

**Комплекс ZooPARK в задачах обеспечения доступа
к геоинформационным ресурсам**
**ZooPARK Complex and Maintenance of Access
to Geoinformation Resources**
**Комплекс ZooPARK у завданнях забезпечення доступу
до геоінформаційних ресурсів**

О. Л. Жижимов

Институт вычислительных технологий Сибирского Отделения РАН, Новосибирск, Россия

Н. А. Мазов

*Институт нефтегазовой геологии и геофизики им. академика А.А. Трофимука Сибирского
Отделения РАН, Новосибирск, Россия*

Oleg Zhizhimov

*Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences,
Novosibirsk, Russia*

Nikolay Mazov

*Trofimuk Institute of Petroleum Geology and Geophysics, Siberian Branch of the Russian Academy
of Sciences, Novosibirsk, Russia*

О. Л. Жижимов

Институт обчислювальних технологій Сибірського Відділення РАН, Новосибірськ, Росія

М. О. Мазов

*Институт нафтогазової геології та геофізики ім. академіка А. Трофимука Сибірського
Відділення РАН, Новосибірськ, Росія*

Рассматривается возможность использования серверного комплекса Z39.50 – ZooPARK для обеспечения доступа к геоинформационным ресурсам в рамках гетерогенных распределенных информационных систем. Рассмотрены некоторые аспекты поиска и извлечения геоинформации в соответствии с международными стандартами. Приведены примеры пользовательских графических интерфейсов.

A possibility of using the server complex Z39.50 - ZooPARK for providing access to geoinformation re-sources within the heterogeneous distributed information systems is considered. Some aspects of search and retrieval of geoinformation according to the international standards are considered. Some examples of user graphic interfaces are given.

Розглядається можливість використання серверного комплексу Z39.50 – ZooPARK для забезпечення доступу до геоінформаційних ресурсів у межах гетерогенних розподілених інформаційних систем. Розглянуто деякі аспекти пошуку та здобування геоінформації у відповідності до міжнародних стандартів. Наведено приклади використання графічних інтерфейсів користувачів.

В последние десятилетия в связи с бурным развитием информационно-вычислительных, телекоммуникационных и спутниковых технологий стало возможным накопление и обработка специализированными организациями и ведомствами большого количества геоинформации, о чем свидетельствует рост числа публикаций как у нас в стране, так и за рубежом. Геоинформация в настоящее время представлена в различных

видах и системах и сейчас как никогда актуальна задача организации системы, позволяющей наиболее эффективно и просто использовать глобальные и локальные сети для интеграции различных информационных ресурсов (библиографических баз данных НТИ, как метаданных, так и полных текстов; различных цифровых аудио-, видео- и фото- коллекций; геоинформационных ресурсов и др.) и их использования в распределенных информационных системах. В Сибирском отделении РАН работам, связанным с интеграцией информационных ресурсов и построением единой распределенной информационной системы уделяется немало внимания в последние годы.

К настоящему моменту в ряде стран, включая, прежде всего, США, Канаду, Австралию и др., уже разработаны жизнеспособные промышленно эксплуатируемые распределенные системы [4-6] на основе стандарта ISO-23950 (протокол Z39.50) [1], включающие сотни серверов Z39.50 [5].

Организация пользовательского и административного доступа к информационным системам, интегрирующим информационные объекты с географическими характеристиками и (или) координатной привязкой, а также организация основных модулей подобных систем, требует специальных подходов, основанных на методах, используемых в геоинформационных системах (ГИС). В качестве информационных объектов при этом могут выступать

- библиографические ресурсы с указанием географической принадлежности
- картографические традиционные и электронные ресурсы
- метаданные в различных схемах, имеющих координатную и/или географическую привязку
- внешние информационные объекты, доступные по стандартным протоколам (объекты специализированных геоинформационных систем, ресурсы публичных WEB-серверов, ресурсы международных сообществ).

