

ЦИФРОВОЙ РЕПОЗИТАРИЙ ДАЛЬНЕВОСТОЧНОГО ГЕОЛОГИЧЕСКОГО ИНСТИТУТА ДВО РАН: ОТКРЫТЫЙ ДОСТУП К НАУЧНЫМ ДАННЫМ ПО ГЕОЛОГИИ ДАЛЬНЕГО ВОСТОКА РОССИИ

А.В. Белоусов

*Дальневосточный геологический институт ДВО РАН
Владивосток*

Открытый доступ к результатам научных исследований — одно из пяти важнейших научных событий 2003 года по версиям The Scientist и Nature, один из семи научных прорывов 2003-го. Открытый доступ — это практика предоставления неограниченного доступа через интернет к библиотечной научной информации. Выделяют два основных технологических направления: журналы открытого доступа и архивы (репозитории) открытого доступа. Журналы открытого доступа это научные журналы, которые доступны онлайн для читателя без финансовых, юридических или технических ограничений. Репозитории предназначены для публикации любой библиографической информации: статей, авторефератов диссертаций, рукописей, отчетов по грантам, учебных материалов и др.

Для хранения библиографических описаний в библиотечных каталогах широко используется формат MARC и его модификации USMARC, RUSMARC и т.д. В цифровых репозиториях распространен формат метаданных Dublin Core, который может быть использован для описания любого веб (видео, изображение, веб страница и др.) или физического (книга, картина и др.) ресурса. В 1999 году Open Archives Initiative был разработан протокол для сбора метаданных OAI-PMH.

Разработано и успешно работает программное обеспечение с открытым кодом для создания и поддержки таких OAI-PMH совместимых архивов. Наиболее популярные из них: E-prints, Dspace.

С 2000 года в Интернет работает Электронная библиотека Дальневосточного геологического института Дальневосточного отделения РАН (ДВГИ ДВО РАН) <http://elibrary.fegi.ru>. На основе этой библиотеки возникла задача создания единой точки доступа к мировым библиотечным научным ресурсам по геологии Дальнего Востока России.

Мировые научные данные публикуются в научных журналах, цифровых репозиториях, полнотекстовых научных базах данных и хранятся в библиотеках. Реализация задачи возможна с использованием интеграции метаданных ресурсов из разнотипных технологических систем.

В настоящий момент существует два различных подхода для интеграции ресурсов на основе метаданных: поиск в распределенных каталогах или сбор метаданных в центральный каталог. Жижимовым решена задача интеграции научных данных из цифрового репозитория DSpace и библиотечных каталогов Ирбис на основе протокола распределенного поиска Z39.50 (система Zoopark) [1].

При проектировании Цифрового репозитория ДВГИ ДВО нами сформулированы следующие основные задачи:

- Создание специализированных тезаурусов;
- Разработка тематического фильтра интегрируемой информации;
- Интеграция данных из научных журналов, цифровых репозитариев, полнотекстовых научных баз данных и библиотек;
- Организация хранения: метаданных, полнотекстовых данных;
- Реализация подсистемы многокритериального поиска, в том числе тематического;
- Реализация некоторых дополнительных функций взаимодействия с пользователем, в том числе и таких как добавление пользователем авторских публикаций;
- Организация вывода результатов поиска, метаданных и самих данных.

Реализованная функциональная схема Системы представлена на рис. 1.

Для создания репозитория выбрана программная среда Dspace, которая обладает богатой функциональностью, вследствие чего, она широко распространена. Удобная система каталогов, наличие клиента и сервера OAI-PMH, полнотекстовый поиск, основанный на поисковом инструменте Apache Lucene или Apache Solr, разграничение прав и поддержка протокола доступа LDAP, открытый код и большое сообщество пользователей и разработчиков по всему миру, возможность хранить цифровой материал любого типа.

