

ДОСТУП К БАЗАМ ДАННЫХ ISIS ИЗ INTERNET И ПОСТРОЕНИЕ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ*

О. Л. ЖИЖИМОВ, Н. А. МАЗОВ

*Объединенный институт геологии, геофизики и минералогии СО РАН
Новосибирск, Россия*

e-mail: zhizhim@uiggm.nsc.ru, mazov@uiggm.nsc.ru

А. С. ФРОЛОВ

Государственный научный центр "Вектор", Новосибирск, Россия

e-mail: fas@uiggm.nsc.ru

A model of the distributed informational system is suggested making access to the CDS/ISIS data bases possible via Internet. The model contains four logical functional levels allowing the implementation on different computer models and standard protocols of the data exchange. The reference to the attraction-mode version of the system is given.

1. Введение

Развитие коммуникационной инфраструктуры научных подразделений в Академии наук, в том числе в Сибирском отделении, создало предпосылки интенсивного применения в научных исследованиях технологий, ориентированных на сетевой доступ к различным информационным ресурсам, как внутренним (институтские), так и внешним (ресурсы других организаций, ресурсы Internet). При этом возможность доступа к ресурсам Internet не только существенно расширила возможности информационного обеспечения, но и привела к необходимости согласования внутренних ресурсов со стандартами, принятыми в мире. С другой стороны, из-за ограниченности средств сегодня практически все научно-исследовательские институты не имеют возможности самостоятельно приобретать всю необходимую научную литературу, банки и базы данных на машинных носителях. Даже крупные региональные библиотеки не могут обеспечить своих пользователей всей необходимой информацией. Особенно это ощутимо для научных центров, удаленных от Москвы. Выход из подобной ситуации — создание центров, кумулирующих информационные ресурсы по конкретным научным направлениям и предоставляющих пользователям в сетевом доступе.

Необходимо отметить, что в настоящее время во многих организациях различных стран ведутся работы по перенесению библиотечной информации на электронные носи-

*© О. Л. Жижимов, Н. А. Мазов, А. С. Фролов, 1997.

тели и обеспечению к ней массового доступа, в том числе с использованием компьютерных сетей различных масштабов [1, 2]. По причине лишь недавнего открытия услуг сети Internet в России в библиотечных сетевых технологиях основной упор до сих пор делался на локальные сети в среде ОС Novell NetWare, не обеспечивающие доступ к распределённой информации в глобальных масштабах. Как отмечают авторы работ [1–3], на настоящий момент разработка библиотечных сетевых технологий с использованием услуг сети Internet является одним из наиболее перспективных направлений этой области (см. также [4]).

В настоящей работе описывается один из возможных способов доступа к данным ISIS, распределенным по мини-серверам, через стандартный сервис сети Internet — WWW и построения на его основе распределенной информационной системы с централизованным управлением ресурсами. Описываемая система частично реализована на сети NT-серверов ОИГГиМ СО РАН и находится в настоящее время в режиме интенсивного тестирования (<http://проxy.uiggm.nsc.ru>).

Толчком для создания подобного программного комплекса в ОИГГиМ СО РАН явилась задача организации унифицированного сетевого доступа к библиографическим базам данных института, находящимся под управлением интегрированной информационно-поисковой системы CDS/ISIS/М версии 3.0 (далее — ISIS) [5–7], имеющейся также в библиотеках академических институтов СО РАН, в ГПНТБ СО РАН, с использованием возможностей сети Internet и предоставляемого этой сетью сервиса World Wide Web. При этом принимались во внимание следующие необходимые условия:

- обеспечение надежной коммуникационной дисциплины или дисциплины обмена данными на основе протокола TCP/IP;

- эффективная по времени и потребляемым системным и сетевым ресурсам реализация функции системы управления базами данных (СУБД) ISIS;

- наличие удобного русскоязычного пользовательского интерфейса, в том числе графического;

- переносимость (что подразумевает доступность исходного кода) либо наличие версий программного продукта для установки на различные аппаратные платформы в средах ОС UNIX, Windows NT, Windows 95;

- низкая стоимость.

Эти условия определяют требования к информационной системе, предназначенной для обеспечения доступа к библиографическим базам данных через Internet с использованием широко распространенных Web-браузеров типа Netscape или MS Internet Explorer:

- предоставление каждому пользователю графического интерфейса для организации поиска и просмотра информации из библиографических баз данных на основе стандартных протоколов HTTP [8, 9], HTML [11];

- централизованное хранение информации о всех доступных библиографических ресурсах с возможностью динамического обновления;

- доступ к базам данных в CDS/ISIS в своем исходном виде по модели <клиент-сервер> [10] с возможностью форматного вывода информации;

- корректную работу с кириллическими текстами и поисковыми полями;

- возможность управления доступом к информационной системе в целом и к конкретным базам данных на уровне IP адресов, сетей и конкретных пользователей;

- сбор статистической информации по работе с базами данных;

- возможность расширения для организации доступа к другим информационным ресурсам.

2. Общая схема информационной системы

Согласно опыту, полученному в ОИГГиМ СО РАН в результате разработки и эксплуатации собственной системы доступа к библиографическим базам данных [12], и требованиям, предъявляемым к подобной системе, распределенная информационная система представляется в виде трех-, четырехуровневой структуры с разграничением функций на каждом уровне и фиксацией протоколов межуровневого потока данных. Разграничение на логически замкнутые функциональные уровни необходимо для возможности их реализации на разных физических серверах и добавления в дальнейшем новых возможностей (рис. 1).

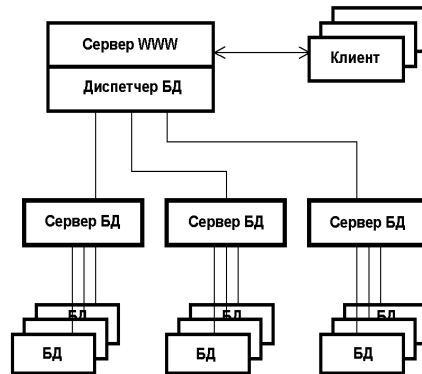


Рис. 1. Функциональная схема информационной системы (уровни 1–3).

Уровень 1. Собственно данные представляют собой обычные файлы данных в формате, необходимом для работы сервера БД. В случае CDS/ISIS данные хранятся в виде набора файлов в отдельном каталоге для каждой БД в кодировке, принятой в MS-DOS. Кроме собственно данных, каталог может включать информацию о predetermined форматах для отображения данных и файл заголовка для расширенного названия БД.

Уровень 2. *Сервер баз данных* реализует основные функции выборки информации из БД. Для публичной информационной системы эти функции сводятся к следующим:

- получение запроса с уровня 3;
- логический разбор строки запроса;
- исполнение запроса;
- возврат данных на уровень 3.

В соответствии с этим сервер БД обрабатывает следующие запросы.

Информационный — запрос на информацию о конкретной базе данных. Во входном потоке — идентификатор базы данных сервера БД, в выходном — заголовок, количество записей и комментарий указанной БД, описание поле БД.

Словарный — запрос на список ключевых слов с параметрами. Во входном потоке — идентификатор БД, шаблон ключевого слова, порядковый номер ключевого слова, количество слов в выходном буфере, в выходном — список затребованных ключевых слов и их частота.

Форматный — запрос на предоставление списка predetermined форматов вывода данных. Во входном потоке — идентификатор БД, в выходном — пронумерованный список predetermined форматов для данной БД.

Основной — запрос на предоставление данных в требуемом формате с параметрами. Во входном потоке — идентификатор БД, строка запроса в стиле ISIS, номер записи на-

чала вывода, количество записей для вывода, идентификатор формата, в выходном — форматированная выборка из БД.

Служебный — запрос на номер версии сервера БД. В выходном потоке — номер версии текущего сервера БД, пронумерованный список доступных БД, идентификатор внутренней кодировки сервера БД.

Сервер баз данных функционирует в многозадачной операционной системе в виде сервисной службы (Windows NT) или даيمона (UNIX). Обмен информацией с уровнем 1 происходит через файловую систему (локальную или сетевую), с уровнем 3 — по протоколам TCP через фиксированный программный порт. В последнем случае для лучшей межплатформенной совместимости данные передаются только в текстовом виде.

Уровень 3. Сервер WWW с модулем управления серверами БД — *диспетчер БД* — предназначен для обработки запросов пользователей, формирования запросов к серверам БД и возврата клиентам полученной информации по протоколу HTTP и спецификациям HTML. Оптимальным вариантом является Windows NT 4.0 + IIS 3.0 с поддержкой JAVA и ASP (Active Server Pages) ввиду тесной интеграции IIS с операционной системой и возможностью организации многопоточной обработки данных сравнительно простыми и дешевыми средствами. Управляющий модуль (диспетчер БД) может быть реализован в виде динамической библиотеки и (или) набора объектов ASP. В эксплуатируемом сегодня варианте (MS IIS 3.0) модуль представляет собой DLL, использующую интерфейс ISAPI. При этом диспетчер БД выполняет следующие функции:

- хранение и предоставление пользователям текущей информации о доступных БД;
- формирование запросов к серверам БД и возвращение клиентам полученной информации в требуемой кодировке;
- хранение информации о правах доступа на каждую доступную БД и проверка их для каждого пользователя;
- учет и сбор статистики обращений к БД в соответствии с текущими установками;
- синхронизация версий серверов БД и их обновление;
- при наличии уровня 4 передача служебной информации о себе и о поддерживаемых базах данных на уровень 4.

Для организации полнофункциональной системы достаточно перечисленных трех уровней. Однако при построении территориально распределенной системы с ярко выраженными районами и ненадежными линиями связи между ними желательно локализовать все три уровня в каждом районе с интеграцией последних на уровне 4.

Уровень 4. *Главный диспетчер (ГД)* информационной системы представляет собой сервер WWW, функционально идентичный серверу уровня 3, но наделенный дополнительной функцией хранения информации о всей информационной системе в целом. В идеальном случае каждый из серверов уровня 3 должен быть готов взять на себя роль главного диспетчера. Основная задача ГД — получить информацию о конфигурации каждого сервера уровня 3 и растаждировать ее по всем серверам.

Таким образом, общая схема распределенной информационной системы состоит из четырех логических уровней. Организация их взаимодействия представлена на рис. 2.

3. Сегодняшняя реализация информационной системы

Описанная выше схема распределенной информационной системы доступа к библиографическим базам данных частично реализована на NT-серверах ОИГГиМ СО РАН

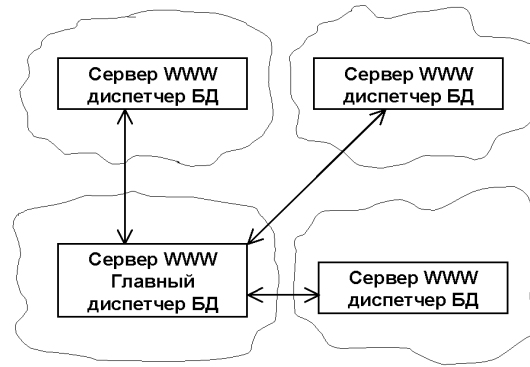


Рис. 2. Интеграция диспетчеров БД на уровне 4.

(<http://proxy.uiggm.nsc.ru>) и находится сегодня в режиме интенсивного тестирования.

