

Н.В. Голиков, И.А. Сударикова, Ю.И. Молородов
**Распределенная web-система для работы
с атмосферными аэрозолями**

Институт вычислительных технологий СО РАН, г. Новосибирск

Поступила в редакцию 27.11.2005 г.

Представлено краткое описание информационно-вычислительной Интернет-системы «Атлас “Атмосферные аэрозоли Сибири”». Рассматриваются методы и технологии создания распределенной web-системы, которая обеспечивает систематизацию и структуризацию исходных данных по атмосферному аэрозолю и предоставляет максимально возможную унификацию доступа к гетерогенным информационным ресурсам (состав которых может со временем изменяться).

Введение

Атмосферный аэрозоль (АА) является неотъемлемой составной частью атмосферы и играет решающую роль во многих атмосферных процессах (облако- и осадкообразование, радиационный теплообмен, видимость). Он оказывает существенное влияние на качество окружающей среды, климат, химию и физику атмосферы, обладает многообразием форм и физико-химических свойств. Чрезвычайно широк диапазон размеров аэрозольных частиