理运行速度较慢的索引重组时使用（通过刷新索引获得的）最新文档刷新搜索服务器。但是，索引构建是非常耗费资源的进程。构建索引时，将消耗大量的系统内存和磁盘空间。如果启用此选项，并且磁盘或内存空间不足，系统总体性能就会下降。

过程

要指定系统的索引构建选项：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在索引页上，单击**配置建立索引选项**。
4. 在系统级的建立索引选项页上，输入可以共享系统资源并且并行处理其索引构建请求的集合数。

共享建立索引资源的集合数不能超出企业搜索系统中的集合数。例如，如果有 5 个集合，则必须输入小于或等于 5 的数值。

5. 如果有足够的系统资源来支持各个集合的多个并发索引构建，则可以选择允许并发运行索引刷新请求和重组请求的选项。
6. 单击**确定**。

影响索引的可搜索视图的选项

在对文档建立索引后，可以指定选项，以控制用户可如何搜索文档并查看搜索结果中的文档。

要指定选项以影响用户查看索引的方式，可执行以下任务：

- 配置对在查询项中使用通配符的支持。可以将通配符查询支持构建到索引中，也可以指定在查询处理期间扩展查询项的选项。

- 配置作用域以限制用户可搜索的文档范围。当用户搜索集合时，他们将仅搜索属于该作用域的文档，而不是搜索整个索引。
- 折叠搜索结果中来自相同来源的文档。可以将索引中符合同一 URI 或 URI 模式的文档分组到一起，并在搜索结果中只显示排在第一位的结果文档（用户可以指定选项以查看折叠起来的结果文档）。
- 从索引中除去 URI。有时可能需要临时阻止用户搜索索引中的特定文档。

用于搜索文档的建立索引选项

在配置用于搜索已搜寻的数据或外部源的选项时，或者在将 XML 和 HTML 元数据元素映射至搜索字段时，指定如何搜索文档以及如何在搜索结果中显示那些文档。

在编辑包含字段的搜寻器的选项时，可以指定下列选项以控制是否能够搜索某个字段、如何搜索该字段以及是否可以在搜索结果中返回该字段：

- 自由文本搜索
- 字段搜索
- 参数搜索
- 搜索结果
- 文档内容

在编辑外部源的选项时，将字段标记为文档内容字段的选项不可用。

在配置解析器并指定要将 XML 元素和 HTML 元数据元素映射至索引中的字段时，请指定是否可以在字段搜索、参数搜索或搜索结果中使用那些字段。

指定的选项将与文档一起存储在索引中。它们使您能够限制用户可以查询的内容以及用户在搜索结果中能够看到的内容。

自由文本搜索

企业搜索索引是一种全文索引，其内容来自各种数据源。可以通过以自然语言指定简单查询来搜索内容。搜索进程将搜索字段和文档内容以查找与查询相关的文档。

示例：

自由文本搜索可以象以下查询一样简单：

bicycle chain

为了指示哪些词必须出现在文档中或必须不出现在文档中，可以包括特殊表示法。例如，可以在词的前面放置加号（+）以指定文档必须包含该词才能找到匹配项。在词的前面放置减号（-）以将包含该词的文档排除在搜索结果之外。将两个或多个单词括在引号（"）中以搜索精确的短语。

示例：

在以下自由文本查询中，仅当文档包含精确的短语 science fiction 并且不包含 robot 一词时才会发生匹配：

+ "science fiction" -robot

字段搜索

字段搜索使您能够将查询对象约束为文档的特定字段和元数据。例如，可以指定某些词必须存在于文档的标题中。

要在企业搜索中指定字段搜索，则在查询中包括字段名以及必须存在于该字段中的词或短语。

示例:

以下查询搜索 title 字段中必须包含 ibm 一词和短语 enterprise search 的文档:

```
title:ibm title:"enterprise search"
```

要按字段名搜索字段，在配置搜寻器时必须使该字段可用于字段搜索。

参数搜索

参数搜索是一种字段搜索类型，它使您能够对数字和日期字段以及元数据进行比较查询或求值查询。例如，可以搜索具有特定大小的文档或在特定日期之后撰写的文档。还可以搜索属性值大于、小于或等于指定值的文档。

示例 1:

以下查询搜索价格正好为 50 美元（或为 price 字段建立索引的任何货币单位）的商品:

```
#price::=50
```

示例 2:

以下查询搜索文件大小大于 1024 但小于或等于 2048 的文档:

```
#filesize::>1024<=2048
```

要使用参数查询来搜索字段，在配置搜寻器时必须使该字段可用于参数搜索。

搜索结果

可以搜索某些字段但不将它们显示在搜索结果中，即使不查询某个字段也可以将该字段显示在搜索结果中。例如，可能需要查询财务数据以获取有意义的报表，但现在想在显示了职员姓名的结果中显示职员薪酬。

文档内容

通过使用已被标记为文档内容字段的字段的内容，可以使文档与基于模型的类别相关联，并使它们与指定了基于文档内容的规则的类别相关联。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Query syntax”

第 147 页的『企业搜索的搜索应用程序』

搜索应用程序使您能够对企业搜索系统中的集合和外部源进行搜索。可以创建任意数目的搜索应用程序，并且单个搜索应用程序可以搜索任意数目的集合和外部源。

查询中的通配符

可以允许用户在查询项中包括通配符并搜索与指定模式匹配的项。

通配符查询项是包含星号 (*) 的查询项。当用户提交包括通配符的查询时，搜索结果将包括索引中所有与查询项匹配的文档，再加上索引中所有与通配符表示的模式匹配的文档。例如，查询项 `sea*` 中的结尾通配符可以与 `search`、`season` 和 `seals` 匹配。

在为索引配置通配符选项时，请选择是否允许用户在查询中指定通配符以及在允许的情况下如何提供此支持：

- 可以允许在文档的所有部分中搜索与通配符模式匹配的项，也可以将模式匹配限于字段。
- 可以使所有字段都支持包含通配符的查询，也可以将模式匹配限于指定的字段。
- 可以将通配符限制为只能作为查询项中的最后一个字符出现（结尾通配符），也可以允许通配符在查询项中的任何位置出现。（通配符不能出现在字段名中。）
- 根据允许的通配符出现的位置，可以选择查询项扩展方式（在索引中将包含通配符的查询项扩展为所有匹配的项）。可以让索引存储所有可能的项扩展，也可以让搜索进程在查询处理期间扩展项。

对通配符设置所作的任何更改都将在下次重组索引后生效。

索引扩展

要在索引中包括项的扩展，应指定在一个词中必须要有多少前导字符与查询项中的通配符模式匹配才能使匹配发生。只有至少包含此字符数（不包括 *）的查询项才会返回结果。例如，如果指定 4，则查询项必须至少指定 4 个字符才能使匹配发生。

如果指定 4，则单词 `technology` 与查询项 `tech*` 和查询项 `techno*` 匹配，但与查询项 `te*` 不匹配。

刷新索引或重组索引后，将为文档中的原始项以及每个项的所有可能扩展建立索引。此方法的一个优点就是，在查询处理期间无需再花额外的时间用于扩展查询项。但此方法会增加索引的大小，这意味着必须要有足够的系统资源来容纳较大的索引。

如果集合相对较小，或者用于构建索引的空间和时间不如查询响应时间重要，则此方法是最有用的。例如，可能会选择此方法来搜索目录或职员目录。

仅当启用了结尾通配符支持时，此方法才可用。如果允许通配符在查询项中的任何位置出现，就不能选择在索引中包括项扩展的选项。

查询扩展

要在用户提交包含通配符的查询时展开查询并应用模式匹配规则，应指定一个查询项有多少种变化来构成一个匹配项。例如，如果指定 50，则一个查询项最多有 50 种变化能够成为该查询项的匹配项。

以下举例说明了此示例：查询项 `tech*` 与单词 `technical`、`technique`、`technology` 以及最多 50 个以字符 `tech` 开始的不同的单词匹配。

虽然查询扩展对索引大小只有微弱的影响，但它会降低查询性能。搜索进程必须反复查找通配符查询项的所有可能扩展项，直到达到通配符设置中指定的限制为止。

如果集合的大小相对较大且必须尽量减少构建索引的空间和时间，则此方法是最有用的。例如，可以对电子邮件存储库使用此方法（在电子邮件存储库中，索引必须与快速更改的文档保持一致，但查询响应时间却不那么重要）。

无论是启用了结尾通配符支持还是允许通配符在查询项中的任何位置出现，此方法都可用。

在查询中支持通配符

通配查询项的扩展项集包含索引中所有可以通过将通配符替换为任意字符序列获取的项。这个扩展项集是按如下方式确定的：

- 如果集合支持通配符在查询项中的任何位置出现，则将任何包含通配符的查询项解释成通配项。
- 扩展项集最多包含由企业搜索管理员配置的最大扩展项数。如果索引包含的扩展项超出此数目，则忽略额外的扩展项。（搜索结果会指示是否忽略了任何通配扩展项。）
- 如果已将通配符支持限于一组字段，则扩展项集只包含在其中一个指定字段中出现的项。项只需要在索引中至少一个文档的其中一个字段中出现。
- 如果该查询项是字段项，则通配符必须出现在字段说明符后面（例如，`fieldname:*sphere`）。字段名不能包含冒号（:）。
- 如果已将通配符支持限于一组字段，则通配查询项中的字段名必须是企业搜索管理控制台中指定的其中一个字段。否则，就无法找到该项的扩展项。
- 仅对纯文本项支持通配符，而对 XML 元素名、属性名或属性值不支持通配符。不支持仅包含通配符的项。

配置通配符查询选项

在为企业搜索集合配置建立索引选项时，可以指定是否要使用户能够在查询项中指定通配符。

在开始之前

要配置通配符选项，您必须是企业搜索管理员角色的成员或者是索引所属的集合的集合管理员。

关于本任务

指定通配符选项时，所作的更改将在下次重组索引后生效。

过程

要配置对通配符查询的支持：

1. 编辑集合，选择索引页，然后单击**配置通配符选项**。
2. 在通配符选项页上，选中**支持通配符查询**复选框。
3. 使用在**不搜索字段的查询中支持通配符**复选框来指定在搜索非字段文本的查询中是否支持通配符。例如，仅当选中了此复选框时，自由文本查询 `tech*`（它不搜索指定的字段）才会返回扩展的结果（例如 `technology` 或 `technique`）。
4. 指定哪些字段支持通配符：
 - 要使文档中的所有字段都能够支持包含通配符的查询，请选择**所有字段**。

- 要将通配符支持限于某些字段，请选择**特定字段**，然后输入字段名。将仅对指定的字段返回扩展的结果。例如，仅当指定 `author` 字段支持通配符时，查询 `author:john*` 才会返回扩展的结果。
5. 指定通配符是必须在查询项的最后一个位置中出现（结尾通配符）还是通配符不受限制并可以在查询项中的任何位置出现。

在选择通配符位置和类型时，还必须指定如何启用通配符支持。在配置通配符选项时，单击**帮助**可获得详细信息。

6. 单击**确定**。

作用域

当想要向用户显示集合的受限视图时，请配置作用域。

作用域是索引中的一组相关 **URI**。配置作用域时，限制了用户在集合中可以看到文档。当用户搜索集合时，他们将仅搜索该作用域内的文档，而不是搜索整个索引。要使用此功能，搜索应用程序必须支持搜索作用域。

创建作用域时，应指定索引中用户能搜索的 **URI** 范围。限制用户可以搜索的文档有助于确保搜索结果中的文档特定于用户所查找的信息。

例如，可以创建一个包括“技术支持”部门的 **URI** 的作用域和另一个包括“人力资源”部门的 **URI** 的作用域。如果搜索应用程序支持作用域，则“技术支持”部门中的用户将从“技术支持”作用域中检索文档，而“人力资源”部门中的用户将从“人力资源”作用域中检索文档。

虽然创建太多的作用域会影响性能，但是，可以创建任意数目的作用域。请将作用域配置成大多数搜索请求仅需要对一两个作用域进行过滤。因为作用域可以包含整个 **URI** 或 **URI** 模式，所以同一文档可以属于多个作用域。

配置作用域时，可能需要将索引重组两次才能使更改生效。如果在第一次对集合执行索引重组前配置作用域，用户就能够搜索该集合，但他们在搜索结果中将看不到作用域数据。这时，需要再次重组索引以确保搜索结果反映了作用域中的 **URI** 范围。

如果在至少执行一次重组后配置作用域，更改就会在下次索引重组后生效。

配置作用域

为企业搜索集合配置作用域时，指定索引中允许用户搜索的文档范围的 **URI** 或 **URI** 模式。

在开始之前

要配置作用域，您必须是企业搜索管理员角色的成员或者是作用域所属的集合的集合管理员。

关于本任务

如果搜索应用程序启用了对作用域的支持，则用户在搜索集合时可以仅搜索与定义作用域边界的 **URI** 相匹配的文档。

配置作用域时，可能需要将索引重组两次才能使更改生效。如果在第一次对集合执行索引重组前配置作用域，用户就能够搜索该集合，但他们在搜索结果中将看不到作用域数据。这时，需要再次重组索引以确保搜索结果反映了作用域中的 URI 范围。

如果在至少执行一次重组后配置作用域，更改就会在下次索引重组后生效。

过程

要配置作用域：

1. 编辑集合，选择索引页，然后单击**配置作用域**。
2. 在作用域页上，单击**创建作用域**。
3. 指定作用域名以及用于定义作用域边界的 URI 和 URI 模式。也可以指定要排除在作用域之外的 URI 和 URI 模式。
4. 单击**确定**。

新作用域将与属于此集合的其他作用域一起列在作用域页中。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识 (URI) 都指示了将文档添加到该集合的搜寻器的类型。

折叠的 URI

企业搜索可以对搜索结果进行组织，以使来自 URI 前缀相同的 Web 和 NNTP 源的文档在搜索结果中分组到一起并折叠起来。

当结果处于折叠状态时，排在第一位的结果通常显示在最左边。其后的一个或多个排名较低的结果将在该结果下方分组并缩进。

要将带有不同 URI 前缀的结果文档折叠到一组中，可将 URI 前缀与创建的组名关联。例如，有 3 个服务器用于管理财务数据，可以在搜索结果中将来自所有这 3 个服务器的文档分组到一起，并将排名较低的结果折叠在排在第一位的结果文档之下。

搜索应用程序可以使用 URI 前缀或组名来在搜索结果中将文档折叠起来。在企业搜索的样本搜索应用程序中，将显示前两个搜寻结果文档。如果返回了两个以上带有相同的 URI 前缀（或者属于同一 URI 组）的结果文档，可以选择选项来查看折叠的结果。

用户可以使用企业搜索查询语法 (`samegroupas:URI prefix`) 来搜索指定的 URI 前缀所在的组中的所有文档。

如何组织 URI 前缀和组名

当使用管理控制台来配置折叠搜索结果的规则时，应指定想要折叠的文档的 URI 前缀，并有选择性地将其 URI 前缀与组名关联。

配置的 URI 前缀的顺序非常重要。索引服务器在计算集合中每个 URI 的值时会使用此 URI 前缀顺序。对于每个 URI：

1. 索引服务器扫描规则中的所有 URI 前缀，以按顺序折叠搜索结果。
2. 当索引服务器找到第一个与索引中的文档前缀相匹配的 URI 前缀时，它会关联该组名（或该 URI 前缀，如果规则未指定组名的话）作为文档的附加搜索项。

3. 如果文档无法与 URI 前缀匹配, 则:

- 对于 Web URI, 索引服务器使用该 URL 的主机名作为 URI 前缀。
- 对于 NNTP URI, 索引服务器使用引用头的值中的第一个消息标识作为 URI 前缀。

在将 URI 前缀添加到要在搜索结果中折叠的那些 URI 前缀的列表之后, 必须以索引服务器扫描该 URI 前缀的顺序放置它, 并在可能时将它作为附加搜索项与索引中的文档进行关联:

- 如果添加了 URI 前缀但未使其与组名关联, 则可以选择各个 URI 前缀并将它在列表中上移或下移。
- 如果添加了 URI 前缀并使其与组名关联, 则只要将某个 URI 前缀在列表中上移或下移, 就必须移动属于同一组的整组 URI 前缀。(组中 URI 前缀的顺序无关紧要; 选择一个 URI 前缀将自动选择整组。)

折叠搜索结果中的 URI

可以指定选项以便对来自 URI 前缀相同的 Web 和 NNTP 源的结果文档进行分组并将它们折叠起来。也可以创建一个组名以便能够将带有不同 URI 前缀的结果文档折叠到一起。

在开始之前

要指定用于折叠搜索结果的选项, 您必须是企业搜索管理员角色的成员或者是集合的集合管理员。



关于本任务

您所作的有关折叠搜索结果的更改在下次进行索引重组前不会生效。

过程

要指定用于折叠搜索结果的选项:

1. 编辑集合, 选择索引页, 然后单击**折叠搜索结果**。
2. 在折叠搜索结果页上, 单击**添加 URI 前缀**。
3. 在添加用于折叠结果的 URI 前缀页上, 输入搜索结果中要折叠的文档的 URI 前缀。
例如:

```
http://finance/ROI/  
http://server1.com/finance/
```
4. 可以输入要与此 URI 前缀相关联的描述性组名。要将来自数个源的结果文档折叠到一个组中, 请在添加每个 URI 前缀时输入同一个组名。
5. 单击**确定**。
6. 在折叠搜索结果页上, 按照索引服务器扫描新规则的顺序来安排这些规则的位置:
 - 如果添加了 URI 前缀并且未使其与组名相关联, 则新的 URI 前缀将显示在列表底部。使用方向键来将其移至正确的位置。
 - 如果已使新的 URI 前缀与某个组名相关联, 则新的 URI 前缀将显示在属于同一个组的 URI 前缀集合的底部。使用方向键来将整组 URI 前缀移至正确的位置。
7. 要更改 URI 前缀或组名, 请选择 URI 前缀并单击  **编辑**。
8. 要从列表中除去 URI 前缀, 请选择 URI 前缀并单击  **除去**。

从索引中除去 URI

要防止用户搜索集中的文档，可以从索引中除去那些文档的 URI。

在开始之前

要从索引中除去 URI，您必须是企业搜索管理员角色的成员或该集合的集合管理员。

关于本任务

如果指定标准 URI，则用户在搜索结果中就看不到该 URI。但是，如果用户提交同一个查询，并且该查询的结果文档位于搜索高速缓存中，则会继续在搜索结果中返回所除去的 URI 的高速缓存结果页面。在下次刷新索引或重组索引之前，不会刷新搜索高速缓存，并且不会从索引中除去该 URI。

如果指定一个模式以除去多个 URI，则用户在搜索结果中将继续能够看到与该模式相匹配的 URI，直到下次刷新索引或重组索引为止。

当从索引中除去 URI 时，并不会从搜寻空间中除去它。当搜寻器下次搜寻文档时，该 URI 将被构建到索引中并且再次可搜索。要从搜寻空间中除去某个 URI，必须更新搜寻规则以排除该文档，然后停止并重新启动搜寻器。

过程

要从索引中除去特定文档的 URI:

1. 编辑集合，选择索引页，然后单击**从索引中除去 URI**。
2. 在从索引中除去 URI 页上，输入要从索引中除去的 URI（或 URI 模式）。

例如:

```
http://domain.org/hr/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

3. 单击**确定**。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识（URI）都指示了将文档添加到该集合的搜寻器的类型。

企业搜索的搜索服务器管理

可以对搜索服务器指定的选项包括使用高速缓存空间来返回搜索结果、控制搜索结果中文档摘要的最大长度、关联定制同义词字典以提高搜索质量以及当查询包含某些查询项时在搜索结果中返回预定义的 URI。

当用户提交查询时，搜索服务器使用索引来快速地找到相关文档。搜索服务器使用企业搜索数据商店（它包含已解析的和已标记化的数据）来检索相关文档的元数据。元数据可以包括（但不限于）文档 URI、标题、描述、日期以及数据类型等等。

在为集合配置搜索服务器时，可以指定对查询处理方式有影响的选项，包括可能影响查询性能的选项：

配置搜索高速缓存

要优化查询性能，可以指定要存储在高速缓存中的结果文档，也可以配置要为高速缓存的文档分配的空间量。

配置文档摘要的最大显示长度

大多数结果文档都显示了文档内容摘要，以帮助用户决定某个文档是否是用户想要检索的那个文档。可以指定将有多少空间可用于搜索结果中以显示此摘要信息。

指定另一缺省语言

在创建集合时指定了用于搜索集合中的文档的缺省语言，但也可以根据需要指定另一种语言。

关联定制字典

如果应用程序开发者创建了定制同义词字典，无用词字典或升降字字典，则可以指定当用户搜索集合时要使用的字典。

配置快速链接

可以预先确定要对某些关键字和短语返回的 URI。当用户在查询中指定了那些关键字或短语时，预定义的 URI 将与搜索结果一起返回。除了返回搜索服务器通过搜索索引返回的 URI 以外，还将返回快速链接 URI。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

第 138 页的『定制升降字字典』

为了提高搜索结果的质量，可以通过创建定制升降字字典来影响文档在搜索结果中的排名。

搜索高速缓存

当搜索服务器上的负载相对较高时，可以通过对搜索结果进行高速缓存来提高性能。

当搜索服务器处理搜索请求时，它们首先检查同一个查询的结果是否已存在于高速缓存中。如果搜索服务器找到适当的结果文档，则它们可以快速地搜索结果返回给用户。如果搜索服务器无法找到适当的结果文档，则它们将搜索索引。

当搜索高速缓存变满时，将以循环方式清除最旧的结果文档（即不经常使用的查询的结果文档）以便为新的搜索结果腾出空间。

从企业搜索管理控制台中，可以启用搜索高速缓存并指定高速缓存的容量（可同时对多少个查询的结果进行高速缓存）。

对搜索高速缓存设置进行更改时，必须重新启动搜索服务器才能使更改生效。

配置搜索高速缓存

可以对集合启用或禁用搜索高速缓存。也可以指定选项以控制搜索高速缓存的大小。

在开始之前

要为集合配置搜索高速缓存，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要配置搜索高速缓存：

1. 编辑集合，选择搜索页，然后单击**配置搜索服务器选项**。
2. 在搜索服务器选项页上，选中**使用搜索高速缓存**复选框。
3. 在**高速缓存中的最大条目数**字段中，输入搜索高速缓存可以存储的搜索结果集的最大数目。
4. 单击**确定**。
5. 要使更改生效，请监视搜索服务器并重新启动服务器进程。

定制同义词字典

要提高搜索结果的质量，可以使用户在搜索集合时能搜索查询项的同义词。

如果创建同义词字典，将它添加到企业搜索系统中，并将它与集合关联，用户就可以在搜索集合时搜索包含查询项同义词的文档。通过这种方式扩展查询，用户就更有可能找到感兴趣的所有文档，而不只是精确匹配查询项的那些文档。由于在创建同义词字典时定义了哪些单词互为同义词，从而有助于确保用户不必指定查询项的所有变化形式就可查找相关文档。

例如，您的公司可能会使用首字母缩写词和简写来指代部门、设备等，或者您的集合中的文档可能会包含特定于您的行业的词汇表。通过创建同义词字典，就可以确保包括首字母缩写词（例如，ACL）的查询会返回涉及那个首字母缩写词的扩展词条（例如，ACLs、access control lists、access controls 等）的文档。

企业搜索查询语言通过允许用户对查询项附加顿化音运算符来支持同义词。例如，查询 ~WAS 可能会返回涉及 WebSphere Application Server 的文档。应用程序开发者也可以通过不需要特殊语法的查询属性来提供同义词支持。

同义词字典包含了单词的变化形式，并具有以下特征：

- 单词不是特定于某种语言的，可在各种语言中使用。每个集合只有一个同义词字典。

- 单词未发生词尾变化。所有可能的词尾变化都必须添加到同义词列表中。例如，词尾变化可以是某个单词的单数和复数形式（例如，ACL 和 ACLs）。

添加到同义词字典的大多数词条都是语义严格相等的，这意味着如果词条 A 是词条 B 的同义词，则 B 也就是 A 的同义词。每次在查询中使用 A 时都会使用 B，反之亦然。

但也可以添加对应于某个词条的不同使用情况的词条，包括该词条的词性或更具体的变化形式。例如，可以有一个同义词组，包括了 building 和 house，而另一组则包括了 bank、shore 和 credit union。

词条之间的关系越不严格，搜索结果就会越大，但有些搜索结果可能与查询不相关。搜索和索引 API 提供的方法使用户可以在提交搜索请求时选择适当的同义词，这些方法还向用户显示了哪些查询项被扩展为哪些同义词。

要创建同义词字典，集合专家需要创建一个 XML 格式的同义词列表（或与应用程序开发者一起合作创建 XML 文件）。必须使用 WebSphere II OmniFind Edition 附带提供的工具来将 XML 文件转换为二进制文件（.dic）文件。

企业搜索管理员将此二进制文件上载到系统并为它指定显示名。集合管理员在配置集合的搜索服务器选项时，可选择用于搜索集合文档的同义词字典。

限制： 在将定制的同义词字典添加到系统后，就不能再编辑它。要修订可用于集合的同义词，必须：

1. 更新源 XML 文件。
2. 将 XML 源文件转换为新的字典文件。
3. 从使用旧同义词字典的集合中除去该字典。
4. 从系统中删除这个旧的同义词字典。
5. 将新的同义词字典添加到系统中。
6. 使新的同义词字典与要使用该字典的集合相关联。

相关概念

“Text Analysis Integration”中的“Synonym support in search applications”

相关任务

“Text Analysis Integration”中的“Creating an XML file for synonyms”

“Text Analysis Integration”中的“Creating a synonym dictionary”

将同义词字典添加到系统中


如果为搜索集中的文档创建了定制同义词字典，则必须使那些字典与企业搜索系统相关联。以后，可以选择要使用哪个同义词字典来搜索集合。

在开始之前

要添加定制同义词字典以便与企业搜索查询配合使用，您必须是企业搜索管理员角色的成员。

过程

要使同义词与企业搜索系统相关联：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在搜索页上，单击**配置同义词字典**。
4. 在配置同义词字典页上，单击**添加同义词字典**。
5. 在添加同义词字典页上，为同义词字典输入唯一的显示名，并可以选择输入描述。
6. 指定 .dic 文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。
7. 单击**确定**。定制同义词字典这就被添加到企业搜索系统中并可用于搜索集合。

使同义词字典与集合相关联

如果有同义词字典与企业搜索系统相关联，在搜索集合时就可以选择使用同义词字典。如果查询项与字典中的某个术语相匹配，则搜索结果还将返回包含该术语的同义词的结果文档。

在开始之前

要为集合选择同义词字典，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要使同义词字典与集合相关联：

1. 编辑集合，选择搜索页，然后单击**配置搜索服务器选项**。
2. 在搜索服务器选项页上的**同义词字典名字段**中，选择用户查询此集合时要使用的同义词字典。

可用同义词字典的列表包含所有已添加到企业搜索系统中的同义词字典。

3. 单击**确定**。

定制无用词字典

为了提高搜索结果的质量，可以指定在查询处理期间要从查询项中自动除去某些词。

无用词字典包含特定于企业的常用词条，这些词条作为查询项用处不大。通过从查询中排除这些词，可以确保用户不会被大量仅仅略微相关的结果文档困扰（只返回与其他查询项匹配的文档）。在查询处理期间，搜索服务器将从查询中除去无用词。除去的词包括定制字典中的无用词以及为企业搜索预定义的无用词（例如常用前置词和文章）。

缺省情况下，WebSphere II OmniFind Edition 执行特定于语言的无用词识别功能。此过程从查询中除去诸如 a 和 the 之类的常用词。只需要为特定于企业或特定于域的用词定义定制无用词字典。

要创建无用词字典，集合专家需要创建一个 XML 格式的用词列表（或与应用程序开发者一起合作创建 XML 文件）。必须使用 WebSphere II OmniFind Edition 附带提供的工具来将 XML 文件转换为二进制文件（.dic）文件。

企业搜索管理员将此二进制文件上载到系统并为它指定显示名。集合管理员在配置集合的搜索服务器选项时，可选择用于搜索集合文档的无用词字典。

限制： 在将定制无用词字典添加到系统后，就不能再编辑它。要修订可用于查询处理的无用词，必须：

1. 更新源 XML 文件。
2. 将 XML 源文件转换为新的字典文件。
3. 从使用旧无用词字典的集合中除去该字典。
4. 从系统中删除这个旧的无用词字典。
5. 将新的无用词字典添加到系统中。
6. 使新的无用词字典与要使用该字典的集合相关联。

相关概念

“Text Analysis Integration”中的“Custom stop word dictionaries”

相关任务

“Text Analysis Integration”中的“Creating an XML file for stop words”

“Text Analysis Integration”中的“Creating a stop word dictionary”

将无用词字典添加到系统中


如果创建了定制无用词字典以便从查询中除去某些词，则必须使那些字典与企业搜索系统相关联。以后，可以选择要使用哪个无用词字典来搜索集合。

在开始之前

要将定制无用词字典添加到系统中，您必须是企业搜索管理员角色的成员。

过程

要使定制无用词与企业搜索系统相关联：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在搜索页上，单击**配置无用词字典**。
4. 在配置无用词字典页上，单击**添加无用词字典**。
5. 在添加无用词字典页上，为字典输入唯一的显示名。
6. 指定 .dic 文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。
7. 单击**确定**。定制无用词字典这就被添加到企业搜索系统中并可用于搜索集合。

使无用词字典与集合相关联

如果无用词字典与企业搜索系统相关联，在搜索集合时就可以选择使用无用词字典。如果查询项与该字典中的某个项匹配，在处理该查询前就会从该查询中除去该项。

在开始之前

要为集合选择无用词字典，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要使无用词字典与集合相关联：

1. 编辑集合，选择搜索页，然后单击**配置搜索服务器选项**。
2. 在搜索服务器选项页上的**无用词字典名字段**中，选择用户查询此集合时要使用的无用词字典。

可用字典的列表包含所有已添加到企业搜索系统中的无用词字典。

3. 单击**确定**。

动态摘要

动态摘要是一种技术，它确定结果文档的哪些短语最能够表达用户正在搜索的概念。

对于企业搜索，动态摘要尝试捕获文档中包含各种搜索项的句子。将选择少数几个句子或句子的一部分并将它们显示在搜索结果中。搜索项将在搜索结果的 **HTML** 版本中突出显示。

为集合配置搜索服务器选项时，可以指定搜索结果中文档摘要的最大显示长度。由于摘要中包含了突出显示的字符，所以返回给搜索应用程序的缓冲区将大于指定的最大值。但是，显示长度不会超出指定的最大值，尽管摘要可能较短（这取决于从源文档中抽取的摘要数据）。

在管理控制台中定制文档摘要

通过在企业搜索管理控制台中指定搜索服务器选项，可以定制文档摘要中显示的信息量。

在开始之前

要控制集合摘要的显示长度，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

指定的文档摘要最大显示长度值与指定的每个摘要可以包含的句子数共同起作用。导致文档摘要最短的值具有优先权。

例如，如果指定 4 个句子的限制，则文档摘要只包含 4 个句子，即使显示长度所允许的字符数多于那些句子中的总字符数亦如此。再例如，如果指定 10 个句子的限制，并对显示长度限制指定 500 个字符，则文档摘要包含的句子可能不到 10 个。

过程

要配置文档摘要显示长度：

1. 编辑集合，选择搜索页，然后单击**配置搜索服务器选项**。
2. 在搜索服务器选项页上，指定文档摘要的最大显示长度。当用户查看搜索结果时，文档摘要不会超出指定的值。

3. 指定每个文档摘要可以包含的句子数（摘要可以包含多达 10 个句子）。
4. 单击**确定**。
5. 要使更改生效，请监视搜索服务器并重新启动服务器进程。

通过编辑属性定制文档摘要

企业搜索查询的每个结果文档都包括一个摘要。可以通过编辑属性文件来定制每个摘要包含的信息量。

关于本任务

可以通过更改

`ES_NODE_ROOT/master_config/collection_ID.runtime.node1/runtime-generic.properties` 文件中以下属性的值来定制搜索结果描述:

MinWordsPerSentence

描述语句可包含的最小单词数。缺省值是 4。

MaxWordsPerSentence

描述语句可包含的最大单词数。缺省值是 20。

NumberOfReturnedSentences

组成一个文档描述的句子数。缺省值是 5。

MaxSentencesPerDocument

在创建描述的过程中，一个文档中将被视为候选值的最大句数。缺省值是 1000。

过程

要定制搜索结果中的文档摘要:

1. 在搜索服务器上，作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 使用文本编辑器来编辑以下文件，其中 `coll_ID` 是在创建集合时对集合指定的标识（或由系统分配的标识）:

`ES_NODE_ROOT/master_config/coll_ID.runtime.node1/runtime-generic.properties`

提示：要确定集合名及其标识之间的映射，请参阅

`ES_NODE_ROOT/master_config/collections.ini` 文件。

3. 更改想要定制的属性，然后保存并退出该文件。
4. 停止搜索服务器，然后重新启动它以应用更改。

使用快速链接

快速链接是每当用户提交包括特定词和短语的查询时在搜索结果中返回的文档。

使用企业搜索管理控制台来配置集合的快速链接。

快速链接

快速链接使您能向用户提供指向特定文档的链接，这些文档已预先确定为将与查询项相关。

快速链接是一个 URI，当查询包含特定的词或短语时，企业搜索自动地将此 URI 包括在搜索结果中。通常，快速链接 URI 出现在结果列表顶部，这有助于确保用户能够看到预先确定与查询相关的文档。

快速链接与其他搜索结果一起返回。搜索进程在索引中搜索与查询项相匹配的文档，并返回那些文档的 URI 和快速链接 URI。

在配置快速链接时，可以为 URI 指定描述性的标题和摘要，以帮助用户识别该文档和快速地确定该文档是否是他们想要检索的文档。

例如，对于 URI `http://www.ibm.com/education/us/`，可以使用诸如 `IBM Education in the United States` 之类的标题并提供摘要 `Solutions, products, and resources for professionals, educators, and students in the United States`。

要在企业搜索集合中使用快速链接，用于显示快速链接的选项在搜索应用程序中必须是可用的。用户通过使用某些搜索应用程序，可能可以允许或禁止在搜索集合时返回快速链接。

配置快速链接

要为企业搜索集合创建快速链接，使文档的 URI 与一些关键字相关联，这些关键字将导致该文档包括在搜索结果中。

在开始之前

要配置快速链接，您必须是企业搜索管理员角色的成员或者是该快速链接所属的集合的集合管理员。

关于本任务

要获取有关如何对快速链接指定关键字和 URI 的示例，请在创建或编辑快速链接时单击**帮助**。

不需要重新启动搜索服务器就可以使更改生效。

过程

要配置快速链接：

1. 编辑集合，选择搜索页，然后单击**配置快速链接**。
2. 在快速链接页上，单击**创建快速链接**。
3. 指定导致在搜索结果中返回此快速链接的关键字和短语、预先确定与此查询相关的文档的 URI 以及此快速链接的其他选项。

每行可以指定一个关键字、若干个关键字或一个短语（括在引号中的两个或多个单词）。用空格将关键字分隔开（不能使用逗号来对关键字定界）。按 **Enter** 键以开始一个新行。

4. 单击**确定**。

新的快速链接将与其他属于此集合的快速链接一起列在快速链接页上。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识 (URI) 都指示了将文档添加到该集合的搜寻器的类型。

企业搜索中的文档排名

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

搜索服务器支持一种丰富的查询语法并使用若干种技术来生成最相关的搜索结果，例如基于文本的打分和静态排名。可以扩展缺省排名行为，这是通过配置对文档在搜索结果中的重要程度有影响的选项实现的：

- 可以创建定制升降字字典以影响包含指定升降字的文档在搜索结果中的排名。
- 可以影响与指定 URI 模式匹配的文档的分数。
- 可以影响包含映射至升降类的字段的文档的分数。

相关概念

第 141 页的『基于升降类的文档排名』

通过将字段映射至升降类，可以影响文档在搜索结果中的排名。

第 140 页的『基于 URI 模式的文档排名』

通过对 URI 模式指定升降因子，可以提高或降低文档的重要程度。

第 138 页的『定制升降字字典』

为了提高搜索结果的质量，可以通过创建定制升降字字典来影响文档在搜索结果中的排名。

基于文本的打分

企业搜索动态地计算与查询中的项相匹配的每个文档的分数。

要计算与查询相匹配的每个文档的文本分数，企业搜索要考虑许多因素，如：

- 每个查询项在整个集合中的出现频率。通常，出现在大多数文档中的查询项对文档分数的贡献要比出现在更有限的文档集合的查询项要小。
- 每个查询项在匹配文档中的出现次数。通常，查询项在文档中的出现次数越多，它的分数就越高。
- 查询项在每个匹配文档中出现时的接近性。通常，在文档中以相互较为接近的方式出现的查询项对文档分数的贡献比以相互比较疏远的方式出现的相同查询项的贡献大。
- 在每个匹配文档中出现查询项的上下文。例如，出现在文档标题中的查询项对文档分数的贡献比出现在文档明文中的同一查询项贡献大。

每个文档的长度以及它的词汇表的丰富程度也是确定其分数的因素。

静态排名

对于某些类型的文档，可以关联一个静态排名系数，该系数将提高那些文档在搜索结果中的重要程度。

创建集合时，指定是否要使一个静态排名系数与集合中的文档相关联。对于 Web 内容，从其他文档指向一个文档的链接数目以及那些链接的起源可以提高该文档在搜索结果中的相关性。

对于包含日期字段或日期元数据的文档，可以使用文档的日期来提高它的相关性。例如，Nntp 新闻组中的新文章可能与旧文章更具相关性。如果数据源包含多个日期值，则可以选择哪个日期值对于确定文档在数据源中的相关性来说最重要。

如果对集合使用静态排名，则确保不要将使用不同排名类型的数据源混合在同一个集合中。例如，如果要将指向文档的链接用作静态排名系数，则确保集合仅包含 Web 文档。在将使用不同排名模型的源组合在同一个集合中时，搜索质量可能会下降。

还必须确保文档包含使得静态排名能够被应用的字段和值。例如，如果指定要将文档日期用作在集合中对文档进行排名的系数，并且文档未包含日期字段或属性，则搜索质量可能会下降。

相关任务

第 190 页的『从 WebSphere Portal 迁移集合』

要将集合从 WebSphere Portal 迁移至企业搜索，请在 WebSphere Portal 中准备集合，然后使用迁移向导来迁移它们。

定制升降字典

为了提高搜索结果的质量，可以通过创建定制升降字典来影响文档在搜索结果中的排名。

如果查询指定了升降字典中的词，包含该词的文档的重要程度就会根据该字典中对该词配置的升降因子提高或降低。

升降因子的范围是 -10 到 10。在查询处理期间，搜索服务器将提高包含带有正升降因子的词的文档的重要程度，并降低包含带有负升降因子的词的文档的重要程度。

例如，文档在与升降值较高的查询条件匹配时的排名比未应用升降值时的排名高。（只有经过升降的查询项才会对文档在搜索结果中的排名产生影响，升降因子仅仅是影响文档分数的一个因素。）

创建字典时，可以对任意数目的词指定相同的升降因子。字典可以包含单个词的项或多个词的项（多个词的项与短语匹配）。

如果在使用 OR 运算符的查询中指定了具有升降值权重的词（例如：this | that），就会计算查询项的平均权重值。计算得到的总分数将用于出现的所有 OR 查询操作数（不会为不同的 OR 查询操作数计算不同的分数）。

字段查询项不支持基于升降字典的升降处理。对查询项进行解析时，将只使用查询文本（而不使用字段名）来计算文档分数。（要将升降因子应用于字段中出现的查询项，可以将字段名映射至升降类。）

要创建升降字典，集合专家需要创建一个 XML 格式的升降列表（或与应用程序开发者一起合作创建 XML 文件）。必须使用 WebSphere II OmniFind Edition 附带提供的工具来将 XML 文件转换为二进制文件（.dic）文件。

企业搜索管理员将此二进制文件上载到系统并为它指定显示名。集合管理员在配置集合的搜索服务器选项时，可选择用于搜索集合文档的升降字典。

限制：在将定制的升降字典添加到系统后，就不能再编辑它。要修订可用于查询处理的升降字，必须：

1. 更新源 XML 文件。
2. 将 XML 源文件转换为新的字典文件。
3. 从使用旧升降字字典的集合中除去该字典。
4. 从系统中删除这个旧的升降字字典。
5. 将新的升降字字典添加到系统中。
6. 使新的升降字字典与要使用该字典的集合相关联。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

“Text Analysis Integration”中的“Custom boost word dictionaries”

相关任务

“Text Analysis Integration”中的“Creating an XML file for boost words”

“Text Analysis Integration”中的“Creating a boost word dictionary”

将升降字字典添加到系统中


如果创建了定制升降字字典，则必须使那些字典与企业搜索系统相关联。以后，您可以选择要使用哪个升降字字典来搜索集合。

在开始之前

要将定制升降字字典添加到系统中，您必须是企业搜索管理员角色的成员。

过程

要使定制升降字与企业搜索系统相关联：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在搜索页上，单击**配置升降字字典**。
4. 在配置升降字字典页上，单击**添加升降字字典**。
5. 在添加升降字字典页上，为字典输入唯一的显示名，并可以选择输入描述。
6. 指定 .dic 文件的位置。如果该文件小于 8 MB，它可以在您的系统上。如果该文件大于 8 MB，它必须在索引服务器上。
7. 单击**确定**。这样就将定制升降字字典添加到企业搜索系统中并可用于搜索集合。

使升降字字典与集合相关联

如果有升降字字典与企业搜索系统相关联，在搜索集合时就可以选择使用升降字字典。如果查询项与该字典中的某个项匹配，包含该项的文档的重要程度就会根据该字典中对该项指定的升降因子提高或降低。

在开始之前

要为集合选择升降字字典，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要使升降字典与集合相关联:

1. 编辑集合, 选择搜索页, 然后单击**配置搜索服务器选项**。
2. 在搜索服务器选项页上的**升降字典名字段**中, 选择用户查询此集合时要使用的升降字典。

可用字典的列表包含所有已添加到企业搜索系统中的升降字典。

3. 单击**确定**。

基于 URI 模式的文档排名

通过对 URI 模式指定升降因子, 可以提高或降低文档的重要程度。

在将所有文档添加至索引时都会对其指定缺省的静态排名分数。根据是否对集合启用了静态排名功能以及启用此功能时指定的静态排名类型(按文档日期排名, 或者, 对于 Web 文档来说, 按其他文档链接到该文档的数目排名)的不同, 这个缺省分数也会有所变化。

可以通过对 URI 模式指定升降因子来影响文档的相对重要程度。升降因子与缺省静态排名分数以及其他因素共同确定文档的最终静态分数。

配置的 URI 模式的顺序非常重要。索引服务器在计算集合中每个文档的值时, 它按照 URI 模式的列示顺序来对这些 URI 模式进行求值。对于每个 URI:

1. 索引服务器按顺序扫描各个 URI 模式。
2. 当索引服务器找到第一个与索引中的文档匹配的 URI 模式时, 它将为该 URI 模式配置的升降因子应用于该文档。
3. 如果某个文档无法与 URI 模式匹配, 就会使用缺省静态排名分数。

为 URI 模式配置升降因子后, 必须按照要让索引服务器扫描 URI 模式的顺序来安排它的位置。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时, 搜索进程将返回查询项和查询条件的最相关结果。

影响与 URI 模式匹配的文档的分数

通过将升降因子应用于缺省静态排名分数, 可以提高或降低与 URI 模式匹配的文档的重要程度。

在开始之前

要影响与 URI 模式匹配的文档的重要程度, 您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

配置的升降因子与缺省静态排名分数共同用于计算所有与指定 URI 模式匹配的文档的新静态分数。

升降因子只影响静态分数，并且升降因子仅仅是其中一个影响文档最终排名计算的因素。例如，如果某个文档是大量链接的目标（这导致初始分数较高），则不作为链接目标的文档的排名始终较低。

过程



要影响与 URI 模式相匹配的文档的分数：

1. 编辑集合，选择索引页，然后单击**通过 URI 模式匹配影响分数**。
2. 在通过 URI 模式匹配影响分数页上，单击**添加 URI 模式**。
3. 输入要在搜索结果中提高或降低重要程度的文档的 URI 模式。 例如：

```
http://domain.org/hr/*
db2://*ROI*
*/afs/*
```
4. 请为升降因子输入 -10 与 10 之间的值。 将根据这个升降因子来计算与 URI 模式匹配的所有文档的最终静态分数。
5. 单击**确定**。
6. 在通过 URI 模式匹配影响分数页上，按照索引服务器扫描新 URI 模式的顺序来安排这些 URI 模式的位置。

索引服务器将按照 URI 的列示顺序来计算静态排名分数。为了获得最佳结果，请首先列示较为具体的 URI。在以下示例中，/forms 子目录与 http://www.ibm.com/hr/* URI 模式匹配。为了确保正确地计算 /forms 子目录中的文档的分数，应该首先列示 /forms 子目录的 URI 模式：

```
http://www.ibm.com/hr/forms/* 8
http://www.ibm.com/hr/* -2
```

7. 要更改 URI 模式或升降因子，请选择该 URI 模式，然后单击  **编辑**。
8. 要从列表中除去某个 URI 模式，请选择该 URI 模式，然后单击  **除去**。
9. 要将升降因子应用于先前已建立了索引的文档，请重组索引。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识（URI）都指示了将文档添加到该集合的搜寻器的类型。

基于升降类的文档排名

通过将字段映射至升降类，可以影响文档在搜索结果中的排名。

对文档进行解析时，解析器根据文档标记所属的字段来对这些标记指定升降类。这些升降类包括在索引中，在查询求值期间，它们用来计算分数，这些分数将影响结果文档的排名。

为了影响分数的计算方式，可以为升降类配置数字升降因子。如果查询项与映射至升降类的字段中的标记匹配，则该标记的这一次出现将影响该文档的总分数。该分数是通过应用该升降类中配置的升降因子计算的。

例如，可以提升或降低标题字段的分数。当查询项出现在标题中时，这次出现对文档分数的贡献就会提高，从而提高该文档在搜索结果中的排名。

要影响文档排名，请使用企业搜索管理控制台来为升降类指定升降因子以及将字段映射至升降类。已经为企业搜索预先配置了 16 个升降类。其中，8 个升降类设计成用于内容字段，另外 8 个升降类设计成用于元数据字段。可以编辑与缺省升降类相关联的分数，也可以使不同的或其他的字段与升降类相关联。

如果更改字段映射，则必须再次搜寻和解析文档以便可以将更改应用于先前已建立了索引的文档。如果更改了对升降类指定的因子，则请监视搜索服务器，然后停止并重新启动搜索服务器进程以使更改生效。

重复文档检测

在将字段映射至升降类时，必须指定是否应该使用该字段来检测重复文档：

- 如果要使用某个字段来检测重复文档，就将该字段视为内容字段，只有为内容字段设计的升降类才可供选择。
- 如果不使用某个字段来检测重复文档，就将该字段视为元数据字段，只有为元数据字段设计的升降类才可供选择。在这种情况下，将把两个只有指定字段不同的文档视为重复的。

高和低重新调用值

对查询进行求值时，搜索进程会估算将要返回的结果文档数。阈值确定了查询是拥有低重新调用值还是拥有高重新调用值：

低重新调用

如果估算的结果文档数低于低阈值，则将该查询视为低重新调用查询。

高重新调用

如果估算的结果文档数高于高阈值，则将该查询视为高重新调用查询。

混合 如果估算的文档数介于两个阈值之间，则该查询的重新调用值是两个阈值的混合。

每个升降类都指定在查询处理期间与低重新调用查询和高重新调用查询相关联的升降因子。低升降因子影响低重新调用查询的相对重要程度，高升降因子影响高重新调用查询的相对重要程度。两个升降因子的混合影响拥有混合重新调用值的查询的相对重要程度。

升降因子值控制文档中出现的每个查询项的相对重要程度。文档中每出现一次查询项都根据相应的升降因子进行计数。

在为集合配置升降类时，可以编辑缺省升降因子。例如，可以指定升降因子以确保标题字段中出现的查询项的计数比正常文本中多 5 倍。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

将字段映射至升降类

可以通过将字段名映射至升降类来影响字段的相对重要程度。

在开始之前

要将字段映射至升降类，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

系统使用升降因子来影响这样的文档的排名：这些文档在映射至升降类的字段中包含查询项。

企业搜索保留了一些字段名以便可以计算没有任何其他定义特征的字段和文本的分数（例如，不是标题字段的字段以及未强调的正规文本）。可以将其他字段映射至保留字段使用的升降类，但不能编辑或删除保留字段。

过程



要将字段映射至升降类：

1. 编辑集合，选择解析页，然后单击**将字段映射至升降类**。
2. 在将字段映射至升降类页上，单击**添加字段**。
3. 在将字段添加至升降类页上，输入要映射至升降类的字段名。

可以指定已搜寻的源或外部源中的字段名、从 XML 元素映射的字段名、从 HTML 元数据元素映射的字段名或者其中一个预定义字段名。

4. 指定是否要使用此字段来检测重复的文档。如果选中此复选框，则可用升降类列表将包含应用于内容字段的类。如果清除此复选框，则可用升降类列表将包含应用于元数据字段的类。
5. 选择升降类，然后单击**确定**。

添加的字段将显示在将字段映射至升降类页上。可以选择选项以编辑升降类和配置不同的升降因子，从而确定包含此字段的文档的分数。

6. 要更改某个字段是否用于检测重复文档或者要将该字段映射至另一个升降类，请单击  **编辑**。（不能编辑保留给企业搜索使用的字段。）
7. 要从升降类中除去字段，请单击  **除去**。（不能除去保留给企业搜索使用的字段。）
8. 要将更改应用于先前已建立了索引的文档，必须再次对那些文档执行搜寻、解析并建立索引。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

配置升降类的升降因子

为升降类配置的升降因子表示结果文档中的特定字段与查询的估算相关程度。对于拥有高升降因子的升降类来说，它们可以提高包含映射至该升降类的字段的结果文档的重要程度。

在开始之前


要为升降类配置升降因子，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

对于包含映射至升降类的字段的结果文档，系统将使用为升降类配置的升降因子、缺省静态排名分数和其他因素来计算新分数。

过程

要为升降类配置升降因子：

1. 编辑集合，选择解析页，然后单击**将字段映射至升降类**。
2. 在将字段映射至升降类页上，单击**编辑升降类**。
3. 在升降类页上，找到要更改的升降类，然后单击  **编辑**。
4. 在编辑升降类页上，指定新的高低升降因子值。可以对这两个因子输入同一个值。
5. 单击**确定**。
6. 要使更改生效，请监视搜索服务器并选择用于停止和重新启动搜索进程的图标。当用户提交查询时，对于包含映射至此升降类的字段的结果文档，将根据新的升降因子确定那些文档的相对重要程度。

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

缺省升降类值

企业搜索提供了 16 个升降类，可以使用这些升降类来影响文档在搜索结果中的排名。

为了计算没有任何其他定义特征的字段和文本的分数，保留了下列字段供企业搜索使用：

```
es_special_field.regular_text  
es_special_field.default_field
```

可以将其他字段映射至保留字段使用的升降类，但不能编辑或删除保留字段。

对于所有其他字段来说，可以对系统计算文档排名时使用的升降因子进行编辑。还可以将任意数目的字段映射至任何升降类，包括由保留字段使用的升降类。

下表列示了缺省配置中的升降类名、拥有低重新调用值的查询的缺省升降因子、拥有高重新调用值的查询的缺省升降因子以及映射至升降类的预定义字段名。

根据创建集合时为集合选择的静态排名方法的不同，缺省升降因子也会有所变化。选项包括不进行静态排名、由链接至文档的链接数确定的排名（对于 Web 源）或者由文档日期确定的排名。

表 3. 缺省升降类值

缺省高低升降因子				
升降类名	不进行静态排名	文档链接数	文档日期	预定义字段映射
内容类 A	低: 4 高: 2	低: 6 高: 1	低: 4 高: 2	es_special_field.regular_text
内容类 B	低: 5 高: 4	低: 7 高: 3	低: 5 高: 4	es_special_field.html_emphasized_text 包括这些 HTML 元素: b、big、caption、dfn、em、h4、h5、h6 和 strong
内容类 C	低: 7 高: 4	低: 9 高: 3	低: 7 高: 4	es_special_field.html_headers 包括这些 HTML 元素: h1、h2 和 h3
内容类 D	低: 2 高: 5	低: 1 高: 5	低: 2 高: 5	标题
内容类 E	低: 1 高: 1	低: 5 高: 10	低: 1 高: 1	es_special_field.anchor
内容类 F	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.anchor_same_dir
内容类 G	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.anchor_same_host
内容类 H	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
元数据类 A	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.default_field
元数据类 B	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
元数据类 C	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
元数据类 D	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	
元数据类 E	低: 1 高: 1	低: 5 高: 1	低: 1 高: 1	关键字
元数据类 F	低: 1 高: 1	低: 3 高: 1	低: 1 高: 1	es_special_field.urlhost
元数据类 G	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	es_special_field.urlpath
元数据类 H	低: 1 高: 1	低: 1 高: 1	低: 1 高: 1	描述

相关概念

第 137 页的『企业搜索中的文档排名』

当用户搜索集合时，搜索进程将返回查询项和查询条件的最相关结果。

企业搜索的搜索应用程序

搜索应用程序使您能够对企业搜索系统中的集合和外部源进行搜索。可以创建任意数目的搜索应用程序，并且单个搜索应用程序可以搜索任意数目的集合和外部源。

样本搜索应用程序

样本搜索应用程序演示了许多可用于企业搜索的搜索和检索功能。样本应用程序也是一个工作示例，它演示可以如何使用 IBM 搜索和索引 API (SI-API) 来构建反映企业目标的交互式定制搜索应用程序。

除非更改缺省配置文件中的属性，否则样本搜索应用程序允许搜索企业搜索系统中的所有活动集合和外部源。在将新集合和外部源提供给用户之前，可以使用样本搜索应用程序来对它们进行测试。

样本搜索应用程序自动与所有集合和外部源相关联。在生产环境中，企业搜索管理员控制允许哪些搜索应用程序搜索各个集合。

定制搜索应用程序

可以在 IBM WebSphere Application Server 环境中运行作为独立 Web 应用程序创建的搜索应用程序，也可以在 IBM WebSphere Portal 环境中将它们作为 portlet 启动。通过使用搜索和索引 API，可以设计与样本搜索应用程序类似并且工作方式在这两种环境中完全相同的搜索应用程序。

相关概念

第 119 页的『用于搜索文档的建立索引选项』

在配置用于搜索已搜寻的数据或外部源的选项时，或者在将 XML 和 HTML 元数据元素映射至搜索字段时，指定如何搜索文档以及如何搜索结果中显示那些文档。

第 171 页的『通过搜索应用程序标识实现的安全性』

要提供集合级安全性，请指定哪些搜索应用程序可以搜索每个集合和外部源。

第 172 页的『文档级安全性』

如果创建集合时对集合启用了安全性，则可以配置文档级安全性控制。文档级安全性确保搜索集合的用户只能够访问他们被允许查看的文档。

第 182 页的『通过 Portal Search Engine 实现的文档级安全性』

可以使用 IBM WebSphere Portal Search Engine 来在用户搜索企业搜索集合时强制执行文档级安全性。

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API overview”

“Programming Guide and API Reference for Enterprise Search”中的“Query syntax”

将搜索应用程序与集合相关联

在可以使用新的搜索应用程序之前，必须使该搜索应用程序与它可以搜索的集合相关联。

在开始之前

要使搜索应用程序与它们可以搜索的集合相关联，您必须是企业搜索管理员角色的成员。

过程

要使搜索应用程序与一个或多个集合相关联：

1. 在管理控制台的工具栏中单击**安全性**。
2. 在搜索应用程序页上，单击**配置搜索应用程序**。
3. 在配置搜索应用程序页上，单击**添加搜索应用程序**。
4. 输入搜索应用程序的名称。
5. 选择该应用程序可以搜索的集合：
 - 如果要让该搜索应用程序访问所有已添加至系统的集合，则单击**所有集合和外部源**。
 - 如果要让该搜索应用程序仅访问指定的集合，则单击**特定集合和外部源**。

当选择此选项时，将显示一组集合名和外部源名。对于该应用程序能够搜索的每个集合，选中**选择复选框**。

6. 单击**确定**。

样本搜索应用程序功能

企业搜索的样本搜索应用程序演示了大部分可构建到定制搜索应用程序中的搜索功能。

可以使用样本搜索应用程序来每次搜索一个、数个或所有集合和外部源。除非修改了缺省应用程序属性，否则可以使用此应用程序来搜索企业搜索系统中的所有集合和外部源。

查询功能

使用这些功能，可以：

- 指定简单的自由文本查询。
- 指定更复杂的查询以提高搜索结果的精确度。例如，可以搜索文档中的特定字段或使用查询语法来搜索包含或不包含特定词或短语的文档。
- 指定所要搜索的集合和外部源。
- 搜索特定源类型或所有源类型。
- 搜索特定类型的文档。例如，可以只搜索 Microsoft Word 文档或只搜索可移植文档格式 (PDF) 文档。
- 指定查询项使用的语言。也可以指定要搜索的文档的语言。
- 搜索集合的特定子集。例如，搜索应用程序可以限制只能查看预定义范围的文档（作用域），也可以提交仅搜索属于指定类别的文档的查询。
- 扩展查询以包括查询项同义词。如果已有同义词字典与此集合关联，则搜索结果中将返回包含查询项同义词的文档。

搜索结果功能

使用这些功能，可以：

- 查看与查询项匹配的搜索结果。
- 控制在每页上显示多少个结果文档以及向前和向后浏览结果集。
- 隐藏和显示有关结果文档的详细信息。例如，可以查看文档的简要描述或者查看诸如每个结果文档中的字段名之类的详细信息。
- 折叠来自相同来源的文档。例如，如果一个源返回了 100 个文档，则前两个最相关的文档将在结果集中被分组到一起。通过选择选项查看来自同一来源的更多文档，可以查看余下的 98 个文档。
- 按相关性、按升序排列的文档日期或按降序排列的文档日期对文档进行排序。
- 当在查询字符串中检测到可能拼错的词时，向您提供拼写更正建议。
- 查看有关结果文档所属类别的信息（如果集合使用类别的话），以及仅浏览属于特定类别的文档。
- 指定附加查询项以便在搜索结果中进行搜索。

文档检索功能

使用这些功能，可以：

- 通过单击文档 URI 来检索文档。如果有客户机应用程序可用，则还可能使用本机查看器来查看结果文档。例如，如果安装了 Notes 客户机应用程序，则可以使用该应用程序来查看来自 Lotus Notes 数据库的文档。

如果为搜寻器配置了文档级安全性，则只有有权访问安全内容的用户才能检索文档。

- 通过单击快速链接来检索文档。快速链接将关键字与 URI 相关联。如果查询包含指定的关键字，则相关联的 URI（已预先确定它们与那些关键字高度相关）将出现在搜索结果顶部。

样本搜索应用程序属性

可以编辑样本搜索应用程序的 `config.properties` 文件以指定环境选项、更改应用程序外观和控制用户启动搜索应用程序后可以使用的选项。

环境参数

可以指定选项以控制搜索应用程序的操作。

applicationName

指定有效搜索应用程序的名称。缺省值是 `Default`。

如果要将另一个搜索应用程序用作缺省应用程序，请更改缺省值。

提示：当应用程序名是 `Default` 时，可以使用样本搜索应用程序来在一个查询中搜索所有集合和外部源。

hostname

指定已配置为支持 WebSphere Application Server 实例的 Web 服务器的标准主机名。缺省值是 `localhost`。

要确保搜索应用程序工作正常，请将缺省值更改为 WebSphere Application Server 所要使用的标准主机名。例如，如果计算机主机名是 MyMachine 并且已将 Web 服务器配置为使用 www.ibm.com 作为其主机名，则指定 www.ibm.com。

port 指定已配置为支持 WebSphere Application Server 实例的 Web 服务器的端口号。缺省值是 80。仅当更改了 Web 服务器实例的端口时才需要更改此缺省值。

timeout

指定搜索请求超时前等待搜索服务器响应的秒数。此数值必须是整数（例如 60，而不能是 60.5 或六十）。如果未指定超时值，缺省值为 30 秒。

username

指定一个用户名，此用户名使企业搜索能够强制实施用户认证并在用户提交搜索请求时向 WebSphere Application Server 认证用户。仅当在 WebSphere Application Server 中启用了全局安全性时才会使用此字段。

password

指定所指定的用户名的密码。

filter 指定一个类，这个类用来检索搜索结果中列示的文档。缺省类是 com.ibm.es.api.filters.SetDocumentURIFilterFetch。仅当要使用定制类来检索文档时，才需要更改此值。

logging.level

指定要记录的详细信息量：

OFF 不记录消息。

INFO 记录参考消息。

FINE 记录错误消息。

ALL 记录所有消息。

源类型图标

可以定制表示结果文档所属数据源类型的图像。在 config.properties 文件中预定义了下列源类型图标，它们标识了安装 WebSphere Information Integrator OmniFind Edition 时支持的搜寻器和外部源。



documentSource.vbr.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Content Edition 搜寻器搜索到的。缺省图标是 /images/sourceVBR.gif。



documentSource.db2.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 DB2 搜寻器搜索到的。缺省图标是 /images/sourceDB2.gif。



documentSource.cm.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 DB2 Content Manager 搜寻器搜索到的。缺省图标是 /images/sourceCM.gif。



documentSource.dominodoc.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Domino Document Manager 搜寻器搜索到的。缺省图标是 /images/sourceDominoDoc.gif。



documentSource.exchange.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Exchange Server 搜寻器搜索到的。缺省图标是 /images/sourceExchange.gif。



documentSource.nntp.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 NNTP 搜寻器搜索到的。缺省图标是 /images/sourceNNTP.gif。



documentSource.notes.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Notes 搜寻器搜索到的。缺省图标是 /images/sourceNotes.gif。



documentSource.quickplace.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 QuickPlace 搜寻器搜索到的。缺省图标是 /images/sourceWorkplace.gif。



documentSource.unixfs.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 UNIX 文件系统 搜寻器搜索到的。缺省图标是 /images/sourceUnixFS.gif。



documentSource.web.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Web 搜寻器搜索到的。缺省图标是 /images/sourceWeb.gif。



documentSource.wps.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 WebSphere Portal 搜寻器搜索到的。缺省图标是 /images/sourceWPS.gif。



documentSource.winfs.icon

指定一个图像文件的路径和名称，该图像文件指示文档是由 Windows 文件系统 搜寻器搜索到的。缺省图标是 /images/sourceWindowsFS.gif。



documentSource.ldap.icon

指定一个图像文件的路径和名称，该图像文件指示文档属于为 LDAP 服务器创建的外部源。缺省图标是 /images/sourceLDAP.gif。



documentSource.jdbc.icon

指定一个图像文件的路径和名称，该图像文件指示文档属于为 Java 数据库连接（JDBC）数据库表创建的外部源。缺省图标是 /images/sourceJDBC.gif。

客户机查看器图标

结果文档可以显示在 Web 浏览器中。如果有结果文档类型的本机客户机应用程序可用，也可以通过客户机查看器来显示该文档。可以定制表示用来显示文档的客户机查看器类型的图像。在以下示例中，Lotus Notes 图标用来指示可以通过 Lotus Notes 客户机应用程序来显示文档：

`client.notes.icon=/images/notes.gif`

在搜索结果中，该图标以及客户机应用程序链接显示成：



客户机查看器

文档字段

对于包含字段的数据源类型来说，可以控制要将哪些字段显示在结果文档中。

fields.URI prefix=*space_separated_list_of_field_names*

在 URI 前缀中，必须通过在冒号字符 (:) 前面加上反斜杠字符 (\) 来对冒号字符进行转义。要使字段名列表延续到下一行，上一行应该以反斜杠字符 (\) 结尾。例如：

```
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
```

字段图标

对于包含字段的数据源类型和文档来说，可以定制表示字段的图像。在 config.properties 文件中预定义了下列字段图标。



field.icon.databasetitle

指定一个图像文件的路径和名称，该图像文件指示字段包含文档标题。缺省图标是 /images/notesdb.gif。



field.icon.databasename

指定一个图像文件的路径和名称，该图像文件指示字段包含文档所属数据库的名称。缺省图标是 /images/db2.gif。



field.icon.tablename

指定一个图像文件的路径和名称，该图像文件指示字段包含文档所属表的名称。缺省图标是 /images/table.gif。



field.icon.directory

指定一个图像文件的路径和名称，该图像文件指示字段包含文档所属目录的名称。缺省图标是 /images/closedFolder.gif。

缺省字段图标

可以指定当没有为搜索结果中显示的字段配置字段图标时要使用的图像。在 config.properties 文件中预定义了下列缺省字段图标。



field.defaultIcon

指定一个图像文件的路径和名称，该图像文件是搜索结果中的字段的缺省图标。缺省图标是 /images/database.gif。

日期字段

可以指定哪些字段是日期字段。此处指定的字段名格式与搜索结果中的日期数据格式类似。日期格式与 Web 浏览器中的语言环境设置匹配。

date.fields=*space_separated_list_of_field_names*

要使字段名列表延续到下一行，上一行应该以反斜杠字符 (\) 结尾。

示例:

```
date.fields=modifieddate createddate
```

文档标题

通过将标题文本替换为更有意义的数数据，可以指定文档的替代标题（即，可以清除标题）。例如，可以指定在搜索结果中不显示 Slide 1，而不必查看带有无意义标注 Slide 1 的文档标题。（可以使用更有意义的字段（例如文件名）来标识结果文档。）

也可以通过从文档标题中除去无意义的词来指定文档替代标题（即，可以截断标题）。例如，如果许多结果文档以 Microsoft Word 开头，则可以通过不显示重复文本来提高搜索结果的可读性。

titles.clean=*comma_separated_list_of_titles*

titles.truncatePrefix=*comma_separated_list_of_prefixes*

以逗号分隔的列表可以包含空格以及其他字符（逗号除外）。要使列表延续到下一行，上一行应该以反斜杠字符 (\) 结尾。

例如:

```
titles.clean=Slide 1, Layout 1, IBM Software Group Presentation Template, \
untitled, Untitled Document, PowerPoint Presentation, \
（此页没有标题）
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

用户首选项缺省值

可以指定搜索应用程序中的首选项页的缺省值。如果用户更改了首选项，则新值仅对该用户的当前会话有效。在 `config.properties` 文件中预定义了下列首选项。

preferences.resultsRange=10

指定每页搜索结果可以列示 10 个结果文档。

preferences.siteCollapsing=Yes

指定以下情况：如果在企业搜索管理控制台中配置了站点折叠规则，就在搜索结果中折叠与 URI 前缀规则匹配的 URI。

preferences.spellCorrections=Yes

指定当用户提交的查询包含拼写可能有误的词时显示建议拼写校正。

preferences.extendedHighlighting=No

指定除了在文档摘要字段中不突出显示查询项，在其他字段（例如文档标题）中也不突出显示查询项。

缺省集合和外部源

可以指定在首选项和高级搜索页中预先选择了哪些集合和外部源。用户可以对缺省集进行编辑，这样，搜索的集合和外部源就可以比缺省情况下搜索的少。如果在此处限制了集合和外部源集，则用户在修改他们的首选项或高级搜索选项时可以选择任何可供搜索应用程序使用的集合或外部源。

preferences.defaultCollections=*

preferences.defaultCollections=*space_separated_list_of_collection_IDs*

指定星号 (*) 以允许搜索所有集合和外部源。(集合和外部源必须与企业搜索管理控制台中的搜索应用程序相关联。) 这是 `config.properties` 文件中的缺省设置。

要对用户未修改其首选项或高级搜索选项时将会搜索的内容进行限制，请指定用户在缺省情况下将会搜索的集合和外部源的集合标识。

例如:

```
preferences.defaultCollections=*\npreferences.defaultCollections=coll_id1 coll_id2
```

附加搜索结果信息

可以定制随搜索结果一起提供的信息量以及控制用户是否能够对搜索结果进行过滤。下列设置是 `config.properties` 文件中的缺省设置。

refreshButton.show=false

控制是否显示刷新按钮以及刷新搜索应用程序的功能。如果将此选项设置为 `true`，用户就能够刷新可供搜索的集合和外部源列表（例如，如果已在企业搜索管理控制台中使该搜索应用程序与其他集合或外部源相关联）。

可以在测试对 `config.properties` 文件所作的更改时显示刷新按钮。保存更改后，可以单击刷新以了解那些更改对搜索应用程序的影响。如果没有刷新按钮，就必须在 WebSphere Application Server 中重新启动 ESSearchApplication 企业应用程序才能使更改生效。

如果没有可供搜索的集合或外部源（例如，如果指定了错误的主机名、搜索服务器未启动或者在 WebSphere Application Server 中未启动 ESSearchServer 企业应用程序），就会自动显示刷新按钮以帮助诊断问题。

extraMessages.show=false

控制是否在搜索结果底部显示警告消息和参考消息的显示区。错误消息总是显示在页面顶部。如果要查看其他消息，则将此选项设置为 `true`。

builtQueryString.show=false

控制是否在结果文档列表前面的区域中显示完全扩展的查询语法。如果要查看已处理的实际查询，请将此选项设置为 `true`。

refineResults.show=true

控制用户是否能够通过指定其他查询项来优化搜索结果。如果将此选项设置为 `true`，在搜索结果页面底部就会显示标注为在结果中搜索的查询框。

filter.showOnTwoLines=true

控制按源类型过滤结果和按文件类型过滤结果的选项在搜索结果中是显示为一行还是两行。（查看搜索结果时，用户可以选择源类型并选择文件类型以便仅查看与所选过滤器匹配的结果文档。）

如果要最大程度地增加可用于显示搜索结果的空间量，可将此属性设置为 `false`。如果想提高过滤器的可读性，特别是当可用过滤器在一行内无法完整显示时，可以将此属性设置为 `true` 以便每个过滤器单独地显示为一行。

定制条幅和徽标

可以定制在搜索应用程序顶部的条幅区中显示的图像。例如，您可能希望使用反映企业品牌的图像来替换 WebSphere II OmniFind Edition 的缺省图像。如果不想显示条幅，则使这两行中的一行或两行成为注释行。 `banner.icon` 属性标识显示在条幅区左侧的图形。 `banner2.icon` 属性标识显示在条幅区右侧的图形。

```
banner.icon=/images/WS_II_OFEdition.gif  
banner2.icon=/images/WS_II_mosaic.gif
```

定制背景图像

可以定制搜索应用程序各页中显示的背景图像。例如，可能想使用反映企业标记的图像替换 WebSphere II OmniFind Edition 的缺省图像。如果不想显示页面背景图像，请使用下列各行中的一行或多行成为注释行。

```
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  
logoff.backgroundImage=/images/IIOF_logout.gif
```

Links

在 `config.properties` 文件中， `Links` 区域中的属性使搜索应用程序各页的名称能够作为链接显示在每一页上，而不是显示在工具栏上或者显示在带有选项卡的页面中。在将搜索应用程序作为 `portlet` 运行并想最大程度地减少门户网站页面上用来显示搜索应用程序的空间量时，查看链接就很有用。

如果喜欢通过选择工具栏选项以及带有选项卡的页面上的选项来进行搜索应用程序导航，请将这些行注释掉。

Search tabs

在 `config.properties` 文件中， `Search tabs` 区域中的属性指定搜索应用程序的搜索视图中的标签页（搜索、高级搜索和类别树）使用的 Java Server Page (JSP) 的名称。除非很熟悉 Java 编程和 JSP，否则不要编辑这些页面。

定制此区域的方式示例包括：

- 将搜索应用程序定向到定制 JSP，那些定制 JSP 提供另一种标签页外观。
- 注释掉“类别树”条目。例如，如果不想为集合配置类别，则不需要在搜索应用程序中显示“类别树”页。
- 将标签页的条目复制到 `config.properties` 文件的 `Toolbars` 区域并注释掉这些行。例如，可能只想显示工具栏，而完全不想显示标签页。

Toolbars

在 `config.properties` 文件中， `Toolbars` 区域中的属性指定搜索应用程序工具栏使用的 Java Server Page (JSP) 的名称。除非很熟悉 Java 编程和 JSP，否则不要编辑这些页面。

定制此区域的方式示例包括:

- 将搜索应用程序定向到定制 JSP, 那些定制 JSP 提供另一种工具栏外观。
- 注释掉不想显示的项的工具栏条目。例如, 您可能不想在工具栏中包括关于页的链接。
- 从 `config.properties` 文件的 `Search tabs` 区域中移走用于显示高级搜索页的功能, 以使此选项仅显示在工具栏上。

有意义的文档类型标注

可以通过将实际文档类型名映射至更简明和有意义的术语来提高文档类型过滤器的可读性。可供搜索应用程序使用的文档类型是由搜索和索引 API (SI-API) 的 `AvailableDocumentTypes` 类定义的。为了便于使用, 在 `config.properties` 文件末尾也列示了可用的文档类型。

documentType.label=space_separated_list_of_document_types

指定在搜索结果文档类型过滤器行中显示的名称以及当用户选择过滤器时要显示的实际文档类型列表。

例如, 可以指定标注 `html` 并将各种 Web 文档的文件扩展名和 MIME 类型映射至该名称。当用户单击 `html` 以过滤搜索结果时, 将只显示具有指定扩展名和 MIME 类型的文档。

在 `config.properties` 文件中预定义了下列文档类型映射。

```
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
```

定制过滤器

可以指定定制查询以对显示的结果文档进行过滤。

filterCustom.label=query_terms

指定在搜索结果的定制过滤器行中显示的名称以及当用户选择该过滤器时用于优化搜索结果的查询。(查看搜索结果时, 用户可以选择定制过滤器以便仅查看与预定义查询匹配的结果文档。)

在以下示例中, 将对搜索结果进行过滤以便仅显示属于人力资源 (hr) 数据库的文档:

```
filterCustom.HR_database_only=databasename::hr
```

当用户单击 `HR_database_only` 以便对搜索结果进行过滤时, 就会处理查询 `databasename::hr`。在显示搜索结果时, 将只列示 `hr` 数据库中的文档。

在 `config.properties` 文件中提供了几个处于注释状态的定制过滤器示例。

编辑样本搜索应用程序属性

企业搜索的样本搜索应用程序能够搜索系统中的所有活动集合和外部源。可以编辑属性文件以指定 Web 服务器环境选项、使用另一个搜索应用程序作为缺省应用程序或者控制搜索应用程序启动时所显示的选项。

关于本任务

安装程序将把企业搜索样本搜索应用程序部署到企业搜索的搜索服务器上的 IBM WebSphere Application Server 中。要配置此搜索应用程序，编辑与该应用程序部署在一起的属性文件 `config.properties`。

要使更改生效，必须在 WebSphere Application Server 中停止 `ESSearchApplication` 企业应用程序，然后重新启动它。

过程

要编辑样本搜索应用程序属性：

1. 使用标准文本编辑器来编辑 `config.properties` 文件。

`config.properties` 文件安装在以下位置中，其中 `ES_INSTALL_ROOT` 是搜索服务器上的 WebSphere II OmniFind Edition 安装目录：

```
ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/  
ESSearchApplication.war/WEB-INF/config.properties
```

2. 编辑属性以指定 Web 服务器环境信息和搜索首选项，然后保存并关闭该文件。（在该文件中，# 符指示注释行。）
3. 停止 `ESSearchApplication` 应用程序，然后重新启动它：
 - a. 在搜索服务器上，启动 WebSphere Application Server 管理控制台。

以下列方式打开管理控制台：

- 使用 Windows 的开始菜单来选择程序。
 - 对于 WebSphere Application Server V5，打开 Web 浏览器并访问 `http://hostname:port/admin`，其中 `hostname` 是搜索服务器的主机名，`port` 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9090。
 - 对于 WebSphere Application Server V6，打开 Web 浏览器并访问 `http://hostname:port/ibm/console`，其中 `hostname` 是搜索服务器的主机名，`port` 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9060。
- b. 当提示输入用户标识和密码时，输入在 WebSphere Application Server 用户注册表中注册的用户名和密码。如果使用本地操作系统作为用户注册表，可指定企业搜索管理员标识和密码。
 - c. 登录到管理控制台后，单击**应用程序**，然后单击**企业应用程序**。
 - d. 选中 `ESSearchApplication` 的复选框，然后单击**停止**。
 - e. 在该应用程序停止后，再次选中该应用程序的复选框，然后单击**启动**。

访问样本搜索应用程序

通过在 Web 浏览器中指定 URL 来访问样本搜索应用程序。

在开始之前

必须为 Web 服务器环境配置样本搜索应用程序。

关于本任务

样本搜索应用程序安装在企业搜索的搜索服务器上。在将集合和外部源提供给用户之前，可以使用提供的此应用程序来对它们进行测试。还可以将样本应用程序用作创建自己的搜索应用程序的模型。

过程

要启动样本搜索应用程序：

1. 在 Web 浏览器中输入搜索应用程序的 URL。 例如：

```
http://SearchServer.com/ESSearchApplication/
```

SearchServer.com 是搜索服务器的主机名。

如果您的 Web 服务器未配置为使用端口 80，也需要指定正确的端口号。例如：

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

2. 如果在 WebSphere Application Server 中启用了安全性，可使用有效的用户标识和密码登录到该应用程序。

如果对任何可供搜索应用程序使用的集合启用了安全性，并且安全集合包含已配置为要在查询处理期间验证用户凭证的搜寻器，则可以配置用户概要文件。在我的概要文件页上，指定用于访问安全域的凭证。于是，可以对那些域进行搜索，而不必登录到那些域。

3. 在搜索页上，选择要搜索的集合和外部源并提交查询。

为样本搜索应用程序启用安全性

如果在 WebSphere Application Server 中启用了全局安全性，并想使用样本搜索应用程序来搜索安全集合，则必须更改样本应用程序和 WebSphere Application Server 中的配置。

在开始之前

- 您必须是企业搜索管理员角色的成员。
- 必须在 WebSphere Application Server 中启用全局安全性。如果将 WebSphere II OmniFind Edition 安装为多服务器配置，则在搜索服务器上启用全局安全性。请参阅 WebSphere Application Server 文档获取如何启用全局安全性的指示信息。
- 如果选择了轻量级目录访问协议 (LDAP) 用于 WebSphere Application Server 中的用户注册表，则必须将企业搜索管理员标识添加到 LDAP 注册表中（此标识和密码是在安装 WebSphere II OmniFind Edition 时指定的）。

例如，如果企业搜索管理员标识是 adminUser，则 LDAP 注册表中该用户的条目可能是 uid=adminUser,ou=Employees,o=IBM,c=US。请参阅 LDAP 服务器文档获取指示信息。

过程

要为企业搜索样本应用程序启用安全性：

1. 更新管理控制台中的搜索应用程序属性：
 - a. 作为拥有企业搜索管理员特权的用户登录到企业搜索管理控制台。
 - b. 单击工具栏上的**安全性**。
 - c. 在搜索应用程序页上，单击**配置搜索应用程序**。
 - d. 单击**添加搜索应用程序**，然后在**搜索应用程序名字段**中输入在安装 WebSphere II OmniFind Edition 时指定的企业搜索管理员标识。
 - e. 确保选择了**所有集合和外部源**，然后单击**确定**。
2. 编辑 config.properties 文件：
 - a. 如果正在使用 UNIX，可打开控制台窗口。如果正在使用 Microsoft Windows，可打开命令提示符窗口。
 - b. 切换到样本搜索应用程序的 WEB-INF 目录。以下命令显示为两行是为了更易阅读；指定该命令时应指定在一行上：

