

УПРАВЛЕНИЕ ДАННЫМИ ЭКСПЕРИМЕНТОВ С ИСПОЛЬЗОВАНИЕМ WEB-ТЕХНОЛОГИЙ

А.П. Григорюк, Л.П. Брагинская
and@opg.sccc.ru, ludmila@opg.sccc.ru

В статье рассматриваются концептуальные основы, архитектура и программное обеспечение web-ориентированной информационно-вычислительной системы для управления структурированными и неструктурированными данными натуральных и вычислительных экспериментов. Система обеспечивает пользователей поисковым, вычислительно-аналитическим и ГИС сервисами для эффективной работы с данными.

В качестве примера практической реализации предложенной архитектуры представлена разработанная авторами доклада информационно-вычислительная система (ИВС) «Вибросейсмическое просвечивание Земли» для управления данными, полученными в ходе экспериментов по активному вибросейсмическому мониторингу. При разработке ИВС был применен ряд достаточно оригинальных решений, позволивших создать систему, ориентированную на работу в сетях intranet/internet с данными в виде числовых рядов большой длины. Данные именно такого типа регистрируются в ходе многих экспериментов в области как геофизики, так и в других областях науки и техники.

Предметная область.

В последние десятилетия широкое развитие получили методы активного геофизического мониторинга, основанные на исследовании отклика литосферы на искусственные воздействия различной физической природы. Определение основных закономерностей изменения геофизических полей, выделение их аномалий является основой решения широкого спектра прогностических задач, способствует лучшему пониманию природы происходящих в литосфере геодинамических процессов.

Одним из направлений таких исследований является метод активной сейсмологии, в котором в качестве источника сейсмических волн используются мощные механические вибраторы. Значительная возмущающая сила, развиваемая источником (до 100 тс), длительные (до нескольких часов) сеансы излучения и прецизионные системы управления и регистрации обеспечивают наблюдаемый волновой эффект (сейсмограммы), сравнимый со средним землетрясением. Проведенные полевые эксперименты показали возможность получения качественных сейсмограмм на расстояниях до 500 км и регистрацию гармонических сигналов на расстояниях до 2000 км [1].

К безусловным достоинствам метода можно отнести контролируемость и повторяемость процесса излучения и полную экологическую безопасность, которая позволяет вести экспериментальные работы даже вблизи населенных пунктов.

Институт вычислительной математики и математической геофизики СО РАН участвует, начиная с 1995 г, в экспериментальных работах по активному мониторингу литосферы в сейсмоопасных районах России. В ИВМиМГ СО РАН в рамках экспериментальных исследований проводился вибросейсмический мониторинг литосферы в Алтае-Саянском регионе, Байкальской рифтовой зоне и Таманской грязевулканической провинции. Разработанные в СО РАН методы активной сейсмологии с 2003 года начали развивать и японские сейсмологи, изготовившие сеймовибраторы собственной конструкции.

Начиная с 2001 г. ИВМиМГ СО РАН проводит экспериментальные работы по регистрации карьерных взрывов в Кузбассе.

В последние годы к методам активной сейсмологии так же относят экспериментальные работы, в которых регистрируется сейсмическое поле природных источников по технологии, определенной конкретной задачей геофизики, например сейсмоэмиссионная томография вулканических структур с использованием сейсмического шума из активной области вулкана. К таким работам можно отнести экспериментальные работы ИВМиМГ СО РАН по мониторингу физических полей вулканических структур, которые были проведены в туннеле вспомогательной штольни Баксанской нейтринной обсерватории в 2010 г.

В результате полевых работ в ИВМиМГ был накоплен уникальный архив данных экспериментов (около 40), включающий архив файлов волновых форм (более 30000 сейсмотрасс) и сопутствующей информации (тип сейсмического источника, параметры излучаемого им сигнала, параметры регистратора, географические координаты источника и регистратора и т.д.) Но архив не был структурирован, файлы волновых форм имели различный формат, определявшийся различными типами применяемых регистраторов, сопутствующая информация хранилась в произвольной форме.

Модель данных

Организация структурированных табличных (реляционных) данных обычно затруднений не вызывает, для этого идеально подходят реляционная модель и язык SQL, лежащие в основе современных СУБД. Сложнее обстоит дело с файлами волновых форм, представляющими собой n-мерные, в общем случае, массивы, которые не могут быть структурированы и поэтому не поддерживаются реляционными СУБД непосредственно. Для работы одновременно с реляционными и нереляционными данными, в основном используют одну из двух архитектур:

- как реляционные, так и нереляционные данные находятся в базе данных;
- реляционные данные находятся в базе данных, а нереляционные данные – в файловых системах или на файловых серверах.

