

IBM WebSphere Information Integrator
OmniFind Edition



Управление поисковой системой предприятия

Версия 8.3

IBM WebSphere Information Integrator
OmniFind Edition



Управление поисковой системой предприятия

Версия 8.3

Перед тем как использовать данный документ и продукт, описанный в нем, прочтите общие сведения под заголовком "Замечания".

Этот документ содержит информацию, принадлежащую IBM. Она предоставляется в соответствии с лицензионным соглашением и защищена законами об авторском праве. Информация в данной публикации не включает никаких гарантий на продукт и никакое из утверждений в данном руководстве не следует понимать подобным образом.

Заказать публикации IBM можно через Интернет или у местного представителя IBM:

- Чтобы заказать публикации через Интернет, вызовите Центр публикаций IBM по адресу www.ibm.com/shop/publications/order.
- Чтобы найти местное представительство IBM, вызовите IBM Directory of Worldwide Contacts по адресу www.ibm.com/planetwide.

Отсылая информацию IBM, вы тем самым даете IBM неисключительное право использовать или распространять эту информацию любым способом, как фирма сочтет нужным, без каких-либо обязательств перед вами.

© Copyright International Business Machines Corporation 2004, 2005. Все права защищены.

Содержание

Об этих темах vii

Для кого предназначены эти темы vii

Что такое поисковая система предприятия? 1

Типы источников данных, поддерживаемые поисковой системой предприятия 2

Обзор компонентов поисковой системы предприятия . 3

Искатели поисковой системы предприятия. 4

Анализаторы поисковой системы предприятия . . . 5

Индексы поисковой системы предприятия 6

Серверы поиска для поисковой системы предприятия 8

Консоль администратора поисковой системы предприятия 9

Слежение за поисковой системой предприятия . . . 9

Файлы журналов поисковой системы предприятия . 10

Настройка поисковой системы предприятия 10

Пример программы поиска для поисковой системы предприятия 11

Поток данных поисковой системы предприятия. . . 12

Обзор управления поисковой системой предприятия 15

Регистрация на консоли администратора 18

Изменение пароля администратора поисковой системы предприятия для конфигурации с одним сервером 19

Изменение пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами 20

Изменение номера порта для поисковой системы предприятия 24

Собрания поисковой системы предприятия 27

Создание собрания при помощи мастера по собраниям 27

Создание собрания при помощи окна Собрания . . . 28

Изменение собрания 30

Удаление собрания 31

Управление искателями поисковой системы предприятия. 33

Создание искателя 35

Изменение свойств искателя 36

Изменение пространства для искателя. 36

Удаление искателя 37

Искатели Content Edition 37

Доступ к репозиториям WebSphere II Content Edition в режиме сервера 38

Доступ к репозиториям WebSphere II Content Edition в непосредственном режиме. 39

Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition 39

Конфигурирование сервера искателей в Windows для WebSphere II Content Edition 40

Искатели DB2 42

Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2. 43

Конфигурирование WebSphere MQ для искателей DB2 46

Конфигурирование сервера искателей в UNIX для использования публикации событий 47

Конфигурирование сервера искателей в Windows для использования публикации событий 48

Искатели DB2 Content Manager 49

Конфигурирование сервера искателей в UNIX для DB2 Content Manager 50

Конфигурирование сервера искателей в Windows для DB2 Content Manager 51

Искатели Domino Document Manager 53

Искатели Exchange Server 54

Проверка доступа к защищенным документам Exchange Server 55

Искатели NNTP 55

Искатели Notes 56

Советы по просмотру баз данных Lotus Domino. . . 58

Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino 59

Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino 60

Конфигурирование серверов, использующих протокол DIIOP 62

Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino 63

Искатели QuickPlace 64

Конфигурирование сервера QuickPlace для использования защиты Local User 66

Конфигурирование Directory Assistance на сервере QuickPlace 67

Искатели файловой системы UNIX 68

Искатели Web. 68

Конфигурация пользовательского агента. 69

Поддержка JavaScript 70

Правила ограничения пространства для искателя Web 71

Проверка соединений URL для искателя Web . . . 75

Задание интервала повторного просмотра для искателя Web 75

Опции посещения URL искателем Web. 76

Как искатель Web обрабатывает произвольные страницы ошибок 76

Поддержка просмотра защищенных сайтов 78

Сайты с доступом через прокси-серверы 80

Управление опознавательными файлами (cookie) 81

Конфигурирование глобального пространства для искателя Web 82

Директивы no-follow и no-index	84
Переопределение директив no-follow и no-index на Web-страницах	85
Искатели WebSphere Portal	86
Внедрение прикладной программы предприятия для искателя WebSphere Portal	86
Копирование URL для просмотра искателем из WebSphere Portal	88
Искатели файловой системы Windows	88
Конфигурирование поддержки для программ приема данных	89
Пользовательские подключаемые модули искателей Форматы URI для индекса поисковой системы предприятия	90 91

Управление анализаторами поисковой системы предприятия . . . 99

Работа с категориями	100
Категории на основе правил	101
Категории на основе модели	103
Деревья категорий	104
Выбор типа категоризации	104
Конфигурирование категорий	105
Работа с полями поиска XML	107
Поля поиска XML	107
Отображение элементов XML на поля поиска	107
Работа с полями поиска HTML	109
Поля поиска HTML	109
Отображение элементов метаданных HTML на поля поиска	110
Пользовательская обработка текста	111
Добавление механизмов анализа текста в систему Связывание механизма текстового анализа с собранием	112 113
Отображение элементов XML на общую структуру анализа	113
Отображение общей структуры анализа на индекс Отображение общей структуры анализа на таблицы JDBC	114 115
Конфигурирование потоков для службы анализатора	116
Включение расширенного анализа для составных терминов	116
Включение поддержки собственного поиска XML Лингвистический анализ документов на китайском, японском и корейском языках	117 118
Сегментация энграмм.	118
Удаление символов новой строки из диапазонов символов не-ASCII.	118
Типы документов, связанные с анализаторами собраний и сеансами Stellent	119
Связывание типов документов с анализатором собраний	120
Правила по умолчанию для служб анализаторов собрания	121
Связывание типов документов с сеансом Stellent Правила анализа по умолчанию для сеансов Stellent.	122 124

Управление индексом поисковой системы предприятия 127

Планирование операций построения индекса	128
Изменение расписания индексирования	129
Включение и выключение расписаний индекса	129
Конфигурирование одновременных операций построения индекса	130
Опции, влияющие на представление индекса для поиска	131
Индексированные опции для поиска документов	131
Символы подстановки в запросах	133
Конфигурирование опций для символов подстановки в запросах	135 136
Области действия	136
Конфигурирование областей действия	137
Сворачивание URI	138
Объединение URI в результатах поиска	139
Удаление URI из индекса.	140

Управление серверами поиска для поисковой системы предприятия . . . 143

Кэши поиска.	143
Конфигурирование кэша поиска	144
Пользовательские словари синонимов	144
Добавление словарей синонимов в систему	146
Связывание словаря синонимов с собранием	146
Пользовательские словари стоп-слов	147
Добавление словарей стоп-слов в систему	148
Связывание словаря стоп-слов с собранием.	148
Динамическое составление сводок.	149
Настройка сводок документов на консоли администратора	149
Настройка сводок документов при помощи редактирования свойств	150
Работа с прямыми ссылками	150
Прямые ссылки	151
Конфигурирование прямых ссылок	151

Ранжирование документов в поисковой системе предприятия . . . 153

Назначение оценок на основе текста	153
Статическое ранжирование	153
Пользовательские словари значимых слов	154
Добавление словарей значимых слов в систему	155
Связывание словаря значимых слов с собранием	156
Ранжирование документов на основе шаблонов URI Влияние на оценки документов, соответствующих шаблонам URI	156 157
Ранжирование документов на основе классов значимости	158
Отображение полей на классы значимости	159
Конфигурирование коэффициентов значимости для классов значимости	160
Значения классов значимости по умолчанию	161

Прикладные программы поиска для поисковой системы предприятия . . . 163

Задание прикладных программ поиска для собраний	163
Функции примера программы поиска.	164

	Свойства примера программы поиска	165
	Редактирование свойств примера программы поиска	174
	Доступ к примеру программы поиска.	175
	Включение защиты для примера программы поиска	176

Внешние источники поисковой системы предприятия 179

	Добавление внешних источников в систему.	179
	Задание прикладных программ поиска для внешних источников	180

Защита поисковой системы предприятия 183

	Административные роли.	184
	Конфигурирование пользователей-администраторов	185
	Сравнение аутентификации и управления доступом	186
	Отключение защиты для прикладной программы предприятия в WebSphere Application Server.	187
	Защита уровня собрания	188
	Анализ повторных документов.	189
	Анализ привязок текста	189
	Индексация текста привязки в ссылках на запрещенные документы	190
	Защита с ID прикладных программ поиска	191
	Защита на уровне документов	191
	Проверка на основе хранимых маркеров защиты	192
	Проверка текущих параметров регистрации при обработке запроса	192
	Обеспечение защиты уровня документов для документов файловых систем Windows	195
	Обеспечение защиты уровня документов для документов Lotus Domino	197
	Отключение защиты на уровне документов.	199

Интеграция поисковой системы предприятия с WebSphere Portal . . . 201

	Защита уровня документов с Portal Search Engine	203
	Внедрение портлета поиска	203
	Конфигурирование портлета поиска и просмотра WebSphere Portal для поисковой системы предприятия	205
	Установка адаптера поисковой системы предприятия для Центра поиска	206
	Установка портлета регистрации поисковой системы предприятия для Центра поиска	207

Перенастройка из WebSphere Portal в поисковую систему предприятия . 209

	Перенос таксономии на основе модели из WebSphere Portal	209
	Перенос собрания из WebSphere Portal	210
	Перенос параметров собраний	212
	Файл журнала мастера по перенастройке	213

Запуск и остановка серверов поисковой системы предприятия . . 215

	Запуск серверов поисковой системы предприятия	215
	Остановка серверов поисковой системы предприятия	217

Слежение за активностью поисковой системы предприятия . . 219

	Оценка числа документов в собрании.	219
	Проверка доступности системных ресурсов	220
	Слежение за собранием	221
	Просмотр подробностей об URI	221
	Отслеживание работы искателей	223
	Просмотр подробностей об активности искателя Web	224
	Подробности потока искателя Web	225
	Активные сайты искателя Web	225
	Скорость просмотра искателя Web	226
	Создание отчетов искателя Web	227
	Коды возврата HTTP искателя Web	228
	Отслеживание работы анализатора	231
	Отслеживание активности индекса для собрания	232
	Слежение за очередью индексации поисковой системы предприятия	233
	Отслеживание работы серверов поиска	234
	Слежение за работой системы приема данных	235
	Отслеживание документов	236
	Конфигурирование файлов журнала для отслеживания документов	236
	Просмотр отчетов об отброшенных документах	237

Файлы журналов и оповещения поисковой системы предприятия . . 239

	Оповещения	239
	Конфигурирование оповещений уровня собрания	240
	Конфигурирование оповещений уровня системы	241
	Конфигурирование файлов журнала	242
	Конфигурирование информации о SMTP-сервере	243
	Получение сообщений журнала по электронной почте	244
	Просмотр файлов журнала	246

Резервное копирование и восстановление поисковой системы предприятия 249

	Резервное копирование поисковой системы предприятия	249
	Восстановление поисковой системы предприятия	250
	Восстановление файлов поисковой системы предприятия на новые серверы	251

Команды, коды возврата и ID сеансов поисковой системы предприятия 253

Документация поисковой системы предприятия 281

Доступность WebSphere II OmniFind Edition 283

Глоссарий терминов поисковой системы предприятия 285

Получение информации о WebSphere Information Integration	297
Отзывы на документацию	299
Как связаться с IBM	301

Товарные знаки	303
Индекс	309

Об этих темах

Эта информация используется при управлении системой IBM WebSphere Information Integrator OmniFind Edition Версии 8.3.

WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) поддерживает технологию, которая называется *поисковой системой предприятия*. Компоненты поиска для поисковой системы предприятия устанавливаются при установке продукта WebSphere II OmniFind Edition. Термин *поисковая система предприятия* используется во всей документации к WebSphere II OmniFind Edition, за исключением ссылок на пути установки и ярлыков на упаковке продукта.

В документации к поисковой системе предприятия описаны следующие темы:

- Введение в понятия, компоненты и возможности поисковой системы предприятия
- Инструкции по созданию собраний, слежению за ними и управлению собраниями
- Информация о том, как анализируются данные, и способы настройки анализа для оптимизации поиска и получения документов
- Информация о том, как индексируются проанализированные данные, и способы управления выполнением индексирования
- Информация о том, как серверы поиска ищут данные, ранжируют и возвращают результаты поиска.
- Информация о связывании пользовательских программ поиска с собраниями
- Обзор различных уровней защиты, доступных в поисковой системе предприятия
- Информация об интеграции поисковой системы предприятия в IBM WebSphere Portal
- Инструкции по перенастройке таксономий и собраний WebSphere Portal в деревьях категорий и собрания поисковой системы предприятия
- Инструкции по резервному копированию и восстановлению системы
- Инструкции по созданию и просмотру файлов журналов
- Инструкции по использованию команд поисковой системы предприятия и интерпретации возвращаемой ими информации

Для кого предназначены эти темы

Эта информация предназначена для администраторов и операторов системы, отвечающих за создание собраний поисковой системы предприятия, наблюдение за ними и управление ими.

Используйте эту информацию при создании собраний, при выборе содержимого для собрания и при конфигурировании опций, позволяющих выполнять поиск в этом содержимом. Кроме того, в этой информации описывается наблюдение за работой собраний и системы, назначение пользователей администраторами поисковой системы предприятия и связывание собраний и внешних источников для поиска с прикладными программами поиска.

Чтобы эффективно использовать эту информацию, надо уметь работать с прикладными программами Web и обладать опытом работы с источниками данных, поиск в которых вы хотите обеспечить при помощи поисковой системы предприятия.

Что такое поисковая система предприятия?

Поисковая система предприятия предоставляет возможность с помощью одного запроса выполнять поиск по любому числу структурированных или неструктурированных источников данных. Поисковая система предприятия обеспечивает быстрый поиск и возвращает сгруппированный и ранжированный набор результатов, позволяющий легко найти необходимую информацию.

Компоненты поисковой системы предприятия, устанавливаемые с IBM WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition), собирают информацию со всего предприятия и позволяют выполнять поиск по этой информации. Вы сможете, введя запрос в браузере, выполнить поиск одновременно по локальным и удаленным базам данных, вспомогательным системам, системам управления содержимым, файловым системам, внутренним и внешним сайтам.

Поисковая система предприятия спроектирована для полной интеграции с существующими системами; она выполняет всю работу по сбору данных из разнородных источников и построению индексов для быстрого извлечения данных. Используя лингвистический анализ и другие типы анализа, поисковая система предприятия возвращает результаты поиска с высоким уровнем соответствия запросу. Вам не придется изучать разные интерфейсы для поиска по различным типам репозиториям.

Можно добавить поддержку поиска в источниках данных, которые вы не хотите включать в индекс поисковой системы предприятия. При помощи функции объединенного поиска, включенной в поисковую систему предприятия, можно выполнять поиск в этих внешних источниках одновременно с поиском в индексированных источниках данных.

Качество поиска

WebSphere II OmniFind Edition поддерживает архитектуру управления неструктурированной информацией IBM (UIMA - Unstructured Information Management Architecture), которая позволяет пользователям находить нужную им информацию. UIMA - это открытая структура, определяющая общий стандартный интерфейс для анализа текста. Всесторонний анализ текста позволяет поисковой системе предприятия выявлять в неструктурированном тексте понятия, латентные значения, связи, факты и другие релевантные данные, которые часто бывают скрыты. Извлеченную при анализе информацию можно использовать для повышения качества результатов поиска или для повышения качества других прикладных программ, например, бизнес-интеллекта и исследования данных.

Защита

Защита - существенный элемент поисковой системы предприятия. Системой могут управлять только авторизованные для этого пользователи. Используя механизмы защиты IBM WebSphere Application Server, можно настроить роли администраторов и предоставлять разным пользователям доступ к разным функциям управления.

Кроме того, можно задать опции установки маркеров защиты данных при сборе данных. Если ваши программы поиска поддерживают защиту, эти маркеры (хранящиеся в индексе вместе с документами) можно использовать для обеспечения

управления доступом, чтобы только пользователи с соответствующими параметрами регистрации могли запрашивать данные и просматривать результаты поиска.

Для некоторых типов источников данных можно конфигурировать опции проверки учетных данных регистрации пользователя с текущим управлением доступом, выполняемой при обработке запросов. Этот дополнительный уровень защиты обеспечивает проверку привилегий пользователей в реальном времени с собственным источником данных. Эта опция обеспечивает защиту от таких экземпляров, где учетные данные пользователя изменились после индексации документа и его маркеров защиты.

Понятия, связанные с данным

Защита поисковой системы предприятия

"Интеграция пользовательского анализа текста" в разделе "Пользовательский анализ текста"

"Unstructured information management architecture (UIMA)" in "Text Analysis Integration"

Типы источников данных, поддерживаемые поисковой системой предприятия

В системе предусмотрена поддержка поиска по различным типам источников данных.

После установки IBM WebSphere Information Integrator OmniFind Edition можно собирать данные из источников данных следующих типов:

- Типы элементов IBM DB2 Content Manager (документы, ресурсы и элементы)
- Базы данных IBM DB2 Universal Database
- Базы данных IBM Domino Document Manager (прежнее название - Domino.Doc)
- Базы данных IBM Lotus Notes
- Базы данных IBM Lotus QuickPlace
- Общедоступные папки Microsoft Exchange Server
- Файловые системы Microsoft Windows
- Группы новостей NNTP
- Файловые системы UNIX
- Репозитории IBM WebSphere Information Integrator Content Edition: классы элементов Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), OpenText Livelink и Portal Document Manager (PDM)
- Таблицы псевдонимов IBM WebSphere Information Integrator для баз данных IBM DB2 Universal Database для z/OS, IBM Informix, Microsoft SQL Server, Oracle и Sybase
- Сайты IBM WebSphere Portal
- Сайты Интернета или внутренней сети

Кроме того, можно добавить поддержку поиска во внешних источниках следующих типов без добавления документов из этих источников в индекс поисковой системы предприятия:

- Таблицы баз данных JDBC (Java database connectivity) (только для баз данных IBM DB2 Universal Database (DB2 UDB) и Oracle). Для каждой таблицы в базе данных JDBC создается отдельный внешний источник данных.
- Серверы LDAP (Lightweight Directory Access Protocol). Для каждого сервера LDAP создается один внешний источник.

Последнюю информацию о поддерживаемых типах источников данных смотрите на сайте WebSphere Information Integrator OmniFind Edition.

Понятия, связанные с данным

Внешние источники поисковой системы предприятия

Обзор компонентов поисковой системы предприятия

Компоненты поисковой системы предприятия собирают данные по предприятию, анализируют и категоризируют информацию и создают индекс, который пользователи используют при поиске.

Собрание поисковой системы предприятия - это набор источников, в которых пользователи могут выполнять поиск при помощи одного запроса. При создании собрания вы задаете источники, которые должны в него входить, и конфигурируете опции, определяющие возможности пользователей при поиске индексированных данных.

Можно создать несколько собраний, каждое из которых может содержать данные из разнообразных источников данных. Например, можно создать собрание, включающее документы из баз данных IBM DB2 Universal Database, IBM Lotus Notes и IBM DB2 Content Manager. Когда пользователи выполняют поиск в этом собрании, результаты поиска могут включать документы из каждого из этих источников данных.

Поддержка объединенного поиска позволяет пользователям выполнять поиск в нескольких собраниях при помощи одного запроса. Результаты поиска могут включать документы из всех собраний и внешних источников в вашей поисковой системе предприятия.

При создании собрания и управлении им выполняются следующие действия:

Сбор данных

Компоненты *искателя* собирают документы из источников данных либо непрерывно, либо по заданному вами расписанию. Частый просмотр информации искателями гарантирует, что у пользователей всегда будет доступ к последним данным.

Анализ данных

Компоненты *анализатора* извлекают текст из документов и выполняют лингвистический и другие типы анализа для каждого документа, обработанного искателем. Подробный анализ содержимого повышает качество результатов поиска.

Индексация данных

Компоненты *индексации* работают на основе регулярных расписаний, добавляя в индекс информацию о новых и измененных документах. Компоненты индексации выполняют также глобальный анализ документов в собрании, улучшая качество результатов поиска.

Поиск данных

Компоненты *поиска* выполняют поиск по индексу и взаимодействуют с программами поиска; они обрабатывают требования поиска и возвращают результаты поиска.

Другие компоненты WebSphere Information Integrator OmniFind Edition позволяют задать предпочтения защиты, отслеживать активность системы и устранять

возникающие ошибки. Кроме того, в продукт включен пример рабочей программы поиска, которую можно использовать как образец для создания своих собственных программ поиска.

Понятия, связанные с данным

- Управление искателями поисковой системы предприятия
- Управление анализаторами поисковой системы предприятия
- Управление индексом поисковой системы предприятия
- Управление серверами поиска для поисковой системы предприятия

Искатели поисковой системы предприятия

Искатели поисковой системы предприятия собирают из источников данных документы, чтобы для них можно было выполнить анализ, индексацию и поиск.

Компонент искателя, поставляемый с WebSphere Information Integrator OmniFind Edition, выполняет следующие функции:

- При конфигурировании искателя процессы *обнаружения* находят информацию об источниках, доступных для просмотра искателем, например, имена всех представлений и папок в базе данных Lotus Notes или имена всех файловых систем на сервере UNIX.
- После выбора источников для просмотра искателем и запуска искателя компоненты искателя собирают из этих источников данные, чтобы их можно было проанализировать и проиндексировать.

В одном собрании может быть несколько искателей, при этом каждый искатель используется для сбора данных из источников данных определенного типа. Например, можно создать три искателя, чтобы в одном собрании объединить данные из файловых систем, баз данных Notes и реляционных баз данных. Можно также создать несколько искателей одного типа и сконфигурировать для них отличающиеся расписания просмотра в соответствии с тем, насколько часто изменяются данные, просматриваемые каждым искателем.

Искатели для источников Web, WebSphere Portal и Network News Transfer Protocol (NNTP) работают непрерывно. После того, как вы зададите универсальные указатели ресурсов (Uniform Resource Locator resource, URL) или группы новостей NNTP для просмотра, искатель периодически будет возвращаться к ним, проверяя новые и измененные данные. Искатели других типов можно запускать и останавливать вручную или задать для них расписания просмотра. При задании расписания для искателя нужно задать время его начального запуска и частоту, с которой он должен посещать источники данных для просмотра новых и измененных документов.

Свойства искателя - это набор правил, регулирующих поведение отдельного искателя во время его просмотра данных. Например, можно задать правила, управляющие использованием искателем системных ресурсов. Набор источников, отобранных для просмотра, составляет *пространство для искателя*. Свойства искателя можно отредактировать в любое время после его создания, изменив правила сбора данных искателем. Можно также отредактировать пространство для искателя, изменив расписание просмотра, добавив новые источники или удалив источники, поиск в которых больше не требуется.

Понятия, связанные с данным

- Управление искателями поисковой системы предприятия

Задачи, связанные с данной

- Отслеживание работы искателей

Анализаторы поисковой системы предприятия

Синтаксический анализатор поисковой системы предприятия анализирует документы, собранные искателем, и подготавливает их для индексирования.

Компонент синтаксического анализа, поставляемый с WebSphere Information Integrator OmniFind Edition, анализирует содержимое документов и их метаданные. Этот компонент сохраняет результаты анализа на складе данных для компонента индексирования. Синтаксический анализатор выполняет следующие задачи:

- Извлекает текст из документов любого формата. Например, синтаксический анализатор извлекает текст из тегов в документах XML и HTML. Используя технологию Stellent for IBM WebSphere Information Integrator OmniFind Edition Outside In Viewer, анализатор извлекает текст и из двоичных файлов, таких как документы Microsoft Word или Adobe Acrobat (PDF).
- Определяет кодовую страницу документа. Анализатор использует эту информацию, чтобы преобразовать весь текст в Unicode, прежде чем выполнять логический анализ.
- Определяет исходный язык документа.
- Применяет правила синтаксического анализа, заданные для собрания. При конфигурировании синтаксического анализатора можно настроить:

Правила отображения полей для документов XML и HTML

Эта возможность позволяет пользователям выполнять поиск по структурированному и неструктурированному содержимому в документах XML и HTML. Если отображать элементы XML или элементы метаданных HTML в поля поиска в индексе поисковой системы предприятия, пользователи смогут указывать в запросах имена полей и выполнять поиск по определенным частям документов XML и HTML. (Запросы с указанием конкретных полей могут давать более точные результаты поиска, чем произвольные запросы текста, выполняющие поиск во всем содержимом документа.)

Категории

Эта возможность позволяет пользователям выполнять поиск по определенной категории документов. Кроме того, пользователи могут выбрать категории в результатах поиска и просмотреть документы по категориям, к которым они относятся.

При создании собрания можно выбрать типы категорий, которые будут в нем использоваться. Если использовать категории *на основе правил*, документы будут разбиваться по категориям на основе определенных вами правил. Сконфигурировать категории на основе правил можно для создаваемых собраний поисковой системы предприятия и для собраний, переносимых из IBM WebSphere Portal.

Если использовать категории *на основе моделей*, документы будут разбиваться по категориям, как задается моделями, существующим в вашей системе WebSphere Portal. Для использования этой опции на сервере индекса поисковой системы предприятия должен быть установлен WebSphere Portal. Необходимо также использовать для управления категориями инструменты категоризации, входящие в WebSphere Portal.

Пользовательский анализ текста

Разработчики могут создавать свои программы для сложного лингвистического анализа данных, в которых выполняется поиск. Эти программы можно подключать к поисковой системе предприятия и

использовать для аннотирования содержимого ваших собраний. Индексация аннотаций делает возможным семантический поиск в собраниях.

Например, пользователи могут выполнять поиск терминов запроса, встречающихся недалеко друг от друга или встречающихся в одном предложении, или поиск различных взаимосвязей между терминами запроса (например, поиск документов, в которых упоминается тот Смит из IBM, который занимается продажами, а не тот Смит из IBM, который инженер).

Поддержка сегментации энграмм

Чтобы облегчить извлечение документов на китайском, японском и корейском, можно разрешить такой метод лексического анализа, как сегментация энграмм. В этом способе анализа пробелы не считаются разделителями слов. (После создания собрания изменить метод сегментации нельзя.)

Поддержка поиска по документам XML с использованием собственного поиска XML

Собственный поиск XML может давать более точные результаты благодаря учету разметки XML. Например, в запросе можно задать, что некоторое слово должно встретиться в определенном элементе XML.

Классы значимости для оценки относительной важности полей

Отображая поля на классы значимости, можно влиять на ранжирование документов в результатах поиска. Например, можно повысить оценку полей названий, чтобы документы, у которых термин запроса входит в название, получали более высокий ранг в результатах поиска.

- Извлекает текст и добавляет маркеры, чтобы облегчить извлечение данных. На этой фазе синтаксический анализатор выполняет следующие задачи:
 - Нормализация символов, например нормализующий перевод в верхний регистр и удаление диакритических знаков, таких как умляуты в немецком.
 - Анализ структуры абзацев, предложений, слов и пустого пространства. Используя лингвистический анализ, анализатор разбивает составные слова на части и назначает маркеры, позволяющие выполнять просмотр словаря и синонимов.

Понятия, связанные с данным

Работа с категориями

поля поиска XML

поля поиска HTML

"Интеграция пользовательского анализа текста" в разделе "Пользовательский анализ текста"

"Включение анализа текста в поисковую систему предприятия" в разделе "Интеграция анализа текста"

Задачи, связанные с данной

Отслеживание работы анализатора

Индексы поисковой системы предприятия

Компоненты индексации поисковой системы предприятия работают по регулярным расписаниям, добавляя в индекс информацию о новых и измененных документах.

Чтобы у пользователей всегда был доступ к последней информации в источниках, где они выполняют поиск, применяется двухэтапное построение индекса:

Реорганизация индекса

При реорганизации перестраивается весь индекс, что гарантирует

оптимальность его структуры. Процессы индексирования считывают все данные, собранные искателями и проанализированные анализаторами.

Обновление индекса

При обновлении в индекс добавляется информация, просмотренная искателями со времени последней реорганизации индекса.

При конфигурировании опций индексации для собрания можно задать расписания для реорганизации и обновления индекса. Периодичность, с которой следует реорганизовывать и обновлять индекс, определяется ресурсами системы и содержимым индексируемых источников - статическим или динамическим.

Чтобы гарантировать доступность новой информации, нужно запланировать достаточно частое обновление индекса. Периодически составляйте расписание реорганизации индекса, позволяющей объединить всю новую информацию, проанализировать новое содержимое и оптимизировать производительность.

Процессы индексации можно также запускать вне расписаний. Например, если вы изменили некоторые правила анализа и хотите, чтобы эти изменения стали доступны вашим программам поиска, можно запустить реорганизацию индекса после повторного просмотра и анализа данных, не дожидаясь начала реорганизации индекса по расписанию.

Чтобы управлять использованием ресурсов, можно задать число собраний, которые могут одновременно использовать процессы индексации и отправлять требования построения индексов. Одновременное построение индексов позволяет добиться того, что реорганизация очень большого индекса не будет блокировать обновление других индексов. Построение индекса может быть ресурсоемким процессом, поэтому чтобы настроить частоту реорганизации и обновления индекса для большой системы, нужно выполнить мониторинг системной нагрузки.

При построении индекса процессы индексации выполняют глобальный анализ документов. Во время этой фазы применяются алгоритмы для идентификации повторяющихся документов, анализа структуры ссылок документов и выполнения специальной обработки текста привязок в документах Web (текста, описывающего страницу назначения в гиперссылке).

Можно задать опции для следующих операций индексации:

- Чтобы разрешить пользователям задавать символы подстановки, можно или встроить в индекс поддержку для расширения терминов запроса, или указать, что расширение терминов запроса надо выполнять при обработке запроса. Выбор опции зависит от того, что именно вы хотите минимизировать - использование ресурсов или время ответа на запрос.
- Можно сконфигурировать области действия. *Область действия* позволяет ограничить данные, которые пользователи могут просмотреть в собрании. Например, можно создать одну область действия, включив в нее все идентификаторы URI для документов из вашего технического отдела, а в другую область действия - URI документов из отдела кадров. Если программа поиска поддерживает области действия, пользователи смогут выполнять поиск документов и получать их только из этих поднаборов документов в собрании.
- Можно задать опции для объединения в результатах поиска документов с одинаковым префиксом URI. Можно также задать имя группы, чтобы в результатах поиска можно было объединять документы с разными префиксами URI.

- После построения индекса можно удалить URI, для которых пользователи не должны выполнять поиск.

Понятия, связанные с данным

Управление индексом поисковой системы предприятия

Символы подстановки в запросах

Области действия

Сворачивание URI

Ранжирование документов на основе шаблонов URI

Задачи, связанные с данной

Планирование операций построения индекса

Конфигурирование одновременных операций построения индекса

Удаление URI из индекса

Отслеживание активности индекса для собрания

Слежение за очередью индексации поисковой системы предприятия

Серверы поиска для поисковой системы предприятия

Серверы поиска для поисковой системы предприятия работают с прикладными программами поиска, они обрабатывают запросы, выполняют поиск по индексу и возвращают результаты поиска.

Серверы поиска для поисковой системы предприятия устанавливаются при установке WebSphere Information Integrator OmniFind Edition. При конфигурировании серверов поиска для собрания можно задать опции поиска по собранию:

- Можно сконфигурировать кэш поиска для хранения часто запрашиваемых результатов поиска. Кэширование поиска может повысить производительность поиска и получения документов.
- Можно задать язык по умолчанию для поиска документов в этом собрании.
- Если ваши разработчики программы создадут пользовательские словари, можно связать эти словари с собраниями:
 - При запросах к собранию, использующему *словарь синонимов*, документы, содержащие синонимы терминов запроса, включаются в результаты поиска.
 - При запросах к собранию, использующему *словарь стоп-слов*, стоп-слова удаляются из запроса перед его обработкой.
 - При запросах к собранию, использующему *словарь значимых слов*, важность документов, содержащих эти слова, увеличивается или уменьшается (в зависимости от показателя значимости, заданного для этих слов в словаре).
- Если вы заранее определили, что определенным запросам соответствуют определенные документы, можно настроить прямые ссылки. *Прямая ссылка* связывает определенный URI с определенными ключевыми словами и словосочетаниями. Если в запросе содержится какое-либо из ключевых слов или словосочетаний, задающих определение прямой ссылки, в результатах поиска автоматически возвращается соответствующий URI.

Для конфигурации с несколькими серверами защита от сбоя доступна также на уровне собрания, а не только на уровне сервера. Если по какой-либо причине собрание на одном из серверов поиска станет недоступно, запросы для этого собрания будут автоматически направлены на другой сервер поиска.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Кэши поиска

- Пользовательские словари синонимов
- Пользовательские словари стоп-слов
- Пользовательские словари значимых слов
- Прямые ссылки
- Задачи, связанные с данной**
- Отслеживание работы серверов поиска

Консоль администратора поисковой системы предприятия

Консоль администратора поисковой системы предприятия запускается в браузере, что позволяет пользователем-администраторам обращаться к ней из любого места в любое время. Механизмы защиты гарантируют, что сделать это смогут только пользователи с правами доступа к функциям управления.

Консоль администратора поисковой системы предприятия устанавливается на серверах поиска при установке WebSphere Information Integrator OmniFind Edition.

Консоль администратора содержит мастера, помогающие выполнять некоторые основные задачи управления. Например, мастер по собраниям помогает создать собрание и позволяет сохранять работу в черновом варианте. Мастера искателей (отдельные для каждого из типов источников данных) помогают выбрать источники, которые вы хотите сделать доступными для поиска.

В других задачах управления можно выбрать отдельные элементы для управления. Например, при редактировании собрания можно выбрать страницу Индекс, чтобы изменить расписание индексации, или страницу Анализ, чтобы изменить правило синтаксического анализа документов XML.

Понятия, связанные с данным

- Обзор управления поисковой системой предприятия
- Административные роли

Задачи, связанные с данной

- Регистрация на консоли администратора

Слежение за поисковой системой предприятия

С помощью консоли администратора поисковой системы предприятия можно выполнять слежение за системой и при необходимости настраивать операции.

После установки WebSphere Information Integrator OmniFind Edition и создания как минимум одного собрания можно просмотреть подробную статистику по каждому из основных процессов (таких как просмотр искателем, анализ, индексация и поиск). В эту информацию входит среднее время отклика и информация о ходе выполнения, например, число документов, просмотренных или проиндексированных в ходе отдельного сеанса просмотра искателем или построения индекса.

Можно остановить и запустить большинство процессов. Например, можно приостановить процесс, изменить его конфигурацию или исправить ошибку и перезапустить обработку, когда вы будете готовы ее продолжить.

Можно также сконфигурировать оповещения, позволяющие получать по электронной почте сообщения при возникновении определенных событий для отслеживаемых процессов. Например, можно получать оповещение, когда время ответа поиска превысило заданный порог.

Если документ был отброшен из поисковой системы предприятия, можно отследить этот документ и выяснить, когда, где и почему он был отброшен. Например, можно определить, что анализатор не смог обработать документ, или что администратор удалил документ из индекса.

Понятия, связанные с данным

Слежение за активностью поисковой системы предприятия

Запуск и остановка серверов поисковой системы предприятия

Файлы журналов поисковой системы предприятия

Файлы журналов поисковой системы предприятия создаются для отдельных собраний и для сеансов системного уровня.

При конфигурировании опций записи в журналы для собрания поисковой системы предприятия или для системы нужно задать типы сообщений, записываемых в журнал (например, сообщения об ошибках и предупреждения). Нужно также задать, как часто система должна переходить от более старых журналов к новым, освобождая место для свежих сообщений. Можно выбрать опции для получения сообщения электронной почты о конкретных сообщениях (в том числе оповещений) или всех сообщениях об ошибках, когда они происходят.

При просмотре файлов журналов можно выбрать файл журнала, который нужно просмотреть (имя файла содержит информацию о времени создания файла и о компоненте, сгенерировавшем сообщения). Можно также задать фильтры просмотра. Например, можно просмотреть только сообщения об ошибках или только сообщения от конкретного сеанса поисковой системы предприятия.

Понятия, связанные с данным

Файлы журналов и оповещения поисковой системы предприятия

Оповещения

Сообщения для поисковой системы предприятия

Задачи, связанные с данной

Конфигурирование файлов журнала

Конфигурирование информации о SMTP-сервере

Получение сообщений журнала по электронной почте

Просмотр файлов журнала

Настройка поисковой системы предприятия

Интерфейсы прикладного программирования для поисковой системы предприятия позволяют создавать пользовательские программы поиска, пользовательские программы для обновления содержимого собраний, пользовательские программы для анализа текста и пользовательские словари синонимов, стоп-слов и значимых слов.

После установки WebSphere Information Integrator OmniFind Edition можно использовать следующие семейства API, позволяющие расширить собрания поисковой системы предприятия:

API поиска и индексации (SI-API)

Этот API используется для построения пользовательских программ поиска и пользовательского интерфейса администратора.

API приема данных

Этот API используется для получения данных от внешних искателей.

Внешние ищатели могут соединяться с компонентом приема данных поисковой системы предприятия и добавлять данные в собрание или удалять данные из собрания.

Подключаемые модули ищателей

API для ищателей Web и других ищателей используются для добавления метаданных в документы при их просмотре ищателем или для привязки маркеров защиты, обеспечивающих выполнение бизнес-правил и правил защиты, принятых в вашей организации.

Можно повысить возможность получения информации, интегрировав пользовательские программы лингвистического анализа в собраниях поисковой системы предприятия. Добавив в систему пользовательские механизмы анализа текста, можно связать их с собраниями. Пользователи, запрашивающие собрание, могут использовать преимущества информации о взаимосвязях между словами, которую пользовательские программы встраивают в индекс. Например, пользователи могут искать понятия и взаимосвязи между терминами, а не просто сами термины.

Можно также повысить возможность получения информации, интегрировав пользовательские словари, отражающие, например, акронимы, сокращения и специфические словарные термины для вашей области деятельности. Добавив в систему пользовательские словари, можно связать их с собраниями. Пользователи, запрашивающие собрание, получают следующие преимущества:

- Если запрос содержит слова, определенные как синонимы, документы, содержащие синонимы терминов запроса, включаются в результаты поиска.
- Если запрос содержит стоп-слова, они удаляются из запроса, чтобы исключить из результатов поиска нерелевантные документы.
- Если запрос содержит значимые слова, документы, содержащие эти слова, получают более высокий или более низкий ранг в результатах поиска в зависимости от показателей значимости, заданных в словаре для этих слов.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Пользовательские словари синонимов

Пользовательские словари стоп-слов

Пользовательские словари значимых слов

"Search and index API overview" (Обзор API поиска и индексирования) в книге "Programming Guide and API Reference for Enterprise Search"

"Data listener" (Система приема данных) в книге "Programming Guide and API Reference for Enterprise Search"

"Crawler plug-ins" (Подключаемые модули ищателей) в книге "Programming Guide and API Reference for Enterprise Search"

Задачи, связанные с данной

Конфигурирование поддержки для программ приема данных

Пример программы поиска для поисковой системы предприятия

Можно использовать этот пример программы поиска как есть или же как образец для создания пользовательских программ поиска.

Пример программы поиска устанавливается при установке WebSphere Information Integrator OmniFind Edition. Этот пример программы поиска демонстрирует

большинство функций поиска и получения, доступных для поисковой системы предприятия. Эта программа - рабочий образец, позволяющий выполнять поиск во всех активных собраниях и внешних источниках в поисковой системе предприятия. Пример программы поиска можно использовать для тестирования новых собраний и внешних источников перед тем, как делать их доступными для пользователей.

Пример программы поиска демонстрирует поддержку объединенного поиска, позволяя одновременно выполнять поиск в нескольких собраниях и внешних источниках.

Если для собрания включена защита, а для системы включено управление идентификацией WebSphere II OmniFind Edition, при использовании программы поиска пользователи могут создавать профили. В профиле пользователя содержатся учетные данные, которые задаются пользователями для регистрации в различных доменах.

При обработке запросов процессы поиска используют сохраненные учетные данные (зашифрованные на защищенном складе, которым управляет WebSphere II OmniFind Edition), чтобы определить, есть ли у пользователя разрешение на поиск в защищенных доменах. Если для домена данные аутентификации отсутствуют или недействительны, документы этого домена исключаются из результатов поиска.

Информацию об использовании примера программы поиска можно получить, нажав кнопку **Справка** во время использования программы. Для создания собственной программы поиска можно использовать API поиска и индексации поисковой системы предприятия.

Понятия, связанные с данным

Прикладные программы поиска для поисковой системы предприятия

Функции примера программы поиска

"Search and index API overview" (Обзор API поиска и индексирования) в книге "Programming Guide and API Reference for Enterprise Search"

Задачи, связанные с данной

Доступ к примеру программы поиска

Редактирование свойств примера программы поиска

Включение защиты для примера программы поиска

Поток данных поисковой системы предприятия

Компоненты поисковой системы предприятия, которые устанавливаются вместе с WebSphere Information Integrator OmniFind Edition, тесно взаимодействуют между собой, чтобы обеспечить поток данных через систему.

Искатели собирают документы из источников данных по всему предприятию. Синтаксический анализатор извлекает из собранных искателем документов полезную информацию и генерирует маркеры, которые позволяют, например, распределять документы по категориям и помогают определить степень соответствия документов терминам в запросе поиска. Индекс хранит данные для эффективного поиска.

Используя браузер и прикладные программы поиска, пользователи выполняют поиск по проиндексированным собраниям и внешним источникам. Прикладная программа поиска может отображать список результатов, по которым пользователь может щелкнуть в браузере; прикладная программа может также быть более сложной и возвращать динамически генерируемое содержание на основе информации из разных источников.

Например, программа поиска каталога может настраивать способ отображения продуктов, удовлетворяющих запросу поиска. С помощью одного запроса можно выполнять поиск по документам в разных источниках данных, например, по документам в репозиториях IBM DB2 Content Manager и Lotus Notes.

Администраторы задают, какие данные надо собирать и как их надо просматривать, анализировать, индексировать и искать. Наблюдая за работой системы, администраторы выполняют также настройку для оптимизации потоков данных.

На следующей диаграмме показан поток информации по поисковой системе предприятия.

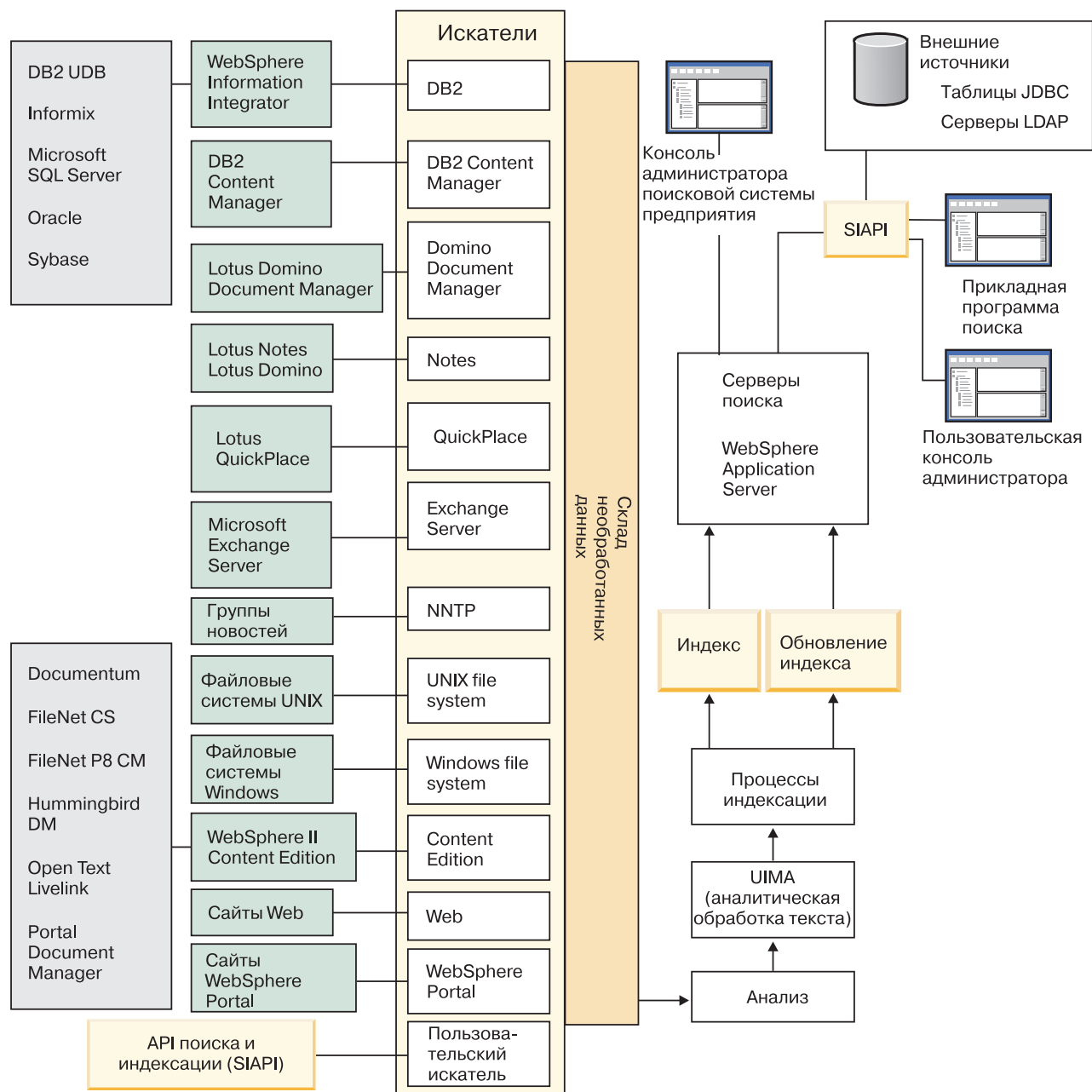


Рисунок 1. Как перемещаются данные по поисковой системе предприятия

Обзор управления поисковой системой предприятия


При помощи консоли управления поисковой системой предприятия можно создавать собрания и внешние источники и управлять ими, запускать и останавливать компоненты, следить за работой системы и файлами журналов, конфигурировать пользователей - администраторов и связывать программы поиска с собраниями.

Окно Собрания

При создании первого собрания и управлении системой следуйте приведенным ниже инструкциям.

Регистрация

Зарегистрируйтесь на консоли управления поисковой системой предприятия. Окно Собрания - это точка входа для создания собраний и управления ими.

Совет: Чтобы получить справку по консоли администратора, нажмите кнопку **Справка** на панели инструментов или **Справка для этой страницы**. При наличии подробных пояснений к сообщениям и описаний действий по исправлению щелкните по значку  **Дополнительная информация** в конце сообщения, чтобы увидеть эти подробности.


Создание собрания

Новое, пустое собрание можно создать одним из следующих способов:

- Чтобы создать собрание при помощи мастера по собраниям, выберите **Мастер по собраниям** и следуйте подсказкам мастера.
- Чтобы создать собрание при помощи окна Собрания, выберите **Создать собрание**, заполните поля на странице Создать собрание и нажмите кнопку **ОК**.

Конфигурирование собрания

Нужно отредактировать новое, пустое собрание, чтобы добавить в него содержимое и задать опции для просмотра данных искателем и доступа к этим данным.

Нажмите для вашего нового собрания кнопку  **Изменить**, затем выберите страницу, чтобы задать опции для этого собрания.

Внимание: Если нажать кнопку Назад или Обновить в браузере Web, можно получить непредсказуемый результат или потерять данные. Когда вы конфигурируете собрания и хотите вернуться на предыдущую страницу или обновить информацию на консоли администратора, следует нажимать кнопки **Предыдущая** и **Обновить** на консоли администратора поисковой системы предприятия, а не кнопки Назад и Обновить в браузере Web.

- На странице Общие можно задать опции, применяемые ко всему собранию:
 - Редактируя общие опции, можно изменить имя или описание собрания, а также изменить ожидаемый размер собрания.
 - Можно просмотреть информацию о собрании, которую нельзя изменить, такую как ID собрания или метод статистического ранжирования для ранжирования документов в результатах поиска.

- Если защита для собрания была включена во время его создания, можно включать и выключать управление защитой на уровне документов.
- На странице Искатель сконфигурируйте как минимум один искатель. Одно собрание может содержать данные из разнообразных источников данных. Для каждого используемого типа источников данных необходимо сконфигурировать как минимум один искатель. При создании искателя сконфигурировать его помогает мастер, зависящий от типа данных этого искателя.
- На странице Анализ можно сконфигурировать опции синтаксического анализа данных, просматриваемых искателем, для более эффективного поиска данных:
 - Можно указать, будет ли выполняться анализ документов в формате XML для поиска собственными средствами поиска XML.
 - Можно связывать документы с категориями, что дает пользователям возможности поиска в поднаборе собрания и просмотра документов в результатах поиска по категориям, к которым принадлежат эти документы.
 - Можно отображать элементы XML и элементы метаданных HTML на поля поиска в индексе, что дает пользователям возможность указывать имена этих полей в запросах и находить необходимые части документов.
 - Если вы добавили в поисковую систему предприятия пользовательские механизмы анализа, можно выбрать один из них для использования с собранием, а затем задать опции обработки текста для повышения возможностей получения информации.
 - Поля можно связать с классами значимости, чтобы изменить ранжирование полей, удовлетворяющих условиям запроса, в результатах поиска.
- На странице Индекс сконфигурируйте расписания для реорганизации и обновления индексов. Запланируйте достаточно частое построение индексов, чтобы у пользователей всегда был доступ к последней информации. Можно также внести следующие необязательные изменения:
 - Позволить пользователям использовать символы подстановки в условиях запросов.
 - Сконфигурировать области действия, чтобы пользователи выполняли поиск в ограниченной части собрания, а не во всех документах, содержащихся в индексе.
 - Объединять результаты поиска, чтобы все документы из одного и того же источника были сгруппированы и объединены в результатах поиска.
 - Присваивать коэффициенты значимости, чтобы изменить порядок ранжирования в результатах поиска документов, соответствующих шаблону URI.
 - Удалять URI из индекса. Например, можно запретить пользователям просматривать определенные документы после создания собрания.
- На странице Поиск можно задать опции для поиска документов в собрании:
 - Можно отвести пространство под кэш для результатов поиска и изменять язык собрания по умолчанию.
 - Если вы добавили в поисковую систему предприятия пользовательские словари для синонимов, стоп-слова и значимые слова, можно выбрать словари для пользователей, выполняющих поиск в собрании.

- Можно указать длину вывода на экран для сводок документов в результатах поиска.
- Если нужно, чтобы отдельные идентификаторы URI выводились автоматически в результатах поиска, когда запрос содержит конкретные ключевые слова или фразы, можно сконфигурировать прямые ссылки.
- На странице Журнал можно:
 - Задать типы записываемых в журнал сообщений и частоту перезапуска файлов журналов.
 - Задать опции для получения оповещений о действиях собрания. Например, можно настроить оповещение, информирующее вас о том, что среднее время отклика поиска превысило определенный предел.
 - Задать опции для получения сообщений по электронной почте, когда в журнал записываются конкретные сообщения или сообщения конкретных типов.
 - Задать опции для информации регистрации, позволяющие определять время, место и причину отбрасывания документа из поисковой системы предприятия.

Запуск компонентов

После задания источников данных, просматриваемых искателем, и опций для сбора и поиска данных можно запустить процессы для построения собрания. Порядок запуска компонентов критичен. Искатели должны предоставить данные до того, как эти данные можно будет проанализировать, анализаторы должны проанализировать предоставленные данные до того, как их можно будет проиндексировать, а индекс должен быть обновлен или реорганизован до того, как серверы поиска смогут начать обрабатывать поисковые запросы.

Окно Внешние источники

Если нужно выполнить поиск в источниках данных без их просмотра искателем или индексирования, нажмите кнопку **Внешние источники** на панели инструментов, чтобы задать опции, делающие источники данных доступными для поиска. Необходимо задать информацию, разрешающую поисковой системе предприятия доступ к базам данных JDBC (Java Database Connectivity) и серверам LDAP (Lightweight Directory Access Protocol). Связав внешние источники с прикладными программами, поиска, вы позволите пользователям выполнять поиск в этих источниках одновременно с поиском в собраниях с обработанными искателем, проанализированными и проиндексированными данными.

Окно Система

Если вы - администратор поисковой системы предприятия, выберите на панели инструментов **Система** и выполните следующие действия (доступ администраторов, операторов и наблюдателей собраний к этому окну может разрешить только администратор поисковой системы предприятия):

- Проверьте доступность системных ресурсов.
- Сконфигурируйте клиентские программы приема данных, дав им возможность изменять собрания.
- Добавьте в систему пользовательские механизмы анализа текста.
- Добавьте в систему пользовательские словари для синонимов, стоп-слова и значимые слова.

- Задайте число собраний для одновременного построения индексов, и укажите, возможно ли одновременное выполнение требований обновления и реорганизации индекса для одного собрания.
- Сконфигурируйте оповещения для событий на уровне системы.
- Задайте опции для записи сообщений, генерируемых сеансами на уровне системы.
- Укажите сведения о вашем почтовом сервере, чтобы получать по электронной почте сообщения о работе поисковой системы предприятия.


Окно Защита

Если вы - администратор поисковой системы предприятия, можно выбрать **Защита**, чтобы задать опции защиты. У администраторов, операторов и наблюдателей собраний нет доступа к этому окну.

Если включена защита на сервере IBM WebSphere Application Server, в окне Защита можно сконфигурировать административные роли. Конфигурирование административных ролей позволяет разрешить управление системой большему числу пользователей, но с ограничением доступа каждого пользователя к конкретным функциям и собраниям.

Пока вы не создали ваши собственные программы поиска, можно использовать пример программы поиска, выполняющей поиск во всех собраниях и внешних источниках. Создав пользовательскую программу поиска, при помощи окна Защита свяжите созданную программу с собраниями и внешними источниками, в которых она может выполнять поиск.

Окно Слежение

В любое время можно щелкнуть по значку  **Слежение**, чтобы начать слежение за компонентами системы или собраний. Если позволяет ваша административная роль, при мониторинге можно также запускать и останавливать процессы компонентов.

Регистрация на консоли администратора

Для управления поисковой системой предприятия надо ввести URL в браузере и зарегистрироваться на консоли администратора.

Прежде, чем вы начнете

Регистрироваться нужно под ID пользователя, авторизованным для доступа к консоли администратора поисковой системы предприятия:

- Если в WebSphere Application Server не включена глобальная защита, только администратор поисковой системы предприятия, заданный при установке WebSphere II OmniFind Edition, может использовать консоль администратора.
- Если вы включили глобальную защиту в WebSphere Application Server, можно использовать консоль администратора поисковой системы предприятия для конфигурирования административных ролей. Конфигурируемые ID пользователей должны существовать в реестре пользователей WebSphere Application Server. При конфигурировании ролей администраторов вы позволяете регистрироваться на консоли администратора большему числу пользователей, но зато можете управлять тем, к каким функциям и собраниям есть доступ у каждого пользователя-администратора.

Порядок действий

Чтобы зарегистрироваться на консоли администратора поисковой системы предприятия:

1. В браузере введите URL консоли администратора. Например:

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

SearchServer.com - имя хоста поискового сервера для поисковой системы предприятия.

В зависимости от конфигурации вашего Web-сервера может потребоваться задание номера порта. Например:

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

2. На странице приветствия введите свои ID пользователя и пароль, после чего нажмите кнопку **Регистрация**.

Откроется окно Собрания - точка входа для управления системой и собраниями.

Если используются роли администраторов, возможные варианты действий и видимые собрания зависят от вашей роли администратора.

Если сеанс не активен в течение некоторого времени, система автоматически отменяет регистрацию. Чтобы продолжить управлять системой, зарегистрируйтесь заново.

Закончив операции управления собраниями, можете выбрать **Выход**, чтобы выйти с консоли. Затем вы можете снова зарегистрироваться с другим ID и паролем или закрыть браузер, чтобы выйти с консоли администратора.

Изменение пароля администратора поисковой системы предприятия для конфигурации с одним сервером

Пароль администратора поисковой системы предприятия хранится в зашифрованном виде. Чтобы изменить этот пароль, используйте сценарий `eschangerpw`.

Прежде, чем вы начнете

ID и пароль администратора поисковой системы предприятия должны быть действительными в вашей операционной системе и обладать полномочиям доступа и конфигурирования для DB2 Universal Database.

Об этой задаче

Пароль для исходного ID администратора поисковой системы предприятия задается при установке WebSphere II OmniFind Edition.

Чтобы изменить пароль, надо запустить сценарий `eschangerpw`, который распространит изменение по всей поисковой системе предприятия. Программа установки создает две переменные среды, которые можно использовать со сценарием `eschangerpw`:

ES_INSTALL_ROOT

Каталог установки поисковой системы предприятия.

ES_NODE_ROOT

Каталог данных поисковой системы предприятия. Пароль для ID администратора поисковой системы предприятия хранится в файле `es.cfg` в этом каталоге.

Поскольку сценарий `eschangerpw` установлен в каталоге `ES_INSTALL_ROOT/bin`, его можно запустить из любой точки системы.

Порядок действий

Для изменения пароля администратора поисковой системы предприятия в конфигурации с одним сервером:

1. Зарегистрируйтесь как администратор поисковой системы предприятия.
2. Введите команду остановки сервера: `esadmin stop`
3. Откройте консоль администратора WebSphere Application Server и остановите программы предприятия `server1` и `ESSearchServer`.
4. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы UNIX или утилитой изменения пароля Microsoft Windows.
5. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 4:

Операционная система	Команда
UNIX	<code>eschangepw.sh новое_значение</code>
Windows	<code>eschangepw новое_значение</code>

6. На консоли администратора WebSphere Application Server запустите программы `server1` и `ESSearchServer`.
7. Повторно запустите слой общей связи (CCL) WebSphere II OmniFind Edition, введя команды:

Операционная система	Команды
UNIX	<code>stopccl.sh</code> , затем <code>startccl.sh -bg</code>
Окно командной строки Windows	<code>stopccl</code> , затем <code>startccl</code>
Администрирование - Службы Windows	<ol style="list-style-type: none">1. Запустите Службы Windows.2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Стоп.3. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Свойства.4. Щелкните по вкладке Вход в систему.5. Введите новый пароль <i>новое_значение</i>, затем нажмите кнопку ОК.6. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Пуск.

8. Введите команду для перезапуска поисковой системы предприятия: `esadmin start`.

Изменение пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами

Пароль администратора поисковой системы предприятия хранится в зашифрованном виде. Чтобы изменить пароль, запустите сценарий `eschangepw`, который изменит его на всех компьютерах в поисковой системе предприятия.

Прежде, чем вы начнете

ID и пароль администратора поисковой системы предприятия должны быть действительными в вашей операционной системе и обладать полномочиям доступа и конфигурирования для DB2 Universal Database.

Пароль для администратора поисковой системы предприятия должен быть одним и тем же на всех компьютерах, принадлежащим данной установке WebSphere II OmniFind Edition.

Об этой задаче

Пароль для исходного ID администратора поисковой системы предприятия задается при установке WebSphere II OmniFind Edition.

Чтобы изменить пароль и распространить изменение по всей поисковой системе предприятия, надо запустить сценарий `eschangerw` на каждом компьютере в конфигурации с несколькими серверами. Программа установки создает две переменные среды, которые можно использовать со сценарием `eschangerw`:

ES_INSTALL_ROOT

Каталог установки поисковой системы предприятия.

ES_NODE_ROOT

Каталог данных поисковой системы предприятия. Пароль для ID администратора поисковой системы предприятия хранится в файле `es.cfg` в этом каталоге.

Поскольку сценарий `eschangerw` установлен в каталоге `ES_INSTALL_ROOT/bin`, его можно запустить из любой точки системы.

Порядок действий

Чтобы изменить пароля администратора поисковой системы предприятия для конфигурации с несколькими серверами:

1. На компьютере, где вы установили сервер индексов, зарегистрируйтесь как администратор поисковой системы предприятия.
2. Введите команду для запуска режима обслуживания системы: `esadmin system maintenance`
3. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы UNIX или утилитой изменения пароля Microsoft Windows.
4. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 3:

Операционная система	Команда
UNIX	<code>eschangerw.sh новое_значение</code>
Windows	<code>eschangerw новое_значение</code>

5. Повторно запустите слой общей связи (CCL) WebSphere II OmniFind Edition, введя команды на сервере индексов:

Операционная система	Команды
UNIX	<code>stopccl.sh</code> , затем <code>startccl.sh -bg</code>
Окно командной строки Windows	<code>stopccl</code> , затем <code>startccl</code>

Операционная система	Команды
Администрирование - Службы Windows	<ol style="list-style-type: none"> 1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Стоп. 3. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Свойства. 4. Щелкните по вкладке Вход в систему. 5. Введите новый пароль <i>новое_значение</i>, затем нажмите кнопку ОК. 6. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Пуск.

6. На компьютере, где вы установили сервер искателя, зарегистрируйтесь как администратор поисковой системы предприятия.
7. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы UNIX или утилитой изменения пароля Windows. Этот пароль должен соответствовать паролю, который вы задали на шаге 3 на стр. 21.
8. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 3 на стр. 21:

Операционная система	Команда
UNIX	<code>eschangepw.sh новое_значение</code>
Windows	<code>eschangepw новое_значение</code>

9. Снова запустите CCL на сервере искателя, введя команды:

Операционная система	Команды
UNIX	<code>stopccl.sh</code> , затем <code>startccl.sh -bg</code>
Окно командной строки Windows	<code>stopccl</code> , затем <code>startccl</code>
Администрирование - Службы Windows	<ol style="list-style-type: none"> 1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Стоп. 3. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Свойства. 4. Щелкните по вкладке Вход в систему. 5. Введите новый пароль <i>новое_значение</i>, затем нажмите кнопку ОК. 6. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Пуск.

10. На одном из компьютеров, где вы установили сервер поиска, зарегистрируйтесь как администратор поисковой системы предприятия.
11. Остановите IBM HTTP Web Server. (Это необходимо, чтобы программа Network Dispatcher не направляла запросы на данный сервер.)

12. Остановите CCL на сервере поиска, введя команды:

Операционная система	Команды
UNIX	stopccl.sh
Окно командной строки Windows	stopccl
Администрирование - Службы Windows	<ol style="list-style-type: none"> 1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Стоп.

13. Откройте консоль администратора WebSphere Application Server и остановите программы server1 и ESSearchServer.

14. Измените пароль системы для ID пользователя администратора поисковой системы предприятия при помощи команд операционной системы UNIX или утилитой изменения пароля Windows. Этот пароль должен соответствовать паролю, который вы задали на шаге 3 на стр. 21.

15. Запустите указанный ниже сценарий, задав в качестве *newValue* пароль, который вы задали на шаге 3 на стр. 21:

Операционная система	Команда
UNIX	eschangepw.sh <i>новое_значение</i>
Windows	eschangepw <i>новое_значение</i>

16. Перезапустите CCL на сервере поиска, введя команды:

Операционная система	Команды
UNIX	startccl.sh -bg
Окно командной строки Windows	startccl
Администрирование - Службы Windows	<ol style="list-style-type: none"> 1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Свойства. 3. Щелкните по вкладке Вход в систему. 4. Введите новый пароль <i>новое_значение</i>, затем нажмите кнопку ОК. 5. Еще раз щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Пуск.

17. На втором сервере поиска повторите шаги с 10 на стр. 22 по 16.

18. Запустите IBM HTTP Web Server.

19. На консоли администратора WebSphere Application Server запустите программы предприятия server1 и ESSearchServer.

20. На компьютере, где вы установили сервер индексов, зарегистрируйтесь как администратор поисковой системы предприятия и введите команду для запуска поисковой системы предприятия: esadmin start

21. Зарегистрируйтесь на консоли администратора поисковой системы предприятия, включите монитор собрания, выберите страницу Поиск, после чего выберите **Стоп** и **Пуск**, чтобы перезапустить серверы поиска. Повторите этот шаг для каждого собрания из вашей поисковой системы предприятия.

Изменение номера порта для поисковой системы предприятия

Если номер порта, используемый поисковой системой предприятия для связи, конфликтует с номером порта, используемым другим продуктом, нужно изменить номер порта поисковой системы предприятия.

Об этой задаче

Номер порта для поисковой системы предприятия задается при установке WebSphere II OmniFind Edition. (В многосерверной конфигурации один и тот же номер порта задается на всех серверах.)

Если данный номер порта нельзя использовать (например, если он назначен для другого продукта на том же сервере), в результате конфликта в файл CCLServer_data.log (здесь data - дата создания файла журнала) будет записано следующее сообщение об ошибке:

```
FFQ00273W Internal warning java.net.BindException: Address already in use:
FFQ00273W Внутреннее предупреждение java.net.BindException: Адрес уже используется:)
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)
```

Порядок действий

Чтобы изменить номер порта, используемого поисковой системой предприятия:

1. Введите команду остановки поисковой системы предприятия:
esadmin stop
2. Перейдите на сервер, где необходимо изменить номер порта, и остановите CCL (common communication layer - общий слой связи), введя следующие команды:

Операционная система	Команды
UNIX	stopccl.sh,
Окно командной строки Windows	stopccl
Администрирование - Службы Windows	1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Стоп .

3. Отредактируйте файл \$ES_NODE_ROOT/nodeinfo/es.cfg (в UNIX) или файл %ES_NODE_ROOT%\nodeinfo\es.cfg (в Windows) (найдите следующее свойство, задайте новое значение для номера порта, а затем сохраните и закройте файл):
CCLPort=*новый_номер_порта*

4. Перезапустите CCL, введя следующие команды:

Операционная система	Команды
UNIX	startccl.sh -bg
Окно командной строки Windows	startccl

Операционная система	Команды
Администрирование - Службы Windows	<ol style="list-style-type: none"> 1. Запустите Службы Windows. 2. Щелкните правой кнопкой по WebSphere Information Integrator OmniFind Edition и выберите Запустить.

5. Перейдите на сервер индексов и выполните указания шага 2 на стр. 24, чтобы остановить CCL.
6. Отредактируйте файл \$ES_NODE_ROOT/nodeinfo/es.cfg (в UNIX) или файл %ES_NODE_ROOT%\nodeinfo\es.cfg (в Windows).
 - a. Найдите следующее свойство (где *имя_компьютера* - имя компьютера, где был изменен номер порта на шаге 3 на стр. 24, а *N* в свойстве *nodeN* - номер, идентифицирующий сервер):


```
nodeN.destination=имя_компьютера
```
 - b. Найдите следующее подсвойство, введите номер порта, заданный вами для сервера на шаге 3 на стр. 24, а затем сохраните и закройте файл:


```
nodeN.port=новый_номер_порта
```
7. Выполните указания шага 4 на стр. 24, чтобы перезапустить CCL.
8. Введите команду для перезапуска поисковой системы предприятия:


```
esadmin start
```

По завершении этой команды номер порта будет изменен на новый на всех серверах поисковой системы предприятия.

Собрания поисковой системы предприятия

Собрание поисковой системы предприятия содержит полный набор источников, в которых пользователи могут выполнять поиск при помощи одного запроса. Используя объединение, пользователи при помощи одного запроса могут выполнять поиск в нескольких собраниях.

При создании собрания вы задаете опции, применяемые ко всему собранию. Новое собрание будет пустым, пока вы не добавите в него содержимое.

Добавить собрания в поисковую систему предприятия можно двумя способами:

- Если вы еще не освоили консоль администратора поисковой системы предприятия или только изучаете, как компоненты собрания работают совместно, создайте собрание при помощи мастера по собраниям. Мастер по собраниям помогает выполнять задачи и позволяет сохранять вашу работу в качестве черного варианта создаваемого собрания.
- Если вы знакомы с консолью администратора поисковой системы предприятия, возможно, вы предпочтете создать собрание, выбирая для управления конкретные страницы в окне Собрания.

После создания собрания, используя элементы управления в окне Собрания, можно изменить собрание, поисковую систему предприятия и опции защиты, а также задать для них параметры слежения.

Объединение собраний

Если в программу поиска встроена поддержка объединения, пользователи могут выполнять поиск одновременно в нескольких собраниях. Объединение позволяет преодолеть предельно допустимый размер для собрания, равный 20 миллионам документов на собрание. Например, пользователь может выполнить поиск в двух собраниях, содержащих по 20 миллионов документов каждое.

Качество поиска зависит от областей действия, генерируемых отдельными собраниями, которые затем сливаются, образуя конечный набор результатов. Результаты будут такими же, как и при передаче двух отдельных запросов поиска с последующим объединением и ранжированием их результатов.

Создание собрания при помощи мастера по собраниям

Если вы не знакомы с поисковой системой предприятия, мастер может помочь вам при создании собрания. Мастер дает подробную информацию о каждом этапе процесса и позволяет сохранять параметры по ходу работы.

Прежде, чем вы начнете

Создать собрание может только пользователь с ролью администратора поисковой системы предприятия.

Добавить содержимое в собрание или задать опции анализа, индексирования и поиска содержимого собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче


При создании собрания можно сохранить его в черновом состоянии. Если собрание находится в черновом состоянии, любой администратор, имеющий право управлять этим собранием, может вносить в него изменения. Например, вы можете попросить администратора собрания, имеющего опыт работы с источниками Lotus Notes, сконфигурировать искатель Notes. Затем администратор собрания, имеющий опыт работы с системами UNIX, может изменить черновое собрание, чтобы сконфигурировать искатель для файловых систем UNIX.

Порядок действий

Чтобы использовать мастер по собраниям для создания собрания:

1. Выберите **Собрания**, чтобы открыть окно Собрания.
2. Выберите **Мастер по собраниям**.
3. Следуйте указаниям мастера, чтобы создать пустое собрание и добавить в него содержимое.

Нужно сконфигурировать общую информацию о собрании и создать хотя бы один искатель. Можно принять значения по умолчанию для оставшихся опций конфигурации или задать опции для вашего нового собрания.

4. Чтобы сохранить собрание до того, как вы закончите его создание, выберите **Сохранить как черновик**.
Ваше собрание находится в списке остальных черновых собраний в окне Собрания. Если для этого собрания включена защита, рядом с именем собрания будет показан значок  **Защита собрания включена**.
5. Чтобы вернуться к собранию, которое вы все еще создаете, выберите **Вернуться в мастер** в окне Собрания.
6. Нажмите кнопку **Закончить**, чтобы создать собрание.
Ваше собрание будет внесено в список остальных собраний в окне Собрания.

После создания собрания нужно запустить процессы искателя, анализатора, индексирования и поиска для этого собрания. Пока вы не будете готовы связать собрание с программами поиска, которые будут искать информацию в нем, можно использовать для поиска в вашем новом собрании пример программы поиска (под именем Default).

Создание собрания при помощи окна Собрания

Для создания пустого собрания служит окно Собрания. После этого можно отредактировать это собрание, указав опции для добавления данных в собрание и создания системы поиска в собрании.

Прежде, чем вы начнете

Создать собрание может только пользователь с ролью администратора поисковой системы предприятия.

Добавить содержимое в собрание или задать опции анализа, индексирования и поиска содержимого собрания может администратор поисковой системы предприятия или администратор этого собрания.


Об этой задаче

Чтобы получить информацию о значениях, которые можно задать для нового собрания, при создании собрания нажмите кнопку **Справка**.

Порядок действий

Чтобы создать собрание из окна Собрания:

1. В окне Собрания выберите **Создать собрание**.
2. На странице Создать собрание введите или выберите информацию для следующих полей:
 - **Имя собрания.** Введите имя, соответствующее содержанию или назначению собрания.
 - **Защита собрания.** Укажите, хотите ли вы включить защиту для этого собрания. После создания собрания этот параметр нельзя будет изменить. Если защита собрания включена, далее можно указать опции для обеспечения управления доступом на уровне отдельных документов.
 - **Важность документов (статическая модель ранжирования).** Укажите стратегию присвоения критерия статического ранжирования, который будет использоваться для упорядочения документов в результатах поиска. После создания собрания изменить это значение нельзя.
 - **Тип категоризации.** Укажите, хотите ли вы искать документы по категориям, к которым они принадлежат.
 - **Используемый язык.** Укажите язык по умолчанию для поиска документов в собрании.
3. Примите для следующих полей значения по умолчанию или задайте опции, которые надо использовать для этого собрания:
 - **Описание.** По умолчанию описание не создается.
 - **Оценка числа документов.** По умолчанию для одного собрания принимается оценка 1 миллион документов. Система использует это значение для оценки необходимых для собрания ресурсов памяти и дискового пространства, но не для ограничения размера самого собрания.
 - **Положение данных собрания.** По умолчанию файлы, имеющие отношение к собранию, находятся на сервере индексов. После создания собрания изменить это значение нельзя.
 - **ID собрания.** По умолчанию ID собрания основывается на имени собрания. После создания собрания изменить это значение нельзя. (Если вы указали пользовательский ID собрания, поисковые программы будут вызывать это собрание по этому идентификатору вместо созданного системой и, возможно, зашифрованного идентификатора.)
 - **Сегментирование энграмм.** По умолчанию используется метод сегментирования пробелами на основе Unicode. Выберите эту опцию, чтобы использовать сегментирование энграмм, только в том случае, когда ваше собрание включает в себя документы на китайском, японском или корейском языках и вы хотите, чтобы анализатор использовал для ограничения слов сегментирование энграмм вместо сегментирования пробелами. После создания собрания изменить это значение нельзя.
4. Нажмите кнопку **ОК**.

В окне Собрания среди других собраний в поисковой системе предприятия появится новое собрание. Если для этого собрания включена защита, рядом с именем собрания будет показан значок  **Защита собрания включена**.

Новое собрание будет пустым, пока вы не добавите в него содержимое. Чтобы добавить содержимое в новое собрание, выберите это собрание в окне Собрания, отредактируйте его, создайте хотя бы один искатель и задайте опции для анализа, индексации и поиска данных.

Затем надо запустить процессы для работы искателя, анализа, индексации и поиска в собрании. До того, как вы сделаете новое собрание доступным для поиска пользователями при помощи пользовательских поисковых программ, можно использовать пример поисковой программы.

Изменение собрания

При редактировании собраний вы задаете информацию о документах, которые надо включить в собрание.

Прежде, чем вы начнете


Редактировать собрание может пользователь с ролью администратора поисковой системы предприятия или администратор этого собрания.

Об этой задаче

При редактировании собрания надо задать опции просмотра искателем источников данных, анализа документов, реорганизации и обновления индекса, поиска в индексированном содержимом и записи в журнал сообщений об ошибках. При создании собрания его надо отредактировать, чтобы добавить в него содержимое. Позднее можно редактировать собрание для изменения содержимого или для изменения способа просмотра информации искателем, ее анализа, индексирования, поиска или записи в журнал.

Порядок действий

Чтобы отредактировать собрание:

1. Выберите на панели инструментов **Собрания**, чтобы открыть окно Собрания.
2. Найдите в списке собраний собрание, которое нужно отредактировать, и выберите  **Удалить** .
3. Измените параметры на любых из следующих страниц:

Общие Задайте общую информацию о собрании и просмотрите опции, которые нельзя изменить. Если при создании собрания для него включена защита, можно сконфигурировать опции управления защитой на уровне документов.

Просмотр

Задайте источники данных, которые должен просматривать искатель, и опции просмотра искателем содержимого. В каждом собрании должен быть как минимум один искатель; в одном собрании могут содержаться данные из нескольких типов источников данных. Для каждого типа источника данных, который нужно включить в собрание, надо сконфигурировать хотя бы по одному искателю.

Анализ

Задайте опции для просмотра документов искателем, их синтаксического и лексического анализа. Можно сконфигурировать категории, позволяющие пользователям выполнять поиск в подмножествах собрания, а также сконфигурировать правила, позволяющие пользователям выполнять поиск по определенным частям документов

XML и HTML. Если в поисковую систему предприятия добавлены пользовательские механизмы тестового анализа, можно выбрать механизм для анализа и составления аннотаций содержимого в этом собрании. Можно также связать поля с классами значимости для изменения ранга документов.

Индекс

Задайте расписания полной реорганизации индекса и обновления индекса с учетом нового и измененного содержимого. Можно также сконфигурировать опции для использования символов подстановки в запросах, для ограничения просмотра собрания диапазоном URI, для объединения результатов поиска с одного Web-сайта и для удаления URI из индекса.

Поиск

Задайте опции для процессов поиска, например, конфигурирования кэша поиска и выбора языка поиска. Кроме того, можно сконфигурировать прямые ссылки - эта функция позволяет возвращать заранее определенные URI, когда пользователь включает в запрос определенные слова и словосочетания. Если в поисковую систему предприятия были добавлены пользовательские словари, можно выбрать словари, которые вы хотите использовать для поиска в этом собрании.

Журнал

Задайте типы сообщений, которые должны записываться в журнал, и опции для создания файлов журнала и перехода к новым файлам журнал. Можно сконфигурировать оповещения, чтобы получать уведомления об определенных событиях, задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов. Можно также указать опции для записи в журнал информации о документах, отброшенных из поисковой системы предприятия.

Удаление собрания

Удаление собрания полностью удаляет всю информацию о собрании из поисковой системы предприятия.

Прежде, чем вы начнете

Удалить собрание может только пользователь с ролью администратора поисковой системы предприятия.

Перед удалением собрания нужно остановить все процессы, связанные с этим собранием.

Об этой задаче

Операция удаления собрания может занять определенное время. Когда вы подтвердите, что хотите удалить это собрание, система удаляет все данные в системе, связанные с этим собранием.

Совет: Сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажмите кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Процесс удаления завершен, когда имя этого собрания более не выводится в списке собраний.

Порядок действий

Чтобы удалить собрание:

1. Выберите **Собрания**, чтобы открыть окно Собрания.
2. В списке собраний найдите собрание, которое хотите удалить, и выберите

 **Удалить.**

Управление искомателями поисковой системы предприятия

Вы конфигурируете искоматели для различных типов данных, которые нужно включить в собрание. Одно собрание может содержать любое число искомателей.

Конфигурирование искомателей

При помощи консоли администратора поисковой системы предприятия можно создавать, редактировать и удалять искоматели. Обычно конфигурированием искомателя занимается специалист по типу данных для искомателя. Например, чтобы настроить искоматель для работы с источниками Lotus Notes, администратор собрания должен быть администратором Notes или работать в тесном сотрудничестве с кем-то, кто знает особенности баз данных, которые будет просматривать искоматель.

При создании искомателя мастер для конкретного типа данных, с которыми будет работать искоматель, поможет задать свойства, управляющие способом использования искомателем системных ресурсов. Этот мастер поможет также выбрать источники, в которых будет выполняться поиск.

В любое время можно внести изменения в существующие искоматели. При необходимости можно изменить свойства искомателя или частей пространства для искомателя. Внести эти изменения помогут также мастера по искомателям.

Заполнение нового искомателя основными значениями

При создании искомателя можно использовать системные значения по умолчанию или скопировать значения, заданных для другого искомателя того же типа. Если взять за основу для нового искомателя существующий искоматель, можно быстро создать несколько искомателей с одинаковыми свойствами, а затем сконфигурировать их, например, для просмотра различных источников или для работы по различным расписаниям.

Создав копии искомателя, можно разделить рабочую нагрузку по нескольким искомателям, использующим одни и те же правила работы с данными. Например, можно создать копию искомателя Notes, чтобы использовать те же свойства и правила просмотра полей искомателем на другом сервере Lotus Notes. Они могут отличаться только наборами баз данных, просматриваемых каждым искомателем, и установками защиты на уровне документов.

Комбинирование типов искомателей в собрании

Искоматели поисковой системы предприятия предназначены для сбора информации из источников данных различных типов. При конфигурировании искомателей для собрания нужно решить, как скомбинировать эти различные типы источников данных, чтобы пользователи могли легко выполнить поиск данных вашего предприятия. Например, чтобы пользователи могли выполнять поиск данных в файловых системах Microsoft Windows и общедоступных папках Microsoft Exchange Server при помощи одного запроса, создайте собрание, включающее искоматели для файловых систем Windows и искоматели Exchange Server.

Если в одном собрании будут работать искоматели нескольких типов, убедитесь, что все они могут использовать один метод статического ранжирования. (Метод статического ранжирования задается при создании собрания.) Например, если вы

объединяете источники Web (для которых для ранжирования используются ссылки на документы) и источники (для которых для ранжирования обычно используются даты документов), это может негативно повлиять на качестве результатов поиска.

Защита на уровне документов

Если при создании собрания вы включили для него защиту, можно сконфигурировать опции защиты на уровне документов. Каждый искатель может связать с просматриваемыми им документами маркеры защиты. Если при конфигурировании искателя задать использование защиты на уровне документов, искатель свяжет заданные вами маркеры защиты с каждым документом, и они будут добавлены с документами в индекс.

Если в пользовательских программах поиска включить защиту, эти программы смогут использовать маркеры защиты, связываемые искателями с документами, для аутентификации пользователей. Эта возможность позволяет ограничить доступ к некоторым документам в собрании и разрешить всем пользователям выполнять поиск в других документах. Например, в одном собрании можно разрешить доступ ко всем документам в общедоступных папках Microsoft Exchange Server всем пользователям, но доступ к документам в базе данных Lotus Notes разрешить только пользователям с конкретными ID пользователей.

Можно применять пользовательские бизнес-правила для определения значения маркеров защиты путем шифрования правил в классе Java. При конфигурировании свойств искателя вы задаете имя подключаемого модуля, который будет использовать искатель при просмотре документов. Маркеры защиты, добавляемые этим подключаемым модулем, сохраняются в индексе и могут использоваться для управления доступом к документам.

При конфигурировании определенных типов искателей можно задать дополнительные элементы управления защитой. Например, можно указать, что вы хотите выполнять проверку пользователей во время обработки запросов. Если включить эту опцию, учетные данные пользователя сравниваются с текущими списками управления доступом, поддерживаемыми источниками данных, в которых будет выполняться поиск. Такая проверка текущих учетных данных может выполняться вместо или помимо проверки на основе маркеров защиты в индексе поисковой системы предприятия.

Составление расписаний искателей

Искатели, создаваемые для источников Web, NNTP и WebSphere Portal, работают постоянно. После запуска таких искателей, как правило, их можно не останавливать, пока не потребуется изменить конфигурацию искателя.

При конфигурировании искателей всех остальных типов нужно составить расписание просмотра ими данных. Для некоторых типов источников данных одно расписание управляет временем обращения искателя ко всем источникам данных в пространстве искателя. Для других источников данных можно задать разные расписания обращения к конкретным источникам данных. Например, можно задать отдельные расписания просмотра для каждой просматриваемой искателем базы данных Lotus Notes.

При конфигурировании расписания задается тип выполняемой работы искателя. Можно запланировать полную обработку искателем всех документов в пространстве работы искателя, запланировать работу искателя, включающую все обновления его рабочего пространства (новые документы, измененные документы и удаленные

документы) или запланировать работу искателя, включающую в себя только новые или измененные документы. Больше всего времени требуется для полной обработки искателем. Работа искателя с удалением удаленных документов требует больше времени чем работа, при которой удаленные документы игнорируются.

При редактировании пространства работы искателя можно задать второе расписание работы искателя. Например, можно сконфигурировать одно расписание для обработки всех документов в пространстве работы искателя в субботу вечером, а другое расписание - для более частой обработки новых и измененных документов.

Создание нескольких расписаний искателя позволяет точнее управлять временем обращения искателя к источникам назначения. Например, для просмотра искателем баз данных в разных часовых поясах можно задать расписание работы искателя на время наиболее вероятного окончания рабочего дня пользователями в этих поясах.

Создание искателя

При создании искателя вы задаете его тип. Мастер помогает задать информацию о данных, которую вы хотите включить в собрание.

Прежде, чем вы начнете

Создать искатель может пользователь с ролью администратора поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

Для собрания нужно создать хотя бы один искатель. Тип создаваемого искателя зависит от типа данных, которые вы хотите включить в это собрание. Мастер для создаваемого типа искателя поможет задать опции искателя. Например, мастер поможет задать опции использования ресурсов системы для искателя. Мастер поможет также выбрать источники данных, которые вы хотите включить в собрание.

Порядок действий

Чтобы создать искатель:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Создать искатель**.
2. Выберите тип искателя и основные значения для искателя:
 - a. Выберите тип искателя, поддерживающий тип данных, в которых вы собираетесь выполнять поиск - сайты, базы данных Lotus Notes или файловые системы UNIX.
Когда вы выберете тип искателя, будут показаны варианты создания искателя.
 - b. Выберите основные значения искателя:

Использовать для нового искателя системные значения по умолчанию

Заполняет исходные значения для этого искателя установочными значениями по умолчанию.

Если вы выбрали эту опцию, нажмите кнопку **Далее**, чтобы начать конфигурирование нового искателя.

Скопировать значения из существующего искателя для нового искателя

Заполняет начальные значения для этого искателя значениями, сконфигурированными для другого искателя этого типа.

Если вы выбрали эту опцию, будет показан список искателей этого типа. Выберите искатель, который вы хотите использовать как основу для нового искателя, и нажмите кнопку **Далее**, чтобы начать конфигурирование нового искателя.

Откроется мастер для создаваемого типа искателя. Чтобы создать искатель, следуйте подсказкам мастера. Чтобы узнать больше об опциях, которые вы можете задать для этого типа искателя, выберите **Помощь** на любой странице мастера.

Ваш новый искатель находится в списке других искателей, принадлежащих этому собранию, на странице Искатель. Если вам нужно внести изменения в искатель, выберите Опции, чтобы изменить свойства искателя и пространство для искателя.

Изменение свойств искателя

Можно изменять информацию о искателе и о том, как он обрабатывает данные. Например, можно изменить использование искателем ресурсов системы.

Прежде, чем вы начнете


Редактировать свойства искателя может только администратор поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

При редактировании свойств искателя выберите **Помощь**, чтобы узнать больше об изменениях, которые вы можете внести. Доступные для изменения свойства зависят от типа искателя.

Порядок действий

Чтобы изменить свойства искателя:

1. Откройте собрание для редактирования, на странице Искатель найдите искатель, который вы хотите отредактировать, и выберите  **Свойства искателя**.
2. Измените свойства искателя и нажмите кнопку **ОК**.
3. Чтобы изменения вступили в силу, остановите и перезапустите искатель. (Если вы изменили только описание искателя, перезапуск не требуется.)

Изменение пространства для искателя

Можно изменять информацию о источниках данных, которые просматривает искатель. Например, можно добавлять источники данных, удалять источники данных, изменять расписание просмотра и правила просмотра документов из определенного источника данных.

Прежде, чем вы начнете


Редактировать пространство искателя может только администратор поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.

Об этой задаче

Чтобы узнать об изменениях, которые вы можете внести для типа искателя, который вы редактируете, выберите **Помощь** во время изменения пространства для искателя.

Порядок действий

Чтобы редактировать пространство для искателя:

1. Откройте собрание для редактирования, на странице Искатель найдите искатель, который вы хотите отредактировать, и выберите  **Пространство для искателя**.
2. Измените пространство для искателя, выбрав опции, которые вы хотите изменить.

Опции, которые можно выбрать, зависят от типа искателя. Для некоторых опций, таких, как добавление источников данных в собрание, открывается мастер для этого типа искателя, который поможет вам изменить пространство для искателя.

3. Чтобы изменения вступили в силу, остановите и перезапустите искатель.

Удаление искателя

При удалении искателя вся информация о нем из поисковой системы предприятия удаляется. Информация, ранее найденная искателем, остается в индексе до тех пор, пока не будет выполнена его реорганизация.

Прежде, чем вы начнете

Удалить искатель может пользователь с ролью администратора поисковой системы предприятия или же администратор собрания, которому принадлежит этот искатель.


Об этой задаче

Операция удаления искателя может занять определенное время. Когда вы подтвердите, что хотите удалить этот искатель, система удаляет все данные в системе, связанные с этим искателем.

Совет: Для этого может потребоваться время, и сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажимайте кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Когда имя искателя исчезнет из списка искателей, процесс удаления завершился.

Порядок действий

Чтобы удалить искатель:

1. Откройте собрание для редактирования и выберите страницу Искатель.
2. Найдите искатель, который вы хотите удалить, и выберите  **Удалить**.

Искатели Content Edition

Чтобы включить репозитории IBM WebSphere Information Integrator Content Edition в собрание поисковой системы предприятия, нужно сконфигурировать искатель Content Edition.

При помощи искателя Content Edition можно просматривать репозитории Documentum, FileNet Panagon Content Services, FileNet P8 Content Manager, Hummingbird Document Management (DM), OpenText Livelink и Portal Document Manager (PDM).

При конфигурировании искателя нужно задать опции для просмотра искателем всех репозиториях в пространстве искателя. Нужно также выбрать типы классов, которые должен просматривать искатель в каждом репозитории.

Чтобы создать или изменить искатель Content Edition, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех репозиториях в пространстве искателя.
- Указать, использует ли искатель при доступе к репозиториям прямой или же серверный режим. Для серверного режима можно также задать информацию, обеспечивающую искателю доступ к серверу прикладных программ Web.
- Выбрать репозитории, которые будет просматривать искатель.
- Задать ID пользователей и пароли для возможности обращения искателя к содержимому в выбранных репозиториях.
- Составить расписания просмотра искателем репозиториях.
- Выбрать в каждом репозитории классы элементов, просматриваемые искателем.
- Задать опции, обеспечивающие доступность поиска для свойств классов элементов. Например, можно исключить из пространства искателя определенные типы документов или указать, что нужен просмотр искателем конкретной версии репозитория.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволят программам поиска применять управление доступом на основе списков управления доступом или маркеров защиты.

Для классов элементов Documentum, FileNet Panagon Content Services и Portal Document Manager можно также выбрать опцию для проверки правильности учетных данных пользователя во время передачи пользователем запроса. В этом случае система сравнивает учетные данные пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных. (Этот тип проверки текущих учетных данных недоступен для других типов репозиториях.)

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Доступ к репозиториям WebSphere II Content Edition в режиме сервера

Для искателя Content Edition можно сконфигурировать доступ к репозиториям в режиме сервера.

В режиме сервера соединитель WebSphere Information Integrator Content Edition, используемый искателем для доступа к данным, устанавливается как прикладная программа предприятия на сервер WebSphere Application Server, и искатель обращается к репозиториям через сервер. Этот подход позволяет воспользоваться преимуществами среды сервера прикладных программ J2EE.

Прежде чем конфигурировать для искателя доступ к репозиториям WebSphere Information Integrator Content Edition в режиме сервера, необходимо запустить сценарий на сервере искателя. Этот сценарий, поставляемый с WebSphere Information Integrator OmniFind Edition, разрешает искателю Content Edition обращаться к репозиториям на сервере.

Перед тем, как при помощи консоли администратора поисковой системы предприятия сконфигурировать искатель Content Edition для использования режима сервера, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition”.
- “Конфигурирование сервера искателей в Windows для WebSphere II Content Edition” на стр. 40.

Доступ к репозиториям WebSphere II Content Edition в непосредственном режиме

Для искателя Content Edition можно сконфигурировать доступ к репозиториям в непосредственном режиме.

В непосредственном режиме искатель использует соединитель WebSphere Information Integrator Content Edition, устанавливаемый на сервере искателя при установке WebSphere II OmniFind Edition.

Порядок действий

Чтобы сконфигурировать в системе возможность доступа к репозиториям в непосредственном режиме:

1. Убедитесь, что переменные среды VBR_HOME и JAVA_HOME в файле *каталог_установки_ice/bin/config.sh* (в UNIX) или в файле *каталог_установки_ice\bin\config.bat* (в Microsoft Windows) задают правильный каталог.
2. Чтобы сконфигурировать запуск консоли администратора WebSphere Information Integrator Content Edition в непосредственном режиме, добавьте системное свойство Java **vbr.as.operationMode=direct** в файл *каталог_установки_ice/bin/Admin.bat* (в UNIX) или в файл *каталог_установки_ice\bin\Admin.bat* (в Windows).
3. Запустите консоль администратора WebSphere Information Integrator Content Edition в непосредственном режиме и сконфигурируйте соединитель для сервера искателя WebSphere II OmniFind Edition. (Смотрите инструкции в документации по WebSphere Information Integrator Content Edition.)
4. Когда вы будете на консоли администратора WebSphere II OmniFind Edition конфигурировать искатель Content Edition, выберите опцию непосредственного режима.

Конфигурирование сервера искателей в UNIX для WebSphere II Content Edition

Если вы устанавливаете WebSphere II OmniFind Edition на компьютере, где запускается IBM AIX, Linux или операционная среда Solaris, и конфигурируете для искателя Content Edition использование режима сервера при обращении к репозиториям, нужно запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю Content Edition обращаться к репозиториям WebSphere Information Integrator Content Edition.

Об этой задаче

Искатель Content Edition использует библиотеки Java WebSphere Information Integrator Content Edition как клиент Java. В режиме сервера для этих библиотек Java требуются библиотеки Java WebSphere Application Server, связанные с EJB. Чтобы искатель Content Edition мог работать с этими библиотеками Java, после установки WebSphere Application Server нужно запустить на сервере искателя сценарий установки, входящий в WebSphere II OmniFind Edition.

WebSphere Information Integrator Content Edition устанавливается на сервере искателя при установке WebSphere II OmniFind Edition. Для возможности использования искателя Content Edition в режиме сервера на сервере искателя нужно скопировать файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition.

Порядок действий

Чтобы сконфигурировать сервер искателя для возможности просмотра репозитория WebSphere Information Integrator Content Edition:

1. Если устанавливается WebSphere II OmniFind Edition в многосерверной конфигурации, установите и свяжите библиотеки Java WebSphere Application Server.
2. На сервере искателя запустите сценарий настройки для искателя Content Edition:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия.
 - b. Запустите следующий сценарий (устанавливаемый в каталоге `$ES_INSTALL_ROOT/bin`) и ответьте на вопросы программы:
`escrvbr.sh`
3. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):

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

4. Скопируйте файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition на сервер искателя.

Скопируйте из:

Файл `vbr_access_services.jar` находится в следующем положении по умолчанию:

корневой_к-лог_уст_was/installedApps/имя_серв/имя_пр-мы

корневой_к-лог_установки_was - каталог установки WebSphere Application Server, *имя_серв* - имя, заданное для этого сервера, *имя_пр-мы* - имя, заданное для прикладной программы WebSphere Information Integrator Content Edition в WebSphere Application Server.

Скопируйте в:

Каталог назначения на сервере искателя - *корневой_каталог_установки_iice/lib*, где *корневой_каталог_установки_iice* - каталог установки WebSphere Information Integrator Content Edition на сервере искателя.

Конфигурирование сервера искателей в Windows для WebSphere II Content Edition

Если вы устанавливаете WebSphere II OmniFind Edition на компьютере Microsoft Windows и конфигурируете для искателя Content Edition использование режима

сервера при обращении к репозиториям, нужно запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю Content Edition обращаться к репозиториям WebSphere Information Integrator Content Edition.

Об этой задаче

Искатель Content Edition использует библиотеки Java WebSphere Information Integrator Content Edition как клиент Java. В режиме сервера для этих библиотек Java требуются библиотеки Java WebSphere Application Server, связанные с EJB. Чтобы искатель Content Edition мог работать с этими библиотеками Java, после установки WebSphere Application Server нужно запустить на сервере искателя сценарий установки, входящий в WebSphere II OmniFind Edition.

WebSphere Information Integrator Content Edition устанавливается на сервере искателя при установке WebSphere II OmniFind Edition. Для возможности использования искателя Content Edition в режиме сервера на сервере искателя нужно скопировать файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition.

Порядок действий

Чтобы сконфигурировать сервер искателя для возможности просмотра репозитория WebSphere Information Integrator Content Edition:

1. Если устанавливается WebSphere II OmniFind Edition в многосерверной конфигурации, установите и свяжите библиотеки Java WebSphere Application Server.
2. На сервере искателя запустите сценарий настройки для искателя Content Edition:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке WebSphere II OmniFind Edition).
 - b. Запустите следующий сценарий (устанавливаемый в каталоге `%ES_INSTALL_ROOT%\bin`) и отвечайте на вопросы программы:
`escrvbr.vbs`
3. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):
 - a. Из командной строки остановите поисковую систему предприятия:
`esadmin stop`
 - b. Выберите **Пуск** → **Программы** → **Администрирование** → **Службы** и перезапустите службу IBM WebSphere Information Integrator OmniFind Edition.
 - c. Из командной строки запустите поисковую систему предприятия:
`esadmin start`
4. Скопируйте файл `vbr_access_services.jar` с сервера WebSphere Information Integrator Content Edition на сервер искателя.

Скопируйте из:

Файл `vbr_access_services.jar` находится в следующем положении по умолчанию:

`корневой_к-лог_уст_was\installedApps\имя_серв\имя_пр-мы`

`корневой_к-лог_уст_was` - каталог установки WebSphere Application Server, `имя_серв` - имя, заданное для этого сервера, `имя_пр-мы` - имя, заданное для прикладной программы WebSphere Information Integrator Content Edition в WebSphere Application Server.

Скопируйте в:

Каталог назначения на сервере искателя -

корневой_каталог_установки_iice\lib, где *корневой_каталог_установки_iice* - каталог установки WebSphere Information Integrator Content Edition на сервере искателя.

Искатели DB2

С помощью искателя DB2 можно добавить базы данных IBM DB2 Universal Database в собрание. При помощи искателя DB2 можно также добавить таблицы псевдонимов, создаваемые вами для IBM DB2 Universal Database for z/OS, IBM Informix, Oracle и Microsoft SQL Server.

Для каждого просматриваемого сервера баз данных нужно сконфигурировать отдельный искатель. При конфигурировании искателя нужно задать опции просмотра искателем всех баз данных на одном сервере. Кроме того, в каждой базе данных надо выбрать конкретные таблицы, которые будет просматривать искатель.

Таблицы, которые вы выбираете для работы искателя, могут быть таблицами базы данных, таблицами псевдонимов или производными таблицами. Искатель DB2 не поддерживает объединенные таблицы.

Публикация событий

Если вы используете WebSphere Information Integrator Event Publisher Edition и связали базы данных, которые хотите обработать искателем, с картами очередей публикации, искатель DB2 может использовать эти карты для просмотра обновлений таблиц этих баз данных.

Карта очередей публикации идентифицирует очередь WebSphere MQ, получающую сообщения XML при публикации обновлений таблицы базы данных. Искатель принимает из этой очереди информацию о публикации событий и обновляет свое пространство работы при обновлении таблиц (при первой обработке таблицы искателем обрабатываются все документы).

Публикация событий позволяет сделать новые и измененные документы доступными для поиска быстрее, чем при работе искателя по расписанию.

Если некоторые или все таблицы сконфигурированы для использования публикации событий, при конфигурировании искателя вы можете задать информацию, которая обеспечит искателю доступ к WebSphere MQ и картам очередей публикации.

Необходимо также проверить, что WebSphere MQ и WebSphere Information Integrator Event Publisher Edition сконфигурированы на сервере, на котором будет работать искатель, и что модуль клиента WebSphere MQ сконфигурирован на сервере искателя. Чтобы использовать публикацию событий с искателем DB2, выполните следующие задачи:

- “Конфигурирование WebSphere MQ для искателей DB2” на стр. 46.
- “Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2” на стр. 43.
- “Конфигурирование сервера искателей в UNIX для использования публикации событий” на стр. 47.
- “Конфигурирование сервера искателей в Windows для использования публикации событий” на стр. 48.

Конфигурирование - обзор

Чтобы создать или изменить искатель DB2, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех баз данных на конкретном сервере баз данных.
- Задать информацию о типах баз данных, которые будет просматривать искатель. Если планируется просматривать удаленные базы данных, не внесенные в каталог локального сервера баз данных, для возможности использования искателя DB2 для просмотра этих баз данных необходимо запустить сервер администратора DB2. При конфигурировании искателя нужно также задать имя хоста и порт удаленного сервера баз данных.
- Выбрать базу данных для просмотра.
- Задать ID пользователей и пароли для возможности обращения искателя к базам данных, где используется управление доступом.
- Составить расписания просмотра искателем баз данных.
- Выбрать в каждой базе данных таблицы, просматриваемые искателем.

Внимание: Чтобы оптимизировать производительность процессов обнаружения (и предотвратить истечение срока ожидания для процесса конфигурирования искателя), выбирайте просмотр искателем всех таблиц, только если в базе данных немного таблиц или в таблицах немного столбцов. Если сейчас выбрать определенные таблицы, впоследствии можно будет изменить пространство для искателя и добавить в собрание дополнительные таблицы.

- Выберите таблицы, которые будут обрабатываться искателем при публикации их обновлений в очереди публикации событий, и укажите информацию для доступа искателя к очереди публикации событий.
- Задать опции, обеспечивающие доступность поиска для столбцов в определенных таблицах. Например, можно включить использование конкретных столбцов в параметрических запросах или задать, какие столбцы будут возвращаться в результатах поиска.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование WebSphere Information Integrator Event Publisher Edition для искателей DB2

Перед тем, как сконфигурировать искатель DB2 для использования публикации событий, проверьте, что на сервере, который будет обрабатываться искателем, сконфигурирована IBM WebSphere Information Integrator Event Publisher Edition.

Об этой задаче

При конфигурировании WebSphere Information Integrator Event Publisher Edition для использования с искателем DB2 руководствуйтесь следующими указаниями:

- В таблицах-источниках для публикации должны быть выбраны как измененные, так и неизменные столбцы.
- Удаленные строки в таблицах-источниках должны быть выбраны для публикации.
- Очередь публикации событий не может совместно использоваться несколькими базами данных.
- У одной базы данных может быть несколько карт очередей и несколько очередей.
- У таблицы должна быть одна публикация XML, связанная с одной картой очередей публикации. (У таблицы не может быть нескольких публикаций XML, связанных с одной картой очередей публикации. У таблицы может быть несколько публикаций XML, если каждая из публикаций XML связана со своей картой очередей публикации.)

Порядок действий

Выполните следующие шаги, чтобы сконфигурировать сервер баз данных так, чтобы искатель DB2 мог получать обновления таблиц, опубликованные в очереди публикации событий. (В качестве помощи при этих шагах используйте документацию к WebSphere Information Integrator Publisher Edition.)

1. Установите WebSphere Information Integrator Event Publisher Edition на сервере баз данных, который будет обрабатываться искателем.
2. Запустите панель запуска центра репликации:

Операционная система	Команда
UNIX	db2rc
Windows	Выберите Пуск → Центр репликации IBM DB2

3. Создайте управляющие таблицы Q Capture:
 - a. Выберите окно панели запуска **Публикация событий**, выберите **Создать управляющие таблицы Q Capture** и нажмите кнопку **Далее**.
 - b. В поле **Сервер Q Capture** выберите из списка доступных серверов баз данных сервер, который вы хотите использовать в качестве сервера Q Capture, и нажмите кнопку **ОК**.
 - c. Укажите ID пользователя и пароль для доступа к выбранному серверу Q Capture. Измените схему Q Capture или примите имя схемы по умолчанию, затем нажмите кнопку **Далее**.
 - d. Укажите имена менеджера очередей и очереди управления и перезапустите очередь, указанную вами при конфигурировании WebSphere MQ на этом сервере баз данных, а затем нажмите кнопку **Далее**.
 - e. Нажмите кнопку **Готово**. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - f. Для опции обработки выберите **Запустить сейчас** и нажмите кнопку **ОК**. Когда появится сообщение о завершении сценариев SQL, нажмите кнопку **Заккрыть**.
4. Создайте публикацию XML:
 - a. На панели запуска Центра репликации выберите окно **Публикация событий**, выберите **Создать публикацию XML** и нажмите кнопку **Далее**.
 - b. На странице **Пуск** нажмите кнопку **Далее**.

- c. На странице Сервер и карта очередей подтвердите правильность сервера Q Capture и схемы Q Capture, затем нажмите кнопку рядом с полем **Карта очередей публикации** и нажмите кнопку **Создать**, чтобы создать карту очередей публикации.
 - d. На странице Общие введите имя карты очередей.
 - e. На странице Свойства укажите имя очереди отправления (например, имя очереди данных, указанное вами при конфигурировании WebSphere MQ на этом сервере), для типа содержимого сообщения выберите либо **Операция со строками** или **Транзакция**, выключите переключатели для отправки сообщений о работоспособности и добавления заголовков сообщений JMS и нажмите кнопку **ОК**.
 - f. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - g. Выберите опцию обработки **Запустить сейчас** и нажмите кнопку **ОК**. После того, как появится сообщение, указывающее, что сценарии SQL завершены, нажмите кнопку **Заккрыть**.
 - h. На странице Выберите карту очередей публикации выберите созданную вами очередь публикации и нажмите кнопку **ОК**.
 - i. На странице Сервер и карта очередей подтвердите правильность имени очереди публикации и нажмите кнопку **Далее**.
 - j. На странице Исходная таблица нажмите кнопку **Добавить**, выберите **Получить все**, выделите таблицу, для которой вы хотите включить публикацию событий, и нажмите кнопку **ОК**, а затем **Далее**.
 - k. На странице Столбцы и строки выберите столбец (или все столбцы), который должен обрабатывать искатель DB2, затем выберите столбцы ключа. На странице, на которой вы выбрали строки для обработки искателем (или задали все строки), выберите опцию публикации удалений из исходной таблицы. После завершения конфигурирования этих опций нажмите кнопку **Далее**.
 - l. На странице Содержимое сообщения выберите данных столбцов опцию включения как измененных, так и не измененных столбцов и опцию Только для новых значений данных. Проверьте, что включен переключатель для автоматического запуска публикаций XML, и нажмите кнопку **Далее**.
 - m. На странице Обзор публикаций XML нажмите кнопку **Далее**.
 - n. На странице Сводка нажмите кнопку **Готово**. Когда появится страница с сообщениями и сценариями SQL, нажмите кнопку **Заккрыть**.
 - o. Выберите опцию обработки **Запустить сейчас** и нажмите кнопку **ОК**. Когда появится сообщение о завершении сценариев SQL, нажмите кнопку **Заккрыть**.
5. Запустите сервер Q Capture:
- a. Закройте панель запуска Центра репликации и запустите Центр репликации.
 - b. В дереве объектов щелкните по **Q-репликация** → **Определения** → **Серверы Q Capture**.
 - c. Щелкните правой кнопкой мыши по значку сконфигурированного вами сервера Q Capture и выберите **Подключить базу данных для Q-репликации**.
 - d. Когда появится сообщение с предупреждением, нажмите кнопку **ОК**.
 - e. Когда появится страница с сообщениями DB2, нажмите кнопку **Заккрыть**.
 - f. В дереве объектов щелкните правой кнопкой мыши по значку для сервера Q Capture и выберите **Запустить программу Q Capture**.
 - g. Выберите опцию обработки **Запустить сейчас**, укажите имя системы, ID пользователя и пароль пользователя DB2, путь к каталогу, в котором хранятся журналы, и имя экземпляра DB2, а затем нажмите кнопку **ОК**.

- h. Когда появится сообщение об отправке требования, нажмите кнопку **Заккрыть**.
- i. В дереве объектов щелкните правой кнопкой мыши по значку для сервера Q Capture и выберите **Проверить состояние**.

Будет показано состояние сервера Q Capture. При возникновении ошибок в сообщении о состоянии будет указано, что сервер предположительно находится в нерабочем состоянии. Для просмотра журналов и определения причин ошибок введите в командной строке команду:

```
asnqcap Capture_Server=имя_сервера_Capture LOGSTDOU=y
```

Конфигурирование WebSphere MQ для искателей DB2

Перед тем, как конфигурировать искатель DB2 для использования публикации событий, проверьте, что на сервере, прием с которого будет осуществлять сервер, сконфигурирована IBM WebSphere MQ.

Прежде, чем вы начнете

Проверьте, что на сервере назначения баз данных установлены DB2 UDB, WebSphere Information Integrator Event Publisher Edition и WebSphere MQ.

Ограничения

Если сервер назначения баз данных установлен на компьютере с Linux, все пользователи DB2 Universal Database (DB2 UDB), пользователи WebSphere MQ и пользователи WebSphere II OmniFind Edition должны задать следующую переменную среды:

```
export LD_ASSUME_KERNEL=2.4.19
```

Эта переменная среды позволяет экспортировать реализацию потоков LinuxThread из любой среды, в которой была выполнена установка, при запуске как управляющих команд WebSphere MQ, так и прикладных программ WebSphere MQ. Для работы WebSphere MQ необходимо, чтобы эта переменная среды была экспортирована.

Об этой задаче

Искатель DB2 поддерживает режим соединения клиентов с сервером WebSphere MQ. Искатель принимает сообщения XML, опубликованные в очереди публикации событий. Искатель не может принимать сообщения XML, транспорт которых осуществляется посредством нескольких очередей.

После того, как вы сконфигурировали WebSphere MQ, искатель DB2 использует имя менеджера очередей, имя очереди, имя хоста сервера, номер порта сервера и имя канала сервера для получения сообщений XML из очереди публикации. Искатель анализирует эти сообщения и обновляет свое пространство при помощи информации об обновленных таблицах.

Порядок действий

Выполните следующие шаги, чтобы сконфигурировать сервер баз данных так, чтобы искатель DB2 мог принимать сообщения от очереди публикации событий. (В качестве помощи при этих шагах используйте документацию к WebSphere MQ.)

1. Зарегистрируйтесь в роли администратора WebSphere MQ и введите следующие команды для создания менеджера очередей и очередей.
 - a. Введите в командной строке команду:

```
crtmqm QM1
```


- b. После появления сообщения Установка завершена введите команду:
`strmqm QM1`
 - c. После появления сообщения 'QM1' запущен введите команду:
`runmqsc QM1`
 - d. После появления сообщения Запускается MQSC для менеджера очередей QM1 введите следующую команду, чтобы создать очередь управления:
`DEFINE QLOCAL('ASN.QM1.ADMINQ')`
 - e. После появления сообщения Очередь WebSphere MQ создана введите следующую команду, чтобы создать очередь перезапуска:
`DEFINE QLOCAL('ASN.QM1.RESTARTQ')`
 - f. После того, как сообщение Очередь WebSphere MQ создана появится снова, введите следующую команду, чтобы создать очередь данных:
`DEFINE QLOCAL('ASN.QM1.DATAQ')`
 - g. После того, как сообщение Очередь WebSphere MQ создана появится снова, введите следующую команду для выхода:
`end`
2. Введите следующую команду, чтобы запустить MQ Listener на сервере баз данных (MQ Listener должен быть запущен, когда вы создаете искатель DB2, использующий публикацию событий). В следующем примере 1414 - номер порта сервера, а канал по умолчанию - SYSTEM.DEF.SVRCONN:
`runmq1sr -m QM1 -t TCP -p 1414 &`
 3. Введите следующие команды для авторизации доступа пользователя DB2 UDB к менеджеру очередей и очередям через интерфейс Message Queuing Interface (MQI) для публикации событий (в этом примере ID пользователя - 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. Введите следующие команды для ID пользователя, который использовался для создания и запуска искателя DB2 с публикацией событий. Эти команды авторизуют данный ID пользователя для доступа к менеджеру очередей и очередям для публикации событий через интерфейс Message Queuing Interface (MQI). В этом примере используется ID пользователя 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`

Конфигурирование сервера искателей в UNIX для использования публикации событий

При установке WebSphere II OmniFind Edition на компьютере с операционной средой IBM AIX, Linux или Solaris на этапе конфигурирования искателя DB2 для использования публикации событий надо запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю обращаться к менеджерам очередей и очередям WebSphere MQ.

Об этой задаче

Искатель DB2 использует модули WebSphere MQ 5.3 для Java Messaging, чтобы обращаться к менеджерам очередей и очередям WebSphere MQ. Эти модули нужно установить на сервере искателя.

Чтобы искатель DB2 мог использовать публикацию событий, надо запустить сценарий установки, который WebSphere II OmniFind Edition помещает на сервер искателя после установки модулей WebSphere MQ.

Порядок действий

Чтобы сконфигурировать сервер искателя для использования публикации событий:

1. Установите модули WebSphere MQ 5.3 для Java Messaging на сервере искателя:
 - a. Зарегистрируйтесь как пользователь root и введите команду:

```
export LD_ASSUME_KERNEL=2.4.19
```
 - b. Вставьте компакт-диск WebSphere MQ в дисковод компакт-дисков.
 - c. Перейдите в каталог, где находятся модули MQ для Java Messaging.
 - d. Введите следующую команду, чтобы установить модули:

```
rpm -i MQSeriesJava-5.3.0-1.i386.rpm
```
2. На сервере искателя запустите сценарий настройки для искателя DB2:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия (этот ID пользователя задается при установке WebSphere II OmniFind Edition).
 - b. Запустите следующий сценарий, который установлен в каталоге `$ES_INSTALL_ROOT/bin`, и ответьте на вопросы программы:

```
escrdb2.sh
```
3. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):

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

Конфигурирование сервера искателей в Windows для использования публикации событий

При установке WebSphere II OmniFind Edition на компьютере с Microsoft Windows при конфигурировании искателя DB2 для использования публикации событий надо запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю обращаться к менеджерам очередей и очередям WebSphere MQ.

Об этой задаче

Искатель DB2 использует модули WebSphere MQ 5.3 для Java Messaging, чтобы обращаться к менеджерам очередей и очередям WebSphere MQ. Эти модули нужно устанавливать на сервере искателя.

Чтобы искатель DB2 мог использовать публикацию событий, надо запустить сценарий установки, который WebSphere II OmniFind Edition помещает на сервер искателя после установки модулей WebSphere MQ.

Порядок действий

Чтобы сконфигурировать сервер искателя для использования публикации событий:

1. Установите модули WebSphere MQ 5.3 для Java Messaging на сервере искателя:
 - a. Вставьте компакт-диск WebSphere MQ в дисковод компакт-дисков.
 - b. Запустите утилиту установки WebSphere MQ.

- c. В окне Выберите возможности продукта выберите опцию установки **Java Messaging**.
2. На сервере искателя запустите сценарий настройки для искателя DB2:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке WebSphere II OmniFind Edition).
 - b. Запустите следующий сценарий (устанавливаемый в каталоге %ES_INSTALL_ROOT%\bin) и отвечайте на вопросы программы:
escrdb2.vbs
3. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):
 - a. Из командной строки остановите поисковую систему предприятия:
esadmin stop
 - b. Выберите **Пуск** → **Программы** → **Администрирование** → **Службы** и перезапустите службу IBM WebSphere Information Integrator OmniFind Edition.
 - c. Из командной строки запустите поисковую систему предприятия:
esadmin start

Искатели DB2 Content Manager

Чтобы включить типы элементов IBM DB2 Content Manager в собрание поисковой системы предприятия, нужно сконфигурировать искатель DB2 Content Manager.

Конфигурация сервера искателя

Для возможности просмотра сервера DB2 Content Manager искателем на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с WebSphere Information Integrator OmniFind Edition, позволяет искателю DB2 Content Manager взаимодействовать с серверами DB2 Content Manager.

Перед тем, как сконфигурировать искатель DB2 Content Manager с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для DB2 Content Manager” на стр. 50.
- “Конфигурирование сервера искателей в Windows для DB2 Content Manager” на стр. 51.

Конфигурирование - обзор

Искатель DB2 Content Manager можно использовать для просмотра любого числа серверов DB2 Content Manager. При конфигурировании искателя нужно задать опции для просмотра искателем всех серверов DB2 Content Manager в пространстве искателя. Нужно также выбрать конкретные типы элементов, которые должен просматривать искатель на каждом сервере.

Чтобы создать или изменить искатель DB2 Content Manager, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют просмотром искателем элементов всех типов на всех серверах DB2 Content Manager в пространстве искателя.
- Выбрать серверы DB2 Content Manager, которые будет просматривать искатель.
- Задать ID пользователей и пароли для доступа искателя к содержимому на серверах DB2 Content Manager.
- Составить расписания просмотра искателем серверов.
- Выбрать типы элементов для просмотра искателем на каждом сервере DB2 Content Manager.
- Задать опции, обеспечивающие доступность поиска для атрибутов в определенных типах элементов. Например, можно исключить из пространства для искателя конкретные типы документов или задать, какие атрибуты будут возвращаться в результатах поиска.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование сервера искателей в UNIX для DB2 Content Manager

При установке WebSphere II OmniFind Edition на компьютере с IBM AIX, Linux или операционной средой Solaris необходимо запустить сценарий для конфигурирования сервера искателя. Этот сценарий позволяет искателю DB2 Content Manager связываться с серверами IBM DB2 Content Manager.

Об этой задаче

Для доступа к серверам DB2 Content Manager искатель DB2 Content Manager использует соединитель Java для DB2 Content Manager Версии 8. Чтобы установить этот соединитель, на сервере искателя надо установить IBM DB2 Information Integrator for Content Версии 8.2 или новее. Чтобы искатель DB2 Content Manager мог работать с DB2 Content Manager, после установки соединителя запустите сценарий настройки, установленный WebSphere II OmniFind Edition на сервере искателя.

Порядок действий

Чтобы сконфигурировать сервер искателя для просмотра серверов DB2 Content Manager:

1. Установите на сервере искателя соединитель Java для DB2 Content Manager Версии 8:
 - a. Зарегистрируйтесь на сервере искателя в качестве пользователя root:


```
su - root
```
 - b. Запустите файл db2profile. Например:


```
./home/db2inst/sqllib/db2profile
```
 - c. Экспортируйте переменную среды JAVAHOME. Например:


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

- d. Добавьте каталог Java в переменную среды PATH:
`export PATH=$PATH:$JAVAHOME/bin`
 - e. Вставьте установочный компакт-диск DB2 Information Integrator for Content и запустите мастер по установке.
 - f. В окне Выбор компонентов выполните следующие действия. (Если вы работаете с Information Integrator for Content Версии 8.3, вы увидите окно Выбор компонентов с опцией установки Пользовательская.)
 - 1) Выберите **Локальные соединители** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - 2) Выберите **Комплекты и примеры соединителей** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - g. Укажите имя базы данных, имя пользователя и пароль для библиотеки DB2 Content Manager и примите значения по умолчанию для остальных окон.
2. Зарегистрируйтесь на сервере искателя под ID пользователя, входящим в группу администраторов DB2.
 3. Добавьте в каталог базу данных удаленного библиотечного сервера DB2 Content Manager и убедитесь, что сервер искателя может соединиться с сервером DB2 Content Manager:


```
db2 catalog tcpip node имя_узла remote имя_хоста server порт
db2 catalog database имя_базы_данных as алиас at node имя_узла
```
 4. Необязательно: Зарегистрируйтесь как пользователь root и проверьте соединение с базой данных:


```
. каталог_установки_Information_Integrator_for_Content/bin/cmbenv81.sh
cd каталог_установки_Information_Integrator_for_Content/samples/java/icm
javac *.java
java SConnectDisconnect имя_базы_данных_ICM
ID_администратора_CM пароль_администратора_CM
```
 5. На сервере искателя запустите сценарий настройки для искателя DB2 Content Manager:
 - a. Перейдите в каталог ES_INSTALL_ROOT/bin:


```
cd $ES_INSTALL_ROOT/bin
```
 - b. Запустите следующий сценарий и отвечайте на вопросы:


```
esrcm.sh
```
 6. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):


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

Конфигурирование сервера искателей в Windows для DB2 Content Manager

При установке WebSphere II OmniFind Edition на компьютере Microsoft Windows нужно запустить сценарий, чтобы сконфигурировать сервер искателя. Этот сценарий позволяет искателю DB2 Content Manager связываться с серверами IBM DB2 Content Manager.

Об этой задаче

Для доступа к серверам DB2 Content Manager искатель DB2 Content Manager использует соединитель Java для DB2 Content Manager Версии 8. Чтобы установить этот соединитель, на сервере искателя надо установить IBM DB2 Information Integrator for Content Версии 8.2 или новее. Чтобы искатель DB2 Content Manager мог работать с DB2 Content Manager, после установки соединителя запустите сценарий настройки, установленный WebSphere II OmniFind Edition на сервере искателя.

Порядок действий

Чтобы сконфигурировать сервер искателя для просмотра серверов DB2 Content Manager:

1. Установите на сервере искателя соединитель Java для DB2 Content Manager Версии 8:
 - a. Вставьте установочный компакт-диск DB2 Information Integrator for Content. Будет автоматически запущена программа установки. Откроется мастер по установке DB2 Content Manager Enterprise Information Portal.
 - b. В окне Выбор компонентов выполните следующие действия. (Если вы работаете с Information Integrator for Content Версии 8.3, вы увидите окно Выбор компонентов с опцией установки Пользовательская.)
 - 1) Выберите **Локальные соединители** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - 2) Выберите **Комплекты и примеры соединителей** из списка **Компоненты**, а затем - **Соединитель Content Manager Версии 8** из списка **Подкомпоненты**.
 - c. Укажите имя базы данных, имя пользователя и пароль для библиотеки DB2 Content Manager и примите значения по умолчанию для остальных окон.
2. Добавьте в каталог базу данных удаленного библиотечного сервера DB2 Content Manager и убедитесь, что сервер искателя может соединиться с сервером DB2 Content Manager. На сервере искателя введите в командной строке следующие команды:

```
db2 catalog tcpip node имя_узла remote имя_хоста server порт
db2 catalog database имя_базы_данных as алиас at node имя_узла
```
3. Необязательно: Проверьте соединение базы данных, введя в командной строке команды:

```
cmbev81.bat
cd каталог_установки_Information_Integrator_for_Content/samples/java/icm
javac *.java
java SConnectDisconnect имя_базы_данных_ICM
ID_администратора_CM пароль_администратора_CM
```
4. На сервере искателя запустите сценарий настройки для искателя DB2 Content Manager:
 - a. Перейдите в каталог ES_INSTALL_ROOT\bin:

```
cd %ES_INSTALL_ROOT%\bin
```
 - b. Запустите следующий сценарий и отвечайте на вопросы:

```
esrcm.vbs
```
5. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):
 - a. Из командной строки остановите поисковую систему предприятия:

```
esadmin stop
```

- b. Выберите **Пуск** → **Программы** → **Администрирование** → **Службы** и перезапустите службу IBM WebSphere Information Integrator OmniFind Edition.
- c. Из командной строки запустите поисковую систему предприятия:
`esadmin start`

Искатели Domino Document Manager

Чтобы включить библиотеки и шкафы Domino Document Manager в собрание поисковой системы предприятия, нужно сконфигурировать искатель Domino Document Manager.

Конфигурация сервера искателя

Если на сервере Domino Document Manager, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с WebSphere Information Integrator OmniFind Edition, позволяет искателю Domino Document Manager взаимодействовать с серверами, использующими NRPC.

Если на сервере Domino Document Manager, который планируется просматривать искателем, используется протокол DIIOP (Domino Internet Inter-ORB Protocol), запускать сценарий установки на сервере искателя не требуется. Однако нужно сконфигурировать сервер Domino Document Manager, чтобы к нему мог обращаться искатель Domino Document Manager.

Если WebSphere II OmniFind Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель Domino Document Manager при помощи консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 59.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 60.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 62.
- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 63.

Конфигурирование - обзор

Искатель Domino Document Manager можно использовать для просмотра любого числа библиотек Domino Document Manager. При создании искателя вы выбираете библиотеки для просмотра искателем с одного сервера Domino Document Manager. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера Domino Document Manager, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все шкафы в библиотеках, выбранных для просмотра или только отдельные шкафы.

Чтобы создать или изменить искатель Domino Document Manager, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать

искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Идентифицировать сервер и протокол связи Domino Document Manager.
- Выбрать библиотеки, которые будет просматривать искатель.
- Составить расписания просмотра искателем библиотек.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все шкафы в библиотеке или только документы, находящиеся в выбранных вами шкафах.
- Задать опции, обеспечивающие доступность поиска для полей в различных библиотеках и шкафах. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности данных регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры регистрации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Искатели Exchange Server

Чтобы включить общедоступные папки Microsoft Exchange Server в собрание поисковой системы предприятия, нужно сконфигурировать искатель Exchange Server.

Искатель Exchange Server можно использовать для просмотра любого числа папок и подпапок на серверах общедоступных папок Exchange Server. При создании искателя нужно выбрать содержимое, просматриваемое искателем на сервере общедоступных папок. Позже можно отредактировать пространство для искателя, добавив в него содержимое с другого сервера общедоступных папок.

Чтобы создать или изменить искатель Exchange Server, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подпапок на всех серверах в пространстве для искателя.
- Задать информацию о сервере общедоступных папок Exchange Server, просматриваемого искателем.

Для возможности обращения искателя к содержимому на сервере нужно задать ID пользователя и пароль. Если сервер использует протокол SSL, можно задать опции для доступа искателя к файлу склада ключей на сервере искателя.

- Составить расписание просмотра искателем сервера общедоступных папок.
- Выбрать папки и подпапки для просмотра.
- Задать опции, обеспечивающие доступность поиска для документов в подпапках. Например, можно исключить из пространства для искателя определенные типы документов.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Проверка доступа к защищенным документам Exchange Server

Чтобы использовать искатель Exchange Server для документов, защищенных брандмауэром, проверьте, что у сервера искателя есть доступ к серверу общедоступной папки Microsoft Exchange Server.

Об этой задаче

Если у сервера искателя нет доступа к защищенному серверу Exchange Server, вы получите от сервера код HTTP 501 (Not Implemented). Кроме того, вы можете получить сообщения о неожиданном ответе HTTP.

Порядок действий

Чтобы обеспечить доступ сервера искателя к документам за брандмауэром:

1. Запустите браузер на сервере искателя.
2. Перейдите на URL сервера общедоступной папки Exchange Server, который должен просматривать искатель. Например:
`http://exchange.yourCompany.com/public/`
3. Проверьте, что можете открыть страницу Exchange Server.

Если у вас нет доступа к серверу Exchange Server, обратитесь к администратору сервера вашей организации.

Искатели NNTP

Чтобы включить группы новостей NNTP в собрание поисковой системы предприятия, нужно сконфигурировать искатель NNTP.

Искатель NNTP можно использовать для просмотра любого числа серверов NNTP. При конфигурировании искателя нужно выбрать группы новостей на каждом просматриваемом сервере. Можно также задать шаблоны для групп новостей, которые нужно исключить. При помощи этого можно разрешить искателю просматривать большую часть групп новостей на сервере и запретить ему просматривать небольшое число групп новостей, которые вы не хотите включать в поиск.

Например, можно задать правила, включающие все группы новостей на некотором сервере NNTP, затем исключить из них группы новостей, имена которых содержат строку `private`.

Чтобы создать или изменить искатель NNTP, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех групп новостей в пространстве искателя.
- Задать шаблоны для включения групп новостей и шаблоны для исключения определенных групп новостей из пространства для искателя.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Искатели Notes

Чтобы включить базы данных IBM Lotus Notes в собрание поисковой системы предприятия, нужно сконфигурировать искатель Notes.

Конфигурация сервера искателя

Если на сервере Lotus Notes, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с WebSphere Information Integrator OmniFind Edition, позволяет искателю Notes взаимодействовать с серверами, использующими NRPC.

Если на сервере Lotus Notes, который планируется просматривать искателем, используется протокол DIIOP (Domino Internet Inter-ORB Protocol), запускать сценарий установки на сервере искателя не требуется. Однако нужно сконфигурировать сервер Lotus Notes, чтобы к нему мог обращаться искатель Notes.

Если WebSphere II OmniFind Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель Notes при помощи консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 59.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 60.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 62.

- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 63.

Защита на уровне документов

Если включена защита собрания и на сервере Lotus Notes, который планируется просматривать искателем, используется протокол NRPC (Notes Remote Procedure Call - удаленный вызов процедур Notes), на сервере искателя нужно сконфигурировать Lotus Domino Trusted Server. Trusted Server используется для обеспечения соблюдения управления доступом на уровне документов. Перед тем, как сделать собрания доступными пользователям для поиска, выполните следующие задачи:

- “Конфигурирование доверенных серверов Lotus Domino для проверки параметров регистрации пользователей” на стр. 197.
- Включение глобальной защиты в WebSphere Application Server и конфигурирование прикладных программ поиска для использования защиты. Этот шаг приводит к тому, что при попытке пользователей работать с программами поиска у них запрашиваются параметры регистрации. После этого серверы поиска могут использовать эти учетные данные для проверки полномочий каждого из пользователей на доступ к документам Lotus Notes.

Конфигурирование - обзор

Искатель Notes может просматривать любое число стандартных баз данных Lotus Notes (файлы .nsf). При создании искателя вы выбираете базы данных или каталоги для просмотра искателем с одного сервера Lotus Notes. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера Lotus Notes, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все базы данных или каталоги на сервере или только конкретные базы данных, представления и папки.

Чтобы создать или изменить искатель Notes, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Указать имя хоста, порт и протокол связи сервера Lotus Notes.
- Выбрать базы данных или каталоги, которые будет просматривать искатель.
- Составить расписания просмотра искателем баз данных или каталогов.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все документы в каталоге, все документы в базе данных или документы в выбранных представлениях и папках базы данных.
- Задать опции, обеспечивающие доступность поиска для полей в различных базах данных, представлениях и папках. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности данных регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры регистрации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Советы по просмотру баз данных Lotus Domino

Перед тем, как конфигурировать искатель Notes, просмотрите указания по обработке искателем баз данных Lotus Domino.

- Базы данных Notes, основанные на стандартных шаблонах (как база данных примера) - самый удачный тип баз данных для работы искателя.
- Искатель Notes использует следующие правила отображения полей:
 - Имена главных полей из стандартных шаблонов Domino зарегистрированы изначально.
 - Значения из полей Notes, указанные в таблице правил отображения, используются в результатах поиска в качестве сводок документов.
 - Значения из полей Notes, не указанные в таблице правил отображения, не используются в сводках документов.
 - Значения из полей Notes, отображенные на поле заголовка, используются в результатах поиска в качестве заголовка документа.
 - Поля в следующей таблице отображаются на имена полей поиска по умолчанию:

Таблица 1. Правила отображения полей по умолчанию

Имя поля базы данных Notes	Имя поля поиска
Title	Title
EventTitle	Title
Subject	Title
Body	Body
Mission	Body
From	Creator
Author	Creator
Keywords	Categories
Categories	Categories
TeamRoomName	Organization
TeamName	Organization
Department	Organization

- Искатель Notes может обрабатывать все типы полей за исключением полей, вычисляемых для вывода.
- Статичный текст и изображения, помещенные в документ Notes, искателем не обрабатываются.
- При конфигурировании искателя включите переключатель **Обрабатывать искателем Все**, чтобы обрабатывать искателем все поля и максимально увеличить

набор данных полей для обработки (для ограничения числа обрабатываемых искателем полей можно использовать поле **Обрабатывать искателем все поля, кроме**).

Чтобы сократить обработку искателем ненужных полей, выключите переключатель **Обрабатывать искателем** для всех полей за исключением тех, которые отображены на поля поиска.

Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino

При установке WebSphere II OmniFind Edition на компьютере с операционной средой IBM AIX, Linux или Solaris, если вы планируете просматривать серверы, которые используют протокол Notes remote procedure call (NRPC), надо запустить сценарий для конфигурирования сервера искателя. Этот сценарий разрешает искателям Notes, QuickPlace и Domino Document Manager обмениваться информацией с серверами базы данных.

Ограничения

Сервер Domino нельзя запустить одновременно и на одном компьютере с искателем Notes, QuickPlace или Domino Document Manager, сконфигурированным для использования протокола NRPC. При попытке запустить один из этих искателей во время работы Domino Server появится сообщение об ошибке, и искатель будет остановлен.

Об этой задаче

Искатели, работающие по протоколу NRPC, используют библиотеки Domino как клиент. Эти библиотеки устанавливаются при установке Lotus Domino Server Версии 6.0.2 или новее на сервере искателя. Чтобы искатели могли работать с библиотеками Domino, запустите сценарий настройки, который WebSphere II OmniFind Edition помещает на сервер искателя после установки библиотек Domino.

Порядок действий

Чтобы сконфигурировать сервер искателя для просмотра серверов Lotus Notes, Lotus QuickPlace и Domino Document Manager:

1. Создайте на сервере искателя пользователя notes и группу notes:
 - a. Зарегистрируйтесь в качестве пользователя root:
`su - root`
 - b. Добавьте пользователя:
`useradd notes`
 - c. Добавьте пароль для этого пользователя:
`passwd notes`
Вас попросят изменить этот пароль.
2. Установите Lotus Domino Server на сервере искателя:
 - a. Вставьте компакт-диск Domino Server Версии 6.0.2 или новее и смонтируйте его. (Если у вас нет этого компакт-диска, можно загрузить его образ.)
 - b. Перейдите в папку для вашей операционной системы.

AIX: `cd /mnt/cdrom/aix`
Linux: `cd /mnt/cdrom/linux`
Solaris: `cd /mnt/cdrom/solaris`

- c. Запустите программу установки:


```
./install
```
 - d. Ответьте на запросы и примите значения по умолчанию или задайте другие параметры установки (например, пути для каталога установки и каталога данных).

При необходимости смотрите информацию об установке Domino Server в документации по Domino.
3. На сервере искателя запустите сценарий настройки, помещенный WebSphere II OmniFind Edition на сервер искателя:
 - a. Зарегистрируйтесь как администратор поисковой системы предприятия (этот ID пользователя задается при установке WebSphere II OmniFind Edition).
 - b. Запустите следующий сценарий, который установлен в каталоге \$ES_INSTALL_ROOT/bin:


```
escrnote.sh
```
 - c. Отвечайте на вопросы программы:
 - Ответьте Y (Да) на следующий вопрос, если Domino Server установлен в каталоге по умолчанию, и N (Нет), если это не так:

Найден путь каталога Lotus Notes /opt/lotus/notes/latest/linux.
Это правильный путь каталога Lotus Notes?

Для AIX путь по умолчанию - /opt/lotus/notes/latest/ibmpow.
Для Linux путь по умолчанию - /opt/lotus/notes/latest/linux.
Для Solaris путь по умолчанию - /opt/lotus/notes/latest/sunspa.
 - Если Domino Server установлен на сервере искателя не в каталоге по умолчанию, задайте путь установки Domino в ответ на следующий вопрос:

Введите путь каталога Lotus Notes

Например, на компьютере Linux это может быть
/opt/lotus/notes/latest/linux.
 - Ответьте на следующий вопрос Д (Да), если каталог данных Domino Server установлен в каталоге по умолчанию, и Н (Нет), если это не так:

Найден путь каталога данных Lotus Notes /local/notesdata.
Это правильный путь каталога данных Lotus Notes?

Путь по умолчанию - /local/notesdata.
 - Если каталог данных Domino Server не внедрен в положение по умолчанию на сервере искателя, задайте путь данных Domino в ответ на следующий вопрос:

Введите путь каталога данных Lotus Notes.
 4. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):


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

Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino

Если вы устанавливаете WebSphere II OmniFind Edition на компьютере с Microsoft Windows и планируете просмотр серверов, использующих протокол Notes Remote remote procedure (NRPC), надо запустить сценарий для конфигурирования сервера

искателя. Этот сценарий разрешает искателям Notes, QuickPlace и Domino Document Manager обмениваться информацией с серверами базы данных.

Ограничения

Сервер Lotus Domino и клиент Lotus Notes не могут работать одновременно и на одном компьютере с искателем Notes, QuickPlace или Domino Document Manager, сконфигурированным для использования протокола NRPC. При попытке запустить один из этих искателей во время работы Domino Server появится сообщение об ошибке, и искатель будет остановлен.

Об этой задаче

Искатели, работающие по протоколу NRPC, используют библиотеки клиента Lotus Notes. Для установки этих библиотек на сервере искателя нужно установить Lotus Notes Выпуск 6.0.2 или новее. Чтобы искатели могли работать с библиотеками клиента Lotus Notes, запустите сценарий настройки, который WebSphere II OmniFind Edition помещает на сервер искателя после установки библиотек клиента Lotus Notes.

Порядок действий

Чтобы сконфигурировать сервер искателя для просмотра серверов Lotus Notes, Lotus QuickPlace и Domino Document Manager:

1. Зарегистрируйтесь на сервере искателя с ID пользователя, входящего в группу администраторов. У этого ID пользователя должны быть полномочия на установку Lotus Notes.
2. Установите Lotus Notes:
 - a. Вставьте в дисковод компакт-диск Lotus Notes Выпуска 6.0.2 или новее. (Если у вас нет этого компакт-диска, можно загрузить его образ.)
 - b. Запустите программу установки: `setup.exe`
 - c. Ответьте на запросы и примите значения по умолчанию или задайте другие параметры установки (например, пути для каталога установки и каталога данных).

Если вам нужна помощь, обратитесь к документации по Lotus Notes.

3. На сервере искателя запустите сценарий настройки, помещенный WebSphere II OmniFind Edition на сервер искателя:
 - a. Зарегистрируйтесь с ID администратора поисковой системы предприятия (этот ID пользователя задается при установке WebSphere II OmniFind Edition).
 - b. Запустите следующий сценарий (он установлен в каталоге `%ES_INSTALL_ROOT%\bin`):
`escrnote.vbs`
 - c. Отвечайте на вопросы программы:
 - Для следующего вопроса ответьте Y, если Lotus Notes установлен в каталог по умолчанию, или N, если Lotus Notes установлен в другой каталог:
Найден каталог Lotus Notes `c:\lotus\notes`.
Это правильный путь каталога Lotus Lotus?

Типичный путь установки на компьютере Windows - `c:\lotus\notes` или `c:\lotus\domino`.

- Если Lotus Notes установлен на сервере искателя не в каталоге по умолчанию, в ответ на следующий запрос задайте путь установки Lotus Notes:

Введите путь каталога Lotus Notes

- Для следующего вопроса ответьте Д, если каталог данных Lotus Notes размещен по умолчанию, или Н, если это не так:
Найден каталог данных Lotus Notes c:\lotus\notes\data.
Это правильный путь каталога данных Lotus Notes?

Типичный путь каталога данных на компьютере Windows -
c:\lotus\notes\data или c:\lotus\domino\data.

- Если каталог данных Lotus Notes размещен на сервере искателя не в положении по умолчанию, в ответ на следующий запрос задайте путь каталога данных:
Введите путь каталога данных Lotus Notes.
4. Остановите и перезапустите поисковую систему предприятия, включая все ее сеансы CCL (common communication layer):
 - a. Из командной строки остановите поисковую систему предприятия:
esadmin stop
 - b. Выберите **Пуск** → **Программы** → **Администрирование** → **Службы** и перезапустите службу IBM WebSphere Information Integrator OmniFind Edition.
 - c. Из командной строки запустите поисковую систему предприятия:
esadmin start

Конфигурирование серверов, использующих протокол DIIOP

Чтобы искатель просматривал серверы, использующие протокол Domino Internet Inter-ORB Protocol (DIIOP), надо сконфигурировать сервер так, чтобы искатели Notes, QuickPlace и Domino Document Manager могли использовать этот протокол.

Прежде, чем вы начнете

На сервере, который вы хотите просматривать, должны быть запущены задачи DIIOP и HTTP.

Порядок действий

Чтобы сконфигурировать серверы, использующие протокол DIIOP:

1. Сконфигурируйте документ server:
 - a. Откройте документ server на сервере Lotus Notes, Lotus QuickPlace или Domino Document Manager, который вы хотите просматривать. Этот документ хранится в каталоге Domino.
 - b. На странице Configuration (Конфигурация) раскройте раздел **server**.
 - c. На странице Security (Защита) в области **Programmability Restrictions** (Ограничения программируемости) задайте соответствующие ограничения защиты для вашей среды в следующих полях:
 - **Run restricted Lotus Script/Java agents** (Запустить агенты Lotus Script/Java с ограничениями)
 - **Run restricted Java/Javascript/COM** (Запустить Java/Javascript/COM с ограничениями)
 - **Run unrestricted Java/Javascript/COM** (Запустить Java/Javascript/COM)
 Например, можно разрешить неограниченный доступ агентов Lotus Script/Java, задав звездочку (*), или же задать имена пользователей, зарегистрированных в Domino Directory, на которых распространяются ограничения Java/Javascript/COM.

Важное замечание: Имена пользователей, заданные вами в этих полях, должны быть доступны для искателя, который вы сконфигурировали для просмотра этого сервера с протоколом DIIOP.

- d. Откройте страницу Internet Protocol (Протокол интернета), затем откройте страницу HTTP и задайте для опции **Allow HTTP clients to browse database** (Разрешить клиентам HTTP просматривать базу данных) значение **Yes** (Да).
2. Сконфигурируйте документ user:
 - a. Откройте документ user на сервере Lotus Notes, Lotus QuickPlace или Domino Document Manager, который вы хотите просматривать. Этот документ хранится в каталоге Domino.
 - b. На странице Basics (Основная информация) в поле **Internet password** (Пароль интернета) задайте пароль.

Если для конфигурирования опций просмотра для этого сервера используется консоль администратора поисковой системы предприятия, задайте этот ID пользователя и пароль на странице, где задается просматриваемый сервер. Искатель использует эти параметры аутентификации для доступа к серверу.
3. Перезапустите задачу DIIOP на сервере.

Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino

Перед использованием искателей Notes, QuickPlace или Domino Document Manager в системе IBM AIX необходимо установить модуль IOCP (I/O completion port, порт завершения ввода-вывода) и сконфигурировать его для использования искателем.

Об этой задаче

Без модуля IOCP при попытке создать искатель возникнет ошибка процессов поиска. Появится сообщение об ошибке:

```
FFQM0105E Получена ошибка от сервера -  
Сообщение: FFQG0024E Перехвачена непредвиденная исключительная ситуация: поиск
```

В файл \$ES_NODE_ROOT/logs/system_ГГГММДД.log записывается приведенное ниже сообщение, включающее ошибку ENOEXEC. (Часть текста сообщения разбита на несколько строк для лучшей читаемости.)

```
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 Перехвачена исключительная ситуация, подробности:  
'java.lang.UnsatisfiedLinkError:  
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:  
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a'  
and a stack trace of 'java.lang.UnsatisfiedLinkError:  
/opt/lotus/notes/65010/ibmpow/liblsxbe_r.a:  
load ENOEXEC on shared library(s) /opt/lotus/notes/latest/ibmpow/libnotes_r.a  
at java.lang.ClassLoader$NativeLibrary.load(Native Method)  
at java.lang.ClassLoader.loadLibrary0(ClassLoader.java:2120)  
at java.lang.ClassLoader.loadLibrary(ClassLoader.java:1998)  
at java.lang.Runtime.loadLibrary0(Runtime.java:824)  
at java.lang.System.loadLibrary(System.java:908)  
at lotus.domino.NotesThread.load(NotesThread.java:306)  
at lotus.domino.NotesThread.checkLoaded(NotesThread.java:327)  
at lotus.domino.NotesThread.sinitThread(NotesThread.java:181)  
at com.ibm.es.crawler.discovery.notes.NotesLibrary$NotesOperation.discover  
(Unknown Source)  
at com.ibm.es.crawler.discovery.api.DiscoveryAPI.discover(Unknown Source)  
at com.ibm.es.control.discovery.server.ComponentDiscoveryW.discover
```

```

(ComponentDiscoveryW.java:72)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:85)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:58)
at sun.reflect.DelegatingMethodAccessorImpl.invoke
(DelegatingMethodAccessorImpl.java:60)
at java.lang.reflect.Method.invoke(Method.java:391)
at com.ibm.es.ccl.sessionwrapper.CallThread.run(CallThread.java:77)

```

Порядок действий

Чтобы установить модуль ИОСР на сервер искателя и проверить правильность установки,

необходимо:

1. Установить модуль ИОСР (bos.iocp.rte) с компакт-диска продукта AIX на сервер искателя.

После установки модуля ИОСР, но до создания искателя Notes, QuickPlace или Domino Document Manager примените программное исправление для модуля.

Инструкции смотрите здесь:

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

2. Чтобы проверить правильность установки модуля ИОСР на сервере искателя, введите команду:

```
$ ls1pp -l bos.iocp.rte
```

Вывод команды ls1pp должен быть аналогичен следующему:

Набор файлов	Уровень	Состояние	Описание

Path: /usr/lib/objrepos			
bos.iocp.rte	5.2.0.10	ВЫПОЛНЕНО	API ИОСР
Путь: /etc/objrepos			
bos.iocp.rte	5.2.0.10	ВЫПОЛНЕНО	API ИОСР

3. Введите следующую команду и проверьте, что порт ИОСР находится в состоянии **Доступен**:

```
$ lsdev -Cc iocp
```

Вывод команды lsdev должен соответствовать следующему:

```
iocp0 Доступен ИОСР
```

4. Если состояние порта ИОСР - **Определен**, измените его на **Доступен**:

- a. Зарегистрируйтесь на сервере искателя как root и введите команду:

```
# smit iocp
```

- b. Выберите **Изменить / Показать характеристики ИОСР** и измените **СОСТОЯНИЕ, КОНФИГУРИРУЕМОЕ ПРИ ПЕРЕЗАПУСКЕ СИСТЕМЫ** с **Определен** на **Доступен**.

- c. Перезагрузите сервер искателя.

- d. Введите команду lsdev еще раз и убедитесь, что состояние порта ИОСР изменилось на **Доступен**.

Искатели QuickPlace

Чтобы включить площадки и комнаты Lotus QuickPlace в собрание поисковой системы предприятия, нужно сконфигурировать искатель QuickPlace.

Конфигурация сервера искателя

Если на сервере QuickPlace, который планируется просматривать искателем, используется протокол NRPC (Notes remote procedure call - удаленный вызов процедур Notes), на сервере искателя нужно запустить сценарий. Этот сценарий, поставляемый с WebSphere Information Integrator OmniFind Edition, позволяет искателю QuickPlace взаимодействовать с серверами, использующими NRPC.

Если на сервере QuickPlace, который планируется просматривать искателем, используется протокол DIIOP (Domino Internet Inter-ORB Protocol), запускать сценарий установки на сервере искателя не требуется. Однако нужно сконфигурировать сервер QuickPlace, чтобы к нему мог обращаться искатель QuickPlace.

Если сервер QuickPlace, на котором вы планируете запустить искатель, использует сервер Lightweight Directory Access Protocol (LDAP), сервер QuickPlace необходимо сконфигурировать для использования протокола DIIOP (сервер QuickPlace не может использовать протокол NRPC для обработки данных LDAP). Необходимо также сконфигурировать базу данных Directory Assistance и сконфигурировать сервер QuickPlace для использования сервера LDAP в качестве вторичного сервера Domino.

Если WebSphere II OmniFind Edition установлен в системе IBM AIX, необходимо проверить, что на сервере искателя установлен и доступен модуль I/O Completion Port.

Перед тем, как сконфигурировать искатель QuickPlace с консоли администратора поисковой системы предприятия, выполните в соответствии с вашей средой следующие задачи:

- “Конфигурирование сервера искателей в UNIX для просмотра источников Lotus Domino” на стр. 59.
- “Конфигурирование сервера искателей в Windows для просмотра источников Lotus Domino” на стр. 60.
- “Конфигурирование серверов, использующих протокол DIIOP” на стр. 62.
- “Конфигурирование сервера QuickPlace для использования защиты Local User” на стр. 66.
- “Конфигурирование Directory Assistance на сервере QuickPlace” на стр. 67.
- “Конфигурирование порта завершения ввода/вывода в AIX для просмотра источников Lotus Domino” на стр. 63.

Конфигурирование - обзор

Искатель QuickPlace можно использовать для просмотра любого числа площадок QuickPlace. При создании искателя вы выбираете площадки для просмотра искателем с одного сервера QuickPlace. Позднее, при редактировании пространства искателя, можно будет добавить документы с другого сервера QuickPlace, который вы хотите включить в это же пространство искателя. При создании или редактировании искателя можно задать, должен ли искатель просматривать все комнаты в площадках, выбранных для просмотра, или же только отдельные комнаты.

Чтобы создать или изменить искатель QuickPlace, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех документов в пространстве искателя.
- Идентифицировать сервер и протокол связи QuickPlace.
- Указать информацию о каталоге пользователя, связанном с сервером (эта информация нужна искателю для соблюдения управления доступом при выполнении пользователями поиска в собрании).
- Выбрать площадки, которые будет просматривать искатель.
- Составить расписания просмотра искателем площадок.
- Выбрать документы, которые будет просматривать искатель. Искатель может просматривать все комнаты на площадке или только документы, находящиеся в выбранных вами комнатах.
- Задать опции, обеспечивающие доступность поиска для полей в различных площадках и комнатах. Например, можно исключить определенные поля из пространства искателя и задать опции для поиска во вложениях.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.
 Можно также выбрать опцию для проверки правильности данных регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры регистрации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование сервера QuickPlace для использования защиты Local User

Если вы планируете конфигурировать искатель QuickPlace для использования опции Local User при реализации защиты, перед созданием искателя надо сконфигурировать Domino Directory на сервере Lotus QuickPlace.

Об этой задаче

При конфигурировании искателя QuickPlace вы выбираете режим защиты, который искатель использует на уровне документов. Если выбрать режим Local User, надо убедиться, что все ID локальных пользователей и локальные группы зарегистрированы в Domino Directory (иерархия Domino Directory должна соответствовать иерархии QuickPlace).

Кроме того, надо убедиться, что ID пользователя и пароль, под которыми будет работать искатель, зарегистрированы в Domino Directory и наделены правами чтения просматриваемой базы данных.

Для использования QuickPlace требуется только имя пользователя. Но для просмотра источников QuickPlace требуется полностью развернутый ID пользователя. Формат развернутого ID пользователя:

имя_пользователя/имя_площадки/QR/имя_домена

Определите при помощи этой процедуры полностью развернутую версию ID пользователя, убедитесь, что у этого ID пользователя есть права чтения базы данных QuickPlace, и добавьте ID пользователя в Domino Directory. В Domino Directory должен быть ID пользователя, под которым искатель будет просматривать базы данных QuickPlace, и все локальные пользователи и группы QuickPlace (иерархия Domino Directory должна соответствовать иерархии QuickPlace).

Порядок действий

Чтобы сконфигурировать сервер QuickPlace для использования защиты Local User:

1. Подтвердите разрешения ID пользователя:
 - a. Откройте документ Server на сервере QuickPlace.
 - b. Откройте страницу Файлы, затем откройте список управления доступом (список ACL) для базы данных, которую будет просматривать искатель.
 - c. Убедитесь, что ID локального пользователя, под которым будет работать искатель, есть в ACL, и что у этого ID пользователя есть разрешение на чтение базы данных.
Полностью развернутую форму этого ID пользователя вы должны были задать на шаге 2.
2. Добавьте пользователя в Domino Directory:
 - a. Откройте документ Server на сервере QuickPlace.
 - b. На странице People and Groups в объекте дерева людей добавьте полностью развернутый ID пользователя, который вы подтвердили на шаге 1.
 - c. В поле **Internet password** (Пароль Интернет) задайте пароль для этого ID пользователя.

Конфигурирование Directory Assistance на сервере QuickPlace

Если вы планируете конфигурировать искатель QuickPlace для использования каталога LDAP при реализации защиты, сначала надо создать базу данных Directory Assistance на сервере Lotus QuickPlace.

Ограничения

На сервере QuickPlace, который вы хотите просматривать, должны быть запущены задачи DPOP и HTTP.

Порядок действий

Чтобы сконфигурировать LDAP Directory Assistance на сервере QuickPlace:

1. 1. Создайте базу данных Directory Assistance:
 - a. Откройте документ Server на сервере QuickPlace.
 - b. Создайте базу данных при помощи шаблона **Directory Assistance(6)**. Этот шаблон находится на сервере.
 - c. Выберите **Добавить Directory Assistance**, чтобы создать документ в базе данных.
 - d. Откройте вкладку Основные и в поле **Тип домена** выберите **LDAP**.
 - e. Откройте вкладку Контексты именования и убедитесь, что переключатель **Доверенные для параметров аутентификации** включен.
 - f. Откройте вкладку LDAP и задайте информацию о сервере LDAP.
 - g. Сохраните и закройте документ Server.

2. Сконфигурируйте сервер QuickPlace для использования базы данных Directory Assistance:
 - a. Откройте документ Server на сервере QuickPlace.
 - b. Откройте вкладку Основные и в поле **Имя базы данных Directory Assistance** задайте имя базы данных, созданной на шаге 1.
 - c. Сохраните и закройте документ Server.
Теперь сервер QuickPlace может использовать сервер LDAP как вторичный каталог Domino.

Искатели файловой системы UNIX

Чтобы включить документы, хранимые в файловых системах UNIX, в собрание поисковой системы предприятия, нужно сконфигурировать искатель для файловых систем UNIX.

Искатель для файловых систем UNIX можно использовать для просмотра любого числа файловых систем UNIX. При конфигурировании искателя нужно выбрать локальные и удаленные каталоги и подкаталоги, которые будет просматривать искатель.

При установке сервера искателя на компьютере Windows этот сервер нельзя использовать для просмотра источников файловых систем UNIX (искатель для файловых систем UNIX не выводится в списке доступных типов искателей).

Чтобы создать или изменить искатель для файловых систем UNIX, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подкаталогов в пространстве искателя.
- Составить расписания просмотра искателем файловых систем.
- Выбрать подкаталоги и уровни подкаталогов, с которыми должен работать искатель.
- Задать опции, обеспечивающие доступность поиска для документов в подкаталогах. Например, можно исключить из пространства для искателя определенные типы документов.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Искатели Web

Чтобы включить страницы с сайтов в собрание поисковой системы предприятия, нужно сконфигурировать искатель Web.

При помощи искателя Web можно просматривать любое число серверов HTTP. Искатель обращается к сайту и читает с него данные. Затем по ссылкам в документах он может просмотреть дополнительные документы. Искатель Web может просматривать и извлекать ссылки с отдельных страниц или из *наборов фреймов* (страниц, созданных с использованием фреймов HTML).

Просматриваемые искателем данные могут быть одного из многих обычных форматов; эти данные искатель может получать из разнообразных источников во внутренней сети или Интернете. К обычным форматам относятся HTML, PDF, Microsoft Word, Lotus WordPro, Extensible Markup Language (XML) и так далее.

Чтобы создать или изменить искатель Web, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Сконфигурировать искатель может только администратор поисковой системы предприятия или администратор собрания, которому принадлежит этот искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех страниц Web в пространстве для искателя.
- Задать правила для разрешения и запрета обращения к определенным сайтам. Задав правила для искателя, можно протестировать их и проверить, что искатель может обращаться к сайтам, которые вы хотите включить в пространство искателя.
- Задайте опции, чтобы включить определенные типы файлов и исключить файлы с определенными расширениями файла.
- Задать правила обработки искателем Web произвольных страниц ошибок.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.
- Задать опции для просмотра искателем сайтов, защищенных паролем (просматриваемые искателем Web-серверы должны использовать для предложения ввода паролей базовую аутентификацию HTTP или формы HTML).
- Задать опции для просмотра искателем сайтов, обслуживаемых прокси-сервером.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурация пользовательского агента

Искатель Web подчиняется протоколу Robots Exclusion. Для просмотра сайта, на котором используется этот протокол, нужно, чтобы в файле robots.txt на этом сайте был разрешен доступ к этому сайту для имени пользовательского агента, заданного в конфигурации искателя Web.

При запуске поисковой системы предприятия искатель Web загружает имя пользовательского агента, заданное в его конфигурации. Прежде чем загружать страницу с сайта, который искатель еще не посещал (или не посещал некоторое время), искатель сначала пытается загрузить файл с именем robots.txt. Этот файл находится в корневом каталоге сайта.

Если файл robots.txt не существует, это значит, что данный сайт открыт для неограниченного просмотра. Если этот файл существует, он указывает, какие области сайта (каталоги) недоступны для искателей. В файле robots.txt задаются разрешения для искателей (искатели идентифицируются по их именам пользовательских агентов).

Протокол Robots Exclusion не является обязательным, но искатели Web поисковой системы предприятия подчиняются ему:

- Если в файле robots.txt есть запись для имени пользовательского агента, заданного в конфигурации искателя Web, этот искатель Web подчиняется ограничениям, наложенным на данный пользовательский агент.
- Если этого имени пользовательского агента нет в файле robots.txt, но в последней записи в файле задано User-agent: * (что означает "любой пользовательский агент") и задано ограничение Disallow: / (что означает "не разрешать никакой доступ, начиная с корневого каталога сайта"), искатель Web не сможет выполнять просмотр этого сайта.
- Если этого имени пользовательского агента нет в файле robots.txt, но в последней записи в файле задано User-agent: * и задано ограничение Allow: /, искателю Web разрешено выполнять просмотр этого сайта.

Администраторы сайтов часто задают такую последнюю запись, запрещающую доступ всем искателям, кроме тех, которым доступ разрешен явным образом. Если вы конфигурируете новый искатель Web и знаете, что на некоторых сайтах, для которых вы хотите выполнять просмотр, используется протокол Robots Exclusion, попросите администраторов этих сайтов добавить записи для вашего искателя в соответствующие файлы robots.txt.

В свойствах искателя Web и во всех файлах robots.txt на интересующих вас сайтах должно быть задано одно и то же имя пользовательского агента.

Если ни один из сайтов, для которых нужно выполнять просмотр, не использует протокол Robots Exclusion, значение, заданное для имени пользовательского агента, обычно несущественно. Однако некоторые серверы прикладных программ, JSP и сервлеты генерируют разные ответы для разных имен пользовательского агента. Например, это могут быть различные ответы для обработки несовместимостей браузеров. В таких случаях имя пользовательского агента, заданное для искателя Web, может быть существенным (независимо от протокола Robots Exclusion). Если нужно выполнять просмотр сайтов таких типов, обратитесь к администраторам этих сайтов, чтобы убедиться, что искателю Web разрешен доступ.

Поддержка JavaScript

Искатель Web для поисковой системы предприятия может находить некоторые ссылки (адреса URL), входящие в разделы JavaScript документов Web.

Искатель Web может находить как относительные, так и абсолютные ссылки. Если документ HTML содержит элемент BASE, искатель использует этот элемент для получения из относительных ссылок абсолютных. В противном случае искатель использует для этого URL самого документа.

Поддержка JavaScript ограничена извлечением ссылок. Искатель не выполняет синтаксический анализ JavaScript, не строит объектную модель документа (Document Object Model - DOM), не интерпретирует и не выполняет операторы JavaScript. Искатель ищет в содержимом документа (в том числе в разделах JavaScript, но не только в них) строки, которые похожи на адреса URL в операторах JavaScript. Это означает две вещи:

- Будут найдены некоторые URL, которые были бы проигнорированы более строгим синтаксическим анализатором HTML. Искатель отвергнет все строки, не являющиеся синтаксически правильными адресами URL, но некоторые из допустимых URL, возвращенных шагом просмотра, могут не представлять интереса для поиска.
- Содержимое документа, генерируемое в JavaScript (например, когда пользователь просматривает страницу и браузер выполняет JavaScript), будет недоступно искателю Web и, следовательно, не будет индексироваться.

Правила ограничения пространства для искателя Web

Чтобы пользователи могли обращаться только к сайтам, на которых вы хотите разрешить им поиск, нужно задать правила, ограничивающие пространство просмотра данных искателем.

Когда искатель Web просматривает Web-страницы, он находит ссылки на другие страницы и помещает эти ссылки в очередь для следующего просмотра. Просмотр и поиск могут повторяться столько раз, сколько позволяют ресурсы времени и памяти. При конфигурировании искателя Web нужно задать URL, с которых искатель должен начинать просмотр. С этих URL (они называются *начальными URL*) искатель Web может прийти до любого документа в Web, перейдя к нему по прямым ссылкам или через промежуточные страницы.

Чтобы ограничить пространство для искателя, сконфигурируйте искатель Web, чтобы он детально просматривал определенные URL и игнорировал ссылки на области, которые вас не интересуют. Поскольку по умолчанию искатель принимает любые обнаруживаемые им URL, необходимо задать правила, задающие URL, которые вы хотите включить в собрание, и исключающие остальные страницы.

Просматриваемую и не просматриваемую искателем Web область можно задать несколькими способами. Вы можете задать:

- Список начальных URL, с которых искатель должен начинать просмотр
- Три типа правил просмотра искателем: правила доменов, IP-адресов и префиксов URL
- Список типов MIME для включения документов
- Список расширений файлов для исключения документов
- Максимальное число каталогов в пути URL

Для правил просмотра используется следующий формат:

действие тип назначение

где действие - запрет или разрешение (forbid или allow), тип - домен (domain), IP-адрес (address) или префикс URL (HTTP или HTTPS), а назначение зависит от значения типа. Для ограничения путей можно использовать в качестве символа подстановки звездочку (*), чтобы задать назначения, соответствующие шаблону.