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. 使用文本编辑器来编辑 config.properties 文件。
 - d. 将 username 属性更改为有效的 WebSphere Application Server 用户名。
 - e. 将 password 属性更改为指定用户的密码。
 - f. 保存并退出该文件。
3. 在 WebSphere Application Server 中重新启动 ESSearchApplication 应用程序：
 - a. 在搜索服务器上，启动 WebSphere Application Server 管理控制台。

以下列方式打开管理控制台：

 - 使用 Windows 的**开始**菜单来选择程序。
 - 对于 WebSphere Application Server V5，打开 Web 浏览器并访问 http://hostname:port/admin，其中 *hostname* 是搜索服务器的主机名，*port* 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9090。
 - 对于 WebSphere Application Server V6，打开 Web 浏览器并访问 http://hostname:port/ibm/console，其中 *hostname* 是搜索服务器的主机名，*port* 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9060。
 - b. 当提示输入用户标识和密码时，输入在 WebSphere Application Server 中启用全局安全性时指定的管理员标识和密码。
 - c. 登录到管理控制台后，单击**应用程序**，然后单击**企业应用程序**。

- d. 选中 ESSearchApplication 的复选框，然后单击**停止**。
- e. 在该应用程序停止后，再次选中 ESSearchApplication 的复选框，然后单击**启动**。

企业搜索外部源

外部源是数据源，企业搜索应用程序能够搜索该数据源，而不需要对该数据源中的文档进行搜寻、解析或建立索引。

可以将下列类型的数据源作为外部源来进行搜索：

- Java 数据库连接 (JDBC) 数据库表 (仅限于 IBM DB2 通用数据库 (DB2 UDB) 和 Oracle 数据库)。将为 JDBC 数据库中的每个表创建一个独立的外部源。
- 轻量级目录访问协议 (LDAP) 服务器。将为每台 LDAP 服务器创建一个外部源。

在配置外部源信息后，必须使其与至少一个搜索应用程序相关联。于是，用户在查询通过对企业搜索数据进行搜索、解析和建立索引而创建的集合时，还可以搜索外部源。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API federators”

将外部源添加到系统中

在将外部源添加到企业搜索系统时，需指定所要添加的外部源的类型。一个向导将帮助指定数据源信息以及数据源搜索方式。

在开始之前

要将外部源添加到系统中，您必须是企业搜索管理员角色的成员。

限制

要将 Oracle 数据库作为外部源来对其进行搜索，必须在企业搜索的搜索服务器上安装 Oracle 客户机程序。

关于本任务

将关于外部源的信息添加到系统后，用户就能够通过企业搜索应用程序来查询该外部源了。可以允许用户搜索轻量级目录访问协议 (LDAP) 服务器和 Java 数据库连接 (JDBC) 数据库表。

在配置 LDAP 服务器时，一个向导将帮助指定使系统能够连接至该服务器的信息以及指定服务器搜索方式选项。

在配置 JDBC 数据库时，一个向导将帮助指定使系统能够连接至该数据库的信息、选择要搜索的表以及指定表数据搜索方式选项。将为添加到系统中的每个表创建独立的可搜索外部源。

过程

要将外部源添加到系统中：

1. 单击**外部源**以打开外部源视图。

2. 单击添加外部源。
3. 选择要添加的外部源的类型（LDAP 服务器或 JDBC 数据库）。
4. 单击下一步以开始配置该外部源。

将打开用于所创建的外部源类型的向导。请遵循该向导的提示来配置外部源。单击该向导中任何页上的**帮助**可了解有关可以指定的选项的更多信息。

在配置 DB2 通用数据库（DB2 UDB）和 Oracle 数据库的连接信息时，下列缺省 JDBC 驱动程序名和位置会对您有所帮助：

DB2: 旧 JDBC 驱动程序

驱动程序名: COM.ibm.db2.jdbc.app.DB2Driver

位置: *db2_install_root*/java/db2java.zip

DB2: 通用 JDBC 驱动程序

驱动程序名: com.ibm.db2.jcc.DB2Driver

位置:

db2_install_root/java/db2jcc.jar

db2_install_root/java/db2jcc_license_cu.jar

Oracle

驱动程序名: oracle.jdbc.driver.OracleDriver

位置:

oracle_home/jdbc/lib/classes12.zip

oracle_home/jdbc/lib/nls_charset12.zip

5. 在指定外部源搜索选项后，单击**完成**。

新的外部源将与已添加至系统的其他外部源一起列示在外部源视图中。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API federators”

使搜索应用程序与外部源相关联

在可以搜索外部源之前，必须至少使一个搜索应用程序与其相关联。

在开始之前

要使搜索应用程序与它们可以搜索的外部源相关联，您必须是企业搜索管理员角色的成员。

过程

要使搜索应用程序与一个或多个外部源相关联：

1. 在管理控制台的工具栏中单击**安全性**。
2. 在搜索应用程序页上，单击**配置搜索应用程序**。
3. 在配置搜索应用程序页上，单击**添加搜索应用程序**。
4. 输入搜索应用程序的名称。
5. 选择该应用程序可以搜索的外部源：

- 如果要让该搜索应用程序访问所有已添加至系统的外部源，则单击**所有集合和外部源**。
- 如果要让该搜索应用程序仅访问指定的外部源，则单击**特定集合和外部源**。

当选择此选项时，将显示一组集合名和外部源名。对于该应用程序能够搜索的每个外部源，选中**选择复选框**。

6. 单击**确定**。

相关概念

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API federators”

企业搜索安全性

企业搜索中的安全性机制使您能够保护来源免遭未经授权的搜索，并可以将管理功能限制为只能由特定用户执行。

借助企业搜索，用户可以搜索各种各样的数据源。要确保只有得到访问内容授权的用户才能执行该操作，并确保只有经过授权的用户才能访问管理控制台，企业搜索在几个级别协调并强制实施安全性。

Web 服务器

第一级安全性是 Web 服务器。如果在 WebSphere Application Server 中启用全局安全性，则可以将用户分配给管理角色并认证试图管理系统的用户。当用户登录到管理控制台时，该用户只能使用他有权管理的功能和集合。

搜索应用程序也可以使用 WebSphere Application Server 中的认证支持来认证使用搜索应用程序搜索集合的用户。

集合级安全性

当您创建集合时，可以在集合级启用安全性。在创建集合后就不能再更改此设置。如果未启用集合级安全性，则稍后不能再指定文档级安全性控制。

如果启用了集合级安全性，则全局分析进程将应用以下特殊规则：

- 为了确保对每个文档的安全性控制进行评估，将独立地对内容重复（或接近重复）的文档建立索引，而不是在规范表示中连带地对它们的内容建立索引。
- 通常，全局分析的锚点文本处理阶段使一个文档（源文档）中显示的文本与另一个不必显示该文本的文档（目标文档）相关联。这使得指定源文档中出现的文本的查询能够检索目标文档。如果允许用户查看目标文档而不允许他们查看源文档，则此类锚点文本处理将会引起安全性风险。启用了集合安全性后，将从索引中排除禁用文档的链接中的锚点文本。仅当文档自己的内容或元数据与查询相匹配时才在搜索结果中返回该文档。

可以在启用集合安全性与搜索质量之间进行权衡。启用集合安全性将减少为每个文档建立索引时所用的信息。副作用是：对于某些查询来说，找到的结果将会减少。

通过应用程序标识，集合级安全性也可用于搜索应用程序。要搜索集合，企业搜索管理员必须将搜索应用程序与它可以搜索的特定集合相关联。然后，可以使用标准访问控制机制来允许或拒绝用户访问搜索应用程序。

文档级安全性

在为集合配置搜寻器时，可以启用文档级安全性。如果选择此选项，搜寻器可将安全性令牌与它搜寻的每个文档相关联。安全性令牌与文档一起存储在索引中。对于某些搜寻器来说，也可指定在查询处理期间使用（本机数据源中配置的）当前访问控制数据来验证用户凭证。

搜索应用程序可以使用安全性令牌和用户凭证来强制实施访问控制。要确保用户仅搜索和检索他们有权访问的文档，搜索应用程序可以在传递给搜索服务器的查询中包括已登录用户的凭证。

集合的安全性超出了企业搜索可用来保护已建立索引的内容的认证和访问控制机制的范畴。还提供了安全措施来防止怀有恶意的用户和未授权的用户访问传输中的数据。例如，搜索服务器使用诸如安全套接字层（SSL）、安全 Shell（SSH）和安全超文本传输协议（HTTPS）之类的协议来与索引服务器以及搜索应用程序通信。

通过加密提供了其他安全性。例如，企业搜索管理员的密码（此密码是在安装产品时指定的）是以加密格式存储的。用户在用户概要文件中指定的密码也以加密格式存储。

为了增强安全性，您需要确保适当地隔离服务器硬件并保护它们免遭未经授权的入侵。通过安装防火墙，可以保护企业搜索服务器免遭通过网络的另一部分进行的入侵。还要确保在企业搜索服务器上没有空闲的已打开端口。配置系统，以使其仅在显式地指定给企业搜索活动和应用程序的端口上侦听请求。

管理角色

企业搜索使用角色的概念来控制对管理控制台中各种功能的访问。

在安装 WebSphere Information Integrator OmniFind Edition（WebSphere II OmniFind Edition）期间，安装程序为企业搜索管理员配置用户标识和密码。您第一次访问管理控制台时，必须作为此用户登录。如果未在 WebSphere Application Server 中启用全局安全性，则此用户标识是唯一可以用来访问企业搜索管理控制台的用户标识。

如果在 WebSphere Application Server 中启用了全局安全性，则可以将其他用户登记为企业搜索管理用户。通过将用户指定到角色，可以限制对特定集合的访问以及控制每个管理用户可以执行的功能。指定给企业搜索中的管理角色的用户标识必须存在于 WebSphere Application Server 用户注册表中。

当管理用户登录时，企业搜索将认证用户标识。在控制台将只出现允许该用户管理的集合和功能。

可以将用户登记到下列管理角色中：

企业搜索管理员

这些用户创建集合并有权对企业搜索系统进行全方位的管理。在安装 WebSphere II OmniFind Edition 时，您为第一个企业搜索管理用户指定用户标识和密码。在第一次登录之后，此用户可以将其他用户分配给企业搜索管理员角色。

集合管理员

这些用户可以编辑、监视和控制特定集合或所有集合的操作。这些用户不能创建集合。集合管理员只有在得到企业搜索管理员的授权后才能监视和执行系统级活动。

操作员 这些用户可以监视和控制特定集合或所有集合的操作。例如，这些用户可以启动和停止集合活动，但他们不能创建集合或编辑集合。操作员只有在得到企业搜索管理员的授权后才能监视和执行系统级活动。

监视员 这些用户可以监视特定集合或所有集合。这些用户不能控制操作（例如启动和停止服务器）、创建集合或编辑集合。监视员只可以观察而不能执行系统级活动，但只要得到企业搜索管理员授权后就可以执行系统级活动。

相关任务

第 16 页的『登录到管理控制台』

要管理企业搜索系统，在 Web 浏览器中指定一个 URL，然后登录到管理控制台。

第 195 页的『启动企业搜索服务器』

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

配置管理用户

通过配置管理角色，可以限制对集合的访问并控制每个管理用户可以执行的功能。

在开始之前

在将用户指定到管理角色之前，确保在 WebSphere Application Server 中启用安全性。还要确保该用户标识存在于 WebSphere Application Server 用户注册表中。

要配置管理用户，您必须是企业搜索管理员角色的成员。

过程

要将用户指定到管理角色：

1. 单击**安全性**以打开安全性视图。
2. 在管理角色页上，单击**添加用户**。
3. 输入要登记的用户的用户标识并选择适当的管理角色。
4. 如果未将此用户登记为企业搜索管理员，则请指定此用户是否能够从**系统工具栏**访问页面。

例如，可以允许某些操作员或集合管理员监视系统级日志文件。

5. 如果未将此用户登记为企业搜索管理员，则请选择此用户可以管理的集合和外部源。

可以选中各个集合和外部源的复选框，也可以使用户能够管理所有集合和外部源。

认证与访问控制

为了保护内容不会被未经授权的用户访问，以及为了控制对管理功能的访问，企业搜索支持用户认证和访问控制。

认证

认证是系统验证用户声称他们自己拥有的身份的过程。由于访问权通常基于请求资源的用户的标识，所以认证对于有效的安全性来说十分关键。

要认证试图访问管理控制台的`用户`，企业搜索利用 WebSphere Application Server 附带的认证支持。

要认证搜索企业搜索集合的`用户`，搜索应用程序可以利用 WebSphere Application Server 中的安全性并实施您选择的`用户凭证`认证方法。通常，`用户凭证`由`用户登录`或`尝试访问搜索应用程序`时传递给搜索应用程序的`用户标识`和`密码`组成。

根据企业中的可用资源和协议，可以通过其他方法来实施用户认证。例如，可以强制要求用户在登录时通过使用智能卡、通过管理数字证书和公共密钥基础结构或通过指定凭单来标识他们自己，以便跟踪他们的认证状态。

访问控制

访问控制指的是在用户标识他们自己并对他们进行认证后限制他们可以执行的操作。访问控制表（ACL）是最常用的限制资源访问的方法。ACL 是用户标识（用户名、组名和用户角色等等）的列表。每个用户标识都与一组许可权相关联，这些许可权定义了用户的权限和特权。

例如，访问控制可以允许或拒绝访问文件服务器上的文件以及控制被允许的用户是否能够读取、创建、编辑或删除该服务器上的文件。

在企业搜索中，所有访问控制都依赖于用户是否有权读取索引中的数据。根据在搜索应用程序中启用访问控制的方式和管理企业搜索时对集合以及搜寻器指定的规则，您可以：

- 允许所有用户搜索集合中的所有文档。
- 允许所有用户搜索由某些搜寻器搜寻的所有文档，并限制对其他搜寻器搜寻的文档的访问。
- 允许特定用户搜索特定的文档。例如，当指定要通过 Notes 搜寻器搜寻的数据库时，可以指定选项以使某些用户能够搜寻某些视图和文件夹中的文档并且不允许其他用户搜索那些文档。

对 WebSphere Application Server 中的企业应用程序禁用安全性

要控制哪些 WebSphere II OmniFind Edition 活动需要用户认证，您可以对 WebSphere Application Server 中的各个企业应用程序单独禁用全局安全性。

关于本任务

WebSphere II OmniFind Edition 安装程序部署了 3 个企业应用程序到 WebSphere Application Server 中：

- ESAdmin 应用程序包含了企业搜索管理控制台的接口。
- ESSearchApplication 应用程序包含了样本搜索应用程序的接口。
- ESSearchServer 应用程序为 WebSphere II OmniFind Edition SIAPI 实施提供了所有远程通信，并使 SIAPI 接口能与搜索服务器通信。

缺省情况下，所有这 3 个企业应用程序都支持 WebSphere Application Server 全局安全性。当这些应用程序检测到全局安全性已启用时，它们就会开始认证接收到的所有请求。

某些公司可能会想要对特定的 WebSphere II OmniFind Edition 企业应用程序启用或禁用安全性。例如，您可能想要认证访问企业搜索管理控制台的所有用户，但不认证使用 WebSphere II OmniFind Edition SIAPI 接口或样本搜索应用程序的用户。

过程

要对特定的企业应用程序禁用安全性：

1. 在搜索服务器上，启动 WebSphere Application Server 管理控制台。

以下列方式打开管理控制台：

- 使用 Windows 的开始菜单来选择程序。
 - 对于 WebSphere Application Server V5，打开 Web 浏览器并访问 `http://hostname:port/admin`，其中 *hostname* 是搜索服务器的主机名，*port* 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9090。
 - 对于 WebSphere Application Server V6，打开 Web 浏览器并访问 `http://hostname:port/ibm/console`，其中 *hostname* 是搜索服务器的主机名，*port* 是 WebSphere Application Server 管理控制台的端口号。通常，管理控制台端口是 9060。
2. 当提示输入用户标识和密码时，输入在 WebSphere Application Server 中启用全局安全性时指定的管理员标识和密码。
 3. 登录到管理控制台后，单击**应用程序**，然后单击**企业应用程序**。
 4. 选中您想要禁用安全性的企业应用程序名称旁的复选框。
 5. 向下滚动，并单击**将安全角色映射至用户 / 组**链接。
 6. 找到 **AllAuthenticated** 角色并选中 **Everyone?** 列下的复选框。
 7. 单击**确定**。
 8. 单击**保存**链接保存您的更改。
 9. 如果您正在使用 WebSphere Network Deployment，则选中与**节点同步更改**复选框。
 10. 单击**保存**。
 11. 再次单击**应用程序**，然后单击**企业应用程序**。
 12. 选中您刚修改的应用程序的复选框，然后单击**停止**。
 13. 在该应用程序停止后，再次选中该应用程序的复选框，然后单击**启动**。

集合级安全性

要提供集合级安全性，请配置用于对内容建立索引的选项以及用于允许搜索应用程序搜索特定集合的选项。

创建集合时，可以选择一个选项以启用集合安全性。如果选择此选项，可以在以后配置文档级安全性控制。启用了集合安全性后，企业搜索全局分析进程也会就对重复文档和文档中的锚点文本建立索引应用不同的规则。

在创建搜索应用程序之后，搜索应用程序标识使您能够指定搜索应用程序可以搜索哪些集合以及哪些用户可以访问搜索应用程序。

重复文档分析

如果启用集合安全性，则全局分析过程不会标识集合中的重复文档。

在全局分析期间，建立索引过程将标识彼此重复或接近重复的文档。然后，它们使所有这些文档与内容的规范表示法相关联。通过允许标识重复文档，可以确保搜索结果不会包含多个内容相同（或接近相同）的文档。

如果在创建集合时启用了集合安全性，则不会标识重复文档，因此它们不与公共规范表示法相关联。而是，独立地对每个文档建立索引。这将确保用户只搜索具有与他们的凭证相匹配的安全性令牌的文档。例如，两个文档在内容上可能很接近，但使用不同的访问控制表来强制执行安全性。

禁用重复文档分析可以增强集合中的文档的安全性，但是，如果用户在搜索结果中接收到同一文档的多个副本，搜索质量就可能会下降。

锚点文本分析

如果启用集合安全性，则全局分析过程将应用特殊的规则以便对 Web 搜寻器搜寻的文档中的锚点文本建立索引。如果未启用集合安全性，可以在配置各个 Web 搜寻器时指定是否想要对禁用文档的链接中的锚点文本建立索引。

锚点文本是超文本链接中的信息，该信息描述该链接所连接的页面。例如，在以下链接中，文本 Query Syntax 是连接至 Web 站点上的 syntax.htm 页面的链接中的锚点文本：

```
<a href="../doc/syntax.htm">Query Syntax</a>
```

通常，Web 搜寻器沿着文档中的链接来搜寻其他文档并将这些链接的页面包括在索引中。在全局分析期间，索引不仅会将锚点文本与它所嵌入的文档（源文档）相关联而且还与目标文档相关联。在上面的示例中，锚点文本 Query Syntax 与目标页面 syntax.htm 以及与包含锚点构造的页面相关联。

如果在创建集合时启用了集合安全性，将禁用锚点文本处理。将不再对文档的锚点文本建立索引，除非它确实出现在文档或文档元数据中。因为禁用文档的锚点文本未与用户有访问权的文档关联，所以此安全性控制可确保不会向用户披露不允许他们访问的文档中的信息。

通过让用户只能搜索那些带有与他们的凭证相匹配的安全性令牌的文档，启用集合安全性能增强 Web 文档的安全性。然而，通过不处理锚点文本，搜索结果可能未包括所有可能与查询相关的文档。

如果未启用集合安全性，可以在配置各个 Web 搜寻器时选择选项，以对禁用文档的链接中的锚点文本建立索引。如果指定要对锚点文本建立索引，则分析过程和建立索引进程将对 Web 搜寻器已检索的所有页面中的锚点文本建立索引。如果指定不对锚点文本建立索引，将从索引中排除禁用文档的链接中的锚点文本。

对指向被禁止文档的链接中的锚点文本建立索引

如果文档包含的链接指向不允许 Web 搜寻器搜寻的文档，则配置 Web 搜寻器时可以指定是否要在索引中保留那些链接的锚点文本。

在开始之前

要配置对锚点文本建立索引的选项，您必须是企业搜索管理员角色的成员或者是想要配置的 Web 搜寻器的集合管理员。


关于本任务

robots.txt 文件或 Web 文档的元数据中的伪指令可能会阻止 Web 搜寻器访问 Web 站点上的文档。如果允许 Web 搜寻器搜索的文档包含指向被禁止文档的链接，则您可以指定如何处理那些链接的锚点文本。

在配置 Web 搜寻器时，可以指定是否要对禁用文档的锚点文本建立索引。为获取最大程度的安全性，应指定您不想对禁用文档的链接中的锚点文本建立索引。但如果不对锚点文本建立索引，搜索结果就可能不会包括所有可能与查询相关的文档。

过程

要启用或禁用对禁用文档的链接中的锚点文本建立索引：

1. 编辑集合，选择搜寻页，找到要配置的 Web 搜寻器，然后单击  搜寻器属性。
2. 单击编辑高级 Web 搜寻器属性。
3. 要对此搜寻器搜寻的所有文档中的锚点文本建立索引，请选中对禁用文档的链接中的锚点文本建立索引复选框。通过搜索指向禁止 Web 搜寻器搜寻的页面链接的锚点文本中的文本，用户将能够了解这些页面。

要从索引中排除禁用文档的链接中的锚点文本，可清除此复选框。这样，用户就不能了解禁止 Web 搜寻器搜寻的那些页面。将把锚点文本与禁用文档一起从索引中排除。

4. 单击确定，然后在 Web 搜寻器属性页上，再次单击确定。
5. 要使更改生效，请停止并重新启动搜寻器。

要将更改应用到先前已建立了索引的文档，必须重新搜寻这些文档以便能再次对它们建立索引。如果先前执行的搜寻已将关于禁用文档的信息添加到索引中，现在将从索引中除去那些信息。

通过搜索应用程序标识实现的安全性

要提供集合级安全性，请指定哪些搜索应用程序可以搜索每个集合和外部源。

所有搜索应用程序都需要将应用程序标识传递给企业搜索 API。企业搜索管理员和搜索应用程序可以使用此标识来对集合和外部源强制实施安全性。

在搜索应用程序可以访问集合或外部源之前，企业搜索管理员必须使该搜索应用程序与它能够搜索的特定集合和外部源相关联。搜索应用程序可以搜索企业搜索系统中的所有集合和外部源，也可以只搜索指定的集合和外部源。

要强制实施访问控制，可以使安全性令牌（例如，用户标识、组标识或用户角色）与搜索应用程序相关联，并且只允许那些用户访问该应用程序。例如，可以限制对启动搜索应用程序的 URL 的访问。

要了解有关搜索应用程序标识以及如何将安全性控制合并到定制搜索应用程序中的更多信息，请参阅企业搜索的搜索和索引 API。

相关概念

第 147 页的『企业搜索的搜索应用程序』

搜索应用程序使您能够对企业搜索系统中的集合和外部源进行搜索。可以创建任意数目的搜索应用程序，并且单个搜索应用程序可以搜索任意数目的集合和外部源。

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API overview”

文档级安全性

如果创建集合时对集合启用了安全性，则可以配置文档级安全性控制。文档级安全性确保搜索集合的用户只能访问他们被允许查看的文档。

要控制对集合中的文档的访问，搜寻器可收集索引的安全性。对于某些类型的数据源来说，还可以在提交查询时验证当前用户凭证。要验证用户的当前凭证，可以将用户概要文件支持构建到定制搜索应用程序中。通过存储用户概要文件，可以使用户能访问文档而不会被多次要求指定凭证。

相关概念

第 147 页的『企业搜索的搜索应用程序』

搜索应用程序使您能够对企业搜索系统中的集合和外部源进行搜索。可以创建任意数目的搜索应用程序，并且单个搜索应用程序可以搜索任意数目的集合和外部源。

第 182 页的『通过 Portal Search Engine 实现的文档级安全性』

可以使用 IBM WebSphere Portal Search Engine 来在用户搜索企业搜索集合时强制执行文档级安全性。

“Programming Guide and API Reference for Enterprise Search”中的“Search and index API security”

通过存储安全性令牌进行验证

如果在创建集合时对集合启用了安全性，则可以通过将安全性数据存储在索引中来配置文档级安全性控制。

缺省情况下，每个文档都被视为公用文档，这表示所有用户都可以搜索它。对于大多数文档类型，可以通过使一个或多个安全性令牌与文档相关联，并将这些令牌与文档一起存储在索引中来达到文档级安全性。配置搜寻器时，可指定要使用安全性令牌来限制哪些用户可以访问由该搜寻器搜寻的文档。

如果数据源类型包括字段，则可以指定想要使用其中一个字段中的值来加强访问控制。如果数据源没有字段，或者您不想使用字段值来加强安全性，或指定的字段未能加强访问控制的值，您可以为搜寻器定义安全性令牌以便与文档相关联。

每个集合的管理员决定搜寻器要与文档相关联的安全性令牌。例如，安全性令牌可能表示用户标识、组标识、用户角色或您确定对数据源有效的任何其他值。如果数据源管理员更新了本机访问控制表，则更新过安全性控制将在下次刷新索引或重组索引时可用。

安全性令牌伴随着文档经过解析、分析和建立索引过程的各个阶段。如果搜索应用程序启用了安全性，则您可使用安全性令牌来控制对文档的访问。搜索集合的用户只能搜索他们的凭证允许他们看到的文档。如果用户的凭证不符合安全性规则，则用户不能搜索受安全性令牌保护的文档。

通过在 Java 类中对规则进行编码，可以应用定制业务规则以确定安全性令牌值。配置搜寻器属性时，请指定搜寻器搜寻文档时要使用的插件名。插件添加的安全性令牌存储在索引中，这些安全性令牌可用来控制对文档的访问。

在查询处理期间验证当前凭证

如果创建集合时对该集合启用了安全性，则某些类型的域允许在用户提交查询时验证该用户的当前凭证。

在配置下列类型的搜寻器时，可以选择一个选项以通过将用户凭证与本机存储库管理的当前访问控制进行比较来验证用户凭证：

- Content Edition 搜寻器（仅限于 Documentum、FileNet Panagon Content Services 和 Portal Document Manager 存储库类型）
- Domino Document Manager 搜寻器
- Notes 搜寻器
- QuickPlace 搜寻器
- WebSphere Portal 搜寻器
- Windows 文件系统 搜寻器

在对查询发出响应之前，搜索服务器与本机存储库进行交互以验证用户的当前许可权，然后从搜索结果中除去用户无权查看的所有文档。

这种强制保证文档级安全性的方法提供了高级别的安全性，因为用户凭证是与当前的安全性数据做比较，而不是与存储在索引中的安全性数据做比较。它还能确保本机存储库安全机制对访问进行控制，而不必考虑那些机制有多复杂。由于文档过滤是实时发生的，所以搜索结果将反映与搜索条件相匹配的每个文档的最新访问控制设置。

此方法的另一优点在于它不会影响索引的大小（不需要额外的空间用于对安全性令牌建立索引）。但由于验证操作需要与本机存储库连接，所以此方法可能会影响查询性能。

要获取最大程度的安全性，并将对查询性能的影响降到最低，应结合将安全性令牌存储在索引中的选项以及验证当前访问控制的选项。当用户提交查询时，验证发生在两个阶段：

- 首先，搜索服务器使用已建立了索引的安全性数据来快速确定用户是否有权访问从中搜寻文档的服务器和数据库（在速度方面对索引进行了优化，响应时间为亚秒级）。
- 接下来，搜索服务器创建一个临时列表，该表只包含服务器上的域中允许用户访问的文档。搜索服务器使用此列表来连接至本机存储库并确定是否允许用户查看所请求的文档。

如果用户有权访问服务器和域，该用户就很有可能有权访问这些文档。但是，这个最终过滤阶段将确保在搜索结果中只返回与用户当前许可权设置匹配的文档。

相关概念

第 175 页的『对 Windows 文件系统文档强制实施文档级安全性』

为了在用户搜索 Windows 文件系统 搜寻器搜寻的文档时验证当前凭证，必须在搜寻器服务器和 Microsoft Windows 服务器上配置域帐户信息。

第 177 页的『对 Lotus Domino 文档强制实施文档级安全性』

如果要搜寻的 Lotus Notes 服务器使用 Notes 远程过程调用（NRPC）协议，则必须对搜寻器服务器进行配置以便可以强制实施文档级访问控制。

相关任务

第 177 页的『配置 Lotus Domino Trusted Server 以验证用户凭证』

要对使用 Notes 远程过程调用 (NRPC) 协议的 Notes 搜寻器搜寻到的文档强制实施安全性, 必须将所要搜寻的 Domino 服务器配置成 Lotus Domino Trusted Server。

用户概要文件和标识管理

通过为企业搜索创建用户概要文件, 用户可以存储凭证, 这样他们就能够搜索安全域。

要搜索要求在提交查询时验证用户凭证的域, 用户必须向搜索应用程序提供他们用来登录到该域的凭证。借助 WebSphere Information Integrator OmniFind Edition 标识管理, 用户可以将任意数目的域的凭证存储在用户概要文件中。凭证以加密形式安全地存储在企业搜索系统中。

当用户使用搜索应用程序时, 他们可以创建用户概要文件并注册他们的凭证。在企业搜索的样本搜索应用程序中, 此功能是由我的概要文件选项提供的 (定制搜索应用程序可能会以别的方式实施此功能)。

用户概要文件存储用户在登录到所要搜索的域时必须指定的各种凭证。如果满足下列全部条件, 用户就可以创建用户概要文件:

- 在 WebSphere Application Server 中启用了全局安全性。
- 在企业搜索管理控制台中启用了 WebSphere II OmniFind Edition 标识管理。
- 在搜索应用程序可搜索的至少一个集合中启用了安全性。
- 至少一个安全集合包含由以下搜寻器搜寻的文档: 该搜寻器通过要求在提交查询时验证用户的当前凭证来强制实施访问控制。
- 在为下列至少一种搜寻器类型配置文档级安全性时选择了要在查询处理期间验证当前凭证的选项:
 - Content Edition (对于某些类型的存储库)
 - Domino Document Manager
 - Notes
 - QuickPlace
 - WebSphere Portal
 - Windows 文件系统

概要文件列示所有可供搜索应用程序使用并要求在查询处理期间验证用户凭证的域。用户可以选择要为哪些域存储凭证。如果用户未指定某个域的凭证, 搜索结果就不会包括该域中的数据源中的文档。

如果未使用 WebSphere II OmniFind Edition 标识管理, 当用户查询要求验证当前凭证的域时, 搜索应用程序必须提供用户的安全上下文 (USC) 字符串。

配置标识管理

可以使用 WebSphere II OmniFind Edition 标识管理来存储用户概要文件。概要文件使用户能够搜索要求在查询处理期间验证用户凭证的域。

在开始之前

要配置标识管理选项, 您必须是企业搜索管理员角色的成员。

关于本任务

如果您在配置搜寻器的文档级安全性选项时指定要在查询处理期间验证用户凭证，用户在查询要求进行验证的域时就必须提供凭证。借助 WebSphere II OmniFind Edition 标识管理，用户可以创建用户概要文件并注册用来登录到安全域的凭证。凭证以加密形式存储在 WebSphere II OmniFind Edition 管理的安全数据库中。

当该用户搜索安全域时，搜索服务器将使用存储的凭证来认证该用户。如果凭证无效，搜索结果就不会包括该安全域中的文档。

过程

要配置标识管理：

1. 单击**安全性**以打开安全性视图。
2. 在搜索应用程序页上，单击**配置标识管理**。
3. 在配置标识管理页上，选中使 WebSphere II OmniFind Edition 能够在用户概要文件中管理用户凭证的复选框。
4. 单击**确定**。

对 Windows 文件系统文档强制实施文档级安全性

为了在用户搜索 Windows 文件系统 搜寻器搜寻的文档时验证当前凭证，必须在搜寻器服务器和 Microsoft Windows 服务器上配置域帐户信息。

在配置 Windows 文件系统搜寻器时，您需要指定是要搜寻本地计算机上的子目录还是要搜寻远程计算机上的子目录。如果对集合启用了安全性，则还可以指定选项以控制对所搜寻的子目录中的文档的访问。

如果您选择通过在用户提交查询时验证当前用户凭证来强制实施访问控制，则必须确保正确地配置域帐户。对于在本地计算机上搜寻到的文件与在远程 Windows 服务器上搜寻到的文件，为这两种文件设置域帐户的要求是不同的。

使用本地访问控制数据来进行验证

在验证当前用户凭证时，系统既使用本地用户帐户信息也使用域帐户信息（如果该计算机从属于某个 Windows 域的话）。要在查询处理期间验证凭证，这两个用户名都必须列示在所搜索的文档的安全性信息中。

本地帐户

对于本地帐户来说，用户名的格式如下所示：

COMPUTER_NAME\USERNAME

用户在登录时只需要指定用户名，但正确指定的 Windows 用户权限指派使用全名。例如，如果本地帐户用户名是 `abcuser`，则帐户全名可以是 `WINSERVER1\abcuser`。

当用户使用搜索应用程序并配置概要文件以搜索本地系统上的安全文档时，他们必须指定用来登录到 Windows 的用户名（例如 `abcuser`）。

域帐户 对于域帐户来说，用户名的格式如下所示：

DOMAIN_NAME\USERNAME

用户在登录时使用以下格式指定此信息:

`USERNAME@DOMAIN NAME`

例如, 如果为某个文件配置用户权限指派并选择域 `WIN1\abcuser`, 该帐户将显示为 `abcuser@win1.company.com`。

当用户使用搜索应用程序并配置概要文件使他们能够搜索安全域中的文档时, 他们必须指定用来登录到 `Windows` 的用户名 (例如 `abcuser@win1.company.com`)。

要在本地计算机上强制验证当前凭证, 搜寻器服务器使用的用户帐户必须拥有下列 `Windows` 用户权限。(要指派用户权限, 请使用 `Windows` 管理工具: **管理工具** → **本地安全策略** → **本地策略** → **本地用户权利指派**。)

- 搜寻器服务器运行时使用的用户标识必须拥有**以操作系统方式操作权限**。(此权限是安装 `WebSphere Information Integrator OmniFind Edition` 时在搜寻器服务器上为企业搜索管理用户配置的。)
- 用户必须拥有**在本地登录**用户权限。

使用远程域访问控制数据来进行验证

对于 `Windows` 操作系统来说, 将任何以 `\\servername` 开头的目录都视为远程目录。例如:

`\\software\utilities\IBM`

要访问远程目录, 用户使用以下格式指定他们的用户名:

`USERNAME@DOMAIN NAME`

当用户使用搜索应用程序并配置概要文件使他们能够搜索远程系统上的安全文档时, 他们必须指定用来访问远程 `Windows` 系统的用户名 (例如 `abcuser@win1.company.com`)。

要在远程计算机上强制验证当前凭证, 用户帐户必须拥有下列 `Windows` 用户权限。(要指派用户权限, 请使用 `Windows` 管理工具: **管理工具** → **域安全策略**。)

- 搜寻器服务器与所要搜索的 `Windows` 服务器必须是同一个域的成员。
- 搜寻器服务器运行时使用的用户标识必须拥有**以操作系统方式操作权限**。(此权限是安装 `WebSphere Information Integrator OmniFind Edition` 时在搜寻器服务器上为企业搜索管理用户配置的。)
- 用户必须拥有**作为批处理作业登录**用户权限。

相关概念

第 173 页的『在查询处理期间验证当前凭证』

如果创建集合时对该集合启用了安全性, 则某些类型的域允许在用户提交查询时验证该用户的当前凭证。

第 79 页的『`Windows` 文件系统搜寻器』

要在企业搜索集中包括 `Microsoft Windows` 文件系统中存储的文档, 必须配置 `Windows` 文件系统 搜寻器。

对 Lotus Domino 文档强制实施文档级安全性

如果要搜寻的 Lotus Notes 服务器使用 Notes 远程过程调用 (NRPC) 协议, 则必须对搜寻器服务器进行配置以便可以强制实施文档级访问控制。

要对在使用 NRPC 协议的 Lotus Notes 服务器上搜寻到的文档强制实施文档级安全性, 必须在搜寻器服务器上安装 Domino 服务器。此 Domino 服务器必须是 Domino 域的成员。请按照 Lotus Domino 文档中的指示信息来安装和配置 Domino 服务器。

还必须完成下列任务以使搜索服务器能够验证搜索安全集合的用户是否有权查看与搜索条件匹配的 Lotus Notes 文档。在将搜索结果返回给用户前, 将从搜索结果中除去该用户无权查看的文档。

- 『配置 Lotus Domino Trusted Server 以验证用户凭证』。
- 在 WebSphere Application Server 中启用全局安全性, 并将搜索应用程序配置为使用安全性。此步骤确保当用户尝试使用搜索应用程序时将被提示指定凭证。然后, 搜索服务器可以使用这些凭证来验证每个用户是否有权访问 Lotus Notes 文档。

相关概念

第 173 页的『在查询处理期间验证当前凭证』

如果创建集合时对该集合启用了安全性, 则某些类型的域允许在用户提交查询时验证该用户的当前凭证。

第 52 页的『Notes 搜寻器』

要在企业搜索集合中包括 IBM Lotus Notes 数据库, 必须配置 Notes 搜寻器。

配置 Lotus Domino Trusted Server 以验证用户凭证

要对使用 Notes 远程过程调用 (NRPC) 协议的 Notes 搜寻器搜寻到的文档强制实施安全性, 必须将所要搜寻的 Domino 服务器配置成 Lotus Domino Trusted Server。

在开始之前

如果要在搜索远程数据库时强制实施文档级安全性, 此过程是必需的。如果要搜索搜寻器服务器的本地数据库, 则此过程不是必需的。

要配置 Trusted Server, 必须在搜寻器上安装 Domino 服务器。此 Domino 服务器必须是 Domino 域的成员。

关于本任务

在为 Notes 搜寻器配置文档级安全性选项时, 您指定是否要通过在用户提交查询时验证当前用户凭证来强制实施访问控制。要强制实施此类安全性, 所要搜寻的 Domino 服务器必须是 Lotus Domino Trusted Server。

当用户搜索要求验证当前用户凭证的域时, Trusted Server 使 Domino 服务器标识能够将上下文切换为当前用户标识。Domino 数据库将打开, 就象是当前用户已打开该数据库一样, 并且将强制实施所有数据库访问控制表信息。

以此方式切换上下文的功能通常仅适用于存储在本地 Domino 服务器的 data 目录中的数据库。从 Lotus Domino V6.5.1 开始, 通过 Trusted Server 提供此功能。要配置 Trusted Server, Domino 管理员指定要信任哪些 Domino 服务器执行敏感的操作, 例如在远程计算机访问数据库时充当另一个用户。

过程

要配置 Trusted Server，请在 Notes 搜寻器搜寻的所有 Domino 服务器上完成下列步骤：

1. 在 Domino 服务器上，使用 Domino 域管理员标识文件来打开 Lotus Domino 管理程序客户机。
2. 选择文件 → 打开服务器。
3. 输入要启用 Trusted Server 功能的 Domino 服务器的名称。
4. 选择配置选项卡。
5. 展开服务器对象，选择当前服务器文档，然后单击编辑服务器。
6. 选择安全性选项卡，滚动到文档底部，找到 **Trusted Server** 条目，然后单击向下方向按钮。
7. 指定下列其中一个选项：