В настоящем докладе рассматриваются возможности комплекса ZooPARK для работы с подобными информационными объектами. Как неоднократно указывалось, архитектура и состав комплекса позволяет реализовать трехуровневую модель распределенной информационной системы, основанной на международных стандартах (Z39.50, WWW, LDAP). При этом ядро сервера, обеспечивающего доступ к информационным ресурсам, допускает динамическое подключение внешних источников данных и прозрачную переадресацию пользовательских запросов к другим информационным системам. Это предоставляет разработчикам гибкие средства для достижения необходимой функциональности разрабатываемых информационных систем, обеспечивая не только платформенную и системную независимость (в рамках POSIX) предлагаемых решений, но и соответствие их принятым международным стандартам.

Ниже рассмотрены основные виды работ с ресурсами ГИС и реализация пользовательских интерфейсов для их выполнения на примере комплекса ZooPARK.

Построение сложных запросов посредством пользовательских интерфейсов встроенного шлюза

При работе с ГИС-ресурсами даже на уровне метаданных возникает задача построения достаточно сложных запросов, особенно, если речь идет о запросах с географическими координатами. Наиболее простое решение – предоставление

возможности графического задания географического примитива как основы для поиска по координатам. Таким примитивом может служить многоугольник, ограничивающий некоторую область, локализирующую искомое множество точек. В комплексе ZooPARK (шлюз Z-GW) реализованы интерфейсы задания поискового геометрического примитива в виде прямоугольника, задающего четыре пары географических координат.

