

Z+SQL: достоинства и недостатки на примере реализации

Жижимов О.Л.

Объединенный Институт Геологии, Геофизики и Минералогии
СО РАН

Применение протокола Z39.50 [1] для сетевого доступа к базам данных дает множество преимуществ при организации распределенных информационных систем. Главные из них - унификация всех процедур и глобальная стандартизация логических элементов взаимодействия субъектов. При этом основной моделью данных, на основании которой строятся схемы и определяются абстрактные структуры записей, является иерархическая модель. Именно в иерархической модели определены такие известные схемы данных как GILS, CIMI, GEO, которые используются при описании метаданных в глобальных информационных системах.

С другой стороны, в качестве систем хранения и управления данными сегодня чаще всего применяются системы реляционного типа на основе различных SQL-серверов (Oracle, Informix, MSSQL, DB2 и т.п.). Используемая в них реляционная модель данных в основе содержит совсем другие принципы, чем иерархическая модель Z39.50. Это усложняет, а иногда делает просто невозможным, отображение данных реляционных таблиц в структуры записей Z39.50, что несомненно ограничивает область применения как унифицированного подхода к работе с базами данных Z39.50, так и использование SQL-серверов при построении гетерогенных распределенных информационных систем.

В свете вышеизложенного очень интересной выглядит идея введения в стандарт Z39.50 возможностей работы с реляционными данными без нарушения общей идеологии (глобальные схемы данных, абстрактные поисковые атрибуты, унифицированный сетевой протокол и т.д.). В феврале 2000 года эта идея воплотилась в расширение стандарта Z39.50 - Z+SQL, которое описывается документом [2].

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

- Новый механизм запросов: type-104.

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

- Новый формат внешнего представления данных: SQL-RS.
- Дополнительные диагностические сообщения.
- Дополнения формата внешнего представления Explain.

Система запросов type-104 (SQL)

Профиль Z+SQL подразумевает использование запросов type-104 в двух вариантах: прямом и абстрактном.

Прямой вариант запроса – обычный запрос SQL в терминах реальных названий таблиц и полей, например

```
select title from collection where title like  
'%stamp%';
```

Допускается использование именованных результирующих наборов. Так, если в результате выполнения предыдущего запроса был создан набор Q1, запрос

```
select title from Q1 where title like '%american%';
```

будет выполнен только на множестве Q1.

Абстрактный вариант запроса отличается от прямого тем, что в нем указываются не реальные названия таблиц и полей, а OIDы и элементы схемы данных. Например, при отображении таблиц предыдущего примера на схему GILS предыдущий запрос можно представить в виде

```
select [(2,1)] from [1.2.840.10003.13.2,  
1.2.840.10003.3.5]  
where [4] like '%stamp%';
```

где указаны OID схемы GILS [1.2.840.10003.13.2], OID набора атрибутов GILS [1.2.840.10003.3.5], выбираемый элемент title [(2,1)] и поисковый атрибут USE title [4]. В абстрактном запросе можно сослаться на наборы элементов, например

```
select [*brief] from [1.2.840.10003.13.2,  
1.2.840.10003.3.5]  
where [4] like '%stamp%';
```

и использовать именованные результирующие наборы. Таким образом, абстрактный вариант запроса совмещает синтаксис SQL с абстрактным представлением данных Z39.50.

Формат представления SQL-RS

Формат представления SQL-RS содержит два обязательных блока: блок описания данных (метаданные) и блок данных

```
SQL-Result ::= SEQUENCE {
  tableDescriptor [0]IMPLICIT SQLTableDescriptor OPTIONAL,
  listOfResultValues [1]IMPLICIT SEQUENCE OF
    SQLRowValue OPTIONAL}
```

В блоке метаданных определяется структура результирующей таблицы: название, имена полей, тип данных и т.п., в блоке данных содержится собственно таблица данных. Допустимые в SQL-RS типы данных соответствуют типам данных SQL-1999, включая коллекции и структуры. Последнее позволяет использовать вложенные таблицы для более адекватного отображения реляционной схемы данных SQL на иерархическую схему данных Z39.50.

Таким образом, Z+SQL позволяет на уровне запросов и формата внешнего представления данных объединить две различные идеологии доступа к базам данных - SQL и Z39.50. Используя этот профиль можно

- организовать обмен данными между SQL-серверами различных производителей;
- обеспечить унифицированный доступ клиентов к реляционным базам данных без применения различных драйверов типа ODBC, шлюзов и другого программного обеспечения промежуточного слоя;
- интегрировать реляционные базы данных в рамках гетерогенных распределенных информационных систем.

Z+SQL и сервер ZooPARK

Для изучения возможностей Z+SQL в сервер ZooPARK [3] была встроена поддержка спецификаций, определенных в [2], как на уровне запросов, так и на уровне формата представления данных SQL-RS. В качестве SQL-сервера использовался MS SQL Server 6.5, соответствующий SQL-92. Для этого был разработан специальный провайдер данных для сервера ZooPARK в соответствии с общими требованиями к динамическим провайдерам данных. Провайдер данных Z-MSSQL обеспечивает

- связь с SQL-сервером MS-SQL Server по протоколу TCP/IP;
- обработку запросов Type-1 (RPN) и Type-104 (SQL);
- конвертацию запросов RPN в запросы SQL;
- представление данных в форматах SQL-RS, SUTRS, GRS-1, XML.

В качестве тестовой базы данных был выбран справочник сотрудников ОИГТиМ СО РАН, представляющий собой набор связанных нормализованных таблиц, отображаемых на схему GILS. Конвертация запросов RPN в SQL осуществлялась при помощи настраиваемых шаблонов с наборами поисковых атрибутов Bib-1 и GILS. Также при помощи шаблонов осуществлялся вывод данных в форматах SUTR, GRS-1 и XML.

Организация доступа к реляционной базе данных по протоколу Z39.50 с поддержкой спецификаций Z+SQL и опытная эксплуатация системы позволили сделать некоторые выводы относительно эффективности описываемой технологии:

1. Использование Z+SQL позволяет организовать достаточно эффективный доступ к реляционным базам данных, не зависящий от конкретного SQL сервера. Эта эффективность максимальна для простых таблиц и, естественно, быстро падает при увеличении числа таблиц, участвующих в запросе.
2. Необходимость поддержки традиционных для Z39.50 запросов RPN (эта поддержка обязательна для всех Z39.50 серверов) и форматов внешнего представления SUTRS, GRS-1 (для традиционных клиентов) не снимает проблем отображения реляционной модели на иерархическую схему данных. Проблема частично решается при отказе от традиционных форматов при работе со специализированными Z+SQL клиентами.
3. Формат SQL-RS с точки зрения клиента практически не вносит в извлекаемую запись дополнительной информации по сравнению с форматом GRS-1. Дополнительная информация относится к некоторым элементам блока метаданных, которая не нужна для обычного просмотра. Следует заметить, что самая важная информация из блока метаданных SQL-RS, например, названия таблиц и полей, может быть включена в элементы метаданных GRS-1. С другой стороны, сам факт включения в метаданные SQL-RS названий реальных таблиц и полей противоречит общей идеологии Z39.50, представление записей в которой основано на схеме данных и абстрактной структуре записи в терминах стандартных меток. Включение же в блок метаданных названий элементов схемы данных вместо названий реальных таблиц и полей делает различия между SQL-RS и GRS-1 практически несущественными. Более того, формат GRS-1 предпочтителен, ибо он является физической реализацией абстрактной структуры записи Z39.50. Единственным случаем, когда последнее несправедливо, является случай сетевого взаимодействия SQL-ориентированного программного обеспечения, например, взаимодействие SQL-сервер – SQL-сервер или SQL-сервер – ODBC-драйвер.
4. Несомненно положительным свойством SQL-RS является возможность формирования вложенных таблиц даже при работе с серверами SQL-92. Это может быть произведено на основе шаблонов и последовательных SQL-запросов при создании структуры SQL-RS сервером Z39.50.

5. Наиболее интересным аспектом Z+SQL является новый стандартный тип запросов для Z39.50 (Type-104) в своем абстрактном варианте. Нетрудно заметить, что в таком виде запросы Type-104 пригодны не только для реляционных СУБД, они полностью вписываются в идеологию Z39.50 и могут конкурировать с традиционными запросами RPN (Type-1, Type-101).
6. Запросы Type-104 в своем прямом варианте фактически ничем не отличаются от запросов Type-0 (запрос в синтаксисе конечной СУБД).

Таким образом, наряду с несомненно положительными свойствами Z+SQL существуют и отрицательные, а также свойства весьма сомнительного достоинства.

Тем не менее, программа Z+SQL для сервера ZooPARK продолжается. В настоящее время идет отладка провайдера данных для MySQL и для Oracle. Кроме этого идут работы по построению распределенной информационной системы с элементами Z+SQL, интегрирующей информацию по кадровому составу СО РАН.

В заключение следует заметить, что многие проблемы построения гетерогенных информационных систем были бы сняты, если бы производители коммерческих SQL-

серверов договорились между собой об едином протоколе обмена данными. В этом смысле Z+SQL – хорошая платформа для таких начинаний, т.к. является международным стандартом. К сожалению, на сегодняшний день поддержка Z+SQL отсутствует как со стороны производителей SQL-серверов, так и со стороны производителей серверов Z39.50. Можно надеяться, что ситуация со временем улучшится, т.к. Z+SQL еще слишком молод и только начинает признаваться мировым сообществом.

Список литературы

- [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.
- [2] Z+SQL Profile. Final as of February 23, 2000. http://archive.dstc.edu.au/DDU/projects/Z3950/Z+SQL/Z+SQL_profile.html
- [3] Жижимов О.Л. Введение в Z39.50. // Новосибирск: Изд-во НГОНБ, 2000 – 196 с., ISBN 5-88742-037-5.