LocalDomainServers

如果要将 Domino 域中的所有服务器都视为 Trusted Server，则选择此选项。

server_name

指定 Domino 服务器的名称，您将能够将其作为 Trusted Server 来对其进行搜寻和搜索。

如果要搜寻的 Domino 服务器在另一个 Domino 域中，则必须指定服务器名称或者选择 **OtherDomainServers** 组。还必须执行一些 Domino 过程以便在 WebSphere II OmniFind Edition Domino 服务器标识文件与其他 Domino 域之间进行交叉认证。请参阅 Domino 服务器文档以了解有关这些过程的信息。

8. 单击保存并关闭以保存更改。
9. 停止并重新启动已能够作为 Trusted Server 工作的远程 Domino 服务器。

相关概念

第 173 页的『在查询处理期间验证当前凭证』

如果创建集合时对该集合启用了安全性，则某些类型的域允许在用户提交查询时验证该用户的当前凭证。

第 52 页的『Notes 搜寻器』

要在企业搜索集合中包括 IBM Lotus Notes 数据库，必须配置 Notes 搜寻器。

禁用文档级安全性

您可以让用户能够搜索集合，而不考虑是否有任何访问控制与索引中的文档相关联。对于支持验证当前凭证的搜寻器来说，还可以让用户能够搜索集合而不在查询处理期间验证当前访问控制。

在开始之前

要配置文档级安全性选项，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

限制

仅当创建集合时对该集合启用了安全性时，才可以指定文档级安全性选项。

关于本任务

在配置搜寻器时，可以使安全性令牌与所搜寻的文档相关联。搜索应用程序可以使用这些令牌（它们存储在索引中）来在用户搜索集合时强制实施访问控制。对于某些搜寻器来说，还可以指定要在用户提交查询时验证文档本机存储库中与文档相关联的当前访问控制。

要除去这些安全限制，可以指定让搜索服务器忽略随查询一起传递的任何安全性令牌。也可以让用户能够查询文档，而不将他们的凭证与当前访问控制作比较。

如果您正在测试新集合，或者需要排除搜索应用程序的问题，则您可能想暂时地禁用文档级安全性。

过程

要禁用文档级访问控制：

1. 编辑集合，选择常规页，然后单击**配置文档级安全性**。
2. 在文档级安全性页上，如果您不想在用户查询集合时使用搜寻器设置的与文档相关联的安全性令牌，则选中**忽略索引中的文档级访问控制**复选框。

搜寻器将继续把安全性令牌添加到文档中，但搜索服务器将忽略那些令牌并允许用户搜索先前受保护的文档。

3. 如果您不想在用户提交查询时验证文档本机存储库中与那些文档相关联的当前访问控制，请选中**在查询处理期间不验证当前凭证**复选框。此复选框仅适用于由支持此功能的搜寻器搜寻的文档。

即使选中此复选框，其他文档级安全性选项也仍然有效。例如，如果您在配置搜寻器时指定了选项以将访问控制存储在索引中，则除非还选中了**忽略索引中的文档级访问控制**复选框，否则仍将应用那些安全性控制。

4. 单击**确定**。

企业搜索与 WebSphere Portal 的集成

可以通过在 WebSphere Portal 和 WebSphere Portal 搜索中心中部署企业搜索 portlet 来扩展 IBM WebSphere Portal 的搜索能力。

集成点

企业搜索 portlet 通过几种方式与 WebSphere Portal 集成:

WebSphere Portal

WebSphere Portal 为用户提供了与应用程序、内容、进程和人员进行交互的单一访问点。WebSphere Portal 框架使新的应用程序（称为 portlet）能够在不影响门户网站中的其他应用程序的情况下被集成和部署。

如果将企业搜索 portlet 部署到 WebSphere Portal 中，则可以从 WebSphere Portal 界面中对企业搜索集合进行搜索。通过 WebSphere Portal 配置，可以确保企业搜索 portlet 具有与 WebSphere Portal 环境中的其他 portlet 相同的外观。

Portal Search Engine

WebSphere Portal Search Engine 搜寻 Web 站点。管理 portlet 使管理员能够构建带索引的集合，而搜索 portlet 使用户能够搜索那些集合。

如果使用 WebSphere Portal V5.0.2 或更新版本，则可以将 Portal Search Engine 集合和分类法的配置信息迁移到企业搜索。

如果使用 WebSphere Portal V5.1 或更新版本，则可以使用 Portal Search Engine 文档搜索 portlet 来对企业搜索集合进行搜索。配置属性使管理员能够根据需要方便地在这两种搜索功能之间切换。

WebSphere Portal 搜索中心

WebSphere Portal 搜索中心为所有可通过 WebSphere Portal 来进行搜索的源提供了中央搜索起始点。搜索中心和通用搜索 portlet 使您能够搜索 WebSphere Portal 内容以及管理员向搜索中心注册的任何其他集合。

搜索中心使用分页界面。可以通过一个公共页面来搜索所有可用集合，也可以选择一个页面来搜索单个集合。例如，提供了用于 Portal Search Engine 索引的页面和用于 Portal Document Management 库的页面。

为了能够从搜索中心中对企业搜索集合进行搜索，WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) 提供了一个适配器和一个注册 portlet。该适配器在搜索中心界面中添加了一个企业搜索选项卡，而注册 portlet 向搜索中心注册该企业搜索 portlet。

必须先安装该适配器和注册 portlet，然后才能使用搜索中心来搜索企业搜索集合。

WebSphere Portal 搜寻器

企业应用程序 (ESPACServer.ear) 是在安装 WebSphere II OmniFind Edition 时安装在搜索服务器上的。在 WebSphere Portal 中部署此企业应用程序之后，可以使用企业搜索管理控制台来配置 WebSphere Portal 搜寻器并将 WebSphere Portal 站点添加至企业搜索集合。

集成的好处

通过支持对更广泛的数据源类型进行搜索，企业搜索增强了 WebSphere Portal 搜索环境。使用企业搜索 portlet，可以搜索 Web 站点以及 WebSphere II OmniFind Edition 支持的所有其他数据源类型。

企业搜索在可伸缩性方面也提供了好处。对于小规模或中等规模的企业（在这些企业中，单台服务器就足以支持搜索和检索工作负载），Portal Search Engine 很有用。要支持企业级的容量，企业搜索将工作负载分布到四台服务器上（一台用于搜寻数据，一台用于对数据进行解析和建立索引，另外两台用于支持搜索和检索过程）。

Portlet 部署概述

可以用于企业搜索的 portlet 取决于使用的 WebSphere Portal 版本：

WebSphere Portal V5.0.2

可以部署企业搜索 portlet 并使用它来对企业搜索集合进行搜索。此 portlet 可以与 Portal Search Engine portlet 共存。

WebSphere Portal V5.1 或更新版本

- 可以部署企业搜索 portlet 并使用它来对企业搜索集合进行搜索。此 portlet 可以与 Portal Search Engine portlet 共存。
- 在部署了企业搜索 portlet 之后，可以在 Portal Search Engine 中配置文档搜索 portlet 以便对企业搜索集合进行搜索。
- 在向 WebSphere Portal 搜索中心注册了企业搜索 portlet 之后，就可以使用搜索中心来对企业搜索集合进行搜索。可以选择一个页面以便仅对企业搜索集合进行搜索，也可以输入对企业搜索集合以及搜索中心中的所有其他可用集合进行搜索的查询。

相关概念

第 189 页的『从 WebSphere Portal 迁移到企业搜索』

企业搜索提供了一个迁移向导，您可以使用该向导来将分类法和集合从 IBM WebSphere Portal 迁移到企业搜索。

相关任务

第 77 页的『为 WebSphere Portal 搜寻器部署企业应用程序』

在创建 WebSphere Portal 搜寻器之前，必须在 WebSphere Portal 中部署企业应用程序 ESPACServer.ear。

通过 Portal Search Engine 实现的文档级安全性

可以使用 IBM WebSphere Portal Search Engine 来在用户搜索企业搜索集合时强制执行文档级安全性。

如果企业搜索搜寻器将组标识安全性令牌与它搜寻的文档相关联，并且如果配置了 Portal Search Engine 的文档搜索 portlet 来搜索企业搜索集合，则 Portal Search Engine 可以派生已登录用户的组标识，并将该组标识的安全性令牌与查询一起传递给企业搜索。安全性令牌确保只在搜索结果中返回用户有权查看的文档。

如果搜寻器将另一类型的安全性令牌与文档相关联（如用户标识或用户角色），并且要在搜索企业搜索集合时强制执行文档级安全性，则必须创建定制搜索 portlet。Portal Search Engine 仅派生组标识的安全性令牌。

相关概念

第 147 页的『企业搜索的搜索应用程序』

搜索应用程序使您能够对企业搜索系统中的集合和外部源进行搜索。可以创建任意数目的搜索应用程序，并且单个搜索应用程序可以搜索任意数目的集合和外部源。

第 172 页的『文档级安全性』

如果创建集合时对集合启用了安全性，则可以配置文档级安全性控制。文档级安全性确保搜索集合的用户只能访问他们被允许查看的文档。

部署“搜索”portlet

“搜索”portlet (ESSearchPortlet.war) 使您能够使用 WebSphere Portal Search Engine 来搜索企业搜索集合。

在开始之前

必须先要在 WebSphere Portal 中为企业搜索部署“搜索”portlet，然后才能将 Portal Search Engine 配置为使用该 portlet。

关于本任务

部署“搜索”portlet 后，可以继续使用 Portal Search Engine 来搜索 WebSphere Portal 中已建立了索引的数据。要使用户能够搜索企业搜索集合，WebSphere Portal 管理员应该编辑 Portal Search Engine 配置属性。可以在这两种搜索功能之间进行切换并使用最能够满足需要的搜索解决方案。

ESSearchPortlet.war 文件是在安装 WebSphere II OmniFind Edition 时安装到搜索服务器上的 ES_INSTALL_ROOT/bin 目录中的。缺省安装路径如下所示：

- UNIX 系统: /opt/IBM/es/bin/ESSearchPortlet.war
- Windows 系统: C:\Program Files\IBM\es\bin\ESSearchPortlet.war

过程

要在 WebSphere Portal 中部署“搜索”portlet:

1. 安装 portlet:

- a. 使用 WebSphere Portal Server 管理员标识来登录到 WebSphere Portal Server。
- b. 从工具栏中选择**管理**。
- c. 从左边的导航区中选择 **Portlet**，然后从 **Portlet** 菜单中选择**安装**。
- d. 单击**浏览**，从系统中选择 ESSearchPortlet.war 文件，然后单击下一步。
- e. 在下一页上，单击**安装**。安装 portlet 完成后，将显示以下消息：APIN0005I: 已成功地安装了 Portlet。

2. 修改 portlet 参数:

- a. 从左边的导航区中选择**管理 Portlet**，然后选择 **IBM 企业搜索**。
- b. 在新图标显示在选择框右边后，选择**修改参数**。
- c. 在 portlet 参数列表中，修改下列参数:

port 将此值设置为 WebSphere II OmniFind Edition 搜索服务器的 Web 服务器端口号。缺省值为 80。

applicationName

将此值设置为企业搜索的有效搜索应用程序的名称。（可用的搜索应用程序的名称列示在企业搜索管理控制台中的“搜索应用程序”页上。）缺省值是 Default，这是在搜索服务器上始终配置的搜索应用程序。

hostname

将此值设置为 WebSphere II OmniFind Edition 搜索服务器的标准主机名。

- d. 单击**保存**以保存更改，然后单击**取消**以退出“修改参数”页。
3. 创建 portlet 的门户网站页面：
 - a. 从左边的导航区中选择**门户网站用户界面**，然后选择**管理页面**。
 - b. 选择**我的门户网站链接**。
 - c. 单击**新建页面**，指定页面标题，然后单击**确定**。创建页面后，将显示以下消息：
APPR0010I: 已成功地创建了 *page_title*。
 - d. 单击**确定**。创建的页面将显示在可用页面列表中。
4. 将“搜索”portlet 添加到页面中：
 - a. 选择已创建的页面的“编辑”图标（铅笔）。
 - b. 单击**添加 portlet**。
 - c. 在**搜索**文本框中，输入企业，然后单击**搜索**。
 - d. 选中 **IBM 企业搜索**旁边的复选框并单击**确定**。
 - e. 单击**完成**。完成此步骤意味着已成功安装了“搜索”portlet 并将该 portlet 添加到 WebSphere Portal 页面中。
5. 访问 portlet：
 - a. 从页面顶部的导航栏中选择**我的门户网站**。
 - b. 选择将与 portlet 添加到系统时指定的页面标题匹配的页面。

第一次访问 portlet 页面时，该页面可能显示得较慢，这是因为系统必须编译 portlet 的 Java Server Page (JSP 文件)。

为企业搜索配置 WebSphere Portal“搜索和浏览”portlet

WebSphere Portal 提供了可用来搜索和浏览本机 WebSphere Portal 集合的 portlet。可以配置此 portlet 以便对企业搜索集合进行搜索。

过程

要配置“搜索和浏览”portlet 以便对企业搜索集合进行搜索：

1. 停止 WebSphere_Portal 服务器实例。
2. 将下列文件从 WebSphere II OmniFind Edition ES_INSTALL_ROOT/lib 目录复制到 WebSphere Portal Server 上的 /WebSphere/PortalServer/shared/app 目录中：
esapi.jar
siapi.jar
3. 启动 WebSphere_Portal 服务器实例，然后使用 WebSphere Portal 管理员标识登录 WebSphere Portal Server。
4. 复制 portlet:

- a. 单击右上角的**管理**，然后展开 **Portlet 管理**对象并选择 **Portlet**。
 - b. 搜索单词搜索。
 - c. 对于“搜索和浏览”portlet，单击**复制 Portlet**，为该 portlet 提供唯一的名称，然后单击**确定**。
5. 配置 portlet:
- a. 单击指定的 portlet 名旁边的**配置 Portlet** 图标。
 - b. 单击向右箭头以转到第二页参数。
 - c. 在**新参数**中，输入 ApplicationInfoId，在**新值**字段中指定有效的 WebSphere II OmniFind Edition 搜索应用程序的名称，然后单击**添加**以添加新参数。

可用的搜索应用程序的名称列示在企业搜索管理控制台中的“搜索应用程序”页上。缺省值是 Default，这是在搜索服务器上始终配置的搜索应用程序。

- d. 删除下列参数:

```
ApplicationInfoId_EXAMPLE
QueryFactoryImp
BrowseFactoryImp
SearchFactoryImp
IndexName
EJB_Example
IIOP_URL_Example
SOAP_URL_Example
EJB
SOAP_URL
IIOP_URL
```

- e. 添加下列参数并指定下面显示的值。单击**添加**以添加每个参数名和值。参数名区分大小写。

```
IMPLEMENT = ES
ApplicationInfoId = search_application_name (例如 Default)
SearchFactoryImp = com.ibm.es.api.search.RemoteSearchFactory
BrowseFactoryImp = com.ibm.es.api.browse.RemoteBrowseFactory
hostname = search_server_host_name (例如 omnifind.ibm.com)
port = search_server_port_number (例如 80)
```

6. 对门户网站页面指定“搜索和浏览”portlet 的新副本。要获取帮助，请参阅 WebSphere Portal 管理文档。
7. 访问 portlet:
 - a. 从门户网站页面顶部的导航栏中选择**我的门户网站**。
 - b. 选择与在步骤 第 184 页的 4 中复制 portlet 时指定的名称匹配的页面。

为搜索中心安装企业搜索适配器

在可以使用 WebSphere Portal 搜索中心来搜索企业搜索集合前，必须安装企业搜索适配器。

关于本任务

该适配器 (ESSearchAdapter.ear) 在搜索中心界面中为企业搜索添加一页。在安装此应用程序和注册 portlet 之后，可以将企业搜索集合添加到搜索中心并使用“通用搜索”Portlet 来搜索那些集合。

ESSearchAdapter.ear 文件是在安装 WebSphere II OmniFind Edition 时安装到搜索服务器上的 ES_INSTALL_ROOT/bin 目录中的。缺省安装路径如下所示:

- UNIX 系统: /opt/IBM/es/bin/ESSearchAdapter.ear
- Windows 系统: C:\Program Files\IBM\es\bin\ESSearchAdapter.ear

过程

要在搜索中心中安装企业搜索适配器:

1. 停止 WebSphere_Portal 服务器实例。
2. 如果 WebSphere Application Server server1 服务器实例尚未启动, 则启动它。
3. 将下列文件从 WebSphere II OmniFind Edition ES_INSTALL_ROOT/lib 目录复制到 WebSphere Portal Server 上的 /WebSphere/PortalServer/shared/app 目录中:

```
esapi.jar  
siapi.jar
```

如果询问是否要覆盖 siapi.jar 文件, 指定是。

4. 在 WebSphere Portal Server 上, 启动 WebSphere Application Server 管理控制台。如果提示您登录, 则进行登录。

以下列方式打开管理控制台:

- 使用 Windows 的开始菜单来选择程序。
 - 对于 WebSphere Application Server V5, 打开 Web 浏览器并访问 `http://hostname:port/admin`, 其中 `hostname` 是搜索服务器的主机名, `port` 是 WebSphere Application Server 管理控制台的端口号。通常, 管理控制台端口是 9090。
 - 对于 WebSphere Application Server V6, 打开 Web 浏览器并访问 `http://hostname:port/ibm/console`, 其中 `hostname` 是搜索服务器的主机名, `port` 是 WebSphere Application Server 管理控制台的端口号。通常, 管理控制台端口是 9060。
5. 单击应用程序, 然后单击安装新应用程序。
 6. 单击浏览, 从系统中选择 ESSearchAdapter.ear, 然后单击两次下一步。如果接收到关于策略文件的警告, 则单击继续。
 7. 单击下一步, 当看到“将模块映射至应用程序服务器”页时, 进行下列选择, 然后单击应用:
 - 在集群和服务器字段中, 选择 **WebSphere:cell=cell_name, node=node_name, server=server1**。
 - 选中 **ESSearchAdapterEJB** 模块旁边的复选框。
 8. 单击下一步直到看到“总结”页为止, 然后单击完成。
 9. 单击保存至主配置链接, 然后单击保存按钮以保存对 WebSphere Application Server 配置所作的更改。
 10. 重新启动 WebSphere Portal Server。

为搜索中心安装企业搜索注册 portlet

在可以使用 WebSphere Portal 搜索中心来搜索企业搜索集合前, 必须安装企业搜索注册 portlet。

关于本任务

| 注册 portlet (ESSearchAdapterPortlet.war) 向搜索中心注册企业搜索 portlet。在安装此
| portlet (以及企业搜索的适配器应用程序) 之后, 可以将企业搜索集合添加到搜索中心
| 中并使用“通用搜索”Portlet 来搜索那些集合。

| ESSearchAdapterPortlet.war 文件是在安装 WebSphere II OmniFind Edition 时安装到搜
| 索服务器上的 ES_INSTALL_ROOT/bin 目录中的。缺省安装路径如下所示:

- UNIX 系统: /opt/IBM/es/bin/ESSearchAdapterPortlet.war
- Windows 系统: C:\Program Files\IBM\es\bin\ESSearchAdapterPortlet.war

| 过程

| 要安装企业搜索注册 portlet:

- | 1. 使用 WebSphere Portal 管理员标识来登录到 WebSphere Portal Server。
- | 2. 从工具栏中选择**管理**。
- | 3. 从左边的导航区中选择 **Portlet**, 然后从 **Portlet** 菜单中选择**安装**。
- | 4. 单击**浏览**, 从系统中选择 ESSearchAdapterPortlet.war 文件, 然后单击**下一步**。
- | 5. 在下一页上, 单击**安装**。安装 portlet 完成后, 将显示以下消息: APIN0005I: 已成功
| 地安装了 Portlet。
- | 6. 将该 portlet (它名为 IBM WebSphere II OmniFind Edition registration portlet for
| enterprise search) 指定给一个门户网站页面。要获取帮助, 请参阅 WebSphere Portal
| 管理文档。

从 WebSphere Portal 迁移到企业搜索

企业搜索提供了一个迁移向导，您可以使用该向导来将分类法和集合从 IBM WebSphere Portal 迁移到企业搜索。

要迁移分类法和集合，请在企业搜索索引服务器上运行迁移向导。在迁移分类法之后，可以将其与企业搜索集合配合使用。还可以使用企业搜索来管理和搜索从 WebSphere Portal 迁移的集合。

在企业搜索中，分类法称为类别树。在迁移基于规则的分类法之后，可以使用企业搜索管理控制台来管理类别树。要将基于模型的分类法与企业搜索配合使用，必须在索引服务器上安装 WebSphere Portal。

如果要迁移分类法和集合，则总是先迁移基于模型的分类法文件，然后再迁移集合。如果不这样做，基于模型的分类将无法与从 WebSphere Portal 迁移的集合配合工作。

相关概念

第 181 页的『企业搜索与 WebSphere Portal 的集成』

可以通过在 WebSphere Portal 和 WebSphere Portal 搜索中心中部署企业搜索 portlet 来扩展 IBM WebSphere Portal 的搜索能力。

相关任务

第 94 页的『配置类别』

可以为集合创建任意数目的类别，并且每个类别可以包含任意数目的规则。规则确定哪些文档自动地与该类别相关联。

从 WebSphere Portal 迁移基于模型的分类法

通过使用 WebSphere Portal 分类法管理 Portlet，可以选择要与企业搜索集合配合使用的基于模型的分类法。已迁移到企业搜索的集合不受新的分类法迁移所影响。

在开始之前

在第一次运行迁移向导之前，请停止企业搜索系统以便可以对配置文件进行更改。第一次运行迁移向导时，必须指定 WebSphere Application Server 和 WebSphere Portal 的安装路径。以后运行迁移向导时，企业搜索系统可以处于活动状态。

关于本任务

要迁移基于模型的分类法，您必须在 WebSphere Portal 中选择分类法并将其导出。然后，使用企业搜索迁移向导来将该分类法迁移到企业搜索。

过程

要将基于模型的分类法从 WebSphere Portal 迁移至企业搜索：

1. 从 WebSphere Portal 分类法管理 portlet 中导出当前基于模型的分类法。分类法由下列 XML 文件组成：

```
synonyms.xml  
titles.xml  
treenodes.xml
```

2. 将这些文件复制到企业搜索索引服务器。
3. 在企业搜索索引服务器上，作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
4. 切换到企业搜索安装目录：

```
UNIX: cd $ES_INSTALL_ROOT/bin  
Windows: cd %ES_INSTALL_ROOT%\bin
```

5. 如果是第一次启动迁移向导，则输入以下命令以停止企业搜索系统：

```
esadmin stop
```

6. 输入以下命令以启动迁移向导，然后单击下一步。

```
UNIX: ./eswpsmigrate.sh  
Windows: eswpsmigrate.bat
```

如果这不是您第一次运行迁移向导，并且企业搜索系统处于活动状态，请确保不要在迁移向导运行时创建企业搜索集合。

7. 选择从 **WebSphere Portal** 导入基于模型的分类法文件，然后单击下一步。
8. 如果是第一次启动迁移向导，则指定 WebSphere Application Server 和 WebSphere Portal 的安装路径。单击下一步。
9. 浏览以获取包含基于模型的分类法文件的目录，选择必须迁移才能使用基于模型的分类法的 XML 文件（synonyms.xml、titles.xml 和 treenodes.xml），然后单击下一步。
10. 如果这是您第一次运行迁移向导，则在迁移完成后输入以下命令以重新启动企业搜索系统：

```
esadmin start
```

如果发生错误，请参阅位于迁移向导安装目录中的 MigrationWizard.log 文件。

相关概念

第 93 页的『基于模型类别』

如果在 IBM WebSphere Portal 系统中使用了基于模型类别，则可以对企业搜索集合继续使用那些类别。

相关任务

第 94 页的『配置类别』

可以为集合创建任意数目的类别，并且每个类别可以包含任意数目的规则。规则确定哪些文档自动地与该类别相关联。

从 WebSphere Portal 迁移集合

要将集合从 WebSphere Portal 迁移至企业搜索，请在 WebSphere Portal 中准备集合，然后使用迁移向导来迁移它们。

在开始之前

如果您计划迁移基于模型的分类法和集合，则必须先迁移基于模型的分类法文件，然后才能使用此过程来迁移集合。如果不这样做，基于模型的分类将无法与从 WebSphere Portal 迁移的集合配合工作。

过程

要将集合从 WebSphere Portal 迁移至企业搜索：

1. 在 WebSphere Portal Search Engine 中，停止您想要迁移的集合中的所有搜寻器进程并核准或拒绝所有暂挂文档。（企业搜索不支持暂挂文档的概念。）
2. 对于每个要迁移的集合，使用 Portal Search Engine portlet 来将设置导出至 XML 文件。
3. 如果将企业搜索索引服务器安装在单独的服务器上，则将导出的 XML 文件复制到索引服务器。
4. 在企业搜索索引服务器上，作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
5. 切换到企业搜索安装目录：

```
UNIX: cd $ES_INSTALL_ROOT/bin
```

```
Windows: cd %ES_INSTALL_ROOT%\bin
```

6. 输入以下命令以启动迁移向导，然后单击下一步。

```
UNIX: ./eswpsmigrate.sh
```

```
Windows: eswpsmigrate.bat
```

7. 选择从 **WebSphere Portal** 中的 **Portal Search Engine** 迁移搜索设置，然后单击下一步。
8. 浏览至包含所导出的 Portal Search Engine 配置文件的目录，选择要迁移的文件，然后单击下一步。将分析并验证所选配置文件。
9. 为每个集合输入以下信息，然后单击下一步以开始将集合迁移至企业搜索：
 - 在企业搜索中使用该集合时要使用的名称。
 - 确定集合的文档重要性的条件。静态排名系数可以是无、基于文档日期或基于从其他 Web 文档到该 Web 文档的链接数。
 - 要用于此集合的类别的类型。可以选择无、基于规则的类别或基于模型的类别。

如果选择基于规则的类别，则将 WebSphere Portal 集合的分类法及规则迁移至企业搜索。

仅当先前已将基于模型的分类法从 WebSphere Portal 迁移至企业搜索时，才可以选择基于模型的类别。

如果在迁移期间发生错误，请参阅位于迁移向导安装目录中的 MigrationWizard.log 文件。

您现在可以使用企业搜索管理控制台来配置对迁移的集合的其他设置。

需求：为迁移的集合配置 Web 搜寻器属性时，必须指定一个电子邮件地址来接收关于搜寻器和搜寻器的用户代理名称的注释。（配置 Web 搜寻器属性时，要获取帮助，请单击帮助。）

10. 从企业搜索管理控制台中为已迁移的集合启动搜寻、解析和建立索引过程。
11. 在确定已迁移的集合在企业搜索中可搜索之后，删除 Portal Search Engine 中的原始集合。
12. 可选：作为 WebSphere Portal 管理员，如果要使用户能够从 WebSphere Portal 中的门户网站搜索已迁移的集合，执行下列步骤。
 - a. 将企业搜索 portlet 部署在 WebSphere Portal 安装中。

在 WebSphere Portal Server 集群中，应该在安装了 WebSphere Application Server 部署管理器的服务器上执行此操作。部署管理器将企业搜索 portlet 分发到 WebSphere Portal Server 集群中的其他服务器。

- b. 将企业搜索 portlet 添加至适当的门户网站页面。

在 WebSphere Portal 中，搜索 portlet 的访问控制由对特定页面及 portlet 的可访问性建模。虽然迁移了集合设置，但 WebSphere Portal Server 管理员必须手工定位 portlet。

相关概念

第 25 页的『企业搜索集合』

企业搜索集合包含用户可以通过单个查询来搜索的整一组源。通过联合，用户可以在一个查询中搜索多个集合。

第 91 页的『基于规则类别』

可以配置规则以控制哪些文档与企业搜索集合中的类别相关联。

第 93 页的『基于模型的类别』

类别树使您能够查看集合中所有基于规则的类别。使用类别树来创建类别、删除类别和编辑使文档与类别相关联的规则。

第 137 页的『静态排名』

对于某些类型的文档，可以关联一个静态排名系数，该系数将提高那些文档在搜索结果中的重要程度。

迁移的集合设置

当从 IBM WebSphere Portal 迁移集合时，迁移向导将为集合和搜寻器创建缺省设置。

如果相同的设置存在于 Portal Search Engine 集合和企业搜索集合中，则当向导将该集合迁移到企业搜索时，它将使用 Portal Search Engine 设置。对于仅存在于企业搜索中的设置，向导使用您在迁移该集合时指定的设置或企业搜索中的集合的缺省设置。

存在于 Portal Search Engine 和企业搜索中的设置

迁移向导为您迁移的每个集合迁移下列设置：

- Portal Search Engine 集合中的 Portal Search Engine 站点
- 集合语言
- 分类法（即类别树）以及用于基于规则的类别的规则（如果企业搜索集合使用基于规则的分类的话）

集合中的每个 Portal Search Engine 站点都被合并到企业搜索 Web 搜寻器中。迁移向导迁移下列搜寻器设置：

- 启动 URL

- 并行搜寻进程数
- 搜寻深度
- 文档检索超时（以秒计）
- 缺省字符集
- 包括和排除搜寻规则

仅存在于企业搜索中的设置

当迁移集合时，您指定关于该集合的信息。迁移向导迁移那些设置并使用企业搜索中的集合的缺省配置来配置所迁移的每个集合。

可以通过使用企业搜索管理控制台来修改集合和 Web 搜寻器配置。显示在圆括号（）中的值是已迁移的数据的缺省设置。

- 集合名
- 文档静态排名策略
- 所使用的分类的类型，如基于规则或无
- 是否使用搜索高速缓存以及搜索高速缓存可以存放的带有搜索结果的查询数目（是，5000）
- 是否监视搜索响应时间并在超出限制时发出警报（是，5 秒）
- 是否使用访问控制（不）
- 用于刷新索引的时间表
- 用于重组索引的时间表
- 日志详细信息级别（所有消息）

迁移向导还为每个 Web 搜寻器创建下列设置：

- 搜寻器名称
- 搜寻器描述
- 最大页面长度
- 文档安全性设置
- 需要搜寻的文档多用途因特网邮件扩展（MIME）类型（如果适用于数据源类型的话）

在启动新迁移的 Web 搜寻器前，请复审所有搜寻器属性和搜寻空间设置并确保指定了所有必需值（必需字段用红色星号标记）。尤其要保证指定一个电子邮件地址来接收关于搜寻器和搜寻器的用户代理名称的注释。配置 Web 搜寻器属性时，要获取帮助，请单击[帮助](#)。

迁移向导日志文件

迁移向导将所有消息写至迁移向导安装目录中的 WpsMigratorLog.log 文件。

对于迁移的每个集合，WpsMigratorLog.log 日志文件包含从 WebSphere Portal Search Engine 读取的所有设置值，并指定将这些设置导入到企业搜索集合中的位置。

启动和停止企业搜索服务器

在创建集合之后，必须启动用于对数据进行搜寻、解析、建立索引和搜索的服务器。在对集合进行更改后，必须停止并重新启动服务器。

大部分企业搜索服务器可以持续运行或根据指定的时间表运行。例如，可以指定用于重组或刷新索引的时间表。在启动用于解析数据和搜索索引的服务器之后，通常只有在更改配置（例如更新类别或增大搜索高速缓存大小）时才需要停止并重新启动它们。

如果更改了集合的内容，或者更改了搜寻器从企业中的来源收集数据的规则，则通常需要停止并重新启动搜寻器才能使更改生效。如果未更改搜寻规则，则搜寻器将持续运行（对于 Web 和 NNTP 搜寻器的情况）或者根据指定的时间表运行。

启动企业搜索服务器

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

在开始之前

配置要搜寻的数据源，然后指定选项以指示如何对数据进行解析、建立索引和搜索。例如，如果要让用户能够在搜索结果中查看类别详细信息，则在启动解析器之前配置类别。

要启动企业搜索服务器，您必须是企业搜索管理员角色的成员、集合的集合管理员或者有权管理该集合的操作员。

必须按正确的顺序为集合启动企业搜索服务器。例如，必须先启动搜寻器并搜寻数据，然后才能对文档进行解析和建立索引。

过程

要启动企业搜索服务器：

1. 要使用企业搜索管理控制台或搜索应用程序：
 - a. 如果 IBM HTTP Server 未在运行，则启动它。
 - b. 如果 ESAdmin 和 ESSearchApplication 企业应用程序未在运行，则使用 WebSphere Application Server 管理控制台来启动它们。
2. 如果企业搜索公共通信层（CCL）未在运行中，则启动它：
 - a. 在索引服务器上，使用安装 WebSphere II OmniFind Edition 时指定的企业搜索管理员用户标识登录。

UNIX 系统

输入以下命令：



```
startccl.sh -bg
```


Windows 系统

选择开始 → 程序 → 管理工具 → 服务，然后重新启动 IBM WebSphere Information Integrator OmniFind Edition 服务。

3. 如果企业搜索系统未在运行，则启动它：
 - a. 在索引服务器上，使用安装 WebSphere II OmniFind Edition 时指定的企业搜索管理员用户标识登录。
 - b. 输入以下命令：


```
esadmin start
```

4. 启动企业搜索管理控制台并作为企业搜索管理员登录。如果使用管理角色，则可以作为集合管理员或对要启动的集合有权限的操作员登录。
5. 在集合视图中，找到要管理的集合，然后单击  监视。
6. 在搜寻页上，对于每个要启动的搜寻器，单击  启动。
 - 如果启动 Web 或 NNTP 搜寻器，则该搜寻器将立即开始搜寻数据。这些类型的搜寻器将持续运行以搜寻和重新搜寻 Web 站点和 NNTP 新闻组上的文档。
 - 如果启动具有其他搜寻器类型的搜寻器，搜寻器会话就会启动。搜寻器将在它的安排日期和时间开始进行搜寻。如果未安排该搜寻器的时间表，或者如果要更早地启动该搜寻器，请监视该搜寻器，然后单击所要搜寻的每个数据源的启动图标。


在搜寻器启动之后，可以让它持续运行。如果已安排了该搜寻器的时间表，则该搜寻器将在已安排的日期和时间再次运行。

7. 在搜寻到数据之后，打开解析页并单击  启动以启动解析器。

可以让解析器持续运行。通常，除非更改了数据的解析方式（如更新类别或 XML 字段映射），否则不需要停止解析器。

8. 可选：要强制启动建立索引的过程，而不是等待到已安排的日期和时间才开始建立索引，请打开索引页，然后，在重组区域中，单击  启动。

可以让建立索引的过程持续运行。在安排的日期和时间，将刷新并重组索引。

9. 在搜索页上，单击  启动。

可以让搜索服务器持续运行。通常，除非更改了搜索高速缓存或文档摘要设置，否则不需要停止搜索服务器。

相关概念

第 13 页的『企业搜索管理概述』

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

第 166 页的『管理角色』

企业搜索使用角色的概念来控制对管理控制台中各种功能的访问。

相关任务

第 16 页的『登录到管理控制台』

要管理企业搜索系统，在 Web 浏览器中指定一个 URL，然后登录到管理控制台。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

停止企业搜索服务器

如果对企业搜索服务器的配置作了更改，或者需要对问题进行故障诊断，则可能需要停止并重新启动企业搜索服务器。

在开始之前

要停止企业搜索服务器，您必须是企业搜索管理员角色的成员、集合的集合管理员或者有权管理该集合的操作员。

关于本任务


可以彼此独立地停止企业搜索服务器。例如，如果停止并重新启动搜寻器以便合并对它的配置所作的更改，则不需要停止并重新启动解析器。

如果想要停止企业搜索系统而不是停止各个服务器，则可以使用企业搜索管理员标识来登录到索引服务器（此标识是在安装 WebSphere II OmniFind Edition 时指定的）。然后，输入以下命令：

```
esadmin stop
```

过程

要停止企业搜索服务器：


1. 在集合视图中，找到要管理的集合，然后单击  **监视**。
2. 在搜寻页上，找到要管理的搜寻器并停止或暂停它。


如果更改了搜寻空间或搜寻器属性，则停止并重新启动搜寻器以便合并更改。如果更改了搜寻空间并且要将更改应用于集合中已包含的文档，则还必须重新搜寻那些文档。


提示： 可能会看到一条消息指示请求的操作已超时，即使该进程仍在后台运行亦如此。要确定该任务是否已完成，请单击管理控制台中的**刷新**（不要单击 Web 浏览器中的**刷新**）。当搜寻器的状态图标指示搜寻器已停止时，表明该过程已完成。

3. 在解析页上，单击  **停止**以停止解析器。

在更改数据解析规则时，需停止并重新启动解析器才能合并更改。那些更改将仅应用于新搜寻的文档。如果要将那些更改应用于索引中已包含的文档，就必须启动完整搜寻以重新搜寻所有文档。这样就可以再次对那些文档进行解析并建立索引。

4. 在索引页上，单击  **停止**以停止正在刷新或重组的索引。

还可以在监视索引队列时停止索引构建。为此，请选择工具栏上的**系统**，打开索引页，然后对要停止构建的索引单击  **停止**。

5. 在搜索页上，单击  **停止**以停止搜索服务器。通常，仅当更改了搜索高速缓存或文档摘要设置时，才需要停止并重新启动搜索服务器。

相关概念

第 13 页的『企业搜索管理概述』

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

监视企业搜索活动

在监视系统和集合活动时，您可以查看各种进程的状态、监视潜在的问题或调整配置以提高性能。

使用企业搜索管理控制台，您可以监视系统并根据需要调整操作。您可以查看每项主要活动（搜寻、解析、建立索引和搜索）的详细统计信息。统计信息包括平均响应时间和进度信息，如在搜寻或索引构建会话中已搜寻或建立索引的文档数。

通过单击图标，可以停止和启动大部分活动。这些操作使您能够暂停活动、对其配置进行更改或对问题进行故障诊断以及在您准备好允许活动继续时重新启动处理。

相关概念

第 13 页的『企业搜索管理概述』

使用企业搜索管理控制台来创建和管理集合以及外部源、启动和停止组件、监视系统活动和日志文件、配置管理用户、将搜索应用程序与集合和外部源相关联以及指定信息以强制实施安全性。

相关任务

第 195 页的『启动企业搜索服务器』

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

第 197 页的『停止企业搜索服务器』

如果对企业搜索服务器的配置作了更改，或者需要对问题进行故障诊断，则可能需要停止并重新启动企业搜索服务器。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

估算集合中的文档数

在创建或编辑企业搜索集合时，可以提供期望该集合包含的文档数估算值。资源管理器使用此数字来估算该集合需要的内存和磁盘资源，但不对集合大小强加限制。

在开始之前

要更改集合的估算大小，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

即使集合增大到您估算的大小，系统也不会停止将文档添加到索引中。当估算的内存和磁盘资源需求量超出系统中当前可用的数量时，资源管理器将发出警告。这些警告使您能够预防将来发生资源不足的问题。

如果对集合配置了警报，并且启用了当索引中的文档数超出限制时进行通知的选项，则缺省限制将符合您对集合中的估算文档数指定的值。监视员使用此数目结合指定的警报阈值百分比，在达到了对集合配置的最大文档数时发送电子邮件。

过程

要为集合可能达到的大小提供估算值：

1. 编辑集合，选择常规页，然后单击**配置常规选项**。
2. 在**估算的文档数**字段中，输入一个数字，表示您希望集合增长到多大。缺省值是 1 000 000 个文档。

检查系统资源的可用性

如果您在创建集合或搜寻器之后但在开始搜寻文档之前检查系统资源，则可以根据当前配置来确定可用的系统资源是否足以支持全力运行 WebSphere II OmniFind Edition。

在开始之前

要检查系统资源，您必须是企业搜索管理员角色的成员。

关于本任务

当您创建集合或搜寻器时，系统会自动检查资源的可用性。您也可以通过选择一个选项来随时检查系统资源的可用性。

系统将把配置集合和搜寻器属性时估算要使用的空间量与可用的系统空间量作比较，然后显示一条消息以通知您资源的可用性情况。该消息指示了哪台服务器可能没有足够的资源、您估算的集合大小所需的空間量以及可用空间量。

如果资源不够，则尝试执行下列其中一项纠正操作：

- 增大消息文本中标识的文件系统的大小。
- 编辑集合的常规选项并对估算的文档数指定小一些的数目。
- 编辑搜寻器属性并对要搜寻的最大文档数、最大页大小和最大线程数指定小一些的数目。
- 对于 Web 搜寻器，编辑搜寻器属性以便对最大活动主机数、最大新文档数和临时存储器中的最大文档数指定小一些的数目。
- 编辑系统的索引属性并减小允许并发运行的索引构建的数目。
- 避免在任何给定时间多个集合中的并发活动进程过多。例如，当来自多个集合的搜寻器和解析器同时运行时，资源可能会受影响。
- 从系统中删除搜寻器。
- 从系统中删除集合。

如果没有警告指示检测到资源可能不足，则表示可能有足够的资源来支持所要搜寻的数据以及为那些数据创建索引。

过程

要确定是否有足够的资源可用于企业搜索系统的当前配置：

1. 单击**系统**以打开系统视图。

2. 在常规页上，单击[检查系统资源](#)。

监视集合


您可以查看关于集合中的每个组件的状态的一般信息，也可以选择选项以查看关于各个组件和 URI 的详细信息。


在开始之前

所有企业搜索管理用户都可以监视集合。要启动或停止组件，或者要启用或禁用时间表，您必须是企业搜索管理员角色的成员、集合的集合管理员或者集合的操作员。

过程

要监视集合：

1. 在集合视图中，找到要监视的集合，然后单击 **监视**。将显示关于每个集合组件的当前状态的信息。

提示：如果您正在编辑集合并且已经打开了常规页，则可以单击 **监视**以切换到用于监视集合的视图。

2. 要查看有关某个 URI 的详细信息，请单击 **URI 详细信息**。

例如，您可能想查看索引是否已包含特定的 URI 或者是否已将 URI 所在的索引复制到搜索服务器。

3. 要监视单个组件并查看关于该组件的活动的详细统计信息，请单击**状态**图标。

相关概念

第 25 页的『企业搜索集合』

企业搜索集合包含用户可以通过单个查询来搜索的整一组源。通过联合，用户可以在一个查询中搜索多个集合。

查看关于 URI 的详细信息

您可以查看有关 URI 的详细信息。您可以查看关于由此 URI 表示的文档是如何被搜寻、被建立索引和被搜索的当前信息和历史信息。

在开始之前

在提交用来查看 URI 报告或将报告发送到电子邮件地址的请求之前，请确保要从中接收信息的组件处于活动状态。例如，要查看有关如何搜寻文档，对文档建立索引和搜索文档的详细信息，请确保 Web 搜寻器、索引服务器和搜索服务器正在运行。要对废弃的文档进行跟踪，请确保已配置用于进行文档跟踪的记录选项。

关于本任务

搜集有关 URI 的信息是一个相当耗时的过程。您可以选择一个选项以查看所请求的信息，然后等待该信息显示。更有效的选项是将报告发送到您指定的电子邮件地址。


在可以接收报告之前，必须确保为企业搜索配置邮件服务器信息。您在系统视图的日志页上配置电子邮件选项时指定此信息。


索引服务器和搜索服务器可以提供关于所有 URI 的信息（例如，该 URI 是位于索引中还是已被复制到搜索服务器）。要查看关于如何搜寻文档的信息，必须指定 Web 搜寻器已搜寻到的文档的 URI。

过程

要查看关于 URI 的详细信息：

1. 在集合视图中，找到要监视的集合，然后单击  **监视**。

提示：如果您正在编辑集合并且已经打开了常规页，则可以单击  **监视**以切换到用于监视集合的视图。

2. 单击  **URI 详细信息**。
3. 在 URI 详细信息页上，输入要查看其信息的 URI。
4. 选中要查看的信息类型的复选框：

搜寻器详细信息（仅对 Web 搜寻器提供）

选中此复选框以查看关于 Web 搜寻器如何搜寻文档的信息以及关于它在搜寻空间中的当前状态的信息。

索引详细信息

选中此复选框以查看是否已对文档建立索引并将其复制到搜索服务器的信息。

搜索详细信息

选中此复选框以查看关于可以如何搜索文档以及该文档是否可搜索的信息。

解析器废弃的文档

选中此复选框以查看先前解析某个文档时是否已从企业搜索系统中废弃了该文档以及废弃原因。

从索引中废弃的文档

选中此复选框以查看先前对某个文档建立索引或进行分析时是否已从企业搜索系统中废弃了该文档以及废弃原因。

5. 要等待报告显示，请单击**查看报告**。
6. 要将报告发送到电子邮件地址以便可以在以后查看，请单击**发送报告**。
 - a. 在发送详细的 URI 报告页上，**要通知的电子邮件地址**字段中输入用于接收报告的电子邮件地址。
 - b. 单击**发送报告**。

相关概念

第 214 页的『文档跟踪』

在各个处理阶段都可能会从系统中废弃文档。您可以指定选项以了解废弃文档的时间以及导致废弃该文档的问题。

相关任务

第 215 页的『查看关于废弃的文档的报告』

您可以查看关于已从企业搜索系统中废弃的文档的详细信息。仅当对集合启用了文档跟踪功能时，才会提供此信息。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识 (URI) 都指示了将文档添加到该集合的搜寻器的类型。

监视搜寻器


您可以查看关于集合中每个搜寻器的状态的一般信息，也可以选择选项以查看关于搜寻器活动的详细信息。


在开始之前


如果您的管理角色限制您只能监视集合，则您可以查看搜寻器统计信息，但是，您不能更改搜寻器的行为（如启动或停止搜寻器）。

过程


要监视搜寻器：

1. 在集合视图中，找到要监视的集合，然后单击  **监视**。
2. 打开搜寻页。

提示：如果您正在编辑集合并且已经打开了搜寻页，则可以单击  **监视**以切换到用于监视搜寻器的视图。

3. 如果该搜寻器正在运行或已暂停，并且您想要查看关于该搜寻器的详细状态信息，请单击  **详细信息**。您看到的统计信息类型随搜寻器类型的不同而有所变化。

如果您的管理角色允许您管理集合的进程，则可以在查看关于搜寻器活动的详细信息时启动、停止和暂停该搜寻器。如果可以安排该搜寻器的时间表，则还可以启用和禁用搜寻时间表。

4. 如果该搜寻器已停止或暂停，并且您想要启动搜寻器会话，则单击  **启动或继续**。

对于 Web 搜寻器：

如果该搜寻器已停止，则该搜寻器将再次开始搜寻并搜寻整个搜寻空间。
如果该搜寻器已暂停，则该搜寻器将从它的暂停目标的开头继续搜寻。

如果要强制使搜寻器立即开始执行完整搜寻，请单击**详细信息**图标，然后单击**启动完整重新搜寻**图标。搜寻器将开始搜寻整个搜寻空间，包括上次搜寻后未更改的页面。例如，如果更改了文档解析规则并想将那些规则应用于先前已建立了索引的文档，则可以重新搜寻所有文档。

对于 NNTP 搜寻器：



如果该搜寻器已停止，则该搜寻器将再次开始搜寻并搜寻整个搜寻空间。
如果该搜寻器已暂停，则该搜寻器将从它的暂停目标的开头继续搜寻。

对于所有其他搜寻器类型：

如果该搜寻器已停止，则该搜寻器将在它的已安排的日期和时间开始进行搜寻。搜寻器第一次搜寻数据源时，搜寻器将执行完整搜寻。当安排的搜寻重复执行时，搜寻器将搜索所有数据源更新（文档添加、删除和修改）或者仅搜寻文档添加和修改。搜寻类型是在搜寻器时间表中配置的。

如果未安排搜寻器的时间表，或者如果要更早地启动搜寻器，则单击**详细信息**图标。然后，在搜寻空间详细信息区域中，单击要启动的搜寻类型的

图标：“完整搜寻”、“全部更新”或“仅新文档和修改过的文档”。对于要搜寻的每个数据源（例如服务器、数据库或子文件夹），必须单击的相应启动图标。

5. 如果搜寻器正在运行，并且您想要停止它，则单击  **停止** 或  **暂停**。该搜寻器将停止搜寻数据，直到您重新启动或继续运行该搜寻器为止。

如果继续运行已暂停的搜寻器，该搜寻器将从它的暂停目标的开头继续搜寻。例如，DB2 搜寻器将从您暂停该搜寻器时正在搜寻的表中的第一行继续进行搜寻。

相关概念

第 31 页的『企业搜索搜寻器管理』

为要包括在集合中的不同类型的数据配置搜寻器。单个集合可以包含任意数目的搜寻器。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

查看有关 Web 搜寻器活动的详细信息


通过查看有关 Web 搜寻器活动的详细信息，可以评估整体性能并根据需要调整 Web 搜寻器属性和搜寻空间定义。


在开始之前


所有企业搜索管理用户都可以监视搜寻器活动。要启动或停止搜寻器，您必须是企业搜索管理员角色的成员、集合的集合管理员或者集合的操作员。

过程

要查看有关 Web 搜寻器的活动的详细信息：

1. 在集合视图中，找到要监视的 Web 搜寻器所在的集合，然后单击  **监视**。
2. 打开搜寻页。

提示：如果正在编辑集合并且已经打开了搜寻页，则可以单击  **监视**以切换到用于监视搜寻器的视图。

3. 如果要监视的 Web 搜寻器正在运行或已暂停，则单击  **详细信息**。
4. 在 Web 搜寻器的详细信息页上，查看或选择下列选项以查看有关搜寻器的当前和过去活动的详细统计信息。
 - a. 单击**线程详细信息**以查看正在搜寻 Web 站点的活动线程数以及正处于不活动状态的线程数。
 - b. 单击**活动站点**以查看有关搜寻器当前正在搜寻的 Web 站点的信息。
 - c. 单击**最近搜寻的 URL**。此信息显示搜寻器当前正在搜寻的内容。当您刷新视图时，如果列表中的项不更改，则表示没有搜寻正在发生。
 - d. 单击**搜寻器历史记录**以查看有关过去的搜寻器活动的报告。
 - e. 在 **URL 状态区域**中，输入要查看其信息的 Web 站点的 URL，然后单击**查看**。

例如，使用此选项来查看某个 URL 是否位于搜寻空间中、它是已被搜寻还是仅仅被发现、它何时应该再次被搜寻以及关于上次搜寻该 Web 站点的尝试的信息。

在关于该 URL 的详细信息显示后，您可以单击**站点历史记录**以查看关于在该 URL 位置处的搜寻器活动的其他统计信息。

Web 搜寻器线程详细信息

您可以监视 Web 搜寻器以查看正在搜寻 Web 站点的活动线程数以及处于不活动状态的线程数。

当您在监视集合时查看关于 Web 搜寻器的详细信息时，可以查看搜寻器线程的状态。您最有可能看到的状态包括：

正在等待

指示线程没有要搜寻的 URL。当线程完成了搜寻并且搜寻器无法找到更多的 URL 以进行足够快速的搜寻时，就会发生这种情况。例如，如果用于控制搜寻器在能够从同一站点检索另一个页面之前必须等待的时间长度的搜寻器属性值太大，则会导致无法足够快地提供 URL。

正在访存

指示线程正在从 Web 站点下载页面。

已完成 指示线程正在将它搜寻到的页面发送到该搜寻器的其余部分，但尚未准备好搜寻另一个 URL。

已暂挂 指示搜寻器已暂停

在理想情况下，所有线程都持续不断地访存页面。如果线程经常处于已完成状态，则表示数据库可能存在吞吐量问题。

如果线程经常处于等待状态，则请查看在搜寻器属性中对**最大活动主机数字**段指定的值。如果此值较小，则表示搜寻空间中可能没有足够的站点来使线程保持繁忙，或者可能没有足够的 URL 符合搜寻条件。可能导致活动程度较低的情况包括 DNS 查询故障和 robot 查询故障。

Web 搜寻器活动站点

您可以监视 Web 搜寻器以查看关于该搜寻器当前正在搜寻的 Web 站点的信息。

当您在监视集合时查看关于 Web 搜寻器的详细信息时，可以查看关于活动站点的统计信息。统计信息显示：

- 搜寻器从其内部数据库读入内存以便此时进行搜寻的 URL 的数目
- 搜寻器迄今为止已尝试搜寻的 URL 的数目
- 对于搜寻器的此次迭代，在将站点取消激活并从内存中除去该站点之前的剩余时间。
- 站点迄今为止在内存中已停留的时间

当搜寻器处理为其配置的搜寻规则时，此信息会不断地更改。理想情况下，已激活的 URL 数接近于为搜寻器属性中的**最大活动主机数字**段配置的值。

如果已激活的 URL 数接近于零，则表示搜寻器无法找到合格的 URL。可能会导致活动程度较低的情况包括 DNS 查询故障、网络连接问题、数据库错误和搜寻空间定义问题。例如：

- 如果已有许多站点长时间停留在内存中，并且只搜寻了很少的 URL，则查找网络连接问题。
- 如果在列表中没有足够的站点，则查找搜寻空间定义问题或 DNS 查找问题。
- 如果站点正在以合理的速率被搜寻，但是在许多 URL 未被搜寻的情况下离开内存，则您应该编辑搜寻器属性并调整 **URL 可以留在内存中的最长时间** 字段中的超时值以使站点在内存中停留更长的时间。

Web 搜寻器搜寻速率

您可以监视 Web 搜寻器以查看关于该搜寻器从 Web 站点下载页面的速率的信息。

当您在监视集合时查看关于 Web 搜寻器的详细信息时，可以查看关于搜寻器搜寻数据的速度（搜寻速率）的统计信息。还可以查看有关自从当前会话开始以来搜寻器已搜寻的 URL 数目的统计信息。

搜寻速率是每秒搜寻的页面数。此数目与您可以为 Web 搜寻器配置的数个属性相关：

- 搜寻器线程数
- 活动站点数
- 搜寻器在能够从同一台 Web 服务器检索另一页面之前必须等待的时间量

如果搜寻器的每个搜寻器线程都有一个活动站点，并且搜寻器在能够从同一台 Web 服务器检索另一页面之前必须等待 2 秒，则搜寻器的搜寻速度不能高于每 2 秒钟每个线程 1 个页面。例如，如果搜寻器使用缺省线程数（200），则搜寻器每秒可以为 200 个线程搜寻 100 个页面。

如果活动站点数是搜寻器线程数的两倍，并且搜寻器在能够从同一台 Web 服务器检索另一页面之前必须等待 2 秒钟，则该搜寻器的速度可以达到每秒钟每个线程 1 个页面。但是，网络下载速度和数据库吞吐量可能会变为限制因素。当搜寻速率与搜寻器线程数、活动站点数和搜寻器等待时间协调一致时，表示搜寻器的性能极佳。

在监视 Web 搜寻器性能时，要检查的另一项因素是该搜寻器从当前搜寻器会话开始后已搜寻的 URL 数。将该数目除以搜寻器已运行的时间总量来计算长期吞吐量的平均值。如果此数目没有增加，则表示该搜寻器已完成或者无法继续。例如，网络连接错误、数据库错误和 DNS 查询故障都可能会阻碍搜寻器的进度。

创建 Web 搜寻器报告

通过查看有关过去的 Web 搜寻器活动的报告，可以评估整体性能并根据需要调整 Web 搜寻器属性和搜寻空间定义。

在开始之前

如果您的管理角色限制您只能监视集合，则您可以查看搜寻器统计信息并创建关于搜寻器活动的报告，但是，您不能更改搜寻器的行为（如启动或停止搜寻器）。

关于本任务


不同类型的报告可以为您提供有关 Web 搜寻器活动的信息。对于某些类型的报告，返回信息的速度与从搜寻器的内部数据库收集该信息的速度一样快。但是，创建站点报告和 HTTP 返回码报告就需要花费一些时间。如果创建这些类型的报告，则可以指定用于接收报告的电子邮件地址，而不是等待结果被返回到企业搜索管理控制台。


有关如何解释报告中的统计信息的信息，请在监视 Web 搜寻器和创建报告时单击**帮助**。

过程

要创建 Web 搜寻器报告：

1. 在集合视图中，找到要监视的 Web 搜寻器所在的集合，然后单击  **监视**。
2. 打开搜寻页。

提示：如果您正在编辑集合并且已经打开了搜寻页，则可以单击  **监视**以切换到用于监视搜寻器的视图。

3. 如果要创建报告的 Web 搜寻器正在运行或已暂停，则单击  **详细信息**。
4. 在 Web 搜寻器的详细信息页上，为要创建的报告类型选择一个选项：
 - 在 **搜寻器状态总结** 区域中，单击 **搜寻器历史记录** 以创建关于该搜寻器以及它发现或搜寻的所有站点的报告。
 - 在 **URL 状态** 区域中，指定要创建报告的特定站点的 URL，单击 **查看**，然后单击 **站点历史记录**。
5. 对于搜寻器历史记录报告和站点历史记录报告，选中要在报告中查看的每项统计信息的复选框，然后单击 **查看报告**。

对于这些类型的统计信息，搜寻器将报告返回给管理控制台的速度与它从其内部数据库中检索信息的速度一样快。

6. 如果您正在创建搜寻器历史记录报告，则指定用于创建站点报告的选项，然后单击 **运行报告**。

此报告是使用您选择包括的统计信息创建的并保存在您指定的文件中（文件名必须是绝对名称）。您可以指定在报告被创建之后您想要接收电子邮件。

7. 如果您正在创建搜寻器历史记录报告，则指定用于创建 HTTP 返回码报告的选项，然后单击 **运行报告**。

此报告提供有关每个站点上分布的 HTTP 返回码数目的信息。此报告将保存在您指定的文件中（文件名必须是绝对名称）。您可以指定在报告被创建之后您想要接收电子邮件。

使用此报告来查看哪些站点返回了大量的 4xx 返回码（指示无法找到页面）、5xx 返回码（指示服务器问题）和 6xx 返回码（指示连接问题）等等。

当搜寻器已有一段时间处于活动状态时（例如，搜寻器已有几个星期处于活动状态），此报告最有用。此报告可以帮助您标识消失的站点、最近到达的站点、有大量 URL 的站点（这可能表示对 Lotus Notes 数据库执行了冗余搜寻）以及带有由 HTTP 服务器提供服务的递归文件系统的站点。如果带有大量 HTTP 返回码的站点没有添加至索引，可以通过从搜寻空间中除去这些站点来提高搜寻器的性能。

Web 搜寻器 HTTP 返回码

在监视 Web 搜寻器时，您可以查看关于搜寻器从它尝试搜寻的页面接收到的 HTTP 返回码的信息。

表摘要

在监视 Web 搜寻器历史记录或监视特定 URL 的状态时，您可以查看关于已返回给搜寻器的 HTTP 返回码的信息。可以使用此信息来管理搜寻空间以及优化搜寻器性能。例如，如果对于某个 URL，搜寻器接收到大量的 HTTP 返回码，并且这些返回码指示无法搜寻该位置处的页面，则可以通过从搜寻空间中除去该 URL 来提高性能。

下表列出了 HTTP 返回码以及 Web 搜寻器对它们的解释。从 100 到 505 的值是标准 HTTP 返回码（有关更多信息，请参阅 <http://www.w3.org/Protocols/rfc2616/rfc2616.html>）。其余 HTTP 返回码是企业搜索和 Web 搜寻器的专有返回码。

表 4. 来自 Web 搜寻器的 HTTP 返回码

代码	描述	代码	描述	代码	描述	代码	描述
NULL	未搜寻	401	未经授权	500	内部服务器错误	700	解析错误（没有头结束）
100	继续	402	需要付款	501	未实施	710	解析错误（头）
200	成功	403	已禁止	502	网关错误	720	解析错误（没有 HTTP 代码）
201	已创建	404	无法找到	503	服务不可用	730	解析错误（主体）
202	已接受	405	方法不被允许	504	网关超时	740	被 robots.txt 文件排除
203	非授权信息	406	不可接受	505	不支持 HTTP 版本	741	robots 暂时不可用
204	没有内容	407	需要代理认证	611	读错误	760	被搜寻空间定义排除
205	复位内容	408	请求超时	612	连接错误	770	协议或非标准系统端口错误
206	部分内容	409	冲突	613	读超时	780	被文件类型排除规则排除
300	多个选项	410	已离开	614	SSL 握手失败	2004	没有索引 META 标记
301	已永久移动	411	需要长度	615	其他读错误	3020	软重定向
302	已找到	412	预备条件失败	616	FBA 异常	4044	被 robots.txt 文件排除
303	请参阅其他	413	请求实体太大	617	编码错误		
304	未被修改	414	请求 URI 太长	680	DNS 查找失败		
305	使用代理	415	不受支持的介质类型				
306	（未使用）	417	预期失败				
307	临时重定向						

表注释

4xx 返回码

400（请求错误）代码比较罕见。按照 HTTP 返回码标准，4xx 代码应该表示客户机（搜寻器）失败。但是，问题通常出自服务器或者搜寻器作为链接接收的 URL 中。例如，某些 Web 服务器不允许尝试从站点根向上导航的 URL（例如 `http://xyz.ibm.com/../../sales`）。其他 Web 服务器对这种向上导航就不会有问题，当搜寻器已位于根位置时，那些 Web 服务器将忽略父目录运算符（`..`）。

某些服务器将站点根请求视为错误，某些旧链接可能会请求已不再被识别或实施的操作。当向应用程序服务器请求它不再提供服务的页面时，它将抛出异常，这将导致 Web 服务器返回 HTTP 返回码 400，原因是该请求已不再被认为是有效请求。

- 615** 表示从 Web 站点下载数据的搜寻器服务器遇到了意外的异常。如果有大量这种类型的返回码，则可能表示搜寻器有问题。

6xx 返回码

除 615 以外，6xx 返回码表示在进行搜寻时可以预期的问题，如超时。下列返回码可能要求执行更正操作：

611、612 和 613

表示站点速度较慢或者网络性能不佳。

- 614** 表示搜寻器无法搜寻安全（HTTPS）站点。如果您相信这些站点应该是可以访问的，则验证是否在搜寻器服务器和目标 Web 服务器上正确地设置了证书。例如，如果站点由已识别的认证中心（CA）认证，则可以将新的 CA 添加至搜寻器所使用的信任密钥库。

并且，请了解在您尝试搜寻的站点上是如何配置自签证书的。搜寻器被配置为接受自签证书。某些站点为根 URL（例如 `http://sales.ibm.com/`）创建自签证书，然后尝试对子域（例如 `http://internal.sales.ibm.com/`）使用该证书。搜寻器无法接受以此方式使用的证书。仅当主题的域名（`sales.ibm.com`）和证书签署者与所请求的页面的域名相匹配时，搜寻器才接受自签证书。

- 616** 表示在重新认证后登录表单仍出现在下载中。
- 617** 表示由于编码字符串（字符集）无效或文档包含无效字节，所以不能从文档的字节内容创建字符串。
- 680** 表示搜寻器无法获取搜寻空间中的主机的 IP 地址，这可能是由网络访问问题导致的。此类错误表示搜寻器无法搜寻整个站点，而不仅仅是无法搜寻某些 URL。大量的此类返回码会大大降低吞吐量。

7xx 返回码

7xx 代码主要是由搜寻空间中的规则引起的：

710 - 730

表示发生的问题导致搜寻器无法执行完整的下载，或者搜寻器在站点上遇到了无效的 HTML 数据。如果您看到大量的这些类型的返回码，则请与企业搜索支持代表联系以获取帮助。

740 或 4044

表示无法对文件的内容建立索引，原因是站点的 `robots.txt` 文件中的限制排除了该文档。

740 表示可以将指向已排除文档的锚点链接包括在索引中。

4044 表示还从索引中排除文档中指向已排除文档的锚点链接。

741 表示站点中有允许进行搜寻的 robots.txt 文件，但是下载失败。如果无法搜寻该 URL 的情况重复发生，则会从搜寻空间中除去该 URL。如果您看到大量的此类返回码，请检查目标站点是暂时不可用还是永久不可用。如果目标站点不再可用，则从搜寻空间中除去它。

其余 7xx 返回码主要是在搜寻器运行一阵后您更改搜寻空间时发生的。这些返回码通常并不表示发生了需要您解决的问题。

3020 表示带有返回码 200 的文档包含了一个位置头，该位置头将用户代理转交给另一 URL。

监视解析器

在将解析器分析的文档添加至企业搜索索引前，如果您需要查看有关那些文档的信息，可以监视解析器。提供的选项使您能够复审统计信息以及管理解析器活动。

在开始之前

如果您的管理角色限制您只能监视集合，则您可以查看解析器的状态，但不能启动或停止解析器。

关于本任务


在监视解析器详细信息时，您查看的是解析器活动的快照，该快照提供了有关特定时间点的解析活动的统计信息。这些统计信息显示了已搜寻的文档数和正在解析或正在等待解析的文档数，并显示了已解析并正在等待存储在索引中的文档数。


当解析器处于活动状态时，消息将提供有关解析器状态的其他信息。例如：


- 解析器当前可能正在解析文档。
- 解析器可能处于空闲状态。解析器将休眠到有更多文档可供解析为止。如果发生错误，解析器就会等待重新启动。如果没有解析器服务可用，解析器就会自行重新启动（例如，如果无法与解析器服务建立连接，或者所有解析器 Java 虚拟机都忙于处理其他集合，解析器就会自动重新启动）。
- 解析器可能处于暂停状态（例如，解析器可能会暂停到索引重组完成为止）。

过程

要监视集合的解析器：

1. 在集合视图中，找到要监视的集合，然后单击  **监视**。
2. 打开解析页。

提示：如果您正在编辑集合并且已经打开了解析页，则可以单击  **监视**以切换到用于监视集合的视图。

3. 如果解析器正在运行并且您想要查看有关解析活动的详细状态信息，请单击  **详细信息**。

如果您的管理角色允许您管理集合的进程，则您在查看关于解析活动的详细信息时，可以启动和停止解析器。

4. 如果解析器已停止并且您想要启动它，则单击  **启动**。

在第一次创建集合时，仅在搜寻器开始搜寻数据之后才启动解析器。这能够确保解析器有可以进行分析 and 分类的数据。除非更改了解析规则，否则可以让解析器持续运行。

5. 如果解析器正在运行，而您想停止它，则单击  **停止**。

如果更改了解析规则，就需要停止并重新启动解析器。例如，如果更改了解析器配置，则必须停止并重新启动解析器，这样才能使更改生效。

监视集合的索引活动


当您需要查看正在构建的索引的进度、启用或禁用索引时间表或启动和停止建立索引的活动时，请监视集合的索引。


在开始之前




所有企业搜索管理用户都可以监视索引活动。要启动或停止索引构建，或者要启用或禁用索引时间表，您必须是企业搜索管理员角色的成员、集合的集合管理员或集合的操作员。

过程

要监视集合的索引：

1. 在集合视图中，找到要监视的集合，然后单击  **监视**。
2. 打开索引页。


提示： 如果正在编辑集合并且已经打开了索引页，则可以单击  **监视**以切换到用于监视集合的视图。

3. 如果已经为索引安排了时间表，但您不希望在安排的日期和时间构建该索引，则单击  **禁用时间表**。在启用该时间表或者启动索引构建过程之前，不会构建该索引。
4. 如果已经为索引安排了时间表，但是禁用了用于构建该索引的时间表，则单击  **启用时间表**。该索引将进行排队，以便在索引时间表中指定的日期和时间进行构建。
5. 如果索引已停止并且您想要启动该索引，则单击  **启动**。

通常，建立索引的操作是定期执行的。如果在构建索引时停止了该索引，或者如果禁用了索引的时间表，则可以单击 **启动**以强制索引构建开始。

6. 如果索引构建处于活动状态并且您想要停止该索引构建，则单击  **停止**。

例如，您可能需要停止索引构建以便在更改集合中使用的分类类型之后强制进行索引重组。

7. 如果索引构建期间发生了错误，则单击  **错误**。

这将显示日志文件的内容页，以便您可以查看关于建立索引错误的其他信息。在该页上，可以选择各个错误消息以查看问题详细信息。

相关概念

第 115 页的『企业搜索索引管理』

要确保用户总是能够访问最新的信息，企业搜索为每个集合创建索引并通过定期地刷新内容和重组内容来维护该索引。

监视企业搜索索引队列

您可以查看索引队列中的所有索引构建的状态、停止正在构建的索引或从队列中删除索引。

在开始之前

要管理索引队列，您必须是企业搜索管理员角色的成员。

关于本任务

可以同时构建多个索引，但每个集合每次在队列中只能有一个索引。在为系统配置索引选项时，您指定可以同时共享队列和索引构建资源的索引数目。

过程


要监视索引队列：

1. 单击**系统**以打开系统视图。
2. 选择索引页。


将显示在索引队列中有索引的集合列表。对于每个索引，您可以查看所构建的索引的类型（刷新或重组）、该索引进入索引队列的时间以及开始构建该索引的时间（如果正在进行构建的话）。


3. 要管理各个索引，请单击**状态**图标。

例如，您可能想查看索引接近完成的程度、查看索引中包含的文档数或禁用索引时间表。

4. 要停止正在构建的索引，请单击  **停止**。

例如，如果您更改了分类规则，则您可能想停止索引刷新，以便可以强制启动索引重组。

要在停止索引构建后开始进行索引构建，您既可以等待该索引在它的下一个安排的开始时间进入索引队列，也可以单击**状态**图标以监视该索引，然后单击  **启动**以刷新或重组该索引。

5. 要从索引队列中除去索引，请单击  **除去**。

相关概念

第 115 页的『企业搜索索引管理』

要确保用户总是能够访问最新的信息，企业搜索为每个集合创建索引并通过定期地刷新内容和重组内容来维护该索引。

监视搜索服务器

您既可以查看特定集合的搜索服务器活动的详细状态信息，也可以查看整个企业搜索系统内的搜索服务器的详细状态信息。


在开始之前


所有企业搜索管理用户都可以监视他们有权管理的集合的搜索服务器。要监视企业搜索系统中的所有搜索服务器，您必须是企业搜索管理员角色的成员。

要启动或停止搜索服务器，您必须是企业搜索管理员角色的成员、集合的集合管理员或者集合的操作员。

过程

1. 要监视单个集合的搜索服务器:

- a. 在集合视图中，找到要监视的集合，然后单击  **监视**。
- b. 打开搜索页。

提示: 如果您正在编辑集合并且已经打开了搜索页，则可以单击  **监视**以切换到用于监视集合的视图。

2. 要监视企业搜索系统中的所有搜索服务器:

- a. 单击 **系统**以打开系统视图。
- b. 选择搜索页。

3. 如果搜索服务器已停止并且您想要启动它，则单击 **启动**。

4. 如果搜索服务器正在运行，而您想停止它，则单击 **停止**。

如果启用或禁用了搜索高速缓存、更改了搜索高速缓存大小或更改了快速链接，就必须停止并重新启动搜索服务器才能使更改生效。

5. 要查看搜索服务器在处理搜索请求时花费的时间的总结，请单击**响应时间历史**。

报告将显示搜索服务器在响应特定日期的搜索请求时耗用的平均时间量（以毫秒计）。

平均响应时间用来衡量系统的执行性能，它与服务质量相对应。响应时间的增加可能意味着系统负载过重。例如，正在搜索的集合的数目和集合大小可能使系统负载过重。

6. 要查看最频繁提交的查询的列表，请单击**常用的查询**。

该报告显示 50 个最频繁提交的查询中的关键字以及用户提交特定查询的次数。

通过复审最频繁的查询，您可以标识快速链接候选。通过创建快速链接，您可以提高许多用户的搜索质量。您可以确保总是在搜索结果中返回高度相关的文档。

您可能还想创建链接以指向回答那些来自企业门户网站的查询的资源。例如，如果用户频繁搜索有关开支帐户的信息，则在内部网首页上包括一个指向讨论开支帐户过程的页面的链接。

7. 要查看最近提交的查询的列表，请单击**最近的查询**。

该报告显示 50 个最近提交的查询中的关键字。

通过复审最近的查询，您可以标识组织中的当前趋势和紧急情况。例如，您可能会看到对某个主题的广泛兴趣。这种广泛的兴趣可能表示该主题需要一个快速链接，或者您需要以其他方式将该主题提供给用户（如在企业门户网站上提供链接）。

监视数据侦听器


监视数据侦听器以查看其状态和查看有关客户机数据侦听器应用程序活动的详细信息。

在开始之前

要监视数据侦听器，您必须是企业搜索管理员角色的成员。

过程

要监视数据侦听器：

1. 单击**系统**以打开系统视图。
2. 在数据侦听器页上，查看状态图标以了解数据侦听器是处于活动状态还是停止状态。
3. 如果数据侦听器正在运行，并且您想查看有关客户机应用程序活动的详细状态信息，请单击  **详细信息**。

数据侦听器详细信息页上的状态图标指示了数据侦听器是正在运行还是已停止。统计信息显示了正在等待处理的请求的数目、正在处理客户机应用程序请求的每个线程的当前状态以及处于指定线程状态的活动线程的数目。

4. 如果更改了数据侦听器的端口号，则单击  **重新启动**。

数据侦听器是在企业搜索系统启动时启动的。除非更改了数据侦听器的端口号，否则不需要重新启动它。

相关任务

第 80 页的『配置对数据侦听器应用程序的支持』

通过使用数据侦听器 API 来创建外部搜寻器，可以扩展企业搜索。定制数据侦听器应用程序可以对集合添加数据、从集合中除去数据或者指示 Web 搜寻器访问和重新访问 URL。

文档跟踪

在各个处理阶段都可能会从系统中废弃文档。您可以指定选项以了解废弃文档的时间以及导致废弃该文档的问题。

如果解析器遇到错误，从而导致无法解析文档，就会记录关于所废弃的文档的消息，该消息包含原因码。（此类错误不会导致从索引中除去旧版本的文档。）

在建立索引阶段可能会废弃文档，在这种情况下也会记录此信息。例如，可以显式地删除 URI 和 URI 模式。文档可能是后来被删除的搜寻器搜寻到的。源文档可能不再存在（负数 HTTP 代码与文档相关联），或者与文档相关联的 HTTP 代码可能是未知的。如果需要全局分析的文档缺少排名信息，也可能会废弃文档。

如果您知道已搜寻某个文档，但该文档未出现在索引中，则可使用企业搜索管理控制台来跟踪系统中的文档流。详细报告会指出废弃文档的时间、位置以及原因。例如，详细报告可能会指出该文档是在全局分析期间被意外废弃的，该报告也可能指出管理员从索引中除去了 URI。

相关任务

第 201 页的『查看关于 URI 的详细信息』

您可以查看有关 URI 的详细信息。您可以查看关于由此 URI 表示的文档是如何被搜寻、被建立索引和被搜索的当前信息和历史信息。

配置用于进行文档跟踪的日志文件

为了确定何时、从何位置以及为何从系统中废弃了文档，您可以配置日志文件以跟踪有关所废弃的文档的信息。

在开始之前

要配置用于跟踪所废弃的文档的选项，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

关于本任务

为了防止日志文件消耗太多的磁盘空间，系统以循环方式使用日志文件，并且，每当当前日期更改时，始终启动一个新的日志文件。如果一个日志文件增大到其最大可允许大小，并且日期未更改，则系统将创建一个新的日志文件。当达到最大日志文件数时，将废弃最旧的日志文件，以便可以创建新的日志文件。

过程

要配置用于进行文档跟踪的日志文件：

1. 编辑集合，选择日志页，然后单击**配置文档跟踪**。
2. 在文档跟踪页上，确保选中了用于进行文档跟踪的复选框。
3. 指定日志文件数，这些日志文件将用来记录关于从系统中废弃的文档的信息。这些日志文件由所有废弃了文档的会话所共享。
4. 单击**确定**。

查看关于废弃的文档的报告

您可以查看关于已从企业搜索系统中废弃的文档的详细信息。仅当对集合启用了文档跟踪功能时，才会提供此信息。

在开始之前

在请求查看关于已废弃的文档的报告或者请求将报告发送到某个电子邮件地址之前，请确保要从中接收信息的会话处于活动状态。例如，要了解解析或建立索引期间废弃的文档的信息，请确保集合的解析器会话和索引会话已启动。

在可以接收报告之前，请确保为企业搜索配置了邮件服务器信息。您在系统视图的日志页上配置电子邮件选项时指定此信息。

关于本任务


搜集有关已废弃的文档的信息是一个相当耗时的过程。您可以选择一个选项以查看信息并等待该信息显示。更有效的选项是将报告发送到您指定的电子邮件地址。


如果废弃了某个文档，报告就会显示废弃该文档的日期和时间、错误的严重性级别、发生问题的组件和会话以及错误消息。

过程

要查看关于已废弃的文档的详细信息：

1. 在集合视图中，找到要监视的集合，然后单击  **监视**。

提示：如果您正在编辑集合并且已经打开了常规页，则可以单击  **监视**以切换到用于监视集合的视图。

2. 单击  **URI 详细信息**。
3. 在 URI 详细信息页上，输入要查看其信息的 URI。
4. 选中要查看的信息类型的复选框：

解析器废弃的文档

选中此复选框以查看先前解析某个文档时是否已将其废弃以及废弃该文档的原因。

从索引中废弃的文档

选中此复选框以查看先前对某个文档建立索引或进行分析时是否已将其废弃以及废弃该文档的原因。

5. 指定查看报告的方式：
 - 要等待报告显示，请单击**查看报告**。
 - 要将报告发送到电子邮件地址以便可以在以后查看，请单击**发送报告**。

在发送详细的 URI 报告页上，在**要通知的电子邮件地址**字段中输入用于接收报告的电子邮件地址，然后单击**发送报告**。

相关任务

第 201 页的『查看关于 URI 的详细信息』

您可以查看有关 URI 的详细信息。您可以查看关于由此 URI 表示的文档是如何被搜寻、被建立索引和被搜索的当前信息和历史信息。

相关参考

第 82 页的『企业搜索索引中的 URI 格式』

企业搜索索引中的每个文档的统一资源标识（URI）都指示了将文档添加到该集合的搜寻器的类型。

企业搜索日志文件和警报

您可以选择要为集合以及为系统记录的消息的类型、指定有关创建和查看日志文件的选项以及指定有关接收警报和接收关于消息的电子邮件的选项。

在正常操作期间，企业搜索组件将日志消息写至公共日志文件。此日志文件位于索引服务器上的 `ES_NODE_ROOT/logs` 目录中。您可以使用管理控制台来查看这个公共日志数据。

如果发生问题，如网络通信故障，则组件将日志消息写至组件安装所在的服务器上的 `logs` 目录。要查看这些本地日志文件，请使用该计算机上的文件查看器，如 UNIX 系统上的 `tail` 实用程序。不能使用管理控制台来查看这些类型的日志文件。

在配置日志文件时，可以选择要记录的消息的类型（如错误消息或警告消息）、指定废弃旧日志文件以便为新日志文件腾出空间的频率、指定日志文件的最大大小以及选择消息的语言。您还可以指定一些选项以便记录每当某些事件、某些消息或某些类型的消息时接收到电子邮件。

在监视日志文件时，您可以选择要打开的日志文件。可以对日志文件的内容进行过滤以便仅查看具有特定严重性级别的消息（如仅查看错误消息）或由特定企业搜索会话生成的消息。在查看日志文件时，您可以查看关于各个消息的详细信息。例如，您可能想查看生成该消息的函数名以及其他可以帮助您执行更正操作（如果有必要的话）的信息。

相关概念

“消息参考”中的“企业搜索消息”

警报

可以将企业搜索配置为每当它检测到发生某些事件时将消息写入日志文件。

由事件触发的消息（称为警报）通知您可能需要解决的问题，如耗尽可用空间的资源。当为企业搜索配置警报时，您指定要让系统监视的情况。每当该情况发生时，系统就自动地将消息写入日志文件。

