

СИНТЕЗ ТЕХНОЛОГИЙ НА ПРИМЕРЕ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПО ГАЗОГИДРАТАМ

О. Л. Жижимов, В. С. Коджесян, Н. А. Мазов

Объединенный институт геологии, геофизики и минералогии СО РАН

Новосибирск, Россия

e-mail: zhizhim@uiggm.nsc.ru

An informational system on gas hydrates is described developed within an international project. This system integrates heterogeneous information and complies with the international standard specifications, in particular with Z39.50.

Цели и задачи проекта

Цель описываемого в этой работе проекта — создание специализированной информационной системы по газогидратам — уникальным объектам, которые являются предметом исследований ученых разных направлений. Эта информационная система должна интегрировать разнородные базы данных со структурированной информацией о газогидратах и различные неструктурированные данные в виде текстов, схем, фотографий и т. п. Отбор информации, ее систематизация и аналитическая переработка составляли часть проекта, за которую отвечали специалисты соответствующих областей знаний. В настоящей работе обсуждается часть проекта, которая касается собственно информационной системы, технологии ее построения и интерфейсов доступа к ней.

Газогидраты представляют собой нечто среднее между твердым веществом и сильно сжатым, замороженным газом. В мире уже открыто много месторождений газогидратов как на суше, так и в Мировом океане на большой глубине. Некоторые газогидраты обладают интересными физико-химическими свойствами. Например, будучи извлеченными с большой глубины, они буквально тают в руках исследователей, испаряются, за несколько минут превращаясь в газ. Кроме чисто научного интереса газогидраты привлекают и своими экономическими перспективами. По оценкам специалистов, в XXI веке газогидраты могут служить основным сырьем для химической промышленности и быть эффективным источником энергии подобно нефти и природному газу. Несмотря на большой интерес к этим уникальным геологическим объектам, информация о них весьма скудна, раздроблена по разным источникам.

Настоящим проектом предполагалось улучшить ситуацию в области систематизации информации о газогидратах.

Технология

Для построения распределенной информационной системы по газогидратам были сформулированы требования, которым она должна удовлетворять:

- обеспечивать хранение данных различных типов: текстов, карт, фотографий, библиографических сведений о научных работах, ссылок на другие ресурсы;
- допускать хранение метаданных, т. е. данных об этих данных;
- допускать пополнение и изменение данных и метаданных;
- поддерживать целостность базы метаданных и их соответствие реальным данным;
- обеспечивать поиск по различным критериям и их комбинации в пределах информации, описанной в базе метаданных;
- поддерживать разнообразное представление информации, включая некоторые специальные формы визуализации, например просмотр карт;
- быть доступной из Интернет через WWW;
- быть распределенной прозрачно для пользователя;
- соответствовать существующим мировым и отраслевым стандартам;
- быть аппаратно независимой;
- быть легко управляемой (каждый сервер информационной системы должен допускать удаленное администрирование);
- быть достаточно живучей (отключение нескольких серверов не должно приводить к краху информационной системы).

Принимая во внимание, что удовлетворить описанным выше требованиям практически невозможно в рамках одной технологии, изначально был выбран подход, основанный на интеграции различных технологий, каждая из которых эффективна в своей части.

В качестве основы информационной системы выбран протокол Z39.50 (ISO 23950) [1]. Его использование позволяет не только остаться в рамках международных стандартов доступа к базам данных, но и достаточно эффективно построить аппаратно независимую распределенную информационную систему как минимум на уровне метаданных с единой системой запросов и стандартных форматов представления информации.

В отличие от уровня метаданных для уровня данных выбрано несколько параллельных возможностей доступа: Z39.50, WWW, ftp, вообще могут использоваться любые другие, определяемые по URL.

В качестве точки доступа к информационной системе со стороны WWW решено использовать шлюз Z39.50-WWW, выполняющий функции клиента Z39.50. Это обеспечивало достаточно простой и доступный механизм не только для построения пользовательских интерфейсов в окружении привычного Web-браузера, но и для создания существенно продвинутых решений на основе Z39.50. Другой принятый способ доступа к информационной системе — применение специализированных клиентов Z39.50. Сегодня в мире насчитывается несколько готовых решений, но все они ориентированы в первую очередь на доступ к библиографической информации, в то время как существует необходимость отображения разнородной информации, в том числе и графической. Тем не менее информационная система допускает использование специализированных клиентов Z39.50. Общая схема информационной системы представлена на рис. 1.

