

Лабораторная работа

Тема: *Построение поискового интерфейса*

Цель работы: *Получить представление об основных подходах при разработке поисковых функций в веб — приложениях ColdFusion*

Краткие теоретические сведения

Для задач поиска в ColdFusion встроена поисковая машины Verity Developers Kit (VDK), поддерживающая собственный поисковый язык Verity. Поиск может проводиться в упрощённом и стандартном режиме. В упрощенном режиме поиск проводится по слову или нескольким словам, а в стандартном режиме задействованы операторы и модификаторы.

Создание поискового интерфейса в приложениях ColdFusion начинается с определения специальной базы данных, хранящей метаданные для файла, группы документов, объекта запроса, называемой коллекцией Verity. Метаданные проходят предварительную оптимизацию для более быстрого поиска. После чего поиск проводится не по документам или объектам запроса, а по коллекциям упорядоченных метаданных.

Общие принципы организации поискового интерфейса в приложении следующие:

1. Создание коллекции Verity
2. Заполнение коллекции Verity метаданными и их индексация
3. Непосредственное осуществление поиска

Создание коллекции Verity

Создать коллекцию можно средствами Администратора ColdFusion или на программном уровне с помощью тега CFCOLLECTION.

Создание коллекции на программном уровне с помощью тега CFCOLLECTION дает возможность приложениям регистрировать свои коллекции, не прибегая к Администратору ColdFusion. Упрощенный синтаксис тега CFCOLLECTION имеет следующий вид:

```
<CFCOLLECTION ACTION          ="create"  
                  CLLECTION    = "имя_коллекции"  
                  PATH         = "файловый_путь_к_коллекции">
```

Атрибут ACTION – указывает тип операции (в данном случае создание коллекции).

Атрибут COLLECTION – содержит имя создаваемой коллекции.

Атрибут PATH — должен содержать физический путь к коллекции, доступный для сервера приложений ColdFusion.

Тег CFCOLLECTION позволяет регистрировать уже существующие коллекции Verity, для этого атрибуту ACTION присваивается значение map. Для отображения существующей коллекции необходимо, чтобы сервер ColdFusion имел доступ к ней через локальную файловую систему.

На рис.1 проиллюстрировано, как можно создать коллекцию с помощью Администратора ColdFusion: DATA & SERVICES – VERITY COLLECTION. В соответствующие поля необходимо ввести имя коллекции и прописать путь к коллекции.