如果您希望直接获得有关情况的通知，则可以指定一些选项以便在记录其中一条被监视消息时接收到电子邮件。

可以为集合级事件以及为发生在系统级的事件配置警报。在集合级，系统可以：

- 监视每个搜寻器所搜寻的文档数目，并在将要达到所允许的最大文档数时发出警报消息。
- 监视正被添加至集合索引的文档数目，并在将要达到所允许的最大文档数时发出警报消息。
- 当响应搜索请求所需的时间超出指定的限制时通知您。

在系统级，系统可以监视每台企业搜索服务器上的磁盘空间，并在可用空间量较低时发出警报消息。

配置集合级警报

通过配置警报，您可以确保每当发生某些集合级事件时都将消息写入日志文件。您还可以在每当记录关于这些事件的消息时接收到电子邮件。

在开始之前

要为集合配置警报，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。

过程

要配置集合级警报：

1. 编辑集合，选择日志页，然后单击**配置警报**。
2. 如果要让系统监视每个搜寻器正在搜寻的文档数，则执行下列步骤：
 - a. 选中**当任何搜寻器搜寻到的文档数达到所允许的最大数目的某个百分比时复选框**。
 - b. 在**百分比**字段中，指定您希望何时记录消息。将此数值作为搜寻器所能够搜寻的最大文档数的百分比指定（在配置搜寻器属性时指定**要搜寻的最大文档数**）。缺省值是 90%。

由于可以为不同的搜寻器配置不同的限制，所以将为每个搜寻器记录单独的消息。例如，如果使用缺省警报阈值，允许 DB2 搜寻器搜寻 2 000 000 个文档，并允许 Notes 搜寻器搜寻 1 000 000 个文档，则当 DB2 搜寻器搜寻了 1 800 000 个文档时将记录一条消息，当 Notes 搜寻器搜寻了 900 000 个文档时将记录另一条消息。

3. 如果要让系统监视正被添加到索引中的文档数，则执行下列步骤：
 - a. 选中**当集合中的文档数达到估算大小的某个百分比时复选框**。
 - b. 在**百分比**字段中，指定您希望何时记录消息。将此数值作为您期望该集合包含的估算文档数的百分比指定。缺省值是 85%。

限制字段显示了该集合的当前估算大小。要更改此值，请打开该集合的常规页，选择选项以配置常规选项，并在**估算的文档数**字段中指定新的值。

警告： 此限制以及您为集合配置的估算文档数仅用于监视集合的增大情况。它们不会对索引可以达到的大小施加绝对限制。

4. 如果要让系统在响应搜索请求所需的时间超出限制时通知您，则执行下列步骤：
 - a. 选中**当搜索响应时间超出限制时复选框**。
 - b. 在**限制**字段时，输入您认为可以接受的秒数作为最大搜索响应时间。

当超出此数目时，系统将写关于此事件的日志消息。例如，如果您保留了缺省值，则每当搜索服务器平均需要 5 秒钟或更长的时间来响应搜索请求时，系统将创建一条日志消息。

典型的响应时间是小于半秒。大于 1 秒的平均时间可能表示需要对操作系统进行调整以提高性能或者搜索服务器配置存在问题。例如，您可能想增加为搜索高速缓存分配的空间量。

5. 单击**确定**。

如果您想要在系统记录关于这些事件的消息时接收电子邮件，则打开日志页，然后单击**配置消息的电子邮件选项**，以便可以指定电子邮件地址。您启用的警报的消息标识将自动被添加到要为其发送电子邮件的消息标识列表中。

在可以接收电子邮件之前，还必须确保配置了邮件服务器信息。要完成此操作，企业搜索管理员必须在工具栏上选择**系统**，打开日志页，然后单击**配置消息的电子邮件选项**。

相关任务

第 221 页的『接收关于已记录消息的电子邮件』

您可以指定一些选项以便每当记录某些消息或某些类型的消息时接收到电子邮件。

配置系统级警报


通过配置警报，您可以确保每当发生某些系统级事件时都将消息写入日志文件。您还可以在每当记录关于这些事件的消息时接收到电子邮件。

在开始之前

要配置系统级警报，您必须是企业搜索管理员。

过程

要配置系统级警报：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在日志页上，单击**配置警报**。
4. 如果要想系统监视企业搜索服务器上的可用空间容量，则选中**当可用文件系统空间的容量达到总空间的某个百分比时复选框**。
5. 在**百分比**字段中，指定要想系统在什么时候通知您服务器上的可用空间容量较低。以文件系统空间总量的百分比形式指定此数值。缺省值是 80%。

如果企业搜索系统设置在多台服务器上，则系统将为每台服务器创建各自的日志消息。例如，当搜寻器服务器上的空间量较低时，将会有一条消息把这个情况通知您；并且还会有单独的消息通知您有关索引服务器和搜索服务器上的空间约束情况。

6. 单击**确定**。

如果您想要在系统记录关于此事件的消息时接收电子邮件，请打开日志页，然后单击**配置消息的电子邮件选项**以便可以指定电子邮件地址和邮件服务器信息。

相关任务

第 221 页的『接收关于已记录消息的电子邮件』

您可以指定一些选项以便每当记录某些消息或某些类型的消息时接收到电子邮件。

配置日志文件

您可以指定要记录的消息的类型并指定选项以指示如何创建日志文件。

在开始之前

要配置集合级日志文件，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。要配置系统级日志文件，您必须是企业搜索管理员。



关于本任务

为了防止日志文件消耗太多的磁盘空间，系统以循环方式使用日志文件，并且，每当当前日期更改时，始终启动一个新的日志文件。如果一个日志文件增大到其最大可允许大小，并且日期未更改，则系统将创建一个新的日志文件。当达到最大日志文件数时，将废弃最旧的日志文件，以便可以创建新的日志文件。

要接收关于已记录的消息的电子邮件，首先指定关于电子邮件的传递方式的信息。然后，指定要对哪些消息接收电子邮件。

过程

要配置企业搜索日志文件：

1. 如果要配置用于创建和循环使用系统级日志文件的选项：
 - a. 单击**系统**以打开系统视图。
 - b. 单击  **编辑**以切换到系统编辑视图。
 - c. 在日志页上，单击**配置日志文件选项**。这就显示了系统级的日志文件选项页。
2. 如果要配置用于创建和循环使用集合级日志文件的选项：
 - a. 在集合视图中，找到要为其指定选项的集合，然后单击  **编辑**。
 - b. 在日志页上，单击**配置日志文件选项**。这就显示了集合级日志文件选项页。
3. 在**要记录的信息的类型**字段中，选择要记录的消息的类型：

仅错误消息

错误消息指示已发生不期望的情况或意外的行为，并且进程无法继续。您必须执行操作来更正问题。

错误消息和警告消息

警告消息指示可能存在冲突或不一致性，但它们不会导致进程停止。此选项是缺省选项。

所有消息

参考消息提供有关系统或当前任务的一般信息，并且不需要执行任何更正操作。

4. 在**每个日志文件的最大大小**字段中，输入每个日志文件的最大兆字节数。缺省值是 5MB。

当日志文件增大到此大小时，将创建新的日志文件，直到达到您允许的最大日志文件数为止。通过使日志文件相对较小，您可以更有效地查看它们。

5. 在**最大日志文件数**字段中，输入要创建的日志文件的最大数目。缺省值是 10。

如果想要确保能够复审较旧的日志消息，则增大此值。如果您对最近的消息更感兴趣，并且不需要维护很长的活动历史，则减小此值。

6. 在**缺省语言环境**字段中，选择要用来记录消息的语言。缺省值是英语。
7. 单击**确定**。

配置 SMTP 服务器信息

在您可以接收关于企业搜索活动的电子邮件之前，您必须配置有关简单电子邮件传输协议（SMTP）服务器的信息。

在开始之前

要配置有关 SMTP 服务器的信息，您必须是企业搜索管理员角色的成员。


关于本任务

有好几项企业搜索管理功能允许您接收电子邮件。在您可以接收来自任何这些功能的电子邮件之前，您必须指定有关 SMTP 服务器的信息：

- 如果您配置集合级警报或系统级警报，则可以在记录那些消息时接收到电子邮件。您也可以在记录其他消息（而不仅仅是由被监视的事件触发的消息）时接收到电子邮件。
- 如果要查看关于索引中的 URI 的详细信息或者关于已从企业搜索系统中废弃的文档的详细信息，则可以通过电子邮件接收报告。
- 如果您监视 Web 搜寻器并指定您想要创建 Web 搜寻器历史报告，则可以在报告被创建时接收电子邮件通知。

过程

要配置有关 SMTP 服务器的信息：

1. 单击**系统**以打开系统视图。
2. 单击  **编辑**以切换到系统编辑视图。
3. 在日志页上，单击**配置消息的电子邮件选项**。
4. 在系统消息的电子邮件选项页上，在**用于传递电子邮件的 SMTP 邮件服务器**字段中，输入要使用的 SMTP 服务器的标准主机名或 IP 地址。

系统将使用此服务器来将电子邮件发送至您指定的地址。

5. 在**检查电子邮件的频率**字段中，指定要让系统以何种频率检查符合条件的消息并发送关于那些消息的电子邮件。

系统将特定电子邮件地址的所有消息组合到一条消息中并按指定的频率发送该消息。

6. 单击**确定**。

接收关于已记录消息的电子邮件

您可以指定一些选项以便每当记录某些消息或某些类型的消息时接收到电子邮件。

在开始之前

要为系统级消息配置电子邮件选项，您必须是企业搜索管理员角色的成员。要为集合级消息配置电子邮件选项，您必须是企业搜索管理员角色的成员或者是集合的集合管理员。


在您可以接收电子邮件之前，您首先必须配置有关简单电子邮件传输协议（SMTP）服务器的信息以便可以传递电子邮件。

关于本任务

在配置警报时，可以选择一个选项以便在发生某些事件时记录消息。如果启用那些选项，您就可以配置一些选项以便每当记录那些消息时自动接收到电子邮件。您还可以指定一些选项以便在记录其他消息（而不仅仅是由事件触发的消息）时接收到电子邮件。

过程

要配置消息的电子邮件选项:

1. 如果您想要接收有关系统消息的电子邮件:
 - a. 单击**系统**以打开系统视图。
 - b. 单击  **编辑**以切换到系统编辑视图。
 - c. 在日志页上，单击**配置消息的电子邮件选项**。
 - d. 在系统消息的电子邮件选项页上，选中**发送关于系统级消息的电子邮件**复选框。
 - e. 在**用于接收电子邮件的电子邮件地址**字段中，输入一个或多个电子邮件地址。通常，企业搜索管理员应该接收关于系统消息的信息。


用逗号来分隔每个地址。例如:

`steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com,`

- f. 如果要接收关于所记录的所有错误消息的电子邮件，请选中**发送关于所有错误消息的电子邮件**复选框。
- g. 如果您希望仅当记录某些系统级消息时才接收电子邮件，请在**发送关于某些消息的电子邮件**区域中输入那些消息的消息标识。每行输入一个消息标识。例如:

`FFQC4819E
FFQ00005E`

缺省情况下，已列出了几个消息标识（单击**帮助**获取这些消息的描述）。

- h. 单击**确定**。
2. 如果您想要接收关于集合消息的电子邮件:
 - a. 单击**集合**以打开集合视图。
 - b. 在集合列表中，找到要配置的集合并单击  **编辑**。
 - c. 在日志页上，单击**配置消息的电子邮件选项**。
 - d. 在集合消息的电子邮件选项页上，选中**发送关于集合级消息的电子邮件**复选框。
 - e. 在**用于接收电子邮件的电子邮件地址**字段中，输入一个或多个电子邮件地址。通常，集合管理员应该接收关于集合级消息的信息。

用逗号来分隔每个地址。例如:

`steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com,`

- f. 如果要接收关于所记录的所有错误消息的电子邮件，请选中**发送关于所有错误消息的电子邮件**复选框。
- g. 如果您希望仅当记录某些集合级消息时才接收电子邮件，请在**发送关于某些消息的电子邮件**区域中输入那些消息的消息标识。每行输入一个消息标识。例如:

FFQC4819E
FFQ00005E

缺省情况下，已列出了几个消息标识（单击[帮助](#)获取这些消息的描述）。

h. 单击**确定**。

相关概念

“消息参考”中的“企业搜索消息”

相关任务

第 218 页的『配置集合级警报』

通过配置警报，您可以确保每当发生某些集合级事件时都将消息写入日志文件。您还可以在每当记录关于这些事件的消息时接收到电子邮件。

第 219 页的『配置系统级警报』

通过配置警报，您可以确保每当发生某些系统级事件时都将消息写入日志文件。您还可以在每当记录关于这些事件的消息时接收到电子邮件。


查看日志文件


您可以查看系统组件和集合组件写入公共日志文件中的日志消息。您还可以指定过滤器以查看特定严重性级别的消息以及来自特定企业搜索会话的消息。

在开始之前

所有企业搜索管理用户都可以查看他们有权管理的集合的日志文件。要查看系统级日志文件，您必须是企业搜索管理员角色的成员或者有权访问**系统**工具栏。

过程

1. 要查看单个集合的日志文件：
 - a. 单击**集合**以打开集合视图。
 - b. 在集合列表中，找到要查看的集合，单击 **监视**，然后打开日志页。

提示：如果您正在编辑集合并且已经打开了日志页，则可以单击  **监视**以切换到用于监视集合的视图。

2. 要查看系统级日志文件：
 - a. 单击**系统**以打开系统视图。
 - b. 选择日志页。
3. 在**日志文件**字段中，选择您想要查看的日志文件。


每个日志文件名都包含日志文件类型（如系统或集合标识）、该文件的创建日期以及一个数字后缀，该后缀指示了在该日期创建文件的顺序。例如：

```
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. 如果您只想查看特定严重性级别的消息，则选中**严重性**字段中的适当复选框。
5. 如果您只想查看来自特定会话的消息，则选中**会话**字段中的适当复选框。
6. 单击**查看日志**。

对于日志文件的内容页上的每条消息，您可以查看该消息的发出日期和时间、消息严重性级别、发出该消息的会话名以及消息标识和错误文本。

可以单击按钮以转到日志文件的第一页、最后一页、上一页或下一页。也可以指定页号并直接转到该页。

7. 要查看有关某条消息的更详细信息，请单击  **详细信息**。

在日志消息详细信息页上，您可以查看发出该消息的企业搜索服务器的主机名、生成该错误的文件名、发生该错误的函数名和行号、进程标识和线程标识。

备份和复原企业搜索系统

备份和复原脚本使您能够备份和复原企业搜索系统。

如果系统由于不可恢复的错误而失败，则必须重新安装 WebSphere Information Integrator OmniFind Edition 并接着运行复原脚本。还可以使用这些脚本来将必需的系统文件复原到一台或多台新服务器上。

这些脚本备份和复原下列文件：

- ES_NODE_ROOT/master_config 目录中的配置文件。
- 搜寻器的数据库文件。
- 索引文件。如果企业搜索索引不在 ES_NODE_ROOT/data 目录中，就无法使用企业搜索脚本来备份和复原索引文件。

备份脚本将在运行该脚本时指定的目录中创建下列子目录。（企业搜索管理员标识必须有权写指定的目录。）

master_config

包含 ES_NODE_ROOT/master_config 目录中的配置文件

database

包含来自搜寻器服务器的数据库文件

data

包含来自索引服务器的索引文件

必须有足够的磁盘空间可用来将企业搜索系统文件备份到另一个目录中。备份和复原脚本不检查文件。启动备份时，大部分系统会话会暂时不可用。搜索进程将继续运行。在重组索引之后，应该启动备份，以便有最新的索引。

在多服务器安装中，请从企业搜索索引服务器备份和复原系统。由于索引服务器创建数据库目录，所以索引服务器可以访问和备份搜寻器服务器上的产品数据库表。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

备份企业搜索系统

可以使用 esbackup.sh 脚本（对于 UNIX）或 esbackup.bat 脚本（对于 Microsoft Windows）来备份企业搜索系统。

限制

当备份和复原脚本运行时，所有系统会话都将停止。为了避免看到不正确或不一致的系统信息，在运行脚本时请不要使用企业搜索管理控制台。

企业搜索管理员标识必须有权写在运行备份脚本时指定的目录。

过程

要备份企业搜索系统:

1. 在索引服务器上, 作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 输入以下命令, 其中 *backup_directory* 是用来备份文件的目录:

```
UNIX: esbackup.sh -c -d -i backup_directory
```

```
Windows: esbackup.bat -c -d -i backup_directory
```

可以指定下列选项:

- c 备份配置文件。这是缺省选项。
- d 备份搜寻器数据商店中已搜寻到的文档 (尚未对这些文档进行解析或建立索引)。
- i 备份索引文件。如果索引文件不在 ES_NODE_ROOT/data 目录中, 就无法使用 esbackup 脚本来备份索引文件。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

复原企业搜索系统

在重新安装 WebSphere II OmniFind Edition 后, 可以使用 esrestore.sh 脚本 (对于 UNIX) 或 esrestore.bat 脚本 (对于 Microsoft Windows) 来复原系统配置文件。

限制

当备份和复原脚本运行时, 所有系统会话都将停止。为了避免看到不正确或不一致的系统信息, 在运行脚本时请不要使用企业搜索管理控制台。

过程

要复原企业搜索系统:

1. 在索引服务器上, 作为企业搜索管理员登录。此用户标识是在安装 WebSphere II OmniFind Edition 时指定的。
2. 如果尚未启动公共通信层 (CCL), 则启动它:

```
UNIX: startccl.sh -bg
```

```
Windows: startccl.bat
```

在 Windows 系统上, 要在后台启动 CCL, 请单击开始 → 程序 → 管理工具 → 服务, 然后重新启动 WebSphere Information Integrator OmniFind Edition 服务。

3. 如果控制器尚未停止, 则停止它:

```
esadmin stop
```

4. 输入以下命令, 其中 *backup_directory* 是备份文件时使用的目录:

```
UNIX: esrestore.sh -c -d -i backup_directory
```

```
Windows: esrestore.bat -c -d -i backup_directory
```

可以指定下列选项:

- c** 复原配置文件。这是缺省选项。
- d** 复原搜寻器数据商店中已搜寻的文档。(尚未对这些文档进行解析或建立索引。)
- i** 复原索引文件。如果索引文件不在 `ES_NODE_ROOT/data` 目录中, 就无法使用 `esrestore` 脚本来复原索引文件。

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

将企业搜索系统文件复原至新服务器

可以从一台企业搜索服务器备份系统文件并将系统文件复原至一台或多台服务器, 而不是将它们复原至当前服务器。

在开始之前

在运行复原脚本之前, 必须在新服务器上安装 `WebSphere II OmniFind Edition`。

限制

当备份和复原脚本运行时, 所有系统会话都将停止。为了避免看到不正确或不一致的系统信息, 在运行脚本时请不要使用企业搜索管理控制台。

企业搜索管理员标识必须有权写在运行备份脚本时指定的目录。

关于本任务

存储在 `ES_NODE_ROOT/master_config/nodes.ini` 文件中的服务器信息未包括在备份文件中。

过程

要将企业搜索系统文件复原至一台或多台新服务器:

1. 在索引服务器上, 作为企业搜索管理员登录。此用户标识是在安装 `WebSphere II OmniFind Edition` 时指定的。
2. 在当前索引服务器(旧索引服务器)上运行备份脚本, 其中 `backup_directory` 是用来备份文件的目录:

UNIX: `esbackup.sh backup_directory`

Windows: `esbackup.bat backup_directory`

3. 使用 FTP 程序来将 `backup_directory` 中的所有文件发送至新的索引服务器。
4. 在新的索引服务器上运行复原脚本:

UNIX: `esrestore.sh backup_directory`

Windows: `esrestore.bat backup_directory`

相关参考

第 229 页的『企业搜索命令、返回码和会话标识』
可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

企业搜索命令、返回码和会话标识

可以使用命令来诊断问题、确定系统不同部分的状态、启动和停止会话或者启动和停止系统。

在多服务器安装中，可以从系统中的任何服务器运行命令。但是，应该从索引服务器运行命令。索引服务器（即控制器服务器）可以访问系统中所有其他服务器上的信息。

大部分命令具有以下格式：

```
esadmin command_name arguments  
esadmin session_ID action -option
```

要了解有关所有命令的更多信息，请输入 `esadmin help`。要了解有关特定命令的更多信息，请输入 `esadmin action help`。

企业搜索 `esadmin` 命令

在一行上输入下列命令。

表 5. 企业搜索 `esadmin` 命令

命令	描述
<code>esadmin <i>crawler_session_id</i> start</code>	启动搜寻器会话。此命令不会启动任何搜寻活动。 样本命令： <code>esadmin col1.WEB1.esadmin start</code> 样本消息和返回码： FFQC5310I WEB1Crawler1 (sid: col1.WEB1.esadmin) 未在运行中。 FFQC5314I 结果: 0
<code>esadmin <i>crawler_session_id</i> startCrawl</code>	开始搜寻。 样本命令： <code>esadmin col3.DB21.esadmin startCrawl</code> 样本消息和返回码： FFQC5303I DB21Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650 FFQC5314I 结果: 0
<code>esadmin <i>crawler_session_id</i> pause</code>	暂停搜寻。 样本命令： <code>esadmin col3.DB21.esadmin pause</code> 样本消息和返回码： FFQC5303I DB21Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650 FFQC5314I 结果: 0

表 5. 企业搜索 *esadmin* 命令 (续)

命令	描述
<code>esadmin crawler_session_id resume</code>	<p>继续搜寻。</p> <p>样本命令:</p> <pre>esadmin col3.DB21.esadmin resume</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650 FFQC5314I 结果: 0</pre>
<code>esadmin crawler_session_id stopCrawl</code>	<p>停止搜寻。</p> <p>样本命令:</p> <pre>esadmin col3.DB21.esadmin stopCrawl</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650 FFQC5314I 结果: 0</pre>
<code>esadmin crawler_session_id stop</code>	<p>停止搜寻器会话。</p> <p>样本命令:</p> <pre>esadmin col3.DB21.esadmin stop</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650 FFQC5314I 结果: 0</pre>
<code>esadmin crawler_session_id getCrawlerStatus</code>	<p>获取搜寻器状态。返回的信息取决于搜寻器是 Web 搜寻器还是所有其他数据源的搜寻器。</p> <p>Web 搜寻器的示例:</p> <pre>esadmin col1.WEB1.esadmin getCrawlerStatus</pre> <p>Web 搜寻器的可能返回码和消息:</p> <pre>FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) 已经在运行中。PID: 23650</pre> <p>非 Web 搜寻器的示例:</p> <pre>esadmin col3.DB21.esadmin getCrawlerStatus</pre> <p>非 Web 搜寻器的可能返回码和消息:</p> <pre>FFQC5303I db2crawler (sid: db2col.DB2_96945) 已经在运行中。PID: 5936</pre> <p>要了解关于返回的状态消息的更多信息，请参阅第 234 页的『状态命令的详细信息』。</p>

表 5. 企业搜索 *esadmin* 命令 (续)

命令	描述
<pre>esadmin dscrawler_session_id getCrawlSpaceStatus</pre>	<p>获取除 Web 搜寻器以外的任何搜寻器的常规搜寻空间状态。</p> <p>样本命令:</p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatus</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650</pre> <p>获取 Web 搜寻器的常规搜寻空间状态。</p> <p>样本命令:</p> <pre>esadmin col11.WEB1.esadmin getCrawlStatus</pre> <p>要了解关于返回的状态消息的更多信息, 请参阅第 234 页的『状态命令的详细信息』。</p>
<pre>esadmin dscrawler_session_id getCrawlSpaceStatusDetail -ts target_server_id</pre> <pre>esadmin webcrawler_session_id getCrawlDetailsPerSite -url string selections num -threshold num</pre>	<p>获取除 Web 搜寻器以外的任何搜寻器的详细搜寻空间状态。如果未指定目标服务器选项, 则将返回所有目标服务器的数据。例如, 如果 DB2 搜寻器搜寻 FOUNTAIN 和 SAMPLE 数据库并且您未指定目标服务器选项, 则将返回 FOUNTAIN 和 SAMPLE 数据库中所有表的状态。</p> <p>样本命令:</p> <pre>esadmin col3.DB21.esadmin getCrawlSpaceStatusDetail -ts FOUNTAIN</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) 已经在运行中。PID: 23650</pre> <p>获取 Web 搜寻器的详细搜寻空间状态。</p> <p>样本命令:</p> <pre>esadmin col11.WEB1.esadmin getCrawlDetailsPerSite</pre> <p>要了解关于返回的状态消息的更多信息, 请参阅第 234 页的『状态命令的详细信息』。</p>
<pre>esadmin startParser -cid collection_ID</pre>	<p>启动解析器。</p> <p>样本命令:</p> <pre>esadmin startParser -cid col1</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (server1) (sid: controller) 已经在运行中。PID: 25917 FFQC5314I 结果: 0</pre>

表 5. 企业搜索 *esadmin* 命令 (续)

命令	描述
<i>esadmin stopParser collection_id</i>	<p>停止解析器。</p> <p>样本命令:</p> <pre>esadmin stopParser -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (node1) (sid: controller) 已经在运行中。PID: 15292 FFQC5312E 执行操作时出错: [stopParser] FFQC4823E 会话解析器驱动程序 - 集合 coll (node1) [coll.parserdriver] 未在运行中。</pre>
<i>esadmin monitor getCollectionParserMonitorStatus -cid collection_ID</i>	<p>获取解析器状态。</p> <p>样本命令:</p> <pre>esadmin monitor getCollectionParserMonitorStatus -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 监视器 (node1) (sid: monitor) 已经在运行中。PID: 12543</pre> <p>要了解关于返回的状态消息的更多信息，请参阅第 234 页的『状态命令的详细信息』。</p>
<i>esadmin startMain -cid collection_id</i>	<p>开始重组索引。</p> <p>样本命令:</p> <pre>esadmin startMain -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (node1) (sid: controller) 已经在运行中。PID: 25917 FFQC5314I 结果: 1117671147056</pre>
<i>esadmin startDelta -cid collection_id</i>	<p>开始刷新索引。</p> <p>样本命令:</p> <pre>esadmin startDelta -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (node1) (sid: controller) 已经在运行中。PID: 4548 FFQC5314I 结果: 1117670603408</pre>

表 5. 企业搜索 *esadmin* 命令 (续)

命令	描述
<pre>esadmin monitor getCollectionIndexMonitorStatus -cid <i>collection_id</i> -buildType [main delta] -numrecords <i>lastNrecords</i></pre>	<p>获取索引的刷新或重组状态。(主索引指的是重组的索引; 增量索引指的是刷新的索引)。选项 <code>numrecords</code> 显示最新的 <code>N</code> 个索引构建状态记录。如果省略了 <code>numrecords</code>, 就会返回最后 20 次索引构建操作的状态。</p> <p>样本命令:</p> <pre>esadmin monitor getCollectionIndexMonitorStatus -cid coll -buildType main -numrecords 4</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 监视器 (node1) (sid: monitor) 已经在运行中。PID: 12649</pre> <p>要了解关于返回的状态消息的更多信息, 请参阅第 234 页的『状态命令的详细信息』。</p>
<pre>esadmin startSearch -cid <i>collection_id</i></pre>	<p>启动搜索服务器进程。</p> <p>样本命令:</p> <pre>esadmin startSearch -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (node1) (sid: controller) 已经在运行中。PID: 25917 FFQC5314I 结果: 0</pre>
<pre>esadmin stopSearch -cid <i>collection_id</i></pre>	<p>停止搜索服务器进程。</p> <p>样本命令:</p> <pre>esadmin stopSearch -cid coll</pre> <p>样本消息和返回码:</p> <pre>FFQC5303I 控制器 (node1) (sid: controller) 已经在运行中。PID: 15292 FFQC5314I 结果: 0</pre>

表 5. 企业搜索 *esadmin* 命令 (续)

命令	描述
<code>esadmin monitor</code> <code>getCollectionSearchMonitorStatus</code> <code>-cid collection_id</code>	获取搜索服务器的状态。 样本命令: <code>esadmin monitor getCollectionSearchMonitorStatus</code> <code>-cid coll</code> 样本消息和返回码: FFQC5303I 监视器 (node1) (sid: monitor) 已经在运行中。PID: 12649
<code>esadmin searchmanager_session_id</code> <code>getStatus -cid collection_id</code>	返回给定搜索服务器上的某个集合的详细搜索索引状态信息。每个搜索服务器都有一个搜索管理器会话。每个搜索管理器会话都负责监视和操作特定搜索服务器上的搜索索引。 样本命令: <code>esadmin searchmanager.node1 getStatus -cid coll</code> 样本消息和返回码: FFQC5303I 搜索管理器 (node1) (sid: searchmanager.node1) 已经在运行中。PID: 15711 FFQC5314I 结果: PID=18390 CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1 要了解关于返回的状态消息的更多信息, 请参阅『状态命令的详细信息』。

状态命令的详细信息

某些命令可以返回详尽的信息。本节描述可以返回的搜寻器状态信息和搜寻空间状态信息。第 229 页的『企业搜索 *esadmin* 命令』一节中的表提供了每个 *esadmin* 命令可能返回的信息。本节描述下列命令返回的信息:

- Web 搜寻器状态
- 非 Web 搜寻器状态
- Web 搜寻器的搜寻空间状态
- 非 Web 搜寻器的搜寻空间状态
- Web 搜寻器的详细搜寻空间状态
- 非 Web 搜寻器的详细搜寻空间状态
- 解析器状态
- 索引构建状态
- 搜索服务器状态
- 详细的搜索服务器状态

Web 搜寻器状态: 当运行此命令以获取 Web 搜寻器状态时, 此命令将返回 XML 文档格式的信息。Web 搜寻器状态命令可以返回以下信息:

```
FFQC5314I 结果: <?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>
```

下表描述了 Web 搜寻器状态命令返回的每个 XML 元素及其可能属性:

表 6. Web 搜寻器状态信息

元素	属性	描述
CrawlerStatus	<ul style="list-style-type: none"> CrawlerThreadStateDist ActiveBucketList CrawlRate RecentlyCrawledURLList NumURLsThisSession 	搜寻器状态。
CrawlerRunLevel 值	<ul style="list-style-type: none"> 字符串 (英文) “Not started”: 搜寻器会话已存在, 但它尚未接收到指示其处理文档的启动消息。 “Started”: 搜寻器正在启动。 “Running”: 搜寻器已完成初始化和启动, 当前正在执行搜寻。 “Paused”: 搜寻器已被指示暂停活动的搜寻操作, 但尚未退出。 “Stopping”: 搜寻器已接收到停止信号, 将要停止。 “Error”: 搜寻器处于不可恢复的状态, 它必须停止并重新启动才能继续执行搜寻。 	关于搜寻器正在执行的操作的信息。
CrawlerThreadState 状态	字符串 (英文)	搜寻器线程活动。此字段显示线程正在执行的操作。

表 6. Web 搜寻器状态信息 (续)

元素	属性	描述
ActiveBucket	<ul style="list-style-type: none"> URL: 字符串 (URL 规范) 正在搜寻的 URL 的协议、主机和端口。 NumActURLs: 整数 (正数) 在使存储区可供搜寻 (激活) 时其中包含的 URL 数。 NumProcURLs: 整数 (非负数) 迄今为止在存储区中已处理 (已搜寻或已拒绝) 的 URL 数。 TimeRem: 整数 离存储区超时还剩的秒数。 Duration: 整数 (非负数) 从激活存储区起已经过的秒数。 	指定的 Web 站点的当前活动。
CrawlRate	Value: 整数 (非负数)	搜寻器吞吐量计量。
	每秒搜寻的页面数 (所有存储区的组合值)。	
RecentlyCrawledURL	URL: 字符串 (URL 规范)	最近搜寻的页面。
	字符串, 它指定已搜寻的协议、主机、端口和文件。	
NumURLsThisSession	Value: 整数 (非负数)	自从此搜寻器实例 (进程) 开始搜寻后已搜寻的 URL 数。

非 Web 搜寻器状态: 当运行此命令以获取非 Web 搜寻器的搜寻器状态时, 此命令将返回 XML 文档格式的信息。非 Web 搜寻器状态命令可以返回以下信息:

```
FFQC5314I 结果: <?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>
```

下列各表描述了除 Web 搜寻器以外的每个企业搜索搜寻器的 XML 元素和属性。此信息是使用搜寻器状态命令返回的。

表 7. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用搜寻器状态命令时返回的元素和属性

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
Status	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)
StatusMessage	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误

表 7. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用搜寻器状态命令时返回的元素和属性 (续)

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
NumberOfServers	搜寻空间中的 NNTP 服务器数。	搜寻空间中的数据库数。	搜寻空间中的数据库数。
NumberOfCompletedServers	已搜寻的 NNTP 服务器数。	已搜寻的数据库数。	已搜寻的数据库数。
NumberOfTargets	搜寻空间中的新闻组数。	搜寻空间中的数据库数。	搜寻空间中的视图和文件夹数。
NumberOfCompletedTargets	已搜寻的新闻组数。	已搜寻的表数。	已搜寻的视图和文件夹数。
NumberOfCompletedRecords	已搜寻的文章数。	已搜寻的记录数。	已搜寻的文档数。
RunningThreads	搜寻器线程数。	搜寻器线程数。	搜寻器线程数。

表 8. 对 Exchange Server 搜寻器、DB2 Content Manager 搜寻器和 Content Edition 搜寻器使用搜寻器状态命令时返回的元素和属性

元素名和属性名	Exchange Server 搜寻器	DB2 Content Manager 搜寻器	Content Edition 搜寻器
Status	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)
StatusMessage	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误
NumberOfServers	搜寻空间中的 Exchange Server 服务器数。	搜寻空间中的 Content Manager 服务器数。	搜寻空间中的存储库数。
NumberOfCompletedServers	已搜寻的 Exchange Server 服务器数。	已搜寻的 Content Manager 服务器数。	已搜寻的存储库数。
NumberOfTargets	搜寻空间中的子文件夹数。	搜寻空间中的项类型数。	搜寻空间中的类数。
NumberOfCompletedTargets	已搜寻的子文件夹数。	已搜寻的项类型数。	已搜寻的项类数。
NumberOfCompletedRecords	已搜寻的文档数。	已搜寻的文档数。	已搜寻的文档数。
RunningThreads	搜寻器线程数。	搜寻器线程数。	搜寻器线程数。

表 9. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用搜寻器状态命令时返回的元素和属性

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Status	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)	状态 (0、1、2 和 -1)
StatusMessage	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误	状态: 0 - 空闲, 1 - 正在运行, 2 - 已暂停, -1 - 错误
NumberOfServers	搜寻空间中的位置数。	搜寻空间中的库数。	固定值 1。
NumberOfCompletedServers	已搜寻的位置数。	已搜寻的库数。	0 或 1 (如果已搜寻了所有子目录)。
NumberOfTargets	搜寻空间中的位置数据库数和空间数据库数。	搜寻空间中的文件柜数。	搜寻空间中的子目录数。
NumberOfCompletedTargets	已搜寻的位置数据库数和空间数据库数。	已搜寻的文件柜数。	已搜寻的子目录数。
NumberOfCompletedRecords	已搜寻的文档数。	已搜寻的文档数。	已搜寻的文件数。
RunningThreads	搜寻器线程数。	搜寻器线程数。	搜寻器线程数。

Web 搜寻器的搜寻空间状态: 当运行此命令以获取 Web 搜寻器的搜寻空间状态时, 此命令将返回 XML 文档格式的信息。Web 搜寻空间状态命令可以返回以下信息:

表 10. Web 搜寻器搜寻空间状态命令的选择掩码值

掩码位	选择
1	原始数据商店中的页面数。
2	已发现的站点数。
4	有 DNS 的站点数。
8	没有 DNS 的站点数。
16	已发现的 URL 数。
32	已保存的唯一页面数。
64	已搜寻的 URL 数。
128	未搜寻的 URL 数。
256	已延误的 URL 数。
512	HTTP 返回码分发。

所有值都表示使用当前内部数据库的所有会话的累积总计:

```
<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>
```

返回的数据包含下列任何或全部元素 (有可能未包含任何元素):

表 11. Web 搜寻器搜寻空间状态命令返回的信息

元素	属性	描述
CrawlerStatus	<ul style="list-style-type: none"> • NumPagesInRDS • NumSitesDiscovered • NumSitesWithDNS • NumSitesWithoutDNS • NumURLsDiscovered • NumUniquePagesSaved • NumURLsCrawled • NumURLsUncrawled • NumURLsOverdue • HTTPCodeDist 	可以快速获取的累积搜寻状态信息（所有会话）。
NumPagesInRDS	Value: 非负整数 原始数据商店（RDS）登台区域当前包含的页面数（仅来自此搜寻器的页面）。	原始数据商店（RDS）的满程度（仅限于由此搜寻器添加的内容）。
NumSitesDiscovered	Value: 非负整数 通过搜寻（或从种子）发现的主机数。	这是搜寻器对所搜寻的域的覆盖范围的测量（主机计数）。
NumSitesWithDNS	Value: 非负整数 有相关 IP 地址的主机数（IP 地址由搜寻器在后台解析）。	这是搜寻器获取由 URL 中的 DNS 名称发现的主机 IP 地址的能力效率的测量。
NumSitesWithoutDNS	Value: 非负整数 没有相关 IP 地址的主机数（IP 地址由搜寻器在后台解析）。	这是搜寻器获取由 URL 中的 DNS 名称发现的主机 IP 地址的能力效率的测量。
NumURLsDiscovered	Value: 非负整数 搜寻器已访问的唯一 URL 数。	这是搜寻器对所搜寻的域的覆盖范围的测量（URL 计数）。
NumUniquePagesSaved	Value: 非负整数 已写入 RDS 的唯一页面数，其他企业搜索组件将对这些页面进行进一步的处理。	此搜寻器对索引大小所起的影响。
NumURLsCrawled	Value: 非负整数 搜寻器已搜寻的唯一 URL 数。	这是搜寻器的数据处理能力的测量（端到端）。由于并不会将搜寻到的所有页面都写入 RDS，所以此数目与已写入 RDS 的页数不同。
NumURLsOverdue	Value: 非负整数 符合重新搜寻条件的唯一 URL 数。	这是搜寻器遍历 Web 空间能力的测量。

非 Web 搜寻器的搜寻空间状态: 当运行此命令以获取非 Web 搜寻器的搜寻空间状态时，此命令将返回 XML 文档格式的信息。非 Web 搜寻空间状态命令可以返回以下信息:

