

*Жижимов О.Л., Федотов А.М., Шокин Ю.И.
ИВТ СО РАН, Новосибирск*

ИНФОРМАЦИОННЫЕ СИСТЕМЫ ИВТ СО РАН: ПРИНЦИПЫ, АРХИТЕКТУРА, РЕАЛИЗАЦИИ¹

Обсуждаются основные принципы построения информационных систем ИВТ СО РАН, которые являются следствием предъявляемым к ним основным требованиям. Рассмотрена архитектура основных элементов этих систем и правила их взаимодействия. Приводится описание действующих реализаций.

Ключевые слова: распределенные информационные системы, базы данных, CRIS, интероперабельность, ZooSPACE, LDAP, SRW/SRU, Z39.50.

Институт вычислительных технологий СО РАН, как и любая другая научная организация, оперирует большим и постоянно возрастающим объемом информации, связанной с его основной деятельностью. Традиционно источником этой информации являются локальные базы данных о научных исследованиях, специализированные (библиографические, аналитические и пр.) базы данных сторонних организаций, порталы и WEB-сайты сети интернет и др. При этом структура этой информации постоянно меняется как по объему, так и по номенклатуре источников, по степени детализации и пр. Поэтому задача создания и поддержки собственных информационных систем, способных интегрировать и обрабатывать вышеуказанную информацию является постоянно актуальной [1-2], как и задача минимизации расходов на поддержку этих систем.

Опыт разработки и эксплуатации информационных систем выдвигает следующие основные требования к подобным системам [2]:

1. системы должны разрабатываться на основе свободно распространяемого программного обеспечения;
2. системы должны иметь возможность импортировать информацию из других систем;
3. системы должны предоставлять информацию в соответствии со стандартными протоколами, схемами и форматами;
4. системы должны интегрировать информацию, специфичную для CRIS-систем (Current Research Information System) [3-4];

¹ Работа выполнена при частичной поддержке РФФИ (проект 12-07-00472), президентской программы «Ведущие научные школы РФ» (ИШ 5006.2014.9) и Министерства образования и науки Российской Федерации (грант № 07.514.11.4130).

5. системы должны иметь полный комплект административных и пользовательских интерфейсов, желательно на основе WEB-технологий;
6. системы должны разграничивать доступ к хранимой информации;
7. системы должны иметь возможность расширения как по номенклатуре хранимых данных, так и по номенклатуре внешних источников информации

Для реализация этих требований проще всего использовать проверенные временем технологические подходы, основанные на использовании свободно распространяемого программного обеспечения и на стандартных сетевых протоколах. При этом речь может идти не об одной информационной системе, а о целом комплексе относительно самостоятельных информационных систем, взаимодействующих между собой в соответствии с фиксированными правилами (протоколами).

В качестве таких самостоятельных информационных систем в ИБТ СО РАН используются:

- База данных "Сотрудники" на основе LDAP-каталога
- База данных "Структура организации" на основе LDAP-каталога
- Информационная CRIS-система (публикации, проекты, конференции, персоны, организации) на основе СУБД PostgreSQL с административными и пользовательскими WEB-интерфейсами и возможностью импорта данных из РИНЦ, Web of Science и др.
- Репозиторий цифровых объектов (полные тексты статей, сборники, изображения и пр.) на основе модернизированной DSpace 5.
- Интегрирующая информационная система на основе ZooSPACE [5-6] для доступа к разнородным информационным ресурсам, в том числе сторонним.
- Публичный WEB-портал организации на основе Apache HTTP Server.

Каталог LDAP хранит публичную информацию о сотрудниках и структуре института, в том числе изображения (фотографии сотрудников) и информацию, необходимую для авторизованной работы в информационных системах ИБТ СО РАН (пароли, ключи и т.п.). Каталог функционирует на основе свободно распространяемых серверов LDAP, данные реплицируются между всеми серверами.

