

# Пример распределенной информационной системы на основе метаданных и международных стандартов

Жижимов О.Л.  
ОИГГиМ СО РАН, Новосибирск, Россия  
[mazov@uiqgm.nsc.ru](mailto:mazov@uiqgm.nsc.ru)

Коджесян В.С.  
ОИГГиМ СО РАН, Новосибирск, Россия  
[vlad@uiqgm.nsc.ru](mailto:vlad@uiqgm.nsc.ru)

Мазов Н.А.  
ОИГГиМ СО РАН, Новосибирск, Россия  
[mazov@uiqgm.nsc.ru](mailto:mazov@uiqgm.nsc.ru)

Технологиям построения распределенных информационных систем в последнее время посвящается много работ как в России, так и за рубежом. При этом применяются различные технологии с различной степенью интеграции компонент и функциональными возможностями. Настоящая работа посвящена описанию одной из информационных систем, создающейся в рамках международного проекта по газгидратам. Авторы надеются, что технологический подход, реализованный в этом проекте, окажется продуктивным для других аналогичных проектов.

## 1 ЦЕЛИ И ЗАДАЧИ ПРОЕКТА

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

На языке специалистов газгидраты представляют со-

бой нечто среднее между твердым веществом и сильно сжатым и замороженным газом. Они реально существуют в природе. Уже открыто несколько месторождений. Ученые геологи находят газгидраты в совершенно разных местах планеты, например, в мировом океане на большой глубине. Некоторые газгидраты 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 газгидраты привлекают и своими экономическими перспективами. Существуют оценки специалистов, что газгидраты могут служить основным сырьем для химической промышленности 21 века и быть эффективным источником энергии подобно нефти и природному газу. Несмотря на большой интерес к столь уникальным объектам, информация о них скудна, раздроблена по разным источникам и не систематизирована.

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

## 2 ТЕХНОЛОГИЯ

Для реализации проекта, т.е. для построения распределенной информационной системы по газгидратам, были сформулированы требования, которым должна удовлетворять эта система, а именно:

- допускать хранение данных различных типов: тексты, карты, фотографии изображения, библиографические сведения о научных работах, ссылки на другие ресурсы;
- допускать хранение метаданных, т.е. данных об этих данных;

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

- 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;
- удовлетворять 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 был выбран протокол Z39.50 (ISO 23950) [1]. Использование этого протокола 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 было выбрано несколько параллельных возможностей доступа: Z39.50, WWW, ftp и, вообще любые другие, определяемые URL. В качестве точки доступа к информационной системе со стороны WWW было решено использовать шлюз Z39.50-WWW, выполняющий функции клиента Z39.50. Этот выбор обеспечивал достаточно простой и демократичный механизм не только для построения пользовательских интерфейсов в окружении привычного Web-браузера, но и для создания существенно продвинутых решений на основе Z39.50.

Другой 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 имеется 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.

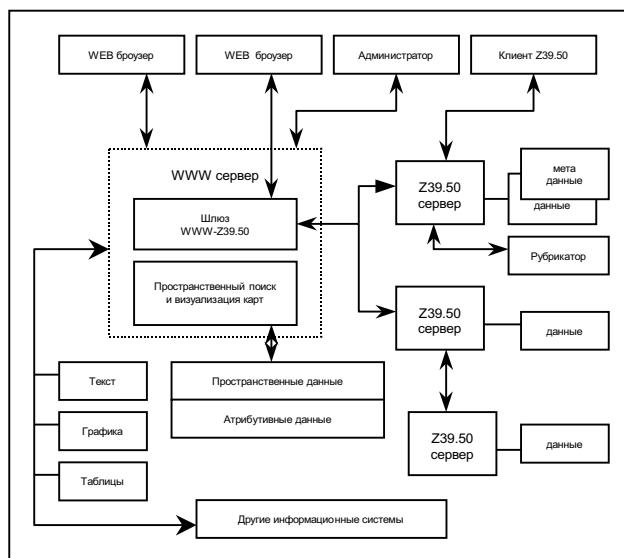


Рис.1: Схема информационной системы

### 3 РЕАЛИЗАЦИЯ

В качестве сервера Z39.50 был использован сервер ZooPARK, разработанный в ОИГГиМ СО РАН [2]. Реализация этого сервера допускает возможность установки практически на любую наиболее распространенную серверную платформу, кроме этого он показал достаточно высокую устойчивость в эксплуатации.

