

ИНФОРМАЦИОННАЯ СИСТЕМА ХРАНЕНИЯ И ОБРАБОТКИ ОКЕАНОЛОГИЧЕСКИХ ДАННЫХ*

Аннотация. Представлена общая архитектура и особенности реализации информационной системы хранения и обработки данных натуральных экспериментов по изучению волновой динамики на разнесенной сети станций на шельфах дальневосточных морей России.

Ключевые слова: информационная система, база данных, экспериментальные наблюдения, волновые процессы.

С целью комплексного изучения волновых процессов в шельфовой зоне Охотского и Японского морей, сотрудниками кафедры «Прикладная математика» Нижегородского государственного технического университета и лаборатории «Волновой динамики и прибрежных течений» Института морской геологии и геофизики ДВО РАН, на протяжении нескольких лет проводились натурные эксперименты [1 – 5], в результате которых удалось собрать огромное количество качественных данных о придонной температуре и давлении. Основным инструментом измерений стал автономный регистратор волнения [6], основанный на измерении пульсации давления под водой, индуцированной поверхностными волнами. За несколько лет эксплуатации прибор зарекомендовал себя как инструмент, обладающий высокой надежностью, простотой обращения, качеством получаемых данных и относительной дешевизной. Результатом работы прибора на протяжении пяти с половиной месяцев является текстовый файл, содержащий время, придонное давление и температуру с дискретностью одна секунда, размер одного файла составляет около 700 мегабайт. За годы проводимых натуральных наблюдений скопился очень большой объем данных (порядка 170 гигабайт), с которыми довольно затруднительно работать, поскольку приходится использовать различные программы и приложения для обработки результатов. Эти программы громоздки и медлительны, что в итоге служит серьезным сдерживающим фактором при анализе данных. Все вышеизложенное свидетельствует о необходимости разработки информационной системы, позволяющей, в первую очередь структурировано хранить данные и предоставлять к ним удобный доступ, а также проводить типовую обработку океанологических данных.

В настоящее время наиболее известными российскими базами океанографических данных являются база ТОИ ДВО РАН «ОКЕАН-1» и «ОКЕАН-2» и база Института океанологии им. П.П. Ширшова РАН. Однако эти базы ориентированы на другие задачи и не позволяют выполнять требуемые операции, такие как получение необходимого временного ряда за заданный промежуток времени и проведение типовой обработки.

Таким образом, для решения, описанных выше задач была разработана информационная система хранения и обработки данных, полученных в результате натуральных измерений волнения, названная OceanData. С точки зрения информационных технологий OceanData представляет собой сложный расчетный программный комплекс с модульной архитектурой. В качестве основных единиц декомпозиции продукта можно выделить следующие:

* © Кузнецов К.И., Куркин А.А.

Работа выполнена при поддержке гранта Президента РФ для государственной поддержки молодых российских ученых – докторов наук (МД-99.2010.5) и гранта РФФИ 10-05-00199а.

- блок хранения данных;
- пользовательский интерфейс представленный в виде сайта;
- блок обработки данных (набор оптимизированных программ);
- блок управления запросами пользователя (набор php-скриптов).

Архитектура системы

В основе архитектуры системы лежит клиент-серверная технология. Соответственно, выделяют клиентскую и серверную стороны приложения. Клиентская сторона приложения функционирует на рабочем месте пользователя, в роли которого выступает браузер персонального компьютера. Серверная сторона функционирует на специализированном комплексе, включающем в себя мощные аппаратные средства, требуемый набор стандартного программного обеспечения, систему управления базами данных и собственно структуры данных. На рис. 1 изображено послойное представление архитектуры системы.

Взаимодействие клиентской и серверной частей приложения осуществляется через сеть – локальную или глобальную. При этом с точки зрения клиента и сервера взаимодействие осуществляется прозрачно, соответственно сетевой компонент включает в себя совокупность необходимого сетевого оборудования, набор программных технологий, обеспечивающих передачу данных между узлами сети, а также собственно протокол или протоколы для обмена запросами и результатами их выполнения.

Система состоит из трех слоев или компонент: Компонент ввода-вывода, компонент прикладной логики и компонент хранения базы данных. При этом компонент прикладной логики находится на промежуточном слое, который является клиентом для базы данных и сервером для пользователя. На стороне пользователя выполняются только операции визуализации и ввода-вывода данных, а всю прикладную логику реализует сервер. Обмен между клиентом и сервером в таких системах осуществляется на уровне команд вывода данных на экран и результатов пользовательского ввода.

