

Применение программного продукта «Яндекс. Сервер» для организации поиска в электронном каталоге библиотеки

Накопленные библиотеками за последние годы огромные массивы информации ставят перед разработчиками задачу организации быстрого и качественного поиска, решение которой возможно с использованием современных поисковых инструментов веб-технологии. Автор рассматривает один из таких инструментов — программный продукт «Яндекс. Сервер», позволяющий организовать оптимальный поиск в электронном каталоге библиотеки с учетом морфологии русского и английского языков, а также различных логических условий.

Ключевые слова: *электронный каталог библиотеки, поисковый инструмент, программный продукт, технологии Яндекса.*



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

Введение

Ядром и основой функционирования информационно-библиотечных систем, а также одним из основных объектов приложения новых информационных технологий в библиотеке является электронный каталог (ЭК). Эффективность использования информационных ресурсов библиотек и качество библиотечного сервиса как на локальном уровне, так и в режиме удаленного доступа во многом определяют лингвистические средства тематического поиска в ЭК, обеспечивающие интеллектуальный доступ к документным массивам. Механизм эффективного поиска в ЭК открывает быстрый и качественный доступ к информационным ресурсам библиотеки. Согласованный и обоснованный выбор лингвистического обеспечения — задача, без решения которой

создание локальных и распределенных электронных каталогов теряет смысл.

Использование в ЭК широкого набора разнообразных по структуре, лексическому наполнению, свойствам лингвистических средств представляет собой устойчивую тенденцию. Но при этом разные библиотеки демонстрируют множество подходов к определению и обоснованию состава и принципов функционирования лингвистического обеспечения. Это вынуждает библиотечное сообщество на новом уровне обратиться к задаче оптимизации комплекса информационно-поисковых языков, т. е. определения состава лингвистического обеспечения и параметров каждого из поисковых языков, необходимых и достаточных для эффективного функционирования системы в целом. Задача оптимизации лингвистического обеспечения ЭК может быть решена на базе поисковых инструментов Интернета.

Быстрый рост и динамичное развитие информационно-коммуникационных технологий (прежде всего веб-технологий) дает возможность разработки совершенного нового подхода к поиску информационных ресурсов библиотек. Чтобы найти нужную информацию в электронных ресурсах библиотек, необходимо найти ее адрес. Для этого существуют специализированные поисковые серверы (роботы индексов (поисковые системы), тематические интернет-каталоги, системы мета-поиска, и т. д.). Веб-технология World Wide Web (WWW) считается специальной технологией подготовки и размещения документов в Интернете. В состав WWW входят и веб-страницы, электронные библиотеки, каталоги и т. д. Все увеличивающиеся информационные ресурсы библиотек ставят перед разработчиками задачу организации быстрого, эффективного и качественного поиска в их масштабном информационном пространстве. В решении данной проблемы на помощь приходят поисковые инструменты.

Использование программного продукта «Яндекс. Сервер» для организации поиска в электронном каталоге библиотек

Поисковые инструменты — это особое программное обеспечение, основная цель которого — организовать наиболее оптимальный и качественный поиск информации для пользователей Интернета. Одним из таких инструментов является «Яндекс. Сервер» — программный продукт для полнотекстового поиска информации на веб-сервере с учетом морфологии русского языка.

«Яндекс. Сервер» состоит из поискового модуля и индексатора.

Поисковой модуль — это подсистема, осуществляющая разбор запроса, полученного от пользователя, поиск документов в индексных файлах, подготовленных ранее индексатором, и формирующая страницы отчета с результатами поиска. Каждой коллекции документов соответствует своя собственная поисковая машина.

Индексатор — это подсистема, анализирующая документы отдельной коллекции и сохраняющая информацию о них в специальных индексных файлах. Каждой коллекции документов соответствует свой набор индексных файлов, который хранится в отдельной директории, и свой собственный индексатор.

Схема поиска информации проста. Пользователь набирает ключевую фразу и активизирует поиск, в результате чего получает подборку документов по сформулированному запросу. Этот список документов ранжируется по определенным критериям так, чтобы вверху списка оказались те документы, которые наиболее соответствуют запросу пользователя (рис. 1).

Одна установка «Яндекс. Сервера» позволяет индексировать и находить нужную информацию сразу на нескольких веб-серверах, причем не обязательно расположенных физически на той же машине, что и «Ян-