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

Поскольку в основу всей информационной системы было положено требование соответствия мировым стандартам, функциональность построенной системы никак не связана с конкретными 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 GILS - это стандартная схема, отображение на которую 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 тэговой структурой GILS. Индексирование произво-

дилось при помощи свободно распространяемого программного обеспечения Zebra от компании IndexData [4], создающего совместимую по структуре базу данных с одним из провайдеров данных сервера ZooPARK. В дальнейшем, при увеличении объема хранимой информации, планируется задействовать 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 задействовать механизм гипертекстовых ссылок при просмотре метаданных, предоставляя простой выход на сами данные.

На Рис. 2 показан интерфейс пользователя при поиске информации в базе метаданных.

Другой информационный слой, который интегрирован с метаданными - географическая информация о по-

ложении и размерах описываемых месторождений газ-гидратов. Для четкой идентификации объекта и его визуализации на карте необходимо использовать данные ГИС (ГеоИнформационные Системы) систем. 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 уже несколько мегабайт. Существуют специализированные системы и форматы для хранения такого рода информации. В описываемой системе был исполь-

зован один из наиболее популярных и стандартных форматов хранения - SHP файлы, атрибутивная же информация была размещена в файлах форма DBF. SHP/DBF - очень распространенные форматы ГИС данных, любая ГИС система может их экспортировать и импортировать. В нашей системе доступ 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 отображения на ней записей из баз метаданных. На Рис. 3 показан интерфейс пользователя при масштабировании карты и географическом поиске в базе метаданных.

Следующий слой - доступ к накопленной ранее библиографической информации. Система позволяет подключать 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. В этом случае в системе необходим доступ к тематическим рубрикам, возможность навигации по ним и проведение параллельного поиска. Это задача также была решена на основе протокола Z39.50 и схемы данных Zthes [5], внедрение которой сейчас в мире только начинается. В схеме данных Zthes присутствуют все элементы, необходимые для построения рубрикаторов или тезаурусов: коды, названия, отношения, что позволяет создать классификационную базу данных, обеспечить спецификации Z39.50 и предоставить через WWW-Z39.50 шлюз интерфейс для перемещений по рубрикам с одновременным поиском в базах метаданных и выходом на найденные записи. Интерфейс навигации по рубрикам показан на Рис. 5.

## 4 ЗАКЛЮЧЕНИЕ

Работы по настоящему проекту еще не закончены. Проект находится в самой интенсивной стадии разработки. Он является международным, в нем принимают участие более 12 организаций из разных стран мира. Выше была описана только технологическая его часть. Существуют и другие не менее интересные и важные стороны проекта: сбор информации, ее систематизация.

На сегодняшний день открыто две точки входа в информационную систему, доступ к которым открыт для всех и все могут принять участие в опытной эксплуатации системы: в России в (г. Новосибирск, ОИГГиМ СО РАН - <http://geolibr.uiggm.nsc.ru/gas>) и в Бельгии (г. Брюссель, Королевский Музей Африки - <http://geol1.africamuseum.be/hydrat>).

В заключении отметим, что описанный выше симбиоз 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, ибо не привязана к конкретной задаче и к 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] ZooPARK модульный сервер Z39.50. Версия 2.36. 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] Mike Taylor. Zthes: A Z39.50 Profile for Thesaurus Navigation.Version 0.3b. <http://lcweb.loc.gov/z3950/agency/profiles/zthes-03>.

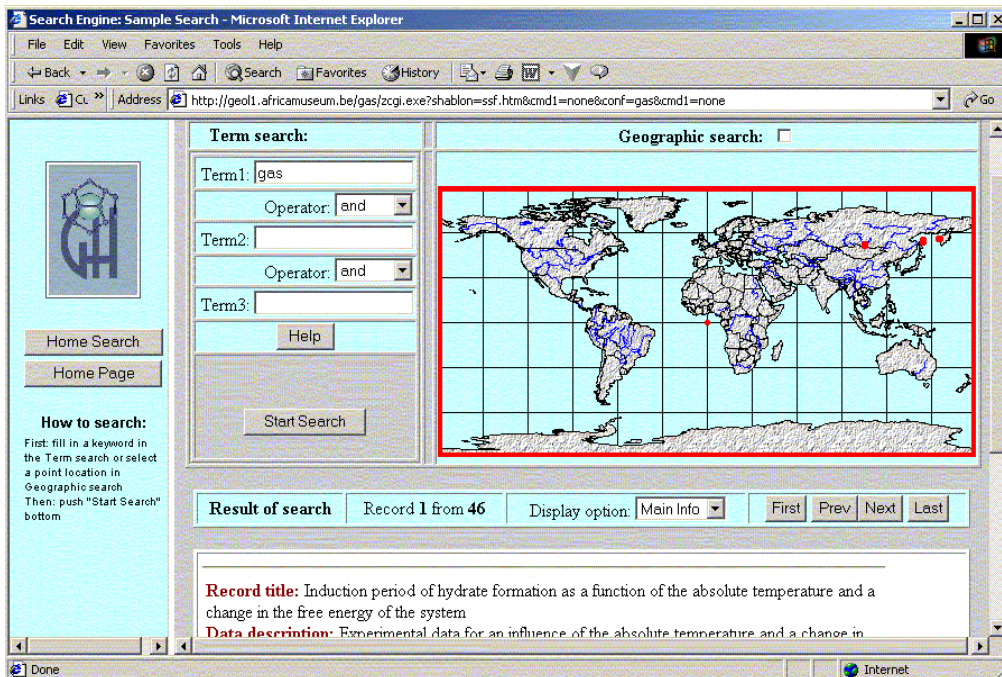


Рис. 2: Интерфейс пользователя при поиске

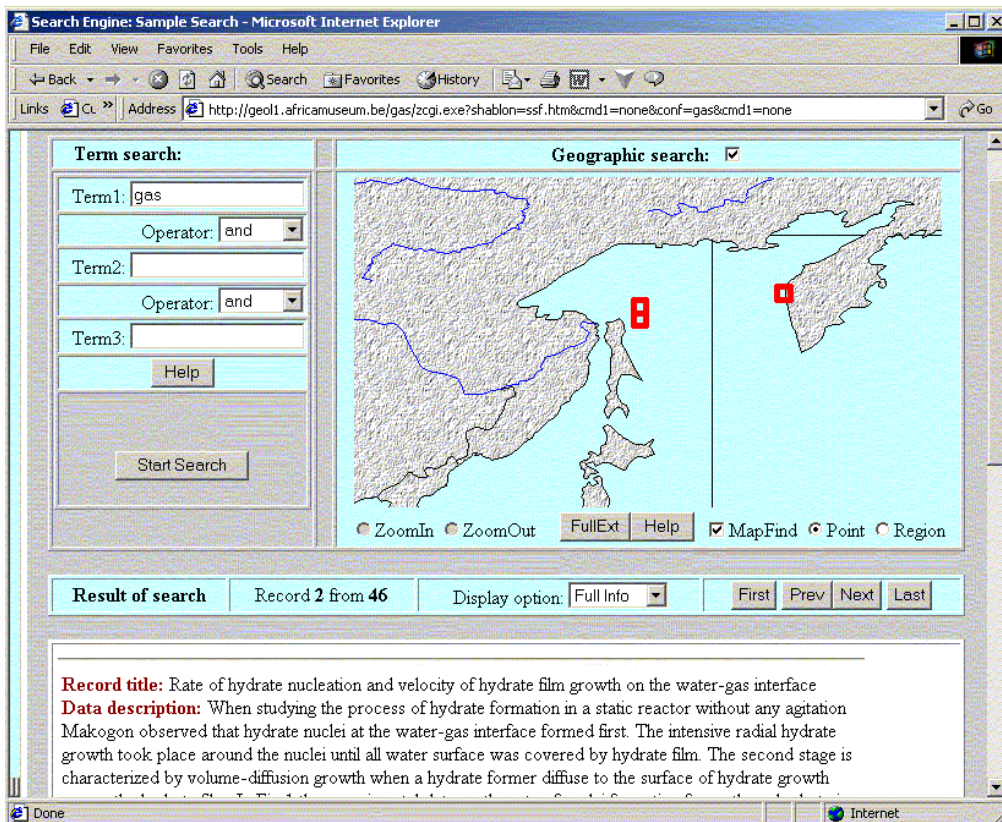


Рис.3: Масштабирование и географический поиск

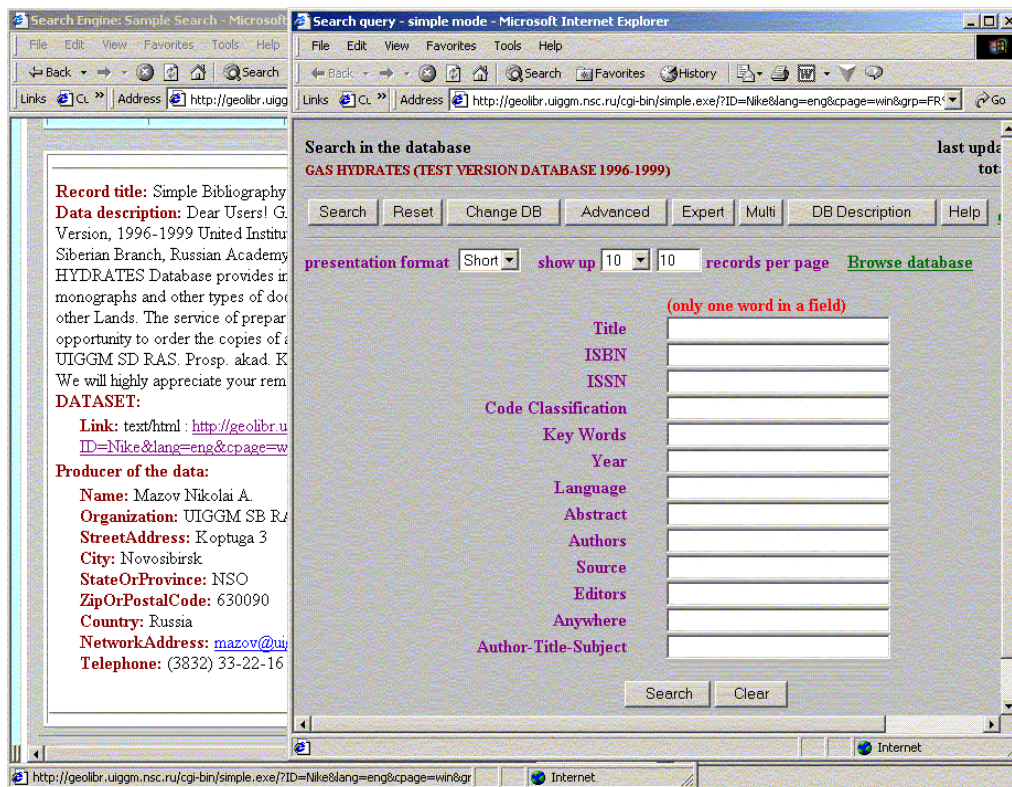


Рис.4: Выход в другую библиографическую информационную систему

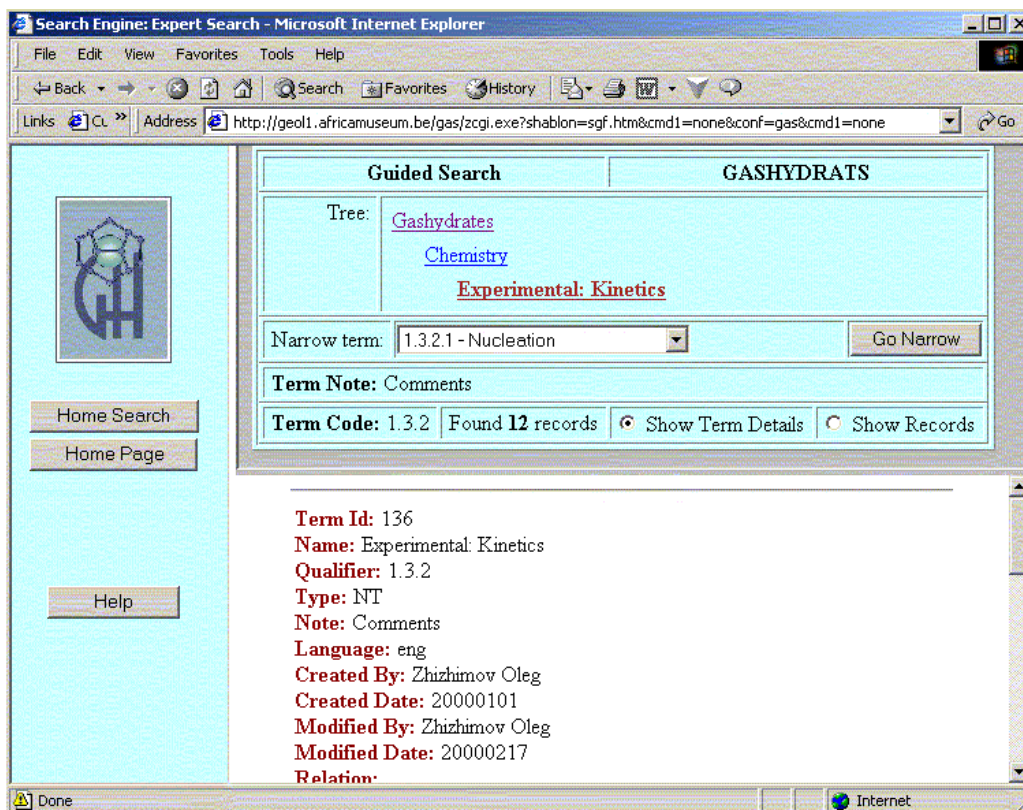


Рис.5: Интерфейс навигации по рубрикатору