Пользователь, набрав в браузере адрес системы <http://hydrolab.nnov.ru/>, получает доступ к пользовательскому интерфейсу системы. Удобный интерфейс позволяет ему сформировать нужные запросы к базе, не требуя при этом специальных знаний. Запрос отправляется на сервер системы, где его обрабатывает php-скрипт. Скрипт проводит интерпретацию и проверку полученных данных, после чего запускает оптимизированное приложение с нужными параметрами. Приложение, общаясь напрямую с базой данных, записывает или читает данные из базы и проводит над ними процедуру обработки.

Далее представлены некоторые особенности реализации отдельных модулей программного обеспечения.

Особенности реализации

Хранение и управление данными осуществляет СУБД MySQL. Основная таблица базы данных – таблица паспорта эксперимента, в которой хранится полная информация о месте и времени проведения эксперимента, а также об элементах датчика, с которого считана информация. С каждой записью в этой таблице ассоциирована таблица, содержащая следующие атрибуты: давление, температуру и время записи.

Поскольку основной задачей системы является структурированное хранение данных об экспериментах и удобный доступ к ним, основными функциями являются импорт и экспорт из базы данных. Остановимся подробнее их реализацию. При импорте данных скорость не так важна, поскольку эта операция используется сравнительно редко. При экспорте скорость обработки играет, куда большую роль. По запросу пользователя программа обращается к базе данных, запрашивая нужный временной ряд. Полученный ряд, сглаживается заданным пользователем окном с помощью алгоритма Кайзера-Бесселя,

потом записывается в текстовый файл, который сжимается архиватором, затем пользователю выдается ссылка на скачивание файла. Поскольку объем данных, которыми приходится часто оперировать достаточно большой (средний размер запрашиваемого файла с данными около 60 мегабайт), необходимо максимально оптимизировать работу с ними. Выборка нужного временного интервала происходит с помощью программы, реализованной на C++. При этом дополнительно используются оптимизированные библиотеки Intel Performance Primitives, позволяющие сократить время выполнения некоторых алгоритмов в десять раз по сравнению с неоптимизированной их реализацией. Существенно снижается также время выполнения операции экспорта из базы данных при записи файла не на жесткий диск, а в оперативную память, так как скорость записи в ПЗУ в десятки раз быстрее. Этому способствует специальная файловая система tmpfs, монтируемая на определенную папку, в которую и будет записываться файл. Данная файловая система позволяет резервировать в оперативной памяти место, определенного объема, после заполнения которого скачиваемые файлы автоматически начинают записываться на жесткий диск.

Общение между сервером и клиентами происходит через php-скрипты. С каждой операцией и методикой ассоциирован свой php-скрипт, проводящий проверку введенных пользователем данных и запускающий оптимизированную процедуру обработки. Подобный подход позволяет легко расширять систему, поскольку для добавления новой методики не требуется изменения уже существующих скриптов и программного кода.

Следующая задача – обзор и анализ данных. Для нее реализована возможность графического представления данных об экспериментах. Это в первую очередь отображение мест постановок датчиков на карте, а также построение графиков давления и придонной температуры. Для отображения данных на карте используется продукт компании Google – Google-Maps, а также средства для разработчиков, позволяющие использовать в своих приложениях этот продукт – API Maps. Отображение на карте позволяет оценить расстояние между датчиками, а также дает возможность оценить пространственную структуру процессов, зарегистрированных датчиками.

Для построения графиков давления и температуры использована программа gnuplot, считывающая данные из файла, находящегося под управлением файловой системы tmpfs, описанной выше. Gnuplot несколько лет занимает лидирующие позиции среди программ построения графиков и используется огромным количеством организаций. В результате работы программы пользователю предоставляются качественные графики, практически готовые к публикации. При построении графиков используются только те данные, которые возможно отобразить на изображении заданного разрешения, в результате время выполнения снижается в несколько раз без потери качества графика.

Пользовательский интерфейс представляет собой сайт, находящийся под управлением «Joomla! CMS». Выбор именно этой системы управления сайтами обусловлен тем, что на сегодняшний день эта CMS является наиболее гибкой и надежной. Ее распространение повлекло за собой появление огромного количества дополнений и плагинов для этой системы, которые позволяют легко добавлять новые функциональные возможности на сайт.

