

ЭЛЕКТРОННЫЕ РЕСУРСЫ В ВУЗАХ: проблемы единого поиска

С каждым годом всё большую долю в комплектовании фондов университетских библиотек приобретают электронные издания. Одним из основных источников электронных изданий являются электронно-библиотечные системы (ЭБС). Однако в процессе их использования становится очевидным, что ни одна ЭБС по своему содержанию не удовлетворяет всем потребностям вузов. Поэтому учебному заведению необходимо подписываться на несколько ЭБС одновременно. Кроме того, нередко вуз создаёт свою собственную электронную библиотеку.

В данной статье рассматривается технология интеграции электронно-библиотечных систем, в том числе электронной библиотеки вуза, с целью создания единой информационно-поисковой системы по фондам библиотеки университета.

Во всех ЭБС поиск устроен по-разному, и возможности его различны. Количество поисковых полей — от двух в ЭБС «Лань» до девяти в IQlib, количество используемых булевых операторов — от одного «И» в «БиблиоТех» до трёх «И, ИЛИ, НЕТ» в ЭБС «Book.ru». Качество морфологического анализа тоже различается в зависимости от используемой системы полнотекстового поиска. В связи с этим для удобства читателей возникает необходимость в разработке единой информационно-поисковой системы как по фонду электронных изданий, имеющихся в университете, так и по фонду электронных изданий, представленных в ЭБС. Наиболее удобным с точки зрения использования и администрирования было бы использовать единую систему ЭБС, в которой централизованно осуществлялись бы функции поиска ресурсов коллекции, контроля доступа, сбора статистики и т.п.

Одним из способов решения обозначенной проблемы может быть использование технологии «Яндекс.Сервер» [1]. Для этого необходимо решить несколько задач:

- выбрать источник данных для индексирования и анализатор содержимого документа;
- определить структуру коллекций «Яндекс.Сервера»;
- разработать схему информационных потоков процесса поиска в электронном каталоге и разработать программные модули для обеспечения его работы.

При решении поставленных задач необходимо учитывать следующие особенности ор-

ганизации поиска в электронном каталоге библиотеки университета:

- возможность ежегодного изменения состава ЭБС;
- обеспечение поиска по заранее определённому перечню ЭБС для различных подразделений, филиалов вуза.

В стандартную поставку «Яндекс.Сервера» входят четыре возможных источника данных:

- 1) `ftds` — для индексирования файловых директорий;
- 2) `webds` — для индексирования веб-страниц;
- 3) `odbcds` — для индексирования данных через интерфейс ODBC;
- 4) `mysqlds` — для индексирования баз данных MySQL.

В качестве источника данных был выбран стандартный источник `webds`. Выбор обусловлен простотой организации документов для представления индексатору и наличием отлаженного механизма для работы с коллекцией университета на основе веб-технологий. В данном случае механизм индексирования работает по принципу «сетевого паука» (робота) [2].

Анализ содержимого документа реализован с помощью `xml`-парсера [3]. Все атрибуты `xml`-файлов являются текстовыми. Анализ содержимого производится по следующим полям: название; авторы; тематика; год издания; специальность; кафедра; дисциплина; ключевые слова.

Обязательным полем является поле «название».

Автор



**Пётр Алексеевич
БОЛДЫРЕВ**

заместитель директора по автоматизации библиотечно-библиографических процессов научной библиотеки Оренбургского государственного университета, кандидат технических наук

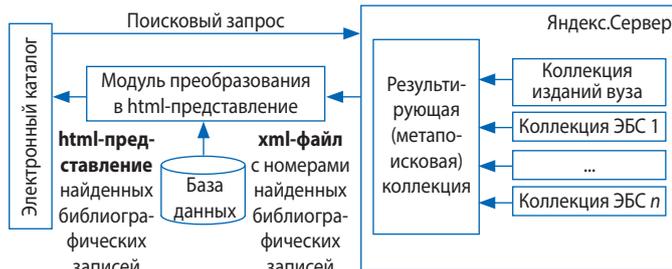


Рис. 1. Схема информационных потоков процесса поиска в электронном каталоге

Каждое из полей имеет относительный вес в тексте внутри xml-элемента, что позволяет указать наиболее важные для поиска поля. Поле «название» имеет самый высокий относительный вес, поле «ключевые слова» – самый низкий.

При разработке структуры коллекций важную роль сыграла необходимость ежегодного изменения состава подключённых ЭБС, а также необходимость обеспечения поиска по заранее определённому перечню ЭБС для различных подразделений, филиалов вуза. По этой причине для каждой ЭБС была создана отдельная коллекция. Результирующая (метапоисковая) коллекция, по которой производится поиск в электронном каталоге, формируется на основе коллекции вуза и проиндексированных коллекций ЭБС. Такой подход позволяет не производить переиндексацию всех коллекций, а также гибко настроить поиск по определённым коллекциям для различных подразделений, филиалов вуза.

