

# ДОСТУП К ТЕЗАУРУСАМ И КЛАССИФИКАЦИОННЫМ СХЕМАМ ИЗ ИНТЕРНЕТ ПО ПРОТОКОЛУ Z39.50

О. Л. ЖИЖИМОВ, Н. А. МАЗОВ

*Объединенный институт геологии, геофизики и минералогии СО РАН*

*Новосибирск, Россия*

e-mail: zhizhim@uiggm.nsc.ru

The use of the data scheme *Zthes* is described providing the access to the databases of thesauruses and classifiers according to protocol Z39.50. An efficient interface of the access to data bases of thesauruses and classifiers integrating them with informational databases including catalogues of electronic libraries has been developed.

В настоящее время как в России, так и за рубежом в информационной практике применяются различные тезаурусы и классификационные схемы. Например, в ЦНСХБ имеется отраслевой тезаурус по сельскому хозяйству — русско-английская версия известного медицинского тезауруса MeSH, используемого в ГЦНМБ; Государственный рубрикатор научно-технической информации (ГРНТИ) используется так или иначе практически во всех информационных органах и библиотеках России. Как правило, ресурсы этого класса оформлены в машиночитаемые базы данных, которые поддерживаются собственным программным обеспечением, поэтому интеграция таких баз данных и совместное их использование в распределенных информационных системах затруднены.

Насколько известно авторам, ряд организаций, располагающих базами данных тезаурусов и классификаторов (БДТК), в настоящее время ведут работы по обеспечению унифицированного сетевого доступа к ним, который бы позволял не только осуществлять просмотр тезаурусов и классификаторов в удобном интерфейсе, но и активно использовать их при поиске соответствующей информации в базах данных (БД), в том числе и в электронных каталогах библиотек.

Если в интеграции ведения, поиска и отображения библиографической информации в настоящее время наметился определенный сдвиг в области форматов, стандартных протоколов передачи и схем данных, то в области тезаурусов и классификационных схем к этому только подходят, о чем свидетельствует факт появления в середине 1999 г. бэта-версии схемы данных *Zthes* для работы с тезаурусом по протоколу Z39.50 [1].

Действительно, с развитием технологий построения больших распределенных информационных систем, включающих множество различных баз данных, актуальным становится поиск 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” по единому протоколу Z39.50 [2].

Опыт авторов по разработке программного обеспечения доступа к БД по протоколу Z39.50 [3, 4] позволил применить его для работы с тезаурусом по наукам о Земле и рубрикатом баз данных ВИНТИ в рамках информационной системы ОИГГиМ СО РАН.

Исходя из общей идеологии Z39.50, доступ к любой базе данных, в том числе и к БДТК, осуществляется 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 он активно обсуждается. Авторами предпринята попытка использования этой схемы для осуществления доступа к БДТК по протоколу Z39.50.

Ниже представлена схема Zthes, которая определяет абстрактную структуру записи БДТК в иерархической схеме (колонка Tag Path содержит полный путь в терминах tagsetM (1), tagsetG (2) и tagsetZthes (4); М — обязательный, О — факультативный, R — повторяемый элементы).

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	termModifiedDate
(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

Каждая запись БДТК должна быть отображена на эту структуру. В частности, одна из статей тезауруса по наукам о Земле (bromine) 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> ADOA65E3 </termId>
    </relation>
    <relation>
        <relationType> RT </relationType>
        <termName> bromine deposits </termName>
        <termId> 4510AE36 </termId>
    </relation>
</zthes>

```

Аналогично можно представить записи БД классификационных схем.

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

- тезаурус по наукам о Земле (grefth);
- тезаурус по сельскому хозяйству ЦНСХБ (agrith);
- рубрикатор ГРНТИ (версия ГПНТБ России) (grnti).

Последние две БДТК находятся в настоящее время в промышленной эксплуатации и были предоставлены руководством соответствующих библиотек авторам для тестирования. Записи этих БДТК были отображены на схему Zthes штатными средствами сервера ZooPARK (v2.20) [5] 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. На рис. 1 и 2 представлены интерфейсы пользователя при работе с тезаурусом и рубрикатором.

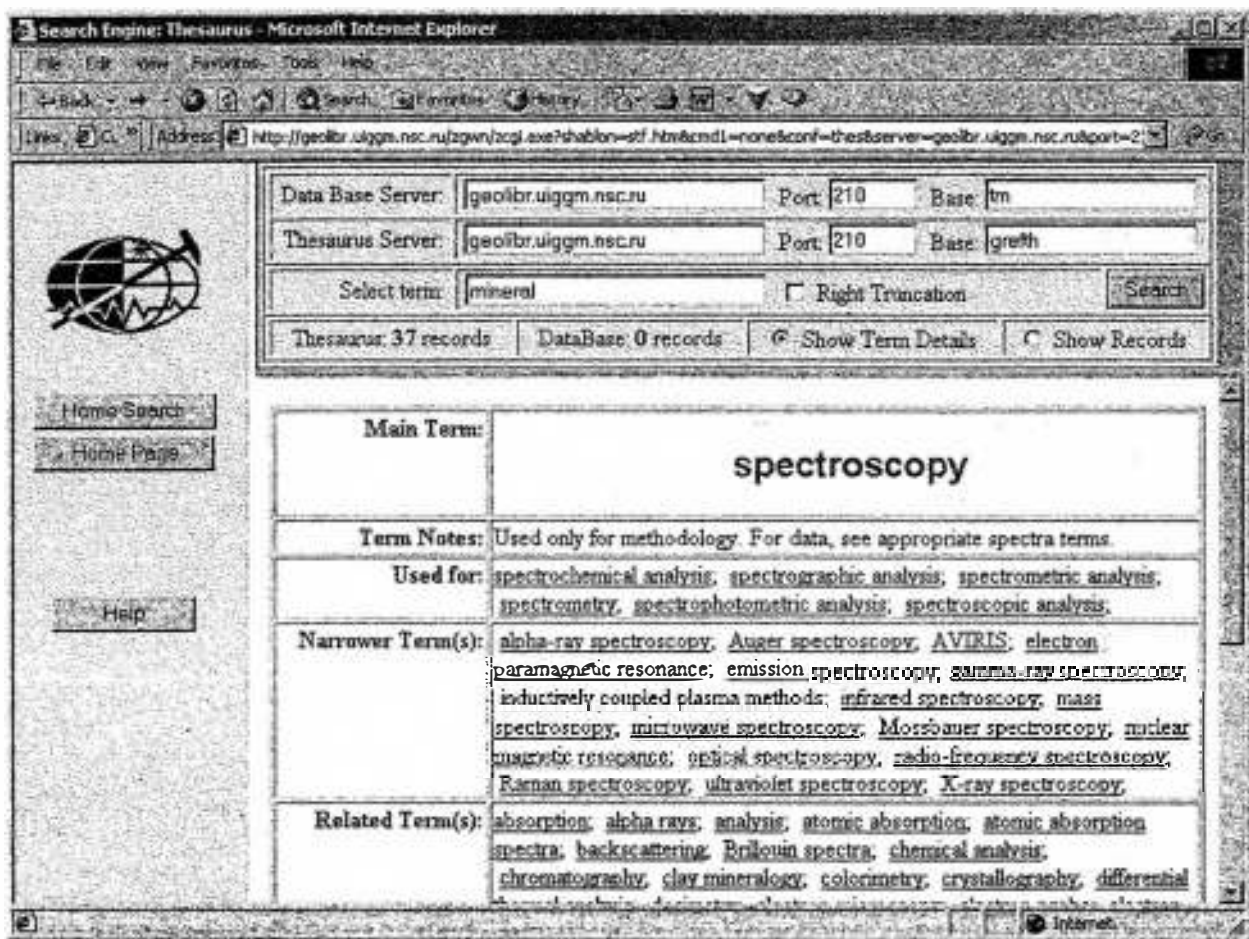


Рис. 1. Интерфейс пользователя при навигации по тезаурусу.

Следует отметить, что сам по себе сетевой доступ к БДТК хотя и представляет интерес, но, на наш взгляд, не несет особого смысла без возможности одновременного выхода в поисковую систему по БД. Иными словами, при просмотре статьи тезауруса или классификационной схемы логично иметь возможность проведения одновременного поиска в БД по соответствующим ключевым словам или кодам рубрик. Именно в этом случае протокол Z39.50 ввиду стандартности 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 (каталоги ГПНТБ России, ГПНТБ СО РАН, БД ВИНТИ и др.).

Результатами проведенных работ явились не только создание 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 следующем:

- применение протокола Z39.50 для доступа к БДТК дает богатые возможности при построении распределенных информационных систем — интеграцию БДТК и БД с обеспечением единого сетевого доступа по стандартным глобальным схемам;
- использование стандартных схем данных позволяет скрыть частные отличия структур различных БДТК и обеспечить единый интерфейс без регенерации последних;