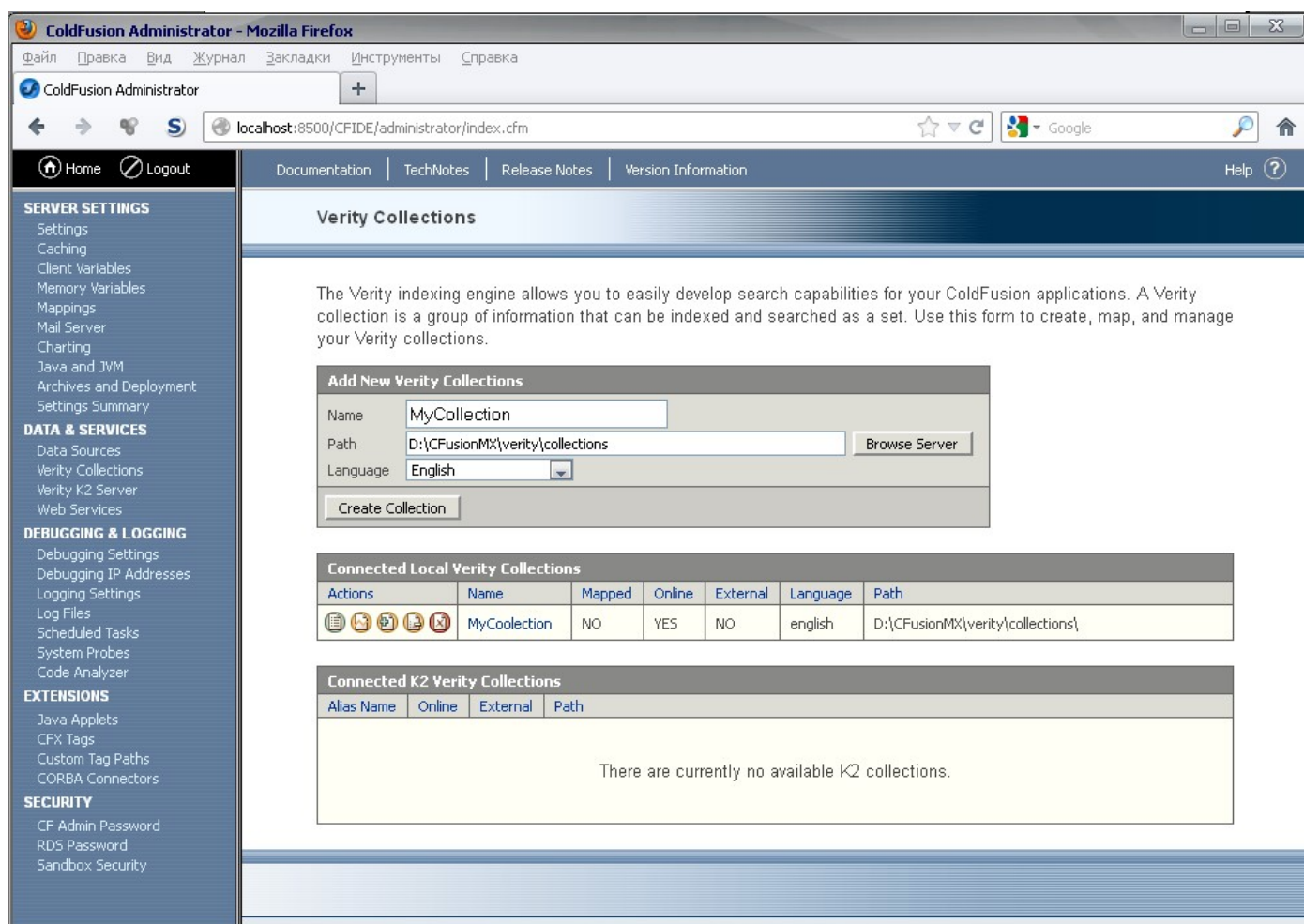


Рис.1

Заполнение и индексация коллекции

После создания коллекции Verity ее необходимо заполнить индексированными данными. Заполнение и индексация происходит одновременно или при помощи средств Администратора ColdFusion, или тегом CFINDEX.

С помощью этого тега можно индексировать документы редакторов Adobe PDF, Microsoft Word, Microsoft Excel, HTML, CFML, XML и др.

Заполнение и индексация коллекции может проводиться по данным файлов директории, отдельному файлу или объекту запроса ColdFusion. Упрощенный синтаксис тега CFINDEX выглядит следующим образом:

<CFINDEX	ACTION	= "update"
	COLLECTION	= "имя_коллекции"
	TYPE	= "тип_индекса"
	TITLE	= "имя_поля"
	KEY	= "уникальный_идентификатор"
	BODY	= "список_имен_полей"
	URLPATH	= "url"
	EXTENSIONS	= "список_расширений_файлов"
	QUERY	= "имя_объекта_запроса">

Атрибут ACTION - определяет тип операции, выполняемой тегом CFINDEX.

Атрибут COLLECTION – содержит тип коллекции.

Атрибут TYPE — определяет тип индекса. Допустимые значения : CUSTOM — индексация объекта запроса; FILE — индексация файла; PATH — индексация файлов файловой директории, расширения которых указаны в атрибуте EXTENSIONS.

Атрибут TITLE — имя поля, содержащего названия индексируемых записей. Используется при индексации объекта запроса.

Атрибут KEY - определяет уникальный идентификатор для каждой записи в индексе. В качестве значений может выступать имя файла (при индексации по файлу), полный путь к директории (при индексации по файловой директории), уникальный идентификатор записей в реляционном наборе (первичный ключ при индексации объекта запроса).

Атрибут BODY — должен содержать разделенный запятыми список полей, которые содержат индексируемые данные. Используется при индексации объекта запроса ColdFusion.

Атрибут URLPATH — должен содержать часть url, которая будет присоединяться к имени файла. Можно использовать, когда TYPE="File" или TYPE="Path". Например: URLPATH = "http: //127 . 0. 0. 1:8500/my/". В этом случае при использовании тега <CFSEARCH> полное значение выводимой колонки URL будет принимать значение: "http: //127 . 0 . 0 . 1:8500/my/" + "имя файла". Если данный атрибут не указывается, при использовании все того же тега <CFSEARCH> переменная URL будет принимать значение "имя файла"

Атрибут **EXTENSIONS** — содержит разделенный запятыми список расширений файлов, участвующих в индексации по файловой директории.

Атрибут **QUERY** — содержит имя объекта запроса, содержащего индексируемые данные.

На рис.2 проиллюстрирован интерфейс ввода настроек для заполнения и индексации коллекции. В соответствующих полях необходимо указать расширения индексируемых файлов директории, путь к индексируемой директории, часть URL, которая будет добавляться к имени файла при формировании URL к выбранному ресурсу, содержащему критерий поиска. Так же можно задать индексацию файлов в поддиректориях, указанной директории.