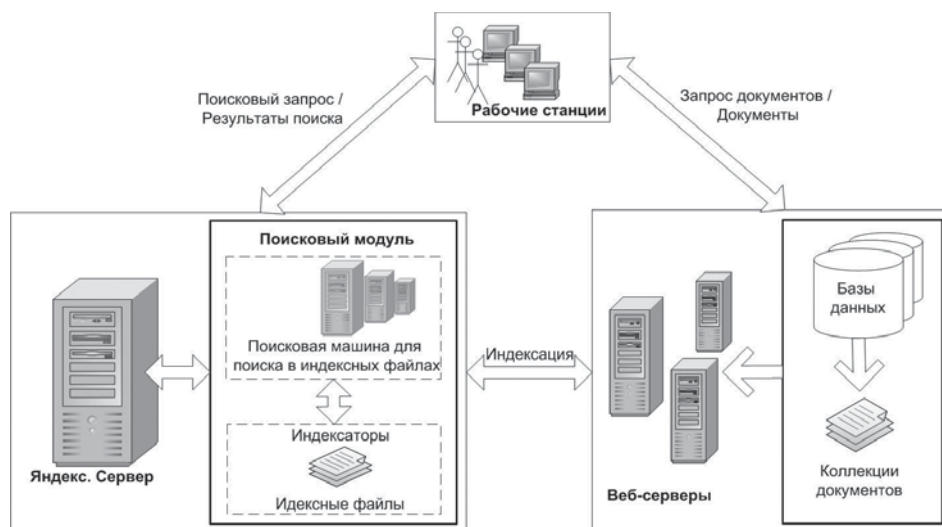


Рис. 1. Архитектура поисковой системы

декс. Сервер». Таким образом, различия в структуре баз данных, в используемых автоматизированных библиотечно-информационных системах и способах организации веб-серверов библиотек не мешает объединению их электронных каталогов.

«Яндекс. Сервер» работает с одной или несколькими коллекциями документов, в которых осуществляется поиск. Поиск в каждой коллекции документов осуществляется независимо. Применяя «Яндекс. Сервер» для организации поиска в ЭЖ такими коллекциями документов выступают базы библиографических записей библиотек, т. е. индексирование и поиск производятся по электронным каталогам библиотек не зависимо друг от друга [3].

Каждая библиографическая запись из образующих коллекцию документов в электронном каталоге содержит множество полей данных. Число полей может меняться в зависимости от типа материала, на который создается библиографическая запись. Для удобства поиска, а также из соображений совместимости электронных каталогов принято использовать стандартный набор атрибутов. В него входит несколько десятков атрибутов, но на практике используется только часть, такие как: автор, заглавие, предметная рубрика, год публикации, место публикации, издательство, ключевые слова, вид издания.

На основе атрибутов задается поисковый запрос. Каждый атрибут соответствует одному или нескольким полям библиографической записи. Такой подход является удобным, так как не требует указывать в поисковом запросе поля библиографической записи, по которым необходимо осуществлять поиск.

Результатом поиска является список библиографических записей, упорядоченных по релевантности. Релевантность учитывает не только количество найденных документов, но и контрастность слов (частота их употребления) и расстояние между словами.

Гибкий механизм настройки «Яндекс. Сервера» позволяет задавать режимы индексирования без остановки поисковой машины. Учитывая большие объемы коллекций библиографических записей, такая возможность является существенной, поскольку производится доиндексация: новые документы добавляются, измененные переиндексируются, а неизменные остаются в текущем состоянии, что экономит значительную часть ресурсов, в том числе временных. При этом нет

ограничений на размер и количество индексированных документов.

Возможности поисковой системы «Яндекс. Сервер»

Язык запросов «Ядекса. Сервера» представляет собой контекстно-свободную грамматику, терминалами которой служат слова естественного языка.

Поиск работает с учетом морфологии русского, английского языков — так же, как работает поиск по Интернету. Например, если задан запрос «идти», то в результате поиска будут найдены ссылки на документы, содержащие слова «идти», «идет», «шел», «шла» и т. д.

При выборе форм слова для поиска можно учитывать части речи. Например, если задан запрос «компьютер» (существительное), то будут найдены библиографические записи со словами «компьютера» или «компьютерами», но не «компьютерный» (прилагательное).

При необходимости для поиска точной формы слова морфологию можно «отключить». Для этого используются специальные символы — кавычки и восклицательный знак. Например, по запросу «!основы [анализа]!электрических» найдутся записи со словами: «основы», «анализа», «электрических», по запросу «!Соловьев» будут найдены записи автора Соловьев, а не, например, Соловьева.

Чтобы исключить библиографические записи, в которых встречается определенное слово, также применяются специальные символы — тильды (~) или один минус (без пробела). Например, при поиске библиографических записей на учебники по английскому языку, автором которых не является Бонк, необходимо задать запрос: «английский учебник~~Бонк».

«Яндекс. Сервер» дает возможность осуществлять поисковые запросы с различными логи-

ческими условиями, например, запросы с логическим условием «или», с исключением слов, запросы с указанием расстояния между словами в предложении и т. п. Так, если точно известно не только расстояние, но и порядок слов запроса, в котором они должны идти в тексте интересующих библиографических записей, необходимо указать между символом «/» (слеш) и числом-расстоянием символ + (плюс) для прямого порядка слов или — (минус) для обратного. Например, если необходимо найти основы математики вычислительной, дискретной и т. д., то задается запрос «основы /+2 математика». Будут найдены записи, где между словами основы и математика есть ровно одно слово посередине.

Индексирование и поиск по различным библиографическим атрибутам позволяет в полной мере раскрывать возможности расширенного поиска. При заполнении нескольких полей учитываются все заданные условия одновременно, при этом для каждого поискового поля действуют такие же правила языка запросов, что и для формы классического поиска. Например, если нужно найти все источники со словом «история» авторов Соловьева и Радугина, не содержащие в заглавии фразу «с древнейших времен», то в поле «Автор» задается «Радугин|Соловьев», в поле «Заглавие» — «История~~древнейший». Расширенный поиск позволяет накладывать дополнительные фильтры на процесс поиска, тем самым сужая и конкретизируя его область [1, 2].

