

[Комиссия Европейского Сообщества](#)



[Министерство культуры Российской Федерации](#)



[Министерство промышленности, науки и технологий Российской Федерации](#)



[Российский Фонд Фундаментальных Исследований](#)

[Институт "Открытое Общество" - Россия](#)



Ежегодная международная конференция
"Электронные изображения и визуальные искусства"

EVA 2000 MOSCOW
Electronic Imaging & the Visual Arts

... EVA 2000 Москва ...

Государственная Третьяковская галерея, 30 октября - 3 ноября 2000
Конференция, выставка мультимедиа, рабочие группы, культурная программа

Организаторы конференции:



[Фирма ВАЗАРИ, Великобритания](#)



[Центр по проблемам информатизации сферы культуры Минкультуры России \(Центр ПИК\)](#)



[Государственная Третьяковская галерея](#)

Техническая поддержка:

THE DOCUMENT COMPANY
XEROX

ACTIVISION
ПРОДУКЦИЯ ДЛЯ ВКЛАДОВ И ПРЕЗЕНТАЦИЙ

Тезаурусы и классификационные схемы в распределенных информационных системах: проблемы и решения

Жижимов О.Л., Мазов Н.А.

Объединенный Институт Геологии, Геофизики и Минералогии

Сибирского Отделения Российской Академии Наук

Адрес: 630090, Новосибирск-90, пр.Коптюга 3

Факс: (383-2) 33-27-92

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

Internet: <http://geolibr.uiggm.nsc.ru>

В настоящее время, как в России, так и за рубежом существует значительное количество различных тезаурусов и классификационных схем, используемых в информационной музейной практике. Огромная работа по классификации музейных предметов проделана в музейном сообществе России. Авторам известна острота настоящей проблемы из материалов конференций, посвященных информатизации музеев, прошедших в России в последнее время [1].

Как правило, во всех организациях, в которых используются ресурсы этого класса, они оформлены в машиночитаемые базы данных, поддерживаемые собственным программным обеспечением, и, как следствие этому, – затруднена интеграция и совместное использование таких баз данных в распределенных информационных системах.

Насколько известно авторам настоящего доклада, ряд организаций, располагающих такими базами данных (далее БДТК – базы данных тезаурусов и классификаторов), в настоящее время ведут работы над тем, чтобы обеспечить унифицированный сетевой доступ к ним.

Цель этих работ - обеспечение возможности не только просмотра тезаурусов и классификаторов в удобном интерфейсе, но и активное использование их при поиске соответствующей информации в специализированных базах данных (далее БД), в том числе и в электронных коллекциях музеев.

Достаточно сказать, что для интеграции ведения, поиска и отображения музейной информации в настоящее время наметился определенный сдвиг в области форматов, стандартных протоколов передачи и схем данных, т.е. правила, которые регламентируют работу с музейной информацией в распределенных информационных системах по протоколу Z39.50, сформулированные в документах CИМІ (Consortium for the Computer Interchange of Museum Information) [2].

Наряду с вышеизложенным в области тезаурусов и классификационных схем, работе в распределенных системах также уделяется большое внимание, о чем свидетельствует факт появления в середине 1999 года бета-версии схемы данных для работы с тезаурусами по протоколу Z39.50 [3].

Действительно, с развитием технологий построения больших распределенных информационных систем, включающих в себя множество различных баз данных, достаточно актуальным становится вопрос поиска информации в БД с использованием тезаурусов и классификационных схем. Более того, в распределенной информационной системе логично обеспечить доступ к БДТК в той же самой технологии, в которой осуществляется доступ к БД, т.е. в технологии “клиент-сервер” с использованием единого протокола Z39.50[4].

Актуальность этого вопроса для информационно-библиотечного и музейного сообщества России, а также предшествующий опыт работы авторов доклада по разработке программного обеспечения доступа к БД по протоколу Z39.50 [5,6], позволил применить протокол Z39.50 для работы с тезаурусами и рубрикаторами в рамках распределенной информационной системы ОИГГиМ СО РАН.

Исходя из общей идеологии Z39.50, доступ к любой базе данных, в том числе и к БДТК, должен осуществляться через единую стандартную схему данных, на которую должны быть корректно отображены все частные структуры БДТК. Проект такой схемы сегодня уже существует - Zthes (OID 1.2.840.10003.13.8) [1] и она активно обсуждается.

Авторами настоящего доклада была предпринята попытка использования схемы данных Zthes для предоставления доступа к БДТК по протоколу Z39.50. Ниже изложены основные результаты этой работы.

В таблице 1 представлена схема Zthes, которая согласно [3] определяет абстрактную структуру записи БДТК в иерархической схеме (колонок Tag Path определяет полный путь в терминах tagsetM (1), tagsetG (2) и tagsetZthes (4)):

Tag Path	Occurrence	Element
(1,14)	M	termId
(2,1)	M	termName
(4,1)	O	termQualifier
(4,2)	O	termType
(2,17)	O	termNote
(2,20)	O	termLanguage
(1,15)	O	termCreatedDate
(1,27)	O	termCreatedBy
(1,16)	O	termModifiedDat
(1,28)	O	termModifiedBy
(4,4)	O, R	postings
(4,4)(2,36)	M	sourceDb
(4,4)(4,5)	O	fieldName
(4,4)(4,6)	M	hitCount
(2,30)	O, R	relation

(2,30)(4,3)	M	relationType
(2,30)(2,36)	O	sourceDb
(2,30)(1,14)	M	termId
(2,30)(2,1)	M	termName
(2,30)(4,1)	O	termQualifier
(2,30)(4,2)	O	termType
(2,30)(2,20)	O	termLanguage

Табл.1. Абстрактная структура записи Zthes. Элементы: М - обязательный, О - факультативный, R – повторяемый.

Каждая запись БДТК должна быть отображена на эту структуру. В частности, одна из статей (**bromine**) тезауруса по наукам о Земле выглядит следующим образом (представление XML):

```

<zthes>
<termId> 549BC38E </termId>
<termName> bromine </termName>
<termType> TT </termType>
<termNote> Chemical element. Use bromine deposits for bromine as a
commodity
</termNote >
<relation>
<relationType> UF </relationType>
<termName> Br </termName>
<termId> F90CF05F </termId>
</relation>
<relation>
<relationType> BT </relationType>
<termName> halogens </termName>
<termId> CCEACFE6 </termId>
</relation>
<relation>
<relationType> NT </relationType>
<termName> bromide ion </termName>
<termId> F8BB2A67 </termId>
</relation>
<relation>
<relationType> RT </relationType>
<termName> brines </termName>
<termId> AD0A65E3 </termId>
</relation>
<relation>
<relationType> RT </relationType>
<termName> bromine deposits </termName>

```

```
<termId> 4510AE36 </termId>
</relation>
</zthes>
```

Аналогичное отображение можно осуществить и для записей БД классификационных схем.

Для апробации реальной работы были выбраны следующие БДТК, хранящиеся в СУБД CDS/ISIS (в скобках – имя базы данных):

- Тезаурус по Наукам о Земле(grefth)
- Рубрикатор ГРНТИ (версия ГПНТБ России)(grnti)

Записи этих БДТК были отображены на схему Zthes штатными средствами сервера ZooPARK (v2.42) [7] и Z-ISIS - провайдера данных CDS/ISIS. С результатом работы можно ознакомиться через Интернет по протоколу Z39.50 на сервере geolibr.uiggm.nsc.ru:210 (имена баз данных указаны выше). Кроме того, доступ к этим БДТК может быть осуществлен через шлюз Z39.50 (<http://geolibr.uiggm.nsc.ru/zgwn/>) с удобным графическим интерфейсом для навигации.

Однако следует отметить, что сам по себе сетевой доступ к БДТК хотя и предоставляет интерес, но, на наш взгляд, не несет особого смысла без возможности одновременного выхода в поисковую систему по БД. Иными словами, просматривая статьи тезауруса или классификационной схемы, хотелось бы проводить одновременный поиск в БД по соответствующим ключевым словам или кодам рубрик. Здесь Z39.50, ввиду стандартизации поискового механизма, дает уникальную возможность подключать к параллельному поиску совершенно различные БД. На вышеуказанном шлюзе продемонстрирована возможность реализации этого механизма на примере рубрикатора ГРНТИ, коды которого сегодня присутствуют во многих БД.

Результатом проведения вышеописанных работ явилось не только появление вполне работоспособного интерфейса доступа к БДТК с их интеграцией с БД (см. ссылку на шлюз выше), но и накопление опыта, суть которого может быть выражена в следующих тезисах:

- Применение протокола Z39.50 для доступа к БДТК дает богатые возможности для построения распределенных информационных систем – интеграция БДТК и БД с обеспечением единого сетевого доступа по стандартным глобальным схемам.
- Использование стандартных схем данных позволяет скрыть частные различия структур различных БДТК и обеспечить единый интерфейс без регенерации последних.

- В распределенных системах можно и нужно организовывать специализированные серверы для хранения БДТК с предоставлением к ним доступа по Z39.50 (чем меньше копий данных, тем проще поддерживать их синхронность).
- При построении конкретных БД следует как можно шире использовать привязку записей к различным тезаурусам и классификационным схемам. Отсутствие этой информации в БД исключает последнюю из единого информационного пространства в распределенной информационной системе.
- Глобальная схема Zthes требует расширения для адекватного отображения информации БДТК.

Последний тезис связан с тем фактом, что в схеме Zthes отсутствует корректная ссылка на базу данных БДТК, из которой выбирается терм. Поле sourceDB, предназначенное для этой цели, содержит лишь имя БД, но не содержит имени сервера и порта. Однако в распределенной системе указание только имени БД недостаточно для ее однозначной идентификации. Расширение схемы позволит снять это ограничение при построении распределенной системы взаимосвязанных БДТК. Существует еще ряд замечаний к схеме Zthes. Поскольку в настоящее время происходит активное обсуждение готовящегося стандарта схемы Zthes, в котором авторы принимают участие, есть надежда, что в следующих версиях схемы эти ограничения будут сняты.