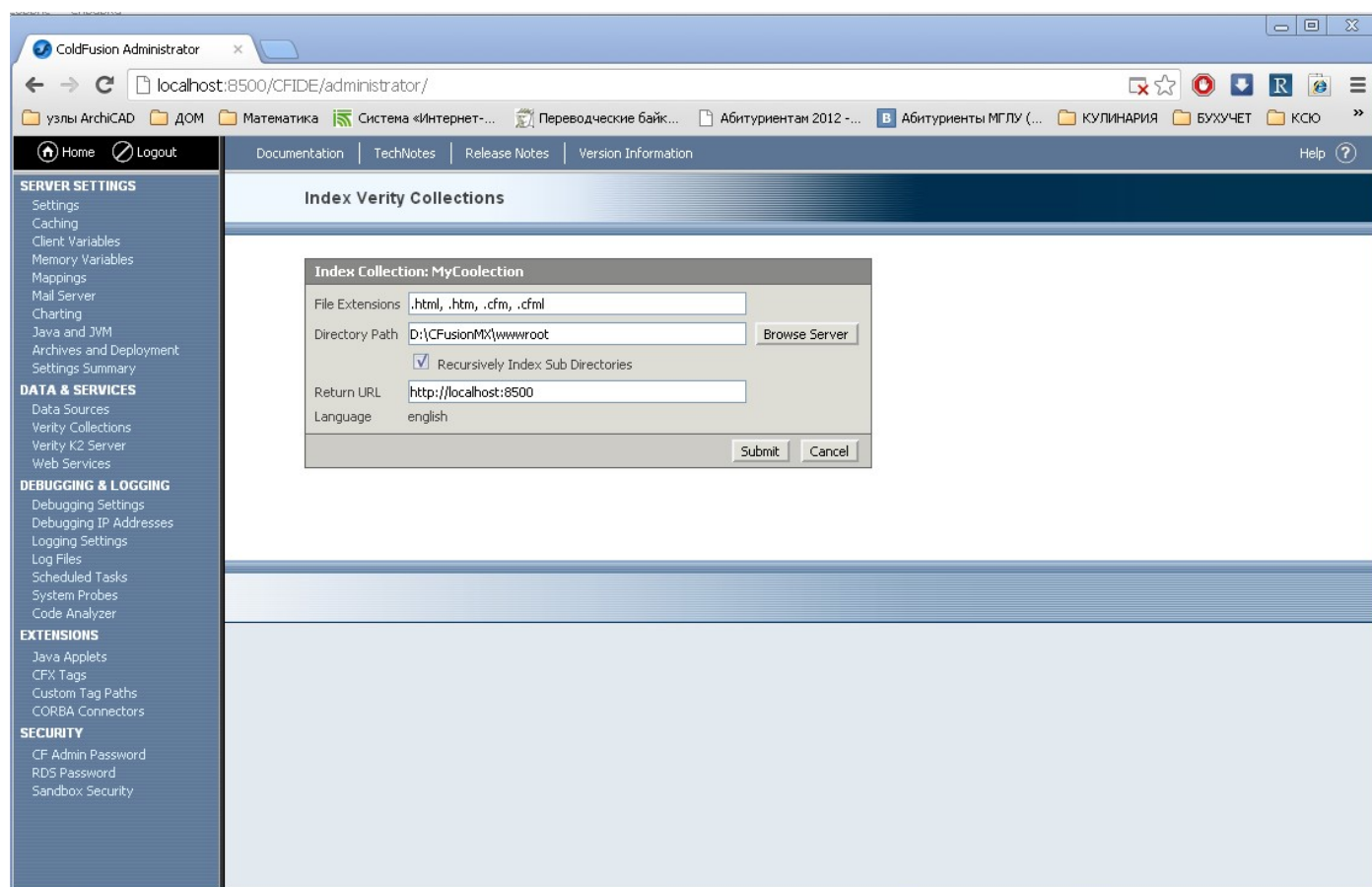


Рис.2

Поиск в коллекциях Verity

После создания коллекции и заполнения ее данными можно выполнять поиск по метаданным коллекции с помощью тега **CFSEARCH**. Упрощенный синтаксис тега имеет вид:

<CFSEARCH	NAME	= "имя_объекта_запроса"
	CLLECTION	= "имя_коллекции"

CRITERIA = "критерий_поиска">

Атрибут NAME — определяет имя объекта поискового запроса.

Атрибут COLLECTION — содержит имя коллекции Verity, по которой будет осуществлен запрос.

Атрибут CRITERIA — содержит ключевое слова или ключевую фразу, в соответствии с которой будет производиться поиск.

Объект поискового запроса будет содержать в себе не только результирующий набор поиска, но и стандартные переменные. Некоторые из них:

Имя_запроса.ColumnList — список имен полей итогового набора.

Имя_запроса.CurrentRow — текущая запись запроса, обрабатываемая тегом CFOUTPUT.

Имя_запроса.RecordCount — общее количество записей в коллекции Verity, в которой производится поиск.

Имя_запроса.Score — степень соответствия критерию поиска.

Имя_запроса.Title — имя записи поискового запроса.

Имя_запроса.URL — гиперссылка, определяемая сочетанием указанного атрибута URLPATH тега <CFINDEX> и именем файла, который содержит искомый фрагмент.

Ход работы

1) В администраторе ColdFusion зарегистрировать коллекцию метаданных, проиндексировав содержимое исходного каталога документов веб-сервера Cold Fusion.

2) Написать интерфейс поиска по файлам созданной коллекции.

3) Создать поисковый интерфейс по содержанию статей, собранных в [таблице News базы данных ProgrammingCF](#) :

а) В администраторе ColdFusion зарегистрировать коллекцию, проиндексировав содержимое соответствующих данных базы данных ProgrammingCF.

б) Написать интерфейс поиска по данным созданной коллекции.

4) Написать интерфейс "Верхняя десятка", создав и заполнив предварительно базу данных критериями поиска.

Контрольные вопросы

Что такое коллекция Verity?

Опишите основные этапы разработки поиска в технологии ColdFusion.

Каким тегом осуществляется создание коллекции Verity ?

Каким тегом осуществляется заполнение и индексация коллекции Verity ?

Каким тегом осуществляется поиск по коллекции Verity в ColdFusion?

Назовите все возможные типы индексации коллекций Verity.

Литература

Брук-Билсон Р. Программирование на Cold Fusion. Для профессионалов. -
СПБ: Питер, 2002