

Организация Z39.50-ресурсов Томского консорциума  
«Открытая электронная библиотека»

**Проблема учета Z39.50-ресурсов**

До появления первых Z39.50-серверов в России вопрос об организации распределенных библиографических баз данных сводился к вопросу о возможностях определенного программного обеспечения или к организации сводных электронных источников (например веб-сайтов). С появлением коммерческих и свободно распространяемых информационных систем, которые предоставляют Z39.50-сервис, специалисты также сталкиваются с проблемой учета и поддержания в актуальном состоянии информации о Z39.50-ресурсах.

Понятие *Z39.50-ресурс* определяется нами как сущность, имеющая описание и обозначающая Z39.50-базу данных.

Частично трудности работы Z39.50-ресурсами решаются с помощью так называемых шлюзов доступа к Z39.50-среде, которые содержат локальное описание ресурсов. Но возникают ситуации, когда, к примеру, удаленный ресурс перемещен в Интернете, изменены название Z39.50-базы данных, точки доступа или политика доступа, а возможно, и что-то другое – все приводит к проблеме учета Z39.50-ресурсов.

**Три подхода к решению проблемы**

Решение проблемы видится нам в выборе формы организации и способа передачи специализированной информации о Z39.50-ресурсе между клиентом и сервером. Под *специализированной информацией о Z39.50-ресурсе* мы понимаем следующее: данные о Z39.50-сервере; данные о БД, удерживаемых Z39.50-сервером; данные о точках доступа к БД; данные о записях, расположенных в БД.

В совокупности специализированная информация о Z39.50-ресурсе может быть представлена как запись базы данных. Таким образом можно сформировать специализированную БД, которая будет содержать записи, описывающие множество Z39.50-ресурсов. Если предположить, что подобная БД полностью определяет Z39.50-ресурсы, тогда различные клиентские приложения, получившие доступ к ней, могут использовать полученную информацию для доступа к тому или иному Z39.50-ресурсу. Безусловно, для поддержания данных в актуальном состоянии специализированную БД требуется обновлять. Такие обновления могут проводиться как при участии человека, так и в автоматическом режиме.

Корпоративное объединение открытой электронной библиотеки (ОЭБ) [1] включает несколько десятков различных баз данных, доступных через протокол Z39.50, и вследствие этого стоит перед проблемой организации информации о Z39.50-ресурсах. Для ее решения рассмотрены три возможных подхода: *Z39.50 Explain* [2]; *Explain Lite* [3]; *ZeeRex* [4].

Рассмотрим основные аспекты использования предложенных подходов для ОЭБ.

**Z39.50 Explain** является частью стандарта Z39.50–1995 [5] и позволяет клиенту получить информацию о сервере как таковом, включая информацию о базах данных, удерживаемых сервером, о поддерживаемых наборах атрибутов (точках доступа) и другие параметры, которые могут быть использованы клиентом в процессе работы с сервером. Сервер, поддерживающий *Z39.50 Explain*, предоставляет доступ (через Z39.50-сервисы поиска и извлечения) к БД с предопределенным названием *IR-Explain-1*. Используя стандартизованный набор атрибутов Exp-1, клиент может извлечь необходимую детальную информацию о сервере. Данные представляются в *Explain*-синтаксисе, который определен стандартом Z39.50–1995.

*Explain*-запись хорошо детализирована и обладает некоторой избыточностью [6]. Подобная запись может быть получена только субъектом Z39.50 и только для конкретного сервера. Другими словами, для доступа к специализированной информации о группе серверов необходимо учредить сессионное соединение с каждым

из серверов, получить *Explain*-записи от каждого из серверов и только потом выполнять Z39.50-задачи с выбранными базами данных.

Безусловно, клиентское приложение может выполнить кэширование *Explain*-записей, что сократит накладные расходы, но это работает для одного клиента и трудно реализуемо в общем виде в силу того, что данные предоставляются в специальном синтаксисе и доступны только для Z39.50-субъектов.

В Томском библиотечном консорциуме насчитывается больше десяти Z39.50-серверов различных производителей и сегодня все серверы предоставляют Z39.50-сервис *Explain*. Наличие такого механизма получения данных о сервере мы рассматриваем как основу для организации Z39.50-ресурсов консорциума и стандартное решение Z39.50 *Explain* для этого – существенное дополнение.

***Explain Lite***. Как результат работы проекта ONE-2 [7] в 2001 г. появилось описание альтернативного, отличного от стандартного механизма, получившего название *Explain Lite* (облегченный *Explain*). Этот механизм (подход) не является частью стандарта Z39.50–1995 и изначально призван решить проблемы, актуальные для самого проекта ONE-2.

Появление *Explain Lite* обусловлено тем, что классический Z39.50 *Explain* не получил широкого распространения среди разработчиков Z39.50-приложений и необходимо было найти решение, которое позволило бы решить задачи проекта ONE-2. Суть подхода заключается в том, чтобы отказаться от использования ASN.1-структур (*Abstract Syntax Notation*) для описания Z39.50-ресурса, а использовать популярный язык разметки XML (*Extensible Markup Language*). Разработчики не ставили перед собой задачу провести реинжиниринг Z39.50 *Explain* модели, а искали более гибкий и более простой подход с точки зрения реализации. Как результат предложено использовать Init-сервис Z39.50 для передачи информации о Z39.50-ресурсах клиенту от сервера. В рамках *InitResponse* клиенту передаются описания ресурсов в виде XML-описаний, которые в дальнейшем используются им для организации взаимодействия с сервером. Такой подход является более гибким по сравнению с классическим Z39.50 *Explain* хотя бы потому, что описания Z39.50-ресурсов уже не «скрыты» в ASN.1-структурах, которые доступны только для Z39.50-субъектов, а XML-описания могут быть доступны поисковым роботам Интернета.

В процессе работы в проекте ОЭБ мы периодически сталкивались с задачей разделения информации о Z39.50-ресурсах между различными приложениями. Например, шлюз доступа, автономный клиент, программа-робот по сбору статистической информации из библиографических баз данных должны использовать одну и ту же информацию о Z39.50-ресурсах. На практике же получалось, что каждое приложение имело свой собственный формат хранения данных о ресурсах, и как временное решение мы использовали в качестве авторитетного источника БД для описания Z39.50-ресурсов, в которую вводили первичные изменения, а в дальнейшем с помощью дополнительных конвертеров преобразовывали в искомый формат для каждого из приложений.

Безусловно, использование *Explain Lite*, предложенного проектом ONE-2, позволило бы нам решить программистские проблемы, и мы могли бы предложить на его базе более качественный Z39.50-сервис, однако его использование затруднительно в силу следующих причин:

- размер APDU (*Application Protocol Data Unit*) потенциально ограничен и это означает, что размер описаний Z39.50-ресурсов также имеет потенциальное ограничение;
- предложенный подход ограничен в использовании наборов атрибутов и схемы данных в силу того, что решались задачи проекта ONE-2;
- не рассматриваются вопросы составления описания в многоязыковой среде;
- появление спецификаций *ZeeRex* (третий подход).

***ZeeRex*** (*Z39.50 Explain, Explained and Re-Engineered in XML*) – это третья попытка решения проблемы. Как следует из названия, этот подход – по сути переработанная классическая модель Z39.50 *Explain*, которая не наследует сложности применения модели Z39.50 *Explain* и исключает системные ограничения, присущие модели *Explain Lite*.

Концептуально *ZeeRex* базируется на обмене записями, которые описывают Z39.50-базу данных. Каждая запись соотносится только с одной БД, собранные вместе записи составляют *ZeeRex*-базу данных. Подобная БД может удерживаться обычным Z39.50-сервером и быть доступной для поиска и извлечения по протоколу

Z39.50. По аналогии с Z39.50-*Explain* на Z39.50-сервере *ZeeRex*-база данных именуется предопределенным именем *IR-Explain-1*. *ZeeRex*-запись формируется в XML-синтаксисе, который выступает в качестве коммуникативного формата передачи *ZeeRex*-записей между приложениями. Совсем необязательно, чтобы данные передавались только через Z39.50-протокол, это может быть другой протокол передачи данных. Полная *ZeeRex*-запись состоит из следующих разделов:

*serverInfo* – информация о Z39.50-сервере – хост, порт и название БД;

*databaseInfo* – информация о БД описательного характера;

*metaInfo* – метаинформация о самой *ZeeRex*-записи (когда создана, последний раз модифицирована, агрегирована);

*indexInfo* – информация о том, как искать в БД (информация о точках доступа);

*recodInfo* – информация о синтаксисе записей БД и наборах элементов записи (имеется в виду полный, краткий и т.д.).

*ZeeRex*-запись не изменяет форму при передаче между приложениями. Этот факт, возможно, привлекателен для разработчиков.

Z39.50-база данных поддерживает следующую комбинацию атрибутов, которые согласуются с архитектурой Z39.50-атрибутов: *Z39.50 Utility Attribute Set*; *Cross-Domain Attribute Set*; *ZeeRex Attribute Set*.

Используя традиционный Z39.50-сервис поиска и извлечения записей и комбинацию наборов атрибутов, Z39.50-клиент может получить полное описание искомой БД в XML-синтаксисе (полную *ZeeRex*-запись).

Сегодня в Интернете не более десятка *ZeeRex*-баз данных и спецификации *ZeeRex* время от времени изменяются. Первый доступный релиз появился 28 февр. 2002 г. Но, несмотря на относительно короткую историю, мы предпочитаем применять именно этот подход, так как считаем его наиболее подходящим и перспективным.

### **Организация Z39.50-ресурсов ОЭБ**

Не всегда академически правильное (красивое) решение может иметь успех на практике, а *ZeeRex* предложен для рассмотрения совсем недавно, но в нем учтены недостатки двух предыдущих подходов.

По большей части мы пытаемся решить следующие задачи:

- представить информацию о Z39.50-ресурсах разнородным клиентам в доступной форме;
- формально описать распределенный ЭК консорциума и, используя дополнительные технические решения, поддерживать описания в актуальном состоянии;
- минимизировать количество начальной информации, необходимой для взаимодействия клиентского приложения с Z39.50-средой.

Для решения поставленных задач мы поддерживаем *ZeeRex*-БД всех ресурсов консорциума на одном из наших серверов. Регулярный сервер консорциума также имеет *ZeeRex*-БД, которая генерируется автоматически в процессе работы сервера и удерживает информацию о локальных БД. Для поддержания в актуальном состоянии *ZeeRex*-БД корпоративного объединения используется программа-робот, которая умеет работать с *ZeeRex*-БД (*IR-Explain-1*). Программа-робот также способна работать с классической *Explain*-базой данных (*IR-Explain-1*) для формирования *ZeeRex*-записи с сервера, который не поддерживает БД *ZeeRex*, но поддерживает классический *Explain*. В этом случае робот формирует запись только на том языке, на котором представлена классическая *Explain*-запись.

На данном этапе *ZeeRex*-описания баз данных ОЭБ доступны по Z39.50-протоколу. Например, унифицированный веб-интерфейс [2] периодически обновляет локальную копию описаний БД посредством Z39.50-протокола, автономный клиент Z39.50 может синхронизировать описания БД по желанию

пользователя [8]. Таким образом, клиентские приложения могут работать с *ZeeRex*-БД и для этого нет необходимости вводить дополнительные расширения Z39.50, так как используется традиционный сервис Z39.50 поиска и извлечения.

### Список источников

1. **Открытая** электронная библиотека г. Томск [Электронный ресурс] / Дмитрий Владимирович Дробышевский, 2002. Режим доступа: <http://www.oel.tomsk.ru>, свободный – Заглавие с экрана.
2. **Z39.50 Part 7** [Explain] [Электронный ресурс] / Z39.50 Maintenance Agency, 1955. Режим доступа: <http://lcweb.loc.gov/z3950/agency/markup/07.html>, свободный – Заглавие с экрана.
3. **Explain Lite** [Электронный ресурс] / Rob Bull, 2001. Режим доступа: <http://www.one-2.org/exp-lite/>, свободный – Заглавие с экрана.
4. **ZeeRex: The Explainable «Explain» Service** [Электронный ресурс] / Mike Raylor, 2002. Режим доступа: <http://explain.Z39.50.org/>, свободный – Заглавие с экрана.
5. **ANSI/NISO Z39.50–1995 Information Retrieval (Z39.50): Application Service and Protocol Specification / Maintenance Agency. Official Text Z39.50 – 1995, July 1995.**
6. **ZeeRex: An Overview of ZeeRex** [Электронный ресурс] / Mike Taylor, 2002. Режим доступа: <http://explain.Z39.50.org/overview/index.html#2.2.1>, свободный – Заглавие с экрана.
7. **ONE-2 Project** [Электронный ресурс] / 2001. Режим доступа: <http://www.one-2.org/>, свободный – Заглавие с экрана.
8. **Колобов О.С., Пачуев К.Е., Татарский Ф.Е.** Технические особенности организации Томского консорциума «Открытая электронная библиотека» // «Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества» (Крым–2003): Сб. тр. Десятой междунар. конф., Судак, Автономная Республика Крым, Украина, 2003.