



Девятая Международная Конференция «Крым 2002»

Конференция проводится в рамках мероприятий ИФЛА 2002 г.

***Библиотеки и ассоциации
в меняющемся мире: новые технологии
и новые формы сотрудничества***

Тема 2002 года:

***Электронные информационные
ресурсы и социальная значимость
библиотек будущего***

***Труды конференции
Том 2***

СУДАК

(основная программа)

Алушта, Бахчисарай, Белогорск, Феодосия, Керчь, Коктебель, Старый Крым

(выездные заседания)

Автономная Республика Крым, Украина

8–16 июня 2002

Москва
Издательство ГПНТБ России
2002

Реляционные базы данных и протокол Z39.50 на примере сервера ZooPARK
Relational Databases and Z39.50 Protocol Exemplified by ZooPARK Server
Реляційні бази даних і протокол Z39.50 на прикладі сервера ZooPARK

Скибин С. В., Жижимов О. Л., Мазов Н. А.
Объединенный Институт Геологии, Геофизики и Минералогии
Сибирского Отделения РАН, Новосибирск, Россия

Sergei V. Skibin, Oleg L. Zhizhimov, Nikolai A. Mazov
Joint Institute of Geology, Geophysics and Mineralogy Siberian Branch
of The Russian Academy of Sciences, Novosibirsk, Russia

Скибін С. В., Жижимов О. Л., Мазов М. А.
Об'єднаний інститут геології, геофізики і мінералогії
Сибірського відділення РАН, Новосибірськ, Росія

Обсуждаются детали доступа к реляционным базам данных по протоколу Z39.50 для сервера ZooPARK. В качестве источников данных рассматриваются широко известные MS SQL Server, MySQL, а также интерфейсы

MS ADO. Приводятся примеры работающих решений для библиографических и кадровых баз данных, а также для тезаурусов и классификационных схем. Доклад ориентирован на специалистов, планирующих обеспечение доступа к информации по протоколу Z39.50 из уже существующих реляционных баз данных.

The details of providing access to relational databases via Z39.50 protocol for the ZooPARK server are discussed. Widely known MS SQL Server, MySQL and MS ADO are considered as the sources of data interfaces. Some examples of working decisions for bibliographic and personnel databases, and also for thesauruses and classification circuits are presented. The presentation is intended for experts supporting access through Z39.50 protocol to data in the relational databases.

Висвітлюються деталі доступу до реляційних баз даних за протоколом Z39.50 для сервера ZooPARK. Як джерела даних розглядаються широко відомі MS SQL Server, MySQL, а також інтерфейси MS ADO. Наводяться приклади діючих вирішень для бібліографічних і кадрових баз даних, а також для тезаурусів і класифікаційних схем. Доповідь орієнтована на фахівців, що планують забезпечення доступу до інформації за протоколом Z39.50 із вже існуючих реляційних баз даних.

Протокол Z39.50 [1] на сегодняшний день является международным стандартом ISO-23950. Он регламентирует сетевой доступ к базам данных, содержащим различного рода информацию, причем не только библиографическую, но и, например, информацию по цифровым коллекциям, метainформацию по ресурсам с пространственной привязкой (карты, космические снимки) и др. Используя этот протокол можно унифицировать работу с тезаурусами и классификационными схемами, интерес к которым непрерывно растет.

Регламентируя правила обмена данными по сети, протокол Z39.50 не описывает способов хранения этих данных. Для хранения данных сегодня чаще всего используются коммерческие СУБД, которые ориентированы на реляционные схемы данных и на язык SQL. Наиболее мощные серверы баз данных сегодня – SQL-серверы различных производителей (Oracle, MS-SQL, Informix, DB2 и т. д.). Эти серверы плохо совместимы между собой: основываясь на стандартном языке SQL, они используют различные сетевые протоколы, что затрудняет построение распределенных гетерогенных информационных систем.

Несмотря на то, что существуют решения унификации доступа к различным реляционным системам управления базами данных (PCУБД), основанные как на унификации API (ODBC, MS ADO), так и на функционировании промежуточных подсистем и серверов (CORBA, Applications Servers), универсальное решение проблемы возможно только при стандартизации сетевого обмена и схем данных. Именно эти моменты хорошо описаны и регламентированы в Z39.50.

Настоящий доклад посвящен механизмам и практике доступа к PCУБД по протоколу Z39.50 для сервера ZooPARK.

Различные PCУБД сегодня – достаточно развитые системы. Обычно существует полный инструментарий, т. е. набор программ, разработанный специально для работы с конкретными базами данных. При возникновении потребностей предоставления данных по протоколу Z39.50, например, для интеграции в мировое библиотечное сообщество, существует возможность настроить некоторые из серверов Z39.50, например ZooPARK, на использование уже созданных баз данных. При таком подходе весь инструментарий (программы для ввода, редактирования, просмотра данных) остается полностью пригодным, но добавляется поддержка протокола Z39.50.

При определении схемы данных и абстрактной структуры записи в Z39.50 принято иерархическое представление данных. Несмотря на то, что такое представление достаточно естественно, реальные СУБД чаще всего ориентированы на реляционную схему и на язык SQL. Почти очевидное решение проблемы совместимости SQL-серверов – использование протокола Z39.50 – наталкивается на проблему отображения внутреннего реляционного представления данных во внешнее иерархическое и обратно. Эта задача практически не решается даже для простых случаев просмотра, не говоря уже о неизбежных потерях при передаче данных от одного SQL-сервера к другому по протоколу Z39.50, когда необходимо сохранять именно реляционную структуру и типы данных.

Именно для решения проблемы однозначной интерпретации клиентом передаваемой по Z39.50 информации из реляционных баз данных был создан профиль Z+SQL [6].

Z+SQL включает в себя следующие дополнения к стандарту Z39.50:

- Новый механизм запросов: type-104 {Z39.50 16 1}
- Новый формат внешнего представления данных: SQL-RS {Z39.50 5 111}
- Дополнительные диагностические сообщения
- Дополнения формата внешнего представления Explain

Однако профиль Z+SQL до сих пор не нашел поддержки не только у производителей РСУБД, но и у разработчиков программного обеспечения Z39.50. Поэтому вопрос о доступе к РСУБД по протоколу Z39.50 обычным для этого стандарта способом (запросы RPN, представление записей в GRS-I, SUTRS и др.) остается актуальным, а также остаются актуальными задачи преобразования запросов RPN в SQL и реляционной структуры записи в иерархическую. Именно эти задачи и способы их решения рассматриваются ниже применительно к серверу ZooPARK.

ZooPARK является модульным сервером Z39.50. Доступ к различным СУБД обеспечивается в нем при помощи динамически загружаемых модулей, называемых провайдерами данных. Все детали функционирования конкретных СУБД скрыты в этих модулях, а сервер ZooPARK взаимодействует с ними при помощи

