



- ▶ Журнал ЭБ
- ▶ О журнале
- ▶ Редакционная коллегия и редакция
- ▶ Что нового?
- ▶ 2014 год
- ▶ 2013 год
- ▶ 2012 год
- ▶ 2011 год
- ▶ 2010 год
- ▶ 2009 год
- ▶ 2008 год
- ▶ 2007 год
- ▶ 2006 год
- ▶ 2005 год
- ▶ 2004 год
- ▶ 2003 год
- ▶ 2002 год
- ▶ 2001 год
- ▶ 2000 год
- ▶ 1999 год
- ▶ 1998 год

▶ ENGLISH

Электронные библиотеки -1999 - Том 2 - Выпуск 2

Модель распределенной информационной системы Сибирского отделения РАН на базе протокола Z39.50

О.Л. Жижимов, Н.А. Мазов

Объединенный Институт Геологии, Геофизики и Минералогии СО РАН

Основные компоненты и общая схема информационной системы

В настоящей работе описывается один из возможных способов построения информационной системы на основе протокола Z39.50 для доступа к распределенным библиографическим базам данных.

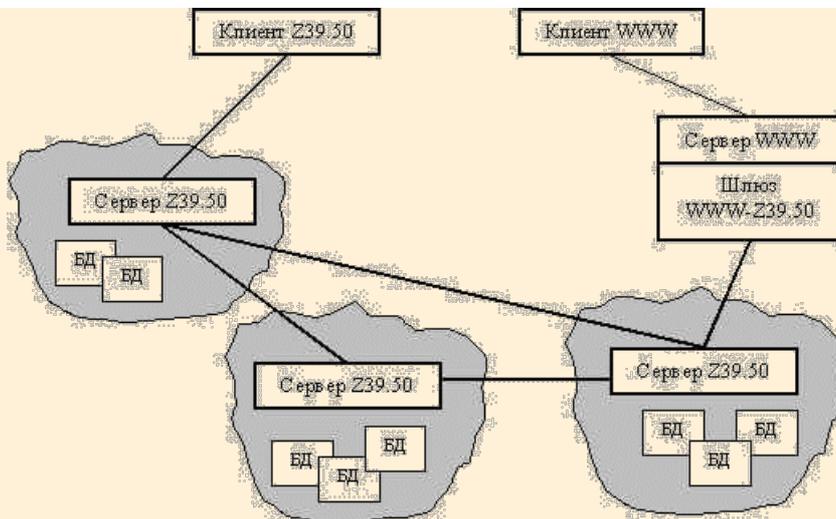
Описываемая система реализована на сети серверов ОИГГИМ СО РАН, ИВТ СО РАН и ГПНТБ СО РАН и находится в настоящее время в режиме опытно-промышленной эксплуатации

(<http://geolibr.uiggm.nsc.ru>, z3950: geolibr.uiggm.nsc.ru:210).

Толчком для создания подобного программного комплекса в ОИГГИМ СО РАН явилась задача организации унифицированного сетевого доступа к библиографическим базам данных Института, находящихся под управлением интегрированной информационно-поисковой системы CDS/ISIS/M версии 3.0 и выше (далее - ISIS), имеющейся также в библиотеках академических институтов СО РАН, в ГПНТБ СО РАН, с использованием возможностей сети Internet и сервиса WWW, предоставляемого этой сетью. При этом принимались во внимание следующие требования, необходимость выполнения которых была продиктована опытом эксплуатации локальных информационных систем:

- **Работа с распределенными данными** – информационная система должна допускать возможность работы с данными, расположенными на разных физических серверах, различных аппаратно-программных платформах и хранящихся в различных внутренних форматах.
- **Логическая группировка данных** – система должна позволять обрабатывать все запросы на логических группах баз данных, полностью скрывая тем самым физическое расположение последних.
- **Абстрактная модель данных** – информационная система строится на основе абстрактной схемы данных, на которую должны быть отображены конкретные базы данных. Это позволяет объединять данные из разнородных систем в одной логической группе.
- **Абстрактная система запросов** – система должна оперировать не конкретным синтаксисом запросов, а его логической сутью на основе абстрактных атрибутов.
- **Метаинформация** – система должна предоставлять полную информацию о себе и о всех своих ресурсах.
- **Разграничение доступа** – система должна быть способна предоставлять различные уровни привилегий для пользователей по доступу к информации
- **Учет и контроль** – система должна уметь собирать статистику по запросам пользователей и вести их бюджеты.
- **Открытость** – система должна легко расширяться и быть основана на открытых стандартах и протоколах.
- **Связь с другими системами** – возможность интегрировать свои ресурсы с ресурсами других информационных систем.
- **Легкость в общении** – для пользователей система должна предоставлять простые интерфейсы доступа к информации. Необходимым элементом системы сегодня является также наличие WEB-шлюза.