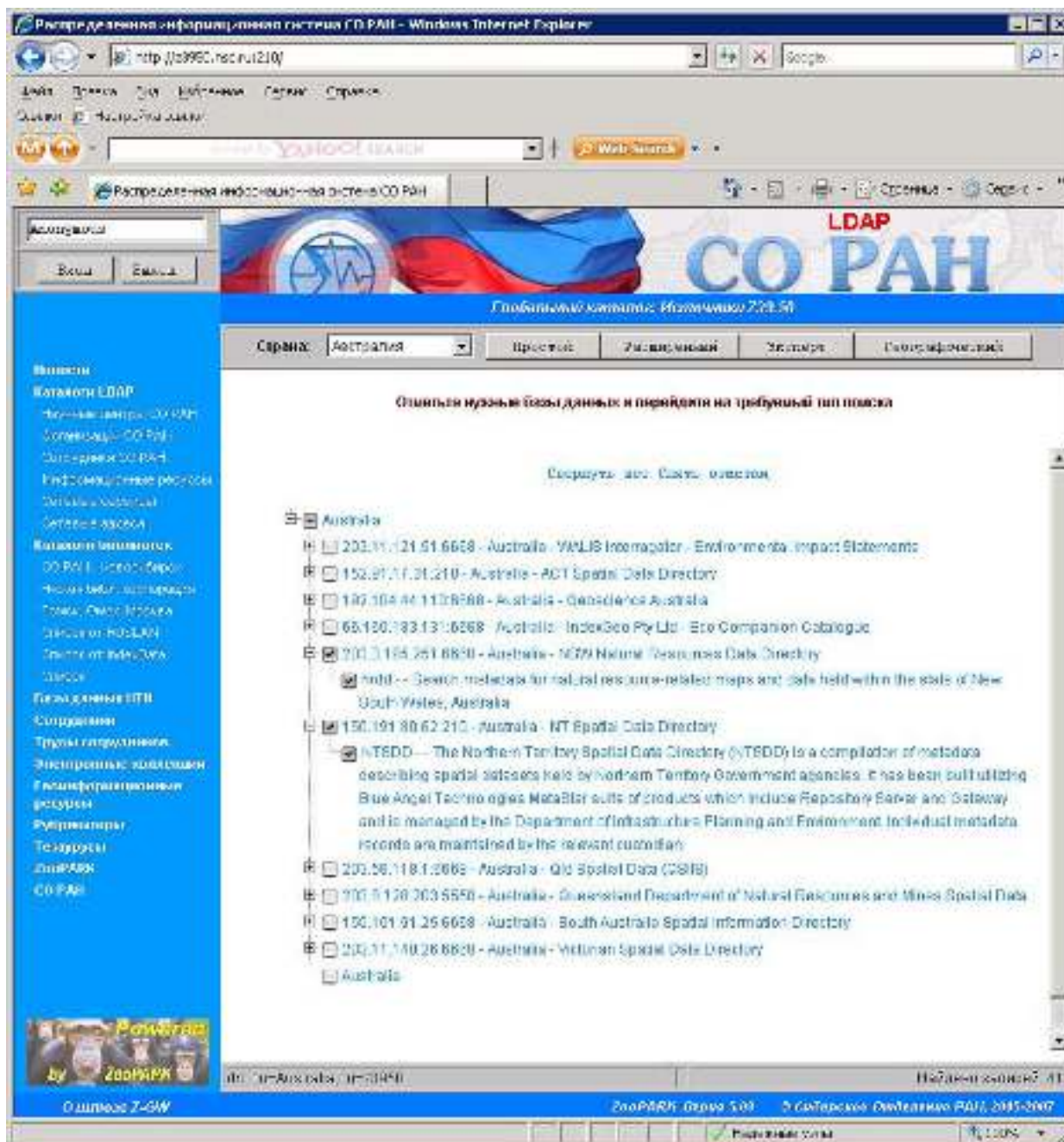


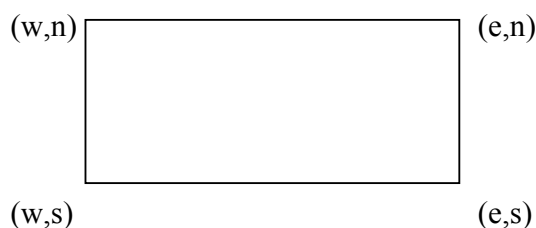
Рис.1 Шлюз Z-GW комплекса ZooPARK: выбор источников данных

В качестве подложки (топоосновы) используется векторная карта или растровый

космический снимок, предоставляемые сервисом GoogleMap. При выборе прямоугольника автоматически формируется RPN-запрос в соответствии с заданным набором поисковых атрибутов. В частности, для набора поисковых атрибутов GEO поисковый запрос RPN в нотации PQF выглядит следующим образом

```
@and @and @and @attr geo 1=2038 @attr 2=2 @attr 4=109 e
@attr geo 1=2039 @attr 2=4 @attr 4=109 w
@attr geo 1=2040 @attr 2=4 @attr 4=109 s
@attr geo 1=2041 @attr 2=2 @attr 4=109 n
```

при задании поискового прямоугольника в виде



Более подробно некоторые вопросы, связанные с поиском информации с географической привязкой рассмотрены в [7].

Просмотр индексов баз данных

Функция просмотра индексов баз данных (сервис SCAN для Z39.50) в сервере атрибутов. Возможности этого сервиса при работе с метаданными ГИС в схемах GEO и GIS такие же, как и при работе с библиографическими данными (схемы RUSmarc, USmarc и т.п.), поэтому такими же остаются и пользовательские интерфейсы. Однако существует весьма существенное отличие, которое вытекает не столько из смены схемы данных, сколько из более строгого контроля используемых типов данных. Дело в том, что в библиографических метаданных тип данных практически никогда не контролируется, более того, как правило, считается, что все данные имеют текстовый тип, независимо от их реального семантического типа. Это, конечно, обедняет поисковые возможности информационных систем, но обычно не сужает типовые поисковые потребности пользователей. Например, в подавляющем большинстве систем считается, что поле «Год издания» является текстовым. Текстовый тип этого поля не позволяет проводить поиск с операторами «больше» или «меньше», но для обычного пользователя этот факт остается незамеченным.

Однако для ГИС индексирование полей, содержащих географические координаты, как полей текстового типа лишено всякого смысла, т.к. в основе запросов по координатам лежат те самые операторы «больше» и «меньше», которые в случае текстового индексирования не работают. Эти поля должны индексироваться как поля вещественного типа.

Обработка поисковых запросов различного типа (RPN, CCL, CQL, SQL и др.)