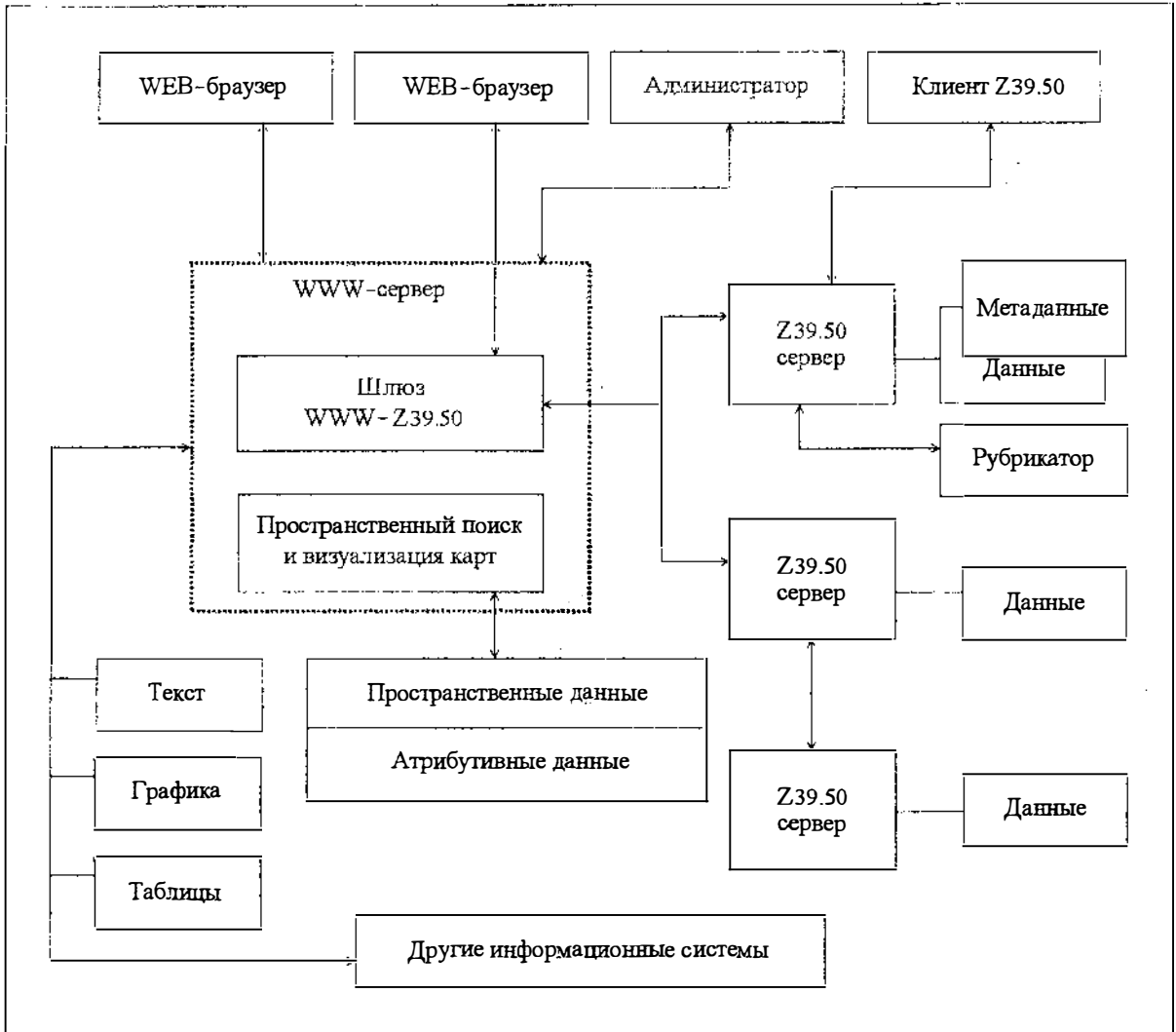


Рис. 1. Схема информационной системы.

Реализация

В качестве сервера Z39.50 использован сервер ZooPARK, разработанный в ОИГГиМ СО РАН [2]. Этот сервер может быть инсталлирован на любой серверной платформе. Кроме того, он показал достаточную устойчивость в эксплуатации.

В качестве WWW-сервера использовался Microsoft IIS 4.0 для Windows NT, на котором выполнялась, помимо всего прочего, задача шлюза Z39.50-WWW, а именно Z-CGI, разработанная также в ОИГГиМ СО РАН.

Поскольку в основу всей информационной системы положено требование соответствия мировым стандартам, функциональность построенной системы никак не связана со способами хранения данных. На уровне метаданных, доступ к которым осуществляется по протоколу Z39.50, информация всегда отображается во внешнее представление в соответствии со схемой GILS (Global Information Locator Service) [3], поскольку она является стандартной и отображение на нее данных из любой СУБД не представляет особой сложности. На первом этапе (опытной эксплуатации) для базы метаданных был выбран вариант хранения

данных в виде обычных текстовых XML-файлов с теговой структурой GILS. Индексирование производилось при помощи свободно распространяемого программного обеспечения Zebra от компании IndexData [4], создающего базу данных, совместимую по структуре с одним из провайдеров данных сервера ZooPARK. В дальнейшем, при увеличении объема хранимой информации, планируется применить реляционную СУБД с соответствующим провайдером данных для сервера ZooPARK. Ниже приведен пример записи базы метаданных в формате XML.

```

<gils>
<controlIdentifier> mgashyd-1020 </controlIdentifier>
<title> Gas composition and isotopic composition
      of gas hydrate from offshore Sakhalin Island,
      Okhotsk Sea
</title>
<originator>
  <name> Ginsburg, G.D. </name>
</originator>
<originator>
  <name> Soloviev, V.A. </name>
</originator>
<publicationDate> 1993 </publicationDate>
<languageOfResource> eng </languageOfResource>
<abstract> Gas hydrate recovered during R/V
  "Geolog Pyotr Antropov" cruise offshore Sakhalin
  Island in Okhotsk Sea in 1991. See table9 in url:
  Analysed gas hydrate recovered from Site 91-02-40
  and Site 91-02-42. Parameters described: gas composition
  (methane, propane,... CO2, N2, He content) and isotopic
  composition (d(13)C-CH(4) PDB, dD-CH(4) SMOW,
  d(13)C-CO(2) PDB)
</abstract>
<controlledSubjectIndex>
  <subjectTermsControlled>
    <controlledTerm> 1.1.2.3 </controlledTerm>
  </subjectTermsControlled>
</controlledSubjectIndex>
<subjectTermsUncontrolled>
<uncontrolledTerm> gas hydrate </uncontrolledTerm>
<uncontrolledTerm> gas composition </uncontrolledTerm>
</subjectTermsUncontrolled>
<spatialDomain>
  <boundingCoordinates>
    <westBoundingCoordinate> 144 </westBoundingCoordinate>
    <eastBoundingCoordinate> 145 </eastBoundingCoordinate>
    <northBoundingCoordinate> 55 /northBoundingCoordinate>
    <southBoundingCoordinate> 54 /southBoundingCoordinate>
  </boundingCoordinates>
<place><placeKeyword> Okhotsk Sea </placeKeyword></place>

```

```

</spatialDomain>
<sourcesOfData> Ginsburg G.D. and Soloviev V.A. Submarine
  Gas Hydrates. VNIIOkeangeologia, St-Petersburg,
  p.215, 1998.
</sourcesOfData>
<originalControlIdentifier> 1020 </originalControlIdentifier>
<url> data/ginssolov98_t9.htm </url>
<recordSource>
  <name> Poort J. </name>
  <organization> RMCA (Tervuren, Belgium) </organization>
</recordSource>
<dateOfLastModification> 19991206 </dateOfLastModification>
<recordReviewDate> 19991230 </recordReviewDate>
</gils>

```

Следующий за метаданными информационный слой системы — собственно данные. Эти данные могут быть любыми: текст, графика, мультимедиа и даже отдельные базы данных с собственными интерфейсами доступа. В описываемой информационной системе доступ к этим данным осуществляется через URL, хранящийся в записи базы метаданных. Это позволяет применить механизм гипертекстовых ссылок при просмотре метаданных, предоставляя простой выход на сами данные. На рис. 2 показан интерфейс пользователя при поиске информации в базе метаданных.