Рис.1. Схема распределенной информационной системы.



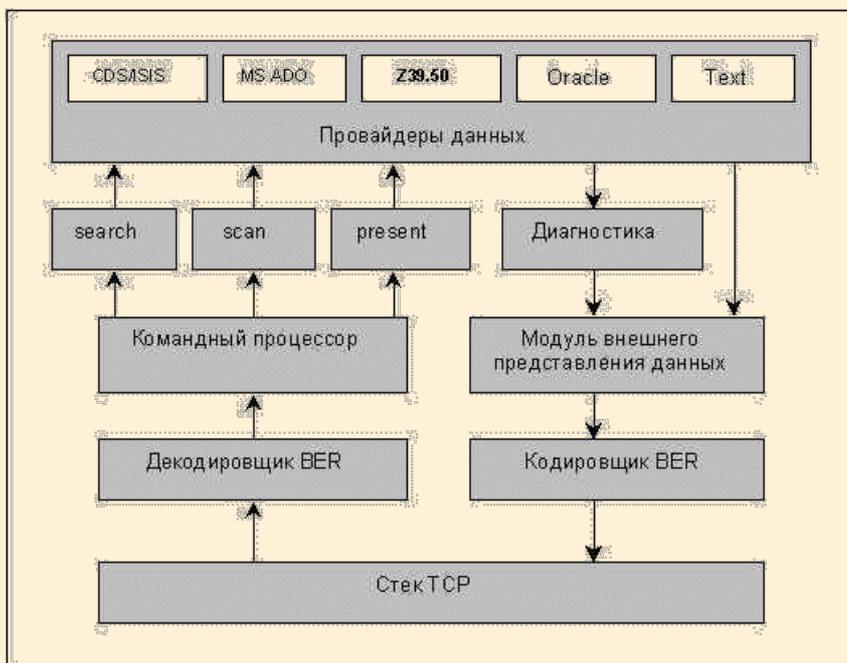
Авторами был построен некий прототип такой информационной системы [1], схема которой представлена на рис.1. Центральное место в информационной системе, несомненно, занимает сервер баз данных, поддерживающий протокол Z39.50. Для этого был разработан модульный сервер Z39.50, допускающий работу с данными различных СУБД. Включение в поисковую систему данных из новой СУБД для описываемого сервера сводится к разработке специального модуля с фиксированным программным интерфейсом (провайдер данных).

Архитектура и основные характеристики сервера

Для информационной системы разработан специальный сервер, удовлетворяющий требованиям:

- Поддержка протокола Z39.50-1995 (v.3) [2,3].
- Работа с различными СУБД.
- Переносимость на различные аппаратные платформы.

Рис. 2. Основные модули сервера Z39.50.



Немаловажную роль при этом играло желание максимально использовать существующие открытые разработки и библиотеки поддержки Z39.50, распространяемые другими разработчиками в исходных текстах. Последнее позволяет применить накопленный мировым сообществом опыт и избежать многих ошибок в реализации базовых функций Z39.50. В результате анализа существующих открытых пакетов выбор был остановлен на библиотеке YAZ компании IndexData [4]. Модули этой библиотеки после модернизации в направлении поддержки национальной специфики легли в основу самого низкого уровня сервера – управление сеансом связи и обработка APDU. Из этого же пакета была заимствована идея организации логики внутреннего командного процессора сервера, однако содержание этого процессора было практически полностью изменено. Более высокие уровни сервера – интерфейс провайдера данных, модули различных провайдеров данных и другие являются полностью оригинальными разработками. Ниже приведены основные характеристики сервера:

- Поддерживаемые аппаратные и программные платформы:
 - Windows NT 4.0 SP3 и выше Windows 2000 Srv (Intel) (вер.1, вер.2)
 - Solaris 2.2 – 2.6 (Intel, Sparc) (вер.2)
 - FreeBSD (Intel) (вер.2)
 - Linux (Intel) (вер.2).
- Внутренняя архитектура:
 - многопоточность (Windows NT);
 - модульная организация с выделением логических уровней обработки запросов и фиксацией межуровневых интерфейсов;
 - возможность расширения за счет создания дополнительных провайдеров данных
- В части поддержки Z39.50:
 - Версия протокола: Z39.50-1995 v.3
 - Система запросов: Type-1 (RPN).
 - Поиск (search).
 - Представление (present) – предоставляет записи в форматах: SUTRS, GRS-1, Usmarc, Explain.
 - Просмотр словаря (scan).
 - Работа с серверными наборами данных: включение в запрос, удаление
 - Работа с базой метаданных Explain – поддержка категорий CategoryInfo, TargetInfo, DatabaseInfo,
 - SchemaInfo, TagSetInfo, AttributeSetInfo, RecordSyntaxInfo и др.
 - Многобазовость на уровне search, present, scan – перечисленные функции выполняются прозрачно для пользователя над группами баз данных безотносительно к их физическому расположению.
 - Работа с виртуальными базами данных – формирование логических групп баз данных, объединенных общими свойствами.
 - Кириллические кодовые таблицы - поддержка сервером трехступенчатой (клиент – сервер - база данных) перекодировки между кодовыми таблицами ALT, ANSI, ISO, MAC, KOI8. Информация о кодовой таблице клиента поступает в момент инициализации сеанса в APDU Init, информация о кодовой таблице базы данных хранится в конфигурационных файлах.
 - Контроль прав доступа к данным.

Функциональные характеристики сервера существенно зависят от возможностей поставщиков (провайдеров) данных. Все провайдеры данных представляют собой динамически загружаемые библиотеки Win32 (для версии 1) или отдельные модули, включаемые в код сервера на стадии компоновки (для версии 2). На сегодняшний день в комплект сервера входят следующие провайдеры данных:

- *Провайдер данных CDS/ISIS (версии 1,2)* - обеспечивает доступ к базам данных CDS/ISIS. В версии 1 сервера доступ к данным организован через динамически загружаемую библиотеку Win32 - модуль WWWISIS (Bireme), в версии 2 сервер содержит оригинальный программный код провайдера данных CDS/ISIS, включающий поисковую машину. Этот провайдер работает с данными, для создания которых используется другое программное обеспечение, например CDS/ISIS v.3.*.
- *Провайдер удаленного доступа Z39.50 (версии 1,2)* - обеспечивает перенаправление запроса к другим серверам по протоколу Z39.50. Этот провайдер данных реализует прозрачный доступ пользователя к распределенным по разным серверам ресурсам.
- *Провайдер доступа к файл-организованным текстовым данным (версия 2)* – основан на коде Zebra (IndexData) и поддерживает доступ к текстовым и MARC базам данных, организованных в виде обычных файлов. Совместно с пакетом индексации Zebra [5] позволяет включить в информационную систему Z39.50 полнотекстовые данные.
- *Провайдер метаданных (версии 1,2)* - обеспечивает доступ к специальной базе данных IR-Explain-1, содержащей (согласно протоколу Z39.50), описание сервера, баз данных, атрибутов, выходных форматов и другую метаинформацию (см.выше).
- *Провайдер данных MS-ADO (версия 1)* - обеспечивает извлечение данных через интерфейсы MS ADO из OLE DB, а, следовательно, из всего множества OLE DB провайдеров данных, в том числе ODBC-совместимых СУБД (например, из MS-SQL, FoxPro).

Шлюз Z39.50-WWW

В качестве универсального клиента описываемой информационной системы авторами был создан шлюз Z39.50-WWW (<http://geolibr.uiggm.nsc.ru/LWS/z39gw/>). Шлюз функционирует в среде Windows NT 4.0 sp3 и выше. Он выполняет функции клиента Z39.50 на сервере WWW (IIS 3.0 и выше) для доступа к серверам баз данных Z39.50 с сохранением контекста сессии. Шлюз представляет собой набор

серверных (ASP) и клиентских (JScript) скриптов, управляющих обычными HTML-формами. Основной функциональный компонент шлюза выполнен в архитектуре COM для локализации сеансовой логики Z39.50 в отдельном модуле (ZCOM), который регистрируется в сессионных переменных IIS для сохранения контекста сессии Z39.50. Это позволяет вызывать интерфейсы ZCOM из обычных страниц ASP и предоставлять пользователю удовлетворительный сервис для доступа к информационной системе. Шлюз позволяет:

- выбирать различные кириллические кодовые таблицы (ALT,ANSI,ISO,KOI8,MAC);
- формировать запросы RPN через интерактивный построитель запросов, используя различные атрибуты;
- проводить поиск в базах данных и просматривать найденные записи в форматах (SUTRS, GRS-1, MARC, Explain);
- выбирать набор элементов и количество просматриваемых записей;
- просматривать содержание индексируемых полей;
- подключаться к любому серверу Z39.50 v.3.

Информационная система ОИГГиМ СО РАН

В настоящее время в ОИГГиМ СО РАН функционирует информационная система, построенная на базе вышеописанного сервера Z39.50, для обеспечения доступа научных сотрудников к библиографической информации (в первую очередь аналитического характера - электронные варианты баз данных ВИНТИ и др.). Она является прототипом большой информационной системы СО РАН и состоит из четырех серверов Z39.50, обеспечивающих доступ к базам данных под управлением CDS/ISIS (библиографическая аналитическая информация), к базам данных MS SQL (справочная информация) и текстовым данным в формате SGML. Информационная система интегрирована в WWW через шлюз Z39.50-WWW, функционирующий в среде сервера IIS 4.0 (Windows NT 4.0 sp3). Шлюз реализован в виде ASP-модулей с применением технологии COM. Доступ к шлюзу открытый, однако, доступ к собственным информационным ресурсам, естественно, ограничен. Поскольку этот шлюз (<http://geolibr.uiggm.nsc.ru/LWS/z39gw>) поддерживает многие функции клиента Z39.50 его можно использовать для работы с любыми серверами Z39.50, которых в мире насчитываются сотни.

Следует отметить, что основной сервер информационной системы ОИГГиМ СО РАН официально зарегистрирован в агентстве Z39.50 (<http://lcweb.loc.gov/z3950/agency/>). Он прошел тестирование и сертификацию на предмет соответствия протоколу Z39.50. Тестирование проводилось на электронном каталоге ГПНТБ СО РАН. Также официально зарегистрирована группа разработчиков сервера с присвоением уникального идентификатора (ID=155), позволяющего в рамках стандарта разрабатывать собственные объекты Z39.50.

Опыт эксплуатации сервера Z39.50 в составе информационной системы ОИГГиМ СО РАН показал его жизнеспособность и эффективность на больших информационных массивах библиографической информации. Для апробации работы сервера в распределенной информационной системе в ГПНТБ СО РАН был установлен Z39.50 сервер версии 1 с доступом к электронному каталогу. При этом тестировалось: возможность перенаправления запросов (работа провайдера Z39.50), логическое объединение пространственно разнесенных и разнородных данных (работа провайдеров CDS/ISIS и MS ADO), время реакции системы. Наряду с этим сервер был успешно апробирован в ГПНТБ России (Москва) на электронном каталоге.

На сегодняшний день ведется опытно-промышленная эксплуатация сервера версии 2 в составе распределенной информационной системы. Серверы установлены в трех организациях СО РАН, при этом используются различные аппаратно-программные платформы (Windows NT, FreeBSD, Solaris-SPARC, Solaris-Intel).

В настоящее время ведутся работы по расширению возможностей сервера в части поддержки дополнительных функций Z39.50 (сортировка, расширенный сервис, управление доступом и др.).

Ссылки

1. Жижимов О.Л., Мазов Н.А., Фролов А.С. Доступ к базам данных ISIS из Internet и построение распределенной информационной системы. Вычислительные технологии, т.2, №3, 1997, стр.45-50.
2. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50-1995, July 1995.
3. Племнек А.И., Усманов Р.Т. Z39.50: открытый доступ к библиографической информации. Научные и технические библиотеки, №8, 1998, стр.24-28.
4. Index Data. YAZ User's Guide and Reference. Version 1.4. (<http://www.indexdata.dk>).
5. Index Data. Zebra Server - Administrators's Guide and Reference. Revision 1.38. (<http://www.indexdata.dk>).

Последнее обновление страницы было произведено: 2003-12-09

Все предложения и пожелания по содержанию и структуре портала направляйте по адресу rdlp@iis.ru

