

ТЕХНОЛОГИИ СОЗДАНИЯ ИНТЕГРИРОВАННЫХ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ В НАУЧНЫХ ПРОЕКТАХ

Гаченко А.С., Ружников Г.М., Хмельнов А.Е.

Институт динамики систем и теории управления СО РАН (ИДСТУ СО РАН)

gachenko@icc.ru

В статье рассматривается реализация основных компонентов инфраструктуры пространственных данных для реализации междисциплинарных научных исследований. Основной целью создания компонентов является обеспечение общей платформы для совместной работы с распределенными сервисами геообработки и пространственными данными. Данная платформа реализует общие для сервисов геообработки функции, например, вызов распределенных сервисов геообработки, систему управления доступом, визуализацию данных и т.д., упрощая публикацию новых сервисов для открытого доступа.

Ключевые слова: инфраструктура пространственных данных, базовые пространственные данные, каталоги метаданных, реестры, органы государственной власти, картографические данные.

This article discusses the implementation of the basic components of spatial data infrastructure for the implementation of interdisciplinary research. The main purpose of the components is to provide a common platform for collaboration with distributed geoprocessing services and spatial data. The platform implements common services, geoprocessing functions, such as calling the distributed geoprocessing services, access control system, data visualization, etc., making it easy to publish new services for public access.

Keywords: infrastructure of spatial data, basic spatial data, catalogs of metadata, registers, public authorities, cartographical data.

Цифровые геопространственные данные активно используются в научных исследованиях и решении большого перечня задач управления территориальным развитием.

Специфика формирования геоинформационных ресурсов увеличила актуальность создания проблемно-ориентированных программных средств, интегрирующих универсальные сетевые технологии с ГИС-технологиями, поддерживающих организацию и работу с пространственно - распределенными данными, т.е. создающих интерактивную среду взаимодействия клиентских приложений с ГИС - сервером.

На начальном этапе развития картографические Web-сервера обеспечивали лишь выбор и просмотр заданного набора картинок в форматах GIF, JPEG или другом графическом формате. Первым среди подобных серверов считается National Atlas Information Service (NAIS) of Canada. Следующим этапом развития систем для Web-просмотра картографической информации было подключение функций СУБД для взаимодействия с базами данных (БД), содержащими атрибутивные данные карт. Данный подход предъявил более высокие требования к программно-аппаратному обеспечению сервера, однако, при этом повысилась эффективность его работы за счет структурированного представления картографических данных.

В институтах Иркутского научного центра (ИНЦ) СО РАН научные исследования базируются на уникальных проблемно и предметно-ориентированных базах пространственных данных по ландшафтам

и геосистемам, картографированию природы, хозяйства и населения Сибири (Институт географии СО РАН), геологической среде и сейсмическим процессам (Институт земной коры СО РАН), геохимии окружающей среды и осадочных бассейнов (Институт геохимии СО РАН), электроэнергетическим и трубопроводным систем (Институт систем энергетики СО РАН), биоразнообразию фауны и флоры оз. Байкал (Лимнологический институт СО РАН), физиологии растений, молекулярной биологии и экологии растительных организмов (Сибирский институт физиологии и биохимии растений СО РАН), дистанционному зондированию поверхности Земли (Институт солнечно-земной физики СО РАН) и т.д.

В настоящее время актуальны работы по разработке и созданию:

- интеллектуальных методов и инструментальных средств создания, анализа интегрированных распределённых информационно-аналитических и вычислительных систем с применением ГИС, GRID и Web-технологий;

- единой интегрированной инфраструктуры проблемно и предметно-ориентированных тематических баз пространственных данных институтов ИИЦ СО РАН;

- современных методов и технологий интеграции разноформатных междисциплинарных данных и результатов исследований, базирующихся на пространственных характеристиках и признаках;

- централизованного хранилища цифровой топоосновы и картографической информации с удалённым Web-доступом пользователей;

- системы сервисов геоданных СО РАН;

- новых методов и технологий исследования и обработки пространственных данных, включая разработку логических методов и методов обработки больших массивов данных;

- ГИС-портала институтов ИИЦ СО РАН;

- методов и средств планирования распределённого решения информационно-вычислительных задач;

- специальных баз проблемно-ориентированных и предметно-ориентированных географических данных и знаний для их размещения в Интернет;

- новых моделей и методов, базирующихся на результатах натурных наблюдений и на эмпирических данных, включая создание методов и технологий обработки данных дистанционного зондирования;

- новых методов и технологий анализа и обработки географических данных и знаний в интегрированных системах геоинформационного картографирования.