Информационная CRIS-система хранит информацию о сотрудниках и их публикациях, о конференциях, связанных с Институтом и сотрудниками, о проектах, выполняемых сотрудниками, об организациях, связанных с проектами, публикациями и конференциями и т.п. Вся информация хранится в СУБД на основе свободно распространяемого ПО PostgreSQL, пользовательские и административные интер-

фейсы реализованы на основе WEB-технологий (приложения PHP сервера Apache). Ввод информации осуществляется как через административные WEB-интерфейсы, так и в пакетном режиме. Допускается импорт данных из систем: РИНЦ, Web of Science, Scopus и др. Авторизованная работа в системе регулируется информацией из LDAP-каталога Института. Доступ к данным возможен не только через WEB-интерфейсы, но и по протоколам OAI-PMH, SRW/SRU, Z39.50. Последнее обеспечивается интегрированием CRIS-системы с системой ZooSPACE. Информационная CRIS-система является поставщиком соответствующей информации для публичного WEB-портала Института.

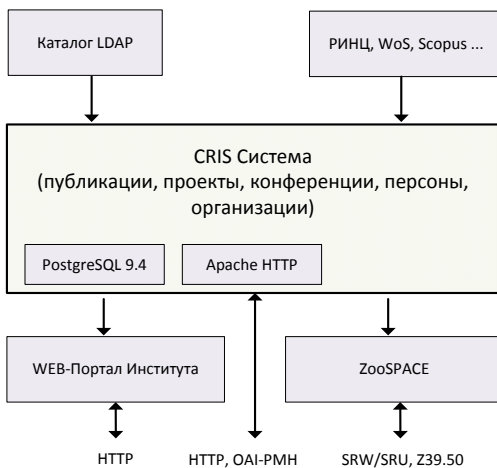


Рисунок 1 - Информационная CRIS-система

Репозиторий цифровых объектов реализован на основе известной системы DSpace (модернизированной для нужд Института) с использованием СУБД PostgreSQL и сервера Apache Tomcat. Данные репозитория структурированы в иерархические разделы и коллекции. Ввод данных осуществляется

- в интерактивном режиме через встроенные WEB-интерфейсы;
 - заимствованием данных из других систем (DOI, PubMet, ArXiv, CiNii, CrossRef и др.), в том числе из CRIS-системы Института.
 - в пакетном режиме - импорт данных в форматах DIM, MEKOF, MARC21, DC и др.;
 - синхронизацией данных по OAI-PMH с внешними источниками.
- Авторизация пользователей осуществляется через каталог LDAP.

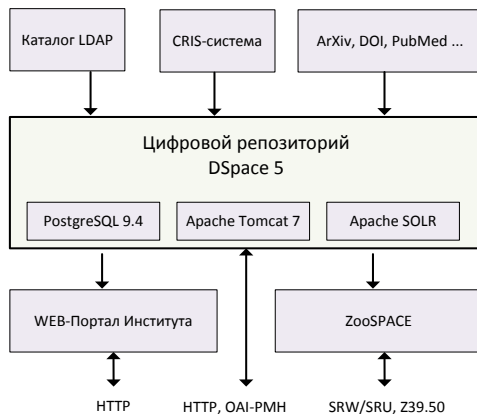


Рисунок 2 - Цифровой репозиторий

Доступ к данным репозитория возможен не только через WEB-интерфейсы DSpace, но и по протоколам OAI-PMH, SOLR, SRW/SRU, Z39.50. Последнее обеспечивается связью DSpace с системой ZooSPACE.

Для более полного соответствия локальным требованиям в базовую систему DSpace внесены многочисленные изменения (расширение схем данных, расширение номенклатуры обменных форматов, возможность работы с географической информацией [7], авторитетный контроль и пр.)

Распределенная информационная система ZooSPACE интегрирует данные из различных информационных источников, обеспечивая доступ к разнородной распределенной информации в соответствии со стандартными протоколами (SRW/SRU, Z39.50) [5-6]. Система функционирует на основе оригинальных серверов ZooPARK-ZS, серверов LDAP и WEB-серверов Apache, обеспечивая сквозной поиск информации в разнородных базах данных, извлечение информации в стандартных схемах и форматах и ее отображение. Система может интегрировать не только локальные ресурсы, но и ресурсы сторонних организаций, доступные по протоколам SRW/SRU или Z39.50.

Публичный WEB-портал Института кроме всего прочего предоставляет публичные WEB-интерфейсы для доступа к данным вышеперечисленных информационных систем и использует эти данные для формирования WEB-страниц.

