

Методы и средства работы с данными в корпоративной электронной библиотеке ИИЦ СО РАН

Вильвер П.Ю., Бычков И.В.

Иркутский научный центр (ИИЦ СО РАН) приступил к работам по созданию WWW-доступа к ресурсам библиотек в конце 1998 г. За прошедшее время была разработана и внедрена первая очередь корпоративной электронной библиотеки (КЭБ), которая обеспечивает:

- ведение картотек и каталогов библиотек ИИЦ СО РАН;
- удаленный доступ к информационным ресурсам библиотек ЦНБ ИИЦ СО РАН;
- режим текущего оповещения на WWW-сервере библиотеки о новых поступлениях;
- удаленный доступ к зарубежным базам данных консорциумов Elsevir, Springer, Academic Press, агентства EBSCO, Баварской государственной библиотеки и др.

На текущий момент информационное обеспечение ученых ИИЦ СО РАН ведется по следующим БД :

- РЖ “Информатика” 1998-1999 гг.;
- РЖ “Охрана окружающей среды” 1995-1999 гг.;
- РЖ “Математика” 1998-1999 гг.;
- Картотека ЦНБ ИИЦ СО РАН;

Общий объем записей ~ 40000 записей.

Библиографические программы, применяемые в настоящее время библиотеками ИИЦ СО РАН, различаются форматами представления данных, не поддерживают или только частично поддерживают международные стандарты библиографических записей. Они ориентированы на использование персональных ЭВМ и локальных сетей с размещением баз данных (БД) на файловом сервере,

что затрудняет использование их в корпоративных сетях ИИЦ СО РАН и Иркутска. Кроме того, не все библиографические программы поддерживают технологию Intranet-сетей и доступа в режиме on-line . Многие из этих программ работают в MS DOS и, как правило, не очень удобны в работе.

Приобретение современных зарубежных программных продуктов требует значительных затрат валютных средств на их покупку, техническую поддержку (поставку исправленных и новых версий) и привязку (адаптацию) к условиям применения в каждой библиотеке.

При переходе на новые программы необходима подготовка требований и создание соответствующих конверторов для переноса БД. Аналогичные проблемы возникают и при переходе на использование отечественных программ. Поэтому методы и средства, используемые при создании КБС ИИЦ СО, отличаются от ставших уже стандартными программ, используемых при автоматизации библиотек:

- CDS/ISIS;
- WWWISIS;
- Разработанный в ГПНТБ России IRBIS;

При создании КБС, вследствие причин изложенных выше, мы пошли следующим путем:

1. Прежде всего, было решено использовать уже существующие в библиотеках программы и созданные базы данных, а также базы данных, которые были созданы в ГПНТБ СО РАН и других библиотеках. Так, например, ГПНТБ ведет базы данных заказанных документов с учетом сигла заказчика в свободно распространяемой Международной организацией UNESCO библиографической программе CDS/ISIS. CDS/ISIS позволяет управлять записями произвольной длины и эффективно хранить информацию и использует международный стандарт формата ISO-2709 для экспорта/импорта данных, что наиболее важно для нас при последующей загрузке библиографических записей в другие БД . Более того, форматы отвечают требованиям международных форматов, используются и поддерживаются многими организациями мира. В распоряжении ИИЦ СО РАН было несколько версий, переданных ему из ГПНТБ

©Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

СО РАН. Наиболее полная версия программы позволяет выводить записи в форматах семейства MARC по выбранным сиглам. Таким образом, при заполнении БД библиотек ИИЦ СО РАН требуется извлечь из БД ГПНТБ записи по определенным сиглам и загрузить их при помощи программы-конвертора. При использовании других источников формирования фондов библиотек ИИЦ СО РАН заполнение БД происходит отдельно в каждой библиотеке, но количество приобретенных таким образом книг невелико. Аналогичен способ использования тематических БД созданных в институтах ИИЦ.