В рамках комплексного проекта информатизации науки и образования в ИИЦ СО РАН создана Интегрированная информационно-вычислительная сеть Иркутского научно-образовательного комплекса (ИИВС ИрНОК) с пропускной способностью магистрали до 1Gb/s, а также региональный узел доступа к Сети передачи данных (СПД) СО РАН.

В 2012 году российской компанией «Т-Платформы» на базе ИДСТУ СО РАН реализован проект создания «под ключ» гибридного вычислительного кластера, названного именем известного ученого, директора-организатора института, академика В.М. Матросова. В качестве вычислительных узлов кластера использована универсальная платформа T-Blade V-Class V205S с процессорами AMD Opteron 6276 «Interlagos». Пиковая производительность системы составляет 33,7 Тфлопс. В 16-ой редакции (от 27.03.2012) списка Топ-50 вычислительный кластер «Академик В.М. Матросов» занял 26-е место.

Суперкомпьютерный центр коллективного пользования, поддерживающий информационно-вычислительное обеспечение фундаментальных и прикладных исследований, проводимых в институтах ИНЦ СО РАН и вузах Байкальского региона.

Таким образом, в ИНЦ СО РАН создана информационно-вычислительная, телекоммуникационная инфраструктура и накоплены уникальные научные проблемно и предметно-ориентированные геопространственные данные, что служит основой формирования Центра поддержки междисциплинарных научных исследований институтов ИНЦ СО РАН.

В целом, отмечается существенный рост потребностей со стороны науки и образования в доступных информационно-вычислительных ресурсах для проведения фундаментальных и прикладных исследований. Это приводит к необходимости интеграции (в том числе, с использованием GRID-технологий) пространственно распределенных хранилищ разнородных данных и вычислительных мощностей, организации коллективного доступа к ним.

В связи с этим большой интерес представляют технологии, позволяющие автоматизировать процесс создания комплексных тематических информационных систем.

В настоящее время в Иркутском научном образовательном комплексе (ИРНОК) проводятся работы по формированию и ведению баз данных поддержки междисциплинарных научных исследований Байкальской природной территории на основе материалов, которые собраны в институтах Иркутском научном центре (ИНЦ) СО РАН.

Разработан и наполняется тематическими и пространственными данными **геопортал** по биоразнообразию Байкальской природной территории.

Основной целью создания геопортала является обеспечение общей платформы для совместной работы с распределенными сервисами геообработки и пространственными данными. Эта платформа реализует систему управления доступа к географическим объектам на карте. Разработаны специализированные сервисы, обеспечивающий следующие функции обработки пространственно-распределенных данных:

- хранение геоданных;
- публикация картографических и реляционных данных;
- поиск ГИС-ресурсов по каталогу геопортала.

Наряду с данными сервисами связанными с геообработкой баз данных существует необходимость в хранении и обмене текстовыми документами (заявки, отчеты, интеграционные проекты).

В Сибирском отделении РАН в рамках развития этого направления разрабатывается **корпоративный каталог**. В его задачи входит:

- каталогизация информационных ресурсов и сетевых сервисов;
- обеспечение единой инфраструктуры аутентификации и авторизации пользователей информационных систем СО РАН и ДВО РАН;
- обеспечение глобальных политик доступа к информационным ресурсам;
- мониторинг доступности информационных ресурсов и сетевых сервисов;

- создание автоматически актуализируемой распределенной справочной системы СО РАН и ДВО РАН;
- обеспечение работы совместных и частных распределенных информационных систем.

Корпоративный каталог разрабатывается как централизованно-распределенная информационная система с разграничением зон ответственности между центрами и организациями отделений РАН. Корпоративный каталог создается на основе семейства LDAP-серверов (Lightweight Directory Access Protocol).

Разрабатываются компоненты инфраструктуры взаимодействия на уровне серверов обмена данными и метаданными. Ведется работа по взаимодействию между серверами на основе LDAP — «облегченный протокол доступа к каталогам» протокол прикладного уровня для доступа к службе каталогов. LDAP — это относительно простой протокол, использующий TCP/IP и позволяющий производить операции авторизации (bind), поиска (search) и сравнения (compare), а также операции добавления, изменения или удаления записей. Обычно LDAP-сервер принимает входящие соединения на порт 389 по протоколам TCP или UDP. Для LDAP-сеансов, инкапсулированных в SSL (англ. (SSL)Secure Sockets Layer — уровень защищённых сокетов)—криптографический протокол, который обеспечивает установление безопасного соединения между клиентом и сервером. На основании протокола SSL 3.0 был разработан и принят стандарт RFC, получивший имя TLS. Протокол обеспечивает конфиденциальность обмена данными между клиентом и сервером, использующим TCP/IP.