Правила доменов

Назначение правила доменов - это имя DNS. Например, можно задать, что искатель должен просмотреть весь домен `www.ibm.com`:

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

Первым символом назначения может быть звездочка, означающая, что правило применимо ко всем именам хостов, заканчивающимся символами, заданными в

остальной части шаблона. Например, можно задать, что искатель не должен просматривать домены, соответствующие следующему шаблону:
forbid domain *.ibm.com

И для явно заданного имени домена, и для шаблона имен доменов при определении соответствия учитывается регистр символов. Например, *.user.ibm.com соответствует joe.user.ibm.com и mary.smith.user.ibm.com, но не соответствует joe.user.IBM.com.

Правило доменов, в котором не задан номер порта, применяется ко всем портам в этом домене. Следующий пример разрешает все порты в домене sales:
allow domain sales.ibm.com

Если в правиле домена указан номер порта, правило применяется только к этому порту. Следующий пример разрешает только порт 443 в домене sales:
allow domain sales.ibm.com:443

Правила префиксов

Правило префиксов управляет просмотром искателем URL, начинающихся с заданной строки. В качестве назначения задается один URL, содержащий обычно одну или несколько звездочек для шаблона. Например, часто задают звездочку в качестве последнего символа в строке префикса.

Правило префиксов позволяет просмотреть искателем весь сайт или его часть. Можно задать путь каталога или шаблон, а затем разрешить или запретить просмотр всего из этой точки в дереве каталогов. Например, следующие правила, работая совместно, разрешают искателю просматривать все в каталоге public на сайте sales.ibm.com, но запрещают ему обращаться ко всем остальным страницам на этом сайте:

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

При задании правила префиксов можно задать сразу несколько звездочек, причем в любой позиции строки префикса, кроме последней. Например, следующее правило запрещает искателю просматривать любые документы в каталогах верхнего уровня на сайте sales.ibm.com, если имя каталога заканчивается на fs. (У вас, например, могут быть смонтированы файловые системы, не содержащие информации, полезной для индекса поиска информации.)
forbid http://sales.ibm.com/*fs/*

Правила адресов

При помощи правила адресов можно управлять просмотром целых хостов или сетей, задавая в качестве назначения задать IP-адрес (только IPv4) и маску сети. Например:
allow address 9.0.0.0 255.0.0.0

Маска сети позволяет задать соответствие шаблону. Чтобы правило адресов применялось к IP-адресу кандидата, IP-адрес в правиле и IP-адрес кандидата должны быть идентичны, за исключением тех мест, где в маске сети стоят нули. Правило адресов задает шаблон, а маска сети определяет значащие биты в шаблоне адреса. Ноль в маске сети действует как символ подстановки и означает, что совпадает любое значение, задаваемое в этой же битовой позиции в адресе.

В предыдущем примере применяется разрешающее правило к любому IP-адресу с 9 в первом октете и любым значением в последних трех октетах.

В качестве заключительного адреса в список правил полезно включить правило `forbid address 0.0.0.0 0.0.0.0`. Это правило соответствует всем адресам, поскольку маска сети делает все биты незначащими (это правило запрещает все адреса, не разрешенные предшествующим правилом в списке правил).

Ограничения для прокси-серверов: Если вы планируете работу искателя с сайтами, обслуживаемыми прокси-серверами, не задавайте правила IP-адресов. Обычно прокси-сервер используется, когда у пользовательского агента (браузера или искателя) нет прямого доступа к сетям, в которых находятся нужные Web-серверы. Например, прокси-сервер HTTP может передавать запросы HTTP от искателя на Web-сервер, а потом ответы - обратно искателю.

Когда Web-искатель использует прокси-сервер, IP-адрес этого прокси-сервера - единственный IP-адрес другого хоста, известный искателю. Если правила IP-адресов разрешают искателю доступ только к IP-адресам подсети, почти все адреса URL будут недоступны и для них будет возвращаться код возврата 760 (указывающий, что это запрещенный адрес в пространстве Web).

Порядок правил просмотра искателем

Искатель в процессе обнаружения и просмотра URL применяет правила просмотра в различное время. Порядок правил важен, но только в пределах каждого типа правил. Важен порядок использования одного правила адресов относительно другого правила адресов, но порядок правил адресов относительно правил префиксов не важен.

В пределах набора правил одного типа искатель проверяет домен, адрес или URL кандидата для каждого правила, с первого заданного правила до последнего, пока не найдет применимое правило. Используется действие, заданное для первого применимого правила.

Зависимость от порядка определяет обычную структуру большей части правил просмотра:

- Набор правил доменов обычно начинается с запрещающих правил, исключающих из пространства искателя отдельные домены. Например, администратор собрания может указать конкретные домены, не содержащие полезной информации.
- После списка запрещающих правил обычно следуют серии разрешающих правил (с символами подстановки), позволяющих искателю посещать любые домены, имена которых заканчиваются одним из имен доменов верхнего уровня, определяющих внутреннюю сеть предприятия (например, `*.ibm.com` или `*.lotus.com`).

Завершите набор правил доменов приведенным ниже правилом по умолчанию, исключающим домены, которые не были разрешены предшествующими правилами:

```
forbid domain *
```

Это последнее правило играет важную роль - оно не дает включить в пространство искателя всю сеть Интернет.

- Набор правил адресов обычно начинается с небольшого числа разрешающих правил, позволяющих искателю просматривать сети верхнего уровня (класса А, В или С) во внутренней сети предприятия.

Правила адресов обычно заканчиваются приведенным ниже правилом, предотвращающим просмотр искателем сайтов вне корпоративной сети.

```
forbid 0.0.0.0 0.0.0.0
```

- Набор правил префиксов обычно самый большой, поскольку содержит произвольные подробные спецификации разрешенных и запрещенных регионов, которые задаются как деревья и поддеревья. Рекомендуется сначала разрешить или запретить наиболее тесно локализованные регионы, а затем задать противоположное правило в более общем шаблоне, чтобы разрешить или запретить все остальное.

Как правило, раздел префиксов не заканчивается обычным правилом. Задание предлагаемых заключительных правил для доменов и адресов гарантирует, что искатель не будет выполнять просмотр вне сети предприятия; это проще, чем проверка префиксов URL.

Искатель сможет применять правила префиксов более эффективно, если сгруппировать эти правила по действиям (запрещающим или разрешающим). Например, вместо коротких, альтернативных друг другу последовательностей разрешающих и запрещающих правил задайте длинную последовательность правил, предусматривающих одно действие, а затем задайте длинную последовательность правил, предусматривающих другое действие. Чтобы получить нужную конфигурацию пространства искателя, разрешающие и запрещающие правила можно чередовать. Но группирование вместе разрешающих правил и группирование запрещающих правил может улучшить производительность искателя.

Расширения файлов, типы MIME и максимальная глубина просмотра искателем

Следующие опции обеспечивают дополнительные способы задания содержимого пространства для искателя. Можно исключать определенные типы документов на основе их расширений и включать определенные типы документов на основе их типов MIME. При задании типов MIME для просмотра искателем учтите, что тип MIME в документах Web часто задается неправильно.

Максимальная глубина просмотра искателем - это число дробных черт в URL от корня сайта. Эта опция позволяет предотвратить просмотр искателем рекурсивных структур файловых систем на неопределенную глубину. Глубина просмотра не соответствует уровням, по которым искатель переходит по ссылкам от одного документа к другому.

Начальные URL

Начальные URL - это URL, с которых искатель начинает просмотр; они включаются в просмотр при каждом запуске искателя. Если начальные URL уже обнаружены, они не будут просматриваться или повторно просматриваться искателем быстрее, чем другие сайты, разрешенные вами в правилах просмотра.

Начальный URL важен только при первом запуске искателя Web, когда пространство искателя пусто. Кроме того, начальный URL важен, если вы добавили в список начальных URL пространства искателя не обнаруженный ранее URL.

В качестве начальных URL нужно задавать полные адреса URL, а не просто имена доменов. Необходимо задать протокол и номер порта, если это не порт 80.

Допустимы следующие начальные URL:

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

Следующий начальный URL недопустим:

`www.ibm.com`

Начальные URL нужно включить в правила просмотра. Например, искатель не сможет начать просмотр с заданного начального URL, если правила просмотра не разрешают его просматривать.

Проверка соединений URL для искателя Web

Здав URL для искателя Web, можно протестировать конфигурацию правил искателя.

Результаты тестирования показывают, доступны ли искателю URL при использовании имени пользовательского агента, заданного в свойствах искателя. Кроме того, результаты тестирования показывают, допускают ли просмотр URL правила исключения (например, просмотр документа может быть запрещен из-за того, что расширение файла соответствует расширению, исключенному из пространства искателя).

Когда сайт просмотрен искателем хотя бы один раз, можно протестировать URL для получения дополнительной информации. Например, отчет тестирования может включать последний код возврата HTTP (который показывает успешность просмотра URL искателем), время последнего просмотра URL, время следующего просмотра согласно расписанию и информацию о том, использует ли пользовательский агент текущий файл robots.txt Web-сервера.

Задание интервала повторного просмотра для искателя Web

Для управления частотой обращения искателя Web к адресам URL нужно задать опции в свойствах искателя Web.

Большинство искателей других типов в поисковой системе предприятия запускаются согласно расписаниям, задаваемым администратором. В отличие от них, искатель Web после запуска обычно работает непрерывно. Для управления частотой повторных посещений URL, уже просмотренных искателем, нужно задать минимальный и максимальный интервалы повторного просмотра.

Если при создании искателя Web или редактировании его свойств используется консоль администратора поисковой системы предприятия, можно выбрать опцию, позволяющую сконфигурировать дополнительные свойства. На странице *Дополнительные свойства искателя Web* можно задать опции для минимального и максимального интервалов повторного просмотра. Искатель Web будет использовать заданные вами значения для вычисления интервала повторного просмотра данных.

При первом просмотре страницы искатель использует дату и время просмотра, а также среднее значение от заданных вами минимального и максимального интервалов повторного просмотра, чтобы установить дату ее повторного просмотра. До этой даты страница просматриваться не будет. Время повторного просмотра страницы после этой даты зависит от нагрузки искателя и соотношения между новыми и старыми URL в пространстве для искателя.

При каждом повторном просмотре данной страницы искатель проверяет, не изменилось ли ее содержимое. Если содержимое изменилось, следующий интервал повторного просмотра будет сокращен, но он никогда не станет короче заданного минимального интервала повторного просмотра. Если содержимое не изменилось, следующий интервал повторного просмотра будет увеличен, но он никогда не станет длиннее заданного максимального интервала повторного просмотра.

Опции посещения URL искателем Web

Вы можете указать искателю Web посетить определенные URL, как только он сможет это сделать.

Если необходимо выполнить обновление информации с конкретных сайтов пространства для искателя, можно открыть страницу Слежение этого искателя, выбрать опцию **Адреса URL для посещения или повторного посещения**, затем задать адреса или шаблоны URL для страниц, которые необходимо просмотреть или просмотреть повторно.

Например, если ваш отдел связи добавляет Web-страницу во внутреннюю сеть или пересматривает страницу для внесения важного изменения, вы можете задать URL новой или измененной страницы. Если искатель запущен, он помещает заданный URL в очередь для просмотра в следующий раз при проверке страниц, ожидающих посещения (обычно каждые десять минут). Если искатель не запущен, он ставит указанный URL в очередь для просмотра в следующий раз, когда будет запущен.

Убедитесь, что правила просмотра позволяют искателю посещать заданные вами URL. Искатель может посетить заданные вами URL раньше, чем при нормальной работе. Однако чтобы такое посещение произошло, должно существовать правило просмотра, разрешающее просмотр данного URL.

Новые просматриваемые искателем данные станут доступны для поиска, когда будет реорганизован или обновлен индекс.

Как искатель Web обрабатывает произвольные страницы ошибок

Можно сконфигурировать искатель Web для обработки настроенных страниц, созданных администраторами сайтов в случаях, когда они не хотят возвращать стандартный код ошибки в ответ на требования получения определенных страниц.

Если сервер HTTP не может вернуть страницу, запрошенную клиентом, он в обычном режиме возвращает ответ, содержащий заголовок и код возврата. Код возврата указывает, что возникла ошибка (например, ошибка 404 - файл не найден). Администраторы некоторых сайтов создают специальные страницы, в которых ошибка объясняется более подробно, и конфигурируют сервер HTTP, чтобы он возвращал именно эти страницы. Такие настроенные страницы называют *произвольными страницами ошибок*.

Произвольные страницы ошибок могут исказить результаты работы искателя Web. Например, вместо заголовка, в котором указана ошибка, искатель получает произвольную страницу ошибки и код возврата 200, означающий успешную загрузку правильной страницы HTML. Но загруженная произвольная страница ошибки не связана с запрошенным URL, и вместо разных страниц возвращается почти одинаковое содержимое страницы ошибки. Такие посторонние по содержанию и почти идентичные страницы искажают индекс и результаты поиска.

Для обработки таких ситуаций можно задать при конфигурировании искателя Web опции для обработки произвольных страниц ошибок. Искателю Web нужна следующая информация о каждом сайте, с которого возвращаются произвольные страницы ошибок:

- Шаблон URL для сайта, для которого используются произвольные страницы ошибок. Этот шаблон URL состоит из протокола (HTTP или HTTPS), имени хоста, номера порта (если он нестандартный) и имени пути. Можно использовать звездочку (*) в качестве символа подстановки, заменяющего один или несколько символов до следующего обычного символа в шаблоне. Задаваемый шаблон регистрозависим.
- Шаблон заголовка для текста, соответствующего тегу <TITLE> документа HTML. При задании этого шаблона можно использовать звездочку (*) в качестве символа подстановки. Задаваемый шаблон регистрозависим.
- Шаблон содержимого для текста, соответствующего содержимому документа HTML. Это содержимое включает не только содержимое тега <BODY> (если тег <BODY> присутствует). Содержимым считается все, что следует в файле после заголовка HTTP. При задании этого шаблона можно использовать звездочку (*) в качестве символа подстановки. Задаваемый шаблон регистрозависим.
- Целочисленное значение, представляющее код возврата для документов, соответствующих заданным шаблонам URL, заголовка и содержимого.

Пример

Следующая конфигурация указывает искателю Web, что он должен сравнить все правильные страницы HTML (код возврата 200), возвращенные с сайта `http://www.mysite.com/hr/*`, с заданными шаблонами заголовка и содержимого. Если тег <TITLE> страницы начинается с "Sorry, the page", а содержимое документа - любой текст (*), искатель обрабатывает такую страницу, как если бы он получил код возврата 404 (страница не найдена).

Таблица 2. Пример обработки страниц ошибок

Шаблон URL	Шаблон заголовка	Шаблон содержимого	Код возврата HTTP
<code>http://www.mysite.com/hr/*</code>	<code>Sorry, the page*</code>	<code>*</code>	404

Для одного сайта можно создать несколько записей для обработки разных кодов возврата. Для каждого кода возврата с данного сайта требуется отдельная запись в конфигурации искателя Web.

Использование символов подстановки

Шаблоны URL, заголовка и содержимого - это выражения с символами подстановки. Символ подстановки (звездочка) соответствует любым символам, идущим до следующего обычного символа. Например:

`*404` означает *любые символы*404

404: `*` означает 404: *любые символы*

`http://*.mysite.com/*` означает `http://любой хост.mysite.com/любой файл`

`*` означает *любые символы*

Влияние на производительность

Сконфигурировав опции для обработки страниц ошибок, вы увеличиваете время, затрачиваемое искателем на обработку, поскольку искателю придется проверить все

успешно полученные страницы. Дополнительное время обработки требуется на проверку соответствия шаблонам, чтобы выяснить, что нужно вернуть - страницу или заменяющий ее код возврата.

Поддержка просмотра защищенных сайтов

Чтобы искатель Web мог обращаться к защищенному содержимому (например, документам, для доступа к которым требуется пароль), можно задать параметры регистрации на консоли администратора поисковой системы предприятия.

Если для ограничения доступа к сайтам Web-сервер использует базовую аутентификацию HTTP или аутентификацию на основе форм HTML, можно задать в конфигурации искателя Web параметры регистрации, чтобы этот искатель Web мог работать со страницами сайтов, защищенным паролями. Можно также задать опции для конфигурирования опознавательных файлов (cookie) вручную.

Сайты, защищенные базовой аутентификацией HTTP

Если для ограничения доступа к сайтам Web-сервер использует базовую аутентификацию HTTP, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.

Чтобы узнать, разрешено ли пользователю (или прикладной программе) обращаться к страницам на некотором сайте, многие Web-серверы используют для идентификации пользователя схему аутентификации клиента, которая называется базовой аутентификацией HTTP. Обычно это интерактивный процесс:

- Когда пользовательский агент HTTP (например, браузер) запрашивает страницу, защищенную базовой аутентификацией HTTP, Web-сервер возвращает код возврата 401, указывающий, что у запрашивающего нет полномочий доступа к запрошенной странице.
- Кроме того, Web-сервер предлагает передать ему параметры регистрации, по которым сервер сможет определить, разрешен ли данному пользователю доступ к защищенному содержимому.
- Браузер выводит пользователю диалоговое окно, в котором предлагает ввести имя пользователя, пароль и другую информацию, необходимую для составления параметров регистрации пользователя.
- Браузер кодирует параметры регистрации и затем включает их в повторное требование защищенной страницы.
- Если эти параметры регистрации успешно прошли проверку, Web-сервер возвращает код возврата 200 и содержимое запрошенной страницы.
- В последующие требования страниц с этого же Web-сервера обычно включаются те же параметры регистрации; таким образом авторизованный пользователь может обращаться к дополнительному защищенному содержимому, не вводя свои параметры регистрации для каждого требования.

Когда пользователь идентифицирован, Web-сервер и пользовательский агент HTTP обычно обмениваются опознавательными сервер файлами (*cookies*), позволяющими сохранять информацию о регистрации пользователя между требованиями HTTP.

Поскольку искатель Web работает не в интерактивном режиме, параметры регистрации для просмотра защищенных паролями страниц нужно задать до начала операции просмотра. При создании искателя Web или редактировании пространства искателя задайте информацию о каждом защищенном сайте, который нужно просматривать.

Чтобы задать такую информацию, нужно работать в тесном взаимодействии с администраторами сайтов или Web-серверов, для защиты которых используется базовая аутентификация HTTP. Администраторы должны сообщить вам требования защиты для сайтов, на которых будет выполняться просмотр, включая всю информацию, используемую для аутентификации искателей Web и определения того, разрешен ли искателю доступ к защищенным страницам.

Если при создании собрания для него включена защита, можно при конфигурировании искателя задать маркеры защиты (такие как ID пользователей, ID групп или роли пользователей) для управления доступом к документам. Искатель Web связывает эти маркеры защиты с каждым документом, просматриваемым им в дереве файловой системы для заданного корневого URL. Эти маркеры используются в дополнение ко всем маркерам защиты уровня документа, сконфигурированным для всего пространства для искателя Web.

Имеет значение порядок, в котором заданы URL. Добавив информацию о защищенном паролем сайте, нужно задать его в соответствии с требуемым порядком обработки искателем. Сначала задайте более конкретные URL, а более общие поместите в конце списка. Когда искатель Web оценивает URL-кандидат, он использует данные аутентификации, которые заданы для первого URL в списке, соответствующего URL-кандидату.

Сайты, защищенные аутентификацией на основе форм

Если для ограничения доступа к сайтам Web-сервер использует формы HTML, можно задать параметры аутентификации, чтобы искатель Web мог обращаться к страницам, защищенным паролем.

Чтобы узнать, разрешено ли пользователю (или прикладной программе) обращаться к страницам на некотором сайте, многие Web-серверы используют для идентификации пользователя формы HTML. Обычно это интерактивный процесс:

- Когда пользовательский агент HTTP (например, браузер) запрашивает страницу, защищенную аутентификацией на основе форм, Web-сервер проверяет, содержит ли требование опознавательные файлы (cookie), позволяющие идентифицировать пользователя.
- Если опознавательных файлов нет, Web-сервер предлагает пользователю ввести данные защиты в форму. Когда пользователь передаст форму, Web-сервер возвращает необходимые опознавательные файлы и разрешает выполнить требование получения защищенной паролем страницы.
- Для последующих требований, в которых есть необходимые опознавательные файлы, также будет разрешено выполнение. Авторизованный пользователь может обращаться к дополнительному защищенному содержимому, не заполняя форму и не вводя свои параметры регистрации для каждого требования.

Поскольку искатель Web работает не в интерактивном режиме, параметры регистрации для просмотра защищенных паролями страниц нужно задать до начала операции просмотра. При создании искателя Web или редактировании пространства искателя задайте информацию о каждом защищенном сайте, который нужно просматривать.

Задаваемые поля соответствуют полям, которые пользователь заполняет в интерактивном режиме по запросу браузера; кроме того, нужно задать все скрытые поля и поля статистики, необходимые для успешной регистрации.

Чтобы задать такую информацию, нужно работать в тесном взаимодействии с администраторами сайтов или Web-серверов, для защиты которых используется

аутентификация на основе форм. Администраторы должны сообщить вам требования защиты для сайтов, на которых будет выполняться просмотр, включая всю информацию, используемую для аутентификации искателей Web и определения того, разрешен ли искателю доступ к защищенным страницам.

Имеет значение порядок, в котором заданы шаблоны URL. Добавив информацию о защищенном паролем сайте, нужно задать его в соответствии с требуемым порядком обработки искателем. В начало списка поместите наиболее конкретные шаблоны URL, а в конец списка - наиболее общие. Когда искатель Web оценивает URL-кандидат, он использует данные формы, которые заданы для первого шаблона URL в списке, соответствующего URL-кандидату.

Сайты с доступом через прокси-серверы

Если искателю Web не разрешен прямой доступ к сети, можно сконфигурировать искатель, чтобы он использовал для доступа к просматриваемому содержимому прокси-сервер HTTP.

Если у компьютера, где должен работать искатель Web, нет доступа к сети TCP/IP (или доступ разрешен только привилегированным процессам), можно сконфигурировать искатель Web, чтобы он использовал прокси-сервер HTTP. Прокси-сервер HTTP - это процесс, ожидающий требований HTTP на заданном порте заданного хоста. Прокси-сервер пересылает требования серверу Web и затем пересылает ответы сервера Web передавшему требованию клиенту (искателю Web). Прокси-сервер может работать на том же компьютере, что и искатель Web, или на другом компьютере.

Если искатель не использует прокси, требование для получения страницы с данным URL отправляется напрямую хосту. Если искатель работает через прокси, такое требование отправляется прокси-серверу.

При создании искателя Web или редактировании пространства для искателя задайте информацию о прокси-серверах, которые этот искатель Web должен использовать при просмотре страниц в домене прокси-сервера. Перед добавлением прокси-сервера в пространство для искателя узнайте следующую информацию:

Домены прокси-сервера

Домены, обслуживаемые этим прокси-сервером. В качестве символа подстановки можно использовать звездочку (*). Например, * соответствует всем доменам, обслуживаемым этим прокси-сервером, а *.resource.com соответствует всем доменам с окончанием resource.com.

Ограничение: Для просмотра искателем прокси-сервера нельзя задавать правила IP-адресов, поскольку IP-адрес прокси-сервера - это единственный IP-адрес, используемый искателем для других хостов. Если правила IP-адресов разрешают искателю доступ только к IP-адресам подсети, почти все адреса URL будут недоступны и для них будет возвращаться код возврата 760 (указывающий, что это запрещенный адрес в пространстве Web).

Имя хоста или IP-адрес прокси-сервера

DNS-имя хоста или IP-адрес (в формате с точками) этого прокси-сервера.

Номер порта прокси-сервера

Номер порта TCP/IP, на котором прокси-сервер ожидает приема требований HTTP прокси.

Добавив прокси-сервер, надо выбрать его и разместить в ряду других прокси-серверов в соответствии с требуемым порядком их обработки искателем. Сначала задайте наиболее конкретные имена доменов, а более общие имена поместите в конец списка. Когда искатель Web проверяет некоторый URL, он использует данные прокси-сервера, заданные для первого домена в списке, соответствующего этому URL. (Адреса URL, которым не соответствует ни одно из правил прокси, считаются напрямую доступными для искателя.)

Управление опознавательными файлами (cookie)

Обычно управление опознавательными файлами (cookie) выполняется автоматически и не требует от администратора поисковой системы предприятия никаких действий. При необходимости можно вручную задать файлы cookie для сеанса просмотра искателем Web.

Опознавательные файлы (cookie) - это внешние маркеры, возвращаемые Web-сервером пользовательскому агенту как часть заголовка ответа HTTP. Они значимы только для сгенерировавшего их Web-сервера и используются для поддержания состояния между требованиями HTTP. Например, при аутентификации клиента Web-сервер может вернуть опознавательный файл, позволяющий серверу определить, что аутентифицируемый пользователь уже зарегистрирован. Наличие опознавательных файлов позволяет пользователю запускать дополнительные требования для страниц на этом Web-сервере без повторной регистрации.

Искатель Web сохраняет опознавательные файлы, получаемые с Web-серверов, и использует их, пока существует экземпляр искателя. Искатель Web хранит опознавательные файлы в файле cookies.ini; искатель перезаписывает этот файл в конце каждого своего сеанса. При остановке искатель Web сохраняет все опознавательные файлы с истекшим сроком действия и перезагружает их при запуске следующего сеанса.

При задании опознавательных файлов вручную сохраните их в отдельном файле, а затем, когда потребуются, объедините с опознавательными файлами в файле cookies.ini. Искатель не отбрасывает опознавательные файлы с истекшим сроком действия, но на случай ошибки, при которой запись всех опознавательных файлов окажется невозможной, лучше застраховаться от потери файлов, заданных вручную. Эти файлы следует объединить с опознавательными файлами, поддерживаемыми искателем автоматически, перед запуском сеанса искателя.

Формат опознавательных файлов (cookie)

Опознавательные файлы, которые вы планируете объединить с файлом cookies.ini поисковой системы предприятия, должны иметь определенный формат.

- Каждая запись должна состоять из одной строки. Допускаются пустые строки и комментарии, но они не будут сохранены в файле cookies.ini.
- Каждая запись должна иметь следующий формат:

`CookieN(длина_cookie,длина_URL)текст_cookie,URL_проверки`

Cookie

Обязательное ключевое слово, указывающее начало записи.

Ключевое слово Cookie не может содержать пробелов, и на конце должна быть присоединена одна цифра - 0, 1 или 2. Эта цифра указывает тип cookie: version-0 (Netscape), version-1 (RFC2109) или version-2 (RFC2965). Списки портов не поддерживаются в файлах cookie RFC2965.

`длина_cookie`

Число символов в соответствующем тексте.

длина_URL

Число символов в соответствующем URL проверки.

текст_cookie

Содержимое, которое будет отправлено на исходный сервер Web. Эта строка (представляющая собой правую часть директивы Set-Cookie в заголовке ответа HTTP) задает пару имя - значение и любое другое содержимое (например, путь, параметры защиты и так далее), отправляемое в качестве cookie. После строки должен следовать разделитель - запятая (,).

URL_проверки

URL, по которому был найден этот опознавательный файл. Этот URL используется для определения, куда отправить запись cookie (например, имя домена и имя пути). URL проверки должен удовлетворять ограничениям защиты и конфиденциальности исходного сервера, установленным для опознавательных файлов.

Для удобства чтения следующий пример разбит на две строки; ваши записи надо задавать в одну строку:

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

Конфигурирование опознавательных файлов (cookie) для искателя Web

Можно вручную задать опознавательные файлы (cookie) для сеанса искателя Web и объединить их с файлами, задаваемыми искателем Web.

Прежде, чем вы начнете

Чтобы вручную сконфигурировать опознавательные файлы (cookie) для искателя Web, нужно быть администратором поисковой системы предприятия.

Порядок действий

Чтобы вручную сконфигурировать опознавательные файлы для искателя Web:

1. С консоли администратора поисковой системы предприятия запустите монитор нужного собрания и остановите искатель Web.
2. Зарегистрируйтесь на сервере искателя как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
3. Перейдите в каталог данных для нужного искателя, (*crawler_session_ID* здесь - ID, назначенный сеансу искателя поисковой системой предприятия). Например:
ES_NODE_ROOT/data/col_56092.WEB_88534
4. Откройте файл cookies.ini, введите свои записи cookie в дополнение к существующим, затем сохраните файл и закройте его. Убедитесь, что ваши опознавательные файлы не переопределяют существующие.
5. С консоли администратора поисковой системы предприятия снова запустите остановленный искатель Web.

Конфигурирование глобального пространства для искателя Web

Для искателей Web можно сконфигурировать глобальное пространство, позволяющее лучше управлять удалением URL из индекса.

Для каждого искателя Web задается пространство, определяющее адреса URL, которые искатель должен (или не должен) просматривать. Если обнаруженный URL входит в пространство для искателя, он сохраняется (в базе данных) для последующего просмотра; если URL не входит в пространство для искателя, он отбрасывается. Если искатель запущен с пустой базой данных, во время его работы определение пространства для искателя и база данных остаются согласованными.

В некоторых случаях искатель может быть остановлен, а его пространство для искателя - сокращено (например, могут быть добавлены новые правила, запрещающие просмотр определенных страниц). При перезапуске искателя его определение пространства для искателя и база данных становятся несогласованными. База данных содержит адреса URL (для некоторых просмотр выполнен, для других - нет), которые не входят в новое, более узкое пространство для искателя.

Если у собрания только один искатель Web, этот искатель может восстановить согласованность, изменив коды возврата HTTP для этих URL на 760 (что означает, что они должны быть исключены) и затребовав удаление этих вновь исключенных страниц из индекса.

Если вы делите пространство для искателя между двумя или несколькими искателями Web (например, чтобы просмотр определенных страниц выполнялся чаще, чем остальных), каждый искатель Web будет работать с собственными таблицами базы данных (исходно пустыми) и будет просматривать свою часть пространства для искателей Web. Затем пространство для исходного искателя сокращается - из него удаляются части пространства, просмотр которых будут выполнять другие искатели. Проблемы возникают, когда исходный искатель пытается восстановить согласованность, удалив из индекса страницы, которые более не входят в его пространство для искателя. Поскольку эти страницы просматриваются теперь другими искателями, они должны оставаться в индексе.

Можно сконфигурировать пространство более высокого уровня - глобальное пространство для искателя, задающее адреса URL, которые исходный искатель не должен просматривать, но удалять их из индекса нельзя. Адреса URL, которые не входят ни в одно из пространств для искателей, по-прежнему будут отмечаться процессом поиска как исключенные и будут удаляться из индекса при повторном просмотре.

Глобальное пространство для искателя определяется в файле конфигурации `global.rules`, который должен находиться в каталоге конфигурации искателя (присутствие файла `global.rules` включает функцию глобального пространства для искателя). Если этот файл существует, он читается во время инициализации искателя. Если этот файл не существует, искатель работает с одноуровневым пространством искателя и удаляет из индекса документы, которые нужно удалить для согласованности определения пространства для этого искателя и базы данных.

Если глобальное пространство искателя определено, оно не влияет на то, как искатель вносит адреса URL в базу данных, но искатель будет затребовать удаление URL из индекса только в том случае, когда этот URL не входит ни в одно из пространств для искателей Web.

Синтаксис файла `global.rules` такой же, как и у файла `crawl.rules`, но он содержит только правила для имен доменов. Это ограничение позволяет разделять пространство для искателей между искателями только на основе DNS-имен хостов, но не на основе IP-адресов или шаблонов префиксов HTTP. Глобальное пространство для искателя не влияет на адреса URL, которые исключены из локального

пространства для искателя правилами префиксов URL или IP-адресов (заданными в файле `crawl.rules`); такие URL будут по-прежнему исключаться.

Глобальное пространство для искателя используется только для того, чтобы предотвратить удаление из индекса адресов URL, которые исключены из пространства для данного искателя его локальным правилом доменов. Следующие правила применяются в указанной последовательности:

1. Если URL из базы данных искателя исключен в соответствии с локальным правилом префиксов или адресов, этому URL назначается код возврата 760 и он удаляется из индекса. Для этого URL не будет выполняться повторный просмотр.
2. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов и глобальное пространство для искателя не определено, этому URL назначается код возврата 760 и он удаляется из индекса. Для этого URL не будет выполняться повторный просмотр.
3. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов, но для него есть явное разрешающее правило в глобальном пространстве для искателя, этому URL назначается код возврата 761. Искатель не будет выполнять для этого URL повторный просмотр, но он не удаляется из индекса (подразумевается, что этот URL входит в локальное пространство для другого искателя).
4. Если URL из базы данных искателя исключен в соответствии с локальным правилом доменов, но для него нет явного разрешающего правила в глобальном пространстве для искателя, этому URL назначается код возврата 760 и он удаляется из индекса.

Поскольку искатель использует определение глобального пространства искателя только для того, чтобы определить, нужно ли удалять адреса URL, уже исключенные из локального пространства для этого искателя, то если в глобальном пространстве искателя нет правила для данного URL, для него по умолчанию запрещен просмотр.

Файл `global.rules` должен существовать в каталоге `master_config` каждого из искателей, работающих в данном глобальном пространстве. Нужно внимательно отредактировать все копии файла `global.rules` и файлы локальных пространств `crawl.rules`, чтобы обеспечить их взаимную согласованность.

Директивы `no-follow` и `no-index`

Можно улучшить качество поиска, указав для искателя Web директивы, управляющие следованием ссылкам на страницах и индексированием страниц.

На некоторых страницах Web есть директивы `no-follow` или `no-index`, указывающие роботам (например, искателю Web), что не надо переходить по ссылкам с этих страниц, не надо включать содержимое этих страниц в индекс или не надо делать ни того, ни другого.

Управление этими параметрами может улучшить качество поиска. Например, на некоторых страницах каталогов могут быть тысячи ссылок и никакой другой полезной информации; искатель должен просматривать эти страницы и переходить с них по ссылкам, но индексирование самих страниц каталога не имеет смысла.

В некоторых случаях нужно, чтобы искатель не переходил вниз по иерархической структуре страниц, но на требуемых конечных страницах есть ссылки и нет директив `no-follow`. Поскольку некоторые из этих страниц генерируются автоматически, у них нет владельцев, которые могли бы вставить нужные директивы.

Чтобы задать правила просмотра таких страниц, надо создать или отредактировать файл конфигурации `followindex.rules`. При задании правил в этом файле руководствуйтесь следующим:

- В конфигулируемых правилах указываются префиксы URL (нельзя задавать сайты Web по IP-адресу или имени хоста DNS).
- Префиксы URL могут содержать звездочки (*) в качестве символов подстановки, что позволяет разрешить или запретить несколько файлов с подобными URL.
- Порядок имеет значение (искатель применяет первое правило, соответствующее проверяемому URL).
- Правила, в явном виде разрешающие и запрещающие переход по ссылкам или индексирование, переопределяют правила, указанные в обрабатываемом документе.

Переопределение директив `no-follow` и `no-index` на Web-страницах

В файле конфигурации можно задать правила, управляющие переходами искателя Web по ссылкам на страницы и индексацией страниц, содержащих директивы `no-follow` или `no-index`.

Прежде, чем вы начнете

Чтобы задать директивы `no-follow` и `no-index` для искателя Web, нужно быть администратором поисковой системы предприятия. Заданные вами директивы переопределяют директивы на страницах, обрабатываемых искателем.

Порядок действий

Чтобы переопределить директивы `no-follow` и `no-index`:

1. С консоли администратора поисковой системы предприятия запустите монитор нужного собрания и остановите искатель Web.
2. Зарегистрируйтесь на сервере искателя как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
3. Перейдите в каталог конфигурации для нужного искателя, (*`crawler_session_ID`* здесь - ID, назначенный сеансу искателя поисковой системой предприятия).
Например:
`ES_NODE_ROOT/master_config/col_56092.WEB_88534`
4. Создайте или отредактируйте файл с именем `followindex.rules`.
5. Введите правила для искателя в приведенном ниже формате, где *префикс_URL* - начальные символы сайтов, для которых вы хотите разрешить (`allow`) или запретить (`forbid`) переход (`follow`) или индексацию (`index`):

```
forbid follow префикс_URL
allow follow префикс_URL
forbid index префикс_URL
allow index префикс_URL
```
6. Сохраните и закройте файл.
7. С консоли администратора поисковой системы предприятия снова запустите остановленный искатель Web.

Искатели WebSphere Portal

Чтобы включить страницы с сайта IBM WebSphere Portal в собрание поисковой системы предприятия, нужно сконфигурировать искатель WebSphere Portal.

Конфигурирование сервера WebSphere Portal

Перед созданием искателя WebSphere Portal необходимо внедрить прикладную программу предприятия ESPACServer ear в WebSphere Portal. Эта программа предприятия устанавливается на поисковых серверах при установке WebSphere Information Integrator OmniFind Edition. Чтобы внедрить программу предприятия, выполните следующие задачи:

- “Внедрение прикладной программы предприятия для искателя WebSphere Portal”.

Конфигурирование - обзор

Можно использовать искатель WebSphere Portal для просмотра одного сайта WebSphere Portal. При конфигурировании искателя нужно задать URL нужного сайта портала. Затем искатель загружает портлеты, доступные для заданного портала.

Чтобы создать или изменить искатель WebSphere Portal, зарегистрируйтесь на консоли администратора поисковой системы предприятия. Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех страниц на сайте.
- Задайте URL сайта портала, с которым должен работать искатель, и информацию, которая позволит искателю соединиться с этим сайтом. Поскольку такие URL могут быть большой длины и содержать символы, не входящие в кодовый набор ASCII, лучше скопировать URL с сервера WebSphere Portal и вставить на консоль администратора поисковой системы предприятия.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности данных регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры регистрации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Внедрение прикладной программы предприятия для искателя WebSphere Portal

Перед созданием искателя WebSphere Portal необходимо внедрить прикладную программу предприятия ESPACServer ear в WebSphere Portal.

Об этой задаче

Файл ESPACServer.ear устанавливается в каталоге ES_INSTALL_ROOT/bin на серверах поиска при установке WebSphere II OmniFind Edition. Пути установки по умолчанию:

Системы UNIX:

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

Системы Windows:

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

Порядок действий

Чтобы внедрить прикладную программу предприятия, которая позволяет искателю WebSphere Portal просматривать сайты WebSphere Portal:

1. Остановите экземпляр сервера WebSphere Portal.
2. Запустите экземпляр сервера WebSphere Application Server server1, если он еще не запущен.
3. На сервере WebSphere Portal запустите консоль администратора WebSphere Application Server. Если вам предложат зарегистрироваться, зарегистрируйтесь.

Чтобы открыть консоль администратора, можно:

- Выбрать программу в меню **Пуск** Windows.
 - В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
 - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
4. Выберите **Программы**, затем **Установить новую программу**.
 5. Нажмите кнопку **Просмотр** и выберите файл ESPACServer.ear в своей системе.
 6. Два раза нажмите кнопку **Далее**. Если появится предупреждение о файлах правил, нажмите кнопку **Продолжить**.
 7. Нажимайте кнопку **Далее**, пока не увидите страницу **Отобразить модули на серверы программ**, выберите значения, как описано ниже, и нажмите кнопку **Применить**:
 - a. В текстовом поле **Кластеры и серверы** выберите **WebSphere:cell=имя_ячейки, node=имя_узла, server=WebSphere Portal**.
 - b. Включите переключатель рядом с модулем **ESPACServer.ear**.
 8. Продолжайте нажимать кнопку **Далее**, пока не увидите страницу **Сводка**, затем нажмите кнопку **Готово**.
 9. Щелкните по ссылке **Сохранить в главной конфигурации**, затем нажмите кнопку **Сохранить**, чтобы сохранить изменения в конфигурации WebSphere Application Server.
 10. Перезапустите сервер WebSphere Portal.

Копирование URL для просмотра искателем из WebSphere Portal

Чтобы снизить вероятность ввода неправильного URL, скопируйте и вставьте в соответствующее поле URL нужного сайта WebSphere Portal при конфигурировании искателя WebSphere Portal.

Об этой задаче

При создании искателя WebSphere Portal вы задаете URL портала на сервере WebSphere Portal, с которым должен работать искатель. Поскольку эти URL - большой длины и обычно содержат символы, не входящие в кодовый набор ASCII, лучше, как описано ниже, скопировать URL с сервера WebSphere Portal и вставить на консоли администратора поисковой системы предприятия.

Порядок действий

Чтобы задать URL для просмотра искателем WebSphere Portal:

1. Выведя страницу Сервер WebSphere Portal для просмотра искателем на консоль администратора поисковой системы предприятия, убедитесь, что сервер WebSphere Portal запущен, и зарегистрируйтесь на WebSphere Portal как администратор.
2. Выберите **Управление** в верхнем правом углу.
3. Выберите **Параметры портала** в области навигации слева, затем выберите **Управление поисковой системой**.
4. На странице Управление собраниями поисковой системы в области Собрания поиска выберите **PortalCollection**. (Можете выбрать и другое собрание, если оно есть.)
5. В области Источники содержимого в собрании выберите **Добавить источник содержимого**.
6. В поле **Тип источника искателя** выберите **Сайт портала**. URL сайта появится в поле **Собирать документы по ссылкам из этого URL**.
7. Скопируйте URL в буфер обмена. (Например, выделите URL и, удерживая клавишу Ctrl, нажмите клавишу Insert.)
8. Вернитесь на консоль администратора поисковой системы предприятия и вставьте скопированный URL в поле **URL сайта WebSphere Portal**.

Искатели файловой системы Windows

Чтобы включить документы, хранимые в файловых системах Microsoft Windows, в собрание поисковой системы предприятия, нужно сконфигурировать искатель для файловой системы для файловых систем Windows.

Искатель для файловых систем Windows можно использовать для просмотра любого числа файловых систем Windows. При конфигурировании искателя нужно выбрать локальные и удаленные каталоги и подкаталоги, которые будет просматривать искатель.

При установке сервера искателя на компьютере UNIX этот сервер нельзя использовать для просмотра источников файловых систем Windows (искатель для файловых систем Windows не выводится в списке доступных типов искателей).

Чтобы создать или изменить искатель для файловых систем Windows, зарегистрируйтесь на консоли администратора поисковой системы предприятия.

Конфигурировать искатель может пользователь с ролью администратора поисковой системы предприятия или администратора собрания, которому принадлежит искатель.

При создании искателя мастер помогает выполнить следующие задачи:

- Задать свойства, управляющие работой искателя с системными ресурсами и их использованием искателем. Свойства искателя управляют способом просмотра искателем всех подкаталогов в пространстве искателя.
- Составить расписания просмотра искателем файловых систем.
- Выбрать подкаталоги для просмотра искателем.

Можно задать, сколько уровней подкаталогов должен просматривать искатель. Для просмотра удаленных файловых систем нужно также задать ID пользователя и пароль, позволяющие искателю обращаться к данным.

- Задать опции, обеспечивающие доступность поиска для документов в подкаталогах. Например, можно исключить из пространства для искателя конкретные типы документов или задать ID пользователя и пароль, чтобы искатель мог обращаться к файлам в определенном каталоге.
- Конфигурирование опций защиты на уровне документов. Если при создании собрания для него была включена защита, искатель может связать данные защиты с документами в индексе. Эти данные позволяют программам поиска применять управление доступом на основе хранимых списков управления доступом или маркеров защиты.

Можно также выбрать опцию для проверки правильности данных регистрации пользователя во время передачи пользователем запроса. В этом случае система сравнивает параметры регистрации пользователя не с данными защиты, внесенными в индекс, а с текущими списками управления доступом, поддерживаемыми исходным источником данных.

Чтобы использовалась защита уровня документов, надо убедиться, что на сервере искателя правильно сконфигурирована информация об учетной записи пользователя и домена.

При создании искателя можно нажать кнопку **Справка**, чтобы узнать о полях в мастере и способах задания информации, необходимой искателю для просмотра данных.

Конфигурирование поддержки для программ приема данных

Можно расширить поисковую систему предприятия, используя API приема данных для создания внешнего искателя. Пользовательские программы приема данных смогут добавлять данные в собрание или будут указывать искателю Web посещать определенные URL или повторно их просматривать.

Прежде, чем вы начнете

Чтобы конфигурировать программы приема данных, надо быть администратором поисковой системы предприятия.

Об этой задаче




Клиентская программа приема данных позволяет выполнять поиск в таких типах источников данных, с которыми не могут работать искатели поисковой системы предприятия по умолчанию. Для использования программы приема данных надо сконфигурировать параметры регистрации, обеспечивающие этой программе доступ к собраниям и их изменение.

Когда клиентская программа приема данных связывается с системой приема данных, она должна передать ID клиентской программы и пароль, а также ID изменяемого собрания. Эта информация должна совпадать с тем, что сконфигурировано для этой программы с консоли администратора.

Система приема данных запускается автоматически при запуске поисковой системы предприятия. Если вы изменили номер порта после того, как сконфигурировали программу с консоли администратора, необходимо перезапустить систему приема данных.

Порядок действий

Чтобы сконфигурировать программы приема данных:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Система приема данных нажмите кнопку **Конфигурировать программы приема данных**.
4. На странице Программы приема данных задайте число потоков, которые система приема данных сможет создавать для обработки требований от клиентских программ, и номер порта, на котором она принимает требования.
5. Выберите **Добавить программу системы приема данных**, чтобы добавить информацию о клиентской программе.
6. На странице Добавить программу приема данных укажите информацию об аутентификации, позволяющую клиентским программам приема данных получать доступ к собраниям поисковой системы предприятия. ID клиентов приема данных должны быть уникальными в поисковой системе предприятия.
7. Выберите собрания, которые сможет обновлять программа приема данных:
 - Чтобы разрешить программе обновлять все собрания, выберите **Все собрания**.
 - Чтобы разрешить программе обновлять только указанные вами собрания, выберите **Отдельные собрания**.Если выбрана эта опция, появляется список имен собраний. Включите переключатели **Выбрать** для всех собраний, которые сможет обновлять эта программа.
8. Нажмите кнопку **ОК**.
9. Если вы изменили номер порта системы приема данных, перезапустите ее:
 - a. Выберите  **Слежение**, чтобы перейти в окно слежения за системой.
 - b. На странице Система приема данных нажмите кнопку  **Перезапустить**.

Пользовательские подключаемые модули искателей

При конфигурировании свойств для искателей можно задать класс Java, обеспечивающий управление доступом на уровне отдельных документов и связь метаданных с документами в индексе поисковой системы предприятия.

Подключаемый модуль содержит класс Java, который вызывается для каждого документа, просматриваемого искателем. Классу Java передается идентификатор документа (URI) из индекса поисковой системы предприятия, маркеры защиты и метаданные. Этот класс возвращает новый или измененный набор маркеров защиты и метаданных или же указывает, что искатель должен игнорировать этот документ.

После первого просмотра искателем всех документов в пространстве искателя подключаемый модуль вызывается только для новых или измененных документов.

Чтобы изменить маркеры защиты и метаданные для документов, включенных в индекс поисковой системы предприятия, но не измененных в исходном источнике данных, запустите полный просмотр искателем всех документов в пространстве искателя, а затем реорганизируйте индекс.

Подключаемый модуль нельзя связать с существующим искателем. Имя и путь класса подключаемого модуля нужно указать при конфигурировании свойств для нового искателя.

Использование подключаемого модуля для обеспечения защиты

Защита на уровне документов реализуется при помощи связывания одного или нескольких элементов защиты (строка с разделителями-запятыми) с каждым документом, просматриваемым искателем. В качестве маркеров защиты обычно используются идентификаторы групп.

По умолчанию каждый документ связан с маркером общедоступности, делающим документ доступным каждому. Этот маркер можно заменить значением, заданным администратором, или значением, извлеченным из поля в просмотренном искателем документе.

Подключаемый модуль позволяет применить ваши собственные бизнес-правила для определения значения маркеров защиты для документов, просмотренных искателем. Маркеры защиты, связанные с документами, хранятся в индексе. При помощи маркеров защиты можно фильтровать соответствующие им документы и добиться возвращения в результатах поиска только тех документов, для просмотра которых у пользователя есть разрешение.

Использование подключаемого модуля для добавления метаданных в документы

Метаданные документов, такие как дата последнего изменения документа, создаются для всех документов, просмотренных искателем. Подключаемый модуль искателя позволяет применить ваши собственные бизнес-правила для определения значения метаданных, которые должны индексироваться для каждого документа.

Метаданные создаются как пара имя - значение. Пользователи могут выполнять поиск в метаданных при помощи свободного текстового запроса или запроса, в котором задается имя поля метаданных.

Подключаемый модуль искателя Web

Интерфейсы прикладного программирования для искателя Web позволяют управлять тем, как искатель просматривает документы и как документы готовятся для анализа. Например, можно добавить поля в заголовок требования HTTP, который будет использоваться, когда искатель затребует документ. После того, как документ просмотрен искателем, но перед его синтаксическим и лексическим анализом можно изменить содержимое, маркеры защиты и метаданные. Можно также остановить отправку документа анализатору.

Форматы URI для индекса поисковой системы предприятия

Универсальный идентификатор ресурса (URI) каждого документа в индексе поисковой системы предприятия определяет тип искателя, добавившего документ в собрание.

При конфигурировании категорий, областей действия и прямых ссылок для собрания можно задавать URI или шаблоны URI. Кроме того, URI задается при удалении документов из индекса или при просмотре подробной информации о состоянии определенного URI.

Чтобы определить URI или шаблоны URI для документа, выполните поиск по собранию. Для получения интересующего вас документа щелкните по его URI в результатах поиска. Можно скопировать URI найденного документа, чтобы использовать его на консоли администратора поисковой системы предприятия. Например, можно задать шаблон URI для автоматической связи документов, соответствующих этому шаблону URI, с определенной категорией поисковой системы предприятия.

Искатели Content Edition

Формат URI для документов, обработанных искателем Content Edition в режиме доступа к серверу:

```
vbr://Имя_сервера/Системный_ID_репозитория/Постоянный_ID_репозитория  
/ID_элемента/ID_версии  
/Тип_элемента/[?Page=Номер_страницы&] Свойства_JNDI
```

Формат URI для документов, обработанных искателем Content Edition в режиме прямого доступа:

```
vbr://Системный_ID_репозитория/Постоянный_ID_репозитория  
/ID_элемента/ID_версии  
/Тип_элемента/[?Page=Номер_страницы]
```

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера WebSphere Information Integrator Content Edition.

Системный_ID_репозитория

Системный ID для репозитория.

Постоянный_ID_репозитория

Постоянный ID для репозитория.

ID_элемента

ID для элемента.

ID_версии

ID версии. Если ID версии пуст, используется последняя версия документа.

Тип_элемента

Тип элемента (CONTENT или FOLDER).

Номер_страницы

Номер страницы.

Свойства_JNDI

Свойства JNDI для клиента прикладных программ J2EE. Свойства бывают двух видов:

java.naming.factory.initial

Имя класса сервера прикладных программ, используемого для создания хэндла EJB.

java.naming.provider.url

URL службы имен для сервера прикладных программ, используемого для получения хэндла EJB.

Примеры

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
```

Искатели DB2

Формат URI для документов, обработанных искателем DB2:

```
db2://Имя_базы_данных/Имя_таблицы  
/Имя_столбца_идентиф_уникальности1/Значение_идентиф_уникальности1  
[/Имя_столбца_идентиф_уникальности2/Значение_идентиф_уникальности2/...  
/Имя_столбца_идентиф_уникальностиN/Значение_идентиф_уникальностиN]
```

Параметры:

Ко всем полям применяется кодировка URL.

Имя_базы_данных

Внутреннее имя базы данных или ее алиас.

Имя_таблицы

Имя таблицы назначения, включая имя схемы.

Имя_столбца_идентиф_уникальности1

Имя первого столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальности1

Значение первого столбца идентификатора уникальности.

Имя_столбца_идентиф_уникальностиN

Имя *n*-ного столбца идентификатора уникальности в таблице.

Значение_идентиф_уникальностиN

Значение *n*-ного столбца идентификатора уникальности.

Примеры

Локальная, каталогизированная база данных:

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

Удаленная, некаталогизированная база данных:

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

Искатели DB2 Content Manager

Формат URI для документов, обработанных искателем DB2 Content Manager:

```
cm://Имя_сервера/Имя_типа_элементов/PID
```

Параметры

Имя_сервера

Имя библиотечного сервера IBM DB2 Content Manager.

Имя_типа_элементов

Имя типа элементов назначения.

PID Постоянный идентификатор DB2 Content Manager.

Пример

```
cm://cmsrvctg/ITEMTYPE1/92+3+ICM8+icmnl_sdb12+ITEMTYPE159+26+A1001001A  
03F27B94411D1831718+A03F27B+94411D183171+14+1018
```

Искатели Domino Document Manager

Формат URI для документов, обработанных искателем Domino Document Manager:

```
dominodoc://Имя_сервера:Номер_порта/ID_реплики_базы_данных/Путь_и_имя_базы_данных  
/Универсальный_ID_представления/Универсальный_ID_документа  
/?AttNo=Номер_вложения&AttName=Имя_файла_вложения
```

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Domino Document Manager.

Номер_порта

Необязательно: Номер порта для сервера Domino Document Manager.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла для базы данных NSF документов на сервере назначения Domino Document Manager.

Универсальный_ID_представления

Универсальный ID представления, используемый для просмотра искателем документов Domino Document Manager.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просмотренном искателем.

Номер_вложения

Необязательно: Порядковый номер, начиная с нуля, для каждого вложения.

Имя_файла_вложения

Необязательно: Первоначальное имя файла вложения.

Примеры

Документ Domino Document Manager:

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

Вложение документа:

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

Искатели Exchange Server

Формат URI для документов, обработанных искателем Exchange Server:

```
exchange://Путь_OWA[?useSSL=true]
```

Параметры

Путь_OWA

Путь Outlook Web Access (OWA) без указания протокола.

useSSL=true

Добавляется, если протокол первоначального пути OWA - HTTPS.

Примеры

Тело документа:

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

Вложение документа:

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

Включена возможность SSL:

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

Искатели Notes

Формат URI для документов, обработанных искателем Notes:

domino://Имя_сервера[:Номер_порта]/ID_реплики_базы_данных/Путь_и_имя_базы_данных
/[Универсальный_ID_представления]/Универсальный_ID_документа
[?AttNo=Номер_вложения&AttName=Имя_файла_вложения]

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Lotus Notes.

Номер_порта

Номер порта сервера Lotus Notes. Задавать номер порта необязательно.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла базы данных NSF на сервере назначения Lotus Notes.

Универсальный_ID_представления

Универсальный ID представления, определенный в удаленной базе данных. Этот ID задается только при выборе документа из представления или папки. Если не указано представление или папка для просмотра искателем (например, если задан просмотр искателем всех документов в базе данных), универсальный ID представления не задается.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просматриваемом искателем.

Номер_вложения

Порядковый номер, начиная от нуля, для каждого вложения. Указывать номер вложения необязательно.

Имя_файла_вложения

Первоначальное имя файла вложения. Указывать имя файла вложения необязательно.

Примеры

Документ, выбранный для просмотра по представлению или папке:

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

Документ, не выбранный для просмотра по представлению или папке:

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

Вложение документа:

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

Искатели QuickPlace

Формат URI для документов, обработанных искателем QuickPlace:

```
quickplace://Имя_сервера:Номер_порта/ID_реплики_базы_данных/Путь_и_имя_базы_данных  
/Универсальный_ID_представления/Универсальный_ID_документа  
/?AttNo=Номер_вложения&AttName=Имя_файла_вложения
```

Параметры

Ко всем полям применяется кодировка URL.

Имя_сервера

Имя сервера Lotus QuickPlace.

Номер_порта

Необязательно: Номер порта для сервера QuickPlace.

ID_реплики_базы_данных

Идентификатор реплики баз данных.

Путь_и_имя_базы_данных

Путь и имя файла для базы данных NSF документов на сервере назначения QuickPlace.

Универсальный_ID_представления

Универсальный ID представления, используемый для просмотра искателем документов QuickPlace.

Универсальный_ID_документа

Универсальный ID документа, определенный в документе, просмотренном искателем.

Номер_вложения

Необязательно: Порядковый номер, начиная с нуля, для каждого вложения.

Имя_файла_вложения

Необязательно: Первоначальное имя файла вложения.

Примеры

Документ:

```
quickplace://1twsvr.ibm.com/49257043000214B3/QuickPlace%5Csamplereplace  
%5CPageLibrary4925704300021490.nsf  
/A7986FD2A9CD47090525670800167225  
/2B02B1DE3A82B2CE49257043001C2498
```

Вложение страницы:

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

Искатели для файловых систем UNIX

Формат URI для документов, обработанных искателем для файловых систем UNIX:

`file:///Имя_каталога/Имя_файла`

Параметры

Ко всем полям применяется кодировка URL.

Имя_каталога

Абсолютный путь каталога.

Имя_файла

Имя файла.

Пример

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

Искатели WebSphere Portal

Формат URI для документов, обработанных искателем WebSphere Portal:

`wps://Путь_доступа_к_WPS?portletDefID=ID_опред_портлета&pageID=Page_ID&useSSL=SSL`

Параметры

Ко всем полям применяется кодировка URL.

Путь_доступа_к_WPS

Путь сервера WebSphere Portal без указания протокола.

ID_опред_портлета

Идентификатор определения портлета для сервера WebSphere Portal.

ID_страницы

Идентификатор страницы для сервера WebSphere Portal.

SSL useSSL=true добавляется, если протокол пути - HTTPS. Иначе добавляется useSSL=false.

Примеры

Тело документа:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+
LKnPy1vM0Y_QjzKCN4g3cQbJgQio-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!
?portletDefID=3_0_3S&pageID=6_0_6J&useSSL=false
```

Примеры

Включена возможность SSL:

```
wps://wpserver.ibm.com:9081/wps/myportal!/ut/p/kcxml/04_Sj9SPykssy0x+
LKnPy1vM0Y_QjzKCN4g3cQbJgQio-pFQAW99X4_83FT9AP2C5IhyR0dFRQD8qHRj/delta
/base64xml/L01DU1kvd0NrQUpORUEvNFBVR0VoQSEvN18wXzZPLzZfMF80RA!!
?portletDefID=7_0_A4&pageID=6_0_6J&useSSL=true
```

Искатели для файловых систем Windows

Форматы URI для документов, обработанных искателем для файловых систем Windows:

`file:///Имя_каталога/Имя_файла`

`file:///Имя_сетевой_папки/Имя_каталога/Имя_файла`

Параметры

Ко всем полям применяется кодировка URL.

Имя_каталога

Абсолютный путь каталога.

Имя_файла

Имя файла.

Имя_сетевой_папки

Только для документов на удаленных серверах: имя общей папки в сети Windows.

Примеры

Локальная файловая система:

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

Сетевая файловая система:

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

Управление анализаторами поисковой системы предприятия

Чтобы усовершенствовать извлечение документов, можно задать опции синтаксического анализа документов и метаданных, лингвистического анализа и категоризации перед добавлением в индекс поисковой системы предприятия.

Доступные опции анализа содержимого документов, необходимые для оптимального извлечения информации:

Задание опций анализа документов на китайском, японском и корейском

Можно задать опции сегментации энграмм для анализа документов на китайском, японском и корейском языках. Можно также удалять символы новой строки из пустого пространства в документах на китайском и японском.

Включение собственного поиска XML

Если собрание содержит документы XML, можно разрешить поиск по этим документам с использованием собственного синтаксиса запроса XML, например, поиск фрагментов XPath и XML. Собственный поиск XML позволяет пользователям задавать запросы на основе взаимосвязей между различными элементами XML.

Конфигурирование категорий

Документы, использующие один и тот же шаблон URI или содержащие определенные слова, можно сгруппировать в категории. При выполнении пользователями поиска в собрании они могут ограничить результаты поиска документами, принадлежащими конкретным категориям.

Конфигурирование полей поиска

Можно отобразить элементы из документов XML на поля поиска в индексе. Можно также отобразить на поля поиска элементы метаданных из документов HTML. Создавая поля поиска в индексе поисковой системы предприятия, вы предоставляете пользователям возможность запроса определенных частей документов XML и HTML и повышаете точность результатов поиска.

Конфигурирование опций обработки текста

Если в поисковую систему предприятия добавлены пользовательские механизмы тестового анализа, можно выбрать механизм анализа для собрания. Связав с собранием механизм анализа, можно задать опции отображения содержимого для его лингвистического анализа и аннотирования. Кроме того, можно задать, как результаты анализа будут отображаться в индексе поисковой системы предприятия или в таблицах базы данных JDBC.

Отображение полей на классы значимости

Можно задать, чтобы в результатах поиска ранг документов, где соответствие терминам запроса найдено в определенных полях, был выше ранга других соответствующих терминам запроса документов. При отображении полей на классы значимости можно задать, какие поля содержимого и метаданных будут значимыми. Можно также сконфигурировать оценки, используемые каждым классом значимости для ранжирования документов.

Работа с категориями

При помощи категорий можно сгруппировать документы с общими характеристиками и находить и получать только документы, соответствующие критерию, заданному для членов данной группы.

Если вы связали документы с категориями и ваши прикладные программы поиска поддерживают такую возможность, пользователи могут выполнять поиск в части собрания, указав имя категории. При выполнении поиска по всему собранию пользователи могут уточнить результаты поиска и просмотреть только документы, принадлежащие той же категории, что и один из документов результата.

Поисковая система предприятия поддерживает два варианта категоризации документов:

На основе правил

Применяйте этот вариант, если нужно связать документы с категориями в соответствии с задаваемыми правилами. Например, можно сгруппировать документы по использованию одинаковых шаблонов URI или по их содержанию (то есть документы, содержащие или не содержащие определенные слова или словосочетания).

На основе моделей

Применяйте этот вариант, если вы используете IBM WebSphere и хотите воспользоваться для поиска в собраниях поисковой системы предприятия предопределенными категориями WebSphere Portal.

Применение изменений в категоризации

Для создания и управления обработкой категорий используется консоль администратора поисковой системы предприятия:

- Тип категоризации выбирается при создании собрания. Возможные варианты: не использовать категории, использовать категории на основе правил или использовать категории на основе моделей.
- Если нужно, тип категоризации можно изменить при конфигурировании для собрания правил синтаксического анализа. Если тип категоризации изменен после просмотра документов искателем и их индексации, качество поиска понизится, пока вы снова не просмотрите все документы искателем, выполните их повторный анализ и реорганизуете индекс.
- Если выбрать использование категорий на основе правил, при помощи консоли администратора можно управлять деревом категорий, категориями и правилами категорий. Если категории или правила категоризации изменены после просмотра документов искателем и их индексации, качество поиска понизится, пока вы снова не просмотрите все документы искателем, выполните их повторный анализ и реорганизуете индекс.

Поля содержимого документа

При создании собрания с категориями на основе модели или с категориями на основе правил, использующими правила для содержимого документов (вместо правил образов URI), убедитесь, что в документах, для которых вводятся категории, имеются поля содержимого.

Категории на основе модели и правила категорий, основанные на содержимом документов, используют поля содержимого документов и игнорируют метаданные документов. Только документы, в которых есть поля содержимого или такие поля,

которые могут быть сконфигурированы как поля содержимого при конфигурировании искателя, могут быть категоризованы при помощи этих опций категоризации.

Если вы конфигурируете категории, основанные на моделях, или создаете правила категорий на основе содержимого документа, эти правила применимы только к частям документа с содержимым. Для следующих типов искателей эти правила применяются к файлам (например, к содержимому файлов UNIX или Microsoft Windows) или к содержимому Web-страниц:

- Exchange Server
- NNTP
- для файловых систем UNIX
- Web
- WebSphere Portal
- для файловых систем Windows

Для следующих типов искателей при конфигурировании опций для отдельных источников данных можно указать, какие поля будут являться полями содержимого документа. При создании искателя эта опция доступна на последней странице мастера искателя. Если искатель уже существует, отредактируйте пространство работы искателя, выберите источник данных для редактирования опций и укажите, какие поля в этом источнике данных назначаются полями содержимого.

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

Категории на основе правил

Вы можете сконфигурировать правила для управления связями документов с категориями в собрании поисковой системы предприятия.

Можно создать категории и правила категорий для собраний, создаваемых в поисковой системе предприятия, а также для категорий, перенастраиваемых из собраний IBM WebSphere Portal. Чтобы сконфигурировать правила для категоризации документов, при создании собрания или при задании для него правил синтаксического анализа нужно указать, что вы хотите использовать категории на основе правил.

Синтаксический анализатор использует задаваемые вами правила для связи документов с одной или несколькими категориями:

- Если для документа выполняется как минимум одно правило в категории, синтаксический анализатор связывает данный документ с этой категорией.
- Если для документа выполняется как минимум одно правило в нескольких категориях, синтаксический анализатор связывает данный документ со всеми этими категориями.
- Если для документа не выполняется ни одно из правил для категории, синтаксический анализатор не связывает документ с категорией. Пользователи могут найти этот документ и получить его, выполнив поиск в собрании, но при помощи поиска в категории получить этот документ невозможно.

Если вы - администратор дерева категорий (или таксономии) для собрания, вы решаете, в какое место иерархии категорий добавить новую категорию. При помощи дерева категорий вы также выбираете категорию, которую нужно отредактировать, а затем добавляете правила для категоризации документов, удаляете правила или изменяете содержимое отдельных правил.

Конфигурируя правило для категоризации документов, вы выбираете, будет ли в поисковой системе предприятия при определении принадлежности документа данной категории использоваться идентификатор URI документа или содержимого в этом документе:

Шаблон URI

Правило URI применяется к URI документа. Вы задаете неполный URI (шаблон), и правило выполняется для тех документов, идентификаторы URI которых соответствуют заданному шаблону.

Например, если задать текст правила `/hr/`, для первого из следующих URI правило выполняется, а для второго - не выполняется:

```
file:///corporate/hr/medicalform.doc  
http://company.com/human_resources/medicalform.htm
```

Поскольку все URI воспринимаются как шаблоны, система игнорирует все знаки звездочки, указанные в качестве символов подстановки в начале или конце шаблона. Например, `*/hr/*` и `/hr/` соответствуют одному и тому же набору URI.

В правилах шаблонов URI регистр символов не учитывается. Если URI содержит пробелы, шаблон URI должен соответствовать правилам кодирования URI поисковой системы предприятия. В следующем примере показаны правильные и неправильные способы задания URI для пути в файловой системе Windows:

```
Неправильный URI: file:///c:/program files/  
Правильный URI: file:///c:/program+files/
```

Содержимое документа

Правила содержимого документа задаются в таком же формате, как запросы. Если документ соответствует запросу, правило для него выполняется. При конфигурировании правила нужно задать слова и фразы, которые должны содержать (или не содержать) документы, а также выбрать язык для применения правил поиска морфологических форм слов.

Например, следующее правило задает, что если документ содержит слово `hr` или фразу `human resources`, для него правило выполняется:

```
hr "human resources"
```

Другой пример: следующее правило задает, что если документ содержит слово `hr`, но не содержит слово `benefits`, для него правило выполняется:

```
+hr -benefits
```

Правила содержимого подвергаются такой же лингвистической нормализации, как и запросы SI-API (Search and Index API - API поиска и индексации). Однако синтаксис правил содержимого поддерживает лишь подмножество операций, доступных в синтаксисе запросов SI-API. В запросах разрешаются только следующие операции:

- + Знак плюс перед термином означает, что этот термин должен присутствовать в документе.

- Знак минус перед термином означает, что этот термин должен отсутствовать в документе.
- " Группа терминов, заключенная в кавычки, означает, что в документе должно присутствовать это словосочетание в точности.

Правила для документов применяются только к частям документа, являющимся содержимым. Для следующих типов искателей эти правила применяются к файлам (например, к содержимому файлов UNIX или Microsoft Windows) или к содержимому Web-страниц:

- Exchange Server
- NNTP
- для файловых систем UNIX
- Web
- WebSphere Portal
- для файловых систем Windows

Для следующих типов искателей при конфигурировании опций для отдельных источников данных можно указать, какие поля будут являться полями содержимого документа. При создании искателя эта опция доступна на последней странице мастера искателя. Если искатель уже существует, отредактируйте пространство работы искателя, выберите источник данных для редактирования опций и укажите, какие поля в этом источнике данных назначаются полями содержимого.

- Content Edition
- DB2 Content Manager
- DB2
- Domino Document Manager
- Notes
- QuickPlace

Категории на основе модели

Если в системе IBM WebSphere Portal используются категории на основе моделей, их можно использовать и с собраниями поисковой системы предприятия.

WebSphere Portal поддерживает заранее заданную таксономию, включающую свыше 2300 тем. Эти темы сгруппированы по основным областям категорий бизнеса, таким как Компьютеры, Финансы и Транспорт. Пользователи WebSphere Portal могут создать программы, автоматически определяющие соответствие документов этим предметным областям, и настроить категории для своих собственных деловых потребностей.

Если вы хотите использовать категории WebSphere Portal с поисковой системой предприятия, вам нужно:

- При помощи мастера по перенастройке импортировать файлы таксономии на основе моделей в поисковую систему предприятия.
- Задать использование категорий на основе моделей при создании собрания или при конфигурировании правил синтаксического анализа для собрания.
- Убедиться, что на сервере индексов поисковой системы предприятия установлен продукт WebSphere Portal.

- Использовать для управления категориями инструменты категоризации, входящие в WebSphere Portal. Управлять категориями на основе моделей с консоли администратора поисковой системы предприятия нельзя.

Деревья категорий

С помощью дерева категорий можно просмотреть все категории на основе правил в собрании. Дерево категорий позволяет создать категории, удалить категории и отредактировать правила, связывающие документы с категориями.

Дерево категорий (другое название - таксономия) организовано иерархически. Дерево начинается с корневой категории, от которой ответвляются все остальные категории. Категорий и подкатегорий может быть произвольное количество, что обеспечит пользователям широту выбора при просмотре и получении документов.

Например, если документ пропускается через правила для нескольких категорий, он связывается со всеми этими категориями. Когда пользователи выполняют поиск в категории или просматривают в результатах поиска документы, принадлежащие определенной категории, принадлежность документа нескольким категориям повышает вероятность нахождения его пользователями.

Управляя деревом категорий, можно управлять числом категорий, которым принадлежат документы, вкладывая новые категории в существующие. При создании категории вы задаете, должна ли она быть создана на корневом уровне или как подкатегория другой категории. При помощи дерева категорий можно также удалить категории из собрания или изменить правила связи документов с категориями. При редактировании категории ее можно переименовать, добавить или удалить правила категоризации или изменить содержимое отдельных правил.

При управлении деревом категорий руководствуйтесь следующим описанием поведения поиска и просмотра:

- Если пользователь выполняет поиск в категории высокого уровня, поиск документов, соответствующих критерию поиска, выполняется в этой категории и всех ее подкатегориях. Если пользователь выполняет поиск в категории, не содержащей дополнительных подкатегорий, поиск выполняется только в этой категории.
- Если пользователь просматривает результаты поиска и выбирает опцию для просмотра документов, принадлежащих конкретной категории, выводятся документы только этой категории. В результатах поиска выводятся также имена всех подкатегорий, что позволяет пользователю перейти с одной категории на другую и одновременно просмотреть поднаборы документов.

Выбор типа категоризации

При выборе типа категоризации вы задаете способ связи документов с категориями, используемый в собрании.

Прежде, чем вы начнете

Изменить тип категоризации может администратор поисковой системы предприятия или администратор изменяемого собрания.

Об этой задаче

Тип категоризации задается при создании собрания. В случае необходимости можно изменить способ категоризации документов для собрания. Можно использовать

категории на основе правил, конфигурируемые специально для собрания, категории на основе моделей, имеющихся в системе IBM WebSphere Portal, или не использовать никаких категорий.

Важное замечание: Если изменить тип категоризации после просмотра искателем данных и создания индекса для собрания, индекс станет несогласованным. Чтобы гарантировать точность результатов поиска, повторно запустите искатель для всех документов в собрании и после анализа документов реорганизуите индекс.

Порядок действий

Чтобы выбрать тип категоризации:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Выбрать дерево категоризации**.
2. На странице Выбрать тип категоризации выберите одну из следующих опций:
 - Нет** Выберите эту опцию, если не хотите категоризировать документы в этом собрании.

На основе правил

Выберите эту опцию, если вы хотите использовать таксономию с правилами категории, сконфигурированными специально для этого собрания.

- Если вы конфигурируете собрание, созданное вами для поиска на предприятии, выберите эту опцию, чтобы задать имена категорий и правила для классификации документов.
- Если вы конфигурируете собрание, перенастроенное из WebSphere Portal, выберите эту опцию, чтобы использовать импортированные категории на основе старых или изменить их.

На основе модели

Выберите эту опцию, если вы хотите связать документы с категориями на основе моделей, существующих в системе WebSphere Portal. Для использования этой опции на сервере индекса поисковой системы предприятия должен быть установлен WebSphere Portal. Необходимо также использовать для управления категориями инструменты категоризации, входящие в WebSphere Portal.

3. Нажмите кнопку **ОК**.

Конфигурирование категорий

Для собрания можно создать сколько угодно категорий, в каждой из которых может быть сколько угодно правил. Правила определяют, какие документы автоматически связываются с категорией.

Прежде, чем вы начнете

Конфигурировать категории может администратор поисковой системы предприятия или администратор собрания, к которому принадлежат эти категории.

В качестве типа категоризации должна быть выбрана опция использования категорий на основе правил.

Чтобы просмотреть примеры задания правил связывания документов с категориями, нажмите кнопку **Справка** во время создания или редактирования категории.

Об этой задаче

Если в прикладных программах поиска включена поддержка категорий, пользователи могут выполнять поиск в подмножестве собрания, задав имя категории. Кроме того, пользователи могут выбрать категорию в результатах поиска и просматривать только документы, относящиеся к выбранной категории.

Важное замечание: Если изменить категории или правила категорий после просмотра искателем данных и создания индекса для собрания, индекс станет несогласованным. Чтобы гарантировать точность результатов поиска, повторно запустите искатель для всех документов в собрании, повторите анализ документов и реорганизируйте индекс.

Порядок действий

Чтобы сконфигурировать категорию:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать дерево категорий**.
2. На странице Дерево категорий выберите место в дереве, куда хотите добавить категорию, затем выберите **Создать категорию**.

Если выбрать корень дерева, новая категория будет создана на уровне корня.

Если выбрать имя категории, новая категория будет вложена в дерево категорий под выбранной категорией.

Откроется мастер, в котором надо задать правила для связывания документов с новой категорией:

- a. На странице Создать категорию введите описательное имя для категории и нажмите кнопку **Далее**.
- b. На странице Создать правила категории нажмите кнопку **Добавить правило**.
- c. На странице Создать правило категории введите в поле **Имя правила** уникальное имя для этого правила. Имя должно быть уникальным для всех категорий в собрании.
- d. Задайте правило, которое нужно использовать для связывания документов с этой категорией, и нажмите кнопку **ОК**.

- Чтобы поисковая система предприятия при определении принадлежности документа к этой категории использовала URI этого документа, выберите **Шаблон URI** и задайте шаблон URI.

Если указанный вами текст существует в URI, документ будет связан с категорией.

Например: `file:///c:/program+files/finance`

- Чтобы поисковая система предприятия при определении принадлежности документа к этой категории анализировала слова в полях содержимого документов, выберите **Содержимое документа**, выберите язык документов и задайте слова, которые должны или не должны входить в содержимое документов. Формат выражений для правил такой же, как и формат запроса (при этом разрешаются только операции запроса +, - и " ").

Если в документе содержатся (или отсутствуют) указанные вами слова, документ включается в данную категорию.

Например: `+finance -accounting +"fiscal year"`

- e. Нажмите кнопку **Готово**.

Новая категория появится на странице Дерево категорий вместе с другими категориями, относящимися к этому собранию.

Работа с полями поиска XML

Чтобы пользователи могли выполнять поиск в определенных частях документов XML, отобразите элементы XML на поля поиска.

Для отображения элементов XML на поля поиска служит консоль администратора поисковой системы предприятия. Linux

Поля поиска XML

Поля поиска XML позволяют пользователям делать запросы по определенным частям документов XML.

Популярность документов в формате XML растет, так как они одновременно содержат полуструктурированный и неструктурированный текст. Структура XML является инкапсулированной и использует контекст, в явном виде заданный элементами XML, между которыми заключен текст. Например, имя автора может быть указано в следующем виде:

```
<author>John Smith</author>
```

В этом контексте текст John Smith определяет автора документа XML.

Поисковая система предприятия может связывать (отображать) элементы XML с именами полей поиска. При конфигурировании опций синтаксического анализа для собрания вы задаете отображение элементов XML на имена полей поиска. Задание отображения элементов XML на поля поиска позволяет пользователям выполнять поиск в этих элементов, указав в запросе имена отображаемых полей. Поисковые запросы для конкретных полей могут дать более точные результаты поиска, чем запросы в виде свободного текста, которые выполняют поиск по всему содержимому документов.

Например, если в собрании есть документы XML, и задано, что элементы title и author будут помечены в индексе как поля поиска, пользователи смогут делать запросы к этим отдельным элементам. При поиске author:Smith будут найдены документы XML, содержащие Smith в элементе author.

Отображение элементов XML на поля поиска

При отображении элемента XML в поле поиска вы определяете, в каких элементах XML могут выполнять поиск пользователи, задав в запросе имя поля.

Прежде, чем вы начнете

Отображать элементы XML на поля поиска может только администратор поисковой системы предприятия или администратор собрания, к которому относятся документы XML.

Об этой задаче

Когда вы создаете отображение поля XML или добавляете, изменяете или удаляете поля в существующем отображении поля XML, эти изменения вступают в силу после следующей реорганизации индекса. Новые и измененные отображения применяются к данным, проанализированным после перезапуска анализатора. Новые и измененные отображения не действуют на данные, которые уже проанализированы и проиндексированы.

В этой задаче проиллюстрировано, как отобразить записи о персонале и позволить пользователям выполнять запросы непосредственно по некоторым элементам, на примере следующего документа XML.

```
<?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>
```

Порядок действий

Чтобы отобразить элементы XML в этом примере на поля поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Отобразить элементы XML на поля**.
2. На странице Отображения полей XML выберите **Создать отображение XML**. Откроется страница Создать отображение поля XML.
3. В поле **Имя корневого элемента XML** введите имя корневого элемента: `personnel`.
Заданное здесь имя должно в точности совпадать с именем корневого элемента в документах XML, в которых будет выполняться поиск. При анализе и индексировании документов XML поисковая система предприятия выбирает используемое отображение в соответствии с именем корневого элемента.
4. В поле **Имя отображения XML** введите имя для этого набора правил для отображения полей XML.
После создания набора правил отображения XML это имя выводится на странице Отображения полей XML, и его надо выбирать для добавления, удаления или изменения правил отображения.
5. Отобразите элемент XML `jobrole` на поле поиска `jobrole`:
 - a. В поле **Имя поля** введите `jobrole`.
 - b. В поле **Имя элемента XML** введите `jobrole`.
 - c. Чтобы пользователи могли вызывать запросы по полю `jobrole` и просматривать в результатах поиска должности, включите переключатели **Поиск по полям** и **Результаты поиска**.
6. Отобразите на это же поле поиска элемент XML `jobroles`:
 - a. Нажмите кнопку **Добавить поле**, чтобы добавить в список правил отображения полей пустую строку.
 - b. В поле **Имя поля** введите `jobroles`.
 - c. В поле **Имя элемента XML** введите `jobrole`.

Совет: Имена элементов XML не обязаны совпадать с именами полей поиска, и на одно и то же поле поиска можно отобразить несколько элементов XML.

- d. Чтобы пользователи могли вызывать запросы по полю `jobrole` и просматривать в результатах поиска должности, включите переключатели **Поиск по полям** и **Результаты поиска**.

7. Отообразите элемент XML `section` с атрибутом `expertise` на поле поиска с именем `expertise`:
 - a. Нажмите кнопку **Добавить поле**, чтобы добавить в список правил отображения полей пустую строку.
 - b. В поле **Имя поля** введите `expertise`.
 - c. В поле **Имя поля** введите `section`.
 - d. В поле **Имя атрибута XML** введите `id`.
 - e. В поле **Значение атрибута XML** введите `expertise`.
 - f. Чтобы пользователи могли вызывать запросы по полю `expertise` и просматривать в результатах поиска значения компетентности, включите переключатели **Поиск по полям** и **Результаты поиска**.
8. Нажмите кнопку **ОК**.

Примеры:

Чтобы найти всех сотрудников организации, работающих над поисковыми продуктами, задайте следующий запрос:

```
jobrole:search
```

Чтобы найти всех сотрудников организации, компетентных в области лингвистики, задайте следующий запрос:

```
expertise:linguistics
```

Работа с полями поиска HTML

Чтобы пользователи могли выполнять поиск в особых разделах метаданных документов HTML, надо отобразить элементы метаданных HTML на поля поиска в индексе.

