

Новые возможности поисковых средств портала ВГБИЛ

New Capabilities of the Search Vehicles of the Web-Portal of the All-Russia State Library of Foreign Literature

Нові можливості пошукових засобів порталу ВДБІЛ

Колосов К. А.

*Всероссийская государственная библиотека иностранной литературы им. М. И. Рудомино,
Москва, Россия*

Kirill A. Kolosov

Rudomino All-Russia State Library of Foreign Literature, Moscow, Russia

Колосов К. А.

Всеросійська державна бібліотека іноземної літератури ім. М. І. Рудомино, Москва, Росія

Описываются принципы и подходы, используемые при разработке поискового модуля нового WEB-портала ВГБИЛ. В основе лежит применение протокола Z39.50 для поиска по библиографическим базам данных и полнотекстовым документам WEB-сайта. Обработка поисковых запросов пользователей осуществляется с использованием последовательного расширения числа информационных источников. Рассмотрены особенности и преимущества решений Z39.50, используемых в ИРБИС-64 и электронной библиотеке ВГБИЛ.

Principles and approaches that were used to design the search module of the library's new web-portal are described. The module is based on the Z39.50 protocol for the search in the web-site bibliographic databases and full-text documents. Users' search queries are processed by sequential extension of information resources. Specific features and advantages of Z39.50 that are used in IRBIS-64 and the e-library of the Library of Foreign Literature are discussed.

Описуються принципи та підходи, які використовуються при розробці пошукового модуля WEB-порталу ВДБІЛ. В основі лежить застосування протоколу Z39.50 для пошуку в бібліографічних базах даних та повнотекстових документах WEB-сайту.

Обробка пошукових запитів користувачів здійснюється з використанням послідовного розширення кількості інформаційних джерел. Розглянуто особливості та перевага рішень Z39.50, які використовуються в ІРБІС-64 та в електронній бібліотеці ВДБІЛ.

В 2005 году Всероссийская государственная библиотека иностранной литературы им М. И. Рудомино (ВГБИЛ) приступила к разработке своего нового WEB-портала. В настоящее время создан поисковый модуль, который будет предоставлять читателям библиотеки интегрированный поиск по всей совокупности доступных электронных ресурсов:

- электронному каталогу (ОРАС) ВГБИЛ;
- электронной библиотеке ВГБИЛ;
- информационным страницам и электронным изданиям WEB-сайта ВГБИЛ;
- тематическим базам данных, ведущимся во ВГБИЛ;
- электронным ресурсам библиотечных корпораций, в работе которых участвует ВГБИЛ;
- электронным публикациям и базам данных EBSCO.

При разработке WEB-портала применяются общие принципы Интернет/Инtranet технологии, обеспечивающие единообразный механизм доступа к portalу как из внутренней сети библиотеки, так и из Интернет, но с разграничением прав доступа для внешних и внутренних пользователей. Для пользователей Интернет полностью доступны электронный каталог библиотеки, страницы WEB-сайта и поиск по каталогам библиотечных корпораций, то есть электронные ресурсы, не содержащие полных текстов электронных документов по причинам ограничений, связанных с обеспечением авторских прав и лицензионных соглашений. В то же время, читатель в залах библиотеки получит доступ ко всей совокупности электронных ресурсов, включая статьи из баз данных EBSCO, электронной библиотеки ВГБИЛ и других доступных источников.

В основе технической реализации поискового модуля лежит использование решений на основе протокола Z39.50, разработанных во ВГБИЛ совместно с ГПНТБ России для проектов корпоративной сети московских библиотек и системы автоматизации библиотек ИРБИС-64, а также программных модулей, разработанных во ВГБИЛ в процессе работы в консорциуме РИБК.

Формирование поисковых запросов и представление результатов поиска осуществляется через программный шлюз HTTP-Z39.50. В отличие от предыдущей версии данные из всех источников представляются в кодировке UNICODE (UTF-8). Кроме того, была изменена логика обработки пользовательских запросов. Если в предыдущей версии шлюза пользователь должен был всегда указывать каталоги, участвующие в

поиске, то теперь в основном режиме поиск производится по всем подключенным каталогам одновременно. Для формирования первичного поискового запроса используется всего одно поле по аналогии с популярными поисковыми системами Интернет и порталом РИБК. При обработке первичного поискового запроса используются только базы данных, расположенные во ВГБИЛ. Пользователь получает список, содержащий количество найденных документов в каждой базе данных. После этого он может перейти к просмотру найденных записей, расширить число подключенных источников за счет серверов Z39.50 корпоративной сети московских библиотек и РИБК, а также базе данных EBSCO. Если необходимо сформировать сложный поисковый запрос, то можно использовать интерфейс расширенного поиска.