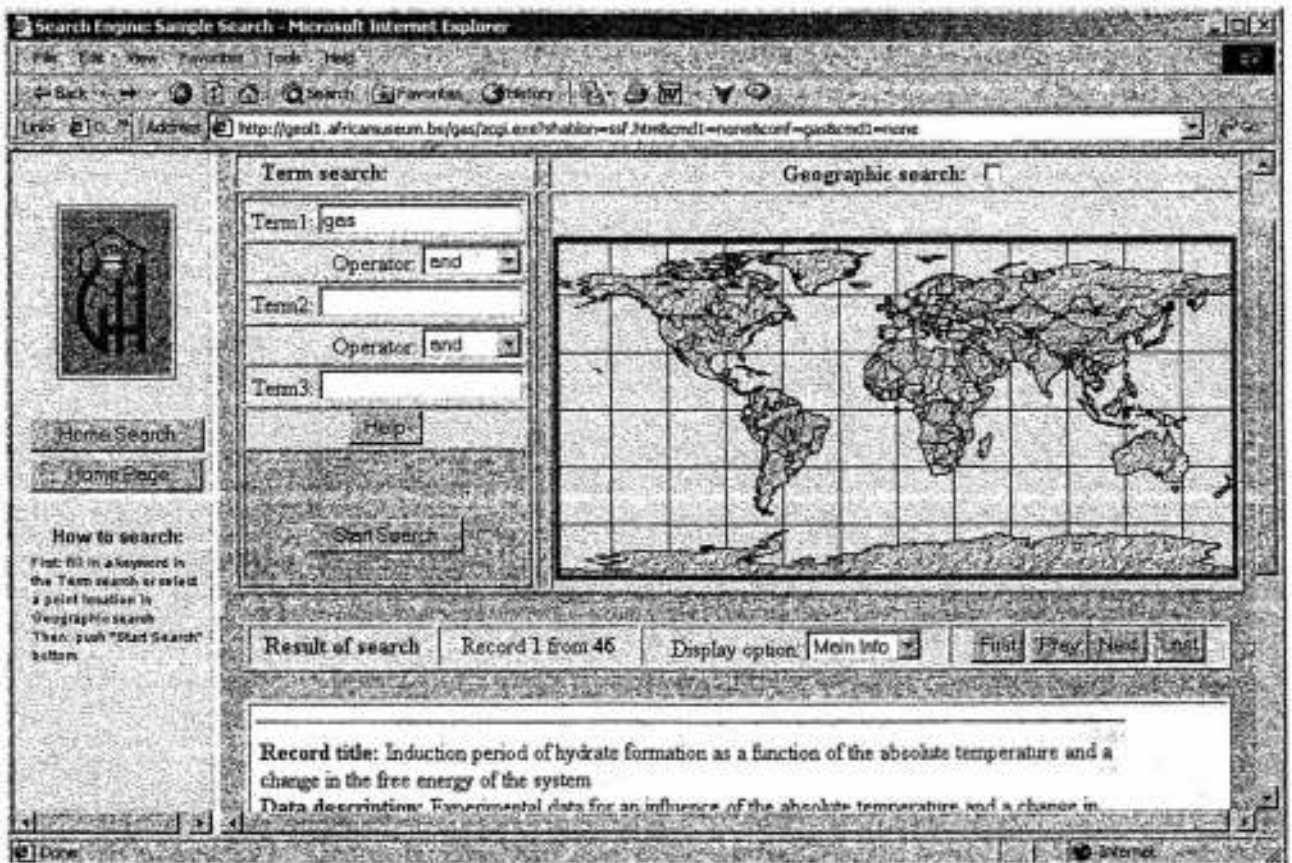


Рис. 2. Интерфейс пользователя при поиске.