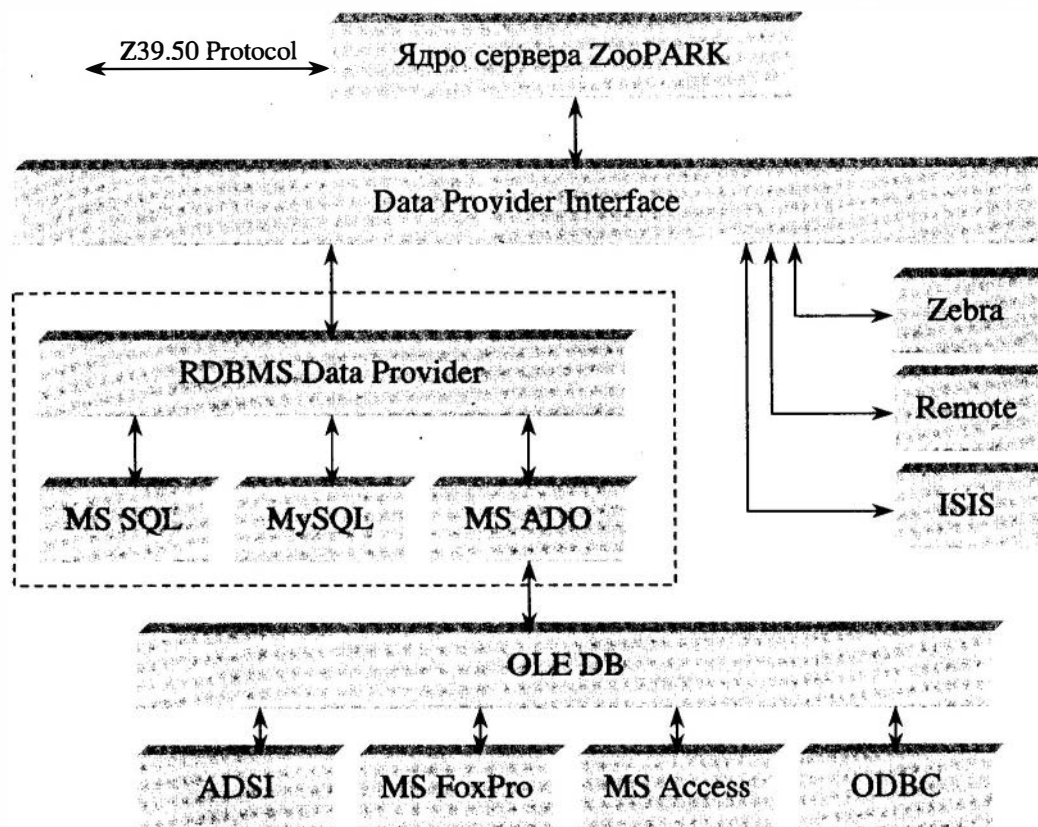


Рис.1. Функциональная схема доступа к данным сервера Zoo-

унифицированного интерфейса [5]. Доступ к РСУБД в сервере ZooPARK осуществляется также через общий интерфейс, однако специфика РСУБД позволяет выделить дополнительный слой, обобщающий функции различных провайдеров данных РСУБД и унифицировать промежуточный интерфейс.

Исходя из анализа существующих РСУБД, был создан прототип провайдера реляционных данных, на основании которого можно создать провайдеры практически для любой РСУБД. Помимо этого существует провайдер данных для MS ADO (ActiveX Data Objects), обеспечивающий доступ через интерфейсы MS ADO и ODBC к РСУБД и другим источникам данных.

На рис. 1 показана функциональная схема сервера ZooPARK при обеспечении доступа к различным СУБД, в том числе и РСУБД. Отдельно выделен провайдер данных MS ADO. Из рисунка видно, что сервер ZooPARK является промежуточным звеном между конечным пользователем и данными. Таким образом, пользователь через механизм Z39.50 получает доступ к информации по стандартным схемам данных, осуществляя поиск по стандартным наборам поисковых атрибутов, извлекая записи в требуемых форматах внешнего представления (XML, GRS-I, HTML и пр.). При этом, естественно, от пользователя полностью скрывается информация о реализации конкретной СУБД, в том числе и РСУБД.

Для обеспечения работоспособности данной схемы необходим, во-первых, механизм преобразования запросов RPN в запросы SQL и установления соответствия между поисковыми атрибутами и структурой конкретной базы данных. В настоящей версии сервера ZooPARK это достигается через текстовые настроеч-

ные файлы для каждой БД и конвертера RPNtoSQL, примерный код которого поставляется как часть тестового провайдера данных вместе с сервером ZooPARK.