В репозитарий ДВГИ ДВО РАН интегрируются метаданные из мировых институциональных репозитариев по протоколу OAI-PMH, библиотечных каталогов по протоколу Z39.50 [3][4] или используя функцию экспорта системы управления библиотекой. Проектом предусмотрен сбор метаданных из полнотекстовых научных баз данных, таких как: Научная электронная библиотека eLIBRARY.RU и ScienceDirect.

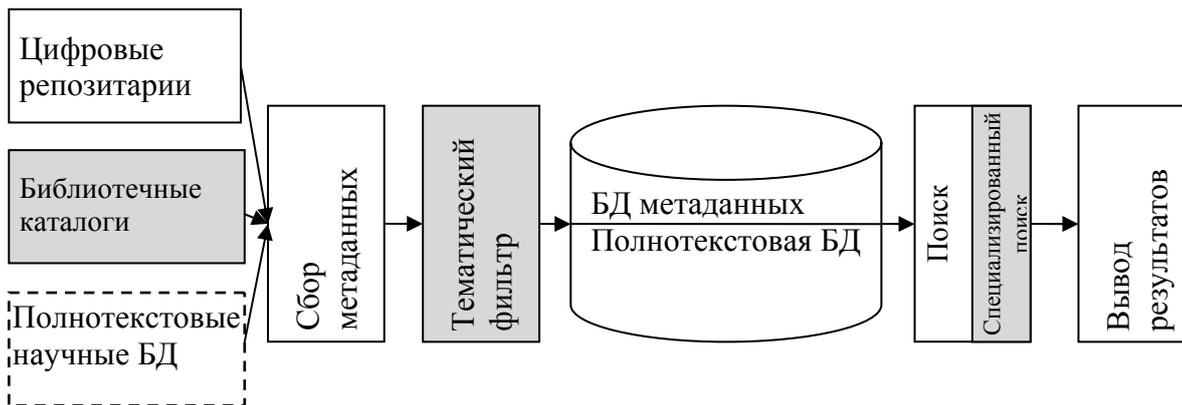


Рис. 1. Функциональная схема системы «Цифровой репозиторий ДВГИ ДВО РАН по геологии Дальнего Востока России»

Система адаптирована для геологии Дальнего Востока России внесением в неё тематических тезаурусов: геологического и географического, взятых из предметного указателя тематической монографии [5], а также стратиграфического (международная стратиграфическая шкала, принятая в 2009 году международной стратиграфической комиссией).

Тезаурус выполняет следующие функции в информационной системе [2]:

- предоставляет значение для преобразования естественного языка автора, индексатора и пользователя в словарь, который может быть использован для индексирования и поиска;
- способствует единообразию в терминах;
- указывает на семантические отношения между терминами;
- предоставляет последовательную и ясную иерархию в навигационной системе, чтобы помочь пользователю находить желаемый объект.

В репозитории организован фильтр собранных метаданных на основе морфологического поиска (stemming) по словарю геологических объектов. Полученные метаданные добавляются в базу данных в формате Dublin Core.