WEB-портал является основным средством доступа к базам данных электронного каталога ВГБИЛ, работа над новой версией которого ведётся отделом автоматизации. В отличие от системы TINLIB, используемой до настоящего времени в ВГБИЛ, новая версия АБИС создаётся на основе СУБД ORACLE, что обеспечивает широкие возможности для поддержки UNICODE и средств представления данных в коммуникативных форматах, включая RUSMARC.

Другим крупным источником информационных ресурсов ВГБИЛ является электронная библиотека, работа над которой ведётся отделами информационных технологий и электронных публикаций. В основе электронной библиотеки лежит использование СУБД MySQL для хранения библиографических описаний электронных ресурсов.

Для обеспечения универсального доступа ко всем информационным источникам отделом информационных технологий ВГБИЛ разрабатывается провайдер данных, обеспечивающий взаимодействие сервера Z39.50 и СУБД ORACLE и MySQL на основе использования языка SQL-запросов и модулей сервера Z64. Сервер Z64 был разработан отделом информационных технологий ВГБИЛ совместно с ГПНТБ России и в своей базовой версии содержит провайдер данных для системы автоматизации библиотек ИРБИС-64. На основе такой конфигурации во ВГБИЛ обеспечивается поддержка баз данных Z39.50 ВГБИЛ и некоторых баз данных в режиме хостинга: парламентской библиотеки (3 базы данных) и ГПНТБ (сводный каталог по научно-технической тематике).

Функциональными особенностями сервера Z64 являются:

- Встроенный провайдер данных ИРБИС-64;
- Неограниченное число подключаемых баз данных;
- Встроенный контроль прав доступа пользователей на основе анализа IP-адресов клиентов;
- Поддержка форматов представления данных ISO-2709 (RUSMARC, UNIMARC, MARC-21), SUTRS, XML;
- Возможность настройки форматов представления данных путем редактирования внешних форматных файлов, написанных на языке форматирования ISIS;
- Поддержка кодировки UTF-8;
- Поддержка набора поисковых атрибутов bib-1 (с возможностью их редактирования), а также соответствие требованиям профайла Bath (требования по совместимости возможностей поиска).

Одной из задач WEB-портала является поддержка полнотекстового поиска по информационным страницам и электронным изданиям WEB-сайта ВГБИЛ. Существует несколько подходов к решению данной задачи: использование индексации с помощью локальной версии системы Яндекс, индексация содержимого сайта, скопированного в базу данных (например, полнотекстовый поиск в ИРБИС) и др. С целью обеспечения совместимости технических решений WEB-портала был выбран вариант использования для полнотекстового поиска отдельного сервера Z39.50 на основе пакета ISITE. Несмотря на то, что данный пакет был разработан очень давно, он позволяет производить индексацию больших массивов полнотекстовых документов без необходимости их перекачки в базу данных и обеспечивать поиск средствами запросов протокола Z39.50. Проведенная доработка пакета ISITE позволила устранить ряд проблем, связанных с кодировками, отличными от Latin-1, а средства WEB-портала позволяют при выводе найденных документов выделять цветом искомые слова, аналогично интерфейсам поисковых порталов Интернет.

Таким образом, новый поисковый портал ВГБИЛ будет предоставлять пользователям современный и удобный доступ к множеству электронных источников, включая полнотекстовые электронные документы. По мере изменения потребностей пользователей список информационных источников портала легко редактировать, так как портал взаимодействует со всеми источниками информации по протоколу Z39.50. В то же время пользователи не догадываются об использовании этого протокола, так как их общение с системой происходит с помощью стандартного WEB-браузера. Организация поиска по различным информационным источникам, включая полнотекстовые базы данных, серверы Z39.50, открытые архивы OAI, вызывает интерес у многих разработчиков. Помимо дальнейшего развития программных протоколов всё большее значение приобретает оптимизация состава информационных источников, участвующих в обработке поисковых запросов, за счет динамического формирования так называемых целевых поисковых групп по регио-

нальным и тематическим признакам. В разрабатываемом WEB-портале ВГБИЛ используется начальный уровень подобной оптимизации, при котором последовательно расширяется область охвата информационных источников: от локального к региональному и к полнотекстовым источникам в Интернет (EBSCO). По мере расширения числа доступных полнотекстовых источников информации будет совершенствоваться и алгоритм оптимизации поиска на портале ВГБИЛ. Портал ориентирован на читателей библиотеки, для которых интерес представляет не просто библиографическая информация, а сами источники, то есть книги, доступные в данной библиотеке и библиотеках ближайшего региона, а также электронные документы и издания, доступные для прочтения на экране монитора.