2. Доступ к информации обеспечивается посредством Web-модуля корпоративной библиотечной системы. При создании Web-модуля системы учитывалась возможность использования различных программ просмотра (Internet Explorer, Netscape Navigator). Аналогично широко распространенной программе WWWISIS, для организации доступа к БД и создания страниц используется интерфейс CGI (Common Gateway Interface). Были созданы CGI-скрипты, обеспечивающие доступ к БД (возможность удаленной работы), динамический вывод информации по запросам. Выбор программных средств, для создания CGI-скриптов достаточно широк — Perl, Delphi, C++, другие языки программирования, (что является одной из причин выбора данной технологии.). Для СУБД с внутренними языками программирования существуют варианты использования этого языка для генерации документов.
3. Для обеспечения быстрого поиска в БД была нормализована структура исходных библиографических записей и, таким образом, в поисковых процедурах не используются пустые поля, оптимизирована структура запросов. В настоящее время система позволяет обеспечить поиск в различных режимах, в зависимости от степени сложности поиска.
4. Для обеспечения совместного использования информационных ресурсов, разрабатываемых различными подразделениями ИИЦ СО РАН, предложена концепция единого информационного хранилища (ИХ) с периодическим обновлением данных (библиографических, полнотекстовых и т. д.). Реализация ИХ позволяет на первом этапе создания КЭБ обеспечить необходимые требования по надежности, эффективности поиска, и т. д. Данный вариант WWW-доступа к базам данных свободен от недостатков схем с периодическим преобразованием содержимого БД в статические WWW-страницы или динамической генерацией WWW-страниц, таких как большое время обработки запросов, необходимость постоянного доступа к основной базе данных, дополнительная загрузка средств поддержки БД, связанная с обработкой запросов от WWW-сервера. Более того, после установления синхронизации данных информационного хранилища с основными БД возможен перенос пользовательских интерфейсов на информационное хранилище, что существенно повысит надежность и производительность, позволит организовать распределенные рабочие места. Для загрузки содержимо-

го основной БД в информационное хранилище могут использоваться все языки программирования, интегрированные средства, а также специализированные средства перегрузки, поставляемые с SQL-сервером, и продукты поддержки информационных хранилищ.

СРЕДСТВА РАБОТЫ С ДАННЫМИ.

Полнотекстовые, реферативные, библиографические БД в ведутся библиотеках ИИЦ СО РАН с использованием программ “ProCite”, “CDS/ISIS”, “Библиотека-4” и, как уже говорилось выше, поступают периодически из ГПНТБ СО РАН на CD. Если возможности “CDS/ISIS” позволяют вывести записи в любом из требуемых форматов, то “ProCite” хотя и имеет разные возможности вывода записей, но не поддерживает MARC-формат. Использование программы экспорта записей со структурой определяемой пользователем позволяет затем при помощи специальных программ перенести записи в любую СУБД. Чтобы загрузить записи в СУБД Oracle7.3 использовались SQL*Loader (Oracle7.3), Data Migration Wizard (Delphi3.0), специально создаваемые программы загрузки. Следует отметить хорошие возможности утилиты Oracle7.3 - SQL*Loader, позволяющей импортировать в БД со сложной структурой записи из файлов данных. Утилита SQL*Loader является гибким и настраиваемым средством, поддерживающим функции SQL, загрузку сегментированных объектов баз данных (отдельных разделов сегментированных таблиц, несегментированных таблиц). Используя NLS (National Language Support — поддержка национального языка) Oracle утилита SQL*Loader имеет возможность интерпретировать и преобразовывать данные со схемами кодировки символов других компьютерных платформ и стран.

При создании КЭБ ИИЦ СО РАН были разработаны (на Delphi v3.0) оригинальные программы загрузки данных, обеспечивающие возможность использования в качестве ИХ не только СУБД Oracle, но и других баз данных доступных через ODBC, а также загрузку в ИХ баз данных библиографических систем без преобразования их в текстовый формат.