Каждый из этих двух подходов имеет свои преимущества и недостатки [2]. В первом случае одна база данных становится удобным централизованным хранилищем для обоих типов данных. Однако нереляционные данные хранятся в формате больших двоичных объектов (BLOB), скорость доступа к этим объектам существенно уступает скорости доступа к файлам. Во втором случае обеспечивается высокая скорость доступа, но усложняется разработка приложений и управление ими, так как приложения должны поддерживать согласованность между записями в базе данных и файлами, связанными с этими записями. Данную проблему можно частично или полностью решить за счет модели данных, обеспечивающей эффективную индексацию файловой системы из базы данных.

При построении концептуальной модели предметной области мы исходили из того, что экспериментально изучаемому объекту может быть приписан определенный набор параметров, соответствующий представлениям исследователей о состоянии и поведении этого объекта. Параметры это то, что можно измерять, наблюдать и изменять в процессе исследований. В процессе экспериментов на изучаемый объект воздействуют некоторые факторы с контролируемыми параметрами, а с помощью сенсоров регистрируется ряд параметров объекта при фиксированных других параметрах. Тогда концептуальную модель эксперимента можно представить с помощью приведенной на Рис.1 ER-диаграммы (диаграмма «сущность-связь») [3].

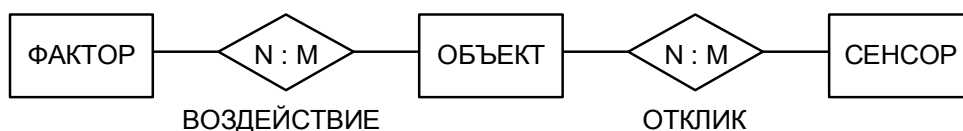


Рис. 1. Концептуальная модель эксперимента

Диаграмма содержит три класса сущностей: ОБЪЕКТ, ФАКТОР и СЕНСОР, каждый класс с набором атрибутов, определяемым конкретной областью исследований. Взаимоотношения сущностей выражаются двумя классами связей: ВОЗДЕЙСТВИЕ и ОТКЛИК. В процессе экспериментов всевозможным сочетаниям экземпляров сущностей ОБЪЕКТ и ФАКТОР ставятся в соответствие экземпляры сущности ОТКЛИК. В случае пассивного эксперимента или наблюдения класс сущностей ФАКТОР может отсутствовать.

В соответствии с [3] для перехода к реляционной модели данных заменим сущности и связи ER-диаграммы на соответствующие отношения R с первичными ключами K и атрибутами A:

ОБЪЕКТ – $R1(\underline{K1}, A_{11}, A_{12}, \dots, A_{1m});$

ФАКТОР – $R2(\underline{K2}, A_{21}, A_{22}, \dots, A_{2n});$

СЕНСОР – $R3(\underline{K3}, A_{31}, A_{32}, \dots, A_{3p});$

ВОЗДЕЙСТВИЕ – $R4(K1, K2);$

ОТКЛИК – $R5(K1, K3).$

Вспомогательные отношения R4 и R5 служат для организации связи типа M:N между отношениями R1, R2 и R1, R3 соответственно. Первичными ключами K могут служить, например, номера (индексы) кортежей соответствующих отношений. Такая модель данных позволяет организовать адресацию файлового архива, имеющего следующую иерархическую структуру:

root / K1 / K2 / K3 / <файлы данных>

Данная структура полностью соответствует естественной древовидной структуре файловой системы. Результирующее отношение, формируемое в результате выполнения запроса, предполагающего обращение к файловому архиву, должно содержать набор ключевых атрибутов K1, K2, K3. Таким образом, пользователи полностью абстрагируются от имен или шаблонов имен файлов/каталогов, работая только с атрибутами, каталогизирующими свойства и происхождение каждого файла.

Как и модель эксперимента, модель данных является обобщенной, ее необходимо адаптировать для каждой конкретной научной области и вида экспериментов. Например, если исследуемые объекты достаточно сложны, целесообразно будет представить их в виде системы более простых объектов. На модели это можно проделать с помощью декомпозиции отношения ОБЪЕКТ в соответствии с правилами реляционной алгебры Кодда.