Интерфейс реализован с применением технологии AJAX (*Asynchronous Javascript and XML*), позволяющей производить запросы на сервер без перезагрузки всей страницы. На рис. 2 представлен процесс построения графика нужного временного ряда.

Интерфейс предоставляет возможность авторизации пользователей и предоставление им доступа к различным функциям системы. Незарегистрированные пользователи смогут посмотреть расположение датчиков на карте и построить график нужного временного ряда. Функция экспорта из базы данных и добавления новых экспериментов доступны только авторизованным пользователям.

Заключение

В статье представлена общая архитектура и особенности реализации системы хранения и обработки данных натурных экспериментов. Подобная система позволяет структурировать и упорядочить собранные океанологические данные. Удобный доступ с помощью пользовательского интерфейса существенно упрощает работу с ними. Дополнительный плюс системы в том, что доступ к данным возможен с любого компьютера и из любой точки мира, при этом не требуется дополнительного программного обеспечения. Система позволяет пользователю получить данные с любой дискретностью в виде тестового файла с выбранными рядами данных. С такими файлами работает любая программа обработки данных. Подобный подход предоставляет пользователю абсолютную свободу в выборе программного инструмента для дальнейшей работы. Представление мест постановок на карте позволяет проводить пространственный анализ волновых процессов. Большое внимание при разработке системы было уделено скорости обработки, а для максимальной эффективности, применялись последние разработки компании Intel в этой области.

В настоящий момент описанная информационная система прошла тестирование и полностью функционирует, и активно используется сотрудниками НГТУ и ИМГиГ ДВО РАН при проведении натурных экспериментов на шельфе дальневосточных морей России. Дальнейшая работа будет связана с наполнением базы результатами новых экспериментов, добавлением возможности хранения метеоданных и наращиванием функционала системы.

СПИСОК ЛИТЕРАТУРЫ.

1. Чернов А.Г., Куркин А.А., Лухнов А.О., Полухина О.Е., Ковалев П.Д., Ковалев Д.П. Натурный эксперимент по регистрации захваченных волн в районе м. Острый (Охотское море) // Известия АИН им. А.М. Прохорова. Прикладная математика и механика. 2006. Т. 18. С. 91 – 98.
2. Ковалев П.Д., Куркин А.А., Шевченко Г.В., Чернов А.Г., Лухнов А.О. Исследование особенностей гидродинамических условий прилегающего к озеру Изменчивое участка взморья // Известия АИН им. А.М. Прохорова. Прикладная математика и механика. 2007. Т. 20. С. 8 – 16.
3. Ковалев П.Д., Куркин А.А., Лухнов А.О., Полухина О.Е., Чернов А.Г. Организация и проведение натурального эксперимента по изучению волновой динамики на восточном шельфе острова Сахалин // Известия АИН им. А.М. Прохорова. Прикладная математика и механика. 2007. Т. 20. С. 33 – 39.
4. Зайцев А.И., Ковалев Д.П., Куркин А.А., Левин Б.В., Пелиновский Е.Н., Чернов А.Г., Ялчинер А. Невельское цунами 2 августа 2007 года: инструментальные данные и численное моделирование // ДАН. 2008. Т. 421. № 2. С. 249 – 252.
5. Левин Б.В., Чернов А.Г., Шевченко Г.В., Ковалев П.Д., Ковалев Д.П., Куркин А.А., Лихачева О.Н., Шишкин А.А. Первые результаты регистрации длинных волн в диапазоне периодов цунами в районе Курильской гряды на разнесенной сети станций // ДАН. 2009. Т. 427. № 2. С. 239 – 244.
6. Куркин А.А., Лухнов А.О., Чернов А.Г., Полухина О.Е. Проблемы создания аппаратно-программного комплекса для исследования гидродинамики шельфовой зоны // Известия АИН им. А.М. Прохорова. Прикладная математика и механика. 2006. Т. 18. С. 120 – 123.
7. Харрингтон Д. Разработка баз данных. «ДМК Пресс» - 2005 - 269 с.
8. Дуглас Камер, Дэвид Л. Стивенс Сети TCP/IP, Том 3. Разработка приложений типа клиент/сервер для Linux/POSIX. «Вильямс» - 2002 - 592 с.

K. Kuznetsov, A. Kurkin

INFORMATION SYSTEM FOR STORING AND PROCESSING OCEANOLOGICAL DATA

Abstract. The article is devoted to the problem of developing a software for organized storing and processing observed data to study the wave dynamics on a distributed network of stations on the shelves of the Far Eastern seas of Russia.

Key word: information system, data base, in-situ measurements, wave process.

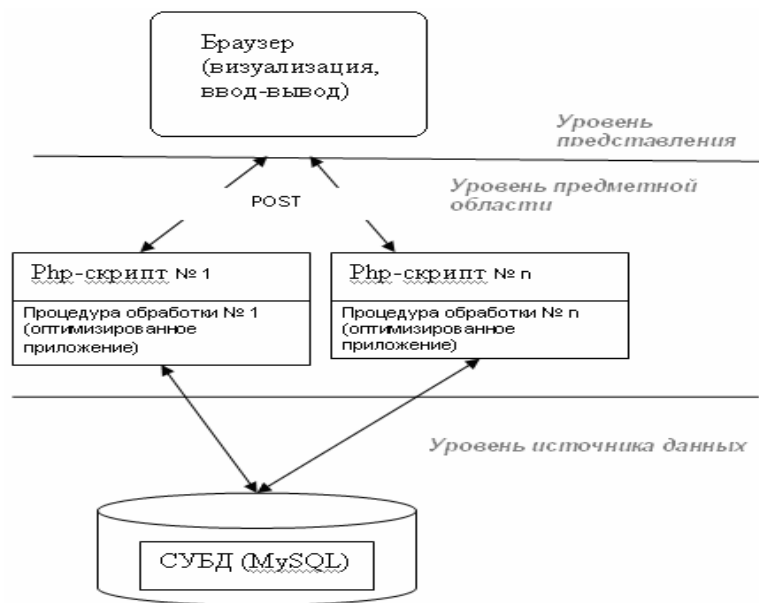


Рис. 1. Послойное представление архитектуры системы

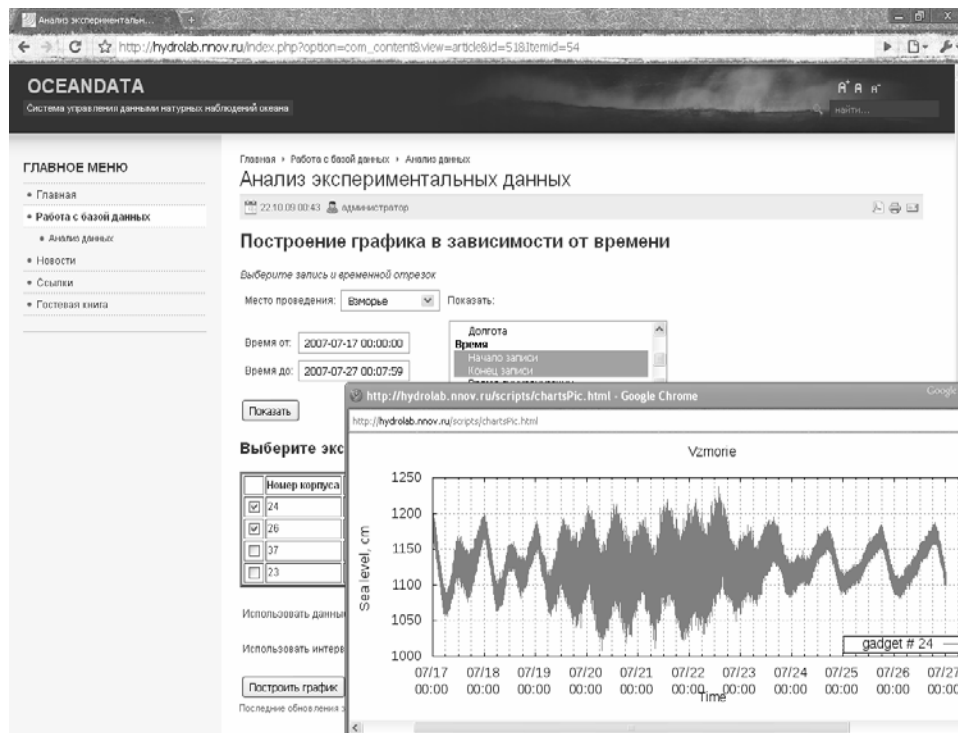


Рис. 2. Пользовательский интерфейс, построение временного ряда.