Другой информационный слой, который интегрирован с метаданными, — географическая информация о положении и размерах описываемых месторождений газогидратов. Для четкой идентификации объекта и его визуализации на карте необходимо использовать данные геоинформационных систем (ГИС). Это могут быть контуры материков, океаны, реки, озера, а также специфическая информация, такая, как геологические складки, формации с их пространственными данными — размерами, площадями, мощностями. Актуальна и атрибутивная информация об объектах — названия, геологический возраст и т. п. Специфика ГИС-данных — большие объемы информации. Так, например, хранение только контура одного геологического объекта может занимать несколько мегабайт. Существуют специализированные системы и форматы для хранения такого рода информации. В описываемой системе использован один из наиболее популярных и стандартных форматов хранения — SHP-файлы, атрибутивная же информация размещена в файлах формата DBF. SHP/DBF — очень распространенные форматы ГИС-данных, любая ГИС может их экспортировать и импортировать. В нашей системе доступ к этой информации осуществляется WEB-сервером и шлюзом Z39.50-WWW через специальный CGI-модуль. Интеграция поисковой подсистемы Z39.50, элементов GILS и ГИС-информации позволила построить пользовательские интерфейсы с реализацией визуального поиска по масштабируемой карте и отображением на ней записей из баз метаданных. На рис. 3 показан