Для отображения элементов метаданных HTML на поля поиска используйте консоль администратора поисковой системы предприятия.

Поля поиска HTML

Поля поиска HTML позволяют пользователям выполнять запросы по атрибутам документов HTML.

Элементы метаданных в документах HTML похожи на атрибуты документов тем, что они предоставляют информацию о документе, о его форматировании и о том, какой доступ к нему разрешен из Web. Например:

```
<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" />
```

Поисковая система предприятия может связывать (отображать) имена элементов метаданных HTML с именами полей поиска. При конфигурировании опций анализа для собрания можно задать отображение элементов метаданных HTML на имена полей поиска. Отображение элементов метаданных HTML на поля поиска позволяет пользователям находить документы с этими элементами, указывая в запросах имена

полей поиска. Поисквые запросы для конкретных полей могут дать более точные результаты поиска, чем запросы в виде свободного текста, которые выполняют поиск по всему содержимому документов.

Например, если в собрании есть документы HTML и включено индексирование элементов метаданных `copyright` и `abstract` в качестве полей поиска, пользователи могут выполнять запросы по этим конкретным элементам. При поиске `copyright:IBM` будут найдены документы HTML, которые содержат строку `IBM` в метаданных `copyright`.

Отображение элементов метаданных HTML на поля поиска

При отображении элемента метаданных HTML в поле поиска вы определяете, в каких элементах метаданных HTML могут выполнять поиск пользователи, задав в запросе имя поля.

Прежде, чем вы начнете

Чтобы отображать элементы метаданных HTML на поля поиска, надо быть администратором поисковой системы предприятия или администратором того собрания, к которому принадлежит этот документ в формате HTML.

Об этой задаче

При создании отображения поля HTML, добавлении, изменении или удалении поля в существующем отображении поля HTML изменения вступают в силу только после остановки и перезапуска анализатора. Изменения будут применяться только к данным, анализ которых будет проводиться после перезапуска анализатора. Чтобы внести изменения для документов, проанализированных и проиндексированных ранее, эти документы надо еще раз обработать искателем и проанализировать, а затем реорганизовать индекс.

Порядок действий

Чтобы отобразить элементы метаданных HTML на поля поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Отобразить метаданные HTML на поля**.
2. На странице Отображения полей HTML выберите **Добавить поле**, чтобы добавить пустую строку в список правил отображения полей.
3. В поле **Имя поля** введите имя, которое вы хотите связать с отображаемым элементом метаданных HTML. Пользователи смогут указывать имя этого поля при запросе документов HTML в этом собрании.
4. В поле **Имя элемента метаданных HTML** введите имя того элемента метаданных HTML, который вы хотите отобразить.
5. Чтобы пользователи могли запрашивать это поле и просматривать его в результатах поиска, включите переключатели **Поиск по полям** и **Результаты поиска**.
6. Если тип данных для поля - `DECIMAL`, `DOUBLE`, `INTEGER`, `SHORT`, `TIME` или `TIMESTAMP` и вы хотите дать пользователям возможность использовать параметрические запросы при поиске этого поля, включите переключатель **Параметрический поиск**.
7. Нажмите кнопку **ОК**.

Примеры:

Теперь пользователи могут запрашивать имена отображенных полей, чтобы искать документы в формате HTML с заданными метаданными. Например, если вы отображали элемент метаданных HTML с именем `description` в поле поиска с именем `abstract`, для поиска документов о компьютерах Thinkpad пользователи могут ввести примерно такой запрос:

```
abstract:thinkpad
```

Пользовательская обработка текста

Можно улучшить качество и точность результатов поиска при помощи интеграции алгоритмов пользовательской обработки текста с собраниями поисковой системы предприятия.

WebSphere Information Integrator OmniFind Edition поддерживает архитектуру управления неструктурированной информацией IBM (Unstructured Information Management Architecture, UIMA) - структуру для создания, просмотра, составления и внедрения функций анализа текста. Разработчики программ создают и тестируют алгоритмы анализа для содержимого поиска, затем создают архив механизмов обработки (файл `.pear`), содержащий все ресурсы, необходимые для использования этого архива в поиске предприятия. Для возможности поиска в собраниях с использованием пользовательских алгоритмов анализа этот архив (содержащий механизм анализа текста) нужно добавить в поисковую систему предприятия.

Компонент логики анализа в механизме анализа текста называется *аннотатором*. Каждый аннотатор выполняет конкретные задачи лингвистического анализа. Механизм обработки текста может содержать любое число аннотаторов или состоять из нескольких механизмов анализа текста, каждый из которых содержит свои собственные пользовательские аннотаторы.

Информация, генерируемая аннотаторами, называется *результатами анализа*. Результаты анализа, соответствующие искомой вами информации, записываются в структуру данных, которая называется *общая структура анализа*.

Чтобы сконфигурировать опции обработки текста, выполните следующие задачи:

- Выберите механизм анализа текста, который нужно использовать для создания аннотаций к документам в собрании.
- Если собрание содержит документы XML со смысловой разметкой, и вы хотите использовать эту разметку в пользовательском анализе текста, можно связать с собранием файлы отображения XML и отобразите вывод отображения XML на общую структуру анализа.

Например, можно отобразить содержимое элементов `<addressee>` и `<customer>` на аннотации `Person` в общей структуре анализа. Эти аннотации станут доступными вашим пользовательским аннотаторам, которые смогут найти дополнительную информацию (например, пол сотрудника). Можно также отобразить аннотации `Person` на индекс поисковой системы предприятия, что позволит пользователям выполнять поиск сотрудников, не зная исходные элементы XML.

Если вы хотите разрешить пользователям задавать в запросах исходные элементы XML, отображения XML определять не нужно. Вместо этого можно сконфигурировать опции синтаксического анализа и включить собственный поиск XML для собрания.

- Отобразите структуры данных в общей структуре анализа на индекс поисковой системы предприятия, что позволит выполнять поиск аннотированных документов с применением семантического поиска.

Например, в зависимости от объектов и взаимосвязей, обнаруживаемых аннотаторами, пользователи могут выполнять поиск понятий, встречающихся в одном и том же предложении (таких как конкретное имя человека и название фирмы), или выполнять поиск ключевого слова и понятия (например, найти термин Alex и номер телефона).

- Отобразите структуры данных в общей структуре анализа на таблицы базы данных, совместимые с JDBC (Java Database Connectivity). Можно отобразить данные на таблицы IBM DB2 Universal Database (DB2 UDB) или Oracle. Этот тип отображения позволяет использовать результаты анализа в программах баз данных, например, в программе исследования данных. Кроме того, он позволяет при помощи запросов SQL выполнять поиск данных вне поисковой системы предприятия.

Добавление механизмов анализа текста в систему

Если вы создали пользовательский механизм анализа текста, чтобы использовать его для поисковой системы предприятия, необходимо сначала добавить этот механизм в систему. Собрания могут использовать этот механизм для анализа документов, составления аннотаций, а также для повышения точности результатов поиска.

Прежде, чем вы начнете

Чтобы добавлять механизмы анализа текста в систему, надо быть администратором поисковой системы предприятия.

Об этой задаче


Разработчики прикладных программ могут создавать архив механизма обработки (файл .rear), который следует при анализе текста шаблону UIMA. Этот архив содержит все ресурсы, необходимые для выполнения поиска в собраниях поисковой системы предприятия. Для возможности поиска в собраниях с использованием пользовательских алгоритмов анализа этот архив (содержащий механизм анализа текста) нужно добавить в поисковую систему предприятия.

После добавления механизма анализа текста в систему можно изменить его имя для вывода и выбрать опцию для просмотра источника XML. (Этот источник XML показывает, какая информация генерируется данным анализом.)

Если механизм анализа текста связан с собранием, его нельзя удалить из системы.

Порядок действий

Чтобы добавить пользовательский механизм анализа текста в поисковую систему предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Анализ выберите **Конфигурировать механизмы анализа текста**.
4. На странице Механизмы анализа текста выберите **Добавить механизм анализа текста**.
5. На странице Добавить механизм анализа текста введите описательное имя для нового механизма. Система использует это имя для идентификации механизма анализа текста на консоли администратора.

6. Задайте положение файла .pear. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
7. Нажмите кнопку **ОК**. Новый механизм анализа текста появится в списке на странице Механизмы анализа текста.

Связывание механизма текстового анализа с собранием

Если с поисковой системой предприятия связаны пользовательские механизмы анализа текста, можно выбрать один из них для использования с собранием. Это позволит пользователям делать семантические запросы при поиске в собрании, а также повысит качество и точность результатов поиска.

Прежде, чем вы начнете

Чтобы связать механизм анализа текста с собранием, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Об этой задаче

Если с этим собранием уже связан один из механизмов анализа текста, при связывании другого механизма анализа произойдет следующее:

- Если выбрать опцию **Без пользовательского анализа**, все отображения анализа текста, ранее определенные для этого собрания, будут сброшены. Собрание станет использовать системные значения по умолчанию.
- Если выбрать имя другого пользовательского механизма анализа текста, все отображения анализа текста, ранее определенные для собрания, будут сохранены. Например, если вы замените механизм_1 на механизм_2, механизм_2 унаследует файлы отображения XML, сконфигурированные для механизм_1.

Порядок действий

Чтобы связать механизм анализа текста с собранием:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. Выберите **Выбрать механизм анализа текста**. Если в систему поиска предприятия не были добавлены пользовательские механизмы анализа текста или если собрание использует алгоритмы анализа по умолчанию, именем этого механизма будет **По умолчанию**.
3. На странице Выберите механизм анализа текста для этого собрания выберите имя механизма, который будет использоваться в этом собрании. Если нет доступных механизмов анализа текста, или выбрана опция **Без пользовательского анализа**, при подготовке документов для индекса анализатор применяет правила анализа текста по умолчанию.
4. Нажмите кнопку **ОК**.

Отображение элементов XML на общую структуру анализа

Если собрание содержит документы XML со смысловой разметкой, и вы хотите при помощи этой разметки разрешить пользователям выполнять поиск в индексе или таблицах JDBC поисковой системы предприятия с использованием семантического поиска, нужные элементы XML можно отобразить на общую структуру анализа.

Прежде, чем вы начнете

Чтобы отображать элементы XML на общую структуру анализа, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Чтобы включить для процессов пользовательского механизма анализа текста доступ к конкретным элементам в документах XML, или отобразить несколько элементов XML на общий тип для использования в семантическом поиске, можно создать пользовательские файлы отображения XML. Файлы отображения XML должны следовать шаблону UIMA для анализа текста.

При добавлении файлов отображения XML в собрание, где используется пользовательский механизм анализа текста, нужно включить отображение элементов XML в исходных документах на аннотации в общей структуре анализа. Затем эти аннотации сможет использовать пользовательский механизм анализа текста. После этого результаты анализа (в общей структуре анализа) можно отобразить на индекс и позволить пользователям делать запросы к аннотациям при поиске в собрании с применением семантического поиска.

Например, можно отобразить содержимое элементов `addressee` и `customer` на аннотации `Person` в общей структуре анализа. Эти аннотации станут доступными вашим пользовательским аннотаторам, которые смогут найти дополнительную информацию (например, пол сотрудника). Можно также отобразить аннотации `Person` на индекс поисковой системы предприятия, что позволит пользователям выполнять поиск сотрудников, не зная исходные элементы XML.

Если вы хотите разрешить пользователям задавать в запросах исходные элементы XML, отображения XML определять не нужно. Вместо этого можно сконфигурировать опции синтаксического анализа и включить собственный поиск XML для собрания.

Порядок действий

Чтобы отобразить элементы XML на общую структуру анализа:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. В области **Отобразить элементы XML на общую структуру анализа**, нажмите кнопку **Добавить отображение**.
3. На странице **Отобразить элементы XML на общую структуру анализа** введите для файла отображения XML описательное имя для вывода.
4. Задайте положение файла. Если файл отображения XML меньше 8 Мбайт, можно ввести его путь или выбрать этот файл. Если файл отображения XML больше 8 Мбайт, убедитесь, что он есть на сервере индексов, и задайте полный путь файла.
5. Нажмите кнопку **ОК**. Новый файл отображения XML будет добавлен в список файлов отображения XML на странице Опции обработки текста.

Отображение общей структуры анализа на индекс

Можно задать общую структуру анализа, которая будет использоваться при запросе пользователями собрания при помощи семантического поиска.

Прежде, чем вы начнете

Чтобы отображать общие структуры анализа в индексе, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Об этой задаче

Отображая результаты общую структуру анализа на индекс поисковой системы предприятия, вы позволяете пользователям задавать семантически точные запросы, что повышает качество результатов поиска.

Например, в зависимости от объектов и взаимосвязей, обнаруживаемых аннотаторами, пользователи могут выполнять поиск понятий, встречающихся в одном и том же предложении (таких как конкретное имя сотрудника и конкурента), или слова и понятия (например, найти термин Alex и номер телефона).

Порядок действий

Чтобы отобразить структуру анализа на индекс:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.
2. В области **Отобразить одну общую структуру анализа на индекс** нажмите кнопку **Выбрать общую структуру анализа**.
3. На странице **Выбрать общую структуру анализа** для этого собрания выберите отображение для использования с индексом поисковой системы предприятия:
 - Чтобы использовать для отображения структуры данных в общей структуре анализа на индекс поисковой системы предприятия системные правила по умолчанию, выберите **По умолчанию**.
 - Чтобы отобразить на индекс конкретную общую структуру анализа, задайте положение нужного файла общей структуры анализа. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
4. Нажмите кнопку **ОК**. Заданная вами общая структура анализа появится на странице Опции обработки текста.

Отображение общей структуры анализа на таблицы JDBC

Можно задать общую структуру анализа, которая будет отображаться на таблицы JDBC для использования в программах баз данных.

Прежде, чем вы начнете

Отображать общие структуры анализа на базы данных Java Database Connectivity (JDBC) может администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Отображая общую структуру анализа на таблицы JDBC, вы делаете данные доступными для использования прикладными программами базы данных. Например, пользователи могут задавать запросы SQL вне поисковой системы предприятия для выполнения поиска в аннотациях, добавленных общей структурой анализа. Эту информацию можно также использовать для дополнительной обработки текста, например, в программах исследования данных.

Порядок действий

Чтобы отобразить общую структуру анализа на таблицы JDBC:

1. Отредактируйте собрание, откройте страницу Анализ и выберите **Конфигурировать опции обработки текста**.

2. В области **Отобразить общие структуры анализа на таблицы JDBC** выберите **Добавить отображение**.
3. На странице **Отобразить общую структуру анализа на таблицы JDBC** введите описательное имя общей структуры анализа, которая будет использоваться для отображения информации на таблицы JDBC.
4. Задайте положение файла общей структуры анализа. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
5. Нажмите кнопку **ОК**. Этот новый файл будет добавлен в список файлов на странице **Опции обработки текста**. Будут также показаны дата и время добавления файла в систему.

Конфигурирование потоков для службы анализатора

Если хватает ресурсов памяти, можно увеличить число потоков, доступных синтаксическому анализатору для анализа документов.

Прежде, чем вы начнете

Число потоков анализатора может иметь смысл увеличить, если у вас большое число собраний. Убедитесь, что в системе достаточно памяти для поддержки дополнительных потоков. Анализатору с одним потоком требуется 200 Мбайт памяти. Каждому дополнительному потоку требуется 50 Мбайт памяти.

Конфигурировать число потоков, запускаемых для анализатора, может администратор поисковой системы предприятия или администратор собрания.

Порядок действий

Чтобы сконфигурировать число потоков синтаксического анализатора:

1. Откройте собрание для редактирования, откройте страницу **Анализ** и нажмите кнопку **Конфигурировать опции анализа**.
2. Задайте число потоков анализатора, запускаемых при запуске анализатора, и нажмите кнопку **ОК**.
3. Перезапустите анализатор.

Включение расширенного анализа для составных терминов

Можно улучшить качество поиска, разрешив синтаксическому анализатору использовать дополнительный анализ сложных терминов. Дополнительный анализ разделяет сложные термины на составные части, и каждая часть обрабатывается как отдельный термин.

Прежде, чем вы начнете

Чтобы задавать опции для синтаксического анализа сложных терминов, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Об этой задаче

В некоторых языках принято писать термины из нескольких корней в одно слово без пробелов (*сложные термины*). Дополнительный анализ и разделение сложных терминов полезны при поиске в таких языках, как немецкий, и критически важны в языках типа корейского.

Если вы включите дополнительный анализ для сложных терминов, пользователи смогут искать термины в составе сложных терминов без использования символов подстановки в терминах запроса. Например, поиск немецкого `Organ` (орган) может вернуть документы, содержащие термины `Organspender` (донор органов), но не вернет документы, содержащие `Organisation` (организация). Если при запросе с символом подстановки `Organ*` возможен возврат любой строки, начинающейся с `Organ`, при данном поиске признаются только те соответствия, где найдено полное слово - компонент сложного термина.

Термины пользовательских словарей, например, словаря синонимов и словаря значимых слов, также применяются к составным частям, используемым как отдельные слова в запросе.

Порядок действий

Чтобы разрешить дополнительный анализ сложных терминов:

1. Вызовите редактирование собрания, выберите Синтаксический анализ и щелкните по **Конфигурировать опции синтаксического анализа**.
2. Включите переключатель **Разрешить дополнительный анализ сложных терминов** и нажмите кнопку **ОК**.

Включение поддержки собственного поиска XML

Если собрание содержит документы XML, можно разрешить пользователям использовать при поиске документов разметку XML, включив для собрания поддержку собственного поиска XML.

Прежде, чем вы начнете

Включить поддержку поиска документов XML с использованием собственного поиска XML может только администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Собственный поиск XML, например, поиск XPath или фрагментов XML, может обеспечить более точные результаты поиска, благодаря использованию разметки XML документов в запросе. Например, пользователи могут задать, чтобы термин запроса встречался в определенном элементе или атрибуте XML.

Например, счета от продавца компьютеров в формате XML могут содержать записи `<order>` (заказ), включающие элементы `<company>` (компания) и `<computertype>` (тип компьютера). Если вы ищете найти счета, содержащие заказы на портативные компьютеры (notebook) фирмы IBM, поиск по ключевым словам IBM и notebook можно дать вам документы с информацией о портативных компьютерах Dell и о моделях настольных компьютеров IBM. Поиск XML позволяет задать, чтобы слово IBM присутствовало в элементе `<company>`, слово notebook - в элементе `<computertype>` и чтобы оба элемента находились в одном и том же элементе `<order>`. Так вы получите заказы, относящиеся конкретно к портативным компьютерам IBM.

Порядок действий

Чтобы разрешить пользователям выполнять поиск в собрании, используя собственный поиск XML:

1. Откройте собрание для редактирования, откройте страницу Анализ и нажмите кнопку **Конфигурировать опции анализа**.
2. Выберите **Разрешить пользователям выполнять поиск документов XML с использованием собственного поиска XML**.
3. Нажмите кнопку **ОК**.

Лингвистический анализ документов на китайском, японском и корейском языках

Чтобы улучшить обработку документов, написанных на китайском, японском и корейском языках, можно задать опции лингвистического анализа.

Можно указать анализатору, чтобы для лексического анализа документов на китайском, японском и корейском языках он использовал метод сегментации энграмм. Кроме того, можно сконфигурировать анализатор таким образом, чтобы для документов на китайском и японском языках он удалял символы новой строки из пустого пространства.

Сегментация энграмм

При создании собрания вы выбираете тип лексического анализа для анализа документов на языках, где для разделения слов не используется пробел.

При сегментации на основе пробелов Unicode в качестве разделителей между словами используются пробелы. При использовании сегментации энграмм перекрывающиеся последовательности из любого количества символов рассматриваются как отдельные слова. Для языков типа китайского, японского или корейского, в которых не используются пробелы в качестве разделителей слов, сегментация энграмм может дать более хорошие результаты при поиске, чем сегментация на основе пробелов Unicode.

Метод сегментации, используемый для анализа документов, задается при создании собрания. После создания собрания можно узнать метод сегментации (в опциях синтаксического анализа), но нельзя изменить его.

Удаление символов новой строки из диапазонов символов не-ASCII

Для языков, в которых для разделения слов не используются пробелы (например, китайского или японского), можно сконфигурировать анализатор, чтобы он удалял символы пробела, вызывающие переход на новую строку.

Прежде, чем вы начнете

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Если для собрания включена эта опция и документ написан на китайском или японском языке, анализатор будет удалять последовательности символов пробелов,

разделяющие двухбуквенные символы. Символы букв должны быть из диапазона символов Unicode, не входящих в набор ASCII. Удаляются следующие символы:

- Табуляция (0x09)
- Символ перевода строки, LF (0x0A)
- Символ возврата каретки, CR (0x0D)

Чтобы изменения вступили в силу, остановите и перезапустите анализатор. Чтобы изменения были применены для документов, которые уже проанализированы и сохранены в индексе, выполните повторный просмотр искателем и анализ этих документов, а затем реорганизируйте индекс.

Порядок действий

Чтобы удалять символы новой строки из пустого пространства:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. При помощи текстового редактора отредактируйте следующий файл (*ID_собрания* - это ID, заданный для собрания или назначенный системой при создании собрания):
`ID_собрания.parserdriver/collection.properties`
3. Измените значение свойства `removeCjNewlineChars` с `false` на `true`.

Типы документов, связанные с анализаторами собраний и сеансами Stellent

Чтобы обеспечить точный и эффективный анализ документов в пространстве искателя можно создать файлы конфигурации и задать в них, какие типы документов будут анализироваться анализатором собрания, а какие - сеансом фильтрации документов Stellent.

В собрании поисковой системы предприятия большинство форматов документов обрабатывается встроенными анализаторами HTML или XML. Некоторые типы документов обычно не анализируются (например, документы Postscript), другие типы документов обрабатываются функциями анализа Stellent (например, типы документов Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus Freelance, Lotus 123, PDF, RT и Ichitaro).

Поскольку метаданные бывают неточны, иногда документы в формате простого текста и HTML ошибочно направляются в сеанс Stellent, а оттуда обратно в один из встроенных анализаторов, что снижает производительность. Чтобы избежать такой ситуации, можно создать файлы конфигурации, задающие, где и как должны обрабатываться различные типы документов.

Чтобы связать типы документов с анализатором собрания и сеансом Stellent, надо выполнить следующие задачи:

1. Задание типов документов для анализатора собрания. Этот шаг включает создание файла конфигурации, в котором типы документов отображаются на анализатор, используемый собранием. Можно создать по одному такому файлу конфигурации для каждого собрания.
2. Задание типов документов для сеанса Stellent. Этот шаг включает в себя создание файла конфигурации, в котором типы документов отображаются на фильтры

документов Stellent, используемые собранием. Можно создать по одному такому файлу конфигурации для каждого собрания.

3. Остановка и перезапуск анализатора. Чтобы изменения вступили в силу, с консоли управления поисковой системы предприятия включите монитор собрания, для которого задали типы документов, после чего остановите и перезапустите анализатор.

Связывание типов документов с анализатором собраний

Чтобы связать типы документов с анализатором собраний, нужно создать файл конфигурации. С консоли администратора поиска эта задача не поддерживается.

Прежде, чем вы начнете

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Если файл конфигурации не существует, анализатор собраний использует правила службы анализатора по умолчанию. Если файл конфигурации существует, правила в этом файле задают:

- Какие документы отправлять в сеанс Stellent в зависимости от расширения файла или типа содержимого.
- Как анализировать документы, тип которых неизвестен из-за неполных метаданных.

Формат этого файла - последовательность строк, где каждая строка задает одно из следующих правил:

EXTENSION *расширение анализатор*

Все документы, URL которых заканчиваются указанным расширением, будут обрабатываться заданным анализатором. Не включайте в расширение точку; регистр символов при сравнении не учитывается.

CONTENTTYPE *тип/подтип анализатор*

Заданным анализатором будут обрабатываться все документы, тип содержимого которых соответствует указанному типу/подтипу. Тип/подтип содержимого документа будет соответствовать условию поиска, если тип документа эквивалентен значению тип, а подтип эквивалентен значению подтип (или если в качестве значения подтип задан символ подстановки (звездочка, *)).

UNKNOWN *анализатор*

Заданным анализатором будут обрабатываться все документы, у которых неизвестны расширение и тип содержимого (то есть документы, не сделанные доступными искателем).

DEFAULT *анализатор*

Заданным анализатором будут обрабатываться все документы, не обрабатываемые другими правилами.

Для всех всех случаев должен быть задан *анализатор* html, xml, stellent или none, где none означает, что документы данного типа обрабатываться не будут.

Если документу соответствует несколько правил, приоритет будет выше у более конкретного правила, независимо от их порядка:

- Правило EXTENSION более конкретно, чем правило CONTENTTYPE.
- Правило CONTENTTYPE, включающее подтип, более конкретно, чем то же правило с символом подстановки. Например, приоритет правила для типа содержимого application/postscript выше приоритета правила для типа application/*.
- Для одного расширения или типа содержимого не должно существовать двух правил. В этом случае приоритет правил зависит от реализации.

Порядок действий

Чтобы связать типы документов с анализатором собраний:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. Создайте файл конфигурации, как показано ниже (*ID_собрания* это ID конфигурируемого вами собрания):
ES_NODE_ROOT/master_config/*ID_собрания*.parserdriver/parserTypes.cfg
3. При помощи текстового редактора отредактируйте файл и задайте правила правила служб анализатора, затем сохраните и закройте этот файл.
4. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Пример

В следующем примере встроенный анализатор HTML обрабатывает все документы с расширением txt, htm или html, с типом содержимого начинающимся с text/ или с неизвестным расширением и типом содержимого. Встроенный анализатор XML обрабатывает все документы с расширением xml или типом содержимого text/xml. Все остальные документы, в том числе с типом содержимого, начинающимся с application/, посылаются в сеанс Stellent.

```
EXTENSION doc stellent
EXTENSION txt html
EXTENSION htm html
EXTENSION html html
EXTENSION xml xml
EXTENSION ps none
CONTENTTYPE text/xml xml
CONTENTTYPE text/* html
CONTENTTYPE application/* stellent
UNKNOWN html
DEFAULT stellent
```

Правила по умолчанию для служб анализаторов собрания

Если не создан файл конфигурации, в котором типы файлов и типы содержимого назначены искателю для собрания, при анализе документов будут использоваться правила по умолчанию.

По умолчанию анализатор собрания использует следующие правила:

```
EXTENSION pdf stellent
EXTENSION ppt stellent
EXTENSION prz stellent
EXTENSION lwp stellent
EXTENSION doc stellent
EXTENSION rtf stellent
EXTENSION xls stellent
```

```
EXTENSION 123 stellent
EXTENSION vsd stellent
EXTENSION vdx stellent
EXTENSION jxw stellent
EXTENSION jsw stellent
EXTENSION jtw stellent
EXTENSION jaw stellent
EXTENSION juw stellent
EXTENSION jbw stellent
EXTENSION jvw stellent
EXTENSION jfw stellent
EXTENSION jtt stellent
EXTENSION jtd stellent
EXTENSION jttd stellent
EXTENSION jttdc stellent
EXTENSION jttdx 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
```

Связывание типов документов с сеансом Stellent

Чтобы задать, какие типы документов должны анализировать фильтры документов Stellent, нужно создать файл конфигурации. С консоли администратора поиска эта задача не поддерживается.

Прежде, чем вы начнете

Для выполнения этой задачи нужно зарегистрироваться в качестве администратора поисковой системы предприятия.

Об этой задаче

Файл конфигурации задает:

- Файлы каких типов будет анализировать сеанс Stellent. Тип файла соответствует одному из типов файлов, распознаваемых библиотекой Stellent.
- Файлы каких типов будут отправляться назад анализатору собрания на обработку одним из встроенных анализаторов. (Эта операция необходима, поскольку анализатор собраний может отправить документ сеансу Stellent по ошибке, из-за неточных метаданных.)
- Файлы каких типов будут отклонены из-за отсутствия их поддержки в поисковой системе предприятия.

Если файл конфигурации задан, но не существует, запуск анализатора завершится неудачно. Если файл конфигурации не задан в файле `stellent.properties` для свойства `OutsideInSupportedTypes`, для сеансов Stellent будут использоваться правила синтаксического анализа по умолчанию.

Файл конфигурации содержит список типов документов и необходимые способы их обработки. Формат этого файла - последовательность строк, где каждая строка задает одно из следующих правил:

```
accept DEFAULT
accept ALL тип_документа
accept тип тип_документа
native DEFAULT
native тип тип_документа
reject тип
```

Где:

тип_документа

- значение, которое должно использоваться для маркера запроса `doctype`. Можно выполнять поиск документов по их типам. Например, для поиска документов PDF пользователь может задать `$doctype::pdf`.

тип - одно из значений `FI_` в библиотеке Stellent, а *тип_документа* - значение, которое должно использоваться для маркера `doctype`, если применяется правило.

DEFAULT

- означает, что в список принимаемых или собственных типов, в зависимости от типа правила, включается список типов по умолчанию. Эта опция позволяет не заменять конфигурацию по умолчанию, а вместо этого расширить ее.

ALL

- означает, что принимаются все типы с заданным маркером `doctype`, не перечисленные явно в списке.

Правила в файле конфигурации обрабатываются следующим образом:

- Если для значения *тип* задано правило `accept`, тип принимается (в том числе и типы списка по умолчанию, если задано `accept DEFAULT`).
- Если для значения *тип* задано правило `reject`, тип не принимается.
- Если задано правило `accept ALL`, тип принимается.
- Иначе тип не принимается.

Если тип документов принимается, то используется значение *тип_документа*, заданное в примененном правиле. Это значение отсылается назад анализатору собрания вместе с проанализированным содержанием. Если тип документа не принимается, выполняется следующее:

- Если для значения *тип* существует собственное правило (включая правила синтаксического анализа по умолчанию, если задано `native DEFAULT`), документ отсылается назад встроенному анализатору вместе со значением *тип_документа* для маркера `doctype`, задаваемого данным правилом. Значение *тип_документа* должно быть `txt`, `htm` или `xm1` (задающее соответственно простой текст, HTML или XML).
- Иначе документ отклоняется и не анализируется.

Порядок действий

Чтобы связать типы документов с сеансом Stellent:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. Отредактируйте файл `ES_NODE_ROOT/master_config/ID_собрания.stellent/stellent.properties` (где `ID_собрания` - ID конфигурируемого вами собрания).
3. Для свойства `OutsideInSupportedTypes` задайте абсолютный путь создаваемого вами файла конфигурации.
 Например, для одного собрания можно задать следующий файл конфигурации (и хранить его с другими файлами для конкретных собраний):
`ES_NODE_ROOT/master_config/ID_собрания.stellent/stellenttypes.cfg`
 Другой пример: для возможности использования одинаковых параметров со всеми собраниями можно создать следующий файл конфигурации (и хранить его с другими файлами системного уровня). (При использовании этого способа обязательно укажите его путь в файле `stellent.properties` для каждого собрания, как указано на шаге 2.)
`ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg`
4. При помощи текстового редактора создайте файл конфигурации и задайте правила анализа Stellent, затем сохраните и закройте этот файл.
5. Чтобы изменения вступили в силу, с консоли администратора поисковой системы предприятия включите монитор анализатора для собрания и перезапустите анализатор.

Пример

В следующем файле конфигурации сеанс Stellent принимает, кроме форматов списка по умолчанию, формат Microsoft Visio.

```
accept DEFAULT
accept FI_VISIO3 visio
accept FI_VISIO4 visio
accept FI_VISIO5 visio
accept FI_VISIO6 visio
```

В следующем файле конфигурации поддерживаются и доступны для поиска документы Postscript с типом документов `ps`, формат XPM отсылается назад встроенному анализатору текста, формат изображений PNG отклоняется, а также принимаются и делаются доступными для поиска все остальные типы файлов, с типом документов `other`.

```
accept DEFAULT
accept FI_POSTSCRIPT ps
native FI_XPIXMAP txt
accept ALL other
reject FI_PNG
```

Правила анализа по умолчанию для сеансов Stellent

Если не создан файл конфигурации, в котором типы файлов назначены фильтрам документов сеансов Stellent, при анализе документов будут использоваться правила по умолчанию.

По умолчанию сеанс Stellent использует следующие правила:

```
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_JAPANESEEUJ	htm
NATIVE FI_HTML_CHINESEBIG5	htm
NATIVE FI_HTML_CHINESEEUJ	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_JAPANESEJIS	xml
NATIVE FI_WML_JAPANESEEUC	xml
NATIVE FI_WML_KOREANHANGUL	xml
NATIVE FI_WML_LATIN2	xml
NATIVE FI_HTMLUNICODE	htm
NATIVE FI_XML_DOCTYPE_HTML	htm
NATIVE FI_XHTML	htm
NATIVE FI_ASCII	txt
NATIVE FI_ANSI	txt
NATIVE FI_UNICODE	txt
NATIVE FI_ASCII8	txt
NATIVE FI_ANSI8	txt
NATIVE FI_MAC	txt
NATIVE FI_MAC8	txt
NATIVE FI_SHIFTJIS	txt
NATIVE FI_CHINESEGB	txt
NATIVE FI_HANGEUL	txt
NATIVE FI_CHINESEBIG5	txt
NATIVE FI_LATIN2	txt
NATIVE FI_JAPANESE_EUC	txt
NATIVE FI_HEBREW_OLDCODE	txt
NATIVE FI_HEBREW_PC8	txt
NATIVE FI_HEBREW_E0	txt
NATIVE FI_HEBREW_WINDOWS	txt
NATIVE FI_ARABIC_710	txt
NATIVE FI_ARABIC_720	txt
NATIVE FI_ARABIC_WINDOWS	txt
NATIVE FI_7BITTEXT	txt
NATIVE FI_JAPANESE_JIS	txt
NATIVE FI_CENTRALEU_1250	txt
NATIVE FI_UTF8	txt
NATIVE FI_EBCDIC_37	txt
NATIVE FI_EBCDIC_273	txt
NATIVE FI_EBCDIC_277	txt
NATIVE FI_EBCDIC_278	txt
NATIVE FI_EBCDIC_280	txt
NATIVE FI_EBCDIC_284	txt
NATIVE FI_EBCDIC_285	txt
NATIVE FI_EBCDIC_297	txt
NATIVE FI_EBCDIC_500	txt
NATIVE FI_EBCDIC_870	txt
NATIVE FI_EBCDIC_871	txt
NATIVE FI_EBCDIC_1026	txt

Управление индексом поисковой системы предприятия

Чтобы у пользователей поисковой системы предприятия всегда был доступ к самой свежей информации, для всех собраний создаются индексы, и их содержимое периодически обновляется и реорганизуется.

Чтобы в данных, собранных искателями, можно было выполнять поиск, необходимо создать индексы. При первом создании собрания поисковая система предприятия создает индекс для всех данных, ранее собранных искателями. Когда искатели обрабатывают новые и измененные источники данных, поисковая система предприятия обновляет индексы, добавляя в них новое содержимое. В конечном счете обновленное содержимое необходимо объединить с основным индексом. Такой процесс объединения называется реорганизацией. При обновлении или реорганизации индекса новое содержимое копируется на поисковые серверы и становится доступным для поиска.

Искатели собирают данные непрерывно или периодически в соответствии с расписанием. Частое обновление индекса позволяет пользователям выполнять поиск наиболее свежих данных. Однако постоянно обновляемый индекс требует реорганизации. Чем больше становится обновляемый индекс, тем больше он потребляет системных ресурсов. Поэтому для сохранения оптимальной производительности индекс следует периодически реорганизовывать.

На то, как часто необходимо реорганизовывать индекс, влияют:

- Системные ресурсы (пространство в файловой системе, скорость процессора и оперативная память)
- Сколько документов должно быть просмотрено и перепросмотрено искателями
- Тип данных, обрабатываемых искателями
- Как часто вы изменяете правила категорий (изменения вступают в действие только после реорганизации индекса)
- Как часто вы запускаете искатель принудительно (вместо запуска в запланированное время)
- Как часто внешние искатели удаляют или добавляют URI (искатели этого типа взаимодействуют с поисковой системой предприятия через API программы приема данных)

Для собраний с несколькими миллионами документов, построенных преимущественно из Web-документов, следует реорганизовывать индекс примерно раз в день, а обновлять индекс - каждые 1-2 часа.

Для поддержания оперативного индекса, в котором можно выполнять поиск, необходимо выполнять следующие задачи:

- Задавать расписания для обновления и реорганизации индекса
- Изменять расписания для индексов
- Включать и отключать расписания для индексов
- Конфигурировать одновременное построение индексов

Для задания опций, влияющих на вид индекса для пользователя, можно выполнять следующие задачи:

- Конфигурировать поддержку символов подстановки в запросах

- Конфигурировать области действия для ограничения диапазона документов, в которых пользователи могут выполнять поиск
- Объединять в результатах поиска документы из одного источника
- Удалять URI из индекса

Планирование операций построения индекса

Можно запланировать реорганизацию индекса и обновление индекса с занесением в него новых данных.

Прежде, чем вы начнете

Запланировать построение индекса для собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Чтобы быть уверенным, что пользователи всегда обращаются к самой свежей информации в источниках при поиске, запланируйте регулярную реорганизацию и обновление индекса. При реорганизации перестраивается весь индекс. Процессы индексирования считывают все данные, собранные искателями и проанализированные анализаторами. При обновлении информация, найденная со времени последней реорганизации индекса, становится доступной для поиска.

По умолчанию выбрана опция планирования построения индекса. Эта опция сообщает процессу планировщика, что надо запланировать задачи обновления и реорганизации индекса каждый раз при запуске системы. Можно отключить переключатель **Включать при запуске системы** в любой момент, если вы хотите отменить запуск запланированного построения индекса. Например, может потребоваться отключить планирование для диагностики возникших проблем.

Порядок действий

Чтобы запланировать построение индекса:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Планировать построения индекса**.
2. Чтобы задать, как часто надо обновлять индекс, добавляя в него новое содержимое, определите следующие опции на странице Планирование построения индекса в области **Задайте расписание обновления индекса**:
 - a. В области **Начиная с** в полях **Год, Месяц, День, Час** и **Минута** задайте, когда индекс надо обновить в первый раз.
 - b. В области **Интервал обновления** в полях **дни** и **часы** определите частоту обновления индекса.
Рекомендуется обновлять индекс часто, каждый час или два. В зависимости от того, как часто изменяется содержание источника, задайте больший или меньший промежуток времени. Например, можно задать обновление каждый час (0 дней и 1 час) или каждые 12 часов (0 дней и 12 часов).
3. Чтобы задать, как часто индекс должен полностью перестраиваться, определите следующие опции в области **Задайте расписание реорганизации индекса**:
 - a. В области **Начиная с** в полях **Год, Месяц, День, Час** и **Минута** задайте, когда индекс надо реорганизовать в первый раз.
 - b. В области **Интервал обновления** в полях **дни** и **часы** определите частоту реорганизации индекса.

Рекомендуется регулярно, по крайней мере каждые 24 часа реорганизовывать индекс. В зависимости от того, как часто изменяется содержание источника, задайте больший или меньший промежуток времени. Например, можно задать реорганизацию каждые 12 часов (0 дней и 12 часов) или каждые два с половиной дня (2 дня и 12 часов).

4. Нажмите кнопку **ОК**.

Изменение расписания индексирования

Можно изменить расписание реорганизации или обновления индекса.

Прежде, чем вы начнете

Изменить расписание индексирования может администратор поисковой системы предприятия или администратор данного собрания.

Порядок действий

Чтобы изменить расписание индексирования:

1. Вызовите редактирование собрания, выберите страницу **Индекс** и измените соответствующие значения в полях **Месяц**, **День**, **Год** и **Час**. Задайте частоту обновления индекса с добавлением нового содержимого и частоту его реорганизации.
2. Нажмите кнопку **Применить**.

Включение и выключение расписаний индекса

Можно включать и выключать расписания для обновления и реорганизации индекса.

Прежде, чем вы начнете

Включать или выключать расписание индексирования может администратор поисковой системы предприятия или администратор данного собрания.

Об этой задаче

Выключить расписание для индекса можно, если требуется, чтобы запланированное построение этого индекса не запускалось. Например, может понадобиться отключить расписание, чтобы построение индекса не запускалось в запланированные дату и время, если в это время требуется исправление ошибок.

Включать и выключать расписание можно при редактировании собрания, а также при слежении за собранием.

Порядок действий

1. Чтобы включить или выключить расписание для индекса при редактировании собрания, выполните следующие действия:
 - a. Вызовите редактирование собрания, которое вы хотите изменить.
 - b. На странице **Индекс** включите или выключите переключатель **Включать при запуске системы**, чтобы включить или выключить расписание для обновления индекса.
 - c. Чтобы включить или выключить реорганизацию индекса, включите или выключите переключатель **Включать при запуске системы**.
 - d. Нажмите кнопку **Применить**.

2. Чтобы включить или выключить расписание для индекса при слежении за собранием, выполните следующие действия:
 - a. Включите слежение за собранием, которое вы хотите изменить.
 - b. На странице Индекс, если индекс включен в расписание, но вы не хотите, чтобы построение началось в запланированный момент времени, выберите **Отключить расписание**. Построение индекса не будет выполняться до тех пор, пока вы не включите расписание или не нажмете кнопку **Запуск** для запуска процесса построения индекса.
 - c. Если индекс внесен в расписание, но расписание для построения отключено, выберите **Включить расписание**.
Индекс будет поставлен в очередь на построение на дату и время, указанные в расписании индексирования.

Конфигурирование одновременных операций построения индекса

Управляя использованием ресурсов индексации, вы задаете, для скольких собраний могут одновременно обрабатываться требования построения индекса. При достаточных ресурсах системы можно повысить качество поиска, разрешив обновление индекса собрания при его реорганизации.

Прежде, чем вы начнете

Задавать опции построения индексов для системы может только администратор поисковой системы предприятия.

Об этой задаче

Поисковая система предприятия может создавать несколько индексов одновременно, разделяя ресурсы между собраниями, что позволяет параллельную обработку требований построения индекса для нескольких собраний. Путем совместного использования процессов можно добиться того, что реорганизация очень большого индекса не будет блокировать доступность других индексов, ожидающих построения в очереди.


Когда затребовано или запланировано построение индекса, этот индекс ставится в очередь индексирования, где и дожидается обработки. Так как у каждого собрания есть собственный индекс, одновременно в очереди индексирования может находиться несколько запросов на построение индекса. При настройке опций индексации для системы вы задаете, сколько собраний может совместно использовать ресурсы индексации для параллельной обработки требований.

Кроме того, можно задать обработку требований обновления индекса собрания параллельно реорганизации индекса собрания. Если включить эту опцию, поисковые серверы будут получать последние документы (в ходе обновления индекса) во время более длительного процесса реорганизации индекса. С другой стороны, построение индекса требует большого количества ресурсов. Во время построения индекса используется большой объем системной памяти и дискового пространства. Если включить эту опцию при нехватке дискового пространства или памяти, возможно снижение производительности системы в целом.

Порядок действий

Чтобы задать опции построения индекса для системы:

1. Выберите **Система**, чтобы открыть окно Система.

2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Индекс нажмите кнопку **Конфигурировать опции индексирования**.
4. На странице Опции индексации уровня системы введите число собраний, для которых разрешено совместное использование ресурсов системы и параллельная обработка требований построения индекса.
Число собраний, совместно использующих ресурсы индексации, не может превышать число собраний в вашей поисковой системе предприятия. Например, если у вас есть пять собраний, нужно ввести число не больше пяти.
5. При достаточных ресурсах системы для поддержки параллельного построения индекса для нескольких собраний вы можете включить опцию, которая разрешает параллельное выполнение требований обновления и реорганизации индекса.
6. Нажмите кнопку **ОК**.

Опции, влияющие на представление индекса для поиска

После индексации документов вы можете задавать опции, управляющие возможностями пользователей по поиску документов и просмотру найденных документов.

Чтобы задать опции, влияющие на представление индекса для пользователя, можно выполнить следующие задачи:

- Сконфигурировать поддержку терминов запроса с символами подстановки. Можно встроить поддержку запросов с символами подстановки в индекс, а можно задать опции расширения терминов запроса при обработке запроса.
- Задать области действия, чтобы ограничить диапазон документов, в которых пользователи могут выполнять поиск. Когда пользователи выполняют поиск в этом собрании, поиск выполняется только в документах из области действия, а не по всему индексу.
- Объединять найденные документы из одного источника. Можно группировать документы, соответствующие некоторому URI или шаблону URI в индексе, и показывать только первые найденные документы (пользователи смогут задавать опции, чтобы увидеть объединенные документы).
- Удалять URI из индекса. Иногда нужно предотвратить поиск по некоторым документам в индексе.

Индексированные опции для поиска документов

При конфигурировании опций для поиска в данных искателя или внешних источниках или при отображении элементов метаданных XML и HTML на поля поиска можно задать, как выполнять поиск документов и как выводить документы в результатах поиска.

При редактировании опций для искателей, содержащих поля, доступны следующие опции управления, задающие, можно ли выполнить поиск в поле, как его можно выполнить и может ли поле быть возвращено в результатах поиска:

- Свободный текстовый поиск
- Поиск по полям
- Параметрический поиск
- Результаты поиска
- Содержимое документа

При редактировании опций для внешних источников опция, позволяющая пометить поле как поле содержимого документа, недоступна.

Если при конфигурировании синтаксического анализатора задается отображение элементов метаданных HTML и XML на поля в индексе, задайте, могут ли использоваться эти поля при поиске по полям, параметрическом поиске или в результатах поиска.

Заданные вами опции хранятся в индексе с документами. Они позволяют ограничить данные, доступные пользователям для запроса, и данные, доступные им для просмотра в результатах поиска.

Свободный текстовый поиск

Индекс поисковой системы предприятия - это полнотекстовый индекс с содержимым с разных источников данных. Можно выполнить поиск по содержимому, введя простой запрос на естественном языке. Процесс поиска выполнит поиск по полям и содержимому, чтобы найти документы, соответствующие запросу.

Пример:

Свободный текстовый поиск может быть очень простым, например как следующий запрос:

велосипедная цепь

Чтобы указать, какие слова должны и какие не должны встречаться в документе, можно использовать специальную запись. Например, перед словом можно поставить знак плюс (+), это означает, что найденный документ должен содержать это слово. Если перед словом поставить знак минус (-), содержащие это слово документы будут исключены из результатов поиска. Для поиска точных словосочетаний заключите два или несколько слов в двойные кавычки ("").

Пример:

Следующему свободному текстовому запросу будут соответствовать только документы, содержащие точную фразу научная фантастика и не содержащие слова робот:

+ "научная фантастика" -робот

Поиск по полям

Поиск по полям позволяет ограничить область поиска определенными полями и метаданными документа. Например, можно потребовать, чтобы определенные слова содержались в заголовке документа.

Чтобы воспользоваться поиском по полям в поиске предприятия, включите в ваш запрос имя поля и слово или словосочетание, которые должны содержаться в этом поле.

Пример:

Следующий запрос ищет документы, содержащие слово `ibm` и словосочетание поисковая система предприятия в поле заголовка:

```
title:ibm title:"поисковая система предприятия"
```

Чтобы можно было использовать поиск по определенному полю, при конфигурировании искателя надо разрешить поиск по полям для этого поля.

Параметрический поиск

Параметрический поиск - это тип поиска по полям, позволяющий выполнять сравнительные и вычислительные запросы для числовых полей, полей даты и метаданных. Например, можно выполнить поиск документов определенного размера или документов, созданных после определенной даты. Кроме того, можно выполнить поиск документов с атрибутами большими, меньшими или равными определенному значению.

Пример 1:

Следующий запрос находит объекты, которые стоят ровно 50 долларов (или 50 единиц той валюты, которая используется для поля цена):

```
#price:=50
```

Пример 2:

Следующий запрос находит документы, у которых размер файла превосходит 1024, но не превосходит 2048:

```
#filesize:>1024<=2048
```

Чтобы можно было использовать параметрический поиск по определенному полю, при конфигурировании искателя надо разрешить параметрический поиск для этого поля.

Результаты поиска

Можно выполнить поиск в некоторых полях, но не выводить эти поля в результатах поиска, или вывести поле в результатах поиска, даже если это поле не запрашивалось. Например, вам может понадобиться запрос к финансовым данным, чтобы получить содержательный отчет, но вы не хотите выводить зарплаты сотрудников в результатах рядом с их фамилиями.

Содержимое документа

При помощи содержимого поля, помеченного как поле содержимого документа, можно связать документ с категорией на основе модели и с категорией на основе правил для содержимого документа.

Символы подстановки в запросах

Можно разрешить пользователям включать в термины поиска символы подстановки и искать слова, соответствующие определенным шаблонам.

Термин запроса с символами подстановки - это термин, содержащий символ звездочки (*). Когда пользователь передает запрос с символами подстановки, в результаты поиска включаются все документы в индексе, в которых есть термин запроса, и все документы в индексе, в которых есть слова, соответствующие шаблону с символом подстановки. Например, символ подстановки на конце термина запроса мор* может соответствовать словам море, морковь и мораль.

При конфигурировании опций символов подстановки для индекса выбирают, разрешать ли пользователям задавать символы подстановки в запросах, и, если символы подстановки разрешены, как должна осуществляться поддержка символов подстановки:

- Можно разрешить поиск слов по шаблону с символами подстановки во всех частях документа, а можно ограничить поиск по шаблону только полями.
- Можно разрешить поддержку запросов с символами подстановки для всех полей, а можно ограничить поиск по шаблону заданными полями.
- Можно задать, что символ подстановки разрешается только в последней позиции термина запроса (конечный символ подстановки), а можно разрешить символ подстановки в любом месте термина запроса. (В имени поля символ подстановки не допускается.)
- В зависимости от разрешенных мест для символов подстановки можно выбрать варианты расширения терминов запроса (термины запросов с символами подстановки расширяются до всех соответствующих терминов из индекса). Можно хранить все возможные версии расширения терминов в индексе, а можно выполнять расширение терминов при обработке запроса.

Все изменения параметров символов подстановки начинают действовать при следующей реорганизации индекса.

Расширение в индексе

Чтобы включать расширения терминов в индекс, нужно задать, сколько начальных символов слова должны совпадать с символами шаблона в термине запроса, чтобы слово считалось соответствующим этому шаблону. Результативными будут только те термины запроса, которые содержат не меньшее число символов (не считая *). Например, если задать 4, для фиксации совпадения термин запроса должен содержать минимум четыре символа.

Если ввести 4, слово технология будет соответствовать терминам запроса техн* и техно*, но не те*.

При обновлении или реорганизации индекса индексируются не только исходные термины, но и все возможные расширения каждого термина в документах. Преимущество этого подхода в том, что при обработке запроса дополнительное время на расширение терминов не требуется. Однако при этом увеличивается размер индекса, поэтому нужно иметь достаточный объем системных ресурсов для индекса большего размера.

Такой подход лучше использовать для собраний относительно небольшого размера или когда пространство и время, затрачиваемые при построении индекса, не столь важны, как время ответа на запрос. Например, этот подход можно выбрать для поиска в каталоге или списке работников.

Данный подход применим только тогда, когда выбрана поддержка символов подстановки на конце слова. Если включить поддержку символов подстановки в любом месте термина запроса, опция, которая вставляет варианты расширения терминов в индекс, будет недоступна.

Расширение в запросе

Чтобы расширение терминов запросов и применение правил соответствия шаблону происходили при передаче пользователем запроса, содержащего символы подстановки, нужно задать, сколько вариантов термина запроса образуют множество

терминов для поиска. Например, если задать число 50, в качестве терминов для поиска могут использоваться до 50 расширений термина запроса.

Так термину запроса `техн*` будут соответствовать слова `технический`, `техника`, `технология` и другие слова (не более 50 в нашем примере), начинающиеся с `техн`.

Расширение в запросе слабо влияет на размер индекса, но может снизить производительность запросов. Процессы поиска должны проверить все возможные расширения термина запроса с символами подстановки вплоть до предельного числа таких расширений, заданных в параметрах символов подстановки.

Такой подход лучше использовать для собраний относительно большого размера, а также когда необходимо минимизировать пространство и время, затрачиваемые при построении индекса. Например, этот подход можно выбрать для репозитория электронной почты, для которых индекс должен постоянно соответствовать быстро меняющемуся набору документов, а время ответа на запрос не столь критично.

Данный подход применим независимо от того, включена ли поддержка только конечных символов подстановки или же символов подстановки в любом месте термина запроса.

Поддержка символов подстановки в запросах

Набор расширений термина запроса с символами подстановки содержит все термины индекса, которые можно получить заменой символа подстановки на произвольную последовательность символов. Этот набор определяется так:

- Если собрание поддерживает символы подстановки в любом месте термина запроса, всякий термин запроса со звездочкой интерпретируется как термин с символом подстановки.
- Набор содержит не больше максимального числа расширений, сконфигурированного администратором поисковой системы предприятия. Если в индексе есть расширения сверх этого числа, они игнорируются. (В результатах поиска указывается, если некоторые расширения символов подстановки были проигнорированы.)
- Если поддержка символов подстановки ограничивается набором полей, набор расширений содержит только термины, которые есть в одном из заданных полей. Достаточно присутствия термина в одном поле одного документа данного индекса.
- Если термин запроса - термин поля, символ подстановки должен находиться после спецификатора поля (например, `имя_поля:*сфера`). Имя поля не может содержать двоеточие (:).
- Если поддержка символов подстановки ограничивается набором полей, имя поля в термине запроса с символами подстановки должно быть одним из имен, заданных на консоли администратора поисковой системы предприятия. В противном случае для этого термина никакие расширения найдены не будут.
- Символы подстановки поддерживаются только в терминах с простым текстом, но не в именах элементов XML, именах атрибутов или значениях атрибутов. Не поддерживается термин, состоящий из одного лишь символа подстановки.

Конфигурирование опций для символов подстановки в запросах

При конфигурировании опций индексирования для собрания поисковой системы предприятия можно указать, разрешить ли пользователям включать в условия запросов символы подстановки.

Прежде, чем вы начнете

Чтобы сконфигурировать опции для символов подстановки, надо быть администратором поисковой системы предприятия или администратором собрания, к которому принадлежит индекс.

Об этой задаче

Изменения, вносимые при задании опций для символов подстановки, вступают в силу при следующей реорганизации индекса.

Порядок действий

Чтобы сконфигурировать поддержку символов подстановки в запросах:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Конфигурировать опции для символов подстановки**.
2. На странице Опции для символов подстановки включите переключатель **Поддерживать символы подстановки в запросах**.
3. При помощи переключателя **Поддерживать символы подстановки в запросах, не выполняющих поиск в полях** задайте, будут ли символы подстановки поддерживаться в запросах, выполняющих поиск текста вне полей. Например, свободный текстовый запрос техн*, не выполняющий поиск в явно заданном поле, будет возвращать развернутые результаты (такие как технология или техника), только если включен этот переключатель.
4. Задайте, в каких полях будут поддерживаться символы подстановки:
 - Чтобы включить поддержку запросов, содержащих символы подстановки, для всех полей в документе, выберите **Все поля**.
 - Чтобы ограничить поддержку символов подстановки для некоторых полей, выберите **Конкретные поля**, а затем введите нужные имена полей. Развернутые результаты будут возвращаться только для заданных здесь полей. Например, запрос `author:john*` будет возвращать развернутые результаты, только если задать поддержку символов подстановки для поля `author`.
5. Задайте, должен ли символ подстановки стоять только в конце термина запроса (концевой символ подстановки) или же он может применяться без ограничений, в любом месте термина запроса.

При выборе позиции и типа символа подстановки нужно также задать способ включения поддержки символов подстановки. При конфигурировании опций для символов подстановки можно получить подробную информацию, нажав кнопку **Справка**.
6. Нажмите кнопку **ОК**.

Области действия

Если нужно показать пользователям сокращенное представление собрания, сконфигурируйте область действия.

Область действия - это группа связанных URI в индексе. При конфигурировании области действия вы ограничиваете видимые для пользователей документы собрания. Когда пользователи выполняют поиск в этом собрании, поиск выполняется только в документах из области действия, а не по всему индексу. Чтобы использовать данную возможность, ваши программы поиска должны включать поддержку областей поиска.

Создавая область действия, вы задаете диапазон URI в индексе, в котором пользователи смогут выполнять поиск. Ограничение документов, в которых пользователи могут выполнять поиск, повышает вероятность того, что документы в результатах поиска будут соответствовать запросам пользователей.

Например, можно создать одну область действия, включающую в себя URI для отдела технической поддержки, и другую область действия с URI для отдела кадров. Если ваши поисковые программы поддерживают области действия, пользователи в отделе технической поддержки будут получать документы из области действия для отдела технической поддержки, а пользователи из отдела кадров - из области действия для отдела кадров.

Можно создать любое число областей действия, хотя создание слишком большого числа областей действия может повлиять на производительность. Конфигурируйте области действия таким образом, чтобы большинство требований поиска фильтровалось только по одной или двум областям действия. Поскольку области действия могут содержать целые URI или шаблоны URI, один и тот же документ может принадлежать к нескольким областям действия.

После конфигурирования областей действия может потребоваться дважды реорганизовать индекс, чтобы изменения вступили в силу. Если сконфигурировать области действия перед первой реорганизацией индекса для собрания, пользователи смогут выполнить поиск в собрании, но не смогут увидеть данные областей действия в результатах поиска. Чтобы результаты поиска гарантированно отражали диапазон URI в области действия, нужно реорганизовать индекс повторно.

Если сконфигурировать области действия хотя бы после одной реорганизации индекса, изменения вступят в силу после следующей реорганизации индекса.

Конфигурирование областей действия

При конфигурировании области действия для собрания поисковой системы предприятия задаются URI или шаблоны URI, определяющие документы в индексе, для которых пользователям разрешено выполнять поиск.

Прежде, чем вы начнете

Конфигурировать области действия может только администратор поисковой системы предприятия или администратор собрания, к которому принадлежат эти области действия.

Об этой задаче

Если ваши прикладные программы поиска поддерживают области действия, пользователи смогут выполнять поиск только в документах, соответствующих заданным URI (эти URI определяют границы области действия для поиска в собрании).

После конфигурирования областей действия может потребоваться дважды реорганизовать индекс, чтобы изменения вступили в силу. Если сконфигурировать области действия перед первой реорганизацией индекса для собрания, пользователи смогут выполнить поиск в собрании, но не смогут увидеть данные областей действия в результатах поиска. Чтобы результаты поиска гарантированно отражали диапазон URI в области действия, нужно реорганизовать индекс повторно.

Если сконфигурировать области действия хотя бы после одной реорганизации индекса, изменения вступят в силу после следующей реорганизации индекса.

Порядок действий

Чтобы сконфигурировать область действия:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать области действия**.
2. На странице Области действия выберите **Создать область действия**.
3. Задайте имя для области действия и URI и шаблоны URI, определяющие границы этой области действия. Можно также задать URI и шаблоны URI, исключаемые из области действия.
4. Нажмите кнопку **ОК**.

Новая область действия появится в списке на странице Области действия вместе с другими областями действия для этого собрания.

Сворачивание URI

Поисковая система предприятия может расположить результаты поиска так, чтобы документы из источников Web и NNTP с одним префиксом URI выводились в результатах поиска одной свернутой группой.

При объединении результатов первый результат обычно выводится по левому краю страницы. Другие результаты с более низким рангом группируются и выводятся под первым результатом со сдвигом вправо.

Для объединения в одну группу документов результатов с разными префиксами URI можно связать префиксы URI с именем созданной вами группы. Например, если у вас есть три сервера для управления финансовыми данными, можно сгруппировать в результатах поиска документы со всех трех серверов и разместить результаты с более низким рангом под результатами с более высоким рангом.

Поисковые прикладные программы могут объединять документы в результатах поиска с помощью префиксов URI или имен групп. В поисковой программе примера поисковой системы предприятия показаны два верхних документа - результата поиска. Если возвращено более двух документов с одним префиксом URI (или относящихся к одной группе URI), можно выбрать опцию для просмотра остальных объединенных результатов.

Пользователи могут использовать синтаксис запросов поисковой системы предприятия (`samegroupas:префикс URI`) для поиска всех документов, находящихся в той же самой группе, что и указанный префикс URI.

Как организовать префиксы URI и имена групп

Используя консоль администратора для конфигурирования правил объединения результатов поиска, вы задаете префиксы URI документов, которые хотите объединять; кроме того, можно связать префиксы URI с именем группы.

Порядок, в котором заданы префиксы URI, имеет значение. Серверы индекса используют порядок префиксов URI при вычислении значения каждого URI в собрании. Для каждого URI:

1. Сервер индекса последовательно просматривает префиксы URI в правилах объединения результатов поиска.

2. Когда сервер индекса находит первый префикс URI, соответствующий префиксу документа в индексе, он добавляет имя группы (или префикс URI, если в правиле не задано имя группы) в качестве дополнительного термина поиска для документа.
3. Если для документа не найден соответствующий префикс URI:
 - Для URI Web сервер индекса использует в качестве префикса URI имя хоста URL.
 - Для URI NNTP сервер индекса использует в качестве префикса URI ID первого сообщения в значении заголовка ссылки.

Добавив префикс URI в список префиксов для объединения результатов поиска, нужно поместить этот префикс URI в место списка, соответствующее желаемому порядку, в котором сервер индекса будет проверять его и, возможно, задавать в качестве дополнительного термина поиска для документов в индексе:

- Если вы добавили префикс URI и не связали его с именем группы, можно выбрать этот отдельный префикс URI и переместить его вверх или вниз в списке.
- Если вы добавили префикс URI и связали его с именем группы, при перемещении этого префикс URI вверх или вниз в списке перемещается вся группа префиксов URI. Порядок префиксов URI внутри группы не имеет значения; при выборе отдельного префикса URI автоматически выбирается вся группа.)

Объединение URI в результатах поиска

Можно задать опции для группировки и объединения документов результатов из источников Web и NNTP с одинаковым префиксом URI. Можно также задать имя группы для объединения полученных документов с разными префиксами URI.

Прежде, чем вы начнете

Чтобы задавать опции для объединения результатов поиска, надо быть администратором поисковой системы предприятия или администратором данного собрания.



Об этой задаче

Изменения, внесенные в объединении результатов поиска, вступают в действие только после следующей реорганизации индекса.

Порядок действий

Чтобы задать опции для объединения результатов поиска:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Объединить результаты поиска**.
2. На странице Объединить результаты поиска нажмите кнопку **Добавить префикс URI**.
3. На странице Добавить префикс URI для объединения результатов введите префикс URI для документов, которые вы хотите объединять в результатах поиска. Например:
`http://finance/ROI/`
`http://server1.com/finance/`
4. Можно ввести описательное имя группы, которое вы хотите связать с этим префиксом URI. Чтобы объединить в одну группу документы, найденные в разных источниках, при добавлении каждого из префиксов URI вводите одно и то же имя группы.

5. Нажмите кнопку **ОК**.
6. На странице Объединить результаты поиска разместите новое правило в соответствии с требуемым порядком его просмотра сервером индексов:
 - Если вы добавили префикс URI и не связали его с именем группы, этот новый префикс URI появится в нижней части списка. Переместите его в необходимое положение при помощи клавиш со стрелками.
 - Если вы связали новый префикс URI с именем группы, этот новый префикс URI появится в нижней части набора префиксов URI, принадлежащих к этой группе. Используйте клавиши со стрелками, чтобы переместить всю группу префиксов URI в необходимое положение.
7. Чтобы изменить префикс URI или имя группы, выберите этот префикс URI и нажмите кнопку  **Правка**.
8. Чтобы удалить префикс URI из списка, выберите этот префикс URI и нажмите кнопку  **Удалить**.

Удаление URI из индекса

Чтобы запретить пользователям поиск некоторых документов в собрании, можно удалить URI этих документов из индекса.

Прежде, чем вы начнете

Для удаления URI из индекса надо быть администратором поисковой системы предприятия или администратором этого собрания.

Об этой задаче

Если указать полный URI, пользователи перестанут видеть этот URI в результатах поиска. Однако если пользователь повторяет один из предыдущих запросов, и результаты поиска для этого запроса находятся в кэше поиска, результаты с удаленным URI будут продолжать возвращаться в результатах поиска. Кэш поиска не обновляется, и этот URI не удаляется из индекса до следующего его обновления или реорганизации.

Если для удаления нескольких URI вы указали шаблон, пользователи будут видеть эти URI в результатах поиска до следующего обновления или реорганизации индекса.

При удалении URI из индекса он не удаляется из пространства искателя. Когда искатель в следующий раз обнаружит этот документ, он снова впишет его в индекс, и документ станет доступен для поиска опять. Чтобы удалить URI из пространства для искателя, надо изменить правила искателя, исключив из них источник этого документа, после этого остановить и перезапустить искатель.

Порядок действий

Чтобы удалить из индекса URI для определенных документов:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Удалить URI из индекса**.
2. На странице Удалить URI из индекса введите URI (или шаблоны URI), которые хотите удалить из индекса.

Например:

```
http://domain.org/ru/*  
db2://knowledgeManagement/ROI*  
cm://enterprise/finance*
```

3. Нажмите кнопку **OK**.

Управление серверами поиска для поисковой системы предприятия

Для серверов поиска можно задать опции, управляющие: использованием пространства кэша для возвращаемых результатов поиска, максимальной длиной имен для вывода сводок документов в результатах поиска, связыванием пользовательских словарей для улучшения результатов поисков и возвращением в результатах поиска предварительно определенных URI при появлении в запросе определенных терминов.

Когда пользователь передает запрос, серверы поиска используют индекс для быстрого обнаружения нужных документов. Чтобы получить для нужных документов метаданные, серверы поиска используют склад поисковой системы предприятия, содержащий данные синтаксического и лексического анализа. Метаданные могут содержать идентификатор URI, заголовок, описание, дату, тип данных документов, а также другую информацию.

При конфигурировании серверов поиска для собрания можно задать опции, управляющие обработкой запросов, включая опции, которые могут повысить производительность запросов:

Конфигурирование кэша поиска

Для оптимизации производительности можно указать, что документы результатов должны помещаться в кэш, и сконфигурировать объем пространства, выделяемый для документов в кэше.

Конфигурирование максимальной длины вывода для сводок документов

Большинство документов результатов содержат сводку содержимого документа, которая помогает пользователям решить, нужен ли им полученный документ. Можно указать, какой объем пространства использовать в результатах поиска для вывода этой информации сводки.

Задание другого языка по умолчанию

Когда создается собрание, задается язык по умолчанию для поиска документов в этом собрании, но при необходимости можно указать другой язык.

Связывание пользовательских словарей

Если ваши разработчики прикладных программ создали пользовательские словари синонимов, стоп-слов или значимых слов, можно задать использование этих словарей при выполнении пользователями поиска в собрании.

Конфигурирование прямых ссылок

Можно заранее задать идентификаторы URI, возвращаемые для определенных ключевых слов и фраз. Если пользователи укажут эти ключевые слова или фразы в запросе, в результатах поиска будет возвращен заранее заданный URI. Эти URI прямых ссылок возвращаются в добавление к URI, возвращаемым серверами поиска при поиске в индексе.

Кэши поиска

Если число документов в результатах потенциально велико, можно повысить производительность, кэшируя результаты поиска.

Когда серверы поиска обрабатывают требования поиска, они сначала проверяют, нет ли в кэше результатов для такого же запроса. Если поисковые серверы обнаружат подходящие документы результатов, они могут быстро вернуть результаты поиска пользователю. Если поисковые серверы не обнаружат подходящих документов, они будут выполнять поиск в индексе.

При заполнении кэша поиска из него удаляются самые старые документы результатов, а также документы для редко обрабатываемых запросов, тем самым освобождая место для новых результатов поиска.

С консоли администратора поисковой системы предприятия можно включить поддержку кэширования и задать емкость кэша (число запросов, результаты которых могут кэшироваться одновременно).

При изменении параметров кэша поиска необходимо перезапустить поисковые серверы, чтобы изменения вступили в действие.

Конфигурирование кэша поиска

Вы можете включить или отключить кэш поиска для собрания. Кроме того, можно задать опции, управляющие размером кэша поиска.

Прежде, чем вы начнете

Конфигурировать кэш поиска для собрания может пользователь с ролью администратора поисковой системы предприятия или администратор собрания.

Порядок действий

Чтобы сконфигурировать кэш поиска:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. На странице Опции сервера поиска включите переключатель **Использовать кэш поиска**.
3. В поле **Максимальное число элементов в кэше** задайте максимальное число наборов результатов поиска, которые могут находиться в кэше поиска.
4. Нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, запустите мониторинг серверов поиска и перезапустите процессы сервера.

Пользовательские словари синонимов

Для повышения качества результатов поиска можно разрешить пользователям при поиске в собрании искать также синонимы терминов запроса.

Если вы создали словарь синонимов, добавили его в систему поисковой системы предприятия и связали с собранием, пользователи могут при поиске в этом собрании искать документы, содержащие синонимы терминов их запросов. Расширяя запросы таким образом, пользователи могут с большей вероятностью найти все интересующие их документы, а не только документы, точно соответствующие заданным терминам запроса. Поскольку при создании словаря синонимов вы определяете, какие слова являются синонимами, пользователи смогут найти нужные документы, не задавая всех вариантов терминов запроса.

Например, в организации могут использоваться акронимы и сокращения для обозначения отделов организации, оборудования и т.п., или же документы в собраниях могут содержать специальные термины вашей области деятельности. Создав словарь синонимов, можно сделать, чтобы для запроса, содержащего акроним (например, ACL), возвращались документы, содержащие другие формы этого термина (например, ACLs, списки управления доступом, элементы управления доступом и т.д.).

Язык запросов поисковой системы предприятия поддерживает использование синонимов, позволяя пользователям добавлять символ тильды (~) перед термином запроса. Например, запрос ~WAS может возвращать документы, содержащие "WebSphere Application Server". Разработчики прикладных программ могут также сделать синонимы доступными, используя свойства запроса (при этом не требуется специальный синтаксис).

Словари синонимов содержат различные варианты слов; эти словари обладают следующими особенностями:

- Слова не относятся к определенному языку и могут использоваться в различных языках. У собрания есть только один словарь синонимов.
- Слова задаются в основной словарной форме. Все возможные формы нужно добавить в список синонимов. Например, такими формами могут быть формы различных падежей или множественного числа.

Большинство терминов, добавляемых в словарь синонимов - это точные семантические эквиваленты, то есть если термин А - синоним термина В, то и термин В - синоним термина А. В любом запросе вместо термина А можно использовать термин В и наоборот.

Однако можно также добавить термины, соответствующие различным использованиям термина, включая общие и более частные варианты термина. Например, можно создать одну группу синонимов, содержащую слова строение и дом, и другую группу, содержащую слова банки, сберкассы и тара.

Чем менее строго соответствуют друг другу термины, тем шире набор результатов поиска, но при этом некоторые из результатов поиска могут оказаться не относящимися к делу документами. API поиска и индексации содержит методы, позволяющие пользователю выбирать подходящие синонимы при задании требования поиска, и методы, показывающие пользователям, какие термины запроса были расширены при помощи синонимов.

Для создания словаря синонимов специалист по предметной области, для которой предназначено собрание, должен создать список синонимов в формате XML (или попросить разработчика прикладной программы, чтобы он создал файл XML). Для преобразования файла XML в двоичный файл (.dic) надо использовать инструмент, поставляемый вместе с WebSphere II OmniFind Edition.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь синонимов для поиска документов в собрании, когда конфигурируют опции серверов поиска для собрания.

Ограничение: Добавленный в систему пользовательский словарь синонимов нельзя редактировать. Чтобы исправить синонимы, доступные для собрания, нужно:

1. Изменить исходный файл XML.

2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить старый словарь синонимов из собраний, в которых он используется.
4. Удалить старый словарь синонимов из системы.
5. Добавить в систему новый словарь синонимов.
6. Связать новый словарь синонимов с собраниями, которые будут его использовать.

Добавление словарей синонимов в систему


Если вы создаете пользовательские словари синонимов для поиска документов в собрании, необходимо связать эти словари с поисковой системой предприятия. После этого можно выбрать, какой из словарей синонимов надо использовать для поиска в собрании.

Прежде, чем вы начнете

Чтобы добавлять пользовательские словари синонимов для использования с запросами поисковой системы предприятия, надо быть администратором поисковой системы предприятия.

Порядок действий

Чтобы связать словари синонимов с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Поиск выберите **Конфигурировать словари синонимов**.
4. На странице Конфигурировать словари синонимов нажмите кнопку **Добавить словарь синонимов**.
5. На странице Добавить словарь синонимов введите для этого словаря синонимов уникальное имя для вывода и (необязательно) описание.
6. Задайте положение файла .dic. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
7. Нажмите кнопку **ОК**. Пользовательский словарь синонимов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря синонимов с собранием

Если с поисковой системой предприятия связаны словари синонимов, можно выбрать один из них для использования при поиске в собрании. Если условие запроса совпадает с термином в этом словаре, в результатах поиска будут возвращаться также документы, содержащие синонимы.

Прежде, чем вы начнете

Чтобы выбрать словарь синонимов для собрания, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Порядок действий

Чтобы связать словарь синонимов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.

2. В поле **Имя словаря синонимов** на странице Опции сервера поиска выберите словарь синонимов, который надо использовать при запросах пользователей к данному собранию.

Список доступных словарей синонимов включает в себя все словари синонимов, добавленные в поисковую систему предприятия.

3. Нажмите кнопку **ОК**.

Пользовательские словари стоп-слов

Для повышения качества результатов поиска можно указать, что определенные слова будут автоматически удаляться из условий запроса во время его обработки.

Словарь стоп-слов содержит специфичные для предприятия слова, которые используются часто и поэтому бесполезны в качестве условий поиска. Исключая эти слова из запросов, можно добиться того, что пользователи не будут завалены найденными документами с очень малой релевантностью (возвращены будут только документы, соответствующие остальным терминам запроса). Во время обработки запросов серверы поиска удаляют стоп-слова из запросов. Удаляются стоп-слова из вашего пользовательского словаря и стоп-слова, предварительно определенные для поисковой системы предприятия (такие как обычные предлоги и артикли).

WebSphere II OmniFind Edition выполняет распознавание специфичных для разных языков стоп-слов по умолчанию. Этот процесс удаляет из запроса обычные часто встречающиеся слова, такие как *a* и *the*. В ваш пользовательский словарь стоп-слов надо вносить только стоп-слова, специфичные для предприятия или домена.

Для создания словаря стоп-слов специалист по предметной области, для которой предназначено собрание, должен создать список стоп-слов в формате XML (или попросить разработчика прикладной программы, чтобы он создал файл XML). Для преобразования этого файла XML в двоичный файл (.dic) надо использовать инструмент, поставляемый вместе с WebSphere II OmniFind Edition.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь стоп-слов для поиска документов в собрании, когда конфигурируют опции серверов поиска для собрания.

Ограничение: После добавления систему пользовательский словарь стоп-слов нельзя редактировать. Чтобы изменить стоп-слова, доступные при обработке запроса, необходимо:

1. Изменить исходный файл XML.
2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить старый словарь стоп-слов из собраний, в которых он используется.
4. Удалить старый словарь стоп-слов из системы.
5. Добавить в систему новый словарь стоп-слов.
6. Связать новый словарь стоп-слов с собраниями, которые будут его использовать.

Добавление словарей стоп-слов в систему


Если вы создаете пользовательские словари стоп-слов для удаления слов из запросов, необходимо связать эти словари с поисковой системой предприятия. После этого можно выбрать, какой из словарей стоп-слов вы хотите использовать для поиска в собрании.

Прежде, чем вы начнете

Чтобы добавлять пользовательские словари стоп-слов в систему, надо быть администратором поисковой системы предприятия.

Порядок действий

Чтобы связать пользовательские стоп-слова с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Поиск нажмите кнопку **Конфигурировать словари стоп-слов**.
4. На странице Конфигурировать словари стоп-слов нажмите кнопку **Добавить словарь стоп-слов**.
5. На странице Добавить словарь стоп-слов введите для этого словаря уникальное имя для вывода.
6. Задайте положение файла .dic. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
7. Нажмите кнопку **ОК**. Пользовательский словарь стоп-слов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря стоп-слов с собранием

Если с поисковой системой предприятия связаны словари стоп-слов, можно выбрать один из них для использования при поиске в собрании. Если условие запроса совпадает с термином в этом словаре, такой термин удаляется из запроса перед его обработкой.

Прежде, чем вы начнете

Чтобы выбрать словарь стоп-слов для собрания, надо быть администратором поисковой системы предприятия или администратором этого собрания.

Порядок действий

Чтобы связать словарь стоп-слов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. В поле **Имя словаря стоп-слов** на странице Опции сервера поиска выберите словарь стоп-слов, который вы хотите использовать при запросах пользователей к данному собранию.
Список доступных словарей включает в себя все словари стоп-слов, добавленные в поисковую систему предприятия.
3. Нажмите кнопку **ОК**.

Динамическое составление сводок

Динамическое составление сводок - технология, которая определяет, какие словосочетания в результирующем документе наиболее полно соответствуют понятиям, поиск которых выполняет пользователь.

Для поисковой системы предприятия динамическое составление сводок пытается найти в документе предложения, содержащие много поисковых терминов. Выбранные несколько предложений или частей предложений показываются в результатах поиска. Поисковые термины при выводе результатов поиска в формате HTML выделяются.

При конфигурировании опций сервера поиска для собрания можно задать максимальную длину для вывода сводок документов в результатах поиска. Поскольку сводка содержит символы выделения, буфер, возвращаемый прикладной программе поиска, будет больше заданного максимума. Однако число выводимых символов не будет превышать заданного максимума, хотя сводка может оказаться и короче (в зависимости от сводных данных, извлеченных из исходного документа).

Настройка сводок документов на консоли администратора

Можно настроить объем информации, выводимой в сводках документов, задав опции для поискового сервера с консоли администратора поисковой системы предприятия.

Прежде, чем вы начнете

Задать размер выводимой информации сводок собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Значение, заданное для максимального размера выводимой информации сводок документов, действует совместно со значением, заданным для числа предложений, которое может содержать сводка каждого документа. Применяется то ограничение, которое дает меньшую длину сводки документа.

Например, если задать, что сводка может содержать не более четырех предложений, в сводке будет только четыре предложения, даже если максимальная длина сводки в символах больше общего числа символов в этих предложениях. Другой пример - предельное значение из 10 предложений в сочетании с пределом длины для вывода в 500 символов может привести к получению сводки документа, содержащей менее 10 предложений.

Порядок действий

Чтобы сконфигурировать длину выводимой информации сводок документов:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. На странице Опции сервера поиска задайте максимальную длину выводимой информации сводок документов. При просмотре результатов поиска пользователи будут видеть сводки, длина которых не превышает заданного значения.
3. Задайте, сколько предложений может содержаться в сводке каждого документа (сводки могут содержать до 10 предложений).

4. Нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, запустите мониторинг серверов поиска и перезапустите процессы сервера.

Настройка сводок документов при помощи редактирования свойств

Для каждого документа в результатах поиска для запроса поисковой системы предприятия выводится сводка. Отредактировав файл свойств, можно настроить объем информации, выводимой в каждой из этих сводок. by editing a properties file.

Об этой задаче

Можно настроить описания в результатах поиска, изменив значения следующих свойств в файле `ES_NODE_ROOT/master_config/ID_собрания.runtime.node1/runtime-generic.properties`:

MinWordsPerSentence

Минимальное число слов в предложении описания. Значение по умолчанию - 4.

MaxWordsPerSentence

Максимальное число слов в предложении описания. Значение по умолчанию - 20.

NumberOfReturnedSentences

Число предложений в описании документа. Значение по умолчанию - 5.

MaxSentencesPerDocument

Максимальное число предложений в документе, которые при создании описания рассматриваются как кандидаты для внесения в описание. Значение по умолчанию - 1000.

Порядок действий

Чтобы настроить сводки документов в результатах поиска:

1. Зарегистрируйтесь на серверах поиска в качестве администратора поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. При помощи текстового редактора отредактируйте следующий файл (*ID_собрания* - это ID, заданный для собрания или назначенный системой при создании собрания):