```
FFQC5314I 结果: <?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>

```

下列各表描述了除 Web 搜寻器以外的每个企业搜索搜寻器的 XML 元素和属性。此信息是使用搜寻空间状态命令返回的。对于 Notes 搜寻器来说，当聚集级别为 0 时，Server@Name 是服务器名称 + 数据库名称。当聚集级别为 1 时，Server@Name 是服务器名称 + 目录名。

表 12. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用搜寻器状态命令时返回的元素和属性

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
Server@Name	新闻服务器名称	数据库名称	数据库名称或目录名
Server/Status	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • 5: 已安排时间表 • -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • 5: 已安排时间表 • -1: 错误
Server/StatusMessage	<ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • 5: 已安排时间表 • -1: 错误 	<ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • 5: 已安排时间表 • -1: 错误 	<ul style="list-style-type: none"> • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • 5: 已安排时间表 • -1: 错误
Server/NumberOfTargets	搜寻空间中的新闻组数。	搜寻空间中的数据库数。	搜寻空间中的视图和文件夹数。
Server/NumberOf CompletedTargets	已搜寻的新闻组数。	已搜寻的表数。	已搜寻的视图和文件夹数。
Server/NumberOfErrors	无。	错误数。	错误数。
Server/StartTime	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。
Server/EndTime	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。

表 12. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用搜寻器状态命令时返回的元素和属性 (续)

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
Server/ScheduleConfigured	无。	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。 	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。
Server/ScheduleTime	无。	安排时间 (如果适用的话)。	安排时间 (如果适用的话)。
Server/TotalTime	总时间 (如果适用的话)。	总时间 (如果适用的话)。	总时间 (如果适用的话)。
Server/AggregationLevel	无。	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0 和 1: <ul style="list-style-type: none"> 0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。) Notes 搜寻器以目录方式搜寻文档。

表 13. 对 Exchange Server 搜寻器、DB2 Content Manager 搜寻器和 Content Edition 搜寻器使用搜寻空间状态命令时返回的元素和属性

元素名和属性名	Exchange Server 搜寻器	DB2 Content Manager 搜寻器	Content Edition 搜寻器
Server@Name	Exchange Server 服务器名称。	DB2 Content Manager 服务器。	存储库名。
Server/Status	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误

表 13. 对 Exchange Server 搜寻器、DB2 Content Manager 搜寻器和 Content Edition 搜寻器使用搜寻空间状态命令时返回的元素和属性 (续)

元素名和属性名	Exchange Server 搜寻器	DB2 Content Manager 搜寻器	Content Edition 搜寻器
Server/StatusMessage	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误
Server/NumberOfTargets	搜寻空间中的子文件夹数。	搜寻空间中的项类型数。	搜寻空间中的项类数。
Server/NumberOfCompletedTargets	已搜寻的子文件夹数。	已搜寻的项类型数。	已搜寻的项类数。
Server/NumberOfErrors	错误数。	错误数。	错误数。
Server/StartTime	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。
Server/EndTime	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。
Server/ScheduleConfigured	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。 	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。 	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。
Server/ScheduleTime	安排时间 (如果适用的话)。	安排时间 (如果适用的话)。	安排时间 (如果适用的话)。
Server/TotalTime	总时间 (如果适用的话)。	总时间 (如果适用的话)。	总时间 (如果适用的话)。
Server/AggregationLevel	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)

表 14. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用搜寻空间状态命令时返回的元素和属性

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Server@Name	位置目录	库数据库	固定值 localhost。

表 14. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用搜寻空间状态命令时返回的元素和属性 (续)

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Server/Status	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	状态 (0、1、2、3、4、5 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误
Server/StatusMessage	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 5: 已安排时间表 -1: 错误
Server/NumberOfTargets	搜寻空间中的位置数据库数和空间数据库数。	搜寻空间中的文件柜数。	搜寻空间中的子目录数。
Server/NumberOfCompletedTargets	已搜寻的位置数据库数和空间数据库数。	已搜寻的文件柜数。	搜寻空间中的子目录数。
Server/NumberOfErrors	错误数。	错误数。	错误数。
Server/StartTime	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。	开始时间 (如果适用的话)。
Server/EndTime	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。	结束时间 (如果适用的话)。
Server/ScheduleConfigured	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。 	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。 	0、1 和 2 <ul style="list-style-type: none"> 0: 未将搜寻器配置为根据搜寻器配置文件执行搜寻。 1: 已经为搜寻器配置了时间表, 但此会话禁用了该时间表。 2: 已经为搜寻器配置了时间表, 并且此会话启用了该时间表。
Server/ScheduleTime	安排时间 (如果适用的话)。	安排时间 (如果适用的话)。	安排时间 (如果适用的话)。
Server/TotalTime	总时间 (如果适用的话)。	总时间 (如果适用的话)。	总时间 (如果适用的话)。

表 14. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用搜寻空间状态命令时返回的元素和属性 (续)

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Server/AggregationLevel	0 和 1 <ul style="list-style-type: none"> 0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。) Notes 搜寻器以目录方式搜寻文档。 	0 和 1 <ul style="list-style-type: none"> 0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。) Notes 搜寻器以目录方式搜寻文档。 	0 和 1 <ul style="list-style-type: none"> 0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。) Notes 搜寻器以目录方式搜寻文档。

Web 搜寻器的详细搜寻空间状态: 当运行此命令以获取 Web 搜寻器的详细搜寻空间状态时, 此命令将返回 XML 文档格式的信息。详细搜寻空间状态命令可以返回以下信息:

表 15. Web 搜寻器详细搜寻空间状态命令的选择掩码值

掩码位	选择
1	原始数据商店中的页面数。
2	已发现的站点数。
4	有 DNS 的站点数。
8	没有 DNS 的站点数。
16	已发现的 URL 数。
32	已保存的唯一页面数。
64	已搜寻的 URL 数。
128	未搜寻的 URL 数。
256	已延误的 URL 数。
512	HTTP 返回码分发。

返回的信息的样本:

```
<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>

```

下表描述了返回的 Web 搜寻器详细搜寻空间状态的每个字段:

表 16. Web 搜寻器详细搜寻空间状态命令返回的信息

元素	属性	描述
CrawlDetailsPerSite	<ul style="list-style-type: none"> LastActivationTime: LastActivationDuration: IPAddressList: RobotsContent: HTTPCodeDist: 	可以快速获取的站点详细状态信息。
Site	URL	站点根页的 URL。
NumURLsDiscovered	Value	在该站点上已发现的 URL 数。
NumURLsOverdue	Value	该站点上符合重新搜寻条件的 URL 数。
NumURLsCrawled	Value	在该站点上已搜寻的 URL 数。
NumURLsUncrawled	Value	在该站点上尚未搜寻的 URL 数。
NumURLsOverdueBy	Threshold 和 Value: 整数 (正数或负数) 用于查找已延误指定时间量 (负数 = 当前时间之前的秒数) 的重新搜寻或者在将来的指定时间间隔内 (正数 = 当前时间之后的秒数) 即将进行的重新搜寻的阈值。	符合重新搜寻条件至少已达某个秒数 (阈值) 或者在接下来的这个秒数 (阈值) 内将会符合重新搜寻条件的 URL 数。
NumURLsActivated	Value	上次扫描此站点期间读入内存并提供给搜寻器线程使用的 URL 数。
LastActivationTime	Value	上次将此站点的 URL 读入内存后经过的秒数。
LastActivationDuration	Value	此站点的 URL 上次在内存中并可供搜寻器线程使用的秒数。
IPAddressList	IPAddress	此站点的服务器主机的所有已知 IP 地址。
IPAddress	Value	此站点的服务器主机的 IPv4 点分地址。
RobotsContent	Text	robots 文件中的文本 (如果该文件包含任何文本的话)。
HTTPCodeDist	HTTPCode	对此站点尝试的下载所分发的 HTTP 代码。
HTTPCode	Code: 整数 HTTP 返回码或另一个内部代码。	搜寻此站点期间特定 HTTP 返回码的出现次数。

非 Web 搜寻器的详细搜寻空间状态: 当运行此命令以获取非 Web 搜寻器的详细搜寻空间状态时, 此命令将返回 XML 文档格式的信息。非 Web 搜寻器的详细搜寻空间状态命令可以返回以下信息:

```
FFQC5314I 结果: <?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. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用详细搜寻状态命令时返回的元素和属性

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
Target@Name	新闻组名	表名	视图或文件夹名
Target@CrawlType	无。	0 和 1: • 0: 主动搜寻 (正常) • 1: 被动搜寻 (DB2 事件发布)	0
Target/Status	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误
Target/StatusMessage	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误
Target/NumberOf Records	服务器上的上一个文章号。	已搜寻的记录数。	已搜寻的文档数。
Target/NumberOf CompletedRecords	已搜寻的文章数。	已搜寻的记录数。	已搜寻的文档数。
Target/NumberOf InsertedRecords	新记入的文章数。	已插入的记录数。	已插入的记录数。
Target/NumberOf UpdatedRecords	无。	已更新的记录数。	已更新的记录数。
Target/NumberOf DeletedRecords	无。	已删除的记录数。	已删除的记录数。
Target/StartTime	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。

表 17. 对 NNTP 搜寻器、DB2 搜寻器和 Notes 搜寻器使用详细搜寻状态命令时返回的元素和属性 (续)

元素名和属性名	NNTP 搜寻器	DB2 搜寻器	Notes 搜寻器
Target/EndTime	搜寻的完成日期和时间。	搜寻的完成日期和时间。	搜寻的完成日期和时间。
Target/TotalTime	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。
Target/AggregationLevel	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0 和 1: <ul style="list-style-type: none"> 0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。) Notes 搜寻器以目录方式搜寻文档。
Target/LastUpdatedTime	无。	上次更新时间: <ul style="list-style-type: none"> 0: 主动搜寻 (正常) 1: 被动搜寻 (DB2 事件发布) 	无。
Target/LastResetTime	无。	上次复位统计信息时间: <ul style="list-style-type: none"> 0: 主动搜寻 (正常) 1: 被动搜寻 (DB2 事件发布) 	无。

表 18. 对 Exchange Server 搜寻器、DB2 Content Manager 搜寻器和 Content Edition 搜寻器使用详细搜寻空间状态命令时返回的元素和属性

元素名和属性名	Exchange Server 搜寻器	DB2 Content Manager 搜寻器	Content Edition 搜寻器
Target@Name	子文件夹名	项类型名	项类名
Target@CrawlType	0	0	0
Target/Status	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	状态 (0、1、2、3、4 和 -1) <ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误
Target/StatusMessage	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误 	<ul style="list-style-type: none"> 0: 未搜寻 1: 正在搜寻 2: 已完成 (未安排时间表) 3: 正在等待 4: 已暂停 -1: 错误
Target/NumberOf Records	无。	无。	无。

表 18. 对 Exchange Server 搜寻器、DB2 Content Manager 搜寻器和 Content Edition 搜寻器使用详细搜寻空间状态命令时返回的元素和属性 (续)

元素名和属性名	Exchange Server 搜寻器	DB2 Content Manager 搜寻器	Content Edition 搜寻器
Target/NumberOf CompletedRecords	已搜寻的文档数。	已搜寻的文档数。	已搜寻的文档数。
Target/NumberOf InsertedRecords	已插入的记录数。	已插入的记录数。	已插入的记录数。
Target/NumberOf UpdatedRecords	无。	已更新的记录数。	已更新的记录数。
Target/NumberOf DeletedRecords	无。	已删除的记录数。	已删除的记录数。
Target/StartTime	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。
Target/EndTime	搜寻的完成日期和时间。	搜寻的完成日期和时间。	搜寻的完成日期和时间。
Target/TotalTime	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。
Target/AggregationLevel	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)
Target/LastUpdatedTime	无。	无。	无。
Target/LastResetTime	无。	无。	无。

表 19. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用详细搜寻空间状态命令时返回的元素和属性

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Target@Name	位置数据库名称或空间数据库名称	文件柜数据库名称	子目录名
Target@CrawlType	0	0	0
Target/Status	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	状态 (0、1、2、3、4 和 -1) • 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误
Target/StatusMessage	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误	• 0: 未搜寻 • 1: 正在搜寻 • 2: 已完成 (未安排时间表) • 3: 正在等待 • 4: 已暂停 • -1: 错误

表 19. 对 QuickPlace 搜寻器、Domino Document Manager 搜寻器、UNIX 文件系统搜寻器和 Windows 文件系统搜寻器使用详细搜寻空间状态命令时返回的元素和属性 (续)

元素名和属性名	QuickPlace 搜寻器	Domino Document Manager 搜寻器	UNIX 和 Windows 文件系统搜寻器
Target/NumberOf Records	无。	无。	无。
Target/NumberOf CompletedRecords	已搜寻的文档数。	已搜寻的文档数。	已搜寻的文件数。
Target/NumberOf InsertedRecords	已插入的记录数。	已插入的记录数。	已插入的记录数。
Target/NumberOf UpdatedRecords	已搜寻的位置数据库数和空间数据库数。	已搜寻的文件柜数。	搜寻空间中的子目录数。
Target/NumberOf DeletedRecords	已更新的记录数。	已更新的记录数。	已更新的记录数。
Target/StartTime	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。	搜寻器的上次启动日期和时间。
Target/EndTime	搜寻的完成日期和时间。	搜寻的完成日期和时间。	搜寻的完成日期和时间。
Target/TotalTime	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。	搜寻器在执行搜寻时耗用的时间量。
Target/AggregationLevel	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)	0: Notes 搜寻器以正常方式搜寻文档。(除 Notes 搜寻器以外的其他搜寻器始终返回 0。)
Target/LastUpdatedTime	无。	无。	无。
Target/LastResetTime	无。	无。	无。

解析器状态: 当运行此命令以获取解析器状态时, 此命令将返回 XML 文档格式的信息。解析器状态命令可以返回以下信息:

```
FFQC5314I 结果: <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>
```

下表描述了解析器状态命令返回的信息的 XML 元素:

表 20. 解析器状态命令的元素

元素	描述
Status	如果此集合的解析器会话正在运行中, 则为 1, 如果解析器会话已停止, 则为 0。
State	可能的状态是: Idle、Restart、Parsing、Resuming、Stopped、Initializing、Pause 和 NoParserServiceIsAvailable。 状态 Idle 表示解析器已休眠 N 分钟, 它正在等待此集合中的搜寻器提供更多的文档。缺省休眠时间为 300 秒。 状态 Restart 表示解析器正在等待解析 / 标记化 JVM 重新启动。解析 / 标记化 JVM 在一个独立的会话中运行, 这就是最终处理文档的位置。 状态 Parsing 表示解析器正在处理文档。 状态 Pause 表示解析器已被此集合的索引构建会话暂停。状态 Initializing 表示解析器正在启动并初始化其状态。 状态 Resuming 表示此集合的索引构建会话已使解析器由 Pause 状态恢复到 Parsing 状态。 状态 NoParserServiceIsAvailable 表示没有解析 / 标记化 JVM 可用来处理此集合的文档。这意味着所有解析 / 标记化 JVM 都被其他集合使用。
SnapShotTimeStamp	解析器下次从搜寻器表获取已读取的文档的时间, 以 1970 年起的秒数计。
NumberOfDocsToBeIndexed	存储库中此集合的文档数。此数目还包括已标记为要在下次索引构建期间删除的文档。
ParseRate	解析速度, 以每秒文档数计。
ParseRateMBPerHour	解析速度, 以每小时 MB 数计。
NumberOfCpmThreads	解析 / 标记化 JVM 在处理此集合的文档时使用的 CPM 线程数。
ParserServiceSession	正在处理此集合的文档的解析标记化 JVM 的名称。仅当解析器处于“解析”状态时, 此字段才可用。
Name	搜寻器的名称。
Type	搜寻器的类型 (Web、NNTP 和 DB2 等等)。
ParserStatus	<ul style="list-style-type: none"> • 0: 尚未对来自此搜寻器的文档进行解析。 • 1: 当前正在对来自此搜寻器的文档进行解析。 • 2: 已对来自此搜寻器的文档进行解析。
NumberOfDocsToBeParsed	来自此搜寻器的正在等待解析的文档数。
NubmerOfDocsAlreadyParsed	来自此搜寻器的已解析文档数。

索引状态: 当运行此命令以获取索引构建状态时, 此命令将返回 XML 文档格式的信息。索引构建状态命令可以返回以下信息:

```
FFQC5314I 结果: <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>

```

下表描述了索引构建状态命令返回的信息的每个 XML 元素:

表 21. 索引构建状态命令的元素

元素	描述
IndexStatusId	索引状态标识。
StartTime	此次索引构建的开始时间, 以 1970 年起的秒数计。
Progress	此次索引构建的完成百分比。
CurrentPhase	<ul style="list-style-type: none"> • 1: 存储重写阶段 • 2: 全局分析阶段 • 3: 索引构建阶段
TotalPhase	此次索引构建的阶段数。此值当前是 3。
IndexCopyProgress	此次索引复制的完成百分比。索引复制进程将已构建的索引从索引构建服务器复制至搜索服务器。
CurrentServer	索引复制操作的复制目标搜索服务器。
TotalServer	要将索引复制到的搜索服务器数。
IndexCopyTime	所有索引构建阶段的总时间。
Progress	0 表示索引构建和复制操作成功, 非零数值表示错误消息码。
StopTime	索引构建(所有阶段)和索引复制操作的结束时间。
TotalTime	开始时间与停止时间之间的时间段。
NumberOfDocuments	索引中的文档数。
CurrentIndexWildcardSupport	下次执行索引构建操作时要使用的通配符设置。可能的值是 None、QueryExpansion 或 IndexExpansion。
ScheduleStatus	<ul style="list-style-type: none"> • 如果未对此集合和索引类型启用时间表, 则为 0。 • 如果已对此集合和索引类型启用时间表, 则为 1。
ScheduledTimeEnabled	此集合和索引类型的索引构建操作的下次运行时间, 以 1970 年起的秒数计。

搜索服务器状态: 当运行此命令以获取搜索服务器状态时, 此命令将返回 XML 文档格式的信息。搜索服务器状态命令可以返回以下信息:

```

FFQC5314I 结果: <?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>

```

下表描述了搜索服务器状态命令返回的信息的 XML 元素:

表 22. 搜索服务器状态命令的元素

元素	描述
SearchStatusName	搜索管理器会话的名称和标识，该搜索管理器会话正在监视和维护此集合的搜索索引。
HostName	正在运行搜索索引的服务器的主机名。
Status	<ul style="list-style-type: none"> 如果此集合的搜索索引未在运行中，则为 0。 如果此集合的搜索索引正在运行中，则为 1。

详细搜索服务器状态： 返回搜索服务器状态的命令可以返回下列信息：

```
FFQC5303I 搜索管理器 (node1) (sid: searchmanager.node1)
已经在运行中。PID: 15711
FFQC5314I 结果: PID=18390
CacheHits=3
QueryRate=1
Port=44008
SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1
```

下表描述了详细搜索服务器状态命令返回的信息中的各个项：

表 23. 详细搜索服务器状态命令返回的各个项

项	描述
CacheHits	从搜索高速缓存中检索到的结果数。
QueryRate	在上一时间间隔内接收到的查询数。缺省情况下，时间间隔是 5 分钟。
Port	搜索索引在侦听或接收查询时使用的端口号。
SessionId	此集合的搜索索引的会话标识。
CacheHitRate	从搜索高速缓存中检索到的结果数，以所有搜索结果的百分比计。
ResponseTime	指定时间间隔内的平均响应时间，以毫秒计。（缺省值是 5 分钟。）
Status	<ul style="list-style-type: none"> 如果此集合的搜索索引未在运行中，则为 0。 如果此集合的搜索索引正在运行中，则为 1。
SessionName	此集合的搜索索引的会话名。

esadmin 命令的返回码

esadmin 命令可以返回下列代码：

表 24. esadmin 命令的返回码

代码	名称	描述
0	CODE_ERROR_NONE	已成功地完成了命令。
102	CODE_ERROR_INSTANTIATION_EXCEPTION	实例化命令句柄时出错。
103	CODE_ERROR_ACCESS_EXCEPTION	实例化命令句柄时发生非法访问错误。
104	CODE_ERROR_EXECUTE_EXCEPTION	
105	CODE_ERROR_THROWABLE	
106	CODE_ERROR_NO_SUCH_METHOD	

表 24. esadmin 命令的返回码 (续)

代码	名称	描述
107	CODE_ERROR_INVALID_SESSION	
108	CODE_ERROR_INVALID_PARAMETER	
109	CODE_ERROR_SESSION_NOT_RUNNING	

获取会话标识

使用 esadmin check 命令来显示企业搜索组件及其相应会话标识的列表。下表显示了公共会话、它们的标识、它们所在的服务器以及会话状态的列表。

表 25. 会话名、源服务器、会话标识和会话状态的示例

会话	运行会话的服务器	会话标识	会话状态
configmanager	索引服务器	10433	已启动
controller	索引服务器	10464	已启动
customcommunication	索引服务器	无	无
datalistener	索引服务器	10582	已启动
discovery	索引服务器	10649	已启动
monitor	索引服务器	10682	已启动
parserservice	索引服务器	10718	已启动
resource.node1	索引服务器	10759	已启动
samplecpp	索引服务器	10827	已启动
sampletest	索引服务器	10857	已启动
scheduler	索引服务器	10889	已启动
searchmanager.node1	索引服务器	10927	已启动
utilities.node1	索引服务器	10384	已启动

相关概念

第 225 页的『备份和复原企业搜索系统』

备份和复原脚本使您能够备份和复原企业搜索系统。

第 199 页的『监视企业搜索活动』

在监视系统和集合活动时，您可以查看各种进程的状态、监视潜在的问题或调整配置以提高性能。

相关任务

第 203 页的『监视搜寻器』

您可以查看关于集合中每个搜寻器的状态的一般信息，也可以选择选项以查看关于搜寻器活动的详细信息。

第 195 页的『启动企业搜索服务器』

为了使用户能够搜索集合，必须启动系统进程，然后启动对该集合进行搜寻、解析、建立索引和搜索的服务器。

第 197 页的『停止企业搜索服务器』

如果对企业搜索服务器的配置作了更改，或者需要对问题进行故障诊断，则可能需要停止并重新启动企业搜索服务器。

企业搜索文档

您可以阅读 PDF 或 HTML 格式的 WebSphere Information Integrator OmniFind Edition (企业搜索) 文档。

WebSphere Information Integrator OmniFind Edition 安装程序可以自动安装信息中心。安装程序在搜索服务器上安装信息中心。如果未安装信息中心, 当您单击“帮助”时, 将打开 IBM Web 站点上的信息中心。要查看 HTML 格式的企业搜索主题, 请启动信息中心。

要查看 PDF 文档, 请转到 docs/locale/pdf。例如, 要查找英文文档, 请转到 docs/en_US/pdf。您还可以查看 WebSphere Information Integrator OmniFind Edition 支持站点上提供的最新 PDF 文档。

下表显示了可用的文档、文件名和位置。

表 26. PDF 格式的企业搜索文档

标题	标题	标题
《企业搜索安装指南》(信息中心也提供了此文档的主题)	iiysi.pdf	docs/locale/pdf/
<i>Administering Enterprise Search</i> (信息中心也提供了此文档的主题)	iiysa.pdf	docs/locale/pdf/
<i>Programming Guide and API Reference for Enterprise Search</i> (信息中心也提供了此文档的主题)	iiysp.pdf	docs/locale/pdf/
<i>Messages for Enterprise Search</i> (信息中心也提供了此文档的主题)	iiysm.pdf	docs/locale/pdf/
<i>Installation Requirements for Enterprise Search</i> (信息中心也提供了此文档的主题)	iiysr.txt 或 iiysr.htm	docs/locale/ (也可以从“第一步”程序中打开此文件。)
《发行说明》	iiysn.pdf	仅在 IBM WebSphere Information Integrator OmniFind Edition 文档 Web 站点上提供。
WebSphere Information Integrator 信息中心	无	

WebSphere II OmniFind Edition 辅助功能选项

IBM WebSphere Information Integrator OmniFind Edition 的用户界面和文档都是可访问的。

安装程序

可以通过键盘快捷方式来使用 WebSphere II OmniFind Edition 安装程序。下表描述了一些键盘快捷方式。

表 27. 安装程序的键盘快捷方式

操作	快捷方式
突出显示单选按钮	方向键
选择单选按钮	Tab 键
突出显示按钮	Tab 键
选择按钮	Enter 键
转到下一个或上一个窗口或取消	通过按 Tab 键和按 Enter 键来突出显示按钮
使活动窗口不活动	Ctrl + Alt + Esc

企业搜索管理控制台和信息中心

管理控制台和信息中心都是基于浏览器的界面，您可以在 Microsoft Internet Explorer 或 Mozilla FireFox 中查看这些界面。请参阅 Internet Explorer 或 FireFox 的联机帮助以获取键盘快捷方式列表以及浏览器的其它辅助功能选项功能部件。

PDF 文档

您可以查看所有 PDF 格式的企业搜索文档。可以使用 Adobe Acrobat V6.0 来打开 PDF 文档。PDF 文档是结构化的文档，大部分屏幕阅读器都可以阅读这些文档。

企业搜索术语词汇表

本词汇表定义企业搜索界面和文档使用的术语。

访问控制表 (**access control list**)

这个列表标识能够访问相关对象的用户并指定用户对该对象的访问权。

管理角色 (**administrative role**)

用户的分类，它确定用户在企业搜索管理控制台中能够执行的功能。角色还确定用户可以管理哪些集合。

分析引擎 (**analysis engine**)

请参阅文本分析引擎 (**text analysis engine**) 。

分析结果 (**analysis results**)

注释器生成的信息。分析结果与您想要搜索的信息相对应，它被写入称为公共分析结构的数据结构中。

注释 (**annotation**)

有关文本范围的信息。例如，注释可以指出表示公司名的文本范围。在 UIMA 中，注释是一种特殊类型的功能部件结构。

注释器 (**annotator**)

负责执行特定语言分析任务并生成和记录注释的软件组件。注释器是分析引擎中的分析逻辑组件。

布尔搜索 (**boolean search**)

一种搜索，在这种搜索中，使用诸如 AND、NOT 和 OR 之类的运算符来将一个或多个搜索项组合起来。

升降类 (**boost class**)

可以影响文档在搜索结果中的相对排名的规范。

升降字 (**boost word**)

可以影响文档在搜索结果中的相应排名的单词。在查询处理期间，包含升降字的文档的重要程度可能会提高或下降，这取决于对该单词预定义的分數。

类别 (**category**)

一组具有相似属性的文档。

类别树 (**category tree**)

这指的是企业搜索管理控制台中显示的类别层次结构。

证书 (**certificate**)

一个数字文档，它将公用密钥与证书所有者的标识绑定，从而使该证书所有者能够通过认证。证书由认证中心发放。

认证中心 (**certificate authority**)

一个组织，它发放证书并对电子交易所涉及的实体（个人或组织）进行认证。认证中心保证交换信息的双方确实拥有他们所声称的身份。

字符规范化 (**character normalization**)

一个过程，在此过程中，将字符的变体格式（例如大写和区别标记）还原为公共格式。

附着词 (clitic)

一个单词，它在语法功能是独立的，但在发音上与另一个词相连。附着词和它所绑定的单词可以书写在一起，也可以分开书写。附着词的常见示例包括英语缩写的最后一部分 (*wouldn't* 或 *you're*)。

集合 (collection)

一组数据源以及用于对那些数据源进行搜寻、解析、建立索引和搜索的选项。

公共分析结构 (common analysis structure)

一个结构，它存储文本分析引擎正在分析的文档。信息以注释形式存储在公共分析结构以及其它功能部件结构中。

公共通信层 (Common Communication Layer, CCL)

一种通信基础结构，它将 WebSphere Information Integrator OmniFind Edition 的各种组件（控制器、解析器、搜寻器和索引器）统一起来。

概念抽取 (concept extraction)

一项搜索功能，它找出文本文档中的重要词汇表词条（例如人员、位置或产品）并生成那些词条的列表。另请参阅主题抽取 (theme extraction)。

搜寻空间 (crawl space)

与指定的模式（例如数据库名称、文件系统路径、域名、IP 地址和 URL）相匹配的一组来源，搜寻器读取这些来源以检索项以便建立索引。

搜寻器 (crawler)

一个软件程序，它从数据源检索文档并收集可用来创建搜索索引的信息。

凭证 (credential)

在认证期间所需的详细信息，它描述用户、任何组关联以及其它与安全性相关的标识属性。可以使用凭证来执行许多服务，例如授权、审计和委托等等。

数据源 (data source)

可以从中检索文档的任何数据存储库，例如 Web、关系数据库和非关系数据库以及内容管理系统。

数据源类型 (data source type)

根据用来访问数据的协议而进行的数据源分组。

出队 (dequeue)

从队列中除去项。

变音符号 (diacritics)

对字母添加的标记，用于更改单词的发音或者对相似的单词进行区分，例如重音符号标记或德语中的变音符号。

发现器 (discoverer)

搜寻器的一项功能，它确定搜寻器可以从哪些数据源检索信息。

专有名称 (distinguished name)

一个名称，它唯一地标识目录中的条目。专有名称由“属性:值”对组成，各对“属性:值”之间由逗号分隔。并且，一组“名称 - 值”对（例如“CN=人员姓名”和“C=国家或地区”）唯一地标识数字证书中的实体。

文档对象模型 (Document Object Model)

一个系统，在该系统中，以对象树形式查看结构化文档（例如 XML 文件），可以通过编程来访问和更新该对象树。

Domino Document Manager 文件柜 (Domino Document Manager cabinet)

用来组织文档的 Domino Document Manager 数据库。文件柜存放 Domino 数据库。

Domino Document Manager 库 (Domino Document Manager library)

一个 Domino Document Manager 数据库，它是 Domino Document Manager 的入口点。

Domino 因特网 ORB 间协议 (Domino Internet Inter-ORB Protocol, DIIOP)

在服务器上运行的一个服务器任务，它与 Domino 对象请求代理配合工作，从而允许在使用 Notes Java 类创建的 Java applet 与 Domino 服务器之间进行通信。浏览器用户和 Domino 服务器使用 DIIOP 来进行通信以及交换对象数据。

动态排名 (dynamic ranking)

一种排名类型，它根据所搜索的文档来分析查询项以确定结果的排名。另请参阅基于文本的打分 (text-based scoring)。与静态排名 (static ranking) 对照。

动态摘要 (dynamic summarization)

一种摘要类型，在此摘要中，搜索项突出显示，并且搜索结果中包含的短语最能表达用户所搜索的文档的概念。与静态摘要 (static summarization) 对照。

排队 (enqueue)

将项放入队列。

企业搜索管理员 (enterprise search administrator)

一种管理角色，它使用户能够管理整个企业搜索系统。

转义字符 (escape character)

一个字符，它取代跟随在它后面的一个或多个字符或者为那些字符选择特殊的含义。

外部数据源 (external data source)

要联合的数据源，WebSphere Information Integrator OmniFind Edition 未对该数据源执行搜寻、进行解析或建立索引。搜索外部数据源的任务委托给那些数据源的查询应用程序编程接口进行。

功能部件路径 (feature path)

一条路径，它用来访问 UIMA 功能部件结构中的功能部件值。

功能部件结构 (feature structure)

用于表示文本分析结果的底层数据结构。功能部件结构是“属性 - 值”结构。每个功能部件结构都是一种类型，每种类型都有一组指定的有效功能部件和属性，这与 Java 类非常相似。

联合搜索 (federated search)

一种搜索功能，它允许跨多个搜索服务执行搜索并返回合并的搜索结果列表。

联合 (federation)

将命名系统组合到一起的过程，执行此过程的目的是使聚集系统能够处理跨命名系统的组合名称。

字段 (field)

记录的最小可标识部分。

字段搜索 (fielded search)

被限制到特定字段的查询。

自由文本搜索 (free text search)

搜索项以自由格式文本表示的一种搜索。

全文索引 (full text index)

一种数据结构，它引用数据项以使搜索能够快速找到包含查询项的文档。

模糊搜索 (fuzzy search)

一种搜索，它返回与搜索项的拼写近似的词。

混合搜索 (hybrid search)

布尔搜索与自由文本搜索的组合。

标识管理 (identity management)

在安全密钥库中对用户凭证进行加密的功能。

索引 (index)

请参阅全文索引 (full text index)。

索引队列 (index queue)

所要处理的索引重组请求或索引刷新请求的列表。

索引刷新 (index refresh)

将新信息添加至企业搜索系统中的现有索引的过程。与索引重组 (index reorganization) 对照。

索引重组 (index reorganization)

在企业搜索系统中构建索引的过程。与索引刷新 (index refresh) 对照。

信息抽取 (information extraction)

一种概念抽取类型，它自动识别文本文档中的重要词汇表词条，例如姓名、术语和表达式。

IP 地址 (IP address)

唯一的 32 位地址，它标识网络中的主机。

Java 数据库连接 (Java Database Connectivity, JDBC)

Java 平台与各种数据库之间的业界连接标准，此连接与数据库无关。JDBC 接口提供了调用级别的 API 来进行基于 SQL 的数据库访问。

JavaScript

浏览器和 Web 服务器中使用的 Web 脚本语言。

JavaServer Pages (JSP)

一种服务器脚本技术，它使 Java 代码能够动态地嵌入在 Web 页面 (HTML 文件) 中并且在该页面接受服务时执行，从而将动态内容返回给客户机。

Java 虚拟机 (Java virtual machine, JVM)

处理器的一种软件实施，它运行编译型 Java 代码 (applet 和应用程序)。

片假名 (Katakana)

一种字符集，它包含一种常用日语语音字母 (共两种) 使用的符号，主要用来按语音书写外来字。

密钥库文件 (keystore file)

一种密钥数据库文件，它包含作为签署者证书存储的公用密钥以及存储在个人证书中的专用密钥。

语言标识 (language identification)

一项企业搜索功能，它确定文档的语言。

词条 (lemma)

单词的规范格式。词条在字形高度变化的语言（例如捷克语）中十分重要。

词条化 (lemmatization)

在字典中查找给定单词的词条的过程。词条化与衍生形态查询的不同点在于，衍生形态查询通过算法进行并且通常不会使用列示语言单词的字典来工作。

词法亲缘关系 (lexical affinity)

在文档中相互意思较为接近的搜索词的关系。词法亲缘关系用来计算结果的相关性。

库 (library)

一种系统对象，它充当其它对象的目录。另请参阅 Domino Document Manager 库 (Domino Document Manager library)。

连字 (ligature)

连接在一起以便作为一个字符出现的两个或更多个字符，例如，将 f 和 i 连接在一起就形成了连字。

轻量级目录访问协议 (Lightweight Directory Access Protocol, LDAP)

一种开放式协议，它使用 TCP/IP 来访问支持 X.500 模型的目录，并且不象更复杂的 X.500 目录访问协议那样需要资源。

语言搜索 (linguistic search)

一种搜索类型，它在将词条还原为基本形态（例如，*mice* 在索引中是 *mouse*）或使用基本形态进行词条扩展（对于复合词）的情况下对文档进行浏览、检索和建立索引。

链接分析 (link analysis)

一种基于文档之间的超链接分析的方法，它用来确定集合中的哪些页面对用户重要。

本地联合器 (local federator)

对一组可搜索对象执行联合的客户端联合器。

Lotus QuickPlace 位置 (Lotus QuickPlace place)

Lotus QuickPlace 提供的 Web 地点，它使地理上分离的参与者能够协作处理项目以及在一个结构化的安全工作空间中进行在线通信。

Lotus QuickPlace 空间 (Lotus QuickPlace room)

Lotus QuickPlace 位置的分区区域，仅供有共同需求并且需要一起工作的授权成员使用。

掩码字符 (masking character)

一个字符，它用来表示位于搜索项开头、中间和末尾的可选字符。通常，掩码字符用来查找索引中的项的变体。另请参阅通配符 (wildcard character)。

MIME 类型 (MIME type)

一种因特网标准，它用来标识正在通过因特网传输的对象的类型。

基于模型的类别 (model-based category)

预定义项的分类法，它用来确定文档的主题，以便可以对该文档建立索引以及与内容类似的文档一起搜索该文档。

监视员 (monitor)

一位企业搜索用户，他有权观察集合级进程。

自然语言查询 (natural language query)

一种搜索类型，它分析所写的表达式（如“Who runs the finance department?”）而不是关键字的简单集合。

换行符 (newline character)

一个控制字符，它导致打印位置或显示位置向下移动一行。某些系统需要多个字符。

n-gram 分段 (n-gram segmentation)

一种分析方法，此方法将给定数目字符的重叠序列视为单个单词，而不是象基于 Unicode 的空格分段那样使用空格来对单词进行定界。

no-follow 伪指令 (no-follow directive)

Web 页面中的一条伪指令，它指示机器人（例如 Web 搜寻器）不要顺着那些页面中的链接前进。

no-index 伪指令 (no-index directive)

Web 页面中的一条伪指令，它指示机器人（例如 Web 搜寻器）不要将那些页面的内容包括在索引中。

Notes 远程过程调用 (Notes remote procedure call, NRPC)

Lotus Notes 的用于进行全部 Notes 到 Notes 通信的体系结构层。

操作员 (operator)

一位企业搜索用户，他有权观察、启动和停止集合级进程。

参数搜索 (parametric search)

一种搜索类型，它在指定的范围内查找包含数值或属性（例如日期、整数或其它数字数据类型）的对象。

解析器 (parser)

一个程序，它对已添加至企业搜索数据商店的文档进行解释。解析器从文档中抽取信息并对那些信息进行准备以便建立索引、进行搜索和进行检索。

位置 (place)

门户网站中的可视虚拟位置，个人和小组在这个位置接触以进行协作。在门户网站中，每个用户都有用于完成私人工作的个人位置，个人和小组能够访问各种共享位置，这些共享位置可以是公用位置或受限位置。另请参阅 Lotus QuickPlace 位置 (Lotus QuickPlace place)。

流行排名 (popular ranking)

一种排名类型，它根据文档的受欢迎程度来提高对文档的现有排名。

处理引擎归档 (processing engine archive)

一个 .pear zip 归档文件，它包含 UIMA 分析引擎以及将其用于企业搜索中的定制分析所需的所有资源。

相近搜索 (proximity search)

一种搜索类型，它在同一个句子、段落和文档中查找某些词。

代理服务器 (proxy server)

一台服务器，它充当应用程序或 Web 服务器所主管的 HTTP Web 请求的中介。在企业中，代理服务器用于替代内容服务器。

快速链接 (quick link)

URI 与关键字和短语之间的关联。

排名 (ranking)

对查询所返回的搜索结果中的每个文档指定一个整数值的过程。文档在搜索结果中的顺序基于它们与查询的相关程度。排名越高表示匹配得越接近。另请参阅动态排名 (dynamic ranking) 和静态排名 (static ranking) 。

远程联合器 (remote federator)

对一组可搜索对象执行联合的服务器联合器。

Robots Exclusion Protocol 协议 (Robots Exclusion Protocol)

一种协议，它允许 Web 站点管理员向访问机器人指出不允许该机器人访问站点的那些部分。

空间 (room)

一个程序，它允许用户创建由其他用户读取的文档、响应其他用户的意见以及查看项目状态和截止日期。用户还可以与同一个空间中的其他用户交谈。另请参阅 Lotus QuickPlace 空间 (Lotus QuickPlace room) 。

基于规则类别 (rule-based category)

规则创建的类别指定哪些文档与哪些类别相关联。例如，您可以定义规则以使包含或不包含某些词的文档或者与 URI 模式相匹配的文档与特定类别相关联。

作用域 (scope)

用于定义搜索请求范围的一组相关 URI。

搜索应用程序 (search application)

一个程序，它处理查询、搜索索引、返回搜索结果并检索企业搜索系统中的集合的源文档。

搜索高速缓存 (search cache)

一个缓冲区，它存储先前搜索请求的数据和结果。

搜索引擎 (search engine)

一个程序，它接受搜索请求并将文档列表返回给用户。

搜索索引文件 (search index files)

搜索引擎中用于存储索引的一组文件。

搜索结果 (search results)

与搜索请求相匹配的文档的列表。

安全套接字层 (Secure Sockets Layer, SSL)

提供通信保密功能的安全性协议。

安全性令牌 (security token)

关于用来授予对集合中的文档的访问权的标识及安全性的信息。不同的数据源类型支持不同类型的安全性令牌。示例包括用户角色、用户标识、组标识以及其它可以用来控制对内容的访问的信息。

种子 URL (seed URL)

搜寻起始点。

分段 (segmentation)

一个过程，路径控制通过此过程将基本信息单元划分为更小的单元 (称为 BIU 段) ，以便适应相邻服务器中较小的缓冲区大小。

servlet

一种在 Web 服务器上运行的 Java 程序，它生成动态内容来响应 Web 客户端请求，从而扩展服务器的功能。Servlet 通常用来将数据库连接到 Web。

软错误页面 (soft error page)

一种特殊的页面，它在 HTTP 服务器无法返回客户机所请求的页面时详细说明问题并将 HTTP 服务器配置为返回这些页面，而不是返回只包含头（带有指示问题的返回码）的响应。

静态排名 (static ranking)