Вторая проблема – преобразование извлекаемых записей из реляционной схемы хранения данных в иерархическую в соответствии со спецификациями Z39.50. При таком преобразовании может потребоваться произвести несколько взаимосвязанных запросов к базе данных. Для этих целей сервер ZooPARK содержит специальный интерпретатор для обработки ASP-подобных текстовых файлов, в которых описывается логика формирования иерархической XML-записи для преобразования последней в экспортируемую внешнюю структуру в соответствии с требуемым форматом [3].

Ниже приведен пример фрагмента такого файла для схемы ZTHES:

```
<zthes>
  <local-control-number> $r.id$ </local-control-number>
  <termQualifier> $r.id$ </termQualifier>
  <language>eng</language>
  <Title> $r.term$ </Title>
  <termType>$r.type$</termType>
  <Description> $r.def$ </Description>
  <$ sql q select a.id id, a.name term, 'BT' rtype from term a,term2term b where
    a.id=b.term1_id and b.term2_id='$r.id$' $>
  <$ for q $>
  <relation>
    <relationType> $q.rtype$ </relationType>
    <local-control-number> $q.id$ </local-control-number>
    <termQualifier> $q.id$ </termQualifier>
    <Title> $q.term$ </Title>
  </relation>
  <$ close q $>

  <$ sql q select b.term2_id id, a.name term, 'NT' rtype from term a,term2term b where
    a.id=b.term2_id and b.term1_id='$r.id$' $>
  <$ for q $>
  <relation>
    <relationType> $q.rtype$ </relationType>
    <local-control-number> $q.id$ </local-control-number>
    <termQualifier> $q.id$ </termQualifier>
    <Title> $q.term$</Title>
  </relation>
  <$ close q $>
...
</zthes>
```

На этом примере показано, как формируется иерархическая структура записи. Жирным шрифтом выделены команды выполнения и обработки вложенных запросов. Сервер ZooPARK использует эти команды при обработке соответствующих схем данных. Информация после выполнения вложенного запроса подставляется в конкретные позиции в соответствии с шаблоном. В результате получается структура, соответствующая конфигурационному файлу-шаблону, но содержащая реальную информацию.

На следующем этапе информация преобразуется к требуемому формату представления данных (GRS-1, XML, HTML, RTF, SUTRS и др.) [2], устанавливается требуемая кириллическая кодировка. Затем данные передаются клиенту Z39.50. Клиентом может быть как специализированная программа, так и WWW-шлюз Z39.50.

Можно привести несколько примеров применения описанных механизмов обработки данных в реально функционирующих информационных системах.

1. **Информация о сотрудниках** (<http://z3950-2.uiggm.nsc.ru/zgwk/kadrs.htm>) была получена из MS SQL Server с использованием специально разработанной схемы данных PERSONS и может быть довольно обширной, но ограничена из соображений конфиденциальности.
2. **Тезаурусы и классификационные схемы** (см. http://z3950-2.uiggm.nsc.ru/zgwc/thes_start.htm), доступные через шлюз по протоколу Z39.50. На указанном шлюзе представлены всевозможные рубрикаторы и классификационные схемы. GEMET, MSC2000 находятся на MS SQL Server, Gene Ontology – на MySQL.

Библиографические описания на примере организации доступа по протоколу Z39.50 к данным простой библиотечной системы на основе FoxPro, где задействован провайдер данных MS ADO. Ниже приведен пример настроечного файла для формирования структуры записи RUSMARC через XML.

```

<rusmarc>
  <marker>
    <status>n</status>
    <codes>a$r.bu$##</codes>
    <user>z##</user>
    <future>4</future>
  </marker>

<001><#>TestVFP\$r.sid$</#></001>
<101><0#><a>rus</a></0#></101>
<105><##><a>y |||000yy</a></##></105>
<200><1#><a>$r.bl$</a></1#></200>
<210><##>
  <a>$r.tl$</a>
  <c>$r.il$</c>
  <d>$r.yl$</d>
  </##></210>
<205><##><a>$r.ih$</a></##></205>
<215><##><a>$r.kh$</a></##></215>
<330><##><a>$r.am$</a></##></330>

  <$ sql key select a.sname from skey a, keyw b where a.sid=b.ikey and
    b.isid=$r.sid$ $>
<$ for key $>
<606><##><a>$key.sname$</a></##></606>
<$ close key $>

<$ sql rub select a.sname from sr a, srub b where a.rsid=b.irub and
  b.isid=$r.sid$ $>
<$ for rub $>
<615><##><a>$rub.sname$</a></##></615>
<$ close rub $>

<$ sql cl select a.*, 'local' as local from smsc a where a.isid=$r.sid$ $>
<$ for cl $>
<686><##>
  <a>$cl.sname$</a>
  <2>$ci.local$</2>
  </##></686>
<$ close cl $>

<$ sql au select a.* from saut a where a.isid=$r.sid$ and a.izon=2 $>
<700><##><a>$au.sname$</a></##></700>
<$ next au $>

  <$ for au $>
<701><##><a>$au.sname$</a></##></701>
<$ close au $>

<$ sql aul select a.* from saut a where a.isid=$r.sid$ and a.izon=3 $>
<$ for aul $>
<702><##><a>$aul.sname$</a></##></702>
<$ close aul $>

  <$ sql auo select a.* from sorg a where a.isid=$r.sid$ and a.izon=4 $>
<710><##><a>$auo.sname$</a></##></710>
<$ next auo $>

  <$ for auo $>
<711><##><a>$auo.sname$</a></##></711>
<$ close auo $>

  <801><##><c>$r.dl$</c></##></801>
<899><##>
  <h>$r.pol$</h>
  <j>$r.azk$</j>
  <b>Тестовая база данных для MS ADO (Visual FoxPro) 2001</b>
  </##></899>
</rusmarc>

```

В докладе все перечисленные выше примеры демонстрируются на реально работающих системах.

Таким образом, описанные механизмы интеграции РСУБД на базе сервера ZooPARK позволяют построить довольно эффективные решения для распределенных информационных систем. Дальнейшее развитие описанного подхода мы видим в шлифовке внутренних интерфейсов сервера и провайдеров данных, а также в привлечении в качестве интерпретатора формирования структуры записи внешних модулей на основе Perl и PHP. Мы надеемся, что доклад будет интересен специалистам, использующим или планирующим использовать протокол Z39.50 и РСУБД.

Литература

1. 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 (<http://lcweb.loc.gov/z3950/agency>).
2. Жижимов О. Л. Введение в Z39.50. Новосибирск: Изд-во НГОНБ, 200, – 220 с. ISBN 5-88742-046-4. (<http://geolibr.uigmm.nsc.ru/docs/Z39.50/Z-Intro>).
3. Жижимов О. Л., Скибин С. В. О доступе к реляционным СУБД по протоколу Z39.50 // *Электронные библиотеки: перспективные методы и технологии. электронные коллекции. Сборник трудов конференции. Сборник трудов Третьей Всероссийской конференции. Петрозаводск, 2001г.*
4. Скибин С. В., Жижимов О. Л. Прототип унифицированного провайдера данных реляционных СУБД для сервера Z39.50 ZooPARK // *Труды конференции молодых ученых, посвященной 10-летию ИВТ СО РАН, Новосибирск, 25-26 декабря 2000 г. Т. 1: Информационные технологии, задачи поддержки принятия решений, Новосибирск, 2001, 66-69, библиогр.*
5. ZooPARK модульный сервер Z39.50. Версия 2. 59. ОИГГИМ СО РАН, 2002. <http://geolibr.uigmm.nsc.ru/doklads/Z-docs/ZooPARK.doc>
6. Z+SQL Profile. Final as of February 23, 2000 (<http://lcweb.loc.gov/z3950/agency>).