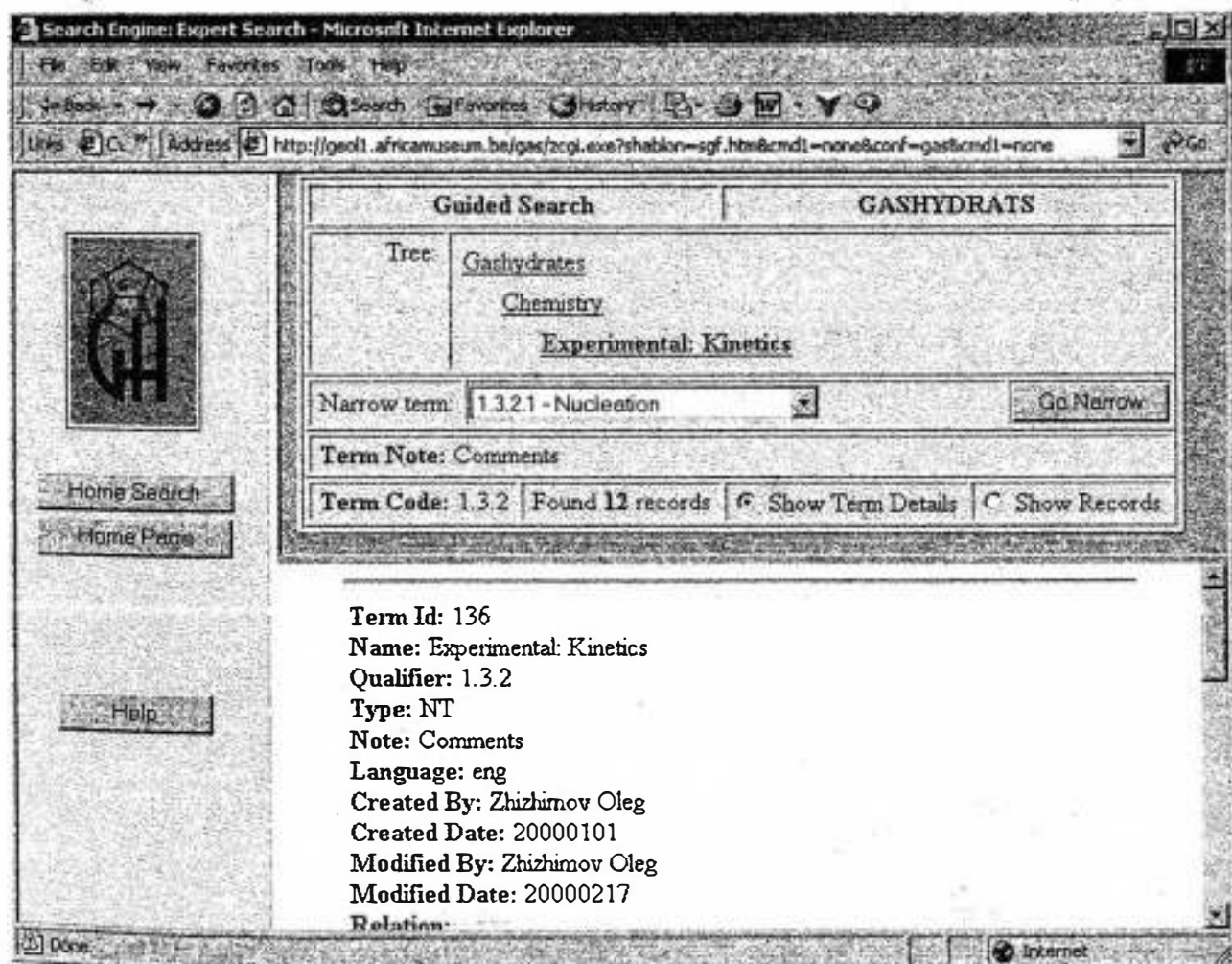


Рис. 3. Масштабирование и географический поиск.

интерфейс пользователя при масштабировании карты и географическом поиске в базе метаданных.

Следующий слой — доступ к накопленной ранее библиографической информации. Система позволяет использовать эту информацию в двух вариантах:

— осуществлять прямой поиск в любых библиографических базах данных, доступных по протоколу Z39.50;

— обращаться к специализированным библиографическим приложениям через URL, ссылка на который присутствует в GILS, для активизации расширенных возможностей этих приложений.

На рис. 4 показан интерфейс пользователя при выходе по URL в другую информационную систему.

Наконец, необходимым элементом информационной системы является классификация информации. Система должна обеспечивать доступ к тематическим рубрикам, возможность навигации по ним и проведение параллельного поиска. Эти задачи также были решены на основе протокола Z39.50 и схемы данных Zthes [5]. Внедрение этой схемы в мире только начинается. В схеме данных Zthes присутствуют все элементы, необходимые для построения рубрикаторов или тезаурусов: коды, названия, отношения. Это