Каталоги такого рода, как правило, содержат статические и редко изменяемые элементы, так как каталоги изначально оптимизированы для очень быстрого отклика на запросы поиска и чтения данных.

Такие каталоги полностью структурированы. Каждый элемент данных имеет имя, которое одновременно определяет положение элемента в иерархии каталога. Каждый атрибут элемента, как правило, может иметь несколько значений и это является нормальным поведением, в отличие от обычных баз данных.

Каталоги являются очень специфическими системами хранения данных. Их удобно использовать для иерархически скомпонованных объектов. Каталоги могут быть реплицированы между несколькими серверами, для организации удобного доступа и распределения нагрузки. Текстовая информация хорошо подходит для каталогов, так как легко поддается поиску, но данные могут быть представлены и в любой другой форме.

Основная цель разрабатываемого сервиса направлена на взаимодействие между каталогами Дальнего востока, Иркутска и Новосибирска. Корпоративный каталог отделений РАН создается для поддержки различных сетевых сервисов и приложений.

Разработаны компоненты инфраструктуры взаимодействия на уровне серверов данных и метаданных. Проведена работа по отработке взаимодействия между серверами на основе протоколов LDAP и Z39.50.

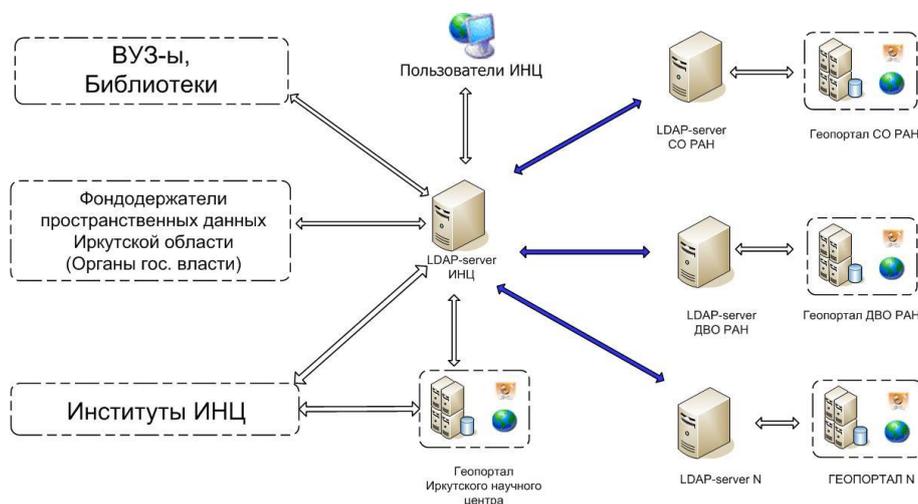


Рис. 1. Структура взаимодействия LDAP-серверов с банками данных фондодержателей Иркутской области в ИДСТУ СО РАН на основе программно-аппаратной платформы, включающей систему хранения данных (СХД) SAN ReadyStorage 3994 развернут периферийный LDAP-сервер на специально выделенной для этих задач виртуальной машине. На сервере виртуальных машин выделены необходимые ресурсы под задачи связанные с процессами репликации данных между LDAP-серверами СО РАН и ДВО РАН. Существует центральный сервер каталогов, хранящий сводную информацию. Информация на центральном сервере изменяется в результате непосредственного редактирования, а также в результате репликаций с периферийных серверов. Каждый периферийный сервер может хранить все дерево корпоративного каталога или его часть. Основной узел каталога РАН находится на сервере <http://elib.sbras.ru:8080/jsrui/> и доступен по протоколу LDAP порт 1389 (LDAPS порт 1636). На LDAP-сервере ИДСТУ СО РАН развернут узел ZooSPACE разрабатываемый в ИВТ СО РАН.