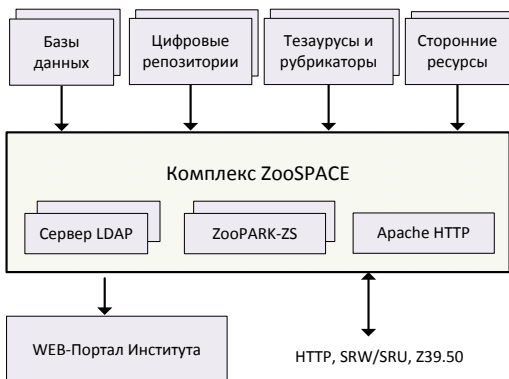


Рисунок 3 - Программный комплекс ZooSPACE

Организация корпоративных информационных систем на основе изложенных выше принципов с одной стороны защищает эти системы от политики поставщиков проприетарного программного обеспечения, а с другой - позволяет строить информационные системы по модульному принципу, используя для межмодульного обмена стандартные сетевые протоколы. Последнее делает информационные системы открытыми и, следовательно, доступными для всего сообщества.

Библиографический список

1. Елепов Б.С., Жижимов О.Л., Федотов А.М., Шокин Ю.И. Интеграция информационных ресурсов Сибирского отделения РАН как шаг к формированию единого научно-образовательного информационного пространства // Теория и практика общественно-научной информации. - 2014. - № 22. - С.21-32. - ISSN 0236-3917.
2. Шокин Ю.И., Федотов А.М., Жижимов О.Л., Федотова О.А. Система управления электронными библиотеками в ИРИС СО РАН // Инфраструктура научных информационных ресурсов и систем: Сборник научных статей Четвертого Всероссийского симпозиума / Под редакцией доктора технических наук Е.Б. Кудашева, доктора физико-математических наук В.А. Серебрякова. - 2014. - Т.1. - Москва: Вычислительный центр РАН. - С.11-39. - ISBN: 978-5-19601-103-6.
3. Guskov A.E., Zhizhimov O.L., Kikhtenko V., Skachkov D.M., Kosyakov D. RuCRIS: A Pilot CERIF based System to Aggregate Heterogeneous Data of Russian Research Projects // Procedia Computer Science. - 2014. - Vol.33. - P.163-167. - EISSN 1877-0509.
4. Chudlarsky, Tomas; Dvorak, Jan: A National CRIS Infrastructure as the Cornerstone of Transparency in the Research Domain. In: Jeffery, Keith G; Dvorak, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 9-17. ISBN 978-80-86742-33-5.

5. Жижимов О.Л., Федотов А.М., Шокин Ю.И. Технологическая платформа массовой интеграции гетерогенных данных // Вестник Новосибирского государственного университета. Серия: Информационные технологии. - 2013. - Т.11. - № 1. - С.24-41. - ISSN 1818-7900.
6. Жижимов О.Л. Новые возможности платформы ZooSPACE для задач интеграции распределенных данных [Электронный ресурс] // Распределенные информационные и вычислительные ресурсы (DICR-2014): материалы XV Российской конференции с международным участием (электронное издание). - 2014. - Новосибирск: Институт вычислительных технологий Сибирского отделения РАН. - Гос. регистр. №: 0321500379. - <http://conf.nsc.ru/files/conferences/dicr2014/fulltext/248947/250122/ZooSPACE-New1.pdf>
7. Скачков Д.М., Жижимов О.Л. Технология географического поиска информации в «негеографических» информационных системах // Научные и организационно-технологические основы интеграции цифровых информационных ресурсов . - 2013. - Санкт-Петербург: ФГБУ «Президентская библиотека имени Б. Н. Ельцина». - С.74-101. - ISBN 978-5-905273-39-1.

*Zhizhimov O.L., Fedotov A.M., Shokin Yu.I.
Institute of Computational Technologies
Siberian Branch of Russian Academy of Sciences (ICT SB RAS)
Novosibirsk*

INFORMATION SYSTEMS ICT SB RAS: PRINCIPLES, ARCHITECTURE, IMPLEMENTATIONS

We discuss the basic principles of information systems ICT SB RAS, which are a consequence of requirements imposed on them with the essential requirements. The architecture of the main elements of these systems and the rules for their interaction are considered. The implement them are described.

Key words: distributed information systems, databases, CRIS, interoperability, ZooSPACE, LDAP, SRW, SRU, Z39.50..