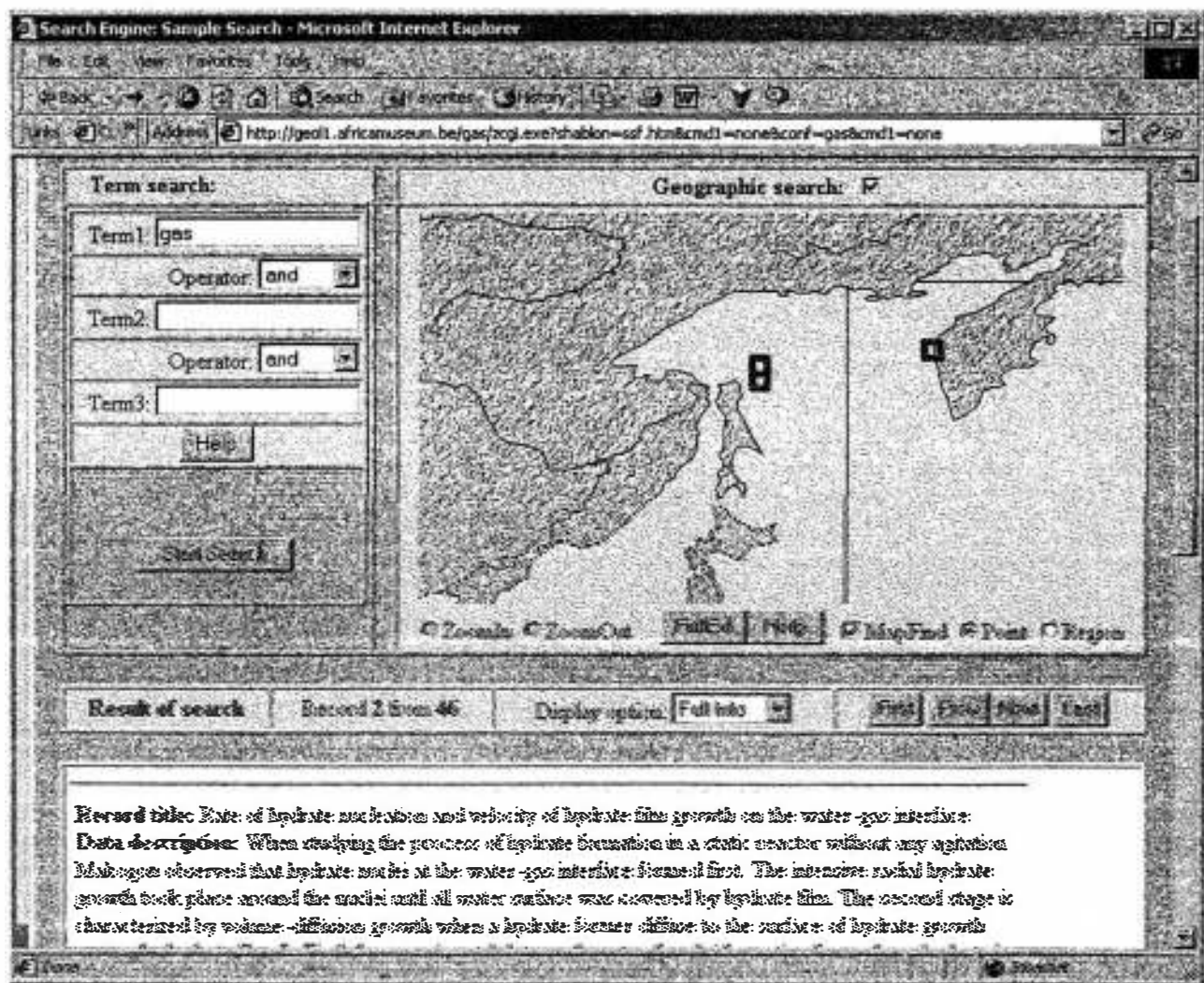


Рис. 4. Выход в другую библиографическую информационную систему.