Текущая версия сервера ZooPARK поддерживает обработку указанных типов запросов. В отличие от библиотечных систем, при работе с ГИС чаще всего используются запросы SQL. Это связано с подавляющим распространением в ГИС реляционных СУБД, для которых SQL является более «естественным» языком запросов. Более того, чаще всего используются специфические расширения SQL для оптимизации работы с геоданными.



Рис.2 Шлюз Z-GW: поиск информации в выбранных источниках данных

Эти расширения различны у разных производителей СУБД (ORACLE, MS SQL Server и др.) и комплексом ZooPARK сегодня не поддерживаются.

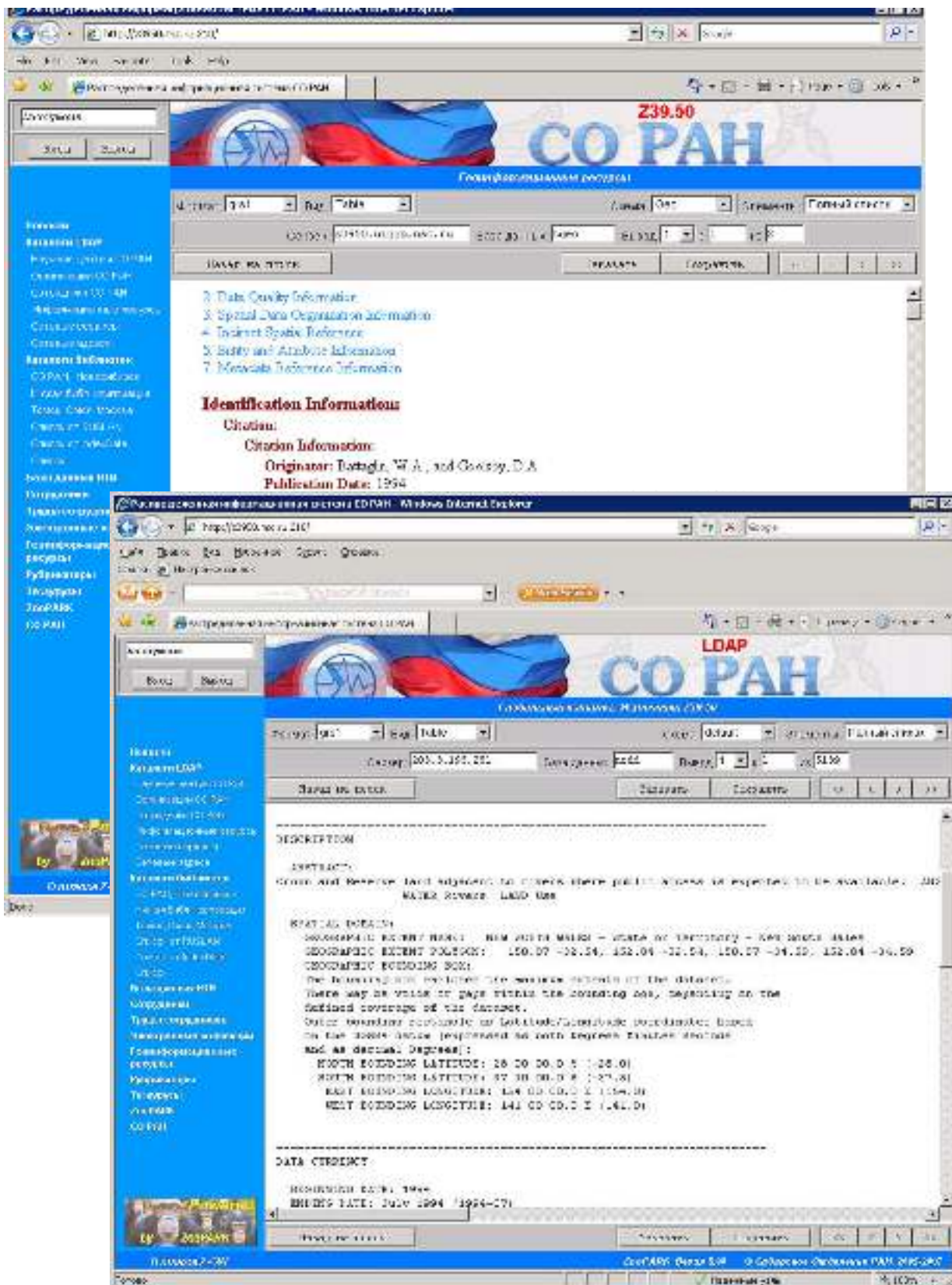


Рис. 3 Шлюз Z-GW: представление данных GEO

Однако в дальнейшем поддержка некоторых расширений SQL для ГИС будет включена на уровне провайдеров данных.