Результатом поиска является список библиографических записей, упорядоченных по релевантности. Кроме информации об авторе и названии библиографические записи включают в себя: предметные рубрики, ключевые слова; краткую аннотацию документа, если она присутствует в записи; дополнительные точки доступа; данные о местонахождении документа.

Эффективный поиск в электронном каталоге позволяет организовать оперативный и качественный доступ к информационным ресурсам библиотеки.

СВОДНЫЙ ЭЛЕКТРОННЫЙ КАТАЛОГ
ВУЗОВСКИХ БИБЛИОТЕК ОРЕНБУРГА И ОРЕНБУРГСКОЙ ОБЛАСТИ

Поиск О проекте Справка Стать участником

Классический поиск Расширенный поиск Руководство пользователя

В корзине: 0
Корзина заказа

Скрыть форму поиска

Я ищу

Тип издания

Искать в библиотеках всех участников

Поиск осуществлен с использованием [Яндекс.Сервера](#) Общее количество записей в каталоге 1 517 700
Общее количество участников в каталоге 14

Оренбургский государственный университет
(общее количество записей: 652 749, дата последнего обновления: 08.11.2011, найдено: 121)

Государственное учреждение культуры «Оренбургская областная универсальная научная библиотека им. Н.К. Крупской»
(общее количество записей: 586 224, дата последнего обновления: 14.10.2011, найдено: 75)

1 2 3 4 ... следующая →

1. + 517 Далецкий, Ю. Л.
Стохастические уравнения и дифференциальная геометрия / Ю.П. Далецкий, Я.И. Белополюская. - Киев : Выш. шк., 1989. - 293,[2] с. с.
ОУНБ КХ-1

Рис. 2. Сводный электронный каталог вузовских библиотек Оренбурга и Оренбургской области

Проект, созданный на основе поисковой системы «Яндекс. Сервер»

Использование поисковой системы «Яндекс. Сервер» позволило в 2010 г. согласно решению Совета ректоров вузов Оренбургской области от 29 апреля 2009 г. организовать «Сводный электронный каталог библиотек Оренбурга и Оренбургской области» (СвЭК) в виде распределенных баз данных (рис. 2). Целью создания СвЭК является использование потенциала библиотек в интересах развития науки, культуры и образования, в поддержке отечественного информационного пространства. СвЭК обеспечивает доступ населению и сотрудникам библиотек Оренбуржья, России и зарубежных стран к библиографическим записям изданий, хранящихся в библиотеках города и области, дает возможность поиска библиографических записей интересующих изданий, а также информацию об их местонахождении.

Работа в рамках данного проекта продолжается, но уже сегодня возможен доступ к СвЭК по адресу — <http://svek56.ru/>. На 1 ноября 2011 г. в СвЭК представлен книжный фонд 14 библиотек общим объемом более 1,5 млн библиографических записей. Для удобства пользователей в СвЭК реализована «Корзина» — функция отбора интересующих изда-

ний из списка результатов поиска для дальнейшей распечатки требований. Вся необходимая информация для библиотек, желающих стать участниками СвЭК, размещена в разделе «Стать участником», содержащем соглашение о сотрудничестве и требования к техническому оснащению библиотеки. Для специалистов одной из первоочередных стоит задача реализация функций сбора статистических показателей работы СвЭК с целью анализа и определения дальнейших действий по его развитию. Таким образом, создание СвЭК позволяет библиотекам региона выйти на новый уровень профессионального взаимодействия, способствует созданию нового регионального информационного ресурса.

Список источников

1. «Яндекс. Сервер» — порядок на вашем веб-сервере [Электронный ресурс]. — Режим доступа: <http://company.yandex.ru/technology/server/>
2. «Яндекс. Сервер» 2010. Руководство по установке и эксплуатации [Электронный ресурс]. — Режим доступа: <http://api.yandex.ru/server/doc/concepts/About.xml>
3. Шевченко В. Информационный сервис : быстро, эффективно, надежно : об изменении традиционной схемы организации обслуживания // Библиотека. — 2010. — № 8. — С. 51—53.

Анонс



Головко С.И. Креатив-технологии в библиотечной сфере : монография. — М. : Пашков дом, 2012. — 168 с.

В книге С.И. Головки предложены и подробно рассмотрены оптимальные варианты использования креативных технологий в образовании, развитии и поддержке библиотечных специалистов нового поколения в условиях инновационного развития общества. Большое внимание уделено интерактивным формам обучения, формирования профессиональной компетентности, стимулирования познавательной активности и преобразовательной деятельности библиотечных лидеров, способных к принятию решений, максимально способствующих обновлению формальной и содержательной деятельности в библиотечной сфере.

119019 Москва, ул. Воздвиженка, 3/5,
Российская государственная библиотека,
Издательство «Пашков дом»
e-mail: pashkov_dom@rsl.ru
факс: (495) 695-59-53