Управление экспериментальными данными

На основе предложенной модели данных нами была разработана база данных и веб-ориентированная информационно-вычислительная система «Вибросейсмическое просвещение Земли» (<http://org.sscs.ru/db>) [4], которая позволила организовать эффективное взаимодействие сотрудников виртуальной лаборатории через Интернет и обеспечить их основными научными сервисами.

ИВС реализует следующие основные функции:

- получение подробной информации по любому из проведенных экспериментов (метаданные);
- поиск в базе данных одновременно по 18 параметрам вибропросвечивания (типы источников, вид и параметры сигналов, географические координаты и др.), поисковая форма системы показана на Рис. 2;
- интерактивный on-line анализ (корреляционный, спектральный, спектрально-временной и т.д.) найденных сейсмотрасс с отображением результатов непосредственно в веб-браузере пользователя;
- построение по результатам поиска интерактивных карт и спутниковых снимков с обозначенными источниками и регистраторами сейсмических волн.

Рис. 2. Поисковая форма ИВС

Структурная схема информационно-вычислительной системы приведена на Рис. 3. Пользователи взаимодействуют с системой с помощью стандартного веб-браузера, посылая запросы на поиск и анализ данных. В запросе на поиск указываются интересующие пользователя значения параметров исследуемого объекта, воздействующих на объект факторов и параметры сенсоров, регистрирующих данные. Запрос на анализ должен содержать перечень процедур анализа, которые будут применены к найденным данным и параметры этих процедур.

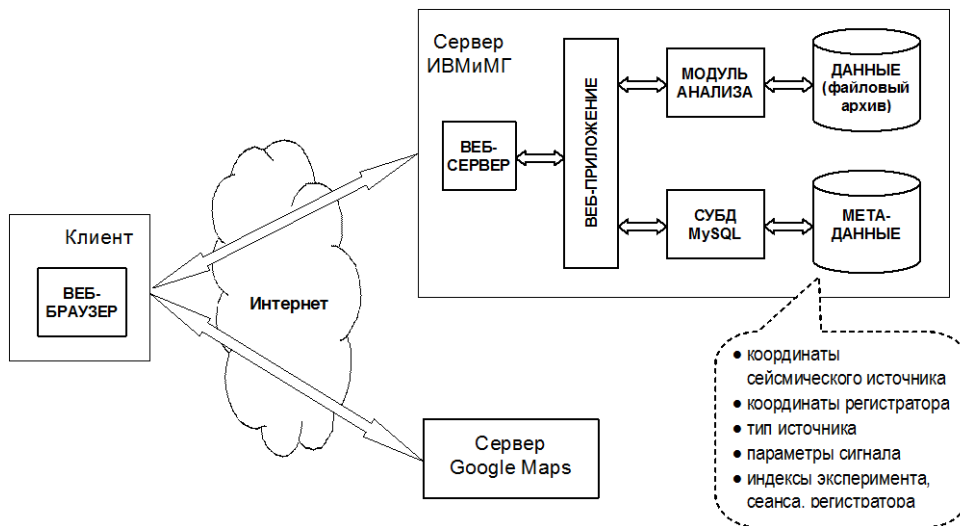


Рис. 3. Структурная схема ИВС

В результате выполнения запросов данные извлекаются из файлового архива и передаются модулю анализа. Это приложение, выполняющее анализ данных в соответствии с алгоритмами, применяемыми в конкретной области экспериментальных исследований. Для обеспечения достаточного для online режима быстродействия приложение написано на языке C++ с использованием программных библиотек Intel Performance Libraries [5]. Полученные в результате анализа числовые массивы возвращаются веб-приложению, которое «на лету» формирует графики, таблицы, текст и отправляет всё это пользователю в виде готовой веб-страницы. Пример веб-страницы с результатами поиска и анализа данных приведен на Рис. 4.