Поиск информации в многобазовом окружении

Серверный комплекс ZooPARK позволяет проводить поиск информации и просмотр индексов в многобазовом окружении независимо от схемы данных. При многобазовом поиске для каждого источника данных активизируется свой поток, что фактически распараллеливает эту процедуру, особенно в случае распределенной системы, когда разные источники данных находятся на физически разных серверах. Это распараллеливание реализовано прозрачным для пользователя образом как для сервера Z39.50, так и для шлюза Z-GW.

Здесь следует заметить, что в мире существует интеграция не только библиотечных ресурсов. Существуют распределенные информационные системы, включающие десятки серверов, предоставляющие доступ к ГИС метаданным. На рис.1 продемонстрирован интерфейс выбора источников данных из иерархического дерева международных ресурсов Z39.50 и SIP для поиска метаданных. Это дерево построено на основе публичной информации, полученной из интернет, упорядоченной и помещенной в каталог LDAP. Работа с этим каталогом возможна как из шлюза Z-GW, так и из специализированных программ-клиентов. Технология LDAP выбрана для обеспечения общей интероперабельности.

На рис. 2 представлен результат поиска одновременно по двум источникам международной системы GEO метаданных.

Извлечение информации в нужных схемах и форматах

Для работы со структурированными метаданными в схемах GILS и GEO [2,3] комплекс ZooPARK обеспечивает извлечение информации и представление ее как в структурированном виде в форматах GRS-1, XML, HTML, SQL-RS, ISO2709, так и в неструктурированном формате SUTRS. Извлечение происходит как всех элементов схемы данных, так и именованных подмножеств (именованных наборов элементов) схемы данных в соответствии с прикладными профилями GEO или GILS.

Визуализация информации через встроенный шлюз

Шлюз Z-GW, являющийся составной частью комплекса ZooPARK, позволяет просматривать извлеченные метаданные в разных структурированных и неструктурированных форматах (см. выше) и переходить по гипертекстовым ссылкам при их наличии соответствующих элементов данных текущей записи. Последнее возможно для структурированного формата GRS-1. Для формата XML происходит обычная процедура визуализации записи в соответствии с включенными в эту запись инструкциями. На Рис.3 продемонстрирован просмотр записи GEO, полученной из международного источника, в форматах GRS-1 и SUTRS при помощи шлюза Z-GW.

Таким образом, программный комплекс ZooPARK, включающий как компоненты для организации доступа к разнородным базам данных, так и компоненты организации пользовательских интерфейсов на основе распространенных WEB-технологий, обладает базовой функциональностью, необходимой для построения систем для работы с геоинформацией. Прототипы таких информационных систем сегодня создаются в Сибирском отделении РАН.

Литература

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. Application Profile for the Government Information Locator Service (GILS), Version 2, November 24, 1997. (http://www.gils.net/prof_v2.html).
3. Douglas D. Nebert. Z39.50 Application Profile for Geospatial Metadata or "GEO"/ Version 2.2 / U.S. Federal Geographic Data Committee (<http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm>)
4. Clearinghouse (<http://www.fgdc.gov/clearinghouse/clearinghouse.html>).
5. Websites Associated with Clearinghouse (http://clearinghouse4.fgdc.gov/registry/-clearinghouse_sites.html).
6. Hourly Status of NSDI Clearinghouse Nodes (<http://registry.gsdi.org/serverstatus>).
7. Жижимов О.Л. Мазов Н.А. Об использовании географических координат при поиске библиографической информации // Научные и технические библиотеки. - 2009. - № 1. - С. 54-60.