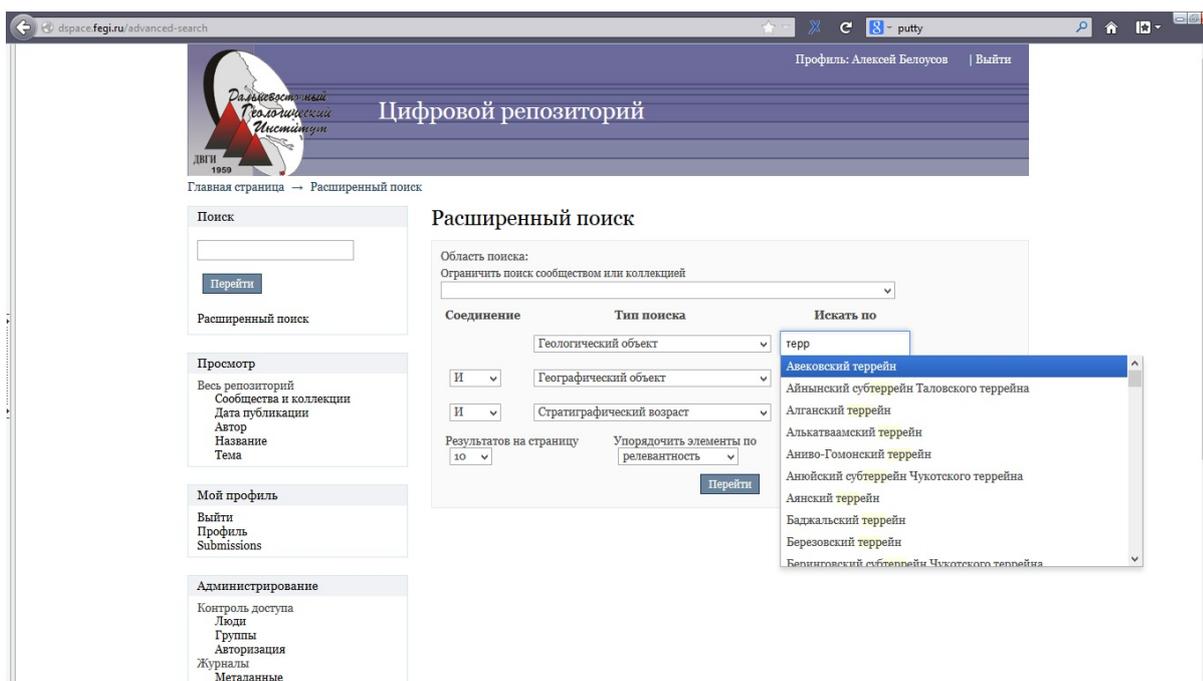


Рис. 2. Интерфейс расширенного поиска Системы с использованием геологического тезауруса

В Системе реализованы некоторые дополнительные функции взаимодействия с пользователем, в том числе и такие как добавление пользователем авторских публикаций.

Поиск в Системе реализован в двух режимах: стандартный и расширенный. В расширенном режиме доступен многокритериальный поиск по 13 параметрам, в том числе и по таким параметрам как: географические и геологические объекты Дальнего Востока России, стратиграфический возраст (рис. 2).

Реализация ресурса <http://dspace.fegi.ru> позволяет организовать единую точку доступа к библиотечным данным по геологии Дальнего Востока России.

Работа выполняется при финансовой поддержке гранта ДВО РАН № 13-III-B-08-010.

ЛИТЕРАТУРА

1. Интеграция разнородных данных в задачах исследования природных экосистем / Жижимов О.Л., Молородов Ю.И., Пестунов И.А., Смирнов В.В., Федотов А.М. // Вестник Новосибирского государственного университета. 2011. Том 9. № 3. С. 67—74.
2. Nguyen N.T., Kowalczyk R. Role of Thesauri in the Information management in the Web-Based Services and Systems // Lecture Notes in Computer Science. 2011. Vol. 6560. P. 25 —49.
3. Kaczmarek J., Naun C.C. A statewide metasearch service using OAI // Library Hi Tech. 2005. Vol. 23. P. 576 —586.
4. Guidelines for preparing a Z39.50/SRU target to enable metadata harvesting [Электронный ресурс] / TELplus. The European Library [сайт]. URL: [http://cyberdoc.univ-lemans.fr/PUB/CfU/Journee_UNIMARC_Lyon/TELplus-D2.3_v1.0\[1\].pdf](http://cyberdoc.univ-lemans.fr/PUB/CfU/Journee_UNIMARC_Lyon/TELplus-D2.3_v1.0[1].pdf).
5. Геодинамика, магматизм и металлогения Востока России: Монография / отв. ред. А.И. Ханчук. Владивосток: Дальнаука, 2006. XXX с.