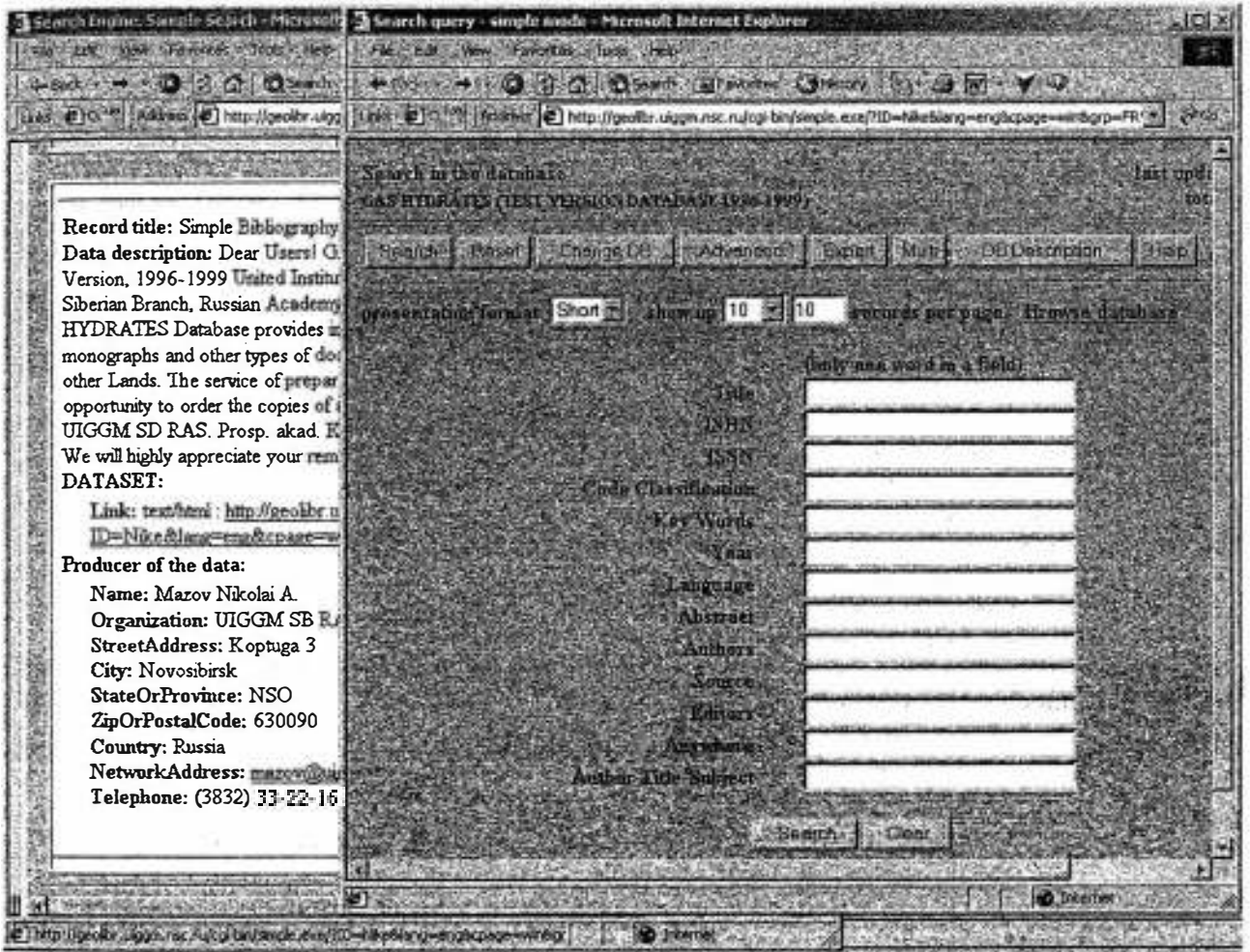


Рис. 5. Интерфейс навигации по рубриктору.

позволяет создать классификационную базу данных, обеспечить спецификации Z39.50 и предоставить через шлюз WWW-Z39.50 интерфейс для перемещений по рубриктору с одновременным поиском в базах метаданных и выходом на найденные записи. Интерфейс навигации по рубриктору показан на рис. 5.

Заключение

Работы по настоящему международному проекту еще не закончены, сейчас он находится в самой интенсивной стадии разработки (в нем принимают участие более 12 организаций из разных стран мира).

Кроме описанной технологии существуют и другие важные стороны проекта, такие, как сбор информации, ее систематизация. На сегодняшний день существуют две точки входа в информационную систему, доступ к которым открыт, и все желающие могут принять участие в опытной эксплуатации системы: в России (г. Новосибирск, ОИГГиМ СО РАН, <http://geolibr.uiggm.nsc.ru/gas>) и в Бельгии (г. Брюссель, Королевский музей Африки, <http://geoll.africamuseum.be/hydrat>).

Отметим, что описанный выше синтез различных технологий в рамках одной информационной системы оказался продуктивным. Созданная распределенная информационная

система позволяет решать широкий круг задач доступа к разнородным данным, предоставляет разнообразные пользовательские интерфейсы и может тиражироваться, так как не привязана к конкретной задаче и конкретному наполнению данными. Система может служить прототипом для информационных систем подобного класса.

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] ZOOПARK модульный сервер Z39.50. Версия 2.20. ОИГГиМ СО РАН.
<http://geolibr.uiggm.nsc.ru/doklads/Z-docs/ZooPARK.doc>
- [3] GLOBAL Information Locator Service. <http://www.gils.net>
- [4] INDEX Data. YAZ User's Guide and Reference. Version 1.4. <http://www.indexdata.dk>
- [5] TAYLOR M. Zthes: A Z39.50 Profile for Thesaurus Navigation. Version 0.3b.
<http://lcweb.loc.gov/z3950/agency/profiles/zthes-03>

Поступила в редакцию 18 апреля 2000 г.