LDAP-сервер функционирует в Интегрированной информационно-вычислительной сети (ИИВС) Иркутского регионального научно-образовательного центра, объединяющей в своей инфраструктуре локально-вычислительные сети институтов Иркутского научного центра, государственных высших учебных заведений г. Иркутска, а также учреждений ВСНЦ РАН. ИИВС ИРНОК работает на базе высокоскоростной оптоволоконной инфраструктуры. Основной технологией передачи данных в сети на канальном уровне является GigabitEthernet. Некоторые участники сети ввиду своей удаленности от опорных узлов подключены через высокоскоростные каналы связи, арендуемые у коммерческих провайдеров. Связность с российскими и зарубежными глобальными сетями обеспечивается внешним каналом связи, пропускной способностью 120 Мбит/сек.

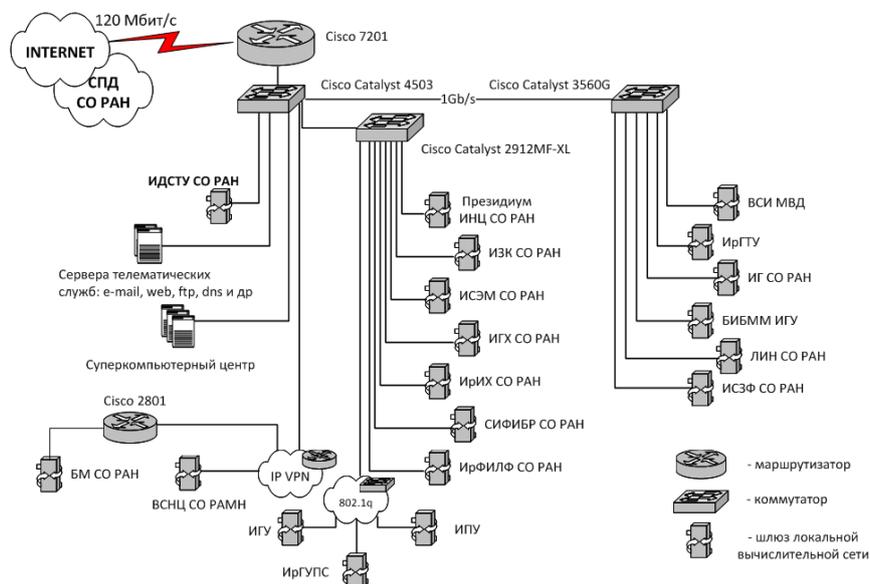


Рис. 2 ИИВС ИРНОК

Данная телекоммуникационная инфраструктура позволяет оперативно подключать новые научные организации для эффективного информационного обмена при проведении научных исследований.

Например, в 2013 году проведены работы по установке библиотечной каталожной распределенной системы IRBIS64 на выделенный для этой цели сервер, который обеспечивает поддержку библиотечных ресурсов ИИЦ.

Сервер IRBIS64 предназначен для осуществления доступа пользователей Интернет к электронным каталогам собственных библиотек и сторонним библиографическим базам данных распределенной системы автоматизации библиотек IRBIS. Система IRBIS представляет собой типовое интегрированное решение в области автоматизации библиотечных технологий и предназначена для использования в библиотеках любого типа и профиля для использования в качестве одного из основных компонентов библиотечных Интернет-серверов и Интернет-комплексов. Система полностью отвечает международным требованиям, предъявляемым к таким системам, и поддерживает все отечественные библиографические стандарты и форматы.

Базовые операции IRBIS64:

Поиск в произвольной базе данных, имеющей структуру IRBIS64 по неограниченному числу полей, по любым элементам описания и их комбинаций, с применением логики «И», «ИЛИ» и «ФРАЗА ЦЕЛИКОМ», с возможностями определения префиксов и квалификаторов поисковых терминов, грамматической нормализации слов русского языка и применения аппарата усечений.

Уточняющий поиск в результатах предыдущего поиска по условию (последовательный поиск).
Сортировка результатов поиска по условиям.

Использование при поиске статических словарей и рубрикаторов, включенных в поисковые формы, с возможностью комбинирования элементов словарей с любыми другими поисковыми предписаниями.

Использование динамических словарей баз данных, с возможностью получения списка терминов словаря и последующего поиска по выбранным терминам; навигация по словарям, включая задание начала сканирования словарей по первым символам, а также в терминах «следующие»-«предыдущие».

Показ найденных записей в стандартных форматах, включая информационный и в виде каталожной карточки.

Данные загружены в региональный информационный узел корпоративного каталога, количество записей в библиотечном каталоге составляет порядка 80 000.

Заключение

В статье изложен подход для организации инфраструктуры и создания интегрированных информационно-аналитических систем, а также описаны области применения систем подобного рода. Изложены оригинальные подходы и методы по их созданию.