Для организации работы «Яндекс.Сервера» необходимо настроить механизм индексирования и поиска.

Схема информационных потоков процесса поиска в электронном каталоге научной библиотеки представлена на рис. 1.

Поисковый запрос, поступивший из электронного каталога, обрабатывается «Яндекс.Сервером». Поиск осуществляется по всем коллекциям, указанным в метапоисковой коллекции. «Яндекс.Сервер» возвращает результат поиска в виде xml-файла, в котором имеются номера библиографических записей. Индексирование начинается со стартовой страницы, на которой расположены ссылки на xml-файлы с описанием изданий (рис. 2). Стартовая страница формируется на основе записей, полученных из базы данных, при этом в индексировании участвуют только вновь добавленные записи и записи, в которые были внесены изменения. Это позволяет снизить нагрузку на «Яндекс.Сервер». На последнем этапе индексирования

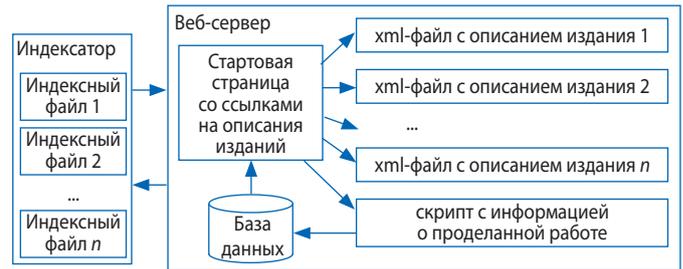


Рис. 2. Схема индексирования xml-файлов с описанием изданий

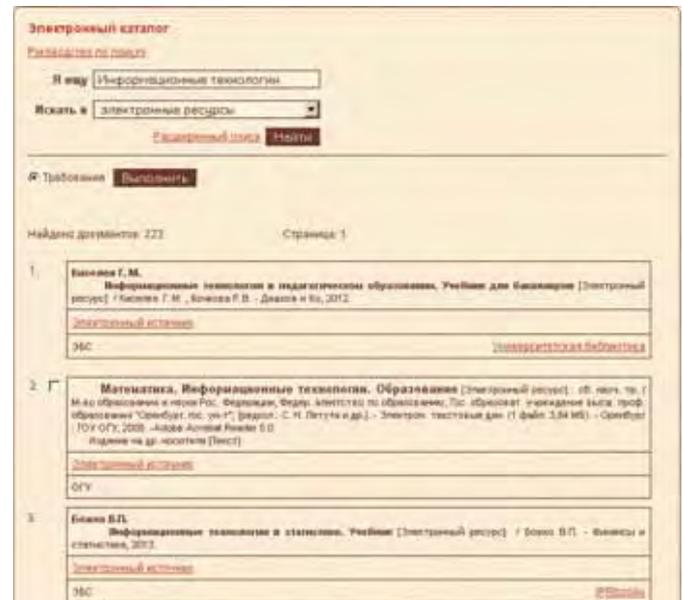


Рис. 3. Электронный каталог

ирования индекатор посещает пустую страницу, на которой выполняется скрипт, записывающий информацию о проделанной работе в базу данных. Полученная таким образом информация позволяет сформировать стартовую страницу при последующем индексировании.

Для преобразования xml-файла и получения необходимой информации из базы данных по номерам библиографических записей используется модуль преобразования в html-представление. Результат работы этого модуля отражается в электронном каталоге (рис. 3).

При организации работы механизма индексирования необходимо учитывать проиндексированные ранее записи, для того чтобы снизить нагрузку на «Яндекс.Сервер». С этой целью после окончания индексирования в базу данных записывается информация о дате последнего индексирования и количестве проиндексированных документов.

Таким образом, на основе технологии «Яндекс.Сервер» нам удалось организовать поиск в электронном каталоге научной библиотеки как по фонду электронных изданий, имеющихся в университете, так и по фонду электронных изданий, представленных в ЭБС, с возможностью изменения состава подключённых ЭБС и возможностью поиска по заранее определённому перечню ЭБС для различных подразделений и филиалов вуза. ■

ЛИТЕРАТУРА

1. Болдырев П.А. Применение программного продукта «Яндекс. Сервер» для организации поиска в электронном каталоге библиотеки / П.А. Болдырев // Библиотечковедение. – 2012. – № 3. – С. 47–51.
2. <http://api.yandex.ru/server/>
3. <http://help.yandex.ru/webmaster/?id=995324>
4. <http://api.yandex.ru/server/doc/concepts/y-ds-parsers.xml>