```
ES_NODE_ROOT/master_config/ID_собрания.runtime.node1/runtime-generic.properties
```

Совет: Чтобы узнать, какой ID у собрания с данным именем, посмотрите файл `ES_NODE_ROOT/master_config/collections.ini`.

3. Измените свойства, которые хотите настроить, затем сохраните файл и закройте редактор.
4. Остановите и перезапустите серверы поиска, чтобы изменения вступили в силу.

Работа с прямыми ссылками

Прямые ссылки - это документы, возвращаемые в результатах поиска, когда пользователь вводит запрос, включающий в себя определенные слова и словосочетания.

Для конфигурирования прямых ссылок для собрания используется консоль администратора поисковой системы предприятия.

Прямые ссылки

Прямые ссылки позволяют выдавать пользователям ссылки на те документы, которые заранее определены как соответствующие данным терминам запроса.

Прямая ссылка - это URI, которое поисковая система предприятия автоматически включает в результаты поиска, если запрос содержит определенные слова или фразы. Обычно URI прямых ссылок появляются в верхней части списка результатов, чтобы пользователи увидели документы, которые вы заранее определили как соответствующие данному запросу.

Прямые ссылки связи возвращаются в дополнение к другим результатам поиска. Процессы поиска ищут в индексе документы, удовлетворяющие условиям запроса, и возвращают URI для этих документов вместе с URI прямых ссылок.

При конфигурировании прямой ссылки можно указать для URI содержательный заголовок и резюме, чтобы пользователи могли быстро определить, тот ли это, который они хотели получить.

Например, для URI <http://www.ibm.com/education/us/> можно использовать заголовок Обучение IBM в США, и резюме Решения, продукты и ресурсы для профессионалов, обучающего персонала и студентов в США.

Чтобы можно было использовать прямые ссылки в собраниях поисковой системы предприятия, в программе поиска должна быть задана опция вывода прямых ссылок. В некоторых поисковых программах у пользователей может быть возможность включать и отключать возвращение прямых ссылок при поиске в собрании.

Конфигурирование прямых ссылок

Чтобы создать прямую ссылку для собрания поисковой системы предприятия, свяжите URI документа с ключевыми словами, для которых этот документ будет включаться в результаты поиска.

Прежде, чем вы начнете

Для конфигурирования прямых ссылок надо быть администратором поисковой системы предприятия или администратором собрания, к которому принадлежит эта прямая ссылка.

Об этой задаче

Чтобы увидеть примеры задания ключевых слов и URI для прямых ссылок, при создании или редактировании прямой ссылки выберите **Справка**.

Чтобы изменения вступили в силу, перезапуск серверов поиска не требуется.

Порядок действий

Чтобы сконфигурировать прямую ссылку:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать прямые ссылки**.
2. На странице Прямые ссылки выберите **Создать прямую ссылку**.

3. Задайте ключевые слова и фразы, для которых в результатах поиска будет возвращаться эта прямая ссылка, URI документа, который вы заранее определяете как документ для такого запроса, и другие опции для этой прямой ссылки.

Можно задать одно ключевое слово, несколько ключевых слов или одну фразу (два или более слов в кавычках) на строке. Разделяйте ключевые слова пробелами (не запятыми!). Чтобы начать новую строку, нажмите клавишу Enter.

4. Нажмите кнопку **OK**.

Новая прямая ссылка появится в списке на странице Прямые ссылки вместе с другими прямыми ссылками для этого собрания.

Ранжирование документов в поисковой системе предприятия

Когда пользователь выполняет поиск в собрании, процессы поиска возвращают наиболее релевантные результаты для терминов и условий запроса.

Серверы поиска поддерживают сложный синтаксис запросов и используют несколько способов получения наиболее релевантных результатов поиска. Можно дополнить поведение ранжирования по умолчанию, сконфигурировав опции, управляющие важностью документов в результатах поиска:

- Можно создать пользовательские словари значимых слов для управления ранжированием документов, содержащих заданные слова, в результатах поиска.
- Можно управлять оценками документов, соответствующих заданному шаблону URI.
- Можно управлять оценками документов с полями, отображенными на классы значимости.

Назначение оценок на основе текста

Поисковая система предприятия динамически вычисляет балл для каждого документа, соответствующего условиям запроса.

Для вычисления оценки текста каждого соответствующего запросу документа поисковая система предприятия учитывает много факторов, таких как:

- Частота каждого термина запроса во всем собрании. Обычно термины, которые встречаются в большинстве документов, дают меньший вклад в балл документа, чем термины запроса, реже встречающиеся в документах.
- Число появлений каждого термина запроса в данном документе. Как правило, чем больше вхождений термина запроса в документе, тем выше будет оценка этого документа.
- Близость расположения терминов запроса в документе. Как правило, термины запроса, расположенные в документе близко друг от друга, дают больший вклад в балл документа, чем те же самые термины, расположенные дальше друг от друга.
- Контекст, в котором появляются термины запроса в документе. Например, если термин появляется в заголовке документа, он дает больший вклад в оценку документа, чем такой же термин в тексте документа.

Длина документа и разнообразие слов в документе тоже учитываются при определении его оценки.

Статическое ранжирование

Для некоторых типов документов можно назначить фактор статического ранжирования, увеличивающий важность этих документов относительно других результатов поиска.

При создании собрания нужно указать, назначать ли документам из этого собрания фактор статического ранжирования. Для содержимого Web число ссылок на документ из других документов и сами документы, ссылающиеся на этот документ, могут увеличить релевантность документа относительно других результатов поиска.

Для документов, содержащих поля или метаданные с датами, можно увеличить релевантность, используя дату документа. Например, недавние статьи в группах новостей NNTP могут быть более релевантны, чем старые. Если источник данных содержит несколько полей с датами, можно выбрать, какое из них лучше всего определяет релевантность документов в источнике данных.

Если в собрании используется статическое ранжирование, не следует смешивать в одном собрании источники данных, использующие разные типы ранжирования. Например, для использования в качестве фактора статического ранжирования ссылок на документ необходимо, чтобы собрание содержало только документы Web. Наличие в одном собрании документов с разными моделями ранжирования может ухудшить качество поиска.

Кроме того, необходимо, чтобы в документах были поля и значения, позволяющие применить статическое ранжирование. Например, если в качестве фактора ранжирования документов в собрании задана дата документа, а в документах нет полей или атрибутов с датами, качество поиска может ухудшиться.

Пользовательские словари значимых слов

Чтобы улучшить качество результатов поиска, можно изменить порядок ранжирования документов в результатах поиска, создав пользовательский словарь значимых слов.

Если в запросе задано слово, включенное в словарь значимых слов, важность документов, содержащих это слово, возрастет или уменьшится в зависимости от коэффициента значимости, назначенного в словаре этому слову.

Коэффициенты значимости принимают значения от -10 до 10. При обработке запросов серверы поиска рассматривают как более важные документы с положительными коэффициентами значимости и как менее важные - документы с отрицательными коэффициентами значимости.

Например, документ, соответствующий терминам запроса с высокими коэффициентами значимости, получает более высокий ранг, чем без использования коэффициентов значимости. (Вклад в ранжирование документа в результатах поиска вносят только термины запроса с указанной значимостью; коэффициент значимости - единственная величина, влияющая на ранг документа.)

При создании словаря можно задать для любого числа слов одинаковые коэффициенты значимости. В словарь можно включать термины из одного или нескольких слов (термины из нескольких слов выступают в поиске в качестве словосочетаний).

Если слово с весовым вкладом, заданным через коэффициент значимости, входит в запрос с операцией OR (например: этот | тот), для терминов этого запроса вычисляется взвешенное среднее. Полученная в итоге сводная оценка используется для всех вхождений операндов запроса OR (частные оценки для отдельных операндов запроса OR не вычисляются).

Оценка значимости, основанная на словарях значимых слов, не поддерживается для терминов поиска по полям. При анализе терминов запроса для численной оценки важности документа используется только текст запроса, но не имя поля. (Чтобы применить коэффициенты значимости к терминам запроса в полях, надо отобразить имена полей на классы значимости.)

Для создания словаря значимых слов специалист по предметной области, для которой предназначено собрание, должен создать список значимых слов в формате XML (или попросить разработчика прикладной программы, чтобы он создал файл XML). Для преобразования этого файла XML в двоичный файл (.dic) надо использовать инструмент, поставляемый вместе с WebSphere II OmniFind Edition.

Администратор поисковой системы предприятия должен загрузить этот двоичный файл в систему и назначить ему имя для вывода. Администраторы собраний могут выбрать словарь значимых слов, используемый для поиска документов в собрании, когда конфигурируют опции поискового сервера для собрания.

Ограничение: Добавленный в систему пользовательский словарь значимых слов нельзя редактировать. Чтобы исправить значимые слова, доступные для обработки запросов, надо:

1. Изменить исходный файл XML.
2. Преобразовать этот исходный файл XML в новый файл словаря.
3. Удалить словарь значимых слов из собраний, в которых он используется.
4. Удалить старый словарь значимых слов из системы.
5. Добавить в систему новый словарь значимых слов.
6. Связать новый словарь значимых слов с собраниями, в которых он должен использоваться.

Добавление словарей значимых слов в систему


Если вы создаете пользовательские словари значимых слов, необходимо связать эти словари с поисковой системой предприятия. После этого можно выбрать, какой из словарей значимых слов использовать для поиска в собрании.

Прежде, чем вы начнете

Добавлять пользовательские словари значимых слов в систему может администратор поисковой системы предприятия.

Порядок действий

Чтобы связать пользовательские словари значимых слов с поисковой системой предприятия:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Поиск выберите **Конфигурировать словари значимых слов**.
4. На странице Конфигурировать словари значимых слов нажмите кнопку **Добавить словарь значимых слов**.
5. На странице Добавить словарь значимых слов введите для этого словаря значимых слов уникальное имя для вывода и (необязательно) описание.
6. Задайте положение файла .dic. Если размер этого файла меньше 8 Мбайт, файл может находиться в вашей системе. Если размер этого файла больше 8 Мбайт, файл должен находиться на сервере индексов.
7. Нажмите кнопку **ОК**. Пользовательский словарь значимых слов будет добавлен в поисковую систему предприятия и станет доступным при поиске в собраниях.

Связывание словаря значимых слов с собранием

Если с поисковой системой предприятия связаны словари значимых слов, можно выбрать один из них для использования при поиске в собрании. Если термин запроса совпадает с термином в словаре, важность документов, содержащих этот термин, возрастет или уменьшится, в зависимости от коэффициента значимости, назначенного в словаре этому термину.

Прежде, чем вы начнете

Выбрать словарь значимых слов для собрания может администратор поисковой системы предприятия или администратор этого собрания.

Порядок действий

Чтобы связать словарь значимых слов с собранием:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать опции сервера поиска**.
2. В поле **Имя словаря значимых слов** на странице Опции сервера поиска выберите словарь значимых слов, который надо использовать при запросах пользователей к данному собранию.
Список доступных словарей включает в себя все словари значимых слов, добавленные в поисковую систему предприятия.
3. Нажмите кнопку **ОК**.

Ранжирование документов на основе шаблонов URI

Можно повысить или понизить важность документов, назначив коэффициенты значимости для шаблонов URI.

Всем документам при их добавлении в индекс назначается оценка статического ранжирования по умолчанию. Эта оценка зависит от того, включено ли для собрания статическое ранжирование, и, если это так, от типа статического ранжирования - по дате документов или (для документов Web) по числу ссылок на данный документ из других документов.

Относительной важностью документов можно управлять, назначая коэффициенты значимости для шаблонов URI. Коэффициент значимости используется с оценкой статического ранжирования по умолчанию и другими коэффициентами при определении конечной статической оценки документа.

Порядок, в котором сконфигурированы шаблоны URI, имеет значение. Когда сервер индексов вычисляет значение для каждого документа в собрании, он оценивает шаблоны URI в порядке их перечисления в списке. Для каждого URI:

1. Сервер индекса последовательно просматривает шаблоны URI.
2. Когда сервер индекса находит первый шаблон URI, соответствующий документу в индексе, он применяет к этому документу коэффициент значимости, сконфигурированный для данного URI.
3. Если для документа не найден соответствующий шаблон URI, используется оценка статического ранжирования по умолчанию.

Сконфигурировав коэффициент значимости для шаблона URI, разместите его в соответствии с требуемым порядком просмотра шаблонов сервером индексов.

Влияние на оценки документов, соответствующих шаблонам URI

Можно повысить или понизить важность документов, соответствующих шаблону URI, применив коэффициент значимости к оценке статического ранжирования по умолчанию.

Прежде, чем вы начнете

Управлять важностью документов, соответствующих URI, может администратор поисковой системы предприятия или администратор этого собрания.


Об этой задаче


Сконфигурированный коэффициент значимости используется совместно с оценкой статического ранжирования при вычислении новой статической оценки для всех документов, соответствующих заданному шаблону URI.

Коэффициенты значимости изменяют только статические оценки, и эти коэффициенты являются единственными участниками вычисления, определяющими конечный ранг документа. Например, если на один документ много ссылок (что выражается в его высокой начальной оценке), то другой документ, на который ссылки отсутствуют, всегда будет иметь более низкий ранг, чем первый.

Порядок действий

Чтобы повлиять на оценки документов, соответствующих заданному шаблону URI:

1. Отредактируйте собрание, откройте страницу Индекс и выберите **Управление оценками на основе соответствия шаблонам URI**.
2. На странице Управление оценками на основе соответствия шаблонам URI выберите **Добавить шаблон URI**.
3. Введите шаблон для документов, важность которых в результатах поиска вы хотите повысить или понизить. Например:
`http://domain.org/ru/*`
`db2://*ROI*`
`*/afs/*`
4. Введите для коэффициента значимости значение от -10 до 10. На основе этого коэффициента значимости будет вычисляться конечная статическая оценка для всех документов, соответствующих шаблону URI.
5. Нажмите кнопку **ОК**.
6. На странице Управление оценками на основе соответствия шаблонам URI разместите новый шаблон URI в соответствии с требуемым порядком его просмотра сервером индексов.
Сервер индексов вычисляет оценки статического ранжирования в порядке перечисления URI. Для достижения лучших результатов в начало списка ставьте более точно заданные URI. В следующем примере подкаталог /forms соответствует шаблону URI `http://www.ibm.com/ru/*`. Чтобы добиться правильного вычисления оценки для документов в подкаталоге /forms, поставьте шаблон URI для подкаталога /forms первым в списке:
`http://www.ibm.com/ru/forms/* 8`
`http://www.ibm.com/ru/* -2`
7. Чтобы изменить шаблон URI или коэффициент значимости, выберите шаблон URI и нажмите кнопку  **Редактировать**.

8. Чтобы удалить шаблон URI из списка, выберите шаблон URI и нажмите кнопку  **Удалить**.
9. Чтобы применить коэффициенты значимости к ранее проиндексированным документам, реорганизуйте индекс.

Ранжирование документов на основе классов значимости

Отображая поля на классы значимости, можно влиять на порядок ранжирования документов в результатах поиска.

При анализе документов анализатор назначает *классы значимости* маркерам документов в соответствии с полями, к которым принадлежат эти маркеры. Эти классы значимости включаются в индекс и используются для вычисления оценок, оказывающих влияние на ранжирование документов результатов.

Чтобы воздействовать на то, как вычисляются оценки, можно сконфигурировать численные коэффициенты значимости для классов значимости. Если термин запроса совпадает с маркером в поле, отображаемом на класс значимости, вклад этого вхождения маркера будет влиять на итоговую оценку значимости документа. В вычислении этой оценки участвует коэффициент значимости, сконфигурированный в классе значимости.

Пусть, например, вам нужно оценить значимость полей заголовка. Если термин запроса встречается в заголовке, это вхождение будет вносить большой вклад в оценку значимости документа и повысит его ранг в результатах поиска.

Изменить ранжирование документа можно при помощи консоли управления поисковой системой предприятия, задав коэффициенты значимости для классов значимости и отобразив поля на классы значимости. Для поисковой системы предприятия предварительно сконфигурировано шестнадцать классов значимости. Восемь из этих классов значимости предназначены для использования с полями содержимого, а другие восемь - для полей метаданных. Коэффициенты, связанные с классами значимости по умолчанию, можно изменить, кроме того, можно связать с классами значимости другие поля или дополнительные поля.

Если вы изменили отображения полей, надо снова обработать документы искателем и проанализировать их, чтобы внесенные вами изменения можно было применить к ранее проиндексированным документам. Если вы изменили коэффициенты, заданные для классов значимости, включите слежение за серверами поиска, остановите и перезапустите процессы на серверах поиска, чтобы эти изменения вступили в силу.

Обнаружение дубликатов документов

При отображении поля на класс значимости надо указать, используется ли это поле для обнаружения дубликатов документов:

- Если поле используется для поиска дубликатов документов, это поле считается полем содержимого, и для выбора остаются только классы значимости, предназначенные для полей содержимого.
- Если поле не используется для поиска дубликатов документов, это поле считается полем метаданных, и для выбора остаются только классы значимости, предназначенные для полей метаданных. В этом случае два документа, идентичные во всех отношениях, кроме указанного поля, считаются дубликатами друг друга.

Высокие и низкие значения возврата

Во время обработки запроса процессы поиска оценивают число возвращаемых документов. Чтобы определить, характеризуется ли запрос высоким или низким значением возврата, задаются пороговые значения:

Низкий возврат

Если оцениваемое число полученных документов ниже нижнего порогового значения, запрос считается обладающим низким значением возврата.

Высокий возврат

Если оцениваемое число полученных документов выше верхнего порогового значения, запрос считается обладающим высоким значением возврата.

Промежуточное значение

Если оцениваемое число документов находится между этими двумя пороговыми значениями, значение возврата считается средним.

Для каждого класса значимости задаются коэффициенты значимости, которые при обработке запросов связываются с высокими и низкими значениями возврата. Нижний коэффициент значимости влияет на относительную важность запросов с низкими значениями возврата, а верхний коэффициент значимости - на относительную важность запросов с высокими значениями возврата. Сочетание этих двух коэффициентов влияет на относительную важность запросов со средним значением возврата.

Значения коэффициентов значимости управляют относительной важностью каждого из вхождений термина запроса в документе. Каждое найденное в документе вхождение термина запроса считается с учетом соответствующего коэффициента значимости.

При конфигурировании классов значимости для собрания можно отредактировать коэффициенты значимости по умолчанию. Например, можно задать такие коэффициенты значимости, что термины запроса, найденные в полях заголовков, будут учитываться с пятикратным преимуществом по сравнению с обычным текстом.

Отображение полей на классы значимости

Можно изменять относительную важность полей, отображая имена полей на классы значимости.

Прежде, чем вы начнете

Отображать поля на классы значимости может администратор поисковой системы предприятия или администратор данного собрания.



Об этой задаче

Через коэффициенты значимости система влияет на ранжирование документов, содержащих термины запроса в полях, отображенных на классы значимости.

В поисковой системе предприятия некоторые имена полей зарезервированы для вычисления оценок значимости обычных полей и текста (например, полей, не являющихся полями заголовка, и обычного невыделенного текста). Вы можете отображать другие поля на классы значимости, которые используют эти зарезервированные поля, но сами зарезервированные поля редактировать или удалять нельзя.

Порядок действий

Чтобы отобразить поля на классы значимости:

1. Вызовите редактирование собрания, перейдите на страницу Анализ и нажмите кнопку **Отобразить поля на классы значимости**.
2. На странице Отобразить поля на классы значимости нажмите кнопку **Добавить поле**.
3. На странице Добавить поле в класс значимости введите имя поля, которое нужно отобразить на класс значимости.
Вы можете указать имя поля, существующего в источнике, обработанном искателем, или во внешнем источнике, имя поля, отображенного из элемента XML, имя поля, отображенного из элемента метаданных HTML, или же одно из заранее определенных имен полей.
4. Укажите, используется ли это поле для обнаружения дубликатов документов. Если этот переключатель включен, список доступных классов значимости будет содержать классы, относящиеся к полям содержимого. Если этот переключатель выключен, список доступных классов значимости будет содержать классы, относящиеся к полям метаданных.
5. Выберите класс значимости и нажмите кнопку **ОК**.
Добавленное вами поле будет выведено на странице Отобразить поля на классы значимости. Вы можете выбрать опцию редактирования классов значимости и задать другие коэффициенты значимости для оценивания документов, содержащих это поле.
6. Чтобы задать, используется ли поле для обнаружения дубликатов документов или отобразить это поле на другой класс значимости, выберите  **Редактировать**. (Редактировать поля, зарезервированные поисковой системой предприятия, нельзя.)
7. Чтобы удалить поле из класса значимости, щелкните по  **Удалить**. (Удалять поля, зарезервированные поисковой системой предприятия, нельзя.)
8. Чтобы применить изменения к ранее индексированным документам, повторно обработайте их искателем, проанализируйте и проиндексируйте их.

Конфигурирование коэффициентов значимости для классов значимости

Коэффициенты значимости, сконфигурированные вами для классов значимости, отражают вашу оценку вклада конкретных полей в найденных документах в выполнение запроса. Классы значимости с высокими коэффициентами значимости могут повысить важность найденных документов с полями, отображаемыми на эти классы значимости.

Прежде, чем вы начнете


Конфигурировать коэффициенты значимости для классов значимости может администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Для перерасчета важности найденных документов с полями, отображаемыми на классы значимости, в системе используются коэффициенты значимости, сконфигурированные для класса значимости, коэффициент статического ранжирования по умолчанию и другие коэффициенты.

Порядок действий

Чтобы сконфигурировать коэффициенты значимости для классов значимости:

1. Вызовите редактирование собрания, перейдите на страницу Анализ и нажмите кнопку **Отобразить поля на классы значимости**.
2. На странице Отобразить поля на классы значимости нажмите кнопку **Редактировать классы значимости**.
3. На странице Классы значимости найдите класс значимости, который нужно изменить, и выберите  **Редактировать**.
4. На странице Редактировать класс значимости задайте новые значения для верхних и нижних коэффициентов значимости. Для обоих коэффициентов можно ввести одинаковые значения.
5. Нажмите кнопку **ОК**.
6. Чтобы изменения вступили в силу, включите слежение за серверами поиска и выберите значки для остановки и перезапуска процессов поиска. При выполнении пользователями запросов относительная важность найденных документов с полями, отображаемыми на данный класс значимости, будет определяться новыми коэффициентами значимости.

Значения классов значимости по умолчанию

В системе поиска предприятия есть 16 классов значимости, с помощью которых можно влиять на порядок ранжирования документов в результатах поиска.

Для вычисления оценок для полей и текста, у которых нет каких-либо других отличительных характеристик, для использования в поисковой системе предприятия зарезервированы следующие поля:

```
es_special_field.regular_text  
es_special_field.default_field
```

Вы можете отображать другие поля на классы значимости, которые используют эти зарезервированные поля, но сами зарезервированные поля редактировать или удалять нельзя.

Для всех остальных полей коэффициенты значимости, используемые системой для вычисления рангов документов, можно редактировать. Можно также отобразить любое число полей на любые классы значимости, включая те из них, которые используются зарезервированными полями.

Ниже в таблице перечислены имена классов значимости, коэффициенты значимости по умолчанию для запросов с низкими значениями возврата, коэффициенты значимости по умолчанию для запросов с высокими значениями возврата и имена заранее определенных полей, отображаемых на классы значимости в конфигурации по умолчанию.

Коэффициенты значимости по умолчанию изменяются в соответствии с методом статического ранжирования, выбранным для собрания во время создания данного собрания. При этом возможны опции без статического ранжирования, с ранжированием по числу ссылок на документ (для источников Web) или с ранжированием по датам документов.

Таблица 3. Значения классов значимости по умолчанию

Нижние и верхние коэффициенты значимости по умолчанию				
Имя класса значимости	Без статического ранжирования	По ссылкам на документы	По датам документов	Предопределенные для отображения поля
Класс содержимого А	Нижний: 4 Верхний: 2	Нижний: 6 Верхний: 1	Нижний: 4 Верхний: 2	es_special_field.regular_text
Класс содержимого В	Нижний: 5 Верхний: 4	Нижний: 7 Верхний: 3	Нижний: 5 Верхний: 4	es_special_field.html_emphasized_text Включает элементы HTML: b, big, caption, dfn, em, h4, h5, h6, strong
Класс содержимого С	Нижний: 7 Верхний: 4	Нижний: 9 Верхний: 3	Нижний: 7 Верхний: 4	es_special_field.html_headers Включает элементы HTML: h1, h2, h3
Класс содержимого D	Нижний: 2 Верхний: 5	Нижний: 1 Верхний: 5	Нижний: 2 Верхний: 5	title
Класс содержимого E	Нижний: 1 Верхний: 1	Нижний: 5 Верхний: 10	Нижний: 1 Верхний: 1	es_special_field.anchor
Класс содержимого F	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.anchor_same_dir
Класс содержимого G	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.anchor_same_host
Класс содержимого H	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных А	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.default_field
Класс метаданных В	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных С	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных D	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	
Класс метаданных E	Нижний: 1 Верхний: 1	Нижний: 5 Верхний: 1	Нижний: 1 Верхний: 1	keywords
Класс метаданных F	Нижний: 1 Верхний: 1	Нижний: 3 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.urlhost
Класс метаданных G	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	es_special_field.urlpath
Класс метаданных H	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	Нижний: 1 Верхний: 1	описание

Прикладные программы поиска для поисковой системы предприятия

Прикладная программа поиска позволяет выполнять поиск в собраниях и внешних источниках поисковой системы предприятия. Можно создать любое число прикладных программ поиска, каждая из которых будет выполнять поиск в любом числе собраний и внешних источников.

Пример прикладной программы поиска

Этот пример программы поиска демонстрирует многие функции поиска и получения, доступные для поисковой системы предприятия. Кроме того, пример прикладной программы поиска - работающий пример, иллюстрирующий использование API поиска и индексации IBM (SI-API, Search and Index API) для построения пользовательских интерактивных прикладных программ поиска, приспособленных для задач вашего предприятия.

Если не изменить свойства в файле конфигурации по умолчанию, пример прикладной программы поиска позволяет выполнять поиск во всех активных собраниях и внешних источниках в поисковой системе предприятия. Пример прикладной программы поиска можно использовать для тестирования новых собраний и внешних источников перед тем, как делать их доступными для пользователей.

Пример программы поиска автоматически связывается со всеми собраниями и внешними источниками. В производственной среде администраторы поисковой системы предприятия решают, каким программам поиска разрешено выполнять поиск в тех или иных собраниях.

Пользовательские прикладные программы поиска

Созданные вами прикладные программы поиска можно выполнять как самостоятельные прикладные программы Web в среде IBM WebSphere Application Server или как портлеты в среде IBM WebSphere Portal. При помощи API поиска и индексации можно разработать прикладные программы поиска, которые, как и пример программы поиска, будут одинаково работать в обеих средах.

Задание прикладных программ поиска для собраний

Чтобы использовать новую прикладную программу поиска, сначала надо связать ее с собраниями, в которых она может выполнять поиск.

Прежде, чем вы начнете

Задать для прикладных программ поиска собрания, в которых они могут выполнять поиск, может пользователь с ролью администратора поисковой системы предприятия.

Порядок действий

Чтобы связать прикладную программу поиска с одним или несколькими собраниями:

1. Выберите **Защита** на панели инструментов консоли администратора.

2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать прикладные программы поиска**.
3. На странице Конфигурировать прикладные программы поиска нажмите кнопку **Добавить прикладную программу поиска**.
4. Введите имя прикладной программы поиска.
5. Выберите собрания, в которых прикладная программа поиска сможет выполнять поиск:
 - Нажмите кнопку **Все собрания и внешние источники**, если вы хотите, чтобы прикладная программа поиска работала со всеми собраниями, которые вы добавили в систему.
 - Чтобы разрешить программе поиска доступ только к указанным собраниям, выберите **Отдельные собрания и внешние источники**.
Если выбрана эта опция, появляется список имен собраний и имен внешних источников. Включите переключатели **Выбрать** для всех собраний, в которых может выполнять поиск эта прикладная программа.
6. Нажмите кнопку **ОК**.

Функции примера программы поиска

Пример программы поиска для поисковой системы предприятия иллюстрирует большинство функций поиска, которые можно встраивать в пользовательские программы поиска.

Пример программы поиска можно использовать для поиска по одному, нескольким или всем собраниям и внешним источникам одновременно. Если не изменены свойства программы по умолчанию, эту программу можно использовать для поиска во всех собраниях и внешних источниках в поисковой системе предприятия.

Функции запроса

Эти функции позволяют:

- Задавать простые запросы в виде свободного текста.
- Задавать более сложные запросы для улучшения точности результатов поиска. Например, можно выполнить поиск по определенным полям документа или при помощи синтаксиса запроса найти документы, содержащие или не содержащие конкретные слова и словосочетания.
- Задавать, в каких собраниях и внешних источниках выполнять поиск.
- Выполнять поиск в источниках конкретных типов или же всех типов.
- Выполнять поиск в документах определенных типов. Например, можно выполнить поиск только в документах Microsoft Word или только в документах PDF.
- Указывать, на каком языке заданы термины запроса. Можно также задать языки документов, в которых будет выполняться поиск.
- Выполнять поиск в определенных подмножествах собрания. Например, программа поиска может ограничить видимость до заранее определенного диапазона документов (области действия), или передать запрос, выполняющий поиск только в документах названной категории.
- Расширять запрос, добавляя синонимы терминов запроса. Если с собранием связан словарь синонимов, документы, содержащие синонимы терминов запроса, возвращаются в результатах поиска.

Функции результатов поиска

Эти функции позволяют:

- Посмотреть результаты поиска, соответствующие вашим терминам запроса.
- Задавать число найденных документов на каждой странице и прокручивать набор результатов вперед и назад.
- Скрывать или выводить подробности о документах результатов. Например, можно просмотреть краткие описания документов или такие подробности, как имена полей в каждом документе результатов.
- Объединять документы из одного источника. Например, если из одного источника возвращено 100 документов, в наборе результатов выводится группа из двух максимально соответствующих документов. Чтобы просмотреть остальные 98 документов, надо выбрать опцию для просмотра большего числа документов из того же источника.
- Сортировать документы по релевантности, по дате документа в порядке возрастания или по дате документа в порядке убывания.
- Выбирать из вариантов исправлений орфографии в случае обнаружения в строке запроса слов с возможными ошибками.
- Просматривать информацию о категориях, к которым принадлежит найденный документ (если собрание использует категории), и просматривать только документы, принадлежащие к определенной категории.
- Задавать дополнительные термины запросов в результатах поиска.

Функции получения документов

Эти функции позволяют:

- Получать документ, щелкнув по URI документа. Кроме того, если доступны клиентские программы, документы результатов, скорее всего, можно будет просматривать в собственной программе просмотра. Например, если установлена программа клиента Notes, с ее помощью можно просматривать документы из базы данных Notes.
Если сконфигурирована защита уровня документов для искателя, получать документы могут только пользователи с правами доступа к защищенному содержимому.
- Получать документ, щелкнув по прямой ссылке. Прямая ссылка связывает ключевые слова с определенными URI. Если запрос содержит заданные ключевые слова, соответствующие им URI (их высокая релевантность для этих ключевых слов определена заранее) выводятся в начале списка результатов поиска.

Свойства примера программы поиска

Чтобы задать опции для вашей среды, изменить внешний вид программы и выбрать опции, доступные пользователям после ее запуска, можно отредактировать файл `config.properties` для примера программы поиска.

Параметры среды

Вы можете задать опции, управляющие работой программы поиска.

applicationName

Задает имя допустимой программы поиска. Значение по умолчанию - `Default`.

Измените это значение, если хотите в качестве программы по умолчанию использовать другую программу.

Совет: Оставив имя программы Default, можно использовать программу примера поиска, выполняющую поиск во всех собраниях и внешних источниках при помощи одного запроса.

hostname

Задаёт полное имя хоста для Web-сервера, сконфигурированного для поддержки вашего экземпляра WebSphere Application Server. По умолчанию используется значение localhost.

Чтобы программа поиска работала правильно, измените значение по умолчанию на полное имя хоста, которое задано в конфигурации WebSphere Application Server. Например, если имя хоста компьютера - MyMachine, а Web-сервер сконфигурирован использовать в качестве имени хоста этого компьютера www.ibm.com, задайте www.ibm.com.

port Задаёт номер порта Web-сервера, который сконфигурирован для поддержки вашего экземпляра WebSphere Application Server. Значение по умолчанию - 80. Изменяйте значение по умолчанию только в том случае, если вы изменяете порт для вашего экземпляра Web-сервера.

timeout

Задаёт срок ожидания ответа от поискового сервера в секундах для требования поиска. Надо ввести целое число (например, 60, но не 60,5 и не "шестьдесят"). Если срок ожидания не задан, по умолчанию используется значение 30 секунд.

username

Задаёт имя пользователя, позволяющее поисковой системе предприятия проводить аутентификацию для пользователей с WebSphere Application Server, когда они отправляют требования поиска. Это поле используется, только если в WebSphere Application Server включена глобальная защита.

password

Задаёт пароль для заданного имени пользователя.

filter Задаёт класс для получения документов, перечисленных в списке результатов поиска. Класс по умолчанию - com.ibm.es.api.filters.SetDocumentURIFilterFetch. Это значение надо изменять, только если у вас есть пользовательский класс, который надо использовать для получения документов вместо класса по умолчанию.

logging.level

Задаёт степень подробности журнала:

OFF В журнал не записывается никаких сообщений.

INFO В журнал записываются информационные сообщения.

FINE В журнал записываются сообщения об ошибках.

ALL В журнал записываются все сообщения.

Значки типов источников

Можно настроить изображения для типов источников данных, к которым принадлежат документы - результаты поиска. Перечисленные ниже значки типов источников (соответствующие искателям и внешним источникам, поддерживаемым WebSphere Information Integrator OmniFind Edition после установки) предопределены в файле config.properties.

**documentSource.vbr.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Content Edition. Значок по умолчанию - /images/sourceVBR.gif.

**documentSource.db2.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем DB2. Значок по умолчанию - /images/sourceDB2.gif.

**documentSource.cm.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем DB2 Content Manager. Значок по умолчанию - /images/sourceCM.gif.

**documentSource.dominodoc.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Domino Document Manager. Значок по умолчанию - /images/sourceDominoDoc.gif.

**documentSource.exchange.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Exchange Server. Значок по умолчанию - /images/sourceExchange.gif.

**documentSource.nntp.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем NNTP. Значок по умолчанию - /images/sourceNNTP.gif.

**documentSource.notes.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Notes. Значок по умолчанию - /images/sourceNotes.gif.

**documentSource.quickplace.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем QuickPlace. Значок по умолчанию - /images/sourceWorkplace.gif.

**documentSource.unixfs.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем для файловых систем UNIX. Значок по умолчанию - /images/sourceUnixFS.gif.

**documentSource.web.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем Web. Значок по умолчанию - /images/sourceWeb.gif.

**documentSource.wps.icon**

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем WebSphere Portal. Значок по умолчанию - /images/sourceWPS.gif.



documentSource.winfs.icon

Задает путь и имя файла изображения, означающего, что документ просмотрен искателем для файловых систем Windows. Значок по умолчанию - /images/sourceWindowsFS.gif.



documentSource.ldap.icon

Задает путь и имя файла изображения, означающего, что документ относится к внешнему источнику, созданному для сервера LDAP. Значок по умолчанию - /images/sourceLDAP.gif.



documentSource.jdbc.icon

Задает путь и имя файла изображения, означающего, что документ относится к внешнему источнику - таблице базы данных JDBC (Java Database Connectivity). Значок по умолчанию - /images/sourceJDBC.gif.

Значки клиентских программ просмотра

Документы - результаты поиска могут выводиться в браузере. При наличии клиентской программы, обрабатывающей тип документа - результата поиска, документы этого типа можно выводить также при помощи клиентской программы просмотра. Можно настроить изображения для типов клиентских программ просмотра, используемых для вывода документов. В следующем примере значок Lotus Notes позволяет указать, что документ можно вывести клиентской программой Lotus Notes:

```
client.notes.icon=/images/notes.gif
```

В списке результатов поиска этот значок и ссылка на клиентскую программу выводятся так:



Клиентская программа просмотра

Поля документа

Для типов источников данных, в которых есть поля, можно задать, какие поля выводятся в списке документов - результатов поиска.

fields.URI prefix=список_имен_полей_через_пробел

Перед символом двоеточия (:) в префиксе URI необходимо указывать символ обратной косой черты (\). Чтобы продолжить список имен полей на другой строке, введите в конце предыдущей строки символ обратной косой черты (\). Например:

```
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
```

Значки полей

Для типов источников данных и документов, в которых есть поля, можно настроить изображения, соответствующие полям. В файле config.properties predefinedены следующие значки полей.



field.icon.databasetitle

Задает путь и имя файла изображения, означающего, что поле содержит название документа. Значок по умолчанию - /images/notesdb.gif.



field.icon.databasename

Задает путь и имя файла изображения, означающего, что поле содержит имя базы данных, к которой принадлежит документ. Значок по умолчанию - /images/db2.gif.



field.icon.tablename

Задает путь и имя файла изображения, означающего, что поле содержит имя таблицы, к которой принадлежит документ. Значок по умолчанию - /images/table.gif.



field.icon.directory

Задает путь и имя файла изображения, означающего, что поле содержит имя каталога, к которому принадлежит документ. Значок по умолчанию - /images/closedFolder.gif.

Значок поля по умолчанию

Можно задать изображение, используемое, если для полей, выводимых в списке результатов поиска, не заданы значки полей. В файле config.properties предопределен следующий значок поля по умолчанию.



field.defaultIcon

Задает путь и имя файла изображения, содержащего значок по умолчанию для полей в списке результатов поиска. Значок по умолчанию - /images/database.gif.

Поля дат

Можно задать, какие поля являются полями дат. Данные из полей, имена которых здесь заданы, форматируются в списке результатов поиска как даты. Формат даты соответствует параметрам национальной версии, заданным в браузере.

date.fields=список_имен_полей_через_пробел

Чтобы продолжить список имен полей на другой строке, введите в конце предыдущей строки символ обратной косой черты (\).

Пример:

```
date.fields=modifieddate createddate
```

Названия документов

Вы можете задать альтернативные названия документов путем замены текста названия более информативными данными (то есть названия можно *корректировать*). Например, вместо вывода названий документов с неинформативной меткой Слайд 1 можно задать, что в списке результатов поиска название Слайд 1 должно подавляться. (Для идентификации документа можно использовать более осмысленное поле, например, имя файла.)

Кроме того, можно задать альтернативные названия документов путем удаления неинформативных слов из названий документов (то есть названия можно *сокращать*).

Например, если названия некоторых документов - результатов поиска начинаются с Microsoft Word, можно улучшить читаемость результатов поиска путем подавления повторяющегося текста.

titles.clean=*список_названий_через_запятую*

titles.truncatePrefix=*список_префиксов_через_запятую*

В этих списках через запятую задаются значения, которые могут содержать пробелы и другие символы, кроме запятой. Чтобы продолжить список на другой строке, введите в конце предыдущей строки символ обратной косой черты (\).

Например:

```
titles.clean=Слайд 1, Макет 1, Шаблон презентации IBM Software Group, \
без имени, Документ без имени, Презентация PowerPoint, \
(страница без имени)
```

```
titles.truncatePrefix=Microsoft Word -, Microsoft Powerpoint -
```

Значения по умолчанию для предпочтений пользователя

Можно задать значения по умолчанию для страницы программы поиска Предпочтения. Если пользователь изменяет эти предпочтения, новые значения действуют только в течение текущего сеанса пользователя. В файле config.properties предопределены следующие предпочтения.

preferences.resultsRange=10

Задаёт, что на каждой странице списка результатов поиска может выводиться до 10 документов - результатов поиска.

preferences.siteCollapsing=Yes

Задаёт, что если в консоли администратора поисковой системы предприятия заданы правила сворачивания сайтов, то URI, соответствующие правилам префиксов URI, должны сворачиваться в списке результатов поиска.

preferences.spellCorrections=Yes

Включает вывод вариантов исправления ввода, когда пользователь отправляет запрос с возможной ошибкой в написании.

preferences.extendedHighlighting=No

Задаёт, что в дополнительных полях (например, в названии документа) не должны выделяться термины запроса (кроме поля сводки документа).

Собрания и внешние источники по умолчанию

Можно задать, какие собрания и внешние источники изначально выбраны на страницах Предпочтения и Расширенный поиск. Пользователи могут отредактировать набор по умолчанию, чтобы выполнять поиск не во всех собраниях и внешних источниках, сделанных доступными по умолчанию. Если здесь ограничить набор собраний и внешних источников, пользователи могут выбрать любое собрание или внешний источник, доступные для программы поиска, при изменении предпочтений или расширенных опций поиска.

preferences.defaultCollections=*

preferences.defaultCollections=*список_ID_собраний_через_пробел*

Укажите символ звездочки (*), чтобы разрешить поиск во всех собраниях и внешних источниках. (Собрания и внешние источники надо связать с программой поиска на консоли администратора поисковой системы предприятия.) Это значение по умолчанию, заданное в файле config.properties.

Чтобы ограничить область поиска, доступную пользователям, если они не изменят предпочтения или расширенные опции поиска, укажите ID собраний для собраний и внешних источников, в которых пользователи будут выполнять поиск по умолчанию.

Например:

```
preferences.defaultCollections=*  
preferences.defaultCollections=coll_id1 coll_id2
```

Дополнительная информация для результатов поиска

Можно настроить объем информации, выводимой в списке результатов поиска, а также включить или выключить возможность фильтрации результатов поиска пользователями. По умолчанию в файле `config.properties` заданы следующие значения.

refreshButton.show=false

Включает или выключает вывод кнопки **Обновить** и возможность обновления программы поиска. Если задать для этой опции значение `true`, пользователи смогут обновлять список собраний и внешних источников, доступных для поиска (например, если программа поиска была связана с дополнительными собраниями или внешними источниками на консоли администратора поисковой системы предприятия).

Возможно, вы захотите включить вывод кнопки **Обновить** для проверки изменений, сделанных в файле `config.properties`. Сохранив изменения, можно нажать кнопку **Обновить**, чтобы посмотреть, как эти изменения повлияют на программу поиска. Если кнопка **Обновить** не выводится, надо перезапустить прикладную программу предприятия `ESSearchApplication` в `WebSphere Application Server`, чтобы изменения вступили в силу.

Если для поиска не доступны никакие собрания или внешние источники (например, если указано неправильное имя хоста, не запущены серверы поиска или в `WebSphere Application Server` не запущена прикладная программа поиска `ESSearchServer`), кнопка **Обновить** выводится автоматически, чтобы помочь исправить эту проблему.

extraMessages.show=false

Управляет выводом области в нижней части результатов поиска, где выводятся предупреждения и информационные сообщения. Сообщения об ошибках всегда выводятся в верхней части страницы. Включите эту опцию, чтобы были видны дополнительные сообщения.

builtQueryString.show=false

Управляет выводом полностью развернутого синтаксиса запроса в области перед списком документов - результатов поиска. Включите эту опцию, чтобы был виден фактически обработанный запрос.

refineResults.show=true

Разрешает или запрещает пользователям уточнять результаты поиска путем задания дополнительных терминов запроса. Если задать для этой опции значение `true`, в нижней части страницы результатов поиска выводится окно ввода запроса с меткой **Поиск в результатах**.

filter.showOnTwoLines=true

Управляет выводом опций фильтрации результатов по типу источника и по типу файла на одной или на двух строках в списке результатов поиска. (При просмотре результатов поиска пользователи могут выбрать тип источника и тип файла для вывода только документов - результатов поиска, соответствующих выбранным фильтрам.)

Чтобы максимизировать пространство, доступное для вывода результатов поиска, задайте для этого свойства значение false. Чтобы улучшить читаемость фильтров, в особенности если доступные фильтры занимают больше одной строки, можно задать для этого свойства значение true, тогда каждый фильтр выводится на отдельной строке.

Настройка баннера и логотипа

Можно настроить изображения, которые выводятся в области баннера в верхней части программы поиска. Например, можно заменить изображения по умолчанию для WebSphere II OmniFind Edition на изображения с символикой вашего предприятия. Чтобы не выводить баннер, закомментируйте одну или обе этих строки. Свойство banner.icon задает рисунок, который выводится в левой части области баннера. Свойство banner2.icon задает рисунок, который выводится в правой части области баннера.

```
banner.icon=/images/WS_II_OFEdition.gif  
banner2.icon=/images/WS_II_mosaic.gif
```

Настройка фонового изображения

Можно настроить изображения, которые выводятся в качестве фона страниц в программе поиска. Например, можно заменить изображения по умолчанию для WebSphere II OmniFind Edition на изображения с символикой вашего предприятия. Чтобы не выводить фоновое изображение, закомментируйте одну или несколько из этих строк.

```
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
```

Ссылки

Свойства в области Ссылки файла config.properties позволяют выводить имена страниц программы поиска в виде ссылок на каждой странице вместо вывода на панели инструментов и на страницах со вкладками. Вывод ссылок полезен, если программа поиска работает как портлет, и нужно минимизировать пространство, которое она занимает на странице портала.

Если вы предпочитаете перемещаться по программе поиска при помощи выбора опций на панели инструментов и на страницах со вкладками, закомментируйте эти строки.

Вкладки поиска

Свойства в области файла config.properties Вкладки поиска задают имена страниц JSP (Java Server Pages), используемых для страниц со вкладками в окне программы поиска Поиски (Поиск, Сложный поиск и Дерево категорий). Не редактируйте эти страницы, если у вас нет опыта программирования на Java и работы с JSP.

Вот некоторые примеры возможной настройки этой области:

- Направление программы поиска на пользовательские страницы JSP, придающие другой внешний вид страницам со вкладками.

- Комментирование записей Дерево категорий. Например, если для собраний не сконфигурированы категории, нет смысла выводить в программе поиска страницу Дерево категорий.
- Копирование записей страниц со вкладками в область файла config.properties Панели инструментов и комментирование этих строк. Например, можно включить только вывод панели инструментов и совсем не выводить страницы со вкладками.

Панели инструментов

Свойства в области файла config.properties Панели инструментов задают имена страниц JSP (Java Server Pages), используемых для панели инструментов программы поиска. Не редактируйте эти страницы, если у вас нет опыта программирования на Java и работы с JSP.

Вот некоторые примеры возможной настройки этой области:

- Направление программы поиска на пользовательские страницы JSP, придающие другой внешний вид панели инструментов.
- Комментирование записей панели инструментов для ссылок, которые вы не хотите выводить. Например, можно отключить на панели инструментов вывод ссылки на страницу О программе.
- Перемещение функции вывода страницы Сложный поиск из области файла config.properties Вкладки поиска, чтобы эта опция была доступна только на панели инструментов.

Содержательные метки типов документов

Можно улучшить читаемость фильтра по типу документов, если вместо фактических имен типов документов выводить более сжатые и выразительные названия. Типы документов, доступные программам поиска, определяются классом SI-API (Search and Index API - интерфейс прикладного программирования для поиска и индексов) AvailableDocumentTypes. Кроме того, в целях удобства доступные типы документов перечислены в конце файла config.properties.

documentType.метка=список_типов_документов_через_пробел

Задаёт имя, которое выводится в списке результатов поиска в строке фильтра по типу документов, и список фактических типов документов, которые выводятся, если пользователь выберет этот фильтр.

Например, можно задать метку html и назначить этому имени расширения файлов и типы MIME для различных документов Web. Если пользователь щелкнет по **html**, чтобы отфильтровать результаты поиска, будут выведены только документы с заданными расширениями и типами MIME.

В файле config.properties предопределены следующие отображения типов документов:

```
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
```

Пользовательские фильтры

Можно задать пользовательские запросы для фильтрации вывода документов - результатов поиска.

filterCustom.метка=термины_запроса

Задаёт имя, которое выводится в списке результатов поиска в строке пользовательского фильтра, и запрос, уточняющий результаты поиска, если пользователь выберет этот фильтр. (При просмотре результатов поиска пользователи могут выбрать пользовательский фильтр для вывода только документов - результатов поиска, соответствующих predetermined запросу.)

В следующем примере результаты поиска фильтруются для вывода только документов, относящихся к базе данных отдела кадров (hr - human resources):
`filterCustom.HR_database_only=databasename:hr`

Если пользователь нажмет **HR_database_only**, чтобы отфильтровать результаты поиска, будет обработан запрос `databasename:hr`. При выводе результатов поиска будут выведены только документы из базы данных hr.

Несколько пользовательских фильтров закомментированы и содержатся в файле `config.properties` в качестве примера.

Редактирование свойств примера программы поиска

Пример прикладной программы поисковой системы предприятия может искать информацию во всех активных собраниях и внешних источниках вашей системы. Можно редактировать файл свойств для задания опций среды вашего Web-сервера, для использования другой программы поиска в качестве программы поиска по умолчанию или для управления составом опций, выводимых при запуске программы поиска.

Об этой задаче

Программа установки внедряет пример программы поисковой системы предприятия в IBM WebSphere Application Server на поисковых серверах для поисковой системы предприятия. Для конфигурирования этой программы поиска надо редактировать файл свойств `config.properties`, который внедряется вместе с ней.

Чтобы изменения вступили в силу, остановите и перезапустите программу поисковой системы предприятия `ESSearchApplication` на сервере прикладных программ WebSphere.

Порядок действий

Чтобы отредактировать пример свойств программы поиска:

1. Редактируйте файл `config.properties` обычным текстовым редактором.
Файл `config.properties` установлен в следующий каталог (`ES_INSTALL_ROOT` здесь - каталог установки WebSphere II OmniFind Edition на поисковом сервере):
`ES_INSTALL_ROOT/installedApps/ESSearchApplication.ear/
ESSearchApplication.war/WEB-INF/config.properties`
2. Отредактируйте свойства, задав информацию о среде Web-сервера и предпочтениях поиска, затем сохраните и закройте файл. (В этом файле знак решетки # указывает на строку комментария.)
3. Остановите и перезапустите программу `ESSearchApplication`:

- a. На сервере поиска запустите консоль администратора WebSphere Application Server.

Чтобы открыть консоль администратора, можно:

- Выбрать программу в меню **Пуск** Windows.
 - В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
 - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
- b. Когда появится приглашение ввести ID пользователя и пароль, введите имя пользователя и пароль, зарегистрированные в реестре пользователей WebSphere Application Server. Если в качестве реестра пользователей используется локальная операционная система, можно задать ID и пароль администратора поисковой системы предприятия.
 - c. После регистрации на консоли администратора выберите **Applications** (Прикладные программы), а затем **Enterprise Applications** (Прикладные программы предприятия).
 - d. Включите переключатель для ESSearchApplication и нажмите кнопку **Остановить**.
 - e. Когда программа будет остановлена, включите ее переключатель еще раз и выберите **Пуск**.

Доступ к примеру программы поиска

Вы обращаетесь к примеру программы поиска, задавая URL в браузере.

Прежде, чем вы начнете

Пример программы поиска надо сконфигурировать для среды вашего Web-сервера.

Об этой задаче

Пример программы поиска установлен на поисковые сервера для поисковой системы предприятия. Можно использовать эту прикладную программу, как есть, для тестирования новых собраний и внешних источников перед тем, как сделать их доступными для пользователей. Можно также использовать эту программу в качестве модели для создания своих собственных программ поиска.

Порядок действий

Чтобы запустить пример программы поиска:

1. Введите URL программы поиска в браузере. Например:

`http://SearchServer.com/ESSearchApplication/`

SearchServer.com - имя хоста поискового сервера.

Кроме того, если для вашего Web-сервера не задано использование порта 80, надо задать допустимый номер порта. Например:

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

2. Если в WebSphere включена защита, зарегистрируйтесь в программе с действующим ID пользователя и паролем.
Если в каких-либо собраниях, доступных для программы поиска, включена защита, и защищенные собрания содержат искатели, сконфигурированные для проверки параметров регистрации пользователя при обработке запросов, можно сконфигурировать профиль пользователя. На странице Мой профиль задайте параметры аутентификации для доступа к защищенным доменам. После этого можно будет выполнять поиск в таких доменах без необходимости регистрироваться в них.
3. На странице Поиск выберите собрания и внешние источники для поиска и отправьте запрос.

Включение защиты для примера программы поиска

Если в WebSphere Application Server включена глобальная защита, и вы хотите использовать пример программы поиска для поиска в защищенных собраниях, нужно изменить параметры конфигурации в примере программы поиска и WebSphere Application Server.

Прежде, чем вы начнете

- Вы должны быть администратором поисковой системы предприятия.
- Нужно включить глобальную защиту в WebSphere Application Server. Если WebSphere II OmniFind Edition установлен в конфигурации с несколькими серверами, включите глобальную защиту на поисковых серверах. Инструкции по включению глобальной защиты смотрите в документации WebSphere Application Server.
- Если в качестве реестра пользователей в WebSphere Application Server выбран Lightweight Directory Access Protocol (LDAP), нужно добавить ID и пароль администратора поисковой системы предприятия в реестр LDAP (эти ID и пароль задаются при установке WebSphere II OmniFind Edition).

Например, если ID администратора поисковой системы предприятия - adminUser, запись для этого пользователя в реестре LDAP может выглядеть так: `uid=adminUser,ou=Employees,o=IBM,c=US`. Смотрите инструкции в документации по серверу LDAP.

Порядок действий

Чтобы включить защиту для примера программы поисковой системы предприятия:

1. С помощью консоли администратора измените свойства этой программы поиска:
 - a. Зарегистрируйтесь на консоли администратора поисковой системы предприятия как пользователь с привилегиями администратора поисковой системы предприятия.
 - b. На панели инструментов выберите **Защита**.
 - c. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать прикладные программы поиска**.
 - d. Нажмите кнопку **Добавить программу поиска** и в поле **Имя программы поиска** введите ID администратора поисковой системы предприятия, заданный при установке WebSphere II OmniFind Edition.
 - e. Убедитесь, что включен переключатель **Все собрания и внешние источники**, и нажмите кнопку **ОК**.
2. Отредактируйте файл `config.properties`:

- a. Если вы используете UNIX, откройте окно консоли. Если вы используете Microsoft Windows, откройте окно командной строки.
- b. Перейдите в каталог WEB-INF для примера программы поиска. В следующих примерах команда показана на двух строках для удобства чтения; задавайте эту команду в одной строке:

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. При помощи текстового редактора отредактируйте файл config.properties.
 - d. Измените значение свойства username на правильное имя пользователя WebSphere Application Server.
 - e. Измените значение свойства password на пароль для заданного пользователя.
 - f. Сохраните файл и закройте текстовый редактор.
3. Перезапустите программу ESSearchApplication в WebSphere Application Server:
- a. На сервере поиска запустите консоль администратора WebSphere Application Server.

Чтобы открыть консоль администратора, можно:

- Выбрать программу в меню **Пуск Windows**.
 - В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
 - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
- b. В окне ввода ID пользователя и пароля введите ID администратора и пароль, заданные при включении в WebSphere Application Server глобальной защиты.
 - c. После регистрации на консоли администратора выберите **Applications** (Прикладные программы), а затем **Enterprise Applications** (Прикладные программы предприятия).
 - d. Включите переключатель для ESSearchApplication и нажмите кнопку **Остановить**.
 - e. Когда эта программа будет остановлена, вновь включите переключатель для ESSearchApplication и нажмите кнопку **Запустить**.

Внешние источники поисковой системы предприятия

Внешний источник - это источник данных, на котором программе поисковой системы предприятия разрешен поиск документов без обработки искателем, синтаксического анализа и индексации.

Типы источников данных, на которых возможен поиск как на внешних источниках:

- Таблицы баз данных JDBC (Java database connectivity) - только для баз данных IBM DB2 Universal Database (DB2 UDB) и Oracle. Для каждой таблицы в базе данных JDBC создается отдельный внешний источник данных.
- Серверы LDAP (Lightweight Directory Access Protocol). Для каждого сервера LDAP создается один внешний источник.

Сконфигурировав информацию о внешнем источнике, свяжите его хотя бы с одной программой поиска. После этого поиск на внешнем источнике будет доступен пользователям наравне с запросами к собраниям, созданным при помощи искателя, синтаксического анализа и индексации данных.

Добавление внешних источников в систему

Добавляя внешний источник в поисковую систему предприятия, вы задаете тип этого источника. Мастер помогает задать информацию об источнике данных и допустимых типах поиска.

Прежде, чем вы начнете

Чтобы добавить внешний источник в систему, надо быть администратором поисковой системы предприятия.

Ограничения

Для поиска в базе данных Oracle как во внешнем источнике необходимо, чтобы на поисковых серверах поисковой системы предприятия была установлена программа клиента Oracle.

Об этой задаче

Добавляя в систему информацию о внешнем источнике, вы разрешаете пользователям запрашивать источник при помощи программ поисковой системы предприятия. Можно разрешить поиск на серверах LDAP (Lightweight Directory Access Protocol) и в таблицах баз данных JDBC (Java database connectivity).

При конфигурировании сервера LDAP запускается мастер, помогающий задать информацию, которая позволяет системе соединиться с сервером, а также опции поиска на сервере.

При конфигурировании базы данных JDBC запускается мастер, помогающий задать информацию, которая позволяет системе соединиться с базой данных, выбрать таблицы, на которых надо обеспечить возможность поиска, и задать опции поиска данных в этих таблицах. Для каждой таблицы, которую вы добавите в систему, создается отдельный внешний источник для поиска.

Порядок действий

Чтобы добавить внешний источник в систему:

1. Выберите **Внешние источники** - откроется окно Внешние источники.
2. Нажмите кнопку **Добавить внешний источник**.
3. Выберите тип внешнего источника, который хотите добавить - сервер LDAP или базу данных JDBC.
4. Нажмите кнопку **Далее**, чтобы начать конфигурирование внешнего источника.

Откроется мастер для создаваемого типа источника. Чтобы сконфигурировать внешний источник, следуйте подсказкам мастера. Чтобы узнать больше о доступных опциях, выберите **Помощь** на любой странице мастера.

Приведенные ниже имена и положения драйвера JDBC по умолчанию могут оказаться полезны при конфигурировании информации о соединении для баз данных DB2 Universal Database (DB2 UDB) и Oracle:

DB2: Драйвер JDBC Legacy

Имя драйвера: `COM.ibm.db2.jdbc.app.DB2Driver`

Положение: *корневой_каталог_установки_db2/java/db2java.zip*

DB2: Universal JDBC Driver

Имя драйвера: `com.ibm.db2.jcc.DB2Driver`

Положение:

корневой_каталог_установки_db2/java/db2jcc.jar

корневой_каталог_установки_db2/java/db2jcc_license_cu.jar

Oracle Имя драйвера: `oracle.jdbc.driver.OracleDriver`

Положение:

домашний_каталог_oracle/jdbc/lib/classes12.zip

домашний_каталог_oracle/jdbc/lib/nls_charset12.zip

5. Задав опции поиска на внешнем источнике, нажмите **Готово**.
Новый внешний источник появится в окне Внешние источники вместе с другими внешними источниками, добавленными в систему.

Задание прикладных программ поиска для внешних источников

Поиск на внешнем источнике станет доступным только после того, как вы свяжете его хотя бы с одной программой поиска.

Прежде, чем вы начнете

Задать для прикладных программ поиска внешние источники, в которых они могут выполнять поиск, может администратор поисковой системы предприятия.

Порядок действий

Чтобы связать прикладную программу поиска с одним или несколькими внешними источниками:

1. Выберите **Защита** на панели инструментов консоли администратора.
2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать прикладные программы поиска**.
3. На странице Конфигурировать прикладные программы поиска нажмите кнопку **Добавить прикладную программу поиска**.
4. Введите имя прикладной программы поиска.

- |
- |
- |
- |
- |
- |
- |
- |
- |
- |
- |
5. Выберите внешние источники, в которых прикладная программа поиска сможет выполнять поиск:
- Чтобы программа поиска обращалась ко всем внешним источникам, которые вы добавляете в систему, выберите **Все собрания и внешние источники**.
 - Чтобы разрешить программе поиска доступ только к указанным вами внешним источникам, выберите **Отдельные собрания и внешние источники**.
- Если выбрана эта опция, появляется список имен собраний и имен внешних источников. Включите переключатели **Выбрать** для всех внешних источников, в которых может выполнять поиск эта прикладная программа.
6. Нажмите кнопку **ОК**.

Защита поисковой системы предприятия

Механизмы защиты в поиске предприятия позволяют защитить источники от поиска неавторизованными пользователями и разрешить выполнение функций управления только определенным пользователям.

Поисковая система предприятия позволяет пользователям выполнять поиск в широком диапазоне источников данных. Чтобы к содержимому обращались только пользователи, авторизованные для доступа к нему, а к консоли администратора - только авторизованные для доступа к ней, поисковая система предприятия координирует и применяет защиту на нескольких уровнях.

Web-сервер

Первый уровень защиты - это Web-сервер. Если в WebSphere Application Server включена глобальная защита, можно назначить пользователям роли администраторов и выполнять аутентификацию пользователей, пытающихся управлять системой. Когда пользователь регистрируется на консоли администратора, этому пользователю доступны только те функции и собрания, для управления которыми он авторизован.

Для аутентификации пользователей, использующих для поиска в собраниях прикладные программы поиска, эти программы поиска могут также использовать поддержку аутентификации из WebSphere Application Server.

Защита уровня собрания

При создании собрания можно включить защиту на уровне собрания. Эту опцию нельзя изменить после создания собрания. Если защита на уровне собрания не включена, позже нельзя будет задать элементы управления защитой на уровне документов.

Когда защита уровня собрания включена, процессы глобального анализа применяют следующие особые правила:

- Чтобы элементы управления защитой работали для всех документов, система индексирует документы с одинаковым (или почти одинаковым) содержимым независимо друг от друга, а не совместно в общем каноническом представлении.
- Во время фазы глобального анализа, выполняющей обработку текста привязки, обычно связывает текст из одного документа (документа-источника) с другим документом (документом назначения), в котором не обязательно есть этот текст. Это позволяет получать документ назначения по запросам, содержащим текст из документа-источника. Такая обработка текста привязки представляет некоторый риск для защиты, если пользователям разрешен просмотр документа назначения, но запрещен просмотр документа-источника. Если включена защита собрания, текст привязки в ссылках из запрещенных документов исключается из индекса. Документ возвращается в результатах поиска, только если запросу соответствуют его собственное содержимое или метаданные.

Платой за включение защиты собрания является снижение качества поиска. Включение защиты собрания уменьшает объем информации, индексируемой для каждого документа. Побочное действие состоит в том, что для некоторых запросов будет найдено меньше результатов.

Кроме того, защита уровня собрания доступна прикладным программам поиска по ID прикладной программы. Для поиска в собраниях администратор поисковой системы предприятия должен связать прикладную программу поиска с конкретными собраниями, в которых она сможет выполнять поиск. Тогда можно будет при помощи стандартных механизмов управления доступом разрешать или запрещать пользователям доступ к прикладным программам поиска.

Защита на уровне документов

При конфигурировании искателей для собрания можно включить защиту уровня документов. Если выбрана эта опция, искатель может связывать с каждым просматриваемым документом маркеры защиты. Маркеры защиты хранятся в индексе вместе с документами. Для некоторых искателей можно также указать, что при обработке запроса нужно проверять параметры регистрации пользователя на соответствие текущим данным управления доступом (заданным на самом источнике данных).

Программы поиска могут использовать для управления доступом маркеры защиты и параметры регистрации пользователей. Чтобы пользователи могли находить и получать только те документы, на доступ к которым у них есть разрешение, программа поиска может включать в запросы, передаваемые поисковым серверам, параметры регистрации пользователя.

Защита собраний - это не только механизмы аутентификации и управления доступом, при помощи которых поисковая система предприятия защищает индексированное содержимое. Кроме этого, есть способы, защищающие данные в процессе передачи от перехвата злонамеренным неавторизованным пользователем. Например, поисковые серверы используют для связи с сервером индексов и с программами поиска такие протоколы, как SSL (Secure Sockets Layer), SSH (Secure Shell) и Secure Hypertext Transfer Protocol (HTTPS).

Дополнительная защита обеспечивается шифрованием. Например, пароль администратора поисковой системы предприятия, заданный при установке продукта, хранится в зашифрованном виде. Пароли, которые пользователи задают в профилях пользователей, также хранятся в зашифрованном формате.

Для повышенной защиты нужно, чтобы аппаратное обеспечение сервера было должным образом изолировано и защищено от неавторизованного вторжения. Установив брандмауэр, можно защитить серверы поисковой системы предприятия от вторжения из другой части вашей сети. Кроме того, на серверах поисковой системы предприятия не должно быть открыто никаких лишних портов. Сконфигурируйте систему так, чтобы она ожидала требований только на портах, в явном виде назначенных для работы поисковой системы предприятия и прикладных программ поиска.

Административные роли

Поисковая система предприятия использует для управления доступом к разным функциям на консоли администратора понятие ролей.

При установке WebSphere Information Integrator OmniFind Edition (WebSphere II OmniFind Edition) программа установки конфигурирует ID пользователя и пароль для администратора поисковой системы предприятия. При первом обращении к консоли администратора надо зарегистрироваться под именем этого пользователя. Если в IBM WebSphere Application Server не включена глобальная защита, этот ID пользователя - единственный ID пользователя, который можно использовать для обращения к консоли администратора поисковой системы предприятия.

Если в WebSphere Application Server включена глобальная защита, можно назначить администраторами поисковой системы предприятия других пользователей. Назначая пользователям роли, можно ограничивать доступ к отдельным собраниям и управлять функциями, выполняемыми отдельными пользователями-администраторами. ID пользователей, назначаемых для ролей администраторов поисковой системы предприятия, должны существовать в реестре пользователей WebSphere Application Server.

При регистрации пользователя-администратора поисковая система предприятия выполняет аутентификацию ID пользователя. На консоли будут доступны только те собрания и функции, которыми разрешено управлять этому пользователю.

Пользователям можно назначать следующие административные роли:

Администратор поисковой системы предприятия

Этот пользователь создает собрания и обладает полномочиями на управление всеми аспектами поисковой системы предприятия. При установке WebSphere II OmniFind Edition вы задаете ID пользователя и пароль для первого пользователя - администратора поисковой системы предприятия. После первой регистрации этот пользователь может назначить других пользователей для роли администратора поисковой системы предприятия.

Администратор собрания

Этот пользователь может изменять определенные собрания или все собрания, а также отслеживать работу этих собраний и управлять их работой. Эти пользователи не могут создавать собрания. Администраторы собраний могут отслеживать и выполнять действия на уровне системы, только если администратор поисковой системы предприятия предоставил им такие полномочия.

Оператор

Этот пользователь может отслеживать работу определенных собраний или всех собраний и управлять их работой. Например, эти пользователи могут запускать и останавливать работу собраний, но не могут создавать или изменять собрания. Оператор может отслеживать и выполнять действия на уровне системы, только если администратор поисковой системы предприятия предоставил ему такие полномочия.

Наблюдатель

Этот пользователь может отслеживать работу определенных собраний или всех собраний. Эти пользователи не могут управлять работой собраний (например, запускать или останавливать серверы), создавать и изменять собрания. Наблюдатель может отслеживать (но не выполнять) действия на уровне системы, только если администратор поисковой системы предприятия предоставил ему такие полномочия.

Конфигурирование пользователей-администраторов

Конфигурируя административные роли, можно ограничивать доступ к собраниям и управлять функциями, выполняемыми отдельными пользователями-администраторами.

Прежде, чем вы начнете

Прежде чем назначать пользователя для административной роли, убедитесь, что в WebSphere Application Server включена защита. Убедитесь также, что этот ID пользователя внесен в реестр пользователей WebSphere Application Server.

Для конфигурирования пользователей-администраторов надо быть администратором поисковой системы предприятия.

Порядок действий

Чтобы назначить пользователям административные роли:

1. Нажмите кнопку **Защита** - откроется окно Защита.
2. На странице Административные роли нажмите кнопку **Добавить пользователя**.
3. Введите ID пользователя для административной роли и выберите соответствующую административную роль.
4. Если этот пользователь - не администратор поисковой системы предприятия, задайте, разрешен ли ему доступ к страницам с панели инструментов **Система**.
Например, вам может понадобиться разрешить некоторым операторам или администраторам собраний отслеживать файлы журнала на уровне системы.
5. Если этот пользователь - не администратор поисковой системы предприятия, выберите собрания и внешние источники, которыми он может управлять.
Можно включить переключатели для отдельных собраний и внешних источников или позволить пользователю управлять всеми собраниями и внешними источниками.

Сравнение аутентификации и управления доступом

Для защиты содержимого от неавторизованных пользователей и для управления доступом к функциям управления поисковая система предприятия поддерживает аутентификацию пользователей и управление доступом.

Аутентификация

Аутентификация - это процесс, используемый системой для проверки, являются ли пользователи теми, за кого они себя выдают. Поскольку обычно доступ предоставляется на основе идентификатора пользователя, обращающегося к ресурсу, аутентификация необходима для эффективной защиты.

Для аутентификации пользователей, пытающихся получить доступ к консоли администратора, поисковая система предприятия использует поддержку аутентификации, входящую в состав WebSphere Application Server.

Для аутентификации пользователей, выполняющих поиск в собраниях поисковой системы предприятия, прикладные программы поиска могут использовать защиту WebSphere Application Server и реализовать предпочтительные способы аутентификации данных регистрации пользователей. Обычно данные регистрации пользователей состоят из ID пользователя и пароля, которые передаются прикладной программе поиска, когда пользователь регистрируется или пытается обратиться к ней.

Аутентификацию пользователей можно реализовать другими способами, в зависимости от ресурсов и протоколов, доступных в предприятии. Например, можно заставить пользователей идентифицировать себя при помощи смарт-карт, при помощи инфраструктуры цифровых сертификатов и общих ключей или путем назначения квитанций при регистрации пользователей, чтобы отслеживать состояние их аутентификации.

Управление доступом

Управление доступом - это ограничение возможных действий пользователей после их идентификации и аутентификации. Наиболее распространенный способ ограничения доступа к ресурсам - список управления доступом (ACL). ACL - это список идентификаторов пользователей (имен пользователей, имен групп, ролей пользователей и т.п.). Каждый идентификатор пользователя связывается с набором разрешений, определяющих права и привилегии пользователей.

Например, элементы управления доступом могут разрешать или запрещать доступ к файлам на файл-сервере и управлять тем, может ли пользователь, у которого есть доступ, читать, создавать, редактировать или удалять файлы на этом сервере.

В поиске предприятия все управление доступом зависит от того, есть ли у пользователя разрешение на чтение данных в индексе. В зависимости от того, каким образом включены элементы управления доступом в прикладных программах поиска, и от правил, заданных для собрания и для искателей при управлении поисковой системой предприятия, можно:

- Разрешить всем пользователям поиск во всех документах в собрании.
- Разрешить всем пользователям поиск в документах, просмотренных определенными искателями, и ограничить доступ к документам, просмотренным другими искателями.
- Разрешить определенным пользователям поиск в определенных документах. Например, при задании баз данных, которые нужно обрабатывать искателем Notes, можно задать опции, позволяющие некоторым пользователям выполнять поиск в определенных представлениях и папках, и запретить поиск этих документов для других пользователей.

Отключение защиты для прикладной программы предприятия в WebSphere Application Server

Для управления аутентификацией пользователей в процессах WebSphere II OmniFind Edition можно отключить в WebSphere Application Server глобальную защиту для отдельных прикладных программ предприятия.

Об этой задаче

Программа установки WebSphere II OmniFind Edition внедряет в WebSphere Application Server три прикладных программы предприятия:

- Программа ESAdmin содержит интерфейс для консоли администратора поисковой системы предприятия.
- Программа ESSearchApplication содержит интерфейс для консоли примера программы поиска.
- Программа ESSearchServer обеспечивает всю удаленную связь для реализации SI-API WebSphere II OmniFind Edition и позволяет интерфейсам SI-API взаимодействовать с серверами поиска.

По умолчанию все три прикладных программы предприятия поддерживают глобальную защиту WebSphere Application Server. Если эти программы определяют, что глобальная защита включена, они начинают аутентифицировать все получаемые ими требования.

В некоторых организациях может потребоваться включить или отключить защиту для отдельных прикладных программ предприятия WebSphere II OmniFind Edition.

Например, можно аутентифицировать всех пользователей с доступом к консоли администратора, но не аутентифицировать пользователей, использующих интерфейсы SIAPI WebSphere II OmniFind Edition или пример программы поиска.

Порядок действий

Чтобы отключить защиту для конкретной прикладной программы предприятия:

1. На сервере поиска запустите консоль администратора WebSphere Application Server.

Чтобы открыть консоль администратора, можно:

- Выбрать программу в меню **Пуск** Windows.
 - В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
 - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
2. В окне ввода ID пользователя и пароля введите ID администратора и пароль, заданные при включении в WebSphere Application Server глобальной защиты.
 3. После регистрации на консоли администратора выберите **Applications** (Прикладные программы), а затем **Enterprise Applications** (Прикладные программы предприятия).
 4. Включите переключатель рядом с именем прикладной программы предприятия, для которой надо отключить защиту.
 5. Прокрутите список вниз и щелкните по ссылке **Map security roles to users/groups** (Отобразить роли защиты для пользователей/групп).
 6. Найдите роль **AllAuthenticated** и включите переключатель под столбцом **Everyone?** (Все?).
 7. Нажмите кнопку **ОК**.
 8. Выберите **Save** (Сохранить), чтобы сохранить внесенные изменения.
 9. При использовании WebSphere Network Deployment включите переключатель **Synchronize changes with Nodes** (Синхронизировать изменения с Nodes).
 10. Нажмите кнопку **Save** (Сохранить).
 11. Еще раз выберите **Applications** (Прикладные программы), а затем **Enterprise Applications** (Прикладные программы предприятия).
 12. Включите переключатель для только что измененной программы и нажмите кнопку **Stop** (Остановить).
 13. После остановки программы снова включите для нее переключатель и нажмите кнопку **Start** (Запустить).

Защита уровня собрания

Для обеспечения защиты на уровне собрания надо сконфигурировать опции индексирования содержимого и опции, позволяющие прикладным программам поиска выполнять поиск в определенных собраниях.

При создании собрания можно выбрать опцию, включающую защиту собрания. Если эта опция выбрана, позже можно сконфигурировать элементы управления защитой

на уровне документов. Если защита собрания включена, процессы глобального анализа поисковой системы предприятия будут также применять различные правила для индексации повторных документов и текста привязки в документах.

После создания прикладной программы поиска ее ID позволяет задать собрания, в которых она может выполнять поиск, и пользователей, имеющих к ней доступ.

Анализ повторных документов

Если включена защита собрания, процессы глобального анализа не определяют повторения документов в собрании.

При глобальном анализе процессы индексирования определяют документы, являющиеся повторениями (точными или приблизительными) друг друга. Затем они связывают все эти документы с одним каноническим представлением содержимого. Если позволить определять повторения документов, в результатах поиска не будет множества документов с одинаковым (или почти одинаковым) содержимым.

Если при создании собрания включить защиту собрания, повторения документов не определяются и, соответственно, не связываются с общим каноническим представлением. Вместо этого каждый документ индексируется отдельно. Это позволяет пользователям выполнять поиск только в документах с маркерами защиты, соответствующими данным регистрации пользователей. Например, содержимое двух документов может быть почти одинаковым, но для защиты они могут использовать разные списки управления доступом.

Отключение анализа повторений документов может улучшить защиту документов в собрании, но если пользователи будут получать в результатах поиска несколько копий одного документа, качество поиска может ухудшиться.

Анализ привязок текста

Если включена защита собрания, процессы глобального анализа применяют специальные правила индексации текста привязки в документах, просматриваемых искателями Web. Если не включить защиту для собрания, можно при конфигурировании отдельных искателей Web выбрать, индексировать ли текст привязки в ссылках на запрещенные документы.

Текст привязки - это информация в гипертекстовой ссылке, описывающая страницу назначения ссылки. Например, в следующей ссылке текст Синтаксис запросов - это текст привязки в ссылке на страницу сайта syntax.htm:

```
<a href="../doc/syntax.htm">Синтаксис запросов</a>
```

Обычно искатель Web переходит по ссылкам в документах для просмотра дополнительных документов и включает в индекс страницы назначения ссылок. Во время глобального анализа процессы индекса связывают текст привязки не только с документом, в котором он находится (документом-источником), но и с документом назначения. В приведенном примере текст привязки Синтаксис запросов связывается со страницей назначения syntax.htm и со страницей, содержащей конструкцию привязки.

Если при создании собрания включить защиту собрания, обработка текста привязки выключается. Текст привязки не будет проиндексирован с документом, если только он не встречается в самом документе или в метаданных документа. Такое управление защитой гарантирует, что пользователям не будет видна информация в документах, к

которым у них нет доступа, так как текст привязки из ссылок на запрещенные документы не будет связан с документами, к которым у пользователей есть доступ.

Включение защиты собрания может улучшить защиту документов Web, позволяя пользователям выполнять поиск только в документах с маркерами защиты, соответствующими данным регистрации пользователей. Однако при отключенной обработке текста привязки среди результатов поиска могут оказаться не все документы, потенциально релевантные для запроса.

Если не включить защиту для собрания, можно при конфигурировании отдельных искателей Web выбрать опцию индексации текста привязки в ссылках на запрещенные документы. Если задана индексация текста привязки, процессы анализа и индексирования индексируют текст привязки для всех страниц, полученных искателем Web. Если задано, что текст привязки не надо индексировать, текст привязки из ссылок на запрещенные документы в индекс не войдет.

Индексация текста привязки в ссылках на запрещенные документы

Если некоторые документы содержат ссылки на документы, которые искателю Web запрещено просматривать, при конфигурировании искателя Web можно указать, нужно ли сохранять в индексе текст привязки для таких ссылок.

Прежде, чем вы начнете

Для конфигурирования опций индексации текста привязки нужно быть администратором поисковой системы предприятия или администратором собрания для искателя Web, который вы хотите сконфигурировать.

Об этой задаче

Директивы в файле robots.txt или метаданные документов Web могут запрещать искателю Web доступ к документам с некоторого сайта. Если документ, который разрешено просматривать искателю Web, содержит ссылки на запрещенные документы, можно задать, как нужно обрабатывать текст привязки для таких ссылок.

При конфигурировании искателя Web можно выбрать, индексировать ли текст привязки в ссылках на запрещенные документы. Для максимальной защиты укажите, что текст привязки в ссылках на запрещенные документы не должен индексироваться. Однако если текст привязки не индексируется, результаты поиска могут содержать не все документы, потенциально релевантные для запроса.

Порядок действий

Чтобы включить или отключить индексацию текста привязки в ссылках на запрещенные документы:

1. Отредактируйте собрание, найдите страницу Искатель, найдите искатель Web, который хотите сконфигурировать, и выберите  **Свойства искателя**.
2. Нажмите кнопку **Изменить дополнительные свойства искателя Web**.
3. Чтобы индексировать текст привязки во всех документах, просматриваемых этим искателем, включите переключатель **Индексировать текст привязки в ссылках на запрещенные документы**. Пользователи смогут узнать о страницах, которые искателю Web не разрешено просматривать, выполняя поиск текста в тексте привязки ссылок, указывающих на такие страницы.

Чтобы исключить из индекса текст привязки в ссылках на запрещенные документы, выключите этот переключатель. Пользователи не смогут узнать о страницах, которые искателю Web не разрешено просматривать. Текст привязки будет исключен из индекса вместе в запрещенными документами.

4. Нажмите кнопку **ОК** и затем, на странице Свойства искателя Web еще раз нажмите кнопку **ОК**.
5. Чтобы изменения вступили в силу, остановите и перезапустите искатель.

Чтобы эти изменения были применены для ранее проиндексированных документов, нужно выполнить повторный просмотр документов и затем заново проиндексировать их. Если предыдущие операции просмотра добавили в индекс информацию о запрещенных документах, эта информация будет удалена из индекса.

Защита с ID прикладных программ поиска

Для обеспечения защиты уровня собрания надо указать, какие прикладные программы поиска могут выполнять поиск в каждом из собраний и внешних источников.

Каждая прикладная программа поиска должны передавать в API поисковой системы предприятия идентификатор прикладной программы. Администратор поисковой системы предприятия и программы поиска могут использовать этот идентификатор для применения защиты собраний и внешних источников.

Чтобы программа поиска могла обращаться к собранию или внешнему источнику, администратор поисковой системы предприятия должен связать эту программу поиска с конкретными собраниями и источниками, в которых она сможет выполнять поиск. Программа поиска может выполнять поиск во всех собраниях и внешних источниках в поисковой системе предприятия или же только в заданных вами собраниях и внешних источниках.

Для применения управления доступом надо связать маркеры защиты (например, ID пользователей, ID групп или роли пользователей) с программой поиска и разрешить доступ к программе только этим пользователям. Например, можно разрешить доступ только к URL запуска прикладной программы поиска.

Дополнительную информацию об ID прикладных программ поиска и о том, как добавить управление защитой в пользовательские программы поиска, смотрите в главе API поиска и индексирования для поисковой системы предприятия.

Защита на уровне документов

Если при создании собрания для него включена защита, можно сконфигурировать элементы управления защитой уровня документов. При защите уровня документов пользователям, выполняющим поиск в собраниях, предоставляется доступ только к тем документам, которые им разрешено просматривать.

Для управления доступом к документам в собрании искатели могут собирать данные защиты для индекса. Для некоторых типов источников данных можно также при передаче запроса проверить текущие параметры регистрации пользователя. Для проверки параметров регистрации текущего пользователя можно встроить в пользовательские программы поиска поддержку профилей пользователей. Если сохранять профили пользователей, пользователи смогут обращаться к документам, не вводя многократно свои параметры регистрации.

Проверка на основе хранимых маркеров защиты

Если при создании собрания для него включена защита, можно сконфигурировать элементы управления защитой уровня документов, сохранив данные защиты в индексе.

По умолчанию все документы считаются общедоступными, то есть разрешенными для поиска всем пользователям. Для большинства типов документов можно обеспечить защиту уровня документов, связав с документами один или несколько маркеров защиты и сохранив эти маркеры с документами в индексе. При конфигурировании искателя вы задаете, что хотите при помощи маркеров защиты указать, каким пользователям разрешен доступ к документам, найденным этим искателем.

Если какой-то тип источников данных содержит поля, можно указать, что для управления доступом нужно применять значения в одном из этих полей. Если в источнике данных нет полей, если вы не хотите использовать значение какого-то поля для целей защиты или если заданное поле не содержит значений, которые можно применять для управления доступом, можно определить маркеры защиты, которые искатель должен связывать с документами.

Администратор выбирает для каждого собрания маркеры защиты, которые искатель должен связывать с документами. Например, маркер защиты может представлять ID пользователя, ID группы, роль пользователя или любое другое значение, определенное вами как действительное для источника данных. Если администратор источника данных изменяет собственный список управления доступом, измененные элементы управления доступом становятся доступными после следующего обновления или реорганизации индекса.

Документы, проходящие этапы синтаксического анализа, лексического анализа и индексирования, сопровождаются маркерами защиты. Если прикладные программы поиска поддерживают защиту, можно использовать маркеры защиты для управления доступом к документам. Пользователи, выполняющие поиск в собрании, могут выполнять поиск только в документах, которые им разрешено просматривать в соответствии с данными регистрации этих пользователей. Если параметры регистрации пользователя не соответствуют правилам защиты, пользователь не сможет выполнять поиск документов, защищенных маркерами защиты.

Можно применить пользовательские логические правила для определения значений маркеров защиты, закодировав эти правила в классе Java. При конфигурировании свойств искателя вы задаете имя подключаемого модуля, который будет использовать искатель при просмотре документов. Маркеры защиты, добавляемые этим подключаемым модулем, сохраняются в индексе и могут использоваться для управления доступом к документам.

Проверка текущих параметров регистрации при обработке запроса

Если при создании собрания для него включена защита, в некоторых типах доменов можно выполнять проверку параметров регистрации текущего пользователя при передаче запроса этим пользователем.

При конфигурировании перечисленных ниже типов искателей можно выбрать опцию проверки параметров регистрации пользователя путем сравнения их с элементами управления доступом, управляемыми исходным репозиторием:

- Искатель Content Edition (только для типов репозитория Documentum, FileNet Panagon Content Services и Portal Document Manager)
- Искатель Domino Document Manager
- Искатель Notes
- Искатель QuickPlace
- Искатель WebSphere Portal
- Искатель для файловых систем Windows

Перед ответом на запрос серверы поиска обращаются к исходным репозиториям для проверки разрешений текущего пользователя, а затем удаляют из результатов поиска все документы, для которых у пользователя нет разрешений на просмотр.

Такой метод применения защиты уровня документов обеспечивает высокий уровень защиты, поскольку параметры регистрации пользователя сравниваются с текущими данными защиты, а не с данными защиты, сохраненными в индексе. Кроме того, такой метод гарантирует, что доступ управляется механизмами защиты исходного репозитория, независимо от степени их сложности. Поскольку фильтрация документов осуществляется в реальном времени, результаты поиска отражают самые свежие значения управления доступом для каждого документа, соответствующего критерию поиска.

Еще одно преимущество такого подхода в том, что размер индекса не увеличивается (в индексе не требуется дополнительное пространство для маркеров защиты). Однако поскольку для проверки приходится соединяться с исходными репозиториями, при таком подходе может снизиться производительность выполнения запросов.

Для получения максимальной защиты и минимизации влияния на производительность выполнения запросов можно использовать опцию сохранения маркеров защиты совместно с опцией проверки текущих элементов управления доступом. Когда пользователь передает запрос, проверка выполняется в два этапа:

- Сначала серверы поиска используют данные защиты в индексе, чтобы узнать, разрешен ли пользователю доступ к серверу и базе данных, в которых найден документ (индекс оптимизирован для максимальной скорости и позволяет получить время ответа менее секунды).
- После этого серверы поиска создают промежуточный список, содержащий только документы из доменов на серверах, к которым разрешен доступ данному пользователю. Серверы поиска используют этот список для соединения с исходными репозиториями и проверки того, разрешено ли данному пользователю просматривать запрошенный документ.

Если пользователю разрешен доступ к серверу и домену, скорее всего, ему разрешен и доступ к документам. Однако этот последний этап фильтрации гарантирует, что в результатах поиска возвращаются только документы, соответствующие текущим разрешениям этого пользователя.

Использование профилей и управление идентификацией

Создавая профили для поисковой системы предприятия, пользователи получают возможность сохранять свои данные регистрации для поиска в защищенных доменах.

При поиске в домене, требующем проверки параметров регистрации пользователей при передаче запроса, пользователи должны передать программе поиска свои данные регистрации для входа в этот домен. При помощи WebSphere Information Integrator OmniFind Edition *управления идентификацией* пользователи могут сохранять в своих профилях параметры регистрации для любого числа доменов. Эти параметры шифруются для защиты и в таком виде хранятся в поисковой системе предприятия.

Пользователи могут создать профили пользователя и зарегистрировать свои параметры регистрации во время работы с программой поиска. В примере программы поиска для поисковой системы предприятия эта возможность обеспечивается опцией **Мой профиль** (в пользовательских программах поиска эта возможность может быть реализована иначе).

В профиле пользователя сохраняются различные параметры регистрации, которые пользователь должен вводить для входа в домены, где выполняется поиск. Пользователи могут создать профиль пользователя, если выполнены все приведенные ниже условия:

- В WebSphere Application Server включена глобальная защита.
- На консоли управления поисковой системой предприятия включено управление идентификацией WebSphere II OmniFind Edition.
- Включена защита как минимум в одном из собраний, в которых программа поиска может выполнять поиск.
- Хотя бы в одном защищенном собрании есть документы, которые просмотрены искателем, применяющим управление доступом путем требования проверки параметров регистрации текущего пользователя при передаче запроса.
- При конфигурировании защиты уровня документов для хотя бы одного из перечисленных ниже искателей включена опция проверки текущих данных регистрации при обработке запросов:
 - Content Edition (для некоторых типов репозиториев)
 - Domino Document Manager
 - Notes
 - QuickPlace
 - WebSphere Portal
 - для файловых систем Windows

В профиле перечисляются все доступные для программы поиска домены, для которых при обработке запроса требуется проверка данных регистрации пользователя. Пользователь может выбрать, для каких доменов он хочет сохранить параметры регистрации. Если пользователь не задал параметры регистрации для какого-либо домена, в результатах поиска не будет документов из источников данных этого домена.

Если управление регистрацией WebSphere II OmniFind Edition не используется, программа поиска должна выдать строку контекста защиты пользователя (user's security context, USC) при выполнении пользователем поиска в доменах, требующих проверки текущих данных регистрации.

Конфигурирование управления идентификацией

Сохранять профили пользователей можно при помощи управления идентификацией WebSphere II OmniFind Edition. Профили позволяют пользователям выполнять поиск в доменах, для которых при обработке запроса требуются проверка параметров регистрации пользователя.

Прежде, чем вы начнете

Для конфигурирования опций управления идентификацией надо быть администратором поисковой системы предприятия.

Об этой задаче

Если вы задали, что параметры регистрации пользователей проверяются во время обработки запроса при конфигурировании опций защиты уровня документа для искателя, пользователи должны будут при запросах в домене, требующем проверки, вводить свои параметры регистрации. С помощью управления идентификацией WebSphere II OmniFind Edition пользователи могут создавать профили и регистрировать параметры регистрации, используемые ими для входа в защищенные домены. Эти параметры регистрации шифруются в защищенной базе данных под управлением WebSphere II OmniFind Edition.

Серверы поиска используют сохраненные параметры регистрации для аутентификации пользователей, выполняющих поиск в защищенном домене. При неверных параметрах регистрации документы из защищенного домена исключаются из результатов поиска.

Порядок действий

Чтобы сконфигурировать управление идентификацией:

1. Нажмите кнопку **Защита** - откроется окно **Защита**.
2. На странице Прикладные программы поиска нажмите кнопку **Сконфигурировать управление идентификацией**.
3. На странице Сконфигурировать управление идентификацией включите переключатель, разрешающий WebSphere II OmniFind Edition управлять параметрами регистрации пользователей в их профилях.
4. Нажмите кнопку **ОК**.

Обеспечение защиты уровня документов для документов файловых систем Windows

Чтобы включить проверку текущих данных аутентификации пользователя при поиске в документах, просмотренных искателем для файловых систем Windows, надо сконфигурировать информацию учетной записи домена и на сервере искателя, и на сервере Microsoft Windows.

При конфигурировании искателя для файловых систем Windows вы задаете, просматривать ли подкаталоги на локальном компьютере или же на удаленном компьютере. Если для собрания включена защита, можно задать также опции управления доступом к документам в просматриваемых подкаталогах.

Если вы решили применить управление доступом путем проверки параметров регистрации текущего пользователя при передаче им запроса, надо проверить, что учетные записи домена правильно сконфигурированы. Требования к конфигурированию учетных записей доменов для файлов, найденных искателем на локальном компьютере, отличаются от требований для файлов, найденных искателем на удаленном сервере Windows.

Проверка с помощью локальных данных управления доступом

Для проверки параметров регистрации текущего пользователя система использует информацию как его локальной учетной записи, так и учетной записи домена (если компьютер принадлежит к домену Windows). Для проверки параметров регистрации при обработке запроса информация защиты должна включать оба имени пользователя - локальное и доменное.

Локальные учетные записи

В локальной учетной записи имя пользователя записывается в формате:

ИМЯ_КОМПЬЮТЕРА\ИМЯ_ПОЛЬЗОВАТЕЛЯ

Для входа в систему пользователи указывают только имя пользователя, но для правильного назначения прав пользователя Windows используется полное имя. Например, если в локальной учетной записи имя пользователя - abcuser, в полной учетной записи оно может быть WINSERVER1\abcuser.

Работая с программами поиска и конфигурируя профили для поиска защищенных документов в локальной системе, пользователи должны указывать имя для регистрации в Windows (например, abcuser).

Доменные учетные записи

В доменной учетной записи имя пользователя записывается в формате:

ИМЯ_ДОМЕНА\ИМЯ_ПОЛЬЗОВАТЕЛЯ

Для входа в систему пользователи вводят эту информацию в формате:

ИМЯ_ПОЛЬЗОВАТЕЛЯ@ИМЯ_ДОМЕНА

Например, если вы сконфигурировали назначения прав пользователей для файла и выбрали домен WIN1\abcuser, учетная запись будет иметь вид abcuser@win1.company.com.

Работая с программами поиска и конфигурируя профили для поиска документов в защищенном домене, пользователи должны указывать имя для регистрации в Windows (например, abcuser@win1.company.com).

Чтобы применить проверку текущих параметров регистрации на локальных компьютерах, в учетных записях пользователей Windows, используемых сервером искателя, должны быть следующие права. (Назначить права пользователям можно с помощью окна Windows Администрирование: **Администрирование** → **Локальная политика безопасности** → **Локальные политики** → **Назначение прав локальным пользователям.**)

- У ID пользователя, от имени которого запущен сервер искателя, должно быть право **Действовать как часть операционной системы.** (Это право конфигурируется для пользователя - администратора поисковой системы предприятия на сервере искателя при установке WebSphere Information Integrator OmniFind Edition.)
- У пользователей должно быть право **Входить локально.**

Проверка с помощью данных управления доступом к удаленному домену

Для операционной системы Windows любой каталог, начинающийся с записи *\\имя_сервера* считается удаленным каталогом. Например:

\\software\utilities\IBM

Для доступа к удаленному каталогу пользователи должны задать свои имена в формате:

ИМЯ_ПОЛЬЗОВАТЕЛЯ@ИМЯ_ДОМЕНА

Работая с программами поиска и конфигурируя профили для поиска защищенных документов в удаленной системе, пользователи должны указывать имя для регистрации в удаленной системе Windows (например, abcuser@win1.company.com).

Чтобы применить проверку текущих параметров регистрации на удаленных компьютерах, в учетных записях пользователей Windows должны быть следующие

права. (Назначить права пользователям можно с помощью окна Windows Администрирование: **Администрирование** → **Политика безопасности домена.**)

- Сервер искателя и просматриваемый сервер Windows должны входить в один домен.
- У ID пользователя, от имени которого запущен сервер искателя, должно быть право **Действовать как часть операционной системы.** (Это право конфигурируется для пользователя - администратора поисковой системы предприятия на сервере искателя при установке WebSphere Information Integrator OmniFind Edition.)
- У пользователей должно быть право **Входить в качестве пакетного задания.**

Обеспечение защиты уровня документов для документов Lotus Domino

Если просматриваемый искателем сервер Lotus Notes использует протокол вызова удаленных процедур Notes (Notes remote procedure call, NRPC), необходимо сконфигурировать сервер искателя так, чтобы можно было применить управление доступом на уровне документов.

Чтобы применить защиту уровня документов к документам, обработанным искателем на сервере Lotus Notes, который использует протокол NRPC, надо установить сервер Domino на сервере искателя. Этот сервер Domino должен быть членом вашего домена Domino. Чтобы установить и сконфигурировать сервер Domino, следуйте инструкциям в документации Lotus Domino.

Необходимо также выполнить следующие задачи, чтобы серверы поиска могли проверить, есть ли у пользователя, выполняющего поиск в защищенном собрании, права просмотра документов Lotus Notes, удовлетворяющих критериям поиска. Документы, на просмотр которых у пользователя нет прав, исключаются из результатов поиска до возвращения их пользователю.

- “Конфигурирование доверенных серверов Lotus Domino для проверки параметров регистрации пользователей”.
- Включение глобальной защиты в WebSphere Application Server и конфигурирование прикладных программ поиска для использования защиты. Этот шаг приводит к тому, что при попытке пользователей работать с программами поиска у них запрашиваются параметры регистрации. С помощью этих параметров серверы поиска могут затем проверять обращения всех пользователей к документам Lotus Notes.

Конфигурирование доверенных серверов Lotus Domino для проверки параметров регистрации пользователей

Чтобы применить защиту для документов, просмотренных искателем Notes, который использует протокол вызова удаленных процедур Notes (Notes remote procedure call, NRPC), серверы Domino надо сконфигурировать, сделав их доверенными серверами Lotus Domino.

Прежде, чем вы начнете

Эта процедура обязательна, если нужно применить защиту на уровне документов при поиске в удаленных базах данных. Для поиска в базах данных, локальных по отношению к серверу искателя, эта процедура необязательна.

Для конфигурирования доверенных серверов на сервере искателя должен быть установлен сервер Domino. Этот сервер Domino должен быть членом вашего домена Domino.

Об этой задаче

При конфигурировании опций защиты на уровне документов для искателя Notes вы задаете, нужно ли применить управление доступом путем проверки параметров регистрации текущего пользователя при передаче им запроса. Для применения этого типа защиты просматриваемые искателем серверы Domino должны быть доверенными серверами Lotus Domino.

Когда пользователи выполняют поиск в доменах, требующих проверки их текущих данных регистрации, доверенный сервер позволяет переключить контекст с ID сервера Domino на ID текущего пользователя. База данных Domino открывается так, как если бы ее открыл текущий пользователь, и для этого пользователя применяется вся информация списка управления доступом к этой базе данных.

Как правило, возможность такого переключения контекстов доступна только для баз данных, хранящихся в каталоге data локального сервера Domino. Начиная с Lotus Domino Версии 6.5.1, эта возможность реализуется через доверенный сервер. При конфигурировании доверенного сервера администратор Domino задает, каким серверам Domino доверяется выполнение ответственных операций, таких как действия от имени другого пользователя при обращении к базе данных с удаленного компьютера.

Порядок действий

Чтобы сконфигурировать доверенный сервер, выполните следующие действия на всех серверах Domino, просматриваемых искателем Notes:

1. На сервере Domino откройте с помощью файла ID администратора домена Domino клиент администратора Lotus Domino.
2. Выберите **Файл** → **Открыть сервер**.
3. Введите имя сервера Domino, для которого нужно включить функциональные возможности доверенного сервера.
4. Выберите вкладку **Конфигурация**.
5. Раскройте объект **Сервер**, выберите документ **Текущий сервер** и нажмите кнопку **Редактировать сервер**.
6. Выберите вкладку **Защита**, прокрутите до конца документа, найдите запись **Доверенные серверы** и нажмите кнопку со стрелкой вниз.
7. Выберите одну из опций:

LocalDomainServers

Выберите эту опцию, если все серверы в домене Domino должны считаться доверенными серверами.

имя_сервера

Задайте имя сервера Domino, который вы хотите просматривать искателем как доверенный сервер.

Если просматриваемый сервер Domino находится в другом домене Domino, надо указать имя этого сервера или выбрать нужную группу **OtherDomainServers**. Надо также выполнить процедуры Domino для перекрестной сертификации файла ID сервера WebSphere II OmniFind Edition Domino с другими доменами Domino. Информацию об этих процедурах смотрите в документации по серверу Domino.

8. Нажмите кнопку **Сохранить и закрыть**, чтобы сохранить внесенные изменения.
9. Остановите и перезапустите удаленные серверы Domino, сконфигурированные вами как доверенные серверы.

Отключение защиты на уровне документов

Можно разрешить пользователям выполнять поиск в собрании независимо от того, связано ли с документами в индексе какое-либо управление доступом. Для искателей, поддерживающих проверку текущих параметров регистрации, можно также разрешить пользователям поиск в собрании без проверки текущих элементов управления доступом во время обработки запросов.

Прежде, чем вы начнете

Чтобы сконфигурировать опции защиты на уровне документов, надо быть администратором поисковой системы предприятия или администратором данного собрания.

Ограничения

Опции защиты на уровне документов можно указывать только тогда, когда защита была включена во время создания собрания.

Об этой задаче

При конфигурировании искателя можно связать с просматриваемыми им документами маркеры защиты. Поисковые программы могут использовать эти маркеры, хранящиеся в индексе, для обеспечения управления доступом при выполнении пользователями поиска в собрании. Для некоторых искателей можно также указать, что при передаче пользователями запросов надо проверять текущие элементы управления доступом, связанные с документами в их исходных репозиториях.

Чтобы снять эти ограничения защиты, можно указать, что серверы поиска должны игнорировать все маркеры защиты, переданные с запросом. Можно также разрешить пользователям запрашивать документы без сравнения их параметров регистрации с текущими элементами управления доступом.

Временное отключение защиты на уровне документов может потребоваться при тестировании нового собрания или при устранении ошибки поисковой программы.

Порядок действий

Чтобы отключить управление доступом на уровне документов:

1. Отредактируйте собрание, перейдите на страницу Общие и нажмите кнопку **Конфигурировать защиту на уровне документов**.
2. На странице Защита на уровне документов включите переключатель **Игнорировать управление доступом на уровне документов в индексе**, если вы не хотите использовать при запросах пользователей к собранию маркеры защиты, которые искатели связывают с документами.
При этом искатели продолжают добавлять маркеры защиты к документам, но поисковые серверы игнорируют эти маркеры, что позволяет пользователям находить ранее защищенные документы.
3. Включите переключатель **Не проверять текущие параметры регистрации во время обработки запроса**, если вы не хотите проверять текущие элементы управления доступом, связанные с документами в исходных репозиториях при передаче запросов пользователями. Этот переключатель доступен только для документов, найденных искателями, которые поддерживают эту возможность.

Если этот переключатель включен, другие опции защиты на уровне документов продолжают действовать. Например, если при конфигурировании искателя вы задали опции для сохранения элементов управления доступом в индексе, это управление защитой будет продолжать действовать, пока не будет включен также переключатель **Игнорировать управление доступом на уровне документов в индексе**.

4. Нажмите кнопку **ОК**.

Интеграция поисковой системы предприятия с WebSphere Portal

Возможности поиска IBM WebSphere Portal можно расширить, разместив портлеты поисковой системы предприятия в WebSphere Portal и WebSphere Portal Search Center.

Точки интеграции

Портлеты поисковой системы предприятия интегрируются с WebSphere Portal несколькими способами:

WebSphere Portal

WebSphere Portal предоставляет пользователям единую точку доступа для взаимодействия с прикладными программами, содержанием, процессами и людьми. Структура WebSphere Portal позволяет интегрировать и размещать новые программы, называемые портлетами, не оказывая влияния на другие программы в этом портале.

Разместив в WebSphere Portal портлет поисковой системы предприятия, можно при помощи интерфейса WebSphere Portal выполнять поиск в собраниях поисковой системы предприятия. Конфигурируя параметры WebSphere Portal, можно добиться того, что портлет поисковой системы предприятия будет иметь такие же вид и поведение, как и остальные портлеты в среде WebSphere Portal.

Portal Search Engine

Механизм поиска WebSphere Portal обрабатывает сайты искателями. Портлеты управления дают администраторам возможность строить индексированные собрания, а портлеты поиска позволяют пользователям выполнять поиск в таких собраниях.

Если вы используете WebSphere Portal Версии 5.0.2 или новее, можно перенастроить информацию о конфигурации собраний и таксономии Portal Search Engine для поисковой системы предприятия.

Если вы используете WebSphere Portal Версии 5.1 или новее, можно использовать портлет поиска документов в Portal Search Engine для поиска в собраниях предприятия. Свойства конфигурации дают администраторам возможность при необходимости быстро переключаться между этими двумя возможностями поиска.

WebSphere Portal Search Center

WebSphere Portal Search Center служит в качестве центральной отправной точки для поиска во всех источниках, которые сделаны доступными для поиска через WebSphere Portal. Search Center и портлет универсального поиска позволяют выполнять поиск в содержимом WebSphere Portal и во всех других собраниях, зарегистрированных администратором в Search Center.

Интерфейс Search Center состоит из нескольких страниц. Можно либо выполнять поиск во всех доступных собраниях посредством одной общей страницы, или же выбрать страницу для поиска в конкретном собрании. Например, есть страница для индексов Portal Search Engine и страница для библиотек Portal Document Management.

Чтобы поиск в собраниях предприятия можно было выполнять из Центра поиска, IBM WebSphere Information Integrator OmniFind Edition (WebSphere II

OmniFind Edition) предоставляет адаптер и портлет регистрации. Адаптер добавляет вкладку для поисковой системы предприятия в интерфейс Центра поиска, а портлет регистрации регистрирует портлет поисковой системы предприятия в Центре поиска.

Адаптер и портлет регистрации должны быть установлены прежде, чем Центр поиска можно будет использовать для поиска в собраниях поисковой системы предприятия.

Искатель WebSphere Portal

На серверах поиска при установке WebSphere II OmniFind Edition устанавливается прикладная программа предприятия ESPACServer.ear. Внедрив эту прикладную программу предприятия в WebSphere Portal, с консоли администратора можно сконфигурировать искатель WebSphere Portal и добавить сайты WebSphere Portal в собрания поисковой системы предприятия.

Преимущества интеграции

Поисковая система предприятия расширяет возможности среды поиска WebSphere Portal, предоставляя поддержку для поиска в большем числе типов источников данных. При помощи портлета поисковой системы предприятия можно выполнять поиск на сайтах и во всех других типах источников данных, поддерживаемых WebSphere II OmniFind Edition.

Поисковая система предприятия дает также преимущества в масштабируемости. Portal Search Engine полезен для малого и среднего бизнеса, в котором для обеспечения рабочей нагрузки по поиску и получению документов достаточно одного поискового сервера. Для поддержки возможностей на уровне предприятия система поисковая система предприятия распределяет рабочую нагрузку между четырьмя серверами (один для работы искателей с данными, один для анализа и индексирования данных и два для поддержки процессов поиска и получения документов).

Обзор размещения портлетов

Какие портлеты можно использовать для поисковой системы предприятия, зависит от используемой вами версии WebSphere Portal.

WebSphere Portal Версии 5.0.2

Можно внедрить портлет поисковой системы предприятия и использовать его для поиска в собраниях предприятия. Этот портлет может существовать вместе с портлетами Portal Search Engine.

WebSphere Portal Версии 5.1 или новее

- Можно внедрить портлет поисковой системы предприятия и использовать его для поиска в собраниях предприятия. Этот портлет может существовать вместе с портлетами Portal Search Engine.
- После внедрения портлета поисковой системы предприятия можно сконфигурировать портлет поиска документов в механизме поиска WebSphere Portal для поиска в собраниях предприятия.
- После регистрации портлета поисковой системы предприятия в Центре поиска WebSphere Portal можно использовать этот Центр для поиска в собраниях предприятия. Можно либо выбрать страницу для поиска только в собраниях предприятия, либо ввести запрос на поиск в собраниях предприятия и во всех других собраниях, доступных в Центре поиска.

Защита уровня документов с Portal Search Engine

IBM WebSphere Portal Search Engine можно использовать для обеспечения защиты на уровне документов при поиске в собраниях поисковой системы пользователями.

Если искатель поисковой системы предприятия связывает маркер защиты ID группы с найденными этим искателем документами и если вы сконфигурировали портлет поиска документов для поиска документов в собраниях предприятия при помощи Portal Search Engine, Portal Search Engine может получить ID группы для зарегистрированного пользователя и передать маркер защиты для этого ID группы при запросе на поиск в поисковой системе предприятия. Маркер защиты позволяет добиться того, что в результатах поиска будут возвращаться только те документы, для просмотра которых у пользователя есть авторизация.

Если искатель связывает с документами другой тип маркера защиты, такой как ID пользователя или роль пользователя, и вы хотите принудительно включить защиту уровня документов при поиске в собраниях поисковой системы предприятия, необходимо создать пользовательский портлет поиска. Portal Search Engine использует только маркеры защиты для ID групп.

Внедрение портлета поиска

Портлет поиска ESSearchPortlet.war позволяет использовать WebSphere Portal Search Engine для собраний поисковой системы предприятия.

Прежде, чем вы начнете

Портлет поисковой системы предприятия необходимо размещать в WebSphere Portal до конфигурирования в Portal Search Engine использования этого портлета.

Об этой задаче

После размещения портлета поиска можно продолжать использовать Portal Search Engine для поиска индексированных данных в WebSphere Portal. Чтобы разрешить пользователям выполнять поиск в собраниях предприятия, администратор WebSphere Portal должен отредактировать свойства в конфигурации Portal Search Engine. Вы можете переключаться между этими двумя возможностями поиска и использовать более подходящее вам решение.

Файл ESSearchPortlet.war устанавливается в каталоге ES_INSTALL_ROOT/bin на серверах поиска при установке WebSphere II OmniFind Edition. Пути установки по умолчанию:

- Системы UNIX: /opt/IBM/es/bin/ESSearchPortlet.war
- Системы Windows: C:\Program Files\IBM\es\bin\ESSearchPortlet.war

Порядок действий

Чтобы внедрить портлет поиска в WebSphere Portal:

1. Установите портлет:
 - a. Зарегистрируйтесь на сервере WebSphere Portal под ID администратора сервера WebSphere Portal.
 - b. Нажмите кнопку **Управление** в полосе инструментов.
 - c. Выберите **Портлеты** в области навигации слева, затем выберите **Установить** в меню **Портлеты**.

- d. Нажмите кнопку **Просмотр**, выберите файл ESSearchPortlet.war в вашей системе и нажмите кнопку **Далее**.
 - e. На следующей странице нажмите кнопку **Установить**. После установки портлета будет выведено сообщение: APIN0005I: Портлеты успешно установлены.
2. Измените параметры портлета:
 - a. Выберите **Управление портлетами** в области навигации слева, затем выберите **IBM Enterprise Search**.
 - b. Когда справа от окна выбора появятся новые значки, выберите **Изменить параметры**.
 - c. В списке параметров портлета измените следующие параметры:

port Задайте номер порта Web-сервера WebSphere II OmniFind Edition Search. Значение по умолчанию - 80.

applicationName
Задайте имя допустимой программы поиска для поисковой системы предприятия. (Имена доступных программ поиска перечислены на странице Программы поиска на консоли управления поисковой системой предприятия.) Значение по умолчанию - Default; это программа поиска, которая всегда сконфигурирована на поисковых серверах.

hostname
Задайте полное имя хоста для поискового сервера WebSphere II OmniFind Edition.
 - d. Нажмите кнопку **Сохранить**, чтобы сохранить изменения, затем нажмите кнопку **Отмена**, чтобы выйти со страницы Изменить параметры.
 3. Создайте страницу портала для портлетов:
 - a. Выберите **Пользовательский интерфейс портала** в области навигации слева, затем выберите **Управление страницами**.
 - b. Выберите ссылку **Мой портал**.
 - c. Выберите **Новая страница**, задайте заголовок для страницы и нажмите кнопку **ОК**. После создания страницы появится сообщение: APPR0010I: страница *заголовок_страницы* успешно создана.
 - d. Нажмите кнопку **ОК**. Созданная страница появится в списке доступных страниц.
 4. Добавьте портлет поиска на страницу:
 - a. Щелкните по значку Правка (значок карандаша) для созданной страницы.
 - b. Выберите **Добавить портлеты**.
 - c. В текстовом поле **Искать**: введите Enterprise и нажмите кнопку **Поиск**.
 - d. Включите переключатель рядом с **IBM Enterprise Search** и нажмите кнопку **ОК**.
 - e. Нажмите кнопку **Готово**. Это действие указывает, что вы успешно установили портлет поиска и добавили его на страницу WebSphere Portal.
 5. Вызовите портлет:
 - a. Выберите **Мой портал** в полосе навигации в верхней части страницы.
 - b. Выберите страницу с заголовком, который вы задали при добавлении портлета в систему.

При первом обращении к странице портлета она может открываться медленно, потому что система должна скомпилировать Java Server Pages (файлы JSP) для портлета.

Конфигурирование портлета поиска и просмотра WebSphere Portal для поисковой системы предприятия

В WebSphere Portal предоставляется портлет, который можно использовать для поиска в собственных собраниях WebSphere Portal и их просмотра. Этот портлет можно сконфигурировать для поиска в собраниях поисковой системы предприятия.

Порядок действий

Чтобы сконфигурировать портлет поиска и просмотра для поиска в собраниях поисковой системы предприятия:

1. Остановите экземпляр сервера WebSphere Portal.
2. Скопируйте приведенные ниже файлы из каталога WebSphere II OmniFind Edition ES_INSTALL_ROOT/lib в каталог /WebSphere/PortalServer/shared/app на сервере WebSphere Portal:

```
esapi.jar  
siapi.jar
```

3. Запустите экземпляр сервера WebSphere Portal и зарегистрируйтесь на сервере WebSphere Portal с ID администратора WebSphere Portal.
4. Скопируйте портлет:
 - a. В верхнем правом углу выберите **Управление**, затем раскройте объект **Управление портлетами** и выберите **Портлеты**.
 - b. Найдите слово поиск.
 - c. Для портлета поиска и просмотра выберите **Копировать портлет**, задайте для портлета уникальное имя и нажмите кнопку **ОК**.
5. Сконфигурируйте портлет:
 - a. Щелкните по значку **Конфигурирование портлета** рядом с заданным вами именем портлета.
 - b. Нажмите кнопку со стрелкой вправо, чтобы перейти на следующую страницу параметров.
 - c. В поле **Новый параметр** введите ApplicationInfoId, в поле **Новое значение** задайте имя допустимой прикладной программы поиска WebSphere II OmniFind Edition, затем нажмите кнопку **Добавить**, чтобы добавить новый параметр.

Список имен доступных прикладных программ поиска выводится на странице Прикладные программы поиска на консоли администратора поисковой системы предприятия. Значение по умолчанию - Default; это программа поиска, которая всегда сконфигурирована на поисковых серверах.

- d. Удалите следующие параметры:

```
ApplicationInfoId_EXAMPLE  
QueryFactoryImp  
BrowseFactoryImp  
SearchFactoryImp  
IndexName  
EJB_Example  
IIOp_URL_Example  
SOAP_URL_Example  
EJB  
SOAP_URL  
IIOp_URL
```

- e. Добавьте следующие параметры и задайте значения, показанные ниже. Нажмите кнопку **Добавить**, чтобы добавить каждое имя и значение параметра. Имена параметров регистрозависимы.

```
IMPLEMENT = ES
ApplicationInfoId = имя_программы_поиска (например, Default)
SearchFactoryImp = com.ibm.es.api.search.RemoteSearchFactory
BrowseFactoryImp = com.ibm.es.api.browse.RemoteBrowseFactory
hostname = имя_хоста_сервера_поиска (например, omnifind.ibm.com)
port = номер_порта_сервера_поиска (например, 80)
```

6. Назначьте новую копию портлета поиска и просмотра для страницы портала. Если потребуется помощь, смотрите документацию по управлению WebSphere Portal.
7. Вызовите портлет:
 - a. В полосе навигации, в верхней части страницы портала, выберите **Мой портал**.
 - b. Выберите страницу в соответствии с именем, заданным вами при копировании портлета на шаге 4 на стр. 205.

Установка адаптера поисковой системы предприятия для Центра поиска

Перед тем, как использовать Центр поиска WebSphere Portal для поиска в собраниях поисковой системы предприятия, необходимо установить адаптер для поисковой системы предприятия.

Об этой задаче

Адаптер ESSearchAdapter.ear добавляет страницу для поисковой системы предприятия в интерфейс Центра поиска. После установки этой программы и портлета регистрации вы можете добавлять собрания поисковой системы предприятия в Центр поиска и выполнять поиск в этих собраниях при помощи портлета универсального поиска.

Файл ESSearchAdapter.ear устанавливается в каталоге ES_INSTALL_ROOT/bin на поисковых серверах при установке WebSphere II OmniFind Edition. Пути установки по умолчанию:

- Системы UNIX: /opt/IBM/es/bin/ESSearchAdapter.ear
- Системы Windows: C:\Program Files\IBM\es\bin\ESSearchAdapter.ear

Порядок действий

Чтобы установить адаптер для поисковой системы предприятия в Центре поиска:

1. Остановите экземпляр сервера WebSphere Portal.
2. Запустите экземпляр сервера WebSphere Application Server server1, если он еще не запущен.
3. Скопируйте приведенные ниже файлы из каталога WebSphere II OmniFind Edition ES_INSTALL_ROOT/lib в каталог /WebSphere/PortalServer/shared/app на сервере WebSphere Portal:
esapi.jar
siapi.jar

Если появится предложение перезаписать файл siapi.jar, выберите **Да**.

4. На сервере WebSphere Portal запустите консоль администратора WebSphere Application Server. Если вам предложат зарегистрироваться, зарегистрируйтесь.

Чтобы открыть консоль администратора, можно:

- Выбрать программу в меню **Пуск Windows**.

- В случае WebSphere Application Server версии 5 откройте браузер и перейдите по адресу `http://имя_хоста:порт/admin`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9090.
 - В случае WebSphere Application Server версии 6 откройте Web и перейдите по адресу `http://имя_хоста:порт/ibm/console`, где *имя_хоста* - имя хоста поискового сервера, а *порт* - номер порта для консоли администратора WebSphere Application Server. Обычно для консоли администратора используется порт 9060.
5. Выберите **Программы**, затем **Установить новую программу**.
 6. Нажмите кнопку **Просмотр**, выберите `ESSearchAdapter.ear` в вашей системе и нажмите кнопку **Далее** два раза. Если появится предупреждение о файлах правил, нажмите кнопку **Продолжить**.
 7. Продолжайте нажимать кнопку **Далее**, пока не увидите страницу **Отобразить модули на серверы программ**, выберите значения, как описано ниже, и нажмите кнопку **Применить**:
 - В текстовом поле **Кластеры и серверы** выберите `WebSphere:cell=имя_ячейки,node=имя_узла,server=сервер1`.
 - Включите переключатель рядом с модулем `ESSearchAdapterEJB`.
 8. Продолжайте нажимать кнопку **Далее**, пока не увидите страницу **Сводка**, затем нажмите кнопку **Готово**.
 9. Щелкните по ссылке **Сохранить в главной конфигурации**, затем нажмите кнопку **Сохранить**, чтобы сохранить изменения в конфигурации WebSphere Application Server.
 10. Перезапустите сервер WebSphere Portal.

Установка портлета регистрации поисковой системы предприятия для Центра поиска

Чтобы Центр поиска WebSphere Portal можно было использовать для поиска в собраниях поисковой системы предприятия, сначала нужно установить портлет регистрации для поисковой системы предприятия.

Об этой задаче

Портлет регистрации `ESSearchAdapterPortlet.war` регистрирует портлет поисковой системы предприятия в Центре поиска. После установки этого портлета (и программы адаптера для поисковой системы предприятия) можно добавить в Центр поиска собрания поисковой системы предприятия и выполнять в них поиск при помощи портлета универсального поиска.

Файл `ESSearchAdapterPortlet.war` устанавливается в каталог `ES_INSTALL_ROOT/bin` на серверах поиска при установке WebSphere II OmniFind Edition. Пути установки по умолчанию:

- В системах UNIX: `/opt/IBM/es/bin/ESSearchAdapterPortlet.war`
- В системах Windows: `C:\Program Files\IBM\es\bin\ESSearchAdapterPortlet.war`

Порядок действий

Чтобы установить портлет регистрации для поисковой системы предприятия:

1. Зарегистрируйтесь на сервере WebSphere Portal с ID администратора WebSphere Portal.
2. Нажмите кнопку **Управление** в полосе инструментов.
3. Выберите **Портлеты** в области навигации слева, затем выберите **Установить** в меню **Портлеты**.
4. Нажмите кнопку **Просмотр**, выберите в системе файл ESSearchAdapterPortlet.war и нажмите кнопку **Далее**.
5. На следующей странице нажмите кнопку **Установить**. После установки портлета будет выведено сообщение: APIN0005I: Портлеты успешно установлены.
6. Назначьте портлет (Портлет регистрации IBM WebSphere II OmniFind Edition для поисковой системы предприятия) для страницы портала. Если потребуются помощь, смотрите документацию по управлению WebSphere Portal.

Перенастройка из WebSphere Portal в поисковую систему предприятия

В поисковой системе предприятия есть мастер по перенастройке, который можно использовать для перенастройки таксономий и собраний из IBM WebSphere Portal в поисковую систему предприятия.

Для перенастройки таксономий и собраний нужно запустить мастер по перенастройке на сервере индексов поисковой системы предприятия. После перенастройки таксономии ее можно использовать с собраниями поисковой системы предприятия. Можно также с помощью поисковой системы предприятия управлять собраниями, перенастроенными из WebSphere Portal, и выполнять поиск в этих собраниях.

В поисковой системе предприятия таксономия называется деревом категорий. После перенастройки таксономии на основе правил для управления деревом категорий можно использовать консоль администратора поисковой системы предприятия. Чтобы в поисковой системе предприятия использовать таксономии на основе моделей, на сервере индексов нужно установить WebSphere Portal.

Перенастраивая таксономии и собрания, в первую очередь всегда перенастраивайте файлы таксономий на основе моделей, а затем собрания. Если не соблюдать это правило, категоризация на основе моделей не будет работать с собраниями, перенастроенными из WebSphere Portal.

Перенос таксономии на основе модели из WebSphere Portal

При помощи портлета управления таксономией WebSphere Portal можно выбрать таксономию на основе модели для использования с собранием поисковой системы предприятия. На собрания, уже перенастроенные в поисковую систему предприятия, перенастройка новой таксономии не влияет.

Прежде, чем вы начнете

Перед первым запуском мастера по перенастройке остановите поисковую систему предприятия, чтобы можно было внести изменения в файлы конфигурации. При первом запуске мастера по перенастройке надо задать путь установки WebSphere Application Server и WebSphere Portal. При первом запуске мастера по перенастройке допустимо, чтобы поисковая система предприятия была активна.

Об этой задаче

Чтобы перенастроить таксономию, ее нужно выбрать и экспортировать из WebSphere Portal. Затем при помощи мастера по перенастройке поисковой системы предприятия нужно перенастроить эту таксономию в поисковую систему предприятия.

Порядок действий

Чтобы перенастроить таксономию на основе модели из WebSphere Portal в поисковую систему предприятия:

1. Экспортируйте текущую таксономию на основе модели из портлета управления таксономией WebSphere Portal. Таксономию составляют следующие файлы XML:

synonyms.xml
titles.xml
treenodes.xml

2. Скопируйте эти файлы на сервер индексов поисковой системы предприятия.
3. На этом сервере зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
4. Перейдите в каталог установки поисковой системы предприятия:

UNIX: cd \$ES_INSTALL_ROOT/bin
Windows: cd %ES_INSTALL_ROOT%\bin

5. Если это первый запуск мастера по перенастройке, введите команду остановки поисковой системы предприятия:

esadmin stop

6. Введите следующую команду, чтобы запустить мастер по перенастройке, затем нажмите кнопку **Далее**.

UNIX: ./eswpsmigrate.sh
Windows: eswpsmigrate.bat

Если это не первый запуск мастера по перенастройке, и поисковая система предприятия активна, учтите, что пока работает мастер по перенастройке, нельзя создавать собрания поисковой системы предприятия.

7. Выберите **Импортировать файлы таксономии на основе модели из WebSphere Portal**, затем нажмите кнопку **Далее**.
8. При первом запуске мастера по перенастройке задайте пути установки сервера WebSphere Application Server и WebSphere Portal. Нажмите кнопку **Далее**.
9. Найдите каталог, содержащий файлы таксономии на основе модели, выберите файлы XML, перенастраиваемые для использования таксономии на основе модели (synonyms.xml, titles.xml и treenodes.xml), затем нажмите кнопку **Далее**.
10. Если это был первый запуск мастера по перенастройке, по завершении работы введите команду перезапуска поисковой системы предприятия:
esadmin start

В случае ошибок смотрите файл MigrationWizard.log, находящийся в каталоге установки мастера по перенастройке.

Перенос собрания из WebSphere Portal

Чтобы перенастроить собрания из WebSphere Portal в поисковую систему предприятия, подготовьте эти собрания в WebSphere Portal, затем перенастройте их при помощи мастера по перенастройке.

Прежде, чем вы начнете

Если вы планируете перенастроить таксономии на основе модели и собрания, перед данной использованием данной процедуры для перенастройки собраний необходимо перенастроить файлы таксономии на основе модели. Если не соблюсти это правило, категоризация на основе моделей не будет работать с собраниями, перенастроенными из WebSphere Portal.

Порядок действий

Чтобы перенастроить собрание из WebSphere Portal в поисковую систему предприятия:

1. В WebSphere Portal Search Engine остановите в перенастраиваемых собраниях все процессы искателей и примите или отклоните все приостановленные документы. (В поисковой системе предприятия понятие приостановленный документ не поддерживается.)
2. Для каждого перенастраиваемого собрания при помощи портлетов Portal Search Engine экспортируйте параметры в файлы XML.
3. Если сервер индексации поисковой системы предприятия установлен на отдельном сервере, скопируйте на него экспортированные файлы XML.
4. На этом сервере зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
5. Перейдите в каталог установки поисковой системы предприятия:

```
UNIX: cd $ES_INSTALL_ROOT/bin
Windows: cd %ES_INSTALL_ROOT%\bin
```

6. Введите следующую команду, чтобы запустить мастер по перенастройке, затем нажмите кнопку **Далее**.

```
UNIX: ./eswpsmigrate.sh
Windows: eswpsmigrate.bat
```

7. Выберите **Перенастроить параметры поиска из Portal Search Engine в WebSphere Portal**, затем нажмите кнопку **Далее**.
8. Найдите каталог, содержащий экспортированные файлы конфигурации Portal Search Engine, выберите файлы для перенастройки, затем нажмите кнопку **Далее**. Выбранные файлы конфигурации будут проанализированы и проверены.
9. Введите для каждого собрания приведенную ниже информацию, затем нажмите кнопку **Далее**, чтобы перенастроить собрания в поисковую систему предприятия:
 - Имя собрания для использования в поисковой системе предприятия.
 - Критерий, по которому будет определяться важность документов для собрания. Можно ничего не выбрать, выбрать коэффициент статического ранжирования на основе состояний документов или коэффициент статического ранжирования на основе ссылок на документы Web из других документов Web.
 - Тип категорий, используемых для данного собрания. Можно ничего не выбрать, выбрать категории на основе правил или категории на основе моделей.

Если выбрать категории на основе правил, в поисковую систему предприятия будут перенастроены таксономия и правила собраний WebSphere Portal.

Категории на основе моделей можно выбрать, только если прежде вы перенастроили из WebSphere Portal в поисковую систему предприятия таксономию на основе модели.

Если при перенастройке происходят ошибки, посмотрите файл MigrationWizard.log, находящийся в каталоге установки мастера по перенастройке.

Теперь можно использовать консоль администратора поисковой системы предприятия, чтобы сконфигурировать дополнительные параметры для перенастроенных собраний.

Требование: При конфигурировании свойств искателя Web для перенастроенного собрания нужно задать адрес электронной почты, на который будут

отправляться комментарии об искателе, и имя пользовательского агента (если при конфигурировании свойств искателя Web потребуется помощь, нажмите кнопку **Справка**).

10. Запустите для перенастроенных собраний процессы искателей, синтаксического анализа и индексации с консоли администратора поисковой системы предприятия.
11. Определив, что перенастроенное собрание доступно для поиска в поисковой системе предприятия, удалите исходное собрание из Portal Search Engine.
12. Необязательно: Как администратор WebSphere Portal, выполните следующие действия, если хотите разрешить пользователям выполнять поиск в перенастроенном собрании с портала в WebSphere Portal.
 - a. Внедрите портлет поисковой системы предприятия в WebSphere Portal.

В кластере сервера WebSphere Portal это можно сделать с сервера, где установлен менеджер внедрения WebSphere Application Server. Менеджер внедрения распространяет портлет поисковой системы предприятия на другие серверы в кластере сервера WebSphere Portal.
 - b. Добавьте портлет поисковой системы предприятия на подходящие страницы портала.

В WebSphere Portal управление доступом для портлета поиска моделируется по доступу к конкретным страницам и портлетам. Параметры собрания перенастраиваются, но портлет поиска должен быть установлен вручную администратором сервера WebSphere Portal.

Перенос параметров собраний

При перенастройке собраний из IBM WebSphere Portal мастер по перенастройке создает параметры по умолчанию для собраний и искателей.

Если один и тот же параметр существует и в собраниях Portal Search Engine, и в собраниях поисковой системы предприятия, при перенастройке собрания в поисковую систему предприятия мастер использует параметр из Portal Search Engine. Для параметров, существующих только в поисковой системе предприятия, мастер использует значения, задаваемые вами при перенастройке собрания или значения, используемые для собраний поисковой системы предприятия по умолчанию.

Параметры, существующие в Portal Search Engine и в поисковой системе предприятия

Мастер по перенастройке переносит для каждого перенастраиваемого собрания следующие параметры:

- Сайты Portal Search Engine в собрании Portal Search Engine
- Язык собрания
- Таксономию (или дерево категорий) и правила для категорий на основе правил, если в собрании поисковой системы предприятия используется категоризация на основе правил

Каждый сайт Portal Search Engine в собрании объединяется в искатель Web поисковой системы предприятия. Мастер по перенастройке переносит следующие параметры искателя:

- Начальный URL
- Число параллельных процессов искателя
- Глубину просмотра

- Срок ожидания (в секундах) при получении документа
- Набор символов по умолчанию
- Правила включения и исключения для просмотра искателем

Параметры, существующие только в поисковой системе предприятия

При перенастройке собрания нужно задать информацию о нем. Мастер по перенастройке перенастраивает эти параметры и использует параметры по умолчанию для собраний в поисковой системе предприятия, конфигурируя каждое перенастраиваемое вами собрание.

Конфигурации собраний и искателей Web можно изменить при помощи консоли администратора поисковой системы предприятия. В круглых скобках () показаны значения по умолчанию для перенастраиваемых данных.

- Имя собрания
- Стратегия статического ранжирования документов
- Тип используемой категоризации, например, на основе правил или ничего
- Используется ли кэш поиска и сколько запросов с результатами поиска может в нем храниться (да, 5000)
- Выполняется ли мониторинг времени отклика поиска и генерируется ли оповещение при превышении предела (да, 5 секунд)
- Используется ли управление доступом (нет)
- Расписание для обновления индекса
- Расписание для реорганизации индекса
- Уровень детализации журнала (все сообщения)

Кроме того, мастер по перенастройке создает для каждого искателя Web следующие параметры:

- Имя искателя
- Описание искателя
- MPL (maximum page length - максимальная длина страницы)
- Параметры защиты документов
- Типы MIME (Multipurpose Internet Mail Extension) для документов, с которыми должен работать искатель (если это применимо к типу источника данных)

Прежде чем запускать перенастроенный искатель Web, проверьте все свойства этого искателя и параметры пространства для искателя и убедитесь, что заданы все необходимые значения (обязательные поля отмечены красной звездочкой). В частности, убедитесь, что заданы адрес электронной почты для получения комментариев об искателе и имя пользовательского агента для искателя. Если при конфигурировании свойств искателя Web потребуется помощь, нажмите кнопку **Справка**.

Файл журнала мастера по перенастройке

Мастер по перенастройке записывает все сообщения в файл WpsMigratorLog.log, находящийся в его каталоге установки.

Для каждого перенастроенного собрания файл журнала WpsMigratorLog.log содержит значения для всех параметров, прочитанных из WebSphere Portal Search Engine, и указывает, в какое место собраний поисковой системы предприятия они были импортированы.

Запуск и остановка серверов поисковой системы предприятия

После создания собрания надо запустить серверы искателей, анализа, индексации и поиска данных. После внесения изменений в собрание остановите и перезапустите эти серверы.

Большинство серверов поисковой системы предприятия могут работать непрерывно или в соответствии с заданными расписаниями. Например, можно задать расписания реорганизации или обновления индекса. После запуска серверов анализа данных и поиска в индексе обычно останавливать и перезапускать их нужно, только если изменены параметры конфигурации (например, изменены категории или увеличен размер кэша поиска).

При изменении содержимого собрания или правил сбора искателями данных из источников в вашем предприятии обычно для вступления изменений в силу нужно остановить и снова запустить искатели. Если не изменять правила искателей, они работают или непрерывно (в случае искателей Web и NNTP), или в соответствии с заданными расписаниями.

Запуск серверов поисковой системы предприятия

Чтобы пользователи могли выполнять поиск в собрании, нужно запустить сначала процессы системы, а затем серверы искателя, анализатора, индексации и поиска для этого собрания.

Прежде, чем вы начнете

Сконфигурируйте источники данных, с которыми должен работать искатель, и задайте опции, определяющие, как выполнять синтаксический анализ, индексацию и поиск данных. Например, если вы хотите, чтобы пользователи могли просматривать в результатах поиска подробную информацию о категориях, сконфигурируйте категории перед запуском анализатора.

Запустить серверы поисковой системы предприятия может только администратор поиска предприятия, администратор собрания или оператор с полномочиями управления этим собранием.

Серверы поисковой системы предприятия нужно запускать для собрания в правильном порядке. Например, сначала необходимо запустить искатель и просмотреть данные; только после этого можно будет выполнять синтаксический анализ документов и индексировать документы.

Порядок действий

Чтобы запустить серверы поисковой системы предприятия:

1. Чтобы использовать консоль администратора поисковой системы предприятия или программу поиска:
 - a. Запустите IBM HTTP Server, если он еще не запущен.
 - b. При помощи консоли администратора WebSphere Application Server запустите прикладные программы предприятия ESAdmin и ESSearchApplication (если они еще не запущены).

2. Если CCL (common communication layer - общий слой связи) поисковой системы предприятия не запущен, запустите его:
 - a. На сервере индексов зарегистрируйтесь под ID администратора поисковой системы предприятия, заданным при установке WebSphere II OmniFind Edition.

Системы UNIX


Введите команду:

```
startccl.sh -bg
```

Системы Windows

Выберите **Пуск** → **Программы** → **Администрирование** → **Службы** и перезапустите службу IBM WebSphere Information Integrator OmniFind Edition.


3. Если поисковая система предприятия не запущена,
 - a. На сервере индексов зарегистрируйтесь под ID администратора поисковой системы предприятия, заданным при установке WebSphere II OmniFind Edition.
 - b. Введите команду:

```
esadmin start
```
4. Запустите консоль администратора поисковой системы предприятия и зарегистрируйтесь как администратор поисковой системы предприятия. Если используются административные роли, можно зарегистрироваться как администратор собрания или как оператор с полномочиями управления собранием, которое нужно запустить.
5. В окне Собрания найдите нужное собрание и нажмите кнопку  **Слежение**.
6. На странице Искатель нажмите для каждого запускаемого искателя кнопку


Запустить.

- Если вы запускаете искатель Web или NNTP, этот искатель сразу начинает просмотр данных. Эти типы искателей работают непрерывно и выполняют просмотр и повторный просмотр сайтов и групп новостей NNTP.
- Если вы запускаете искатель другого типа, запускается сеанс искателя. Искатель начнет просмотр данных согласно дате и времени в расписании. Если вы не задали расписание или хотите запустить искатель раньше, включите мониторинг искателя и щелкните по значку запуска для каждого источника данных, который должен обработать искатель.

После запуска искателя можно позволить ему работать непрерывно. Если искатель есть в расписании, он будет снова запускаться в заданные в расписании даты и моменты времени.

7. После того, как искатель выполнит просмотр данных, откройте страницу Анализ и нажмите кнопку  **Запустить**, чтобы запустить анализатор.

Можно позволить анализатору работать непрерывно. Обычно анализатор нужно останавливать только после внесения изменений в обработку им данных (например, после изменения категорий или отображения полей XML).

8. Необязательно: Чтобы принудительно запустить процесс индексации, не дожидаясь запланированных для этого даты и времени, откройте страницу Индекс и в области **Реорганизация** нажмите кнопку  **Запустить**.

Можно позволить процессу индексации работать непрерывно. Индекс будет обновляться и реорганизовываться в запланированные даты и моменты времени.

9. На странице Поиск нажмите кнопку  **Запустить**.

Можно позволить серверам поиска работать непрерывно. Обычно серверы поиска не требуется останавливать, за исключением случаев изменения параметров кэша поиска или сводки документов.

Остановка серверов поисковой системы предприятия

Остановка и перезапуск сервера поисковой системы предприятия может потребоваться, если вы внесли изменения в его конфигурацию или хотите исправить ошибки.

Прежде, чем вы начнете

Остановить серверы поисковой системы предприятия может только администратор поисковой системы предприятия, администратор собрания или оператор с полномочиями управления этим собранием.

Об этой задаче


Серверы поисковой системы предприятия можно остановить независимо друг от друга. Например, если вы останавливаете и перезапускаете искатель, чтобы он использовал внесенные в его конфигурацию изменения, не нужно останавливать и перезапускать анализатор.

Если нужно остановить всю систему поиска предприятия, а не отдельные ее серверы, зарегистрируйтесь на сервере индексов с ID администратора поиска предприятия (этот ID был задан при установке WebSphere II OmniFind Edition). Затем введите команду:

```
esadmin stop
```

Порядок действий

Чтобы остановить серверы поисковой системы предприятия:


1. В окне Собрания найдите нужное собрание и нажмите кнопку  **Monitor**.
2. На странице Искатель найдите нужный искатель и остановите или приостановите его.


Если вы изменили свойства искателя, остановите и перезапустите искатель, чтобы он использовал эти изменения. Если вы изменили пространство искателя и хотите применить изменения к документам, которые уже содержит собрание, нужно также выполнить их повторный просмотр искателем.


Совет: Сообщение о том, что для затребованной операции превышен срок ожидания, может выводиться, даже если процесс продолжает выполняться в фоновом режиме. Чтобы узнать, завершена ли эта задача, нажмите кнопку **Обновить** на консоли администратора (не нажимайте кнопку **Обновить** в браузере). Процесс завершен, если значок состояния для искателя указывает, что он остановлен.

3. На странице Анализ нажмите кнопку  **Остановить**, чтобы остановить анализатор.

Если вы изменили правила обработки анализатором данных, остановите и перезапустите анализатор, чтобы он использовал эти изменения. Изменения применяются только к новым просматриваемым искателем документам. Если требуется применить изменения к документам, которые уже внесены в индекс, нужно запустить полный просмотр искателем, чтобы выполнить повторный просмотр искателем этих документов. После просмотра можно будет проанализировать и проиндексировать заново.

4. На странице Индекс нажмите кнопку  **Остановить**, чтобы остановить обновляемый в настоящий момент индекс.

Можно также остановить построение индекса при отслеживании очереди заданий. Для этого выберите на панели инструментов **Система**, откройте страницу Индекс и нажмите кнопку  **Остановить** для индекса, построение которого вы хотите остановить.

5. На странице Поиск нажмите кнопку  **Остановить**, чтобы остановить серверы поиска. Обычно остановка и перезапуск серверов поиска требуется только при изменении параметров кэша или сводки документов.

Слежение за активностью поисковой системы предприятия

При слежении за активностью системы и собраний можно просмотреть состояния различных процессов, отследить возможные ошибки и настроить параметры конфигурации для улучшения производительности.

С помощью консоли администратора поисковой системы предприятия можно следить за системой и при необходимости настроить ее работу. Можно просмотреть подробную статистику по каждому из основных процессов (таких как просмотр искателем, анализ, индексация, поиск). В статистику входит среднее время отклика и информация о ходе выполнения, например, число документов, просмотренных искателем или проиндексированных при построении индекса.

Щелкнув по соответствующим значкам, можно остановить и запустить большинство процессов. Эти операции позволяют приостановить процесс, внести изменения в его конфигурацию или исправить ошибку и перезапустить обработку, когда вы будете готовы ее продолжить.

Оценка числа документов в собрании

Создавая или изменяя собрание поисковой системы предприятия необходимо задавать предположительное число документов в этом собрании. Менеджер ресурсов использует это число для оценки необходимых ресурсов памяти и диска; это число не используется для ограничения размера собрания.

Прежде, чем вы начнете

Изменить оценку размера собрания может администратор поисковой системы предприятия или администратор этого собрания.

Об этой задаче

Когда размер собрания достигнет оцененного вами значения, система не прекратит добавлять документы в индекс. Менеджер ресурсов генерирует предупреждения, когда оценка необходимых ресурсов памяти и диска превышает объем ресурсов, доступных в системе в настоящее время. Эти предупреждения позволят предотвратить будущие ошибки из-за нехватки ресурсов.

Если вы сконфигурировали оповещения для собрания и разрешили уведомление, когда число документов в индексе превысит ограничение, по умолчанию это ограничение соответствует значению, заданному как оценка числа документов в собрании. Согласно этому числу и заданному вами процентному порогу оповещения система слежения будет отправлять сообщения электронной почты, когда число документов собрания приблизится к максимальному.

Порядок действий

Чтобы задать оценку размера собрания:

1. Откройте собрание для редактирования, откройте страницу Анализ и нажмите кнопку **Конфигурировать общие опции**.

2. В поле **Примерное число документов** введите ожидаемое число документов в собрании. Значение по умолчанию - 1 миллион документов.

Проверка доступности системных ресурсов

Если проверить системные ресурсы после создания собрания или искателя, но до запуска поиска документов, можно определить, позволяют ли доступные при текущих параметрах конфигурации системные ресурсы использовать WebSphere II OmniFind Edition на полную мощность.

Прежде, чем вы начнете

Проверять системные ресурсы может администратор поисковой системы предприятия.

Об этой задаче

Когда вы создаете собрание или искатель, система автоматически проверяет доступность ресурсов. Можно также включить опцию проверки доступности системных ресурсов в любое удобное время.

Система сравнивает оценку размера пространства, которое будет использовано при заданных свойствах собрания и искателя, с доступным системным пространством, а затем выводит сообщение о доступности ресурсов. В сообщении указывается, у какого из серверов может не хватить ресурсов, оценка необходимого размера собрания и доступное свободное пространство.

Если ресурсов недостаточно, попробуйте исправить ситуацию, выполнив одно из следующих действий:

- Увеличьте размер файловой системы, указанной в тексте сообщения.
- Отредактируйте для собрания основные опции и укажите меньшее число предполагаемых документов.
- Отредактируйте свойства искателя и укажите меньшие значения для максимального числа обрабатываемых им документов, максимального размера страницы и максимального числа потоков.
- Отредактируйте свойства искателя Web, указав меньшие значения для максимального числа активных хостов, максимального числа новых документов и максимального числа документов во временном хранилище.
- Отредактируйте свойства индекса для системы, уменьшив число одновременно разрешенных запусков построений индексов.
- Избегайте одновременной активности слишком большого числа процессов в нескольких собраниях. Например, на ресурсы может оказывать влияние одновременная работа искателей и анализаторов для нескольких собраний.
- Удалите искатели из системы.
- Удалите собрания из системы.

Если сообщения о возможном дефиците ресурсов не обнаружены, скорее всего, ресурсов хватит для размещения и индексирования найденных данных.

Порядок действий

Чтобы определить, достаточно ли ресурсов доступно при текущей конфигурации поисковой системы предприятия:

1. Выберите **Система**, чтобы открыть окно Система.

2. На странице Общие нажмите кнопку **Проверить системные ресурсы**.

Слежение за собранием


Вы можете просмотреть общую информацию о состоянии каждого компонента в собрании или выбрать опции для просмотра подробной информации об отдельных компонентах и URI.


Прежде, чем вы начнете


Все пользователи-администраторы поисковой системы предприятия могут наблюдать за собраниями. Запускать или останавливать компоненты, а также включать и отключать расписания может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

Порядок действий

Чтобы наблюдать за собранием:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**. Появится информация о текущем состоянии каждого из компонентов собрания.

Совет: Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Наблюдатель**, чтобы перейти в окно слежения за собранием.

2. Для просмотра подробной информации об URI выберите  **Подробности URI**. Например, вам может понадобиться узнать, находится ли определенный URI в индексе или скопирован ли на поисковые серверы индекс, в котором находится URI.
3. Для наблюдения за отдельным компонентом и просмотра подробной статистической информации о работе этого компонента щелкните по значку **Состояние**.

Просмотр подробностей об URI

Можно просматривать подробную информацию об URI. Можно просмотреть текущую информацию и хронологическую информацию о том, как выполняется просмотр, индексирование и поиск для документа, представленного этим URI.

Прежде, чем вы начнете

Прежде чем отправить требование на просмотр отчета об URI или послать отчет по адресу электронной почты, убедитесь, что компонент, от которого вы хотите получить информацию, активен. Например, чтобы увидеть подробности о просмотре, индексации и поиске документа, убедитесь, что искатель Web, сервер индексов и серверы поиска запущены. Чтобы отслеживать отброшенный документ, убедитесь, что сконфигурированы опции записи в журналы для отслеживания документов.

Об этой задаче

Процесс сбора информации об URI отнимает много времени. Можно выбрать опцию просмотра затребованной информации и затем подождать ее вывода. Более эффективна опция отправки отчета по указанному вами электронному адресу.


Чтобы можно было получить отчет, нужно сконфигурировать в поисковой системе предприятия информацию о почтовом сервере. Эта информация задается при конфигурировании опций электронной почты на странице Журнал окна Система.


Сервер индексов и серверы поиска могут давать информацию обо всех URI (например, находится ли некоторый URI в индексе и был ли он скопирован на серверы поиска). Чтобы увидеть информацию о том, как просматривался документ, надо задать URI для документа, который просматривался искателем Web.

Порядок действий

Чтобы просмотреть подробности об URI:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.

Совет: Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Наблюдатель**, чтобы перейти в окно слежения за собранием.

2. Нажмите кнопку  **Подробности URI**.
3. На странице Подробности URI введите URI, для которого нужно просмотреть информацию.
4. Включите переключатели для типов информации, которую вы хотите получить:

Подробности искателя (доступно только для искателей Web)

Включите этот переключатель, чтобы увидеть информацию о том, как документ просматривался искателем Web, и информацию о его текущем состоянии в пространстве для искателя.

Подробности индекса

Включите этот переключатель, чтобы увидеть, был ли документ индексирован и скопирован на серверы поиска.

Подробности поиска

Включите этот переключатель, чтобы увидеть информацию о том, доступен ли документ для поиска и как можно выполнить такой поиск.

Документы, отброшенные синтаксическим анализатором

Включите этот переключатель, чтобы увидеть, был ли документ отброшен из поисковой системы предприятия во время его синтаксического анализа, и, если это так, почему он был отброшен.

Документы, отброшенные из индекса

Включите этот переключатель, чтобы увидеть, был ли документ отброшен из поисковой системы предприятия во время его индексирования и лексического анализа, и, если это так, почему он был отброшен.

5. Чтобы вывести на экран отчет, нажмите кнопку **Просмотреть отчет**.
6. Чтобы послать отчет по электронной почте для последующего просмотра, нажмите кнопку **Послать отчет**.
 - a. На странице Послать подробный отчет об URI введите электронный адрес для получения отчета в поле **Адрес электронной почты для уведомления**.
 - b. Нажмите кнопку **Послать отчет**.

Отслеживание работы искателей


Вы можете просматривать общую информацию о состоянии каждого искателя в собрании или выбрать опции для просмотра подробной информации о работе искателей.


Прежде, чем вы начнете


Если ваша административная роль разрешает вам только следить за собранием, вы можете просматривать статистику искателя, но не можете изменять поведение искателя (например, запускать и останавливать его).

Порядок действий

Чтобы следить за искателем:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

Совет: Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель работает или приостановлен, и вы хотите увидеть подробную информацию о состоянии этого искателя, выберите  **Подробности**. Типы выводимых при этом статистических показателей зависят от типа искателя.

Если ваша административная роль позволяет управлять процессами для собрания, вы можете запускать, останавливать и приостанавливать искатель при просмотре подробностей о его действиях. Если для искателя можно задать расписание, вы можете также включить или отключить расписание просмотра.

4. Если искатель остановлен или приостановлен, и вам нужно запустить сеанс искателя, выберите  **Запустить** или **Возобновить**.

Для искателей Web:

Если искатель был остановлен, он начнет просмотр снова по всему пространству для искателя. Если искатель был приостановлен, он возобновит просмотр, начиная с того назначения, где он был приостановлен.

Если вы хотите, чтобы искатель начал полный просмотр немедленно, щелкните по значку **Подробности**, затем по значку **Начать полный повторный просмотр**. Искатель начинает просмотр всего пространства искателя, включая страницы, которые не изменились со времени последнего просмотра. Повторный просмотр всех документов может потребоваться, например, после изменения правил синтаксического анализа документов, если нужно применить эти правила к проиндексированным ранее документам.

Для искателей NNTP:

Если искатель был остановлен, он начнет просмотр снова по всему пространству для искателя. Если искатель был приостановлен, он возобновит просмотр, начиная с того назначения, где он был приостановлен.

Для остальных типов искателей:

Если искатель был остановлен, он начнет просмотр в момент времени, предусмотренный расписанием. При первом просмотре источника данных искатель выполняет полный просмотр. При повторном запуске просмотра

по расписанию искатель просматривает либо все изменения в источнике данных (добавления, удаления и изменения документов), либо только добавления и изменения документов. Тип просмотра конфигурируется в расписании искателя.

Если вы не задали расписания искателя или хотите запустить искатель раньше, щелкните по значку **Подробности**. Затем в области подробностей пространства искателя щелкните по значку типа просмотра, который хотите запустить: полный просмотр, все изменения или только новые и измененные документы. Необходимо выбрать соответствующий значок запуска для каждого источника данных, который нужно просматривать (например, сервера, базы данных, подпапки).

5. Если искатель запущен и нужно остановить его, выберите  **Остановить** или

 **Приостановить**. Искатель прекращает просмотр данных, пока вы его не перезапустите или не возобновите его работу.

Если вы возобновляете работу приостановленного искателя, просмотр начинается с того назначения, где он был приостановлен. Например, искатель DB2 возобновит просмотр с первой строки таблицы, которую просматривал в момент приостановки.

Просмотр подробностей об активности искателя Web


Просматривая подробности об активности искателя Web, можно оценить общую производительность и нужным образом настроить свойства искателя Web и определения пространства для искателя.

Прежде, чем вы начнете


Все пользователи-администраторы могут следить за действиями искателя. Запустить или остановить искатель может администратор поисковой системы предприятия, администратор собрания или оператор этого собрания.

Порядок действий

Чтобы просмотреть подробности о работе искателя Web:

1. В окне Собрания найдите собрание, владеющее искателем Web, за которым вы хотите следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

Совет: Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель Web, который нужно отслеживать, запущен или приостановлен, выберите  **Подробности**.
4. На странице подробностей для искателя Web просмотрите или выберите указанные ниже опции, чтобы увидеть подробную статистику о текущих и прошлых действиях искателя.
 - a. Нажмите кнопку **Подробности потока**, чтобы увидеть, сколько потоков активно просматривают сайты, и сколько потоков неактивно.
 - b. Нажмите кнопку **Активные сайты**, чтобы увидеть информацию о сайтах, в данный момент просматриваемых искателем.
 - c. Нажмите кнопку **Недавно просмотренные URL**. Эта информация показывает последние просмотренные искателем объекты. Если элементы в этом списке не изменяются после обновления окна, искатель не выполняет просмотр.

- d. Нажмите кнопку **Хронология искателя**, чтобы посмотреть отчеты о предыдущих действиях искателя.
- e. В области **Состояние URL** введите URL для сайта, о котором хотите увидеть информацию, затем нажмите кнопку **Посмотреть**.

Например, с помощью этой опции можно узнать, находится ли данный URL в пространстве для искателя, был ли он просмотрен или только найден, когда его следует просматривать снова, а также информацию о последней попытке просмотра сайта.

После того, как будут выведены подробности об URL, нажмите кнопку **Хронология сайта**, чтобы увидеть дополнительную статистическую информацию о деятельности искателя по этому URL.

Подробности потока искателя Web

При наблюдении за работой искателя Web можно видеть число потоков, активно просматривающих сайты Web, и число потоков в состоянии неактивности.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать состояние потоков искателя. Наиболее часто выводятся следующие состояния:

Ожидание

Означает, что у потока нет URL для просмотра. Это состояние возможно, если поток завершил просмотр URL, а искатель не успевает достаточно быстро найти другие URL. Например, если значение свойства искателя, управляющего его временем ожидания перед получением другой страницы с того же сайта, слишком велико, URL могут поступать с недостаточной скоростью.

Выборка

Означает, что поток загружает страницу с сайта.

Завершен

Означает, что поток отправляет остальному искателю просмотренные страницы, но еще не готов к просмотру нового URL.

Приостановлен

Указывает на то, что искатель приостановлен.

В идеальном случае все потоки все время получают страницы. Если потоки часто находятся в состоянии "Завершен", возможно, у базы данных есть проблемы с пропускной способностью.

Если потоки часто находятся в состоянии ожидания, измените значение, заданное в поле **Максимальное число активных хостов** в свойствах искателя. Если это значение невелико, возможно, в пространстве для искателя недостаточно сайтов, чтобы занять все потоки, или недостаточно URL, пригодных для просмотра. Низкая активность может быть вызвана ошибками поиска DNS и ошибками поиска роботов.

Активные сайты искателя Web

За работой искателя Web можно наблюдать, чтобы получать информацию о сайтах, которые в данный момент обрабатывает искатель.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать статистическую информацию об активных сайтах. В статистике показано:

- Сколько URL для просмотра искатель получил в память из внутренней базы данных в этот раз
- Сколько URL искатель попытался просмотреть на данный момент
- Сколько времени остается до деактивации сайта и его удаления из памяти для этой итерации искателя
- Сколько времени сайт уже находится в памяти

Эта информация постоянно изменяется по мере работы искателя по сконфигурированным для него правилам искателя. В идеале число активированных URL должно приближаться к значению, сконфигурированному в поле **Максимальное число активных хостов** в свойствах искателя.

Если число активированных URL близко к нулю, искатель не находит нужных URL. Такая низкая активность может быть вызвана ошибками поиска DNS, проблемами со связью в сети, ошибками в базах данных и ошибками в определении пространства для искателя. Например:

- Если много сайтов долго находятся в памяти, но обработано мало URL, проверьте наличие проблем со связью в сети.
- Если в списке недостаточно сайтов, проверьте наличие ошибок в определении пространства для искателя или ошибок поиска DNS.
- Если искатель просматривает сайты с приемлемой скоростью, но оставляет в памяти много непросмотренных URL, отредактируйте свойства искателя и исправьте значение срока ожидания в поле **Максимальное время хранения URL в памяти**, чтобы сайты дольше оставались в памяти.

Скорость просмотра искателя Web

За работой искателя Web можно наблюдать, чтобы получать информацию о том, с какой скоростью искатель загружает страницы с сайтов.

При просмотре подробной информации об искателе Web во время наблюдения за собранием можно просматривать статистическую информацию о скорости просмотра данных искателем (скорости работы искателя). Кроме того, можно просматривать статистику о числе URL, просмотренных искателем с начала текущего сеанса.

Скорость работы искателя - это число страниц, просматриваемых искателем за секунду. Это число связано с несколькими свойствами, которые можно сконфигурировать для искателя Web:

- Число потоков искателя
- Число активных сайтов
- Период времени, в течение которого искатель должен ожидать перед получением следующей страницы с того же сервера Web.

Если у искателя есть по одному активному сайту на каждый поток искателя, и искатель должен делать паузу в две секунды перед получением очередной страницы с того же сервера Web, он не может просматривать больше одной страницы за две секунды для каждого потока. Например, если искатель использует число потоков по умолчанию (200), эти 200 потоков могут обрабатывать 100 страниц в секунду.

Если активных сайтов в два раза больше, чем потоков искателя, и искатель должен делать паузу в две секунды перед получением очередной страницы с того же сервера Web, он может просматривать одну страницу в секунду для каждого потока. Однако в этом случае ограничивающими факторами становятся скорости загрузки из сети и

пропускная способность базы данных. Индикатором хорошей производительности искателя является скорость работы искателя, соответствующая числу потоков искателя, активных сайтов и времени ожидания искателя.

Еще один параметр, на который следует обратить внимание при слежении за производительностью искателя Web - число URL, просмотренных искателем с начала текущего сеанса искателя. Разделите это число на общее время работы искателем, чтобы получить среднюю пропускную способность за длительное время. Если это число не растет, искатель либо завершил работу, либо не может ее продолжать. Например, ходу работы искателя могут помешать ошибки связи в сети, ошибки баз данных и неудачные попытки поиска DNS.

Создание отчетов искателя Web

Просматривая отчеты о предыдущей активности искателя Web, можно оценить общую производительность и нужным образом настроить свойства искателя Web и определения пространства для искателя.

Прежде, чем вы начнете

Если ваша административная роль позволяет вам только следить за собраниями, вы можете просматривать статистику искателя и создавать отчеты о его активности, но не можете изменять поведение искателя (например, запускать и останавливать его).


Об этой задаче


Информация об активности искателя Web может содержаться в разных типах отчетов. В некоторых типах отчетов информация возвращается сразу же после ее сбора во внутренней базе данных искателя. На создание отчета по сайту и отчета по кодам возврата HTTP требуется время. При создании отчетов этого типа можно указать для их получения электронный адрес, если вы не хотите ждать возвращения результатов на консоли администратора поисковой системы предприятия.


Информацию о том, как интерпретировать статистические показатели в отчетах, можно посмотреть, нажав кнопку **Справка** во время отслеживания искателя Web и создания отчетов.

Порядок действий

Чтобы создать отчеты искателя Web:

1. В окне Собрания найдите собрание, владеющее искателем Web, за которым вы хотите следить, и выберите  **Слежение**.
2. Откройте страницу Искатель.

Совет: Если при редактировании собрания вы уже находитесь на странице Искатель, выберите  **Слежение**, чтобы перейти в окно слежения за искателями.

3. Если искатель Web, для которого нужно создать отчет, работает или приостановлен, выберите  **Подробности**.
4. На странице подробностей для искателя Web выберите опцию для типа отчета, который хотите создать:
 - В области **Сводка состояния искателя** нажмите кнопку **Хронология искателя**, чтобы создать отчеты об искателе и всех найденных или просмотренных им сайтах.

- В области **Состояние URL** задайте URL конкретного сайта, для которого нужно создать отчет, нажмите кнопку **Просмотр** и затем кнопку **Хронология сайта**.
5. Для каждого из отчетов по хронологии искателя и хронологии сайта включите переключатель напротив каждой статистики, которую хотите видеть в отчете, и затем нажмите кнопку **Просмотреть отчет**.
Для этих видов статистик искатель возвратит отчет на консоль администратора сразу после получения информации из своей внутренней базы данных.
 6. Если вы создаете отчет о хронологии искателя, задайте опции для создания отчета по сайту, затем нажмите кнопку **Запустить отчет**.
Этот отчет включает при его создании выбранные вами статистики и сохраняется в указанном вами файле (имя файла должно быть абсолютным). Вы можете задать, чтобы созданный отчет пересылался вам по электронной почте.
 7. Если вы создаете отчет о хронологии искателя, задайте опции для создания отчета по кодам возврата HTTP, затем нажмите кнопку **Запустить отчет**.
Этот отчет содержит информацию о распределении числа кодов возврата HTTP по сайтам. Он сохраняется в указанном вами файле (имя файла должно быть абсолютным). Вы можете задать, чтобы созданный отчет пересылался вам по электронной почте.
Этот отчет позволяет увидеть, например, какие сайты возвращают много кодов возврата 4xx (означающих, что страницы не найдены), кодов возврата 5xx (указывающих на проблемы на сервере), кодов возврата 6xx (указывающих на проблемы со связью) и так далее.
Он наиболее полезен в случае искателей, уже находящихся некоторое время в активном состоянии (например, несколько недель). С его помощью можно выявлять исчезнувшие сайты, вновь появившиеся сайты, сайты со слишком большим количеством URL (что может означать избыточный поиск в базе данных Lotus Notes) и сайты с рекурсивной файловой системой, обслуживаемые сервером HTTP. Если сайты с большим количеством кодов возврата HTTP не попадают в индекс, производительность искателя можно улучшить, удалив эти сайты из пространства для искателя.

Коды возврата HTTP искателя Web

При наблюдении за искателем Web можно просматривать информацию о кодах возврата HTTP, возвращаемых искателем со страниц, которые он пытается просмотреть.

Сводка таблицы

При наблюдении хронологии искателя Web или состояния определенного адреса URL выводится информация о кодах возврата HTTP, возвращенных искателем. Эту информацию можно использовать при управлении пространством для искателя и при оптимизации производительности искателя. Например, если искатель получает для URL множество кодов возврата HTTP, которые говорят о том, что заданные страницы невозможно просмотреть, можно улучшить производительность, удалив этот URL из пространства искателя.

В следующей таблице перечислены коды возврата HTTP и то, как их обрабатывает искатель Web. Значения от 100 до 505 - стандартные коды возврата HTTP (дополнительную информацию смотрите по адресу <http://www.w3.org/Protocols/rfc2616/rfc2616.html>). Остальные коды возврата HTTP - собственные коды возврата HTTP поисковой системы предприятия и искателя Web.

Таблица 4. Код возврата HTTP от искателя Web

Код	Описание	Код	Описание	Код	Описание	Код	Описание
NULL	Не просмотрен	401	Не авторизован	500	Внутренняя ошибка сервера	700	Ошибка анализа (нет конца заголовка)
100	Продолжить	402	Требуется оплата	501	Не реализован	710	Ошибка анализа (заголовок)
200	Успешное завершение	403	Запрещен	502	Неверный шлюз	720	Ошибка анализа (нет кода HTTP)
201	Создан	404	Не найден	503	Служба недоступна	730	Ошибка анализа (тело)
202	Принят	405	Неразрешенный метод	504	Истек срок ожидания шлюза	740	Исключен файлом robots.txt
203	Ненадежная информация	406	Неприемлем	505	Версия HTTP не поддерживается	741	Роботы временно недоступны
204	Нет содержимого	407	Необходима аутентификация прокси	611	Ошибка чтения	760	Исключен определением пространства для искателя
205	Сброс содержимого	408	Истек срок ожидания требования	612	Ошибка соединения	770	Неверный протокол или нестандартный порт системы
206	Частичное содержимое	409	Конфликт	613	Истек срок ожидания при чтении	780	Исключен по типу файла
300	Несколько вариантов выбора	410	Исчез	614	Ошибка при установлении соединения SSL	2004	Нет тега META для индекса
301	Перемещен постоянно	411	Необходима длина	615	Другая ошибка чтения	3020	Программное перенаправление
302	Найден	412	Не выполнено предварительное условие	616	Ненормальный FBA	4044	Исключен файлом robots.txt
303	Смотрите другие	413	Слишком большой объект требования	617	Ошибка кодировки		
304	Не изменен	414	Слишком длинный URI требования	680	Ошибка поиска DNS		
305	Используйте прокси	415	Неподдерживаемый тип мультимедиа				
306	(Не используется)	417	Ожидания не оправдались				
307	Временное перенаправление						

Примечания к таблице

Коды возврата 4xx

Код 400 (неверное требование) встречается редко. В соответствии со стандартом кодов возврата HTTP коды 4xx означают, что ошибка произошла на стороне клиента (то есть искателя). Однако обычно проблема заключается в сервере или в URL, полученном искателем в качестве ссылки. Например, некоторые серверы Web не принимают URL, пытающиеся перейти на верхний уровень по отношению к корневому каталогу сайта (например, <http://xyz.ibm.com/../../sales>). Другие серверы Web нормально обрабатывают такие URL, игнорируя указание родительского каталога, если искатель уже находится в корневом каталоге.

Некоторые серверы воспринимают требования корневого каталога сайта как ошибочные, и некоторые устаревшие ссылки могут требовать выполнения операций, которые уже не распознаются или не поддерживаются. При запросе страницы, которая больше не обслуживается, сервер прикладных программ генерирует исключительную ситуацию, и Web-сервер возвращает код возврата HTTP 400, так как требование уже не считается допустимым.

- 615** Означает, что сервер искателя, загружающий данные с сайтов, неожиданно столкнулся с исключительной ситуацией. Большое число кодов возврата этого типа может означать, что возникла проблема с искателем.

Коды возврата 6xx

За исключением 615, коды возврата 6xx сообщают о возможных проблемах при просмотре сайтов, например, об истечении срока ожидания. Для следующих кодов возврата могут потребоваться действия по исправлению:

611, 612 и 613

Означают медленные сайты или низкую производительность сети.

- 614** Означает, что искатель не может просматривать защищенные сайты (HTTPS). Если вы считаете, что эти сайты должны быть доступны, убедитесь, что на сервере искателя и на Web-сервере назначения установлены правильные сертификаты. Например, если сайт сертифицирован известными сертифицированными (certificate authorities, CA), можно добавить новые CA на склад доверенных сертификатов, используемый искателем.

Кроме того, проверьте, как сконфигурированы самоподписанные сертификаты на сайтах, которые пытается просматривать искатель. Искатель сконфигурирован на принятие самоподписанных сертификатов. Некоторые сайты создают самоподписанный сертификат для корневого URL (например, <http://sales.ibm.com/>), а затем пытаются использовать этот сертификат на поддоменах (например, <http://internal.sales.ibm.com/>). Искатель не принимает сертификаты, используемые таким образом. Он принимает самоподписанные сертификаты, только если имя домена субъекта (sales.ibm.com) и лица, подписавшего сертификат, совпадает с именем домена требуемой страницы.

- 616** Указывает, что в загрузке после переаутентификации по-прежнему есть форма регистрации.
- 617** Указывает на невозможность создания строки из байтового содержания документа из-за неверной кодировки (набора символов) или наличия в документе неверных байтов.
- 680** Означает, что искатель не смог получить IP-адреса для хостов в пространстве искателя, возможно, из-за проблем с доступом в сеть.

Эта ошибка означает, что искатель не смог обработать целые сайты, а не только некоторые URL. Большое число кодов возврата этого типа может значительно снизить пропускную способность.

Коды возврата 7xx

Коды 7xx преимущественно связаны с правилами пространства для искателя:

710 - 730

Означает, что из-за проблем искатель не смог выполнить загрузку полностью или обнаружил на сайте недопустимые данные HTML. Если выводится большое число кодов возврата этого типа, обратитесь за помощью к представителю службы поддержки поисковой системы предприятия.

740 или 4044

Означает, что содержимое файла нельзя индексировать, так как этот документ исключен в соответствии с ограничениями файла robots.txt.

740 Означает, что в индекс можно включать ссылки привязки, указывающие на исключенный документ.

4044 Означает, что ссылки привязки в документах, указывающие на исключенный документ, также исключаются из индекса.

741 Означает, что на сайте есть файл robots.txt, разрешающий просмотр, но его загрузка не удалась. Если искатель несколько раз не может просмотреть URL, этот URL удаляется из пространства искателя. Если коды возврата этого типа попадают часто, проверьте, временно или постоянно недоступен сайт назначения. Если сайт назначения недоступен, удалите его из пространства искателя.

Остальные коды возврата 7xx обычно возвращаются в случае внесения изменений в пространство искателя, когда искатель уже проработал некоторое время. Обычно эти коды возврата не свидетельствуют о проблемах, требующих вашего внимания.

3020 Указывает, что документ с кодом возврата 200 содержит заголовок положения, который перенаправляет пользовательский агент на другой URL.

Отслеживание работы анализатора

Отслеживайте работу анализатора, когда хотите просмотреть информацию о документах, проанализированных анализатором, перед добавлением их в индекс поисковой системы предприятия. Вы можете просматривать статистические данные и управлять активностью анализатора.

Прежде, чем вы начнете

Если ваша административная роль позволяет только следить за собраниями, вы можете просматривать состояние анализатора, но не можете запускать или останавливать анализатор.

Об этой задаче


При отслеживании подробностей анализатора вы видите снимок активности, содержащий статистическую информацию об активности анализатора в конкретный момент времени. Эта статистическая информация показывает число просмотренных, просматриваемых или ожидающих просмотра документов и число документов, которые проанализированы и ожидают сохранения в индексе.


Когда анализатор активен, сообщения дают дополнительную информацию о состоянии анализатора. Например:


- Синтаксический анализатор, возможно, активно работает с документами.
- Синтаксический анализатор, возможно, бездействует. Анализатор бездействует, пока нет новых документов для анализа. Если возникли ошибки, анализатор ожидает перезапуска. Анализатор сам перезапускает себя, если нет доступных служб синтаксического анализа (например, автоматический перезапуск происходит, если не удалось установить соединение со службой анализа или если все виртуальные Java-машины анализатора заняты работой с другими собраниями).
- Анализатор, возможно, приостановлен (например, до завершения реорганизации индекса).

Порядок действий

Чтобы следить за анализатором для собрания:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Анализатор.

Совет: Если при редактировании собрания вы уже находитесь на странице Анализатор, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.

3. Если анализатор работает, и вы хотите увидеть подробную информацию о его активности, выберите  **Подробности**.

Если ваша административная роль позволяет управлять процессами для собрания, можно запускать и останавливать анализатор при просмотре подробностей о его активности.

4. Если анализатор остановлен и нужно запустить его, выберите  **Запустить**.

При первом создании собрания запускайте анализатор только после того, как искатель начнет просмотр данных. Это дает анализатору данные для анализа и категоризации. Если вы не вносите изменения в правила анализа, можно позволить анализатору работать непрерывно.

5. Если анализатор запущен и нужно остановить его, выберите  **Остановить**.

Если в правила анализа внесены изменения, анализатор нужно остановить и перезапустить. Например, если вы изменили конфигурацию анализатора, нужно остановить и перезапустить анализатор, чтобы эти изменения вступили в силу.

Отслеживание активности индекса для собрания


Слежение за индексом для собрания может понадобиться, если вам нужно следить за ходом построения индекса, включить или отключить расписание индекса или же запустить или остановить индексацию.


Прежде, чем вы начнете






Все пользователи-администраторы могут следить за активностью индекса. Запускать или останавливать построение индекса, а также включать и отключать расписание индексирования может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

Порядок действий

Чтобы следить за индексом для собрания:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
2. Откройте страницу Индекс.

Совет: Если при редактировании собрания вы уже находитесь на странице Индекс, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.

3. Если индекс включен в расписание, но вы хотите не начинать построение в запланированный момент времени, выберите  **Отключить расписание**. Построение индекса не начнется, пока вы не включите расписание или не запустите процесс индексирования.
4. Если индекс внесен в расписание, но расписание для построения отключено, выберите  **Включить расписание**. Построение индекса будет поставлено в очередь в дату и время, которые вы задали в расписании индексации.
5. Если индексация остановлена и нужно запустить ее, выберите  **Запустить**. Обычно индексация выполняется регулярно по расписанию. Если вы остановили индексацию во время процесса построения или отключили расписание индексации, нажмите кнопку **Запустить**, чтобы принудительно возобновить построение индекса.
6. Если построение индекса выполняется и нужно остановить его, выберите  **Остановить**. Остановить построение индекса может понадобиться, например, для принудительной реорганизации индекса после изменения используемого в собрании типа категоризации.
7. Если при построении индекса были ошибки, щелкните по  **Ошибка**. Откроется страница Содержимое файла журнала, где можно просмотреть дополнительную информацию об ошибках индексации. На этой странице можно выбрать отдельные сообщения об ошибках и посмотреть подробности, относящиеся к ошибке.

Слежение за очередью индексации поисковой системы предприятия

Вы можете просматривать состояние всех процессов индексации в очереди индексации, останавливать построение индекса, а также удалять индекс из очереди.

Прежде, чем вы начнете

Управлять очередью индексации может пользователь с ролью администратора поисковой системы предприятия.

Об этой задаче




Одновременно можно создавать несколько индексов, но только один индекс для собрания может находиться в очереди в каждый момент времени. При настройке опций индексации для системы вы задаете, сколько индексов может параллельно использовать одну очередь и общие ресурсы индексации.

Порядок действий

Чтобы следить за очередью индексации:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите страницу Индекс.

Появится список собраний, у которых есть индексы в очереди индексации. Для каждого создаваемого индекса можно посмотреть его тип (обновление или реорганизация), время помещения в очередь индексации и время начала построения индекса (если процесс еще выполняется).

3. Чтобы управлять отдельным индексом, щелкните по значку **Состояние**.
Например, вам может понадобиться узнать, насколько индексация близка к завершению, посмотреть, сколько документов содержится в индексе, или отключить расписание индексации.
4. Чтобы остановить процесс построения индекса, выберите  **Остановить**.
Например, если вы изменили правила категорий, вам может понадобиться остановить обновление индекса, чтобы иметь возможность вместо этого принудительно начать реорганизацию этого индекса.
Чтобы возобновить остановленное построение индекса, дождитесь помещения этого индекса в очередь в очередное запланированное время запуска или щелкните по значку **Состояние**, чтобы следить за этим индексом, а затем выберите  **Запустить**, чтобы обновить или реорганизовать индекс.
5. Чтобы удалить индекс из очереди, выберите  **Удалить**.

Отслеживание работы серверов поиска


Вы можете просматривать подробную информацию о состоянии, касающуюся активности сервера поиска, для конкретного собрания или же подробную информацию о состоянии для серверов поиска по всей поисковой системе вашего предприятия.




Прежде, чем вы начнете

Все пользователи-администраторы поисковой системы предприятия могут следить на поисковых серверах за собраниями, на управление которыми у них есть полномочия. Для слежения за всеми серверами поиска в поисковой системе предприятия вы должны быть администратором поисковой системы предприятия.

Запустить или остановить сервер поиска может администратор поисковой системы предприятия, администратор этого собрания или оператор этого собрания.

Порядок действий

1. Чтобы следить за серверами поиска для одного собрания:
 - a. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
 - b. Откройте страницу Поиск.

Совет: Если при редактировании собрания вы уже находитесь на странице Поиск, выберите  **Слежение**, чтобы перейти в окно слежения за собранием.
2. Для слежения за всеми серверами поиска в поисковой системе предприятия:
 - a. Выберите **Система**, чтобы открыть окно Система.
 - b. Выберите страницу Поиск.
3. Если сервер поиска остановлен и нужно запустить его, выберите  **Запустить**.
4. Если сервер поиска запущен и нужно остановить его, выберите  **Остановить**.
Если вы включили или отключили кэш поиска, внесли изменения в размер кэша поиска или в прямые ссылки, нужно остановить и перезапустить серверы поиска, чтобы эти изменения вступили в силу.

5. Чтобы посмотреть сводку о длительности обработки сервером поиска поисковых требований, нажмите кнопку **Время ответа хронология**.

Этот отчет содержит среднее время в миллисекундах, затраченное сервером поиска на ответы на поисковые требования в конкретный день.

Среднее время ответа - это показатель производительности системы, соответствующий качеству обслуживания. Увеличение времени ответа может указывать на перегрузку системы. Например, в системе может быть превышено число собраний, в которых выполняется поиск, а также размер этих собраний.

6. Чтобы посмотреть список часто выполняемых запросов, нажмите кнопку **Частые запросы**.

Этот отчет содержит ключевые слова 50 наиболее частых запросов с указанием частоты каждого запроса.

Просматривая самые частые запросы, можно определить, для каких из них можно задать прямые ссылки. Создание прямых ссылок может способствовать повышению качества поиска для многих пользователей. Это позволяет всегда возвращать в результатах поиска документы с хорошим соответствием его условиям.

Можно также создать ссылки на ресурсы, где эти запросы могут выполняться с портала предприятия. Например, если пользователям часто требуется информация о счетах подотчетных сумм предприятия, добавьте ссылку на страницу с обсуждением процедур составления таких счетов на домашнюю страницу внутренней сети предприятия.

7. Чтобы посмотреть список самых последних запросов, нажмите кнопку **Последние запросы**.

Этот отчет содержит ключевые слова 50 последних выполненных запросов.

Просматривая последние запросы, можно выявлять текущие тенденции в вашей организации, а также ситуации, требующие срочных действий. Например, можно обнаружить всплеск интереса к определенной теме. Эта волна интереса может указывать на необходимость создать прямую ссылку на эту тему или другими способами сделать ее доступной пользователям (например, с помощью ссылки на портале предприятия).

Слежение за работой системы приема данных


Слежение за работой системы приема данных выполняется с целью определения ее состояния и просмотра подробностей работы клиентской программы приема данных.

Прежде, чем вы начнете

Чтобы следить за системой приема данных, необходимо быть администратором поисковой системы предприятия.

Порядок действий

Чтобы следить за системой приема данных:

1. Выберите **Система**, чтобы открыть окно Система.
2. На странице Система приема данных просмотрите значки состояния, чтобы определить, работает система приема данных или остановлена.
3. Если система приема данных работает и вы хотите просмотреть подробную информацию о работе клиентской программы, нажмите кнопку  **Подробности**.
Значки состояния на странице Подробности системы приема данных показывают, работает система приема данных или остановлена. Статистические параметры показывают, сколько требований ожидают обработки, текущее состояние

каждого из потоков, обрабатывающих требования клиентской программы, и сколько потоков активно для данного состояния потоков.

4. Если вы изменили номер порта для системы приема данных, нажмите кнопку

 **Перезапустить.**

Система приема данных запускается при запуске поисковой системы предприятия. Нет необходимости перезапускать систему приема данных, если номер порта не изменялся.

Отслеживание документов

Документы можно отбрасывать из системы на разных стадиях обработки. Вы можете задать опции уведомлений об отбрасывании документов и его причинах.

Если синтаксический анализатор обнаруживает ошибку, препятствующую анализу документа, в журнал записывается сообщение об отброшенном документе с кодом причины. (Ошибки этого типа не приводят к удалению из индекса старых версий документа.)

Документы могут отбрасываться на стадиях индексации, и информация об этом также записывается в журнал. Например, возможно явное удаление шаблонов URI и URI. Документ мог быть просмотрен искателем, который затем был удален. Возможно, что документ источника больше не существует (с документом связан отрицательный код HTTP), или с ним связан неизвестный код HTTP. Кроме того, документы могут быть отброшены, если отсутствует информация о ранге для документа, для которого требуется глобальный анализ.

Если известно, что документ был просмотрен искателем, но не появился в индексе, проследить путь обработки документа в системе можно при помощи консоли администратора поисковой системы предприятия. Подробные отчеты могут рассказать, когда, где и почему документ был отброшен. Например, в отчете может указываться, что документ был неожиданно отброшен при глобальном анализе, или что администратор удалил URI из индекса.

Конфигурирование файлов журнала для отслеживания документов

Чтобы определять, когда, где и почему тот или иной документ был отброшен из системы, можно сконфигурировать запись в файлы журналов информации об отбрасывании документов.

Прежде, чем вы начнете

Конфигурировать опции для отслеживания отброшенных документов может только администратор поисковой системы предприятия или администратор собрания.

Об этой задаче

Чтобы файлы журналов не занимали слишком много дискового пространства, система всегда переходит к новому файлу журнала при изменении текущей даты. Если дата не изменилась, но размер файл журнала достиг максимального разрешенного размера, система создает новый файл журнала. Когда достигнуто максимальное число файлов журнала, самый старый файл журнала отбрасывается, чтобы можно было создать новый файл журнала.

Порядок действий

Чтобы сконфигурировать файлы журнала для отслеживания документов:

1. Вызовите редактирование собрания, выберите страницу Журнал и щелкните по **Конфигурировать отслеживание документов**.
2. На странице Отслеживание документов убедитесь, что включен переключатель отслеживания документов.
3. Задайте число файлов журнала для записи информации о документах, отброшенных из системы. Эти файлы журналов совместно используются всеми сеансами, в которых возможно отбрасывание документов.
4. Нажмите кнопку **ОК**.

Просмотр отчетов об отброшенных документах

Можно просматривать подробную информацию о документах, отброшенных из поисковой системы предприятия. Эта информация доступна только в том случае, если вы включили отслеживание документов для этого собрания.

Прежде, чем вы начнете

Прежде чем отправить требование на просмотр отчета об отброшенных документах или послать отчет по адресу электронной почты, убедитесь, что сеансы, от которых вы хотите получить информацию, активны. Например, чтобы узнать о документах, отброшенных при синтаксическом анализе или индексации, убедитесь, что запущены сеансы анализатора или индексации для этого собрания.

Чтобы можно было получить отчет, сконфигурируйте в поисковой системе предприятия информацию о почтовом сервере. Эта информация задается при конфигурировании опций электронной почты на странице Журнал окна Система.




Об этой задаче

Процесс сбора информации об отброшенных документах отнимает много времени. Можно выбрать опцию просмотра информации и затем подождать ее вывода. Более эффективна опция отправки отчета по указанному вами электронному адресу.

Если документ был отброшен, в отчете указывается дата и время, когда был отброшен документ, уровень серьезности ошибки, компонент и сеанс, в которых возникла ошибка, и сообщение об ошибке.

Порядок действий

Чтобы просмотреть подробности об отброшенных документах:

1. На странице Собрания найдите собрание, за которым нужно следить, и выберите  **Слежение**.
Совет: Если вы редактируете собрание и уже находитесь на странице Общие, можно выбрать  **Наблюдатель**, чтобы перейти в окно слежения за собранием.
2. Нажмите кнопку  **Подробности URI**.
3. На странице Подробности URI введите URI, для которого нужно просмотреть информацию.
4. Включите переключатели для типов информации, которую вы хотите получить:

Документы, отброшенные синтаксическим анализатором

Включите этот переключатель, чтобы видеть, был ли документ отброшен при синтаксическом анализе, а если был, то по какой причине.

Документы, отброшенные из индекса

Включите этот переключатель, чтобы видеть, был ли документ отброшен при индексации или анализе, а если был, то по какой причине.

5. Укажите, в каком виде хотите просматривать отчет:

- Чтобы вывести на экран отчет, нажмите кнопку **Просмотреть отчет**.
- Чтобы послать отчет по электронной почте для последующего просмотра, нажмите кнопку **Послать отчет**.

На странице **Послать подробный отчет об URI** введите электронный адрес для получения отчета в поле **Адрес электронной почты для уведомления** и включите **Послать отчет**.

Файлы журналов и оповещения поисковой системы предприятия

Можно выбрать типы сообщений, записываемых в журнал для собрания и для системы, задать опции для создания и просмотра файлов журналов, получения оповещений и получения электронной почты об этих сообщениях.

В нормальном режиме компоненты поисковой системы предприятия записывают сообщения журнала в общий файл журнала. Этот файл журнала находится в каталоге `ES_NODE_ROOT/logs` на сервере индексов. Для просмотра данных этого общего журнала можно использовать консоль администратора.

В случае ошибки, например, неисправности сетевой связи, каждый компонент поисковой системы предприятия записывает сообщения журнала в каталог `logs` на сервере, где этот компонент установлен. Для просмотра файлов журнала используйте программу просмотра файлов на этом компьютере, например, утилиту `tail` в системе UNIX. Для просмотра этих типов файлов нельзя использовать консоль администратора.

При конфигурировании файлов журнала можно выбрать типы сообщений для записи в журнал (например, сообщения об ошибках или предупреждения), задать частоту отбрасывания старых файлов журнала для освобождения места под новые файлы журнала, задать максимальный размер файлов журнала и выбрать язык сообщений. Можно также задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов.

При слежении за файлами журнала можно выбрать, какой файл журнала открыть. Можно отфильтровать содержимое файла журнала, чтобы просмотреть сообщения только определенного уровня серьезности (например, только сообщения об ошибках) или сообщения, сгенерированные определенным сеансом поисковой системы предприятия. При просмотре файла журнала можно просмотреть подробности об отдельных сообщениях. Например, можно вывести имя функции, сгенерировавшей сообщение, и другую информацию, которая может быть необходима для устранения причины ошибки.

Оповещения

Поисковую систему предприятия можно сконфигурировать так, чтобы при наступлении определенных событий записывать в файл журнала сообщения.

Сообщения, запускаемые событиями (их называют оповещениями), сообщают вам о ситуациях, которые требуют вмешательства, например, исчерпание ресурсом свободного пространства. При конфигурировании оповещений для поисковой системы предприятия вы задаете условия, за которыми должна следить система. Если система обнаруживает такое условие, она автоматически записывает в файл журнала сообщение.

Чтобы получать непосредственное уведомление о каком-либо условии, можно задать опции отправки сообщения по электронной почте при записи сообщения слежения в журнал.

Можно сконфигурировать оповещения для событий уровня собрания и для событий, происходящих на уровне системы. На уровне собрания система может:

- Отслеживать число документов, просматриваемых каждым искателем, и генерировать сообщение оповещения, что вскоре будет достигнуто максимально допустимое число документов.
- Отслеживать число документов, добавляемых в индекс для собраний, и генерировать сообщение оповещения, что вскоре будет достигнуто максимально допустимое число документов.
- Сообщать, что время, необходимое для ответов на требования поиска, превышает заданный вами предел.

На уровне системы система может отслеживать место на диске для каждого сервера поисковой системы предприятия и генерировать сообщение оповещения в случае малого объема свободного пространства.

Конфигурирование оповещений уровня собрания

Конфигурирование оповещений позволяет записывать сообщения в файл журнала при возникновении определенных событий уровня собрания. Кроме того, можно включить отправку сообщений по электронной почте при записи в журнал сообщений об этих событиях.

Прежде, чем вы начнете

Конфигурировать оповещения может пользователь с ролью администратора поисковой системы предприятия или администратор собрания.

Порядок действий

Чтобы сконфигурировать оповещения уровня собрания:

1. Откройте собрание для редактирования, откройте страницу Поиск и нажмите кнопку **Конфигурировать оповещения**.
2. Чтобы система отслеживала число документов, просматриваемых каждым искателем, выполните следующие действия:
 - a. Включите переключатель **Когда число документов, просмотренных любым искателем, достигнет определенного процента от разрешенного максимума**.
 - b. В поле **Проценты** задайте, когда надо записывать сообщение в журнал. Это число задается в виде процентов от максимального числа документов, которые могут просмотреть искатели (**Максимальное число документов для просмотра** задается при конфигурировании свойств искателя). Значение по умолчанию - 90 процентов.

Так как для разных искателей можно настроить разные максимальные значения, для каждого искателя в журнал записываются отдельные сообщения. Например, если вы используете значение порога уведомления по умолчанию, искателю DB2 разрешено просмотреть 2000000 документов, а искателю Notes разрешено просмотреть 1000000 документов; одно сообщение будет записано, когда искатель DB2 просмотрит 1800000 документов, а другое - когда искатель Notes просмотрит 900000 документов.
3. Чтобы система отслеживала число документов, добавляемых в индекс, выполните следующие действия:
 - a. Включите переключатель **Когда число документов в собрании достигнет определенного процента от их предполагаемого числа**.

- b. В поле **Проценты** задайте, когда надо записывать сообщение в журнал. Это число задается в виде процентов от оценки числа документов, которые, как вы ожидаете, будут в этом собрании. Значение по умолчанию - 85 процентов.

В поле **Предел** выводится текущее значение предполагаемого размера собрания. Чтобы изменить это значение, откройте страницу **Общие** для этого собрания, выберите опцию для конфигурирования общих опций и задайте новое значение в поле **Предполагаемое число документов**

Внимание: Это предельное значение и предполагаемое число документов, заданные для собрания, используются только для отслеживания роста размера собрания. Они не задают абсолютного предельного значения размера индекса.

- 4. Чтобы система информировала вас в случае превышения предела времени ответа на требования поиска, выполните следующие действия:
 - a. Включите переключатель **Когда время ответа при поиске превысит предел**.
 - b. В поле **Предел** введите число секунд, которое вы считаете приемлемым в качестве максимального времени ответа при поиске.

Если это значение будет превышено, система запишет сообщение об этом событии в журнал. Например, если вы оставите значение по умолчанию, система будет записывать сообщение в журнал каждый раз, когда среднее время ответа сервера поиска на запросы поиска будет больше или равно пяти секундам.

Типичное время ответа не превышает полсекунды. Если средние значения превышают секунду, это может означать, что операционная система требует настройки для лучшей производительности, или что в настройках конфигурации сервера поиска что-то задано неправильно. Например, может быть нужно увеличить объем, выделяемый для кэша поиска.

- 5. Нажмите кнопку **ОК**.

Чтобы получать сообщение по электронной почте, когда система записывает в журнал сообщения об этих событиях, откройте страницу **Журнал** и выберите **Конфигурировать опции электронной почты для сообщений** и задайте адрес электронной почты. ID сообщений для включенных оповещений автоматически добавляются в список ID сообщений, для которых нужно отправлять сообщение по электронной почте.

Чтобы получать электронную почту, необходимо также сконфигурировать информацию о почтовом сервере. Для этого администратор поисковой системы предприятия должен выбрать на панели инструментов **Система**, открыть страницу **Журнал** и нажать кнопку **Конфигурировать опции электронной почты для сообщений**.

Конфигурирование оповещений уровня системы

Конфигурирование оповещений позволяет записывать сообщения в файл журнала при возникновении определенных событий уровня системы. Кроме того, можно включить отправку сообщений по электронной почте при записи в журнал сообщений об этих событиях.

Прежде, чем вы начнете

Конфигурировать оповещения уровня системы может только администратор поисковой системы предприятия.

Порядок действий

Чтобы сконфигурировать оповещения уровня системы:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Журнал выберите **Конфигурировать оповещения**.
4. Чтобы система отслеживала объем свободного места на серверах поисковой системы предприятия, включите переключатель **Когда объем свободного пространства файловой системы достигнет определенного процента от общего объема**.
5. В поле **Проценты** задайте, когда система должна уведомлять вас о том, что на сервере осталось мало свободного места. Это число указывается в виде процентной доли от общего пространства файловой системы. Значение по умолчанию - 80 процентов.
Если поисковая система предприятия установлена на нескольких серверах, система создает для каждого сервера отдельный журнал сообщений. Например, одно сообщение информирует вас; когда становится мало пространство на сервере искателя, а другие отдельные сообщения - об ограничениях пространства на сервере индекса и поисковых серверах.
6. Нажмите кнопку **ОК**.

Чтобы получать сообщение по электронной почте каждый раз, когда система записывает в журнал сообщение об этом событии, откройте страницу Журнал и выберите **Конфигурировать опции электронной почты для сообщений**, чтобы задать адрес электронной почты и информацию о почтовом сервере.

Конфигурирование файлов журнала

Можно задать типы сообщений, которые надо записывать в журнал, и опции для создания файлов журнала.

Прежде, чем вы начнете

Сконфигурировать файлы журнала уровня собрания может администратор поисковой системы предприятия или администратор этого собрания. Сконфигурировать файлы журнала уровня системы может только администратор поисковой системы предприятия.

Об этой задаче



Чтобы файлы журналов не занимали слишком много дискового пространства, система всегда переходит к новому файлу журнала при изменении текущей даты. Если дата не изменилась, но размер файл журнала достиг максимального разрешенного размера, система создает новый файл журнала. Когда достигнуто максимальное число файлов журнала, самый старый файл журнала отбрасывается, чтобы можно было создать новый файл журнала.

Чтобы получать электронную почту о событиях, записываемых в журнал, надо предварительно указать информацию о доставке электронной почты. Затем надо задать сообщения, для которых надо отправлять электронную почту.

Порядок действий

Чтобы сконфигурировать файлы журнала поисковой системы предприятия:

1. Если надо сконфигурировать опции создания файлов журнала уровня системы и перехода к новым файлам журнала:
 - a. Выберите **Система**, чтобы открыть окно Система.

- b. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
 - c. На странице Журнал выберите **Конфигурировать опции файлов журнала**. Откроется страница Опции файлов журнала уровня системы.
2. Если надо сконфигурировать опции создания файлов журнала уровня собрания и перехода к новым файлам журнала:
 - a. В окне Собрания найдите собрание, для которого нужно задать опции, и щелкните по  **Редактировать**.
 - b. На странице Журнал выберите **Конфигурировать опции файлов журнала**. Откроется страница Опции файлов журнала уровня собрания.
3. В поле **Тип информации для записи в журнал** выберите тип сообщений, которые надо записывать в журнал:

Только сообщения об ошибках

Сообщения об ошибках указывают, что возникла нежелательная ситуация или непредвиденное поведение и процесс не может быть продолжен. Вы должны исправить эту ошибку.

Сообщения об ошибках и предупреждения

Предупреждения указывают возможные конфликты или несовместимости, которые не приводят к остановке процесса. Это - опция по умолчанию.

Все сообщения

Информационные сообщения содержат общую информацию о системе и текущей задаче и не требуют никаких действий по исправлению.

4. В поле **Максимальный размер каждого файла журнала** введите максимальный размер в мегабайтах для каждого файла журнала. Значение по умолчанию - 5 Мбайт.

Когда размер файла журнала достигает этого значения, создается новый файл журнала, пока не будет достигнуто разрешенное вами максимальное число файлов журнала. Сохраняя относительно небольшие файлы журнала, можно облегчить их просмотр.
5. В поле **Максимальное число файлов журнала** введите максимальное число создаваемых файлов журнала. Значение по умолчанию - 10.

Если вы хотите, чтобы можно было просматривать более старые сообщения журнала, увеличьте эти значения. Если вас больше интересуют последние сообщения и вам не нужно хранить длинную хронологию активности, уменьшите это значение.
6. В поле **Национальная версия по умолчанию** выберите язык для записи сообщений в журнал. Значение по умолчанию - Английский.
7. Нажмите кнопку **ОК**.

Конфигурирование информации о SMTP-сервере

Чтобы получать электронные письма о работе поисковой системы предприятия, необходимо сконфигурировать информацию о сервере почтового протокола Simple Mail Transfer Protocol (SMTP-сервере).

Прежде, чем вы начнете

Чтобы сконфигурировать информацию о SMTP-сервере, надо быть администратором поисковой системы предприятия.


Об этой задаче

Получать электронную почту позволяют несколько функций администратора поисковой системы предприятия. Для получения электронной почты в рамках любой из этих функций необходимо указать сведения о вашем SMTP-сервере:

- Если вы сконфигурировали оповещения уровня собрания или уровня системы, можно получать сообщения по электронной почте при каждой записи этих сообщений в журнал. Можно также получать сообщения по электронной почте при записи других сообщений, (не только сообщений, генерируемых событиями, за которыми ведется слежение).
- Если вы хотите просмотреть подробную информацию об URI в индексе или о документе, отброшенном из поисковой системы предприятия, можно получить отчет по электронной почте.
- Если вы выполняете слежение за искателем Web и указали, что хотите создавать хронологические отчеты о работе этого искателя Web, можно получать уведомление по электронной почте после создания отчета.

Порядок действий

Чтобы сконфигурировать информацию о SMTP-сервере:

1. Выберите **Система**, чтобы открыть окно Система.
2. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
3. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
4. На странице Опции электронной почты для сообщений системы в поле **Почтовый SMTP-сервер для доставки электронной почты**, введите полное имя хоста или IP-адрес используемого SMTP-сервера.
Система будет использовать этот сервер для отправки электронной почты по указанным вами адресам.
5. В поле **Частота проверки электронной почты** задайте, как часто система должна проверять наличие заданных сообщений и отправлять о них уведомления по электронной почте.
Система с заданной частотой объединяет все сообщения на один адрес электронной почты в одно сообщение и отправляет это сообщение.
6. Нажмите кнопку **ОК**.

Получение сообщений журнала по электронной почте

Можно задать опции для получения сообщений по электронной почте при записи в журнал определенных сообщений или сообщений определенных типов.

Прежде, чем вы начнете

Конфигурировать опции электронной почты для сообщений уровня системы может только администратор поисковой системы предприятия. Конфигурировать опции электронной почты для сообщений уровня собрания может только администратор поисковой системы предприятия или администратор этого собрания.

Чтобы получать электронную почту, необходимо задать информацию о почтовом сервере Simple Mail Transfer Protocol (SMTP-сервере).



Об этой задаче

При конфигурировании оповещений можно выбрать опцию записи сообщений в журнал при определенных событиях. Если эти опции включены, можно

сконфигурировать опции для автоматического получения электронной почты при записи в журнал этих сообщений. Кроме того, можно задать опции для получения электронной почты при записи других сообщений, а не только сообщений, вызванных событиями.

Порядок действий

Чтобы сконфигурировать опции электронной почты для сообщений:

1. Если вы хотите получать оповещения по электронной почте о сообщениях системы:
 - a. Выберите **Система**, чтобы открыть окно Система.
 - b. Выберите  **Правка**, чтобы перейти в окно редактирования системы.
 - c. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
 - d. На странице Опции электронной почты для сообщений системы включите переключатель **Посылать оповещения по электронной почте о сообщениях уровня системы**.
 - e. В поле **Адрес электронной почты для получения оповещений** введите один или несколько адресов электронной почты. Обычно информацию о сообщениях системы должен получать администратор поисковой системы предприятия. Указывайте адреса через запятую. Например:
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.
 - f. Чтобы получать электронную почту обо всех записываемых в журнал сообщениях об ошибках, включите переключатель **Посылать оповещения обо всех сообщениях об ошибках**.
 - g. Если вы хотите получать оповещения по электронной почте только при записи в журнал определенных сообщений уровня системы, укажите ID этих сообщений в области **Посылать оповещения об определенных сообщениях**. Вводите по одному ID сообщения в строке. Например:
FFQC4819E
FFQ00005E
Несколько ID сообщений входят в список по умолчанию (выберите **Справка**, чтобы посмотреть описания этих сообщений).
 - h. Нажмите кнопку **ОК**.
2. Если вы хотите получать оповещения по электронной почте о сообщениях собрания:
 - a. Выберите **Собрания**, чтобы открыть окно Собрания.
 - b. В списке собраний найдите собрание, которое нужно сконфигурировать, и щелкните по  **Редактировать**.
 - c. На странице Журнал выберите **Конфигурировать опции электронной почты для сообщений**.
 - d. На странице Опции электронной почты для сообщений собрания включите переключатель **Посылать оповещения по электронной почте о сообщениях уровня собрания**.
 - e. В поле **Адрес электронной почты для получения оповещений** введите один или несколько адресов электронной почты. Обычно информацию о сообщениях уровня собрания должен получать администратор собрания. Указывайте адреса через запятую. Например:
steinbeck@us.ibm.com, yeats@ireland.ibm.com, dante@it.ibm.com.

- f. Чтобы получать электронную почту обо всех записываемых в журнал сообщениях об ошибках, включите переключатель **Посылать оповещения обо всех сообщениях об ошибках**.
- g. Если вы хотите получать оповещения по электронной почте только при записи в журнал определенных сообщений уровня собрания, укажите ID этих сообщений в области **Посылать оповещения об определенных сообщениях**. Вводите по одному ID сообщения в строке. Например:
 - FFQC4819E
 - FFQ00005E
 Несколько ID сообщений входят в список по умолчанию (выберите **Справка**, чтобы посмотреть описания этих сообщений).
- h. Нажмите кнопку **ОК**.


Просмотр файлов журнала


Можно просматривать сообщения журнала, которые компоненты системы и собрания записывают в общий файл журнала. Кроме того, можно задать фильтры, чтобы видеть только сообщения некоторого уровня серьезности и только от отдельных сеансов поисковой системы предприятия.

Прежде, чем вы начнете

Все пользователи-администраторы поисковой системы предприятия могут просматривать файлы журнала для собраний, которыми им разрешено управлять. Просматривать файлы журнала уровня системы может только администратор поисковой системы предприятия или пользователь, которому разрешен доступ к панели инструментов **Система**.

Порядок действий

1. Чтобы просмотреть файлы журнала для одного собрания:
 - a. Выберите **Собрания**, чтобы открыть окно Собрания.
 - b. Найдите в списке собраний собрание, которое нужно просмотреть, выберите  **Слежение** и откройте страницу Журнал.

Совет: Если вы редактируете собрание и уже находитесь на странице Журнал, можно выбрать  **Наблюдатель**, чтобы перейти в окно слежения за собранием.
2. Для просмотра файлов журнала уровня системы:
 - a. Выберите **Система**, чтобы открыть окно Система.
 - b. Выберите страницу Журнал.
3. В поле **Файл журнала** выберите файл журнала для просмотра.


Имя каждого файла журнала состоит из типа файла журнала (например, идентификатора системы или собрания), даты создания файла и числового суффикса, определяющего номер файла среди созданных за эту дату. Например:

```
тип_файла_журнала_2005-05-26_1.log
тип_файла_журнала_2005-05-26_2.log
тип_файла_журнала_2005-05-25_1.log
тип_файла_журнала_2005-05-25_2.log
тип_файла_журнала_2005-05-25_3.log
```
4. Чтобы просматривать только сообщения определенных уровней серьезности, включите соответствующие переключатели в поле **Серьезность**.
5. Чтобы просматривать только сообщения от определенных сеансов, включите соответствующие переключатели в поле **Сеанс**.

6. Нажмите кнопку **Просмотреть журнал**.

Для каждого сообщения на странице Содержимое файлов журнала выводится дата и время выдачи сообщения, его уровень серьезности, имя сеанса, выдавшего сообщение, ID сообщения и текст ошибки.

При помощи кнопок можно переходить на первую, последнюю, предыдущую или следующую страницу файла журнала. Кроме того, можно задать номер страницы для перехода сразу на эту страницу.

7. Чтобы просмотреть более подробную информацию о каком-либо сообщении, выберите  **Подробности**.

На странице Подробности сообщения журнала выводится имя хоста сервера поисковой системы предприятия, на котором выдано сообщение, имя файла, сгенерировавшего ошибку, имя функции и номер строки, где произошла ошибка, ID процесса и ID потока.

Резервное копирование и восстановление поисковой системы предприятия

Резервное копирование и восстановление поисковой системы предприятия можно выполнить при помощи сценариев резервного копирования и восстановления.

Если в системе происходит сбой из-за неисправимой ошибки, необходимо переустановить WebSphere Information Integrator OmniFind Edition и затем запустить сценарий восстановления. Эти сценарии позволяют также восстановить необходимые системные файлы на один или несколько новых серверов.

Сценарии выполняют резервное копирование и восстановление следующих файлов:

- Файлов конфигурации из каталога ES_NODE_ROOT/master_config.
- Файлов баз данных для искателей.
- Файлов индексов. Если индексы поисковой системы предприятия не находятся в каталоге ES_NODE_ROOT/data, для резервного копирования и восстановления файлов индексов нельзя использовать сценарии поисковой системы предприятия.

Сценарий резервного копирования создает в каталоге, указанном вами при его запуске, следующие подкаталоги. У ID администратора поисковой системы предприятия должно быть разрешение на запись в указанный вами каталог.)

master_config

Содержит файлы конфигурации из каталога ES_NODE_ROOT/master_config

database

Содержит файлы баз данных с сервера искателя

data

Содержит файлы индексов с сервера индексов

Для резервного копирования файлов поисковой системы предприятия в другой каталог у вас должно быть достаточно доступного дискового пространства. Сценарии резервного копирования и восстановления не проверяют файлы. После запуска резервного копирования большинство сеансов системы станут временно недоступными. Процессы поиска будут продолжать выполняться. Чтобы сохранять свежий индекс, резервное копирование следует начинать после реорганизации индекса.

В системе с несколькими серверами резервное копирование и восстановление системы нужно выполнять с сервера индексов поисковой системы предприятия. Поскольку сервер индексов создает каталог баз данных, он может получить доступ к таблицам базы данных продукта на сервере искателя и выполнить их резервное копирование.

Резервное копирование поисковой системы предприятия

Создать резервную копию поисковой системы предприятия можно при помощи сценариев esbackup.sh для UNIX или esbackup.bat для Microsoft Windows.

Ограничения

На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.

У ID администратора поисковой системы предприятия должно быть разрешение на запись в каталог, задаваемый вами при запуске сценария резервного копирования.

Порядок действий

Чтобы создать резервную копию поисковой системы предприятия:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. Введите следующую команду, где *каталог_резервной_копии* - каталог, куда нужно поместить резервные копии файлов:

UNIX: `esbackup.sh -c -d -i каталог_резервной_копии`

Windows: `esbackup.bat -c -d -i каталог_резервной_копии`

Можно указать следующие опции:

- c Создает резервную копию файлов конфигурации. Это опция по умолчанию.
- d Создает на складе данных искателя резервные копии документов (пока не проанализированных и не проиндексированных).
- i Создает резервную копию файлов индексов. Если файлы индексов не находятся в каталоге `ES_NODE_ROOT/data`, для создания их резервных копий нельзя использовать сценарий `esbackup`.

Восстановление поисковой системы предприятия

Сценарий `esrestore.sh` (для UNIX) или сценарий `esrestore.bat` для (Microsoft Windows) позволяет восстановить файлы конфигурации системы после переустановки WebSphere II OmniFind Edition.

Ограничения

На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.

Порядок действий

Чтобы восстановить поисковую систему предприятия:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.
2. Запустите слой общей связи (common communication layer, CCL), если он еще не запущен:

UNIX: `startccl.sh -bg`

Windows: `startccl.bat`

Чтобы запустить CCL в фоновом режиме в системе Windows, выберите **Пуск** → **Настройка** → **Панель управления** → **Администрирование** → **Службы** и перезапустите службу WebSphere Information Integrator OmniFind Edition.

3. Остановите контроллер, если он еще не остановлен:

```
esadmin stop
```

4. Введите следующую команду (где *каталог_резервной_копии* - каталог созданной вами резервной копии файлов):

```
UNIX: esrestore.sh -c -d -i каталог_резервной_копии
```

```
Windows: esrestore.bat -c -d -i каталог_резервной_копии
```

Можно указать следующие опции:

- c** Восстанавливает файлы конфигурации. Это опция по умолчанию.
- d** Восстанавливает на складе данных искателя документы, обработанные искателем. (Эти документы еще не проанализированы и не проиндексированы.)
- i** Восстанавливает файлы индекса. Если файлов индекса нет в каталоге ES_NODE_ROOT/data, сценарий esrestore нельзя использовать для восстановления файлов индекса.

Восстановление файлов поисковой системы предприятия на новые серверы

Можно создать резервную копию файлов системы на одном сервере поисковой системы предприятия и восстановить их не на тот же сервер, на один или несколько новых серверов.

Прежде, чем вы начнете

Перед запуском сценария восстановления надо установить WebSphere II OmniFind Edition на новые серверы.

Ограничения

На время работы сценариев резервного копирования или восстановления все сеансы системы останавливаются. Чтобы избежать неправильной или несогласованной информации о системе, не используйте консоль администратора поисковой системы предприятия, пока работают эти сценарии.

У ID администратора поисковой системы предприятия должно быть разрешение на запись в каталог, задаваемый вами при запуске сценария резервного копирования.

Об этой задаче

Информация о сервере, хранящаяся в файле ES_NODE_ROOT/master_config/nodes.ini, не включается в файлы резервной копии.

Порядок действий

Чтобы восстановить поисковую систему предприятия на один или несколько новых серверов:

1. На сервере индексов зарегистрируйтесь как администратор поисковой системы предприятия. Этот ID пользователя задается при установке WebSphere II OmniFind Edition.

2. Запустите сценарий резервного копирования на текущем (старом) сервере индексов (где *каталог_резервной_копии* - каталог, куда вы хотите поместить резервную копию файлов):

UNIX: `esbackup.sh каталог_резервной_копии`

Windows: `esbackup.bat каталог_резервной_копии`

3. При помощи программы FTP скопируйте все файлы из каталога *каталог_резервной_копии* на новый сервер индексов.
4. Запустите сценарий восстановления на новом сервере индексов:

UNIX: `esrestore.sh каталог_резервной_копии`

Windows: `esrestore.bat каталог_резервной_копии`

Команды, коды возврата и ID сеансов поисковой системы предприятия

Команды можно использовать для диагностики ошибок, определения состояния различных частей системы, запуска и остановки сеансов, а также запуска и остановки системы.

В установке с несколькими серверами можно запускать команды с любого сервера системы. Однако следует запускать команды с сервера индексов. Сервер индексов или сервер контроллера может обращаться к информации всех серверов в системе.

Формат большинства команд:

```
esadmin команда аргументы  
esadmin ID_сеанса действие -опция
```

Чтобы получить дополнительную информацию обо всех командах, введите `esadmin help`. Чтобы получить дополнительную информацию конкретной команде, введите `esadmin действие help`.

Команды `esadmin` поисковой системы предприятия

Вводите следующие команды в одну строку.

Таблица 5. Команды `esadmin` поисковой системы предприятия

Команда	Описание
<code>esadmin id_сеанса_искателя start</code>	Запускает сеанс искателя. Эта команда не запускает никаких действий искателя. Пример команды: <code>esadmin col1.WEB1.esadmin start</code> Пример сообщений и кодов возврата: FFQC5310I WEBCrawler1 (sid: col1.WEB1.esadmin) не работает. FFQC5314I Результат: 0
<code>esadmin id_сеанса_искателя startCrawl</code>	Запускает просмотр. Пример команды: <code>esadmin col3.DB21.esadmin startCrawl</code> Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650 FFQC5314I Результат: 0
<code>esadmin id_сеанса_искателя pause</code>	Приостанавливает просмотр. Пример команды: <code>esadmin col3.DB21.esadmin pause</code> Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650 FFQC5314I Результат: 0

Таблица 5. Команды esadmin поисковой системы предприятия (продолжение)

Команда	Описание
esadmin id_сеанса_искателя resume	<p>Возобновляет просмотр.</p> <p>Пример команды: esadmin col3.DB21.esadmin resume</p> <p>Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650 FFQC5314I Результат: 0</p>
esadmin id_сеанса_искателя stopCrawl	<p>Останавливает просмотр.</p> <p>Пример команды: esadmin col3.DB21.esadmin stopCrawl</p> <p>Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650 FFQC5314I Результат: 0</p>
esadmin id_сеанса_искателя stop	<p>Останавливает сеанс искателя.</p> <p>Пример команды: esadmin col3.DB21.esadmin stop</p> <p>Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650 FFQC5314I Результат: 0</p>
esadmin id_сеанса_искателя getCrawlerStatus	<p>Получает состояние искателя. Возвращаемая информация зависит от того, является ли искатель искателем Web или же иным искателем.</p> <p>Пример для искателя Web: esadmin col1.WEB1.esadmin getCrawlerStatus</p> <p>Возможные коды возврата и сообщения для искателя Web: FFQC5303I WebCrawler1 (sid: col1.WEB1.esadmin) уже работает. PID: 23650</p> <p>Пример для другого искателя (не Web): esadmin col3.DB21.esadmin getCrawlerStatus</p> <p>Возможные коды возврата и сообщения для другого искателя (не Web): FFQC5303I db2crawler (sid: db2col.DB2_96945) уже работает. PID: 5936</p> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 258.</p>

Таблица 5. Команды *esadmin* поисковой системы предприятия (продолжение)

Команда	Описание
<i>esadmin id_сеанса_искателя</i> <i>getCrawlSpaceStatus</i>	Получает общее состояние пространства для любого искателя, за исключением искателя Web.
<i>esadmin id_сеанса_искателя_web</i> <i>getCrawlStatus</i> <i>-selections значение</i>	Пример команды: <i>esadmin col3.DB21.esadmin getCrawlSpaceStatus</i>
	Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650
	Получает общее состояние пространства для искателя Web.
	Пример команды: <i>esadmin col1.WEB1.esadmin getCrawlStatus</i>
	Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 258.
<i>esadmin id_сеанса_искателя</i> <i>getCrawlSpaceStatusDetail</i> <i>-ts id_сервера_назначения</i>	Получает подробное состояние пространства для любого искателя, за исключением искателя Web. Если не задать сервер назначения, возвращаются данные для всех серверов назначения.
<i>esadmin id_сеанса_искателя_web</i> <i>getCrawlDetailsPerSite</i> <i>-url строка selections число</i> <i>-threshold число</i>	Например, если искатель DB2 просматривает базы данных FOUNTAIN и SAMPLE и вы не задали опцию сервера назначения, возвращается состояние всех таблиц баз данных FOUNTAIN и SAMPLE.
	Пример команды: <i>esadmin col3.DB21.esadmin</i> <i>getCrawlSpaceStatusDetail -ts FOUNTAIN</i>
	Пример сообщений и кодов возврата: FFQC5303I DB2Crawler1 (sid: col3.DB21.esadmin) уже работает. PID: 23650
	Получает подробное состояние пространства для искателя Web.
	Пример команды: <i>esadmin col1.WEB1.esadmin getCrawlDetailsPerSite</i>
	Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 258.
<i>esadmin startParser -cid ID_собрания</i>	Запускает анализатор.
	Пример команды: <i>esadmin startParser -cid col1</i>
	Пример сообщений и кодов возврата: FFQC5303I Контроллер (server1) (sid: controller) уже работает. PID: 25917 FFQC5314I Результат: 0

Таблица 5. Команды esadmin поисковой системы предприятия (продолжение)

Команда	Описание
esadmin stopParser <i>id_собрания</i>	<p>Останавливает анализатор.</p> <p>Пример команды: esadmin stopParser -cid coll</p> <p>Пример сообщений и кодов возврата: FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 15292 FFQC5312E Ошибка при выполнении действия: [stopParser] FFQC4823E Драйвер анализатора сеанса - Собрание coll (node1) [coll.parserdriver] не работает.</p>
esadmin monitor getCollectionParserMonitorStatus -cid <i>ID_собрания</i>	<p>Получает состояние анализатора.</p> <p>Пример команды: esadmin monitor getCollectionParserMonitorStatus -cid coll</p> <p>Пример сообщений и кодов возврата: FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12543</p> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 258.</p>
esadmin startMain -cid <i>collection_id</i>	<p>Запускает реорганизацию индекса.</p> <p>Пример команды: esadmin startMain -cid coll</p> <p>Пример сообщений и кодов возврата: FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 25917 FFQC5314I Результат: 1117671147056</p>
esadmin startDelta -cid <i>id_собрания</i>	<p>Запускает обновление индекса.</p> <p>Пример команды: esadmin startDelta -cid coll</p> <p>Пример сообщений и кодов возврата: FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 4548 FFQC5314I Результат: 1117670603408</p>

Таблица 5. Команды esadmin поисковой системы предприятия (продолжение)

Команда	Описание
<pre>esadmin monitor getCollectionIndexMonitorStatus -cid ID_собрания -buildType [main delta] -numrecords последние_N_записей</pre>	<p>Получает состояние обновления или реорганизации индекса. (main означает реорганизуемый индекс, а delta - обновляемый индекс.) Опция numrecords выводит последние N записей состояния построения индекса. Если опция numrecords опущена, выводится 20 последних записей построения индекса.</p> <p>Пример команды:</p> <pre>esadmin monitor getCollectionIndexMonitorStatus -cid coll -buildType main -numrecords 4</pre> <p>Пример сообщений и кодов возврата:</p> <pre>FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12649</pre> <p>Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния” на стр. 258.</p>
<pre>esadmin startSearch -cid id_собрания</pre>	<p>Запускает процессы сервера поиска.</p> <p>Пример команды:</p> <pre>esadmin startSearch -cid coll</pre> <p>Пример сообщений и кодов возврата:</p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 25917 FFQC5314I Результат: 0</pre>
<pre>esadmin stopSearch -cid id_собрания</pre>	<p>Останавливает процессы сервера поиска.</p> <p>Пример команды:</p> <pre>esadmin stopSearch -cid coll</pre> <p>Пример сообщений и кодов возврата:</p> <pre>FFQC5303I Контроллер (node1) (sid: controller) уже работает. PID: 15292 FFQC5314I Результат: 0</pre>

Таблица 5. Команды esadmin поисковой системы предприятия (продолжение)

Команда	Описание
esadmin monitor getCollectionSearchMonitorStatus -cid ID_собрания	Получает состояние сервера поиска. Пример команды: esadmin monitor getCollectionSearchMonitorStatus -cid coll
esadmin id_сеанса_сервера_поиска getStatus -cid id_собрания	Пример сообщений и кодов возврата: FFQC5303I Монитор (node1) (sid: monitor) уже работает. PID: 12649 Возвращает подробную информацию о состоянии индексов поиска для собрания на заданном сервере поиска. На каждом сервер поиска есть один сеанс менеджера поиска. Каждый сеанс менеджера поиска отвечает за отслеживание и работу индексов поиска на определенном сервере поиска. Пример команды: esadmin searchmanager.node1 getStatus -cid coll
	Пример сообщений и кодов возврата: FFQC5303I Менеджер поиска (node1) (sid: searchmanager.node1) уже работает. PID: 15711 FFQC5314I Результат: PID CacheHits=3 QueryRate=1 Port=44008 SessionId=coll.runtime.node1 CacheHitRate=0.333 ResponseTime=70 Status=1 SessionName=coll.runtime.node1.1 Дополнительную информацию о возвращаемых сообщениях о состоянии смотрите в разделе “Подробная информация для команд состояния”.

Подробная информация для команд состояния

Некоторые команды возвращают обширную информацию. В этом разделе описана информация о состоянии искателя и пространства искателя, которая может быть возвращена. Таблица в разделе “Команды esadmin поисковой системы предприятия” на стр. 253 содержит возможную возвращаемую информацию для каждой из команд esadmin. В этом разделе описана возвращаемая информация для следующих команд:

- Состояние искателя Web
- Состояние других искателей (кроме искателя Web)
- Состояние пространства для искателя Web
- Состояние пространства для других искателей (кроме искателя Web)
- Подробное состояние пространства для искателя Web
- Подробное состояние пространства для других искателей (кроме искателя Web)
- Состояние анализатора
- Состояние построения индекса
- Состояние сервера поиска
- Подробное состояние сервера поиска

Состояние искателя Web: Когда вы вводите команду для получения состояния искателя Web, эта команда возвращает информацию в формате документа XML. Команда состояния искателя Web возвращает следующую информацию:

```
FFQC5314I Результат: <?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>
```

Ниже в таблице описаны все элементы XML и их возможные атрибуты, возвращаемые командой состояния искателя Web:

Таблица 6. Информация состояния искателя Web

Элемент	Атрибуты	Описание
CrawlerStatus	<ul style="list-style-type: none"> CrawlerThreadStateDist ActiveBucketList CrawlRate RecentlyCrawledURLList NumURLsThisSession 	Состояние искателя.
Значение CrawlerRunLevel	<ul style="list-style-type: none"> Строка (английский язык) “Not started”: Сеанс искателя существует, но он еще не получил сообщения запуска, начинающее обработку документов. “Started”: Искатель запущен. “Running”: Искатель завершил инициализацию и запуск, и теперь активно просматривает документы. “Paused”: Искатель получил указание приостановить активный просмотр, но не завершать работу. “Stopping”: Искатель получило сигнал остановки и собирается остановиться. “Error”: Искатель в невосстановимом состоянии; его надо остановить и перезапустить, чтобы продолжить просмотр. 	Информация о текущих действиях искателя.
Состояние CrawlerThreadState	Строка (английский язык)	Действия потока искателя. В это поле показано, что делают потоки.

Таблица 6. Информация состояния искателя Web (продолжение)

Элемент	Атрибуты	Описание
ActiveBucket	<ul style="list-style-type: none"> URL: Строка (спецификация URL) Протокол, хост и номер порта для просматриваемых URL. NumActURLs: Целое (положительное) Число URL в комплекте в момент, когда он стал доступным для просмотра (был активирован). NumProcURLs: Целое (неотрицательное) Число обработанных URL в комплекте (просмотренных или отвергнутых) TimeRem: Целое Оставшийся срок в секундах из времени, отведенного для просмотра комплекта. Duration: Целое (неотрицательное) Время в секундах, прошедшее после активации комплекта. 	Текущее действие на указанном сайте.
CrawlRate	<p>Значение: Целое (неотрицательное)</p> <p>Скорость просмотра в страницах в секунду (для всех комплектов)</p>	Показатель пропускной способности искателя.
RecentlyCrawledURL	<p>URL: Строка (спецификация URL)</p> <p>Строка задает протокол, хост, порт и имя просмотренного файла.</p>	Последняя просмотренная страница.
NumURLsThisSession	Значение: Целое (неотрицательное)	Число URL, просмотренных с момента начала просмотра этим экземпляром искателя (процессом).

Состояние другого искателя (кроме искателя Web): Когда вы вводите команду для получения состояния другого искателя (кроме искателя Web), эта команда возвращает информацию в формате документа XML. Команда состояния искателя non-Web возвращает следующую информацию:

```
FFQC5314I Результат: <?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>
```

Ниже в таблице описаны элементы XML и атрибуты для каждого из искателей поисковой системы предприятия, кроме искателя Web. Эту информацию возвращает команда состояния искателя.

Таблица 7. Элементы и атрибуты для искателя NNTP, искателя DB2 и искателя Notes, возвращаемые командой состояния искателя

Имя элемента и атрибута	Искатель NNTP	Искатель DB2	Искатель Notes
Status	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число серверов NNTP в пространстве для искателя.	Число баз данных в пространстве для искателя.	Число баз данных в пространстве для искателя.
NumberOfCompletedServers	Число просмотренных серверов NNTP.	Число просмотренных баз данных.	Число просмотренных баз данных.
NumberOfTargets	Число групп новостей в пространстве для искателя.	Число таблиц в пространстве для искателя.	Число представлений и папок в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных групп новостей.	Число просмотренных таблиц.	Число просмотренных представлений и папок.
NumberOfCompletedRecords	Число просмотренных статей.	Число просмотренных записей.	Число просмотренных документов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Таблица 8. Элементы и атрибуты для искателя Exchange Server, искателя DB2 Content Manager и искателя Content Edition, возвращаемые командой состояния искателя

Имя элемента и атрибута	Искатель Exchange Server	Искатель DB2 Content Manager	Искатель Content Edition
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число серверов Exchange Server в пространстве для искателя.	Число серверов Content Manager в пространстве для искателя.	Число репозиторий в пространстве для искателя.
NumberOfCompletedServers	Число просмотренных серверов Exchange Server.	Число просмотренных серверов Content Manager.	Число просмотренных репозиторий.
NumberOfTargets	Число подпапок в пространстве для искателя.	Число типов элементов в пространстве для искателя.	Число классов в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных подпапок.	Число просмотренных типов элементов.	Число просмотренных классов элементов.
NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных документов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Таблица 9. Элементы и атрибуты для искателя QuickPlace, искателя Domino Document Manager, искателя файловых систем UNIX и искателя файловых систем Windows, возвращаемые командой состояния искателя

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Состояние	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)	Состояние (0, 1, 2, -1)

Таблица 9. Элементы и атрибуты для искателя QuickPlace, искателя Domino Document Manager, искателя файловых систем UNIX и искателя файловых систем Windows, возвращаемые командой состояния искателя (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
StatusMessage	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка	Состояние: 0 - Бездействие, 1 - Работает, 2 - Приостановлен, -1 - Ошибка
NumberOfServers	Число площадок в пространстве для искателя.	Число библиотек в пространстве для искателя.	Постоянное значение - 1.
NumberOfCompletedServers	Число просмотренных площадок.	Число просмотренных библиотек.	0 или 1, если просмотрены все подкаталоги.
NumberOfTargets	Число баз данных на площадках и в комнатах в пространстве для искателя.	Число шкафов в пространстве для искателя.	Число подкаталогов в пространстве для искателя.
NumberOfCompletedTargets	Число просмотренных баз данных на площадках и в комнатах в пространстве для искателя.	Число просмотренных шкафов.	Число просмотренных подкаталогов.
NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных файлов.
RunningThreads	Число потоков искателя.	Число потоков искателя.	Число потоков искателя.

Состояние пространства для искателя Web: Когда вы вводите команду для получения состояния пространства для искателя Web, эта команда возвращает информацию в формате документа XML. Команда состояния пространства для искателя Web возвращает следующую информацию:

Таблица 10. Значения маски выбора для команды состояния пространства для искателя Web

Бит маски	Выбирает
1	Число страниц на складе необработанных данных.
2	Число обнаруженных сайтов.
4	Число сайтов с записями DNS.
8	Число сайтов без записей DNS.
16	Число обнаруженных URL.
32	Число сохраненных уникальных страниц.
64	Число просмотренных URL.
128	Число непросмотренных URL.
256	Число просроченных URL.
512	Распределение кодов возврата HTTP.

Все значения представляют собой накопленные суммы для всех сеансов, использующих текущую внутреннюю базу данных:

```
<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>

```

Возвращаемые данные могут содержать следующие элементы (возможно, все или же никакие из них):

Таблица 11. Информация, возвращаемая командой состояния пространства для искателя Web

Элемент	Атрибут	Описание
CrawlerStatus	<ul style="list-style-type: none"> NumPagesInRDS NumSitesDiscovered NumSitesWithDNS NumSitesWithoutDNS NumURLsDiscovered NumUniquePagesSaved NumURLsCrawled NumURLsUncrawled NumURLsOverdue HTTPCodeDist 	Информация, которую можно быстро получить из суммарного состояния искателя (по всем сеансам).
NumPagesInRDS	<p>Значение: Неотрицательное целое</p> <p>Сколько страниц сейчас находятся в кэше склада необработанных данных (raw data store, RDS) - только для данного искателя.</p>	Насколько заполнен склад необработанных данных (RDS) - только за счет вклада данного искателя.
NumSitesDiscovered	<p>Значение: Неотрицательное целое</p> <p>Сколько хостов обнаружено при просмотре (или из начальных URL).</p>	Мера покрытия искателем просматриваемого домена (число хостов)
NumSitesWithDNS	<p>Значение: Неотрицательное целое</p> <p>У скольких хостов есть связанные с ними IP-адреса (разрешенные искателем в фоновом режиме).</p>	Показатель того, насколько эффективно искатель может получать IP-адреса для хостов, обнаруженных по именам DNS в URL.
NumSitesWithoutDNS	<p>Значение: Неотрицательное целое</p> <p>У скольких хостов нет связанных с ними IP-адресов (разрешенных искателем в фоновом режиме).</p>	Показатель того, насколько эффективно искатель может получать IP-адреса для хостов, обнаруженных по именам DNS в URL.
NumURLsDiscovered	<p>Значение: Неотрицательное целое</p> <p>Сколько уникальных URL посетил искатель.</p>	Мера покрытия искателем просматриваемого домена (число URL).
NumUniquePagesSaved	<p>Значение: Неотрицательное целое</p> <p>Сколько уникальных страниц записано на RDS для дальнейшей обработки другими компонентами поисковой системы предприятия.</p>	Вклад данного искателя в размер индекса.

Таблица 11. Информация, возвращаемая командой состояния пространства для искателя Web (продолжение)

Элемент	Атрибут	Описание
NumURLsCrawled	Значение: Неотрицательное целое Сколько уникальных URL просмотрел искатель.	Показатель способности искателя обрабатывать данные от начала до конца. Это число отличается от числа страниц, записанных на RDS, поскольку не все просмотренные страницы записываются на RDS.
NumURLsOverdue	Значение: Неотрицательное целое Сколько уникальных URL подлежат повторному просмотру.	Показатель способности искателя обрабатывать пространство Web.

Состояние пространства для других искателей (кроме искателя Web): Когда вы вводите команду для получения состояния пространства для другого искателя (не искателя Web), эта команда возвращает информацию в формате документа XML. Команда состояния пространства для другого искателя (кроме искателя Web) возвращает следующую информацию:

```
FFQC5314I Результат: <?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>
```

Ниже в таблице описаны элементы XML и атрибуты для каждого из искателей поисковой системы предприятия, кроме искателя Web. Эту информацию возвращает команда состояния пространства искателя. Для искателей Notes при уровне сводки 0 Server@Name - это имя сервера + имя базы данных. Если уровень сводки - 1, Server@Name - это имя сервера + имя каталога.

Таблица 12. Элементы и атрибуты для искателя NNTP, искателя DB2 и искателя Notes, возвращаемые командой состояния пространства искателя

Имя элемента и атрибута	Искатель NNTP	Искатель DB2	Искатель Notes
Server@Name	Имя сервера News	Имя базы данных	Имя базы данных или каталога
Server/Status	Состояние: (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка

Таблица 12. Элементы и атрибуты для искателя NNTP, искателя DB2 и искателя Notes, возвращаемые командой состояния пространства искателя (продолжение)

Имя элемента и атрибута	Искатель NNTP	Искатель DB2	Искатель Notes
Server/StatusMessage	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка
Server/NumberOfTargets	Число групп новостей в пространстве для искателя.	Число таблиц в пространстве для искателя.	Число представлений и папок в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных групп новостей.	Число просмотренных таблиц.	Число просмотренных представлений и папок.
Server/NumberOfErrors	Неприменимо.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.
Server/ScheduleConfigured	Неприменимо.	0, 1, 2 <ul style="list-style-type: none"> • 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию. • 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. • 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено. 	0, 1, 2 <ul style="list-style-type: none"> • 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию. • 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. • 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.
Server/ScheduleTime	Неприменимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.

Таблица 12. Элементы и атрибуты для искомателя NNTP, искомателя DB2 и искомателя Notes, возвращаемые командой состояния пространства искомателя (продолжение)

Имя элемента и атрибута	Искоматель NNTP	Искоматель DB2	Искоматель Notes
Server/AggregationLevel	Неприменимо.	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)	0, 1: <ul style="list-style-type: none"> 0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.) Искоматель Notes просматривает документы в режиме каталога.

Таблица 13. Элементы и атрибуты для искомателя Exchange Server, искомателя DB2 Content Manager и искомателя Content Edition, возвращаемые командой состояния пространства для искомателя

Имя элемента и атрибута	Искоматель Exchange Server	Искоматель DB2 Content Manager	Искоматель Content Edition
Server@Name	Имя сервера Exchange Server.	Серверы DB2 Content Manager.	Имя репозитория.
Server/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка
Server/StatusMessage	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка 	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка 	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка
Server/NumberOfTargets	Число подпапок в пространстве для искомателя.	Число типов элементов в пространстве для искомателя.	Число классов элементов в пространстве для искомателя.
Server/NumberOfCompletedTargets	Число просмотренных подпапок.	Число просмотренных типов элементов.	Число просмотренных классов элементов.
Server/NumberOfErrors	Число ошибок.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.

Таблица 13. Элементы и атрибуты для искомателя Exchange Server, искомателя DB2 Content Manager и искомателя Content Edition, возвращаемые командой состояния пространства для искомателя (продолжение)

Имя элемента и атрибута	Искоматель Exchange Server	Искоматель DB2 Content Manager	Искоматель Content Edition
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> 0: Согласно файлам конфигурации искомателя, этот искоматель не сконфигурирован для работы по расписанию. 1: Искоматель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. 2: Искоматель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено. 	0, 1, 2 <ul style="list-style-type: none"> 0: Согласно файлам конфигурации искомателя, этот искоматель не сконфигурирован для работы по расписанию. 1: Искоматель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. 2: Искоматель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено. 	0, 1, 2 <ul style="list-style-type: none"> 0: Согласно файлам конфигурации искомателя, этот искоматель не сконфигурирован для работы по расписанию. 1: Искоматель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. 2: Искоматель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.
Server/AggregationLevel	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)

Таблица 14. Элементы и атрибуты для искомателя QuickPlace, искомателя Domino Document Manager, искомателя файловых систем UNIX и искомателя файловых систем Windows, возвращаемые командой состояния пространства для искомателя

Имя элемента и атрибута	Искоматель QuickPlace	Искоматель Domino Document Manager	Искоматели файловых систем UNIX и Windows
Server@Name	Каталог площадки.	База данных библиотеки.	Постоянное значение - localhost.
Server/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка 	Состояние (0, 1, 2, 3, 4, 5, -1) <ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен 5: Запланирован -1: Ошибка

Таблица 14. Элементы и атрибуты для искателя QuickPlace, искателя Domino Document Manager, искателя файловых систем UNIX и искателя файловых систем Windows, возвращаемые командой состояния пространства для искателя (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Server/StatusMessage	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • 5: Запланирован • -1: Ошибка
Server/NumberOfTargets	Число баз данных на площадках и в комнатах в пространстве для искателя.	Число шкафов в пространстве для искателя.	Число подкаталогов в пространстве для искателя.
Server/NumberOfCompletedTargets	Число просмотренных баз данных на площадках и в комнатах в пространстве для искателя.	Число просмотренных шкафов.	Число подкаталогов в пространстве для искателя.
Server/NumberOfErrors	Число ошибок.	Число ошибок.	Число ошибок.
Server/StartTime	Время запуска, если применимо.	Время запуска, если применимо.	Время запуска, если применимо.
Server/EndTime	Время завершения, если применимо.	Время завершения, если применимо.	Время завершения, если применимо.
Server/ScheduleConfigured	0, 1, 2 <ul style="list-style-type: none"> • 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию. • 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. • 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено. 	0, 1, 2 <ul style="list-style-type: none"> • 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию. • 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. • 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено. 	0, 1, 2 <ul style="list-style-type: none"> • 0: Согласно файлам конфигурации искателя, этот искатель не сконфигурирован для работы по расписанию. • 1: Искатель сконфигурирован для работы по расписанию, но для данного сеанса расписание отключено. • 2: Искатель сконфигурирован для работы по расписанию, и для данного сеанса расписание включено.
Server/ScheduleTime	Время по расписанию, если применимо.	Время по расписанию, если применимо.	Время по расписанию, если применимо.
Server/TotalTime	Общее время, если применимо.	Общее время, если применимо.	Общее время, если применимо.

Таблица 14. Элементы и атрибуты для ищателя QuickPlace, ищателя Domino Document Manager, ищателя файловых систем UNIX и ищателя файловых систем Windows, возвращаемые командой состояния пространства для ищателя (продолжение)

Имя элемента и атрибута	Ищатель QuickPlace	Ищатель Domino Document Manager	Ищатели файловых систем UNIX и Windows
Server/AggregationLevel	0, 1 <ul style="list-style-type: none"> 0: Ищатель Notes просматривает документы в нормальном режиме. (Все прочие ищатели, кроме ищателя Notes, всегда возвращают 0.) Ищатель Notes просматривает документы в режиме каталога. 	0, 1 <ul style="list-style-type: none"> 0: Ищатель Notes просматривает документы в нормальном режиме. (Все прочие ищатели, кроме ищателя Notes, всегда возвращают 0.) Ищатель Notes просматривает документы в режиме каталога. 	0, 1 <ul style="list-style-type: none"> 0: Ищатель Notes просматривает документы в нормальном режиме. (Все прочие ищатели, кроме ищателя Notes, всегда возвращают 0.) Ищатель Notes просматривает документы в режиме каталога.

Подробное состояние пространства для ищателя Web: Когда вы вводите команду для получения подробного состояния пространства для ищателя Web, эта команда возвращает информацию в формате документа XML. Команда подробного состояния пространства для ищателя возвращает следующую информацию:

Таблица 15. Значения маски выбора для команды подробного состояния пространства для ищателя Web

Бит маски	Выбирает
1	Число страниц на складе необработанных данных.
2	Число обнаруженных сайтов.
4	Число сайтов с записями DNS.
8	Число сайтов без записей DNS.
16	Число обнаруженных URL.
32	Число сохраненных уникальных страниц.
64	Число просмотренных URL.
128	Число непросмотренных URL.
256	Число просроченных URL.
512	Распределение кодов возврата HTTP.

Пример возвращаемой информации:

```
<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>

```

Ниже в таблице описано каждое поле, возвращаемой командой подробного состояния пространства для искателя Web:

Таблица 16. Информация, возвращаемая командой подробного состояния пространства для искателя Web

Элемент	Атрибуты	Описание
CrawlDetailsPerSite	<ul style="list-style-type: none"> LastActivationTime: LastActivationDuration: IPAddressList: RobotsContent: HTTPCodeDist: 	Информация, которую можно быстро получить для подробного состояния одного сайта.
Сайт	URL	URL корневой страницы сайта.
NumURLsDiscovered	Значение	Число URL, обнаруженных на сайте.
NumURLsOverdue	Значение	Число URL, подлежащих повторному просмотру на этом сайте.
NumURLsCrawled	Значение	Число URL, просмотренных на сайте.
NumURLsUncrawled	Значение	Число URL, еще не просмотренных на этом сайте.
NumURLsOverdueBy	Порог, Значение: Целое (положительное или отрицательное) Порог в секундах обнаружения пропуска срока повторного просмотра (отрицательный = срок истек до настоящего момента) или предстоящего просмотра (положительный - срок истекает после настоящего момента).	Число URL, подлежащих повторному просмотру в некотором интервале до настоящего момента или в некотором интервале после настоящего момента.
NumURLsActivated	Значение	Число URL, помещенных в память во время последнего просмотра сайта и объявленных доступными для потоков искателя.
LastActivationTime	Значение	Интервал в секундах с момента, когда URL данного сайта последний раз заносились в память.
LastActivationDuration	Значение	Интервал в секундах, когда URL данного сайта последний раз находились в памяти и были доступными для потоков искателя.
IPAddressList	IP-адрес	Все известные IP-адреса с хоста сервера данного сайта.
IPAddress	Значение	Адрес в формате IPv4 с точками для хоста сервера данного сайта.
RobotsContent	Текст	Текст из файла robots, если он есть.

Таблица 16. Информация, возвращаемая командой подробного состояния пространства для искателя Web (продолжение)

Элемент	Атрибуты	Описание
HTTPCodeDist	Код HTTP	Распределение кодов HTTP для попыток загрузки с данного сайта.
HTTPCode	Код: Целое Код возврата HTTP или иной внутренний код.	Сколько раз определенный код возврата HTTP был получен при просмотре данного сайта.

Подробное состояние пространства для других искателей (кроме искателя Web): Когда вы вводите команду для получения подробного состояния пространства для другого искателя (не искателя Web), эта команда возвращает информацию в формате документа XML. Команда подробного состояния пространства для других искателей (кроме искателя Web) возвращает следующую информацию:

```
FFQC5314I Результат: <?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>
```

Таблица 17. Элементы и атрибуты для искателя NNTP, искателя DB2 и искателя Notes, возвращаемые командой подробного состояния пространства искателя

Имя элемента и атрибута	Искатель NNTP	Искатель DB2	Искатель Notes
Target@Name	Имя группы новостей	Имя таблицы	Имя представления или папки
Target@CrawlType	Неприменимо.	0,1: • 0: Активный просмотр (нормальный режим) • 1: Пассивный просмотр (публикация событий DB2)	0
Target/Status	Состояние: (0, 1, 2, 3, 4, -1) • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка	Состояние: (0, 1, 2, 3, 4, -1) • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка	Состояние: (0, 1, 2, 3, 4, -1) • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка

Таблица 17. Элементы и атрибуты для искателя NNTP, искателя DB2 и искателя Notes, возвращаемые командой подробного состояния пространства искателя (продолжение)

Имя элемента и атрибута	Искатель NNTP	Искатель DB2	Искатель Notes
Target/StatusMessage	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка 	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка 	<ul style="list-style-type: none"> 0: Не просмотрен 1: Идет просмотр 2: Завершен (и не запланирован) 3: Ожидание 4: Приостановлен -1: Ошибка
Target/NumberOfRecords	Номер последней статьи на сервере.	Число просмотренных записей.	Число просмотренных документов.
Target/NumberOfCompletedRecords	Число просмотренных статей.	Число просмотренных записей.	Число просмотренных документов.
Target/NumberOfInsertedRecords	Число новых опубликованных статей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOfUpdatedRecords	Неприменимо.	Число измененных записей.	Число измененных записей.
Target/NumberOfDeletedRecords	Неприменимо.	Число удаленных записей.	Число удаленных записей.
Target/StartTime	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.	Дата и время последнего запуска искателя.
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.
Target/AggregationLevel	0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)	0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)	0, 1: <ul style="list-style-type: none"> 0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.) Искатель Notes просматривает документы в режиме каталога.
Target/LastUpdatedTime	Неприменимо.	Время последнего изменения: <ul style="list-style-type: none"> 0: Активный просмотр (нормальный режим) 1: Пассивный просмотр (публикация событий DB2) 	Неприменимо.

Таблица 17. Элементы и атрибуты для искомателя NNTP, искомателя DB2 и искомателя Notes, возвращаемые командой подробного состояния пространства искомателя (продолжение)

Имя элемента и атрибута	Искоматель NNTP	Искоматель DB2	Искоматель Notes
Target/LastResetTime	Неприменимо.	Время последнего сброса статистики: <ul style="list-style-type: none"> • 0: Активный просмотр (нормальный режим) • 1: Пассивный просмотр (публикация событий DB2) 	Неприменимо.

Таблица 18. Элементы и атрибуты для искомателя Exchange Server, искомателя DB2 Content Manager и искомателя Content Edition, возвращаемые командой подробного состояния пространства для искомателя

Имя элемента и атрибута	Искоматель Exchange Server	Искоматель DB2 Content Manager	Искоматель Content Edition
Target@Name	Имя подпапки	Имя типа элементов	Имя класса элементов
Target@CrawlType	0	0	0
Target/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка
Target/StatusMessage	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка
Target/NumberOfRecords	Неприменимо.	Неприменимо.	Неприменимо.
Target/NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных документов.
Target/NumberOfInsertedRecords	Число вставленных записей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOfUpdatedRecords	Неприменимо.	Число измененных записей.	Число измененных записей.
Target/NumberOfDeletedRecords	Неприменимо.	Число удаленных записей.	Число удаленных записей.
Target/StartTime	Дата и время последнего запуска искомателя.	Дата и время последнего запуска искомателя.	Дата и время последнего запуска искомателя.
Target/EndTime	Дата и время завершения работы искомателя.	Дата и время завершения работы искомателя.	Дата и время завершения работы искомателя.
Target/TotalTime	Время, потраченное искомателем на просмотр.	Время, потраченное искомателем на просмотр.	Время, потраченное искомателем на просмотр.

Таблица 18. Элементы и атрибуты для искомателя Exchange Server, искомателя DB2 Content Manager и искомателя Content Edition, возвращаемые командой подробного состояния пространства для искомателя (продолжение)

Имя элемента и атрибута	Искоматель Exchange Server	Искоматель DB2 Content Manager	Искоматель Content Edition
Target/AggregationLevel	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)	0: Искоматель Notes просматривает документы в нормальном режиме. (Все прочие искоматели, кроме искомателя Notes, всегда возвращают 0.)
Target/LastUpdatedTime	Неприменимо.	Неприменимо.	Неприменимо.
Target/LastResetTime	Неприменимо.	Неприменимо.	Неприменимо.

Таблица 19. Элементы и атрибуты для искомателя QuickPlace, искомателя Domino Document Manager, искомателя файловых систем UNIX и искомателя файловых систем Windows, возвращаемые командой подробного состояния пространства для искомателя

Имя элемента и атрибута	Искоматель QuickPlace	Искоматель Domino Document Manager	Искоматели файловых систем UNIX и Windows
Target@Name	Имя базы данных площадки или комнаты	Имя базы данных шкафа	Имя подкаталога
Target@CrawlType	0	0	0
Target/Status	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	Состояние (0, 1, 2, 3, 4, -1) <ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка
Target/StatusMessage	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка 	<ul style="list-style-type: none"> • 0: Не просмотрен • 1: Идет просмотр • 2: Завершен (и не запланирован) • 3: Ожидание • 4: Приостановлен • -1: Ошибка
Target/NumberOfRecords	Неприменимо.	Неприменимо.	Неприменимо.
Target/NumberOfCompletedRecords	Число просмотренных документов.	Число просмотренных документов.	Число просмотренных файлов.
Target/NumberOfInsertedRecords	Число вставленных записей.	Число вставленных записей.	Число вставленных записей.
Target/NumberOfUpdatedRecords	Число просмотренных баз данных на площадках и в комнатах в пространстве для искомателя.	Число просмотренных шкафов.	Число подкаталогов в пространстве для искомателя.
Target/NumberOfDeletedRecords	Число измененных записей.	Число измененных записей.	Число измененных записей.
Target/StartTime	Дата и время последнего запуска искомателя.	Дата и время последнего запуска искомателя.	Дата и время последнего запуска искомателя.

Таблица 19. Элементы и атрибуты для искателя QuickPlace, искателя Domino Document Manager, искателя файловых систем UNIX и искателя файловых систем Windows, возвращаемые командой подробного состояния пространства для искателя (продолжение)

Имя элемента и атрибута	Искатель QuickPlace	Искатель Domino Document Manager	Искатели файловых систем UNIX и Windows
Target/EndTime	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.	Дата и время завершения работы искателя.
Target/TotalTime	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.	Время, потраченное искателем на просмотр.
Target/AggregationLevel	0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)	0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)	0: Искатель Notes просматривает документы в нормальном режиме. (Все прочие искатели, кроме искателя Notes, всегда возвращают 0.)
Target/LastUpdatedTime	Неприменимо.	Неприменимо.	Неприменимо.
Target/LastResetTime	Неприменимо.	Неприменимо.	Неприменимо.

Состояние анализатора: Когда вы вводите команду для получения состояния анализатора, эта команда возвращает информацию в формате документа XML. Команда состояния анализатора возвращает следующую информацию:

```
FFQC5314I Результат: <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>
```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния анализатора:

Таблица 20. Элементы для команды состояния анализатора

Элемент	Описание
Status	1, если сеанс анализатора для собрания запущен, 0, если сеанс анализатора остановлен.

Таблица 20. Элементы для команды состояния анализатора (продолжение)

Элемент	Описание
State	<p>Возможные состояния: Idle (бездействует), Restart (перезапущен), Parsing (идет анализ), Resuming (возобновлен), Stopped (остановлен), Initializing (инициализируется), Pause (приостановлен), NoParserServiceIsAvailable (служба анализатора недоступна).</p> <p>Состояние Idle указывает, что анализатор бездействует в ожидании поступления документов от искателей в данном собрании. Время бездействия по умолчанию - 300 секунд.</p> <p>Состояние Restart указывает, что анализатор ждет перезапуска виртуальной Java-машины для анализа. JVM для анализа работает в отдельном сеансе и выполняет собственно обработку документов.</p> <p>Состояние Parsing указывает, что анализатор обрабатывает документы.</p> <p>Состояние Parsing указывает, что анализатор обрабатывает документы.</p> <p>Состояние Pause указывает, что анализатор приостановлен сеансом построения индекса для данного собрания. Состояние Initializing указывает, что анализатор запускается и инициализирует состояние.</p> <p>Состояние Resuming указывает, что анализатор возобновляет работу из состояния Pause в состояние Parsing в результате сеанса построения индекса для данного собрания.</p> <p>Состояние NoParserServiceIsAvailable указывает, что виртуальная Java-машина для обработки документов данного собрания недоступна. Это означает, что все JVM анализатора используются для других собраний.</p>
SnapShotTimeStamp	Следующий момент чтения документов анализатором из таблиц искателя в секундах, прошедших с 01.01.1970.
NumberOfDocsToBeIndexed	Число документов на складе для данного собрания. В это число входят также документы, помеченные для удаления при следующем построении индекса.
ParseRate	Скорость анализа в документах в секунду.
ParseRateMBPerHour	Скорость анализа в Мбайтах в час.
NumberOfCpmThreads	Число потоков CPM, используемых JVM анализатора для обработки документов данного собрания.
ParserServiceSession	Имя JVM анализатора, обрабатывающей документы для данного собрания. Это поле доступно, только если анализатора находится в состоянии Parsing.
Name	Имя искателя.
Type	Тип искателя (Web, NNTP, DB2 и т.п.)
ParserStatus	<ul style="list-style-type: none"> • 0: Документы данного искателя еще не анализировались. • 1: Документы данного искателя анализируются в текущий момент. • 2: Документы данного искателя проанализированы.
NumberOfDocsToBeParsed	Число документов от данного искателя, ожидающих анализа.
NumberofDocsAlreadyParsed	Число уже проанализированных документов от данного искателя.

Состояние построения индекса: Когда вы вводите команду для получения состояния построения индексах, эта команда возвращает информацию в формате документа XML. Команда состояния построения индекса возвращает следующую информацию:

```
FFQC5314I Результат: <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>

```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния построения индекса:

Таблица 21. Элементы для команды состояния построения индекса

Элемент	Описание
IndexStatusId	ID состояния индекса.
StartTime	Момент запуска построения индекса в секундах, истекших с 01.01.1970.
Ход работы	Процент выполнения для построения индекса.
CurrentPhase	<ul style="list-style-type: none"> 1: стадия перезаписи склада 2: стадия глобального анализа 3: стадия построения индекса
TotalPhase	Общее число фаз построения индекса. В настоящее время это значение - всегда 3.
IndexCopyProgress	Процент выполнения для копирования индекса. Процесс копирования индекса копирует построенный индекс с сервера построения индексов на серверы поиска.
CurrentServer	Сервер поиска, на который в данный момент копируется индекс.
TotalServer	Число серверов поиска, на которые копируется индекс.
IndexCopyTime	Общее время для всех фаз построения индекса.
Progress	0 для успешного построения и копирования индекса или ненулевой код сообщения об ошибке.
StopTime	Время завершения всех фаз построения и копирования индекса.
TotalTime	Интервал между временем запуска и временем завершения.
NumberOfDocuments	Число документов в индексе.
CurrentIndexWildcardSupport	Значение параметра символов подстановки для следующего построения индекса. Возможные значения - None (не используются), QueryExpansion (расширение в запросе) и IndexExpansion (расширение в индексе).
ScheduleStatus	<ul style="list-style-type: none"> 0, если расписание для данного собрания и типа индекса не включено. 1, если расписание для данного собрания и типа индекса включено.
ScheduledTimeEnabled	Следующий момент построения индекса для данного собрания и типа индекса в секундах, прошедших с 01.01.1970.

Состояние сервера поиска: Когда вы вводите команду для получения состояния сервера поиска, эта команда возвращает информацию в формате документа XML. Команда состояния сервера поиска возвращает следующую информацию:

```
FFQC5314I Результат: <?xml version="1.0" encoding='UTF-8'?>
<Monitor Type="Search" Count="1">
<SearchStatus Name="Search Manager (node1)" SearchID=
"searchmanager.node1" HostName="myComputer.svl.ibm.com">
<Status>1</Status>
</SearchStatus>
</Monitor>
```

Ниже в таблице описаны элементы XML для информации, возвращаемой командой состояния сервера поиска:

Таблица 22. Элементы для команды состояния сервера поиска

Элемент	Описание
SearchStatusName	Имя и ID сеанса менеджера поиска, отслеживающего индексы поиска для данного собрания и управляющего ими.
HostName	Имя хоста для сервера, на котором расположен индекс поиска.
Status	<ul style="list-style-type: none"> 0, если индекс поиска для данного собрания не работает. 1, если индекс поиска для данного собрания работает.

Подробное состояние сервера поиска: Команда состояния сервера поиска может возвращать следующую информацию:

```
FFQC5303I Менеджер поиска (node1) (sid: searchmanager.node1)
уже работает. PID: 15711
FFQC5314I Результат: PID
CacheHits=3
QueryRate=1
Port=44008
SessionId=coll.runtime.node1
CacheHitRate=0.333
ResponseTime=70
Status=1
SessionName=coll.runtime.node1.1
```

Ниже в таблице описаны элементы информации, возвращаемой командой подробного состояния сервера поиска:

Таблица 23. Элементы для команды подробного состояния сервера поиска

Элемент	Описание
CacheHits	Число результатов, полученных из кэша поиска.
QueryRate	Число запросов, полученных за последний интервал времени. По умолчанию этот интервал равен 5 минутам.
Port	Номер порта, используемого индексом поиска для ожидания или получения запросов.
SessionId	ID сеанса для индекса поиска этого собрания.
CacheHitRate	Число результатов, полученных из кэша поиска, в виде процентной доли от общего числа результатов.
ResponseTime	Среднее время ответа в миллисекундах за заданный интервал времени. (Интервал по умолчанию - 5 минут.)
Status	<ul style="list-style-type: none"> 0, если индекс поиска для данного собрания не работает. 1, если индекс поиска для данного собрания работает.

Таблица 23. Элементы для команды подробного состояния сервера поиска (продолжение)

Элемент	Описание
SessionName	Имя сеанса для индекса поиска этого собрания.

Коды возврата для команд esadmin

Команды esadmin могут возвращать следующие коды:

Таблица 24. Коды возврата для команд esadmin

Код	Название	Описание
0	CODE_ERROR_NONE	Команда выполнена успешно.
102	CODE_ERROR_INSTANTIATION_EXCEPTION	Ошибка при создании экземпляра обработчика команды.
103	CODE_ERROR_ACCESS_EXCEPTION	Ошибка недопустимого доступа при создании экземпляра обработчика команды.
104	CODE_ERROR_EXECUTE_EXCEPTION	
105	CODE_ERROR_THROWABLE	
106	CODE_ERROR_NO_SUCH_METHOD	
107	CODE_ERROR_INVALID_SESSION	
108	CODE_ERROR_INVALID_PARAMETER	
109	CODE_ERROR_SESSION_NOT_RUNNING	

Получение ID сеансов

Команда esadmin check выводит список компонентов поисковой системы предприятия и соответствующих ID сеансов. В следующей таблице показан список общих сеансов, их ID, серверы, на которых они выполняются, и состояния сеансов.

Таблица 25. Примеры имен сеансов, исходных серверов, ID сеансов и состояний сеансов

Сеанс	Сервер, на котором он выполняется	ID сеанса	Состояние сеанса
configmanager	сервер индексов	10433	Started
controller	сервер индексов	10464	Started
customcommunication	сервер индексов	Неприменимо	Неприменимо
datalistener	сервер индексов	10582	Started
discovery	сервер индексов	10649	Started
monitor	сервер индексов	10682	Started
parserservice	сервер индексов	10718	Started
resource.node1	сервер индексов	10759	Started
samplecpp	сервер индексов	10827	Started
sampletest	сервер индексов	10857	Started
scheduler	сервер индексов	10889	Started
searchmanager.node1	сервер индексов	10927	Started
utilities.node1	сервер индексов	10384	Started

Документация поисковой системы предприятия

Документацию по WebSphere Information Integrator OmniFind Edition (поисковой системе предприятия) можно читать в форматах PDF и HTML.

Программа установки WebSphere Information Integrator OmniFind Edition может автоматически установить информационный центр. Программа установки устанавливает информационный центр на сервере поиска. Если информационный центр не установлен, при нажатии кнопки Справка открывается информационный центр на сайте IBM. Для просмотра тем о поисковой системе предприятия в формате HTML запустите информационный центр.

Для просмотра документов в формате PDF перейдите по адресу docs/национальная_версия/pdf. Например, для просмотра русских версий документов перейдите по адресу docs/ru_RU/pdf. Кроме того, просмотр последней документации в формате PDF возможен на сайте поддержки WebSphere Information Integrator OmniFind Edition.

В следующей таблице перечислена доступная документация с указанием имен файлов и их местоположения.

Таблица 26. Документация по поисковой системе предприятия в формате PDF

Заголовок	Заголовок	Заголовок
<i>Руководство по установке поисковой системы предприятия</i> (темы из этого документа доступны также в информационном центре)	iiysi.pdf	docs/национальная_версия/pdf/
<i>Управление поисковой системой предприятия</i> (Темы из этого документа доступны также в информационном центре)	iiysa.pdf	docs/национальная_версия/pdf/
<i>Programming Guide and API Reference for Enterprise Search</i> (Руководство по программированию и справочник по API поисковой системы предприятия; темы из этого документа доступны также в информационном центре)	iiysp.pdf	docs/национальная_версия/pdf/
<i>Messages for Enterprise Search</i> (Сообщения поисковой системы предприятия; темы из этого документа доступны также в информационном центре)	iiysm.pdf	docs/национальная_версия/pdf/

Таблица 26. Документация по поисковой системе предприятия в формате PDF (продолжение)

Заголовок	Заголовок	Заголовок
<i>Installation Requirements for Enterprise Search</i> (Требования для установки поисковой системы предприятия; темы из этого документа доступны также в информационном центре)	iiysr.txt или iiysr.htm	docs/национальная_версия/ (Этот файл можно запустить также из программы Первые шаги.)
<i>Замечания по выпуску</i>	iiysn.pdf	Доступен только на сайте документации по IBM WebSphere Information Integrator OmniFind Edition.
Информационный центр WebSphere Information Integrator	Неприменимо	

Доступность WebSphere II OmniFind Edition

Пользовательские интерфейсы и документация IBM WebSphere Information Integrator OmniFind Edition поддерживают возможности доступности.

Программа установки

По программе установки WebSphere II OmniFind Edition можно перемещаться при помощи быстрых клавиш. В следующей таблице описаны некоторые быстрые клавиши.

Таблица 27. Быстрые клавиши для программы установки

Действие	Клавиши
Выделить радиокнопку	Клавиша со стрелкой
Выбрать радиокнопку	Клавиша Tab
Выделить кнопку	Клавиша Tab
Выбрать кнопку	Клавиша Enter
Перейти в следующее или предыдущее окно или отменить операцию	Выделите кнопку при помощи клавиши Tab и нажмите Enter
Сделать активное окно неактивным	Ctrl + Alt + Esc

Консоль администратора и информационный центр поисковой системы предприятия

Консоль администратора и информационный центр - интерфейсы на основе браузера, с которыми можно работать в Microsoft Internet Explorer или Mozilla FireFox. Список быстрых клавиш и других функций доступности для вашего браузера смотрите в электронной справке по Internet Explorer или FireFox.

Документация в формате PDF

Всю документацию по поисковой системе предприятия можно просматривать в формате PDF. Средства доступности для документов PDF обеспечивает Adobe Acrobat Версии 6.0. Документы PDF структурированы и должны поддерживаться большинством программ чтения с экрана.

Глоссарий терминов поисковой системы предприятия

В этом глоссарии определяются термины, используемые в интерфейсах и документации поисковой системы предприятия.

IP-адрес (IP address)

Уникальный 32-битный адрес, идентифицирующий хост в сети.

Java Database Connectivity (JDBC)

Промышленный стандарт независимой от баз данных связи между платформой Java и различными базами данных. Интерфейс JDBC содержит API уровня вызовов для доступа к базам данных на основе SQL.

JavaScript

Язык сценариев Web, используемый в браузерах и Web-серверах.

JavaServer Pages (JSP)

Технология создания сценариев на сервере, позволяющая динамически встраивать код Java в страницы Web (файлы HTML) и выполнять его при загрузке страницы для возвращения клиенту динамического содержимого.

Secure Sockets Layer (SSL)

Протокол защиты, обеспечивающий конфиденциальность связи.

Unstructured Information Management Architecture (UIMA)

Архитектура IBM, определяющая структуру для реализации систем анализа неструктурированных данных.

административная роль (administrative role)

Классификация пользователя, определяющая функции, которые пользователь может выполнять на консоли управления поисковой системой предприятия. Роль определяет также, какими собраниями пользователь может управлять.

администратор поисковой системы предприятия (enterprise search administrator)

Административная роль, позволяющая пользователю управлять всей поисковой системой предприятия.

анализ ссылок (link analysis)

Метод, основанный на анализе гиперссылок между документами и помогающий определить, какие страницы в собрании важны для пользователей.

анализ текста (text analysis)

Процесс извлечения из текста семантических признаков и другой информации для повышения возможностей получения данных в собрании.

анализатор (parser)

Программа, интерпретирующая документы, добавленные в склад данных поисковой системы предприятия. Анализатор извлекает из документов информацию и подготавливает эти документы для индексации, поиска и получения.

аннотатор (annotator)

Компонент программного обеспечения, выполняющий конкретные задачи лингвистического анализа, генерирующий и записывающий аннотации. Аннотатор - это компонент логики анализа в механизме анализа.

аннотация (annotation)

Информация о характере текста. Например, в аннотации может быть указано, что текст представляет собой название компании. В архитектуре UIMA аннотация - это особая структура признаков.

аннотирование (summarization)

Процесс включения в результаты поиска предложений, кратко описывающих содержимое документа. Смотрите также динамическое аннотирование и статическое аннотирование.

библиотека (library)

Объект системы, служащий каталогом для других объектов. Смотрите также библиотека Domino Document Manager.

библиотека Domino Document Manager (Domino Document Manager library)

База данных Domino Document Manager, служащая входной точкой для Domino Document Manager.

Виртуальная Java-машина (Java virtual machine, JVM)

Программная реализация процессора, на котором выполняется скомпилированный код Java (апплеты и прикладные программы).

внешний источник данных (external data source)

Источник данных для объединения, который не просматривается искателем, не анализируется и не индексируется WebSphere Information Integrator OmniFind Edition. Поиск во внешних источниках данных делегируется API запросов для этих источников данных.

выделение основ (stemming)

Смотрите выделение основ слов.

выделение основ слов (word stemming)

Процесс лингвистического нормирования, при котором различные формы слова приводятся к общей форме. Например, слова *connections*, *connective*, и *connected* приводятся к форме *connect*.

гибридный поиск (hybrid search)

Сочетание логического поиска и свободного текстового поиска.

дерево категорий (category tree)

Иерархия категорий, которая выводится на консоли управления поисковой системой предприятия.

диакритические знаки (diacritics)

Знаки, добавляемые к буквам для изменения произношения слова или выполняющие смысловоразличительную функцию, например, знак ударения или умляут в немецком языке.

динамическое ранжирование (dynamic ranking)

Тип ранжирования, при котором ранг результата определяется в соответствии с вхождением терминов запроса в анализируемые документы. Смотрите также оценка на основе текста. Сравните со статическим ранжированием.

динамическое составление сводок (dynamic summarization)

Тип создания краткого описания документа, при котором выделяются термины поиска и в результаты поиска включаются фразы, в которых наиболее полно представлены искомые понятия документа. Противоположное понятие - статическое аннотирование.

директива no-follow (no-follow directive)

Директива на странице Web, указывающая роботам (например, искателю Web) не переходить по ссылкам с этой страницы.

директива no-index (no-index directive)

Директива на странице Web, указывающая роботам (например, искателю Web) не включать содержимое этих страниц в индекс.

запрос на естественном языке (natural language query)

Тип поиска, при котором анализируются фразы естественного языка (например, "Кто руководит финансовым отделом?"), а не просто набор ключевых слов.

значимое слово (boost word)

Слово, влияющее на ранг релевантности документа в результатах поиска. При обработке запроса важность документа, содержащего значимые слова, может быть повышена или снижена в зависимости от коэффициента, назначенного этому слову.

извлечение информации (information extraction)

Тип извлечения понятий, при котором в тексте документа автоматически распознаются важные словарные элементы, такие как имена, термины и выражения.

извлечение понятий (concept extraction)

Функция поиска, которая находит в текстовых документах важные словарные элементы (например, имена людей, названия мест или продуктов) и генерирует список таких элементов. Смотрите также извлечение темы.

извлечение темы (theme extraction)

Тип извлечения понятий, при котором в тексте документа автоматически распознаются важные словарные элементы для определения темы документа. Смотрите также извлечение понятий.

индекс (index)

Смотрите полнотекстовый индекс.

искатель (crawler)

Программа, получающая документы из источника данных и собирающая информацию, которая может использоваться для создания индексов поиска.

искатель Web (Web crawler)

Класс программ-роботов, исследующих Web путем получения документов Web и следования по содержащимся в них ссылкам.

источник данных (data source)

Любой репозиторий данных, из которого можно получать документы, например, Web, реляционные и нереляционные базы данных и системы управления содержимым.

катакана (Katakana)

Набор символов в одном из двух японских фонетических алфавитов, применяемом главным образом для фонетической записи иностранных слов.

категория (category)

Группа документов со сходными характеристиками.

категория на основе модели (model-based category)

Таксономия из заранее определенных терминов, используемая для определения темы документа, что позволяет при индексации и поиске обрабатывать этот документ вместе с документами со сходным содержанием.

категория на основе правил (rule-based category)

Категория, созданная по правилам, которые задают, как документы связываются с определенными категориями. Например, можно определить правила, связывающие с конкретными категориями документы, содержащие определенные слова или не содержащие такие слова, или документы, соответствующие шаблону URI.

класс значимости (boost class)

Спецификация, влияющая на относительный ранг документа в результатах поиска.

клитика (clitic)

Синтаксически самостоятельное слово, которое произносится слитно с другим словом. Клитика может писаться как слитно, так и раздельно со словом, с которым она связана фонетически. Типичные примеры клитик в английском языке дает стяжение (*wouldn't* или *you're*).

комната (room)

Программа, с помощью которой пользователи могут создавать документы для чтения другими пользователями, отвечать на комментарии других пользователей и просматривать состояние проектов и контрольные сроки их выполнения. Пользователи могут также обмениваться сообщениями в реальном времени с другими посетителями комнаты. Смотрите также комната Lotus QuickPlace.

комната Lotus QuickPlace (Lotus QuickPlace room)

Часть площадки Lotus QuickPlace с разделами, предназначенная только для авторизованных пользователей с общими интересами и потребностью в коллективной работе.

конечный символ (trailing character)

Символ, занимающий последнюю позицию в слове.

контекстная близость (lexical affinity)

Отношение взаимосвязи слов поиска, расположенных в документе близко друг к другу. Контекстная близость используется для вычисления релевантности результата.

контекстный поиск (proximity search)

Тип поиска, при котором ищется вхождение определенных слов в одном предложении, абзаце или документе.

кэш поиска (search cache)

Буфер, в котором хранятся данные и результаты предыдущих требований поиска.

лексический анализатор (tokenizer)

Программа сегментации текста, просматривающая текст и определяющая, какие последовательности символов можно выделить в качестве элементов текста.

лемма (lemma)

Каноническая форма слова. Леммы важны в языках с развитым словоизменением, например, в русском.

лемматизация (lemmatization)

Процесс поиска леммы в словаре для заданного слова. В отличие от выделения основ, лемматизация носит алгоритмический характер и обычно не предполагает работы со словарем языка.

лигатура (ligature)

Два или несколько символов, объединяемых при печати в один, например, в результате слияния букв f и i получается лигатура .

лингвистический поиск (linguistic search)

Тип поиска, при котором просматриваются, получают и индексируются документы, содержащие термины поиска, которые преобразованы в их базовые лексические формы (например, слово *дни* индексируется как *день*) или дополнены до базовых форм (как в случае сложных слов).

логический поиск (boolean search)

Поиск, в котором один или несколько терминов поиска объединяются при помощи логических операций AND, NOT или OR.

локальный федератор (local federator)

Клиентский федератор, объединяющий набор объектов поиска.

маркер защиты (security token)

Информация идентификации и защиты, используемая для авторизации доступа к документам в собрании. Различные типы источников данных поддерживают разные типы маркеров защиты. Примеры маркеров защиты: роли, ID пользователей, ID групп и другая информация, которая может использоваться для управления доступом к содержимому.

механизм анализа (analysis engine)

Смотрите механизм анализа текста.

механизм анализа текста (text analysis engine)

Компонент программного обеспечения, ответственный за поиск и представление контекстного и смыслового содержимого текста.

механизм поиска (search engine)

Программа, принимающая требование поиска и возвращающая пользователю список документов.

наблюдатель (monitor)

Пользователь поисковой системы предприятия с полномочиями наблюдать за процессами на уровне собрания.

начальный URL (seed URL)

Начальная точка для работы искателя.

нормализация символов (character normalization)

Процесс, при котором различные формы символа, например, символы в верхнем регистре или с диакритическими знаками, приводятся к общей форме.

область действия (scope)

Группа взаимосвязанных URI, определяющая область действия требования поиска.

обновление индекса (index refresh)

Процесс добавления новой информации в существующий индекс в поисковой системе предприятия. Сравните с реорганизацией индекса.

обработка архива механизма (processing engine archive)

Файл zip-архива .reag, включающий механизм анализа UIMA и все необходимые ресурсы, позволяющие выполнять с его помощью пользовательский анализ в поисковой системе предприятия.

общая структура анализа (common analysis structure)

Структура, в которой хранится документ, анализируемый механизмом

анализа текста. Информация в общей структуре анализа хранится в форме аннотаций и других структур признаков.

общий уровень связи (Common Communication Layer, CCL)

Инфраструктура связи, объединяющая различные компоненты (контроллер, анализатор, искатель, индексатор) WebSphere Information Integrator OmniFind Edition.

объединенный поиск (federated search)

Возможность поиска, позволяющая выполнять его в нескольких поисковых службах, возвращая объединенный список результатов поиска.

объектная модель документа (Document Object Model)

Система, в которой структурированный документ, например, файл XML, представлен в виде дерева объектов. Программы могут обращаться к этому дереву и изменять его.

оператор (operator)

Пользователь поисковой системы предприятия с полномочиями наблюдать за процессами на уровне собрания, а также запускать и останавливать эти процессы.

определение языка (language identification)

Функция поисковой системы предприятия, определяющая язык документа.

оценка на основе текста (text-based scoring)

Процесс назначения документу целочисленного значения, указывающего степень соответствия документа терминам запроса (релевантность). Большее значение указывает на большую степень соответствия запросу. Смотрите также динамическое ранжирование.

очередь индексации (index queue)

Список ожидающих обработки требований реорганизации индекса или требований обновления индекса.

параметрический поиск (parametric search)

Тип поиска, при котором ищутся объекты, содержащие числовые значения или атрибуты, например, даты, целочисленные значения или числовых данные других типов в заданном диапазоне.

параметры регистрации (credential)

Запрашиваемая при аутентификации детальная информация, описывающая пользователя, его связи с группами и другие атрибуты идентичности, важные для защиты. Параметры регистрации позволяют выполнять разнообразные операции, например, аутентификацию, аудит и делегирование.

площадка (place)

Виртуальная область, которая видна с портала, служащего для совместной работы индивидуальных пользователей и групп. У каждого пользователя в портале есть персональная площадка для индивидуальной работы, кроме того, у индивидуальных пользователей и групп есть доступ к различным совместно используемым площадкам, которые могут быть как общедоступными, так и защищенными. Смотрите также площадка Lotus QuickPlace.

площадка Lotus QuickPlace (Lotus QuickPlace place)

Узел Web, предоставляемый Lotus QuickPlace и позволяющий географически разбросанным участникам совместно работать над проектами и оперативно общаться в структурированном и защищенном рабочем пространстве.

поиск неполных соответствий (fuzzy search)

Поиск, возвращающий слова, написание которых сходно с термином поиска.

поиск по полю (fielded search)

Запрос, применяемый только к конкретному полю.

поиск со взвешенными терминами (weighted term search)

Запрос, в котором некоторые термины заданы как более важные.

поле (field)

Наименьшая идентифицируемая часть записи.

полнотекстовый индекс (full text index)

Структура данных, содержащая ссылки на элементы данных и позволяющая при поиске быстро находить документы, содержащие термины запроса.

пользовательский агент (user agent)

Прикладная программа, просматривающая данные в Web и оставляющая информацию о себе на сайтах, которые она посетила. В поисковой системе предприятия пользовательским агентом является искатель Web.

поместить в очередь (enqueue)

Поместить элементы в очередь.

прикладная программа поиска (search application)

Программа, обрабатывающая запросы, выполняющая поиск в индексе, возвращающая результаты поиска и получающая исходные документы для собраний в поисковой системе предприятия.

произвольная страница ошибок (soft error page)

Специальная страница с объяснениями, которая отправляется, если сервер HTTP не смог вернуть затребованную клиентом страницу, и при этом сконфигурирован так, чтобы вместо ответа из заголовка с кодом возврата ошибки возвращать данную страницу.

прокси-сервер (proxy server)

Сервер, играющий роль посредника для требований HTTP Web, сгенерированных прикладной программой или сервером Web. Прокси-сервер подменяет внешние контент-серверы внутри предприятия.

пространство для искателя (crawl space)

Набор источников, соответствующих заданным шаблонам (таким как имена баз данных, пути файловой системы, имена доменов, IP-адреса и адреса URL), который искатель читает для получения элементов для индексации.

протокол DIIOP (Domino Internet Inter-ORB Protocol)

Задача сервера, запущенная на сервере и взаимодействующая с Domino Object Request Broker для обеспечения связи между апплетами Java, созданными классами Notes Java и сервером Domino. Пользователи браузеров и серверы Domino используют DIIOP для связи и обмена данными объектов.

протокол NRPC (Notes Remote Procedure Call)

Уровень архитектуры Lotus Notes, служащий для связи всех компонентов Notes друг с другом.

протокол Robots Exclusion (Robots Exclusion Protocol)

Протокол, позволяющий администраторам сайтов указывать посещающим эти сайты роботам, какие участки сайтов роботы не должны посещать.

протокол упрощенного доступа к каталогам (Lightweight Directory Access Protocol, LDAP)

Открытый протокол, который использует TCP/IP для доступа к каталогам, поддерживающим модель X.500. Этот протокол не предъявляет таких требований к ресурсам, как более сложный протокол доступа к каталогам X.500 (X.500 Directory Access Protocol).

прямая ссылка (quick link)

Связь между URI и ключевыми словами и словосочетаниями.

путь признака (feature path)

Путь для доступа к значению признака в структуре признаков UIMA.

ранжирование (ranking)

Процесс назначения каждому документу в результатах запроса целочисленного значения. Порядок документов в результатах поиска определяется степенью их соответствия запросу. Более высокий ранг означает более точное соответствие. Смотрите также динамическое ранжирование и статическое ранжирование.

ранжирование по популярности (popular ranking)

Тип ранжирования, при котором к существующему рангу документа прибавляется значение, основанное на степени его популярности.

результаты анализа (analysis results)

Информация, генерируемая аннотаторами. Результаты анализа, соответствующие информации, которую вы хотите искать, записываются в структуру данных под названием общая структура анализа.

результаты поиска (search results)

Список документов, удовлетворяющих требованию поиска.

реорганизация индекса (index reorganization)

Процесс построения индекса в поисковой системе предприятия. Сравните с обновлением индекса.

свободный текстовый поиск (free text search)

Поиск, при котором термины поиска выражены в виде текста в свободном формате.

сегментация (segmentation)

Процесс, посредством которого механизм управления путем передачи данных делит основные единицы информации (basic information units, BIU) на части меньшего размера, называемые сегментами BIU, чтобы согласовать размер передачи с малыми размерами буферов на ближайших серверах.

сегментация с помощью пробелов на основе Unicode (Unicode-based white space segmentation)

Метод лексического анализа, использующий для различения символов элементов и разделителей свойства символов Unicode.

сегментация энграмм (n-gram segmentation)

Метод анализа, при котором перекрывающиеся последовательности из определенного числа символов рассматриваются как одно слово, а не разбиваются на отдельные слова с помощью пробелов, как при сегментации на основе Unicode.

сервлет (servlet)

Программа Java, выполняемая на Web-сервере и расширяющая его функциональные возможности посредством генерирования динамического содержимого в ответ на требования клиента Web. Сервлеты широко используются для подключения баз данных к Web.

сертификат (certificate)

Цифровой документ, связывающий общий ключ с личностью владельца сертификата, обеспечивая тем самым его аутентификацию. Сертификат выдается сертифицированным.

сертификатор (certificate authority)

Организация, выдающая сертификаты и осуществляющая аутентификацию объектов (лиц и организаций) - участников электронных транзакций. Сертификаторы гарантируют, что стороны, обменивающиеся информацией, действительно являются теми, за кого они себя выдают.

символ маски (masking character)

Символ, представляющий необязательные символы в начале, середине или конце аргумента поиска. Обычно символы маски применяются при поиске вариантов написания аргумента поиска в точном индексе. Смотрите также символ подстановки.

символ новой строки (newline character)

Управляющий символ, вызывающий перемещение позиции печати или вывода на одну строку вниз. В некоторых системах для этого требуется несколько символов.

символ подстановки (wildcard character)

Символ, представляющий необязательные символы в начале, середине или конце аргумента поиска.

система объединения (federation)

Процесс объединения систем имен, в результате которого программы могут обрабатывать составные имена, расширяющие эти системы имен.

словарь синонимов (synonym dictionary)

Словарь, дающий возможность при поиске в собрании искать также синонимы терминов запроса.

собрание (collection)

Набор источников данных и опций для просмотра, анализа, индексации и поиска для этих источников данных.

список управления доступом (access control list)

Список, определяющий пользователей, у которых есть доступ к связанному с ним объекту, и задающий права доступа пользователей к этому объекту.

статическое аннотирование (static summarization)

Тип создания краткого описания, при котором в результаты поиска включается заданное и сохраненное краткое описание документа. Противоположное понятие - динамическое аннотирование.

статическое ранжирование (static ranking)

Тип ранжирования, при котором ранг определяется такими характеристиками документа, как дата, число ссылок на этот документ и т.д. Сравните с динамическим ранжированием.

стоп-слово (stop word)

Часто используемое слово (например, *в*, *на* или *и*), которое прикладная программа поиска игнорирует.

структура признаков (feature structure)

Базовая структура данных, представляющая результат анализа текста. Структура признаков включает атрибуты с их значениями. Каждая структура признаков относится к некоторому типу, а у каждого типа есть заданный набор допустимых признаков или атрибутов, что напоминает классы Java.

таксономия (taxonomy)

Классификация объектов по группам на основе сходства характеристик. В поисковой системе предприятия таксономия организует данные в категории и подкатегории. Смотрите также дерево категорий.

тип MIME (MIME type)

Стандарт Интернета для идентификации типа объекта, передаваемого по Интернету.

тип источников данных (data source type)

Группировка источников данных в соответствии с протоколом, используемым для доступа к данным.

удаление стоп-слов (stop word removal)

Процесс удаления незначущих слов из запроса для повышения релевантности результатов поиска.

удаленный федератор (remote federator)

Федератор на сервере, объединяющий набор объектов поиска.

удалить из очереди (dequeue)

Удалить элементы из очереди.

универсальное имя ресурса (Universal Resource Name, URN)

Элемент протокола Интернета в виде короткой строки символов, удовлетворяющих определенному синтаксису. Эта строка содержит имя или адрес для возможной ссылки на ресурс.

универсальный идентификатор ресурса (Uniform Resource Identifier, URI)

Краткая строка символов, идентифицирующая логический или физический ресурс.

универсальный указатель ресурса (Uniform Resource Locator, URL)

Последовательность символов, представляющая информационные ресурсы на компьютере или в сети (например, в Интернете). Эта последовательность символов содержит краткое имя протокола, используемого для доступа к этому информационному ресурсу, и информацию, используемую этим протоколом для определения положения этого информационного ресурса.

уникальное имя (distinguished name)

Имя, уникальным образом идентифицирующее запись в каталоге. Уникальное имя состоит из разделенных запятыми пар атрибут:значение. Это также набор пар "имя-значение" (например CN=имя лица и C=страна или регион), уникально идентифицирующий объект в цифровом сертификате.

управление идентификацией (identity management)

Возможность шифровки параметров регистрации пользователей в защищенном складе.

файл склада ключей (keystore file)

Файл базы данных ключей, содержащий общие ключи, которые хранятся как сертификаты подписавшего, и секретных ключей, хранимых в персональных сертификатах.

файлы индекса поиска (search index files)

Набор файлов, в которых индекс поиска хранится в механизме поиска.

функция обнаружения (discoverer)

Функция искателя, определяющая, из каких источников искатель может получать информацию.

шкаф Domino Document Manager (Domino Document Manager cabinet)

База данных Domino Document Manager, которая служит для организации документов. В шкафах находятся базы данных Domino.

элемент (token)

Базовая единица текста, индексируемая системой поиска предприятия. Элементы могут быть словами языка или другими единицами текста, допускающими индексирование.

эскейп-символ (escape character)

Символ, изменяющий значений одного или нескольких символов, следующих за ним.

язык XML Path (XML Path Language, XPath)

Язык, который уникальным образом определяет или указывает части исходного документа XML. XPath содержит также базовые возможности для манипуляции со строками, числами и логическими операциями.

Получение информации о WebSphere Information Integration

Информацию о продуктах WebSphere Information Integration можно получить по телефону или через Интернет.

Номера телефонов в Соединенных Штатах:

- Чтобы заказать продукты или получить общую информацию: 1-800-IBM-CALL (1-800-426-2255)
- Чтобы заказать публикации: 1-800-879-2755

Кроме того, вы можете найти информацию о WebSphere Information Integration через Интернет по адресу www.ibm.com/software/data/integration/db2ii/. На этом сайте находится следующая информация:

- Документация продукта
- Загружаемые продукты
- Пакеты FixPack
- Замечания по выпуску и другая документация по поддержке
- Новости о WebSphere Information Integration
- Ссылки на ресурсы Web, такие как бюллетени White Papers и книги Redbooks IBM
- Ссылки на группы новостей и группы пользователей
- Ссылки на сетевые информационные центры для продуктов WebSphere Information Integration
- О заказе книг

Чтобы получить документацию по продукту:

1. Посетите страницу Web www.ibm.com/software/data/integration/db2ii/.
2. Выберите продукт в выпадающем списке, например, WebSphere Information Integrator OmniFind Edition.
3. Щелкните по ссылке Support (Поддержка) у левого края страницы.
4. В разделе Learn (Тема) выберите нужную ссылку. Если для выбранного продукта доступен информационный центр, можете выбрать ссылку на информационный центр. Пример смотрите в разделе рис. 2 на стр. 298.

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](#)

Рисунок 2. Пример ссылок на документацию продукта на сайте поддержки WebSphere Information Integration

Отзывы на документацию

Присылайте ваши комментарии об этой книге или о любой другой документации по IBM WebSphere Information Integration.

Ваши отзывы помогут IBM предоставлять качественную информацию. Присылайте ваши комментарии об этом материале или о любой другой документации по WebSphere Information Integration. Для передачи комментариев можете воспользоваться любым из приведенных ниже способов:

1. При помощи формы для ввода читательских замечаний на странице www.ibm.com/software/awdtools/rcf/.
2. Пошлите ваши отзывы электронной почтой по адресу comments@us.ibm.com. Не забудьте указать название продукта, номер версии, название и номер части (если есть). Если ваши комментарии относятся к определенному месту текста, укажите это место (например, заголовок раздела, номер таблицы или номер страницы).

Как связаться с IBM

Чтобы связаться со службой поддержки заказчиков IBM из США или Канады, позвоните 1-800-IBM-SERV (1-800-426-7378).

Чтобы узнать о доступных вариантах обслуживания, позвоните по одному из следующих номеров:

- В США: 1-888-426-4343
- В Канаде: 1-800-465-9600

Адрес отделения IBM в вашей стране или регионе можно найти на странице IBM Directory of Worldwide Contacts в Интернете по адресу www.ibm.com/planetwide.

Товарные знаки

Ниже перечисляются товарные знаки IBM и некоторые товарные знаки других компаний.

Информацию о товарных знаках IBM смотрите по адресу
<http://www.ibm.com/legal/copytrade.shtml>.

Следующие термины являются товарными знаками или зарегистрированными товарными знаками других компаний:

Java и все товарные знаки и логотипы на основе Java - товарные знаки или зарегистрированные товарные знаки Sun Microsystems, Inc. в Соединенных Штатах и/или других странах.

Microsoft, Windows, Windows NT и логотип Windows - товарные знаки Microsoft Corporation в Соединенных Штатах и в других странах.

Intel, Intel Inside (логотипы), MMX и Pentium - товарные знаки Intel Corporation в Соединенных Штатах и/или других странах.

UNIX - зарегистрированный товарный знак The Open Group в Соединенных Штатах и в других странах.

Linux - товарный знак Линуса Торвальдса в Соединенных Штатах и/или других странах.

Названия других компаний, продуктов и услуг могут быть товарными знаками или марками сервиса других фирм.

Замечания

Это информация о продуктах и услугах, предлагаемых в США. IBM может предлагать описанные продукты, услуги и возможности не во всех странах. Сведения о продуктах и услугах, доступных в настоящее время в вашей стране, можно получить в местном представительстве IBM. Любые ссылки на продукты, программы или услуги IBM не означают явным или неявным образом, что можно использовать только продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любых продуктов, программ и услуг других фирм лежит на пользователе.

Фирма IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Получение этого документа не означает предоставления каких-либо лицензий на эти патенты. Запросы по поводу лицензий следует направлять в письменной форме по адресу: IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране/регионе или направьте запрос в письменной форме по адресу: IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

Следующий абзац не применяется в Великобритании или в любой другой стране/регионе, где подобные заявления противоречат местным законам: КОРПОРАЦИЯ INTERNATIONAL BUSINESS MACHINES ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ “КАК ЕСТЬ” БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, ЯВНЫХ ИЛИ ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОВМЕСТИМОСТИ, РЫНОЧНОЙ ПРИГОДНОСТИ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ, НО НЕ ОГРАНИЧИВАЯСЬ ИМИ. В некоторых странах для определенных сделок подобные оговорки не допускаются; таким образом, это утверждение может не относиться к вам.

Данная информация может содержать технические неточности и типографские опечатки. Периодически в информацию вносятся изменения, они будут включены в новые издания этой публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые ссылки в данной информации на сайты, не принадлежащие IBM, приводятся только для удобства и никоим образом не означают поддержки IBM этих сайтов. Материалы этих сайтов не являются частью данного продукта IBM, и вы можете использовать их только на собственную ответственность.

IBM может использовать или распространять присланную вами информацию любым способом, как фирма сочтет нужным, без каких-либо обязательств перед вами.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

IBM Corporation J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Такая информация может быть предоставлена на определенных условиях (в некоторых случаях к таким условиям может относиться оплата).

Лицензированная программа, описанная в данном документе, и все лицензированные материалы, доступные с ней, предоставляются IBM на условиях IBM Customer Agreement (Соглашения IBM с заказчиком), Международного соглашения о лицензиях на программы IBM или эквивалентного соглашения.

Приведенные данные о производительности измерены в контролируемой среде. Таким образом, результаты, полученные в других операционных средах, могут существенно отличаться от них. Некоторые показатели измерены получены в системах разработки и нет никаких гарантий, что в общедоступных системах эти показатели будут теми же. Более того, некоторые результаты могут быть получены путем экстраполяции. Реальные результаты могут отличаться от них. Пользователи должны проверить данные для своих конкретных сред.

Информация о продуктах других фирм получена от поставщиков этих продуктов, из их опубликованных объявлений или из других общедоступных источников. Фирма IBM не проверяла эти продукты и не может подтвердить точность измерений, совместимость или прочие утверждения о продуктах других фирм. Вопросы о возможностях продуктов других фирм следует направлять поставщикам этих продуктов.

Все утверждения о будущих планах и намерениях IBM могут быть изменены или отменены без уведомлений, и описывают исключительно цели фирмы.

Эта информация содержит примеры данных и отчетов, иллюстрирующие типичные деловые операции. Чтобы эти примеры были правдоподобны, в них включены имена лиц, названия компаний и товаров. Все эти имена и названия вымышлены и любое их сходство с реальными именами и адресами полностью случайно.

ЛИЦЕНЗИЯ НА КОПИРОВАНИЕ:

Эта информация содержит примеры прикладных программ на языках программирования, иллюстрирующих приемы программирования для различных операционных платформ. Разрешается копировать, изменять и распространять эти примеры программ в любой форме без оплаты фирме IBM для целей разработки, использования, сбыта или распространения прикладных программ, соответствующих интерфейсу прикладного программирования операционных платформ, для которых эти примера программ написаны. Эти примеры не были всесторонне проверены во всех возможных условиях. Поэтому IBM не может гарантировать их надежность, пригодность и функционирование. Разрешается копировать, изменять и распространять эти примеры программ в любой форме без оплаты фирме IBM для целей разработки, использования, сбыта или распространения прикладных программ, соответствующих интерфейсам прикладного программирования IBM.

Каждая копия программ примеров или программ, созданных на их основе, должна содержать следующее замечание об авторских правах:

Outside In (®) Viewer Technology, ©1992-2005 Stellent, Chicago, IL., Inc. Все права защищены.

IBM XSLT Processor Лицензированные материалы - собственность IBM ©Copyright
IBM Corp., 1999-2005. Все права защищены.

Индекс

A

- API
 - описание 10
 - поиска и индексации 10, 163
 - система приема данных 10
- API поиска и индексации 10, 163
- API приема данных 10

E

- ES_INSTALL_ROOT, описание 19, 20
- ES_NODE_ROOT, описание 19, 20
- es_special_field.default_field 161
- es_special_field.regular_text field 161
- esbackup.bat, сценарий 249, 251
- esbackup.sh, сценарий 249, 251
- escrcm.sh, сценарий 50
- escrcm.vbs, сценарий 51
- esrestore.bat, сценарий 250, 251
- esrestore.sh, сценарий 250, 251

I

- ID прикладных программ 191
- ID сеансов, поисковая система предприятия 253
- ID собрания, правила синтаксиса 28
- IOCP, конфигурация искателя 63

P

- Portal Search Engine
 - защита уровня документов 203
 - описание 201
 - передача маркеров защиты 203

R

- robots.txt, файл 69

S

- SI-API (API поиска и индексации) 10, 163
- startccl, сценарий 250

U

- UIMA
 - добавление механизмов анализа текста в систему 112
 - общие структуры анализа 114, 115
 - описание 111
 - отображение результатов анализа в индексе 114
 - отображение результатов анализа на таблицы JDBC 115
 - отображение элементов XML 113
 - связывание с собраниями 113

URI

- влияние на статические оценки 156, 157
- области действия 136, 137
- объединенные в результатах поиска 138, 139
- правила категорий 101, 105
- просмотр подробностей 221
- прямые ссылки 151
- удаление из индекса 140
- форматы в поисковой системе предприятия 92

W

- WebSphere II Event Publisher Edition, конфигурация искателя DB2 43
- WebSphere II OmniFind Edition 283
 - API 10
 - ID сеансов 253
 - анализаторы 5
 - диаграмма потока данных 12
 - доступность 283
 - изменение пароля на нескольких серверах 20
 - изменение пароля на одном сервере 19
 - интеграция с WebSphere Portal 201
 - коды возврата 253
 - команды 253
 - компоненты 3
 - консоль администратора 9
 - конфигурирование номера порта 24
 - обзор 1
 - программы поиска 11
 - серверы индексов 6
 - серверы искателей 4
 - серверы поиска 8
 - управление идентификацией 194
- WebSphere MQ, конфигурация искателя DB2 46
- WebSphere MQ, конфигурация сервера искателя 47, 48
- WebSphere Portal
 - внедрение портлета поиска 203
 - интеграция с поисковой системой предприятия 201
 - параметры перенастройки по умолчанию 212
 - перенастройка дерева категорий 209
 - перенастройка собрания 209
 - перенастройка таксономии 209
 - портлет поиска и просмотра 205
 - портлет управления таксономией 209
 - таксономии на основе моделей 209
- WebSphere Portal Search Center
 - адаптер для поисковой системы предприятия 201
 - описание 201
 - портлет регистрации 201, 207

X

- XPath, собственный поиск XML 117

A

- адаптер для поисковой системы предприятия 206
- административные роли
 - администратор поисковой системы предприятия 184, 185
 - администратор собрания 184, 185
 - конфигурирование 185
 - наблюдатель 184, 185
 - оператор 184, 185
 - описание 184
- администратор поисковой системы предприятия
 - изменение пароля на нескольких серверах 20
 - изменение пароля на одном сервере 19
 - конфигурирование роли 185
 - описание 184
- администратор собрания
 - конфигурирование роли 185
 - описание 184
- активные сайты, слежение 224, 225
- анализ текста
 - механизмы анализа текста 112, 113
 - общие структуры анализа 114, 115
 - отображение элементов XML 113
- анализ текста привязки
 - глобальный анализ 189
 - защита собрания 189
 - индексация документов 190
 - описание 183
- анализаторы
 - анализ сложных терминов 116
 - задачи анализа данных 5
 - запуск 231
 - описание 5, 99
 - остановка 231
 - поток 116
 - пустое пространство 118
 - сегментация энграмм 118
 - символы новой строки 118
 - слежение 231
 - собственный поиск XML 117
 - состояние системы 231
 - типы анализа документов 119
 - типы документов для сеансов Stellent 122, 124
 - типы документов для служб анализатора 120, 121
- аннотаторы 111
- аутентификация
 - описание 186
 - отключение для прикладных программ предприятия 187
 - аутентификация на основе форм 78, 79

Б

базовая аутентификация HTTP 78
брандмауэры, работа искателя с документами Exchange Server 55

В

важность документов
в перенастроенных собраниях 210
включение для собрания 28
классы значимости 158, 161
словари значимых слов 156
статическое 153
шаблоны URI 156, 157
включение расписаний индекса 129
внедрение
адаптер для поисковой системы предприятия 206
портлет поиска и просмотра 205
портлет регистрации 207
портлета поиска 203
файл ESPACServer.ear 87
файл ESSearchAdapter.ear 206
файл ESSearchAdapterPortlet.war 207
файл ESSearchPortlet.war 203
внедрение портлета поиска 203
внешние искатели
API приема данных 10
конфигурирование 89
программы приема данных 89
внешние источники
защита ID прикладных программ 191
конфигурирование 179
описание 179
поиск 131
связывание с прикладными программами поиска 180
внешние источники JDBC
драйверы JDBC 179
конфигурирование 179
редактирование 179
удаление 179
внешние источники LDAP
конфигурирование 179
редактирование 179
удаление 179
восстановление
из резервной копии 250
файлов системы на новые серверы 251
восстановление поисковой системы предприятия 249
восстановление системы 249, 250, 251
время ответа при поиске
оповещения 240
слежение 234
выключение расписаний индекса 129

Г

глобальная защита WebSphere
отключение 187
свойства программы поиска 174
глобальное пространство для искателя Web 83

глобальный анализ
анализ текста привязки 183, 189
обнаружение повторений документов 183, 189
описание 6
глубина путей URL 71

Д

деревья категорий
описание 104
перенастройка из WebSphere Portal 209, 210
динамическое ранжирование 153
динамическое составление сводок 149
директивы no-follow
конфигурирование 85
описание 84
директивы no-index
конфигурирование 85
описание 84
доверенные серверы Lotus Domino 197
документация 281
документация в формате PDF 281
документы HTML, поиск 109
Документы XML
поиск 107
собственный поиск XML 117
домены Lotus Domino 197
домены Windows 195
доступность 283

З

запросы с высоким значением возврата
коэффициенты значимости по умолчанию 161
описание 158
запросы с низким значением возврата
коэффициенты значимости по умолчанию 161
описание 158
запуск
анализаторы 231
мастер по перенастройке 209, 210
построение индекса 232
пример программы поиска 175
серверы искателей 223
серверы поиска 234
серверы поисковой системы предприятия 215
система приема данных 235
защита
ID прикладных программ поиска 191
административные роли 185
анализ текста привязки 189
аутентификация 186, 187
включение для собрания 28, 183
глобальная защита WebSphere 187
документы Lotus Domino 197
домены Windows 195
обнаружение повторений документов 189
обход проверки управления доступом на уровне документов 199
описание 183

защита (продолжение)

отключение для прикладной программы предприятия 187
подключаемые модули искателей 90
пример программы поиска 176
профили пользователей 193
управление доступом 186
управление идентификацией 194
уровня документов 191, 192, 194, 199
уровня собрания 188, 199
защита Local User, искатели QuickPlace 66
защита уровня документов
для документов Lotus Domino 197
для файловых систем Windows 195
документы Lotus Domino 197
конфигурация искателей 33
маркеры защиты 192
описание 183, 191
поддержка Portal Search Engine 203
подключаемые модули искателей 90
проверка в реальном времени 192
проверка параметров регистрации текущего пользователя 192
профили пользователей 193
управление доступом через индекс 192
управление идентификацией 194
защита уровня собрания
ID прикладных программ 191
анализ текста привязки 189
включение 28
обнаружение повторений документов 189
описание 183, 188

И

индексы
включение расписания 129, 232
изменение расписания 129
области действия 131, 136
объединенные URI 131, 138, 139
одновременное построение 130
описание 6, 127
оповещения 240
отключение расписания 129, 232
параллельное построение 130
планирование 128
символы подстановки 131, 133, 136
слежение 232, 233
текст привязки 190
удаление URI 131, 140
удаление из очереди 233
форматы URI 92
интервалы повторного просмотра для искателей Web 75
искатели
Content Edition 37, 38, 39
DB2 42
DB2 Content Manager 49
Domino Document Manager 53
Exchange Server 54, 55
NNTP 55
Notes 56
QuickPlace 65
Web 69

- искатели (*продолжение*)
 - WebSphere Portal 86, 87, 88
 - базовые значения для 33
 - включение защиты на уровне документов 33
 - защита уровня документов 191
 - комбинирование типов искателей 33
 - начальные значения для 35
 - обзор конфигурации 33
 - описание 4
 - параметры перенастройки по умолчанию 212
 - планирование 33
 - поддержка для внешних 10, 89
 - подключаемые модули 90
 - Примечания 58
 - программы приема данных 89
 - редактирование пространств для искателей 36
 - редактирование свойств искателя 36
 - слежение 223
 - создание 35
 - состояние системы 223
 - удаление 37
 - файловая система UNIX 68
 - файловая система Windows 88
 - форматы URI 92
- искатели Content Edition
 - конфигурирование в операционной системе AIX 39
 - конфигурирование в операционной системе Linux 39
 - режим сервера 38
- Искатели Content Edition
 - конфигурация 37
 - конфигурирование в Windows 41
 - конфигурирование в операционной среде Solaris 39
 - непосредственный режим 39
 - форматы URI 92
- искатели DB2
 - конфигурация WebSphere II Event Publisher Edition 43
 - конфигурация WebSphere MQ 46
 - конфигурирование публикации событий 47, 48
 - публикация событий 42
 - установка WebSphere MQ 47
 - установка WebSphere MQ в Windows 48
- Искатели DB2
 - конфигурация 42
 - форматы URI 92
- Искатели DB2 Content Manager
 - конфигурация 49
 - конфигурирование в Windows 51
 - конфигурирование в операционной системе AIX 50
 - конфигурирование в операционной системе Linux 50
 - конфигурирование в операционной среде Solaris 50
 - форматы URI 92
- искатели Domino Document Manager
 - конфигурирование в Windows 61
 - конфигурирование в операционной системе Linux 59
- Искатели Domino Document Manager
 - конфигурация 53
 - конфигурация протокола DIIOP 62
 - конфигурирование IOCP 63
 - конфигурирование в операционной системе AIX 59
 - конфигурирование в операционной среде Solaris 59
 - протокол NRPC 59, 61
 - форматы URI 92
- искатели Exchange Server
 - защищенные документы 55
 - конфигурация 54
 - форматы URI 92
- искатели NNTP, конфигурирование 55
- искатели Notes
 - доверенный сервер Lotus Domino 197
 - конфигурация 56
 - конфигурация протокола DIIOP 62
 - конфигурирование IOCP 63
 - конфигурирование в Windows 61
 - конфигурирование в операционной системе AIX 59
 - конфигурирование в операционной системе Linux 59
 - конфигурирование в операционной среде Solaris 59
 - конфигурирование защиты уровня документов 197
 - правила отображения полей 58
 - проверка параметров регистрации текущего пользователя 197
 - протокол NRPC 59, 61
 - советы по использованию 58
 - форматы URI 92
- искатели QuickPlace
 - защита Local User 66
 - конфигурирование Directory Assistance 67
 - конфигурирование в Windows 61
 - конфигурирование пользователя Domino 66
- Искатели QuickPlace
 - конфигурация 65
 - конфигурация протокола DIIOP 62
 - конфигурирование IOCP 63
 - конфигурирование в операционной системе AIX 59
 - конфигурирование в операционной системе Linux 59
 - конфигурирование в операционной среде Solaris 59
 - протокол NRPC 59, 61
 - форматы URI 92
- искатели Web
 - активные сайты 224, 225
 - директивы no-follow 84, 85
 - директивы no-index 84, 85
 - интервалы повторного просмотра 75
 - конфигурирование опознавательных файлов 82
 - начальные URL 76
 - недавно просмотренные URL 224
 - опознавательные файлы (cookie) 81
 - поддержка JavaScript 70
 - подробности потока 224, 225
- искатели Web (*продолжение*)
 - посещение URL при первой возможности 76
 - правила для искателя 71
 - произвольные страницы ошибок 76
 - прокси-серверы 80
 - сайты, защищенные паролями 78, 79
 - скорость работы искателя 226
 - слежение 224
 - создание отчетов 227
 - состояние URL 224
 - состояние системы 224
 - файл followindex.rules 84, 85
 - формат опознавательных файлов 81
 - хронология искателя 224
 - хронология сайта 224
- Искатели Web
 - глобальное пространство для искателя 83
 - конфигурация 69
 - начальные URL 71
 - ограничение пространства для искателя 71
 - пользовательские агенты 69
- искатели WebSphere Portal
 - копирование URL сайтов 88
- Искатели WebSphere Portal
 - внедрение прикладной программы предприятия 87
 - внедрение файла ESPACServer.ear 87
 - конфигурация 86
 - форматы URI 92
- Искатели файловой системы UNIX
 - конфигурация 68
 - форматы URI 92
- искатели файловой системы Windows
 - конфигурирование защиты уровня документов 195
 - форматы URI 92
- Искатели файловой системы Windows
 - конфигурация 88

К

- категории
 - вложенные подкатегории 104
 - деревья категорий 104
 - на основе моделей 103, 104
 - на основе правил 101, 104
 - описание 100
 - перенастройка из WebSphere Portal 209, 210
 - поиск 100
 - создание 105
 - тип категоризации 104
 - форматы URI 92
- категории на основе моделей
 - выбор типа категоризации 104
 - описание 103
- категории на основе правил
 - выбор типа категоризации 104
 - описание 101
 - создание 105
- китайский
 - анализ символов новой строки 118
 - сегментация энграмм 118

- классы значимости
 - запросы с высоким значением возврата 158, 161
 - запросы с низким значением возврата 158, 161
 - значения по умолчанию 161
 - конфигурация 159, 160
 - обнаружение повторений документов 158
 - описание 158
 - отображение полей на 159
- ключевые слова в прямых ссылках 151
- коды возврата HTTP
 - отчет искателя Web 227
 - полученные искателями Web 227
- коды возврата, поисковая система предприятия 253
- команда esadmin 253
- команды, поисковая система предприятия 253
- консоль администратора
 - интерфейс 15
 - описание 9
 - регистрация 18
 - сводка задач 15
- конфигурация SMTP-сервера 243
- конфигурация доверенного сервера 197
- конфигурирование Directory Assistance 67
- конфигурирование пользователя Domino, искателя QuickPlace 66
- конфигурирование портлета поиска и просмотра 205
- корейский
 - анализ сложных терминов 116
 - сегментация энграмм 118
- коэффициенты значимости
 - для классов значимости 159, 161
 - для словарей значимых слов 154
 - для шаблонов URI 156, 157
 - конфигурация класса значимости 158, 160
- кэш поиска
 - конфигурирование 144
 - описание 144

Л

- лингвистическая поддержка
 - пользовательский анализ текста 111
 - семантический поиск 111, 117
 - словари значимых слов 154
 - словари синонимов 144
 - словари стоп-слов 147
 - собственный поиск XML 117

М

- максимальный интервал повторного просмотра 75
- маркеры защиты
 - защита уровня документов 192, 203
 - как запретить для собрания 199
 - конфигурация искателей 192
 - обработка Portal Search Engine 203
- мастер по перенастройке
 - запуск 209, 210

- мастер по перенастройке (*продолжение*)
 - описание 209
 - параметры искателей по умолчанию 212
 - параметры собраний по умолчанию 212
 - собрания 210
 - таксономии на основе моделей 209
 - таксономии на основе правил 210
 - файл журнала 214
- мастер по собраниям 27
- механизмы анализа текста
 - добавление в систему 112
 - описание 111
 - отображение результатов анализа 114, 115
 - отображение элементов XML 113
 - связывание с собраниями 113
- минимальный интервал повторного просмотра 75
- модуль bos.iocp.rte 63
- модуль IOCP, конфигурирование искателя 63

Н

- наблюдатель
 - конфигурирование роли 185
 - описание 184
- начальные URL для искателей Web 71, 76
- недавно просмотренные URL, слежение 224
- непосредственный режим, репозитории Content Edition 39
- номер порта, поисковая система предприятия 24

О

- области действия
 - описание 136
 - поиск 136
 - создание 137
 - форматы URI 92
 - шаблоны URI 136, 137
- обнаружение 4
- обнаружение повторений документов
 - включение защиты 189
 - глобальный анализ 189
 - конфигурация класса значимости 158
 - описание 183
- обновление индекса
 - описание 6, 127
 - планирование 128, 129
- обновление индексов 6, 127
- обработка текста
 - аннотаторы 111
 - механизмы анализа текста 111
 - общие структуры анализа 111
- общие структуры анализа
 - описание 111
 - отображение на индекс 114
 - отображение на таблицы JDBC 115
 - отображение элементов XML на 113
- объединенные URI
 - конфигурирование 139

- объединенные URI (*продолжение*)
 - описание 138
- объединенные результаты поиска
 - конфигурирование 139
 - описание 138
- ограничение пространства для искателя Web 71
- Одновременное построение индексов 130
- окно Защита, описание 15
- окно Система, описание 15
- окно Слежение, описание 15
- Окно Собрания
 - описание 15
 - создание собраний 28
- оператор
 - конфигурирование роли 185
 - описание 184
- операционная система AIX
 - конфигурация искателя Content Edition 39
 - конфигурация искателя DB2 Content Manager 50
 - конфигурация искателя Domino Document Manager 59
 - конфигурация искателя Notes 59
 - конфигурация искателя QuickPlace 59
 - конфигурирование публикации событий 47
- операционная система Linux
 - конфигурация искателя Content Edition 39
 - конфигурация искателя DB2 Content Manager 50
 - конфигурация искателя Domino Document Manager 59
 - конфигурация искателя Notes 59
 - конфигурация искателя QuickPlace 59
 - конфигурирование публикации событий 47
- операционная среда Solaris
 - конфигурирование публикации событий 47
- операционная система Windows
 - конфигурация искателей 61
 - конфигурация искателя Content Edition 41
 - конфигурация искателя DB2 Content Manager 51
 - конфигурирование публикации событий 48
- операционная среда Solaris
 - конфигурация искателя Content Edition 39
 - конфигурация искателя DB2 Content Manager 50
 - конфигурация искателя Domino Document Manager 59
 - конфигурация искателя Notes 59
 - конфигурация искателя QuickPlace 59
- оповещения
 - время ответа при поиске 240
 - документы, просматриваемые искателем 240
 - индексируемые документы 240
 - конфигурация SMTP-сервера 243
 - ограничения на индекс 219
 - описание 239

- оповещения (*продолжение*)
 - опции электронной почты 240, 241
 - получение электронной почты 244
 - свободное место на серверах 241
 - уровня системы 241
 - уровня собрания 219, 240
 - оповещения о свободном месте 241
 - опознавательные файлы для искателя Web
 - конфигурирование 82
 - описание 81
 - формат 81
 - опции поиска
 - для результатов поиска 131
 - параметрический поиск 131
 - поиск по полю 131
 - свободный текстовый поиск 131
 - содержимое документов 131
 - опции содержимого документов 131
 - остановка
 - анализаторы 231
 - построение индекса 232, 233
 - серверы искателей 223
 - серверы поиска 234
 - серверы поисковой системы предприятия 215, 217
 - отброшенные документы
 - конфигурирование файлов журналов 236
 - описание 236
 - отчеты 237
 - отображение
 - общие структуры анализа в индексе 114
 - общие структуры анализа на таблицы JDBC 115
 - полей на классы значимости 159
 - поля поиска HTML 110
 - поля поиска XML 107
 - элементов XML на общие структуры анализа 113
 - отслеживание документов
 - включение 236
 - конфигурирование файлов журналов 236
 - описание 236
 - отключение 236
 - отчеты 237
 - оценка на основе текста 153
 - оценка системных ресурсов 219
 - очередь индексации 233
- П**
- параллельное построение индексов 130
 - параметрический поиск 131
 - пароль администратора
 - изменение на нескольких серверах 20
 - изменение на одном сервере 19
 - пароль, администратор поисковой системы предприятия 19, 20
 - перенастройка
 - собрания 210
 - таксономия на основе модели 209
 - таксономия на основе правил 210
 - планирование
 - искатели 33
 - построение индекса 128, 129
 - повторное посещение URL при первой возможности 76
 - поддержка JavaScript в искателях Web 70
 - подключаемые модули искателей 90
 - подключаемые модули, для искателей 90
 - подробности об URI
 - отброшенные документы 237
 - слежение 221
 - подробности потока, отслеживание 224
 - поиск
 - внешние источники 131
 - документы HTML 109, 110
 - Документы XML 107, 113
 - категории 100
 - прямые ссылки 151
 - собрания 131
 - поиск документации поисковой системы предприятия 281
 - поиск по полю 131
 - поисковая система предприятия
 - API 10
 - ID сеансов 253
 - административные роли 184
 - анализаторы 5, 99
 - восстановление из резервной копии 250
 - восстановление файлов системы 251
 - диаграмма потока данных 12
 - запуск серверов 215
 - защита 183
 - защита уровня документов 191
 - защита уровня собрания 188
 - интеграция с WebSphere Portal 201
 - коды возврата 253
 - команды 253
 - компоненты 3
 - консоль администратора 9
 - конфигурирование номера порта 24
 - обзор 1
 - остановка серверов 215, 217
 - программы поиска 11
 - резервное копирование 249
 - серверы индексов 6, 127
 - серверы искателей 4, 33
 - серверы поиска 8, 143
 - слежение 219
 - сценарии восстановления 249
 - сценарии резервного копирования 249
 - файлы журнала 239
 - форматы URI 92
 - пользовательские агенты 69
 - пользовательский анализ текста
 - механизмы анализа текста 112, 113
 - описание 111
 - отображение общих структур анализа 114, 115
 - отображение результатов анализа в индексе 114
 - отображение результатов анализа на таблицы JDBC 115
 - отображение элементов XML 113
 - поля поиска HTML
 - описание 109
 - отображение элементов 109, 110
 - создание 110
 - поля поиска XML
 - описание 107
 - поля поиска XML (*продолжение*)
 - отображение элементов 107
 - создание 107
 - Поля поиска XML
 - описание 113
 - отображение элементов 113
 - поля, отображение на классы значимости 159
 - портлет управления таксономией 209
 - портлеты
 - описание 201
 - поддерживаемые версии WebSphere Portal 201
 - Поиск 203
 - поиск и просмотр 205
 - поисковая система предприятия 201
 - портлет управления таксономией 209
 - регистрация 207
 - файл ESSearchAdapterPortlet.war 207
 - посещение URL при первой возможности 76
 - последние запросы, слежение 234
 - построение индекса
 - запуск 232
 - одновременное 130
 - остановка 232, 233
 - параллельное 130
 - состояние системы 233
 - поток данных, поисковая система предприятия 12
 - потоки
 - анализатор 116
 - искатель Web 225
 - правила IP-адресов для искателей Web 71
 - правила адресов для искателей Web 71
 - правила доменов для искателей Web 71
 - правила категорий
 - конфигурирование 105
 - содержимое документов 101, 105
 - шаблоны URI 101, 105
 - правила префиксов для искателей Web 71
 - прикладные программы предприятия
 - программа ESAdmin 187
 - программа ESSearchApplication 187
 - программа ESSearchServer 187
 - файл ESPACServer.ear 87
 - файл ESSearchAdapter.ear 206
 - пример программы поиска
 - включение защиты 176
 - внедрение по умолчанию 174
 - глобальная защита WebSphere 176
 - доступ 175
 - запуск 175
 - описание 11, 164
 - отключение защиты 187
 - файл config.properties 165, 174
 - функции поиска 163, 164
 - проверка для запроса 192
 - проверка параметров регистрации текущего пользователя 192, 195, 197
 - проверка системных ресурсов 220
 - программа ESAdmin
 - отключение защиты 187
 - регистрация 18
 - программа ESSearchApplication
 - включение защиты 176
 - отключение защиты 187

- программа ESSearchApplication (продолжение)
 - перезапуск 174, 176
 - файл config.properties 165, 174
- программа ESSearchServer
 - отключение защиты 187
- программа поиска по умолчанию 174
- Программа приема данных
 - конфигурирование 89
 - перезапуск 89
- программы поиска
 - ID прикладных программ 191
 - защита уровня собрания 191
 - описание 11
 - пользовательские 163
 - пример 163, 164
 - связывание с внешними источниками 180
 - связывание с собраниями 163
- произвольные страницы ошибок, искатели Web 76
- прокси-серверы 80
- прокси-серверы HTTP 80
- просмотр
 - подробности об URI 221
 - файлы журнала 246
- пространство для искателя
 - конфигурация искателя Web 71
 - описание 4
 - оповещения 240
 - редактирование 36
- протокол DIIOP, конфигурация искателя 62
- протокол NRPC, конфигурирование искателя 59, 61
- протокол Robots Exclusion 69
- профили пользователей
 - конфигурация 194
 - описание 193
- прямые ссылки
 - описание 151
 - поиск 151
 - создание 151
 - форматы URI 92
- публикация событий
 - конфигурация искателя DB2 43, 46
 - конфигурирование в Windows 48
 - конфигурирование в операционной системе AIX 47
 - конфигурирование в операционной системе Linux 47
 - конфигурирование в операционной среде Solaris 47
 - описание 42

Р

- ранжирование результатов поиска
 - динамическое 153
 - классы значимости 158, 159, 160, 161
 - описание 153
 - оценка на основе текста 153
 - словари значимых слов 156
 - статическое 153
 - шаблоны URI 156, 157

- расширения файлов
 - исключение из пространств для искателей Web 71
 - поддерживаемые анализатором собраний 120, 121
 - поддерживаемые в сеансах Stellent 122, 124
- регистрация на консоли администратора 18
- редактирование
 - программы приема данных 89
 - пространства для искателей 36
 - свойства искателя 36
 - свойства программы поиска 165, 174
 - собрания 30
- режим сервера, репозитории Content Edition 38
- резервная копия системы 249
- резервное копирование поисковой системы предприятия 249
- результаты поиска
 - группировка 138, 139
 - динамическое ранжирование 153
 - динамическое составление сводок 149
 - конфигурация класса значимости 158, 159, 160, 161
 - конфигурирование шаблонов URI 156
 - настройка сводок 149, 150
 - объединение 138, 139
 - описание 153
 - оценка на основе текста 153
 - ранжирование 157
 - расширение символов подстановки 136
 - сводки 149, 150
 - символы подстановки 133
 - статическое ранжирование 153
- реорганизация индекса
 - описание 6, 127
 - планирование 128, 129
- реорганизация индексов 6, 127

С

- сайты, защищенные паролями 78
- аутентификация на основе форм 79
- базовая аутентификация HTTP 78
- свободный текстовый поиск 131
- сводка задач, консоль управления 15
- сводки
 - динамическое 149
 - настройка 149, 150
- сводки документов
 - настройка 149
 - редактирование свойств для 150
- свойства искателя
 - описание 4
 - редактирование 36
- сеансы Stellent
 - связывание типов документов 122
 - типы анализа документов 119
 - типы документов по умолчанию 124
- сегментация энграмм 118
- семантический поиск 111, 113, 117
- серверы анализаторов
 - запуск 215
 - конфигурация потоков 116

- серверы анализаторов (продолжение)
 - остановка 217
- серверы индексов
 - запуск 215
 - остановка 217
- серверы искателей
 - запуск 215, 223
 - остановка 217, 223
- серверы поиска
 - запуск 215, 234
 - кэш поиска 144
 - описание 8, 143
 - остановка 217, 234
 - последние запросы 234
 - связанные с ними словари синонимов 146
 - связывание значимых слов со словарями 156
 - связывание словарей стоп-слов 148
 - связывание со словарями значимых слов 155
 - слежение 234
 - словари значимых слов 154
 - словари синонимов 144, 146
 - словари стоп-слов 147
 - состояние системы 234
 - хронология времени ответа 234
 - частые запросы 234
- символы подстановки
 - расширение запроса 133, 136
 - расширение индекса 133, 136
- синтаксис запроса XML, собственный 117
- система приема данных
 - перезапуск 235
 - слежение 235
- системные ресурсы
 - оценка 219
 - проверка 220
- скорость работы искателя, слежение 226
- слежение
 - Активные сайты искателя Web 225
 - анализаторы 231
 - искатели 223
 - искатели Web 224
 - отброшенные документы 237
 - подробности об URI 221
 - Подробности потока искателя Web 225
 - поисковая система предприятия 219
 - последние запросы 234
 - серверы поиска 234
 - система приема данных 235
 - Скорость просмотра искателя Web 226
 - собрания 221
 - файлы журнала 246
 - хронология времени ответа 234
 - частые запросы 234
- словари значимых слов
 - добавление в систему 155
 - описание 154
 - связывание с собранием 156
- словари синонимов
 - добавление в систему 146
 - описание 144
 - связывание с собранием 146

- словари стоп-слов
 - добавление в систему 148
 - описание 147
 - связывание с собранием 148
- сложные термины, анализ 116
- собрания
 - защита 188
 - защита ID прикладных программ 191
 - защита повторений документов 189
 - защита текста привязки 189
 - обход проверки управления доступом на уровне документов 199
 - описание 3
 - оценка размера 28
 - оценка ресурсов 219
 - параметры перенастройки по умолчанию 212
 - перенастройка из WebSphere Portal 209, 210
 - поиск 131
 - редактирование 30
 - связывание с прикладными программами поиска 163
 - серверы поиска 143
 - синтаксический анализ 99
 - система объединения 27
 - слежение 221
 - создание в окне Собрания 28
 - создание с помощью мастера по собраниям 27
 - состояние системы 221
 - способы создания 27
 - удаление 31
 - файл MigrationWizard.log 210
 - черновое состояние 27
- собрания объединения 27
- собственный поиск XML 117
- соединитель Java для DB2 Content Manager 50, 51
- создание
 - искатели 35
 - категории на основе правил 105
 - области действия 137
 - отчеты искателя Web 227
 - поля поиска HTML 110
 - поля поиска XML 107
 - прямые ссылки 151
 - собрания 27, 28
- сообщения об ошибках
 - конфигурация SMTP-сервера 243
 - получение электронной почты 242, 244
 - просмотр файлов журнала 246
- состояние системы
 - анализаторы 231
 - искатели 223
 - искатели Web 224
 - построение индекса 233
 - серверы поиска 234
 - собрания 221
- статическое ранжирование
 - в перенастроенных собраниях 210
 - включение для собрания 28
 - описание 153
- страница Анализ, описание 15
- страница Журнал, описание 15
- страница Индекс, описание 15

- страница Искатель, описание 15
- страница Поиск, описание 15
- сценарии
 - esbackup.bat 249, 251
 - esbackup.sh 249, 251
 - escrcm.sh 50
 - escrcm.vbs 51
 - escrdb2.sh 47
 - escrdb2.vbs 48
 - escrnote.sh 59
 - escrnote.vbs 61
 - escrvbr.sh 39
 - escrvbr.vbs 41
 - esrestore.bat 250, 251
 - esrestore.sh 250, 251
 - startccl 250
- сценарии восстановления
 - запуск 250, 251
 - описание 249
- сценарии резервного копирования
 - запуск 249, 251
 - описание 249
- сценарий eschangepw 19, 20
- сценарий escrdb2.sh 47
- сценарий escrdb2.vbs 48
- сценарий escrnote.sh 59
- сценарий escrnote.vbs 61
- сценарий escrvbr.sh 39
- сценарий escrvbr.vbs 41

Т

- таксономии, перенастройка из WebSphere Portal 209, 210
- таксономия на основе модели, перенастройка из WebSphere Portal 209
- таксономия на основе правил, перенастройка из WebSphere Portal 209, 210
- тип категоризации
 - выбор 28, 104
 - на основе моделей 103
 - на основе правил 101
- типы MIME, включение в пространства для искателей Web 71
- типы документов
 - для сеансов Stellent 122, 124
 - для служб анализатора 120, 121
 - синтаксический анализ 119
- типы искателей
 - базовые значения для 33
 - объединение в собрании 33
- типы источников данных
 - базы данных DB2 42
 - базы данных Domino Document Manager 53
 - базы данных Notes 56, 58
 - базы данных QuickPlace 65
 - группы новостей NNTP 55
 - общедоступные папки Exchange Server 54
 - поддерживаемые поисковой системой предприятия 2
 - поддержка для внешних 2, 10
 - реляционные базы данных 42
 - репозитории Content Edition 37, 38, 39
 - сайты 69

- типы источников данных *(продолжение)*
 - сайты WebSphere Portal 86
 - типы элементов DB2 Content Manager 49
 - файловые системы UNIX 68
 - файловые системы Windows 88

У

- уведомления по электронной почте
 - конфигурация SMTP-сервера 243
 - о сообщениях 244
 - об оповещениях 244
- удаление
 - индексов из очереди 233
 - искатели 37
 - собрания 31
- удаление URI из индекса 140
- управление доступом
 - защита уровня документов 192, 203
 - как запретить для собрания 199
 - описание 186
 - поддержка Portal Search Engine 203
 - проверка параметров регистрации текущего пользователя 192
 - профили пользователей 193
 - требования для Lotus Domino 197
 - требования для файловых систем Windows 195
- управление идентификацией
 - конфигурация 194
 - профили пользователей 193

Ф

- файл CCLServer_date.log 24
- файл config.properties
 - описания свойств 165
 - редактирование 174
- файл cookies.ini
 - конфигурирование 82
 - описание 81
 - формат 81
- файл crawl.rules 83
- файл es.cfg 19, 20
- файл ESPACServer.ear 87
- файл ESSearchAdapter.ear 206
- файл ESSearchAdapterPortlet.war 207
- файл ESSearchPortlet.war 203
- файл followindex.rules
 - конфигурирование 85
 - описание 84
- файл global.rules 83
- файл parserTypes.cfg 120
- файл runtime-generic.properties 150
- файл stellent.properties 122
- файл stellenttypes.cfg 122
- файл synonyms.xml 209
- файл titles.xml 209
- файл treenodes.xml 209
- файл vbr_access_services.jar 39, 41
- файл WpsMigratorLog.log 214
- файлы EAR
 - ESPACServer.ear 87
 - ESSearchAdapter.ear 206
 - программа ESAdmin 187

- файлы EAR *(продолжение)*
 - программа ESSearchApplication 187
 - программа ESSearchServer 187
- файлы журнала
 - для отслеживания документов 236
 - конфигурация SMTP-сервера 243
 - максимальный размер 242
 - мастер по перенастройке 214
 - описание 239
 - опции электронной почты 244
 - переход к новым 242
 - положение по умолчанию 239
 - просмотр 246
 - слежение 246
 - уровни серьезности 242
 - фильтрация 246
- фрагменты XML, собственный поиск
 - XML 117

Х

- хронологические отчеты искателя
 - описание 224
 - отчет по кодам возврата HTTP 227
 - отчет по сайту 227
 - создание 227
- хронологические отчеты по сайтам
 - описание 224
 - создание 227
- хронология времени ответа,
 - слежение 234

Ц

- Центр поиска WebSphere Portal
 - адаптер для поисковой системы предприятия 206
- Центр поиска для WebSphere Portal
 - адаптер для поисковой системы предприятия 201, 206
 - описание 201
 - портлет регистрации 201, 207

Ч

- частые запросы, слежение 234

Э

- элементы XML
 - отображение на общие структуры анализа 113
 - отображение на поля поиска 107
 - поиск 107, 113

Я

- японский
 - анализ символов новой строки 118
 - сегментация энграмм 118



Напечатано в Дании



GC43-0404-02



Spine information:



WebSphere II OmniFind Edition

Управление поисковой системой предприятия

Версия 8.3