- в распределенных информационных системах можно и нужно организовывать спе-

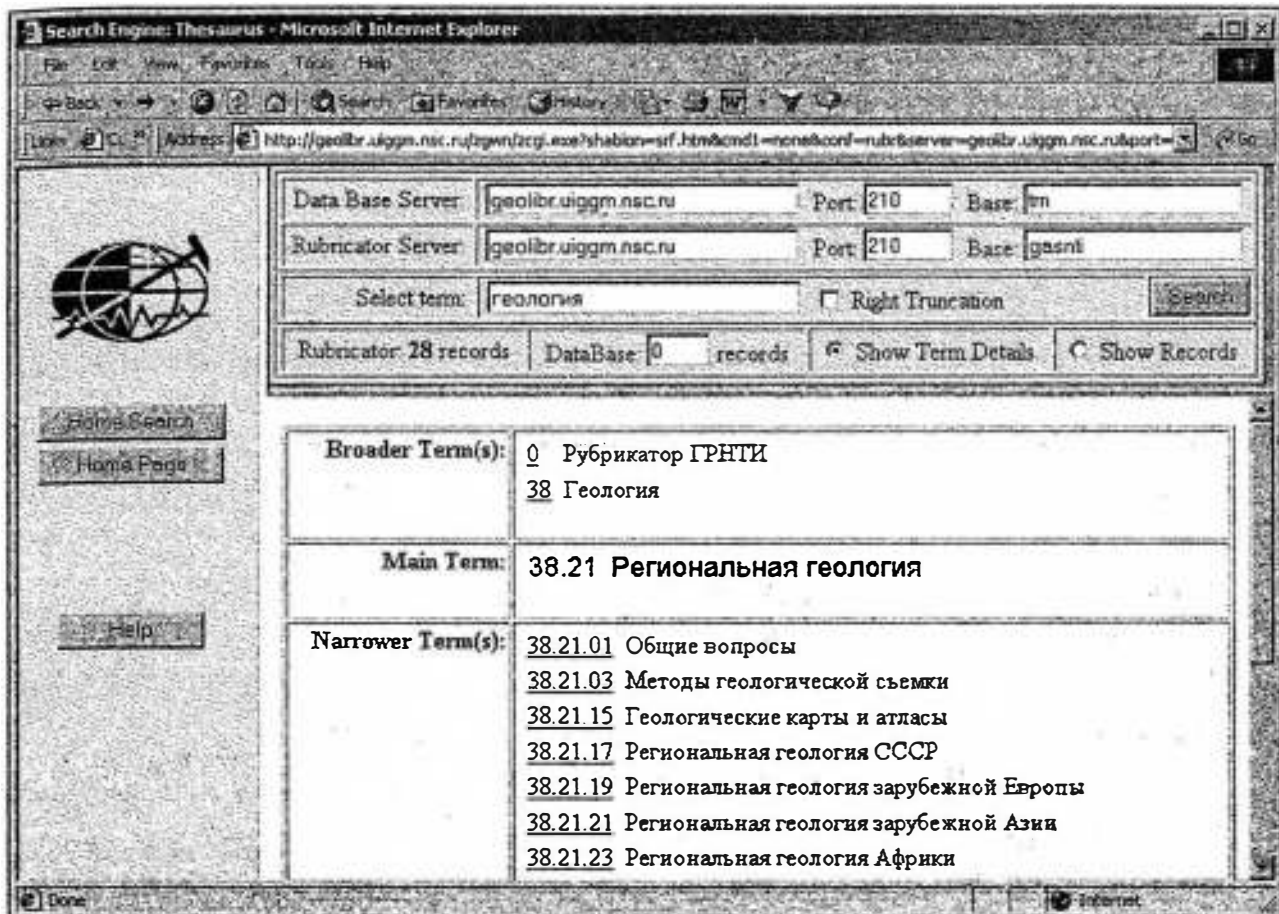


Рис. 2. Интерфейс пользователя при навигации по рубрикатору.

циализированные серверы для хранения БДТК с предоставлением к ним доступа по Z39.50 (чем меньше копий данных, тем проще поддерживать их синхронность);

— при построении библиографических БД следует как можно шире использовать привязку записей к различным тезаурусам и классификационным схемам. Отсутствие этой информации в БД исключает последние из единого информационного пространства в распределенной информационной системе;

— глобальная схема Zthes требует расширения для более адекватного отображения информации БДТК.

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

В заключение отметим, что изложенный в настоящей работе подход для доступа к БДТК позволяет также снять ряд ограничений по поиску информации в БД по протоколу

Z39.50, с которыми иногда сталкиваются пользователи (пользователь может и не знать расшифровок конкретных рубрикационных шифров).

Насколько известно авторам, настоящая разработка доступа к БДТК по протоколу Z39.50 является в России уникальной, хотя подобные работы в мире в последнее время проводятся [6].

## Список литературы

- [1] TAYLOR M. Zthes: A Z39.50 Profile for Thesaurus Navigation. Version 0.3b.  
<http://lcweb.loc.gov/z3950/agency/profiles/zthes-03>.
- [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] Жижимов О. Л., Мазов Н. А., Болванов А. Ю. Опыт построения распределенной информационной системы на базе протокола Z39.50 // Мат. 6-й Междунар. конф. "Крым-99". 1999. Т. 1. С. 249-252.
- [4] Мазов Н. А., Жижимов О. Л. Интеграция Z39.50 и CDS/ISIS: состояние и перспективы развития // Мат. 6-й Междунар. конф. "Крым-99". 1999. Т. 2. С. 249-251.
- [5] ZOOPARK модульный сервер Z39.50. Версия 2.20. ОИГГиМ СО РАН.  
<http://geolibr.uiggm.nsc.ru/doklads/Z-docs/ZooPARK.doc>
- [6] INDEX Data's thesaurus WWW-Z39.50 gateway.  
<http://muffin.indexdata.dk/zthes/tbrowse.zap>

*Поступила в редакцию 18 апреля 2000 г.*