Информационное хранилище реализовано на основе СУБД Oracle v7.3, расположенной на сервере ИИЦ СО РАН. При этом вопросы сохранности, обновления ресурсов, обработки транзакций, организации доступа к данным и т. д. решаются в основном с использованием развитых средств СУБД Oracle.

Поиск реализован при помощи CGI - скриптов, с использованием языка SQL, и обеспечивает выборку информации по полному или частичному соответствию запроса как для конкретного поля базы данных, так и по их совокупности и производится по полям: “Автор”, “Название”, “Дата издания”, “Ключевое слово”. В зависимости от количества выбранных для поиска баз данных и типа поиска (полное соответствие, не полное соответствие, усечение записи сначала, усечение записи с конца) выполняется CGI-скрипт. Поиск может проводиться в конкретном поле (например, поле-Автор или поле-Дата издания) или по всем полям.

Для ведения статистики посещений сервера данные о каждом посетителе автоматически заносятся в специаль-

ную базу данных (посредством CGI-скрипта, использующего переменные окружения), что в дальнейшем дает возможность идентифицировать любого посетителя сайта по его IP-адресу. Формирование подобной базы данных не только о посетителях сервера, но и о посещаемых ими страницах является в определенном смысле основой дальнейшего развития КЭБ. В частности, наличие собираемой в настоящее время информации о пользователях сайта позволит в последующем разработать процедуры автоматического создания списков рассылки по определенным разделам, публикуемым на сервере. Это, в свою очередь, может быть использовано как для информирования заинтересованных пользователей КБС о появлении новых научных работ, о проведении конференций, семинаров и других мероприятий по данной теме, так и, что представляется крайне важным, для организации телекоммуникационных контактов между ними (обсуждение работ в режиме on-line, проведение тематических chat-конференций и т.п.).

Стоит отметить, что аналогичный подход к организации WWW-доступа к ресурсам используется в Российской Государственной библиотеке с той разницей, что там используется другая БД (в РГБ — Paradox, в ИНЦ — Oracle).

Основным направлением развития КЭБ является переход на четырехуровневую модель взаимодействия: “браузер-WWW-сервер-сервер приложений-сервер БД” и усиление активной роли сервера БД.

Это вызвано следующими практическими задачами, возникающими при поиске и хранении информации:

- иметь удаленный доступ к информации с любого рабочего места,
- увеличить скорость поиска информации,
- создавать персональные БД на основе результатов выполнения запросов.

В качестве механизма создания пользовательских БД и дальнейшего их использования предлагаются хранимые процедуры и триггеры. Хранимые процедуры, кроме выполнения вышеперечисленных функций, позволяют перенести программное обеспечение в базу данных. Триггеры, прежде всего, обеспечивают защиту корпоративных данных. Другим преимуществом данного подхода является возможность хранения кода PL/SQL в интерпретированном виде. Также посредством хранимых процедур, при определенных условиях, можно совершать действия, выходящие за рамки обычных привилегий пользователя.

Другим направлением развития библиотечной системы ИНЦ СО РАН является внедрение протокола Z39.50. Наиболее реальным путем будет установка сервера ZooPARK, обеспечивающим работу с распределенными данными с шлюзом к БД Oracle.

Список литературы

- [1] Основы использования WWW - технологий для доступа к существующим базам данных <http://ftp.sibinfo.ru/win/materials/www-db/1.htm>
- [2] Даниэла Флореску, Алон Леви, Альберто Мендельсон. Технологии баз данных для World-Wide Web <http://...>, СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ #04-05/98
- [3] Мазов Н.А. Жижимов О.Л. Отображение электронного каталога ГПНТБ России в Z39.50. Материалы 7-й международной конференции “Крым 2000”, 2000, т.1, стр.221-224.
- [4] Российский коммуникативный формат представления библиографических записей в машиночитаемой форме (Рос. Вариант UNIMARC). СПб.: Изд-во РБН.-1998.