Стержневым пунктом всей системы следует считать разработанный авторами сервер доступа к данным ISIS, реализующий все функции этой СУБД по выборке информации и форматированию. Исполненный в виде динамической библиотеки Win32, этот модуль компактен и управляется описанной выше системой команд, включающей обычную командную строку ISIS. Характерной чертой этого сервера БД является отсутствие функций записи данных и построения индексов. Предполагается, что эти функции избыточны для публичной информационной системы. Такой подход позволяет сильно упростить логику работы сервера и не заботиться о поддержке целостности данных. Что касается функций записи и перестроения индексов, то, по нашему мнению, с ними вполне справляется базовая система ISIS, обладающая всеми перечисленными функциями пополнения данных, но не способная функционировать в многопользовательском сетевом режиме.

Другим элементом информационной системы, заслуживающим отдельного рассмотрения, является диспетчер БД (уровень 3). В ОИГГиМ СО РАН он реализован на сервере Window NT 4.0 sp3 в среде Microsoft IIS 3.0. Такая реализация, на наш взгляд, наиболее экономична ввиду доступности программного обеспечения и его низкой стоимости, с одной стороны, а с другой — наличия достаточно развитых средств организации межпрограммного взаимодействия и удовлетворительной общей производительности системы. Диспетчер БД представляет собой динамическую библиотеку Win32, управляемую со стороны сервера WWW через интерфейс ISAPI — интерфейс доступа к динамическим библиотекам сервера Microsoft IIS.

В заключение авторы приглашают заинтересованных лиц принять участие в тестировании информационной системы.

Список литературы

- [1] ВИСЛЫЙ А. И. Региональные библиотечные компьютерные сети. В *“Матер. междунар. конф. “Крым-95”*, 1, 1995, 176.
- [2] ГОНЧАРОВ М. В. Телекоммуникационные системы ГПНТБ России: современное состояние и перспективы. *Там же*, 2, 1995, 26.

- [3] ЕРОНИНА Е. А. Перспективы развития электронной доставки документов в ГПНТБ России. **Там же**, 44.
- [4] БОБРОВ Л. К. Интернет и информационно-библиотечная система СО РАН: проблемы и перспективы. В *“Матер. 3 международ. конф. “Крым-96”*, **21**, 1996, 10.
- [5] ШРАЙБЕРГ Я. Л. Международная ассоциация пользователей CDS/ISIS: реальное взаимодействие и реальная польза для библиотек. В *“Матер. международ. конф. “Крым-95”*, **1**, 1995, 28.
- [6] МАРШАК Б. И., ГАРБАРЧИК В. А. ГПНТБ России — региональный центр ЮНЕСКО по распространению системы CDS/ISIS. **Там же**, 58.
- [7] КИРИЛЛОВА О. В., МАЗОВ Н. А. Современные технологии в библиотеке академического института. **Там же**, 218.
- [8] HTTP: A protocol for networked information, документ WWW, (<http://www.w3.org/hypertext/WWW/Protocols/HTTP/HTTP2.html>)
- [9] BERNERS-LEE, T., FIELDING, R. T. AND FRYSTYK NIELSEN, H. Hypertext Transfer Protocol — HTTP/1.0, Work in progress (draft-ietf-http-v10-spec-04.txt), MIT/LCS, UC Irvine, October 1995.
- [10] ЛАДЫЖЕНСКИЙ Г. М. Технология клиент — сервер и принципы Upsizing. *Сети*, **3**, №36, 1995, 32.
- [11] BERNERS-LEE, T. AND CONNOLLY, D. Hypertext Markup Language — 2.0, RFC 1866, MIT/W3C, November 1995.
- [12] МАЗОВ Н. А., ЖИЖИМОВ О. Л., ПОМОРЦЕВ А. Ю. Организация доступа к базам данных CDS/ISIS-M из Internet. В *“Матер. 3 международ. конф. “Крым-96”*, **2**, 1996, 177.

Поступила в редакцию 24 апреля 1997 г.