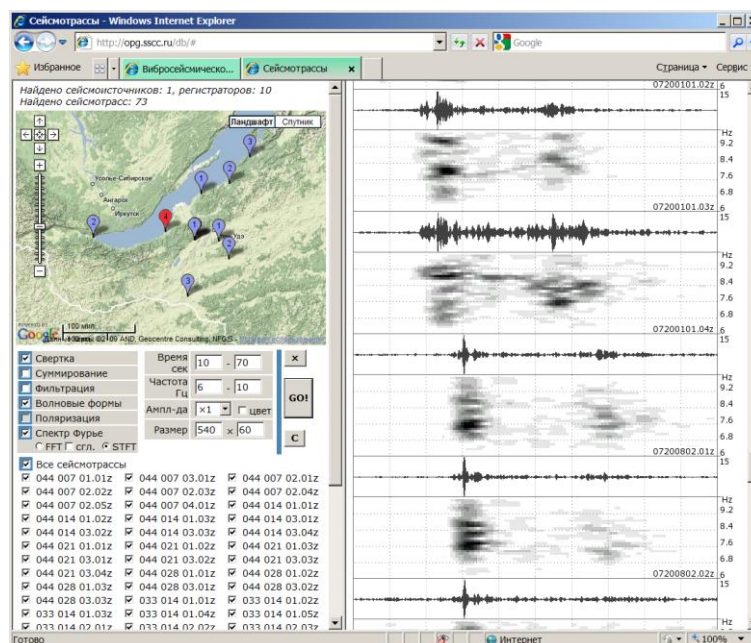


Рис. 4. Отображение результирующей страницы в веб-браузере

В геофизике как и во многих других областях исследователи имеют дело с пространственно обусловленными данными или геопространственными данными. Поэтому архитектура ИВС предусматривает подсистему управления геоданными и

картографическую подсистему на основе сервиса Google Maps [6]. Сервис базируется на данных дистанционного зондирования (спектрональные снимки Landsat, SPOT, Quickbird с разрешением до 0.68м) совмещенных с топографическими картами в проекции Меркатора. Компания Google предоставляет пользователям интерфейс Google Maps API в виде классов объектов JavaScript для генерации карт и нанесения на них собственных маркеров, контуров, а также готовых слоев в формате KML. Данные для отображения могут находиться как непосредственно в коде веб-страниц, так и во внешних XML и KML файлах. Использование технологии AJAX обеспечивает обновление содержимого результирующей веб-страницы без ее перезагрузки, таким образом изменение масштаба и перемещение по карте осуществляется без каких-либо задержек. Схема взаимодействия сервера ИВС, сервера Google Maps и клиентского браузера показана на Рис. 5. Пример карты можно видеть на Рис. 4.

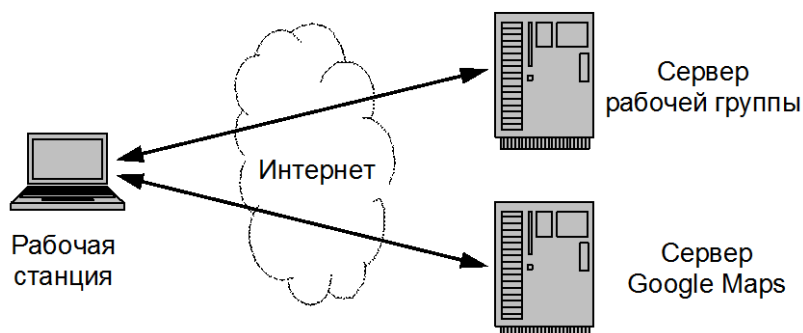


Рис. 5

ЛИТЕРАТУРЫ:

1. Активная сейсмология с мощными вибрационными источниками / Отв.ред. Г.М. Цибульчик. –Новосибирск: ИВМиМГ СО РАН, Филиал «Гео» Издательства СО РАН, 2004.
2. Электронная документация по SQL Server 2008. Управление неструктурированными данными.
http://msdb.ru/Downloads/SQL2008/white_papers/UnstructuredData%20RU.docx
3. Дейт К. Д. Введение в системы баз данных. 7-е изд. – М.: Вильямс, 2001. – 702 с.
4. Григорюк А.П., Брагинская Л.П. Управление данными вибросейсмического мониторинга. // Мониторинг окружающей среды, геоэкология, дистанционные методы зондирования Земли и фотограмметрия. Сб. материалов междунар. науч. конгресса «ГЕО-Сибирь-2007» Т.3. –Новосибирск: СГГА, 2007.
5. Библиотеки Intel Performance Libraries.
<http://www.intel.com/cd/software/products/emea/rus/perflib/358868.htm>
6. Google Maps API Documentation. <http://www.google.com/apis/maps/documentation>