一种排名类型，其中，文档排名因素（例如日期以及指向该文档的链接数目等等）会使排名提高。与动态排名（dynamic ranking）对照。

静态摘要 (static summarization)

一种摘要类型，其中，搜索结果包含了文档的已存储的指定摘要。与动态摘要（dynamic summarization）对照。

衍生形态查询 (stemming)

请参阅单词衍生形态查询（word stemming）。

无用词 (stop word)

被搜索应用程序忽略的常用词，例如 *the*、*an* 或 *and*。

无用词删除 (stop word removal)

一个过程，此过程从查询中除去无用词以忽略常用词并返回更相关的结果。

摘要 (summarization)

在搜索结果中提供简要描述文档内容的句子的过程。另请参阅动态摘要（dynamic summarization）和静态摘要（static summarization）。

同义词字典 (synonym dictionary)

一种字典，它使用户在搜索集合时能够搜索查询项的同义词。

分类法 (taxonomy)

根据相似性将对象分类到组中。在企业搜索中，分类法将数据组织成类别和子类别。另请参阅类别树（category tree）。

文本分析 (text analysis)

从文本中抽取语义和其它信息以提高集合数据可检索性的过程。

文本分析引擎 (text analysis engine)

一种软件组件，它负责查找和表示文本中的上下文和语义内容。

基于文本的打分 (text-based scoring)

对文档指定一个整数值的过程，该整数值表示该文档与查询中的项的相关程度。这个整数值越大，与查询的匹配程度就越接近。另请参阅动态排名（dynamic ranking）。

主题抽取 (theme extraction)

一种概念抽取类型，它自动识别文本文档中的重要词汇表词条以抽取文档的主题。另请参阅概念抽取（concept extraction）。

标记 (token)

基本文本单元，企业搜索对这些文本单元建立索引。标记可以是语言中的单词，也可以是其它适合于建立索引的文本单元。

记号化器 (tokenizer)

一个文本分段程序，它扫描文本并确定是否以及何时可以将一系列字符识别为标记。

末尾字符 (trailing character)

单词的最后一个字符。

基于 Unicode 的空格分段 (Unicode-based white space segmentation)

一种标记化方法，此方法使用 Unicode 字符属性来区分标记和分隔符。

统一资源标识 (Uniform Resource Identifier, URI)

一个简洁的字符串，它标识抽象资源或物理资源。

统一资源定位符 (Uniform Resource Locator, URL)

一个字符序列，它表示计算机上的或网络（如因特网）中的信息资源。这个字符序列包括用来访问该信息资源的协议的简写名以及该协议定位该信息资源时使用的信息。

通用资源名 (Universal Resource Name, URN)

一种因特网协议元素，它包含符合某种语法的短字符串。该字符串由可以用于引用资源的名称或地址组成。

Unstructured Information Management Architecture (UIMA)

一种 IBM 体系结构，它定义无结构数据的分析系统的实施框架。

用户代理 (user agent)

一个应用程序，它浏览 Web 并将关于它自己的信息留在它所访问的站点上。在企业搜索中，Web 搜寻器就是一个用户代理。

Web 搜寻器 (Web crawler)

一类机器人软件，此软件通过检索 Web 文档并顺着该文档中的链接前进来探查 Web。

含权项搜索 (weighted term search)

一种查询，其中，某些项被指定了更高的重要性。

通配符 (wildcard character)

一个字符，它用来表示位于搜索项开头、中间或末尾的可选字符。

单词衍生形态查询 (word stemming)

一个语言规范化过程，在此过程中，将词的变体形式还原为公共形式。例如，诸如 *connections*、*connective* 和 *connected* 之类的单词将还原为 *connect*。

XML 路径语言 (XML Path Language, XPath)

一种语言，它唯一地标识源 XML 文档部分或者对那些部分进行寻址。XPath 还提供了基本设施来处理字符串、数字和布尔运算符。

访问关于 WebSphere Information Integration 的信息

您可以通过打电话或浏览 Web 获得有关 WebSphere Information Integration 产品的信息。

此处提供的电话号码仅在美国有效:

- 要订购产品或了解一般信息: 1-800-IBM-CALL (1-800-426-2255)
- 要订购出版物: 1-800-879-2755

您也可以在 Web 上找到有关 WebSphere Information Integration 的信息, 网址为 www.ibm.com/software/data/integration/db2ii/。此站点提供了下列方面的最新信息:

- 产品文档
- 产品下载
- 修订包
- 发行说明和其它支持文档
- 关于 WebSphere Information Integration 的新闻
- 链接到 Web 资源, 例如白皮书和 IBM Redbooks™
- 链接到新闻组 and 用户组
- 链接到 WebSphere Information Integration 产品的在线信息中心
- 订购书籍

要访问产品文档:

1. 访问 Web 地址 www.ibm.com/software/data/integration/db2ii/。
2. 从下拉列表中选择产品, 例如 WebSphere Information Integrator OmniFind Edition。
3. 单击页面左边的 Support 链接。
4. 在 Learn 部分中, 选择要使用的链接。如果您选择的产品有信息中心, 则可选择信息中心的链接。请参阅第 270 页的图 2 以获取示例。

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](#)

图 2. WebSphere Information Integration Support Web 站点上的产品文档链接示例

提供对文档的意见

请发送您对本资料或其它 IBM WebSphere Information Integration 文档的任何意见。

您的反馈有助于 IBM 提供高质量的信息。请发送您对本资料或其它 WebSphere Information Integration 文档的任何意见。您可以使用下列任何方法来提供意见：

1. 使用 www.ibm.com/software/awdtools/rcf/ 上的在线读者意见表发送您的意见。
2. 通过电子邮件将您的意见发送至 ctscrcf@cn.ibm.com。确保包括产品名、产品的版本号 and 资料的名称及部件号（如果适用的话）。如果您对特定文本有意见，请包括此文本的位置（例如，标题、表号或页码）。

与 IBM 联系

在中国，请致电下列其中一个号码以与 IBM 联系：

- 800-810-1818 或 (010) 84981188 分机 5151，可获得售前客户服务；
- 800-810-1818 或 (010) 84981188 分机 5200，可获得售后客户服务；
- 800-810-1818 或 (010) 84981188 分机 5017，可获得市场营销与销售的信息；

要查找您所在国家或地区的 IBM 营业处，可在网上查看 IBM 全球联系人目录 (Directory of Worldwide Contacts)，网址为：www.ibm.com/planetwide。

商标

本主题列示 IBM 商标和某些非 IBM 商标。

请参阅<http://www.ibm.com/legal/copytrade.shtml>以了解有关 IBM 商标的信息。

下列各项是其他公司的商标或注册商标：

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国和 / 或其他国家或地区的商标或注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和 / 或其他国家或地区的商标。

Intel、Intel Inside（徽标）、MMX 和 Pentium 是 Intel Corporation 在美国和 / 或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Linux 是 Linus Torvalds 在美国和 / 或其他国家或地区的商标。

其他公司、产品或服务名称可能是其他公司的商标或服务标记。

声明

本信息是为在美国提供的产品和服务编写的。IBM 可能未在所有国家或地区提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

有关双字节（DBCS）信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本资料中描述的产品和 / 或程序进行改进和 / 或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：（i）允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及（ii）允许对已经交换的信息进行相互使用，请与下列地址联系：

IBM Corporation J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

所有关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含日常商业运作所使用的数据和报表的示例。为了尽可能全面地说明这些数据和报表，这些示例包括个人、公司、品牌和产品名。所有这些名称都是虚构的，如与实际商业企业所使用的名称和地址有任何雷同，纯属巧合。

版权许可:

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口（API）进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。用户如果是为了按照 IBM 应用程序编程接口开发、使用、经销或分发应用程序，则可以任何形式复制、修改和分发这些样本程序，而无须向 IBM 付费。

凡这些样本程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明:

Outside In (®) Viewer Technology, ©1992-2005 Stellent, Chicago, IL., Inc. All Rights Reserved.

IBM XSLT Processor Licensed Materials - Property of IBM ©Copyright IBM Corp., 1999-2005. All Rights Reserved.

索引

[A]

- 安排时间表
 - 搜寻器 31
 - 索引构建 116, 117
- 安全性
 - 标识管理 174
 - 重复文档检测 169
 - 对集合启用 26, 165
 - 对企业应用程序禁用 168
 - 访问控制 167
 - 管理角色 167
 - 集合级 169, 178
 - 锚点文本分析 170
 - 描述 165
 - 绕过文档级访问控制 178
 - 认证 167, 168
 - 搜索应用程序标识 171
 - 搜寻器插件 81
 - 文档级 172, 173, 174, 178
 - 样本搜索应用程序 158
 - 用户概要文件 174
 - Lotus Domino 文档 177
 - WebSphere 全局安全性 168
 - Windows 域 175
- 安全性令牌
 - 对集合禁用 178
 - 搜寻器配置 172
 - 文档级安全性 172, 182
 - Portal Search Engine 处理 182

[B]

- 备份脚本
 - 描述 225
 - 运行 225, 227
- 备份企业搜索 225
- 本机 XML 搜索 106
- 编辑
 - 集合 27
 - 数据侦听器应用程序 80
 - 搜索应用程序属性 149, 157
 - 搜寻空间 34
 - 搜寻器属性 34
- 标识管理
 - 配置 174
 - 用户概要文件 174
- 并发索引构建 118
- 并行索引构建 118
- 部署
 - 企业搜索适配器 185

- 部署 (续)
 - 注册 portlet 186
 - ESPACServer.ear 文件 77
 - ESSearchAdapterPortlet.war 文件 186
 - ESSearchAdapter.ear 文件 185
 - ESSearchPortlet.war 文件 183
 - “搜索和浏览”portlet 184
 - “搜索”portlet 183

[C]

- 参数搜索 119
- 操作员
 - 角色配置 167
 - 描述 166
- 插件, 用于搜寻器 81
- 查看
 - 日志文件 223
 - URI 详细信息 201
- 查询验证 173
- 查找企业搜索文档 255
- 常用的查询, 监视 213
- 重复文档检测
 - 描述 165
 - 启用安全性 169
 - 全局分析 169
 - 升降类配置 141
- 重组索引 5, 115
- 创建
 - 基于规则的类型 94
 - 集合 25, 26
 - 快速链接 134
 - 搜寻器 33
 - 作用域 123
 - HTML 搜索字段 99
 - Web 搜寻器报告 206
 - XML 搜索字段 96
- 从索引中除去 URI 126
- 错误消息
 - 查看日志文件 223
 - 接收电子邮件, 对于 219, 221
 - SMTP 服务器配置 221

[D]

- 代理服务器 72
- 登录到管理控制台 16
- 低重新调用查询
 - 描述 141
 - 缺省升降因子 144

- 电子邮件通知
 - 对于警报 221
 - 对于消息 221
 - SMTP 服务器配置 221
- 定制文本分析
 - 将分析结果映射至索引 103
 - 将分析结果映射至 JDBC 表 104
 - 描述 99
 - 文本分析引擎 100, 101
 - 映射公共分析结构 103, 104
 - 映射 XML 元素 102
- 动态排名 137
- 动态摘要 (dynamic summarization) 132
- 端口号, 企业搜索 21
- 对搜索结果进行排名
 - 动态 137
 - 基于文本的打分 137
 - 静态 137
 - 描述 137
 - 升降类 141, 143, 144
 - 升降字典 139
 - URI 模式 140

[F]

- 发现 3
- 返回码, 企业搜索 229
- 防火墙, 搜寻 Exchange Server 文档 51
- 访问控制
 - 当前验证 173
 - 对集合禁用 178
 - 描述 167
 - 文档级安全性 172, 182
 - 用户概要文件 174
 - Lotus Domino 的要求 177
 - Portal Search Engine 支持 182
 - Windows 文件系统的要求 175
- 分类法管理 Portlet 189
- 分类法, 从 WebSphere Portal 迁移 189, 190
- 分类规则
 - 配置 94
 - 文档内容 91, 94
 - URI 模式 91, 94
- 分类类型
 - 基于规则的 91
 - 基于模型的 93
 - 选择 26, 94
- 服务器方式, Content Edition 存储库 36
- 辅助功能选项 257
- 复合项, 解析 105

- 复原
 - 从备份 226
 - 系统文件至新服务器 227
- 复原脚本
 - 描述 225
 - 运行 226, 227
- 复原企业搜索 225

[G]

- 高重新调用查询
 - 描述 141
 - 缺省升降因子 144
- 公共分析结构
 - 将 XML 元素映射至 102
 - 描述 99
 - 映射至索引 103
 - 映射至 JDBC 表 104
- 估算系统资源 199
- 管理角色
 - 操作员 166, 167
 - 集合管理员 166, 167
 - 监视员 166, 167
 - 描述 166
 - 配置 167
 - 企业搜索管理员 166, 167
- 管理控制台
 - 登录 16
 - 界面 13
 - 描述 7
 - 任务总结 13
- 管理员密码
 - 在单服务器上更改 17
 - 在多台服务器上更改 18

[H]

- 韩国语
 - 复合项分析 105
 - n-gram 分段 107
- 会话标识, 企业搜索 229
- 活动的 Web 站点, 监视 204, 205

[J]

- 基于表单的认证 70, 71
- 基于规则的分类法, 从 WebSphere Portal 迁移 189, 190
- 基于规则的分类
 - 创建 94
 - 描述 91
 - 选择分类类型 94
- 基于模型的分类法, 从 WebSphere Portal 迁移 189

- 基于模型的类别
 - 描述 93
 - 选择分类类型 94
- 基于文本的打分 137
- 集合
 - 安全性 169
 - 编辑 27
 - 草稿 25
 - 重复文档安全性 169
 - 创建方式 25
 - 从 WebSphere Portal 迁移 189, 190
 - 估算大小 26
 - 估算资源 199
 - 监视 201
 - 解析 89
 - 联合 25
 - 锚点文本安全性 170
 - 描述 2
 - 缺省迁移设置 192
 - 绕过文档级访问控制 178
 - 删除 28
 - 使用集合向导创建 25
 - 使用“集合”视图来创建 26
 - 搜索 119
 - 搜索服务器 127
 - 系统状态 201
 - 应用程序标识安全性 171
 - 与搜索应用程序相关联 147
 - MigrationWizard.log 文件 190
- 集合标识, 语法规则 26
- 集合管理员
 - 角色配置 167
 - 描述 166
- 集合级安全性
 - 重复文档检测 169
 - 锚点文本分析 170
 - 描述 165, 169
 - 启用 26
 - 应用程序标识 171
- 集合向导 25
- 监视
 - 常用的查询 213
 - 集合 201
 - 解析器 210
 - 企业搜索 199
 - 日志文件 223
 - 数据侦听器 214
 - 搜索服务器 213
 - 搜寻器 203
 - 响应时间历史 213
 - 已废弃的文档 215
 - 最近的查询 213
 - URI 详细信息 201
 - Web 搜寻器 204
 - Web 搜寻器活动站点 205
 - Web 搜寻器搜寻速率 206

- 监视 (续)
 - Web 搜寻器线程详细信息 205
- 监视员
 - 角色配置 167
 - 描述 166
- 检查系统资源 200
- 脚本
 - esbackup.bat 225, 227
 - esbackup.sh 225, 227
 - escrcm.sh 46
 - escrcm.vbs 48
 - escrdb2.sh 44
 - escrdb2.vbs 45
 - escrnote.sh 54
 - escrnote.vbs 56
 - escrvbr.sh 37
 - escrvbr.vbs 38
 - esrestore.bat 226, 227
 - esrestore.sh 226, 227
 - startccl 226
- 解析器
 - 本机 XML 搜索 106
 - 复合项分析 105
 - 换行符 107
 - 监视 210
 - 解析器服务的文档类型 108, 109
 - 解析文档类型 107
 - 空格 107
 - 描述 4, 89
 - 启动 210
 - 数据分析任务 4
 - 停止 210
 - 系统状态 210
 - 线程 105
 - n-gram 分段 107
 - Stellent 会话的文档类型 110, 112
- 解析器服务器
 - 启动 195
 - 停止 197
 - 线程配置 105
- 禁用索引时间表 117
- 尽可能快地重新访问 URL 69
- 尽可能快地访问 URL 69
- 警报
 - 电子邮件选项 218, 219
 - 服务器上的可用空间 219
 - 集合级 199, 218
 - 接收电子邮件, 对于 221
 - 描述 217
 - 搜索响应时间 218
 - 搜寻的文档 218
 - 索引限制 199
 - 系统级 219
 - 已建立索引的文档 218
 - SMTP 服务器配置 221

静态排名 (static ranking)
对集合启用 26
描述 137
在迁移的集合中 190

[K]

可用空间警报 219
快速链接
创建 134
描述 134
搜索 134
URI 格式 82
快速链接中的关键字 134

[L]

类别
创建 94
从 WebSphere Portal 迁移 189, 190
分类类型 94
基于规则的 91, 94
基于模型的 93, 94
类别树 93
描述 90
嵌套子类别 93
搜索 90
URI 格式 82
类别树
从 WebSphere Portal 迁移 189, 190
描述 93
联合集合 25

[M]

锚点文本分析
对文档建立索引 170
集合安全性 170
描述 165
全局分析 170
密码保护的 Web 站点 70
基于表单的认证 71
HTTP 基本认证 71
密码, 企业搜索管理员 17, 18
命令, 企业搜索 229

[Q]

启动
解析器 210
企业搜索服务器 195
迁移向导 189, 190
数据侦听器 214
搜索服务器 213
搜寻器服务器 203

启动 (续)
索引构建 211
样本搜索应用程序 158
企业搜索
安全性 165
备份 225
备份脚本 225
从备份复原 226
端口号配置 21
返回码 229
复原脚本 225
复原系统文件 227
概述 1
管理角色 166
管理控制台 7
会话标识 229
集合级安全性 169
监视 199
解析器 4, 89
命令 229
启动服务器 195
日志文件 217
数据流图 10
搜索服务器 6, 127
搜索应用程序 9
搜寻器服务器 3, 31
索引服务器 5, 115
停止服务器 195, 197
文档级安全性 172
与 WebSphere Portal 集成 181
组件 2
API 9
URI 格式 82
企业搜索管理员
角色配置 167
描述 166
在单服务器上更改密码 17
在多台服务器上更改密码 18
企业搜索适配器 185
企业应用程序
ESAdmin 应用程序 168
ESPACServer.ear 文件 77
ESSearchAdapter.ear 文件 185
ESSearchApplication 应用程序 168
ESSearchServer 应用程序 168
启用索引时间表 117
迁移
基于规则的分类法 190
基于模型的分类法 189
集合 190
迁移向导
基于规则的分类法 190
基于模型的分类法 189
集合 190
描述 189
启动 189, 190

迁移向导 (续)
缺省集合设置 192
缺省搜寻器设置 192
日志文件 193
全局分析
重复文档检测 165, 169
锚点文本分析 165, 170
描述 5
全局 Web 搜寻空间 74
缺省搜索应用程序 157

[R]

任务总结, 管理控制台 13
认证
对企业应用程序禁用 168
描述 167
日语
解析换行符 107
n-gram 分段 107
日志文件
查看 223
电子邮件选项 221
过滤 223
监视 223
描述 217
迁移向导 193
缺省位置 217
循环使用 219
严重性级别 219
用于进行文档跟踪 215
最大大小 219
SMTP 服务器配置 221
软错误页面, Web 搜寻器 69

[S]

删除
队列中的索引 212
集合 28
搜寻器 35
升降类
重复文档检测 141
低重新调用查询 141, 144
高重新调用查询 141, 144
将字段映射至 143
描述 141
配置 143, 144
缺省值 144
升降因子
对 URI 模式 140
升降类的 143, 144
升降类配置 141, 144
升降字字典的 138

- 升降字典
 - 描述 138
 - 添加到系统 139
 - 与集合相关联 139
- 事件发布
 - 描述 39
 - 在 AIX 操作系统上设置 44
 - 在 Linux 操作系统上设置 44
 - 在 Solaris 操作环境中设置 44
 - 在 Windows 上设置 45
 - DB2 搜寻器配置 40, 42
- 数据流, 企业搜索系统 10
- 数据源类型
 - 关系数据库 39
 - 企业搜索支持的 2
 - 支持外部 2, 9
 - Content Edition 存储库 35, 36
 - DB2 数据库 39
 - DB2 Content Manager 项类型 46
 - Domino Document Manager 数据库 49
 - Exchange Server 公共文件夹 50
 - NNTP 新闻组 51
 - Notes 数据库 52, 53
 - QuickPlace 数据库 60
 - UNIX 文件系统 62
 - Web 站点 63
 - WebSphere Portal 站点 77
 - Windows 文件系统 79
- 数据侦听器
 - 重新启动 80, 214
 - 监视 214
 - 配置 80
- 数据侦听器 API 9
- 刷新索引 5, 115
- 搜索
 - 集合 119
 - 快速链接 134
 - 类别 90
 - 外部源 119
 - HTML 文档 98, 99
 - XML 文档 96, 102
- 搜索服务器
 - 常用的查询 213
 - 关联升降字典 139
 - 关联同义词字典 130
 - 关联无用词字典 131
 - 监视 213
 - 描述 6, 127
 - 启动 195, 213
 - 升降字典 138
 - 搜索高速缓存 127, 128
 - 停止 197, 213
 - 同义词字典 128, 129
 - 无用词字典 130
 - 系统状态 213
 - 响应时间历史 213
- 搜索服务器 (续)
 - 最近的查询 213
- 搜索高速缓存
 - 描述 127
 - 配置 128
- 搜索和索引 API 9, 147
- 搜索结果
 - 定制摘要 132, 133
 - 动态排名 137
 - 动态摘要 (dynamic summarization) 132
 - 分组 124, 125
 - 基于文本的打分 137
 - 静态排名 (static ranking) 137
 - 描述 137
 - 排名 140
 - 升降类配置 141, 143, 144
 - 通配符 121
 - 通配符扩展 122
 - 摘要 132, 133
 - 折叠 124, 125
 - URI 模式配置 140
- 搜索响应时间
 - 监视 213
 - 警报, 关于 218
- 搜索选项
 - 参数搜索 119
 - 搜索结果的 119
 - 文档内容 119
 - 字段搜索 119
 - 自由文本搜索 119
- 搜索应用程序
 - 定制 147
 - 集合级安全性 171
 - 描述 9
 - 样本 147, 148
 - 应用程序标识 171
 - 与集合相关联 147
 - 与外部源相关联 162
- 搜寻空间
 - 编辑 34
 - 警报, 关于 218
 - 描述 3
 - Web 搜寻器配置 65
- 搜寻器
 - 安排时间表 31
 - 编辑搜寻空间 34
 - 编辑搜寻器属性 34
 - 插件 81
 - 创建 33
 - 的初始值 33
 - 基本值 31
 - 监视 203
 - 描述 3
 - 配置概述 31
 - 启用文档级安全性 31
- 搜寻器 (续)
 - 缺省迁移设置 192
 - 删除 35
 - 数据侦听器应用程序 80
 - 文档级安全性 172
 - 系统状态 203
 - 支持外部 9, 80
 - 组合搜寻器类型 31
 - Content Edition 35, 36
 - DB2 39
 - DB2 Content Manager 46
 - Domino Document Manager 49
 - Exchange Server 50, 51
 - NNTP 51
 - Notes 52, 53
 - QuickPlace 60
 - UNIX 文件系统 62
 - URI 格式 82
 - Web 63
 - WebSphere Portal 77, 78
 - Windows 文件系统 79
- 搜寻器插件 81
- 搜寻器服务器
 - 启动 195, 203
 - 停止 197, 203
- 搜寻器类型
 - 基本值 31
 - 组合在集合中 31
- 搜寻器历史记录报告
 - 创建 206
 - 描述 204
 - 站点报告 206
 - HTTP 返回码报告 206
- 搜寻器属性
 - 编辑 34
 - 描述 3
- 搜寻速率, 监视 206
- 索引
 - 安排时间表 116
 - 并发构建 118
 - 并行构建 118
 - 除去 URI 118, 126
 - 从队列中删除 212
 - 更改时间表 117
 - 监视 211, 212
 - 禁用时间表 117, 211
 - 警报, 关于 218
 - 锚点文本 170
 - 描述 5, 115
 - 启用时间表 117, 211
 - 通配符 118, 121, 122
 - 折叠的 URI 118, 124, 125
 - 作用域 118, 123
 - URI 格式 82
- 索引重组
 - 安排时间表 116, 117

索引重组 (续)
 描述 5, 115
索引队列 (index queue) 212
索引服务器
 启动 195
 停止 197
索引构建
 并发 118
 并行 118
 启动 211
 停止 211, 212
 系统状态 212
索引刷新
 安排时间表 116, 117
 描述 5, 115

[T]

停止
 解析器 210
 企业搜索服务器 195, 197
 搜索服务器 213
 搜寻器服务器 203
 索引构建 211, 212
通配符
 查询扩展 121, 122
 索引扩展 121, 122
同义词典
 描述 128
 添加到系统 129
 与集合相关联 130

[W]

外部搜寻器
 配置 80
 数据侦听器应用程序 80
 数据侦听器 API 9
外部源
 描述 161
 配置 161
 搜索 119
 应用程序标识安全性 171
 与搜索应用程序相关联 162
文本处理
 公共分析结构 99
 文本分析引擎 99
 注释器 99
文本分析
 公共分析结构 103, 104
 文本分析引擎 100, 101
 映射 XML 元素 102
文本分析引擎
 描述 99
 添加到系统 100

文本分析引擎 (续)
 映射分析结果 103, 104
 映射 XML 元素 102
 与集合相关联 101
文档 255
文档重要性
 对集合启用 26
 静态 137
 升降类 141, 144
 升降字典 139
 在迁移的集合中 190
 URI 模式 140
文档跟踪
 报告 215
 禁用 215
 描述 214
 启用 215
 日志文件配置 215
文档级安全性
 安全性令牌 172
 标识管理 174
 当前凭证验证 173
 对 Lotus Domino 文档 177
 对 Windows 文件系统 175
 描述 165, 172
 实时验证 173
 搜寻器插件 81
 搜寻器配置 31
 已建立索引的访问控制 172
 用户概要文件 174
 Lotus Domino 文档 177
 Portal Search Engine 支持 182
文档类型
 解析 107
 解析器服务 108, 109
 Stellent 会话 110, 112
文档内容选项 119
文档摘要
 编辑属性 133
 定制 132
文件扩展名
 从 Web 搜寻空间中排除 65
 集合解析器支持的 108, 109
 Stellent 会话支持的 110, 112
无用词字典
 描述 130
 添加到系统 131
 与集合相关联 131

[X]

系统备份 225
系统复原 225, 226, 227
系统状态
 集合 201
 解析器 210

系统状态 (续)
 搜索服务器 213
 搜寻器 203
 索引构建 212
 Web 搜寻器 204
系统资源
 估算 199
 检查 200
线程
 解析器 105
 Web 搜寻器 205
线程详细信息, 监视 204
限制 Web 搜寻空间 65
响应时间历史, 监视 213

[Y]

验证当前凭证 173, 175, 177
样本搜索应用程序
 访问 158
 禁用安全性 168
 描述 9, 148
 启动 158
 启用安全性 158
 缺省部署 157
 搜索功能 147, 148
 config.properties 文件 149, 157
 WebSphere 全局安全性 158
已废弃的文档
 报告, 关于 215
 描述 214
 日志文件配置 215
应用程序标识 171
映射
 公共分析结构至索引 103
 公共分析结构至 JDBC 表 104
 字段至升降类 143
 HTML 搜索字段 99
 XML 搜索字段 96
 XML 元素到公共分析结构 102
用户代理 64
用户概要文件
 描述 174
 配置 174
用于 Web 搜寻的 cookie
 格式 73
 描述 73
 配置 74
语言支持
 本机 XML 搜索 106
 定制文本分析 99
 升降字典 138
 同义词典 128
 无用词字典 130
 语义搜索 99, 106
 语义搜索 99, 102, 106

[Z]

摘要

定制 132, 133

动态 132

站点历史记录报告

创建 206

描述 204

折叠的搜索结果

描述 124

配置 125

折叠的 URI

描述 124

配置 125

直接方式, Content Edition 存储库 36

中文

解析换行符 107

n-gram 分段 107

注释器 99

字段搜索 119

字段, 映射至升降类 143

自由文本搜索 119

最大重新搜寻时间间隔 68

最近的查询, 监视 213

最近搜寻的 URL, 监视 204

最小重新搜寻时间间隔 68

作用域

创建 123

描述 123

搜索 123

URI 格式 82

URI 模式 123

A

AIX 操作系统

事件发布配置 44

Content Edition 搜寻器配置 37

DB2 Content Manager 搜寻器配置 46

Domino Document Manager 搜寻器配置

54

Notes 搜寻器配置 54

QuickPlace 搜寻器配置 54

API

描述 9

数据侦听器 9

搜索和索引 9, 147

B

bos.iocp.rte 模块 58

C

CCLServer_date.log 文件 21

config.properties 文件

编辑 157

属性描述 149

Content Edition 搜寻器

服务器方式 36

配置 35

在 AIX 操作系统上设置 37

在 Linux 操作系统上设置 37

在 Solaris 操作环境中设置 37

在 Windows 上设置 38

直接方式 36

URI 格式 82

cookies.ini 文件

格式 73

描述 73

配置 74

crawl.rules 文件 74

D

DB2 搜寻器

配置 39

事件发布 39

事件发布配置 44, 45

URI 格式 82

WebSphere II Event Publisher Edition

配置 40

WebSphere MQ 安装 44

WebSphere MQ 配置 42

Windows 上的 WebSphere MQ 安装

45

DB2 Content Manager 搜寻器

配置 46

在 AIX 操作系统上设置 46

在 Linux 操作系统上设置 46

在 Solaris 操作环境中设置 46

在 Windows 上设置 48

URI 格式 82

DIHOP 协议, 搜寻器配置 57

Domino 用户配置, QuickPlace 搜寻器 61

Domino Document Manager 搜寻器

配置 49

在 AIX 操作系统上设置 54

在 Linux 操作系统上设置 54

在 Solaris 操作环境中设置 54

在 Windows 上设置 56

DIHOP 协议配置 57

IOCP 配置 58

NRPC 协议 54, 56

URI 格式 82

E

EAR 文件

ESAdmin 应用程序 168

EAR 文件 (续)

ESPACServer.ear 77

ESSearchAdapter.ear 185

ESSearchApplication 应用程序 168

ESSearchServer 应用程序 168

esadmin 命令 229

ESAdmin 应用程序

登录到 16

禁用安全性 168

esbackup.bat 脚本 225, 227

esbackup.sh 脚本 225, 227

eschangepw 脚本 17, 18

escrcm.sh 脚本 46

escrcm.vbs 脚本 48

escrdb2.sh 脚本 44

escrdb2.vbs 脚本 45

escrnote.sh 脚本 54

escrnote.vbs 脚本 56

escrvbr.sh 脚本 37

escrvbr.vbs 脚本 38

ESPACServer.ear 文件 77

esrestore.bat 脚本 226, 227

esrestore.sh 脚本 226, 227

ESSearchAdapterPortlet.war 文件 186

ESSearchAdapter.ear 文件 185

ESSearchApplication 应用程序

重新启动 157, 158

禁用安全性 168

启用安全性 158

config.properties 文件 149, 157

ESSearchPortlet.war 文件 183

ESSearchServer 应用程序

禁用安全性 168

es.cfg 文件 17, 18

ES_INSTALL_ROOT, 描述 17, 18

ES_NODE_ROOT, 描述 17, 18

es_special_field.default_field 144

es_special_field.regular_text field 144

Exchange Server 搜寻器

安全文档 51

配置 50

URI 格式 82

F

followindex.rules 文件

描述 76

配置 76

G

global.rules 文件 74

H

- HTML 搜索字段
 - 创建 99
 - 将元素映射至 98, 99
 - 描述 98
- HTML 文档, 搜索 98
- HTTP 代理服务器 72
- HTTP 返回码
 - 由 Web 搜寻器接收 206
 - Web 搜寻器报告 206
- HTTP 基本认证 70, 71

I

- IOCP, 搜寻器配置 58
- I/O 完成端口模块, 搜寻器配置 58

J

- Java connector for DB2 Content Manager 46, 48
- JDBC 外部源
 - 编辑 161
 - 配置 161
 - 删除 161
 - JDBC 驱动程序 161

L

- LDAP 外部源
 - 编辑 161
 - 配置 161
 - 删除 161
- Linux 操作系统
 - 事件发布配置 44
 - Content Edition 搜寻器配置 37
 - DB2 Content Manager 搜寻器配置 46
 - Domino Document Manager 搜寻器配置 54
 - Notes 搜寻器配置 54
 - QuickPlace 搜寻器配置 54
 - Solaris 操作环境
 - 事件发布配置 44
 - Lotus Domino 域 177
 - Lotus Domino Trusted Server 177

M

- MIME 类型, 包括在 Web 搜寻空间中 65

N

- NNTP 搜寻器, 配置 51

- Notes 搜寻器
 - 配置 52
 - 使用技巧 53
 - 文档级安全性配置 177
 - 验证当前凭证 177
 - 在 AIX 操作系统上设置 54
 - 在 Linux 操作系统上设置 54
 - 在 Solaris 操作环境中设置 54
 - 在 Windows 上设置 56
 - 字段映射规则 53
 - DIIOP 协议配置 57
 - IOCP 配置 58
 - Lotus Domino Trusted Server 177
 - NRPC 协议 54, 56
 - URI 格式 82
- no-follow 伪指令
 - 描述 76
 - 配置 76
- no-index 伪指令
 - 描述 76
 - 配置 76
- NRPC 协议, 搜寻器配置 54, 56
- n-gram 分段 107

P

- parserTypes.cfg 文件 108
- PDF 文档 255
- Portal Search Engine
 - 传递安全性令牌 182
 - 描述 181
 - 文档级安全性 182
- portlet
 - 分类法管理 Portlet 189
 - 描述 181
 - 企业搜索 181
 - 搜索 183
 - 搜索和浏览 184
 - 注册 186
 - ESSearchAdapterPortlet.war 文件 186
 - WebSphere Portal 的受支持版本 181

Q

- QuickPlace 搜寻器
 - 配置 60
 - 在 AIX 操作系统上设置 54
 - 在 Linux 操作系统上设置 54
 - 在 Solaris 操作环境中设置 54
 - 在 Windows 上设置 56
 - DIIOP 协议配置 57
 - Domino 用户配置 61
 - IOCP 配置 58
 - NRPC 协议 54, 56
 - URI 格式 82

- QuickPlace 搜寻器 (续)
 - “本地用户”安全性 61
 - “目录辅助”配置 62

R

- Robots Exclusion Protocol 协议 64
- robots.txt 文件 64
- runtime-generic.properties 文件 133

S

- SIAPI (搜索和索引 API) 9, 147
- SMTP 服务器配置 221
- Solaris 操作环境
 - Content Edition 搜寻器配置 37
 - DB2 Content Manager 搜寻器配置 46
 - Domino Document Manager 搜寻器配置 54
 - Notes 搜寻器配置 54
 - QuickPlace 搜寻器配置 54
- startccl 脚本 226
- Stellent 会话
 - 关联文档类型 110
 - 解析文档类型 107
 - 缺省文档类型 112
- stellentypes.cfg 文件 110
- stellent.properties 文件 110
- synonyms.xml 文件 189

T

- titles.xml 文件 189
- treenodes.xml 文件 189
- Trusted Server 配置 177

U

- UIMA
 - 公共分析结构 103, 104
 - 将分析结果映射至索引 103
 - 将分析结果映射至 JDBC 表 104
 - 将文本分析引擎添加至系统 100
 - 描述 99
 - 映射 XML 元素 102
 - 与集合相关联 101
- UNIX 文件系统搜寻器
 - 配置 62
 - URI 格式 82
- URI
 - 查看详细信息, 关于 201
 - 从索引中除去 126
 - 分类规则 91, 94
 - 快速链接 134
 - 企业搜索中的格式 82

URI (续)

- 搜索结果中折叠的 124, 125
- 影响静态分数 140
- 作用域 123

URI 详细信息

- 监视 201
- 已废弃的文档 215

URL 路径深度 65

V

vbr_access_services.jar 文件 37, 38

W

Web 搜寻器

- 重新搜寻时间间隔 68
- 创建报告, 关于 206
- 代理服务器 72
- 活动站点 204, 205
- 监视 204
- 尽可能快地访问 URL 69
- 密码保护的 Web 站点 70, 71
- 配置 63
- 起始 URL 65, 69
- 全局搜寻空间 74
- 软错误页面 69
- 搜寻规则 65
- 搜寻器历史记录 204
- 搜寻速率 206
- 系统状态 204
- 线程详细信息 204, 205
- 限制搜寻空间 65
- 用户代理 64
- 站点历史记录 204
- 最近搜寻的 URL 204
- cookie 73
- cookie 格式 73
- cookie 配置 74
- followindex.rules 文件 76
- JavaScript 支持 64
- no-follow 伪指令 76
- no-index 伪指令 76
- URL 状态 204

Web 搜寻器的重新搜寻时间间隔 68

Web 搜寻器的地址规则 65

Web 搜寻器的起始 URL 65, 69

Web 搜寻器的前缀规则 65

Web 搜寻器的域规则 65

Web 搜寻器的 IP 地址规则 65

Web 搜寻器中的 JavaScript 支持 64

WebSphere 全局安全性

- 禁用 168
- 搜索应用程序属性 157

WebSphere II Event Publisher

Edition, DB2 搜寻器配置 40

WebSphere II OmniFind Edition 257

标识管理 174

端口号配置 21

返回码 229

辅助功能选项 257

概述 1

管理控制台 7

会话标识 229

解析器 4

命令 229

数据流图 10

搜索服务器 6

搜索应用程序 9

搜寻器服务器 3

索引服务器 5

与 WebSphere Portal 集成 181

在单服务器上更改密码 17

在多台服务器上更改密码 18

组件 2

API 9

WebSphere MQ, 搜寻器服务器配置 44, 45

WebSphere MQ, DB2 搜寻器配置 42

WebSphere Portal

分类法管理 Portlet 189

分类法迁移 189

基于模型的分类法 189

集合迁移 189

类别树迁移 189

缺省迁移设置 192

与企业搜索集成 181

“搜索和浏览”portlet 184

“搜索”portlet 部署 183

WebSphere Portal 搜索中心

描述 181

企业搜索适配器 181, 185

注册 portlet 181, 186

WebSphere Portal 搜寻器

部署 ESPACServer.ear 文件 77

复制站点 URL 78

配置 77

企业应用程序部署 77

URI 格式 82

Windows 操作系统

事件发布配置 45

搜寻器配置 56

Content Edition 搜寻器配置 38

DB2 Content Manager 搜寻器配置 48

Windows 文件系统搜寻器

配置 79

文档级安全性配置 175

URI 格式 82

Windows 域 175

WpsMigratorLog.log 文件 193

X

XML 查询语法, 本机 106

XML 片段, 本机 XML 搜索 106

XML 搜索字段

创建 96

将元素映射至 96, 102

描述 96, 102

XML 文档

本机 XML 搜索 106

搜索 96

XML 元素

搜索 96, 102

映射至公共分析结构 102

映射至搜索字段 96

XPath, 本机 XML 搜索 106

[特别字符]

“安全性”视图, 描述 13

“本地用户”安全性, QuickPlace 搜寻器 61

“集合”视图

创建集合 26

描述 13

“监视”视图, 描述 13

“解析”页, 描述 13

“目录辅助”配置 62

“日志”页, 描述 13

“搜索和浏览”portlet 配置 184

“搜索”页, 描述 13

“搜索”portlet 部署 183

“搜寻”页, 描述 13

“索引”页, 描述 13

“系统”视图, 描述 13



中国印刷



S151-0015-02



Spine information:



**WebSphere II OmniFind
Edition**

管理企业搜索

版本 8.3