В заключение отметим, что изложенный в настоящем докладе подход для доступа к БДТК позволяет также снять ряд ограничений по поиску информации в БД, доступных по протоколу Z39.50, с которыми иногда сталкиваются пользователи.

Литература

1. <http://www.museum.ru/defruss.htm>
2. The CIMI Profile Release 1.0H A Z39.50 Profile for Cultural Heritage Information <http://www.cimi.org/documents/HarmonizedProfile/HarmonProfile1.htm>
3. Mike Taylor. Zthes: A Z39.50 Profile for Thesaurus Navigation. Version 0.3b. <http://lcweb.loc.gov/z3950/agency/profiles/zthes-03>
4. 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.
5. Жижимов О.Л., Мазов Н.А. Применение протокола Z39.50 в распределенной информационной системе Сибирского отделения РАН // Библиотечно-информационные ресурсы в науке, образовании, культуре и бизнесе (10-16 октября 1999 г., г. Самарканд): Материалы конф., Самарканд, 1999, С. 118-125
6. Жижимов О.Л., Мазов Н.А. Состояние и перспективы использования протокола Z39.50 в информационном сообществе России // Информационное общество, 2000, N 2, С. 39-43
7. ZooPARK модульный сервер Z39.50. Версия 2.42. ОИГГиМ СО РАН. <http://geolibr.uiggm.nsc.ru/doklads/Z-docs/ZooPARK.doc>

Жижимов Олег Львович, к.ф.-м.н., заведующий сектором Информационных Технологий, ОИГГиМ СО РАН
Тел. служ. (383-2) 33-20-05
E-Mail: zhizhim@uiggm.nsc.ru

Профессиональная деятельность:

- **(1998-2000)**

Разработка программного обеспечения Z39.50:

- серверное: - сервер ZooPARK,
- клиентское - Z39.5-0-WWW Gateway Z-CGI,
- Графический клиент Z39.50 для Win32.
- (C, C++, MFC, CGI).

Объединенный Институт Геологии, Геофизики и Минералогии Сибирского Отделения АН.

- **(1997-2000)**

Программное обеспечение для Internet и библиотек (C, CGI, Java, ASP).
Объединенный Институт Геологии, Геофизики и Минералогии Сибирского Отделения АН.

- **(1996 - 1999)**

Разработка информационных систем на основе технологии клиент-сервер.
(MsSQL6.5, FoxPro5.0)
Windows NT server, 30 workstations.
Объединенный Институт Геологии, Геофизики и Минералогии Сибирского Отделения АН.

- **(1993 - 1997)**

Построение и управление ЛВС (4 subnet, class C, 500 hosts)
Объединенный Институт Геологии, Геофизики и Минералогии Сибирского Отделения АН.

- **(1983 - 1993)**

Теоретическое и численное моделирование геофизических процессов.
Объединенный Институт Геологии, Геофизики и Минералогии Сибирского Отделения АН.

- **(1979 - 1983)**

Теоретические исследования в области атомной физики.
Объединенный Институт Геологии, Геофизики и Минералогии Сибирского
Отделения АН.

Образование:

- Киргизский Государственный Университет, Фрунзе (1979), теоретическая и математическая физика.
- Новосибирский Государственный Университет (1982), очная аспирантура, теоретическая и математическая физика.
- Диссертация к.ф.-м.н. (ИЯФ СО РАН), апрель 1983:
Р-нечетное дальное действие атомов и ионов

Мазов Николай Алексеевич, заведующий Информационно-библиотечным
Центром, ОИГГиМ СО РАН
Тел. служ. (383-2) 33-22-16
E-Mail: mazov@uiggm.nsc.ru

Профессиональная деятельность:

- **(1998-2000)**
- Администрирование библиографических баз данных.
- Разработка программного обеспечения для СУБД CDS/ISIS под управлением из WWW и Z39.50

Объединенный Институт Геологии, Геофизики и Минералогии Сибирского
Отделения АН.

- **(1993-1997)**
- Администрирование библиографических баз данных.
- Разработка программного обеспечения для переноса БД с машин серии ЕС в РС.
- Разработка программного обеспечения для работы с БД на РС в локальной сети ОИГГиМ СО РАН.
- Разработка программного обеспечения для библиотеки ОИГГиМ СО РАН.

Объединенный Институт Геологии, Геофизики и Минералогии Сибирского
Отделения АН.

- **(1987 - 1993)**

Разработка информационных систем на машинах ЕС на основе ППП STAIRS для
работы с БД ВИНТИ, ВИМИ.
Государственная публичная научно-техническая библиотека Сибирского Отделения
АН.

- (1980 - 1987)

Разработка и построение АСНТИ ВНИИМБ.

Всесоюзный научно-исследовательский институт молекулярной биологии.

Образование:

Новосибирский Государственный Университет (1980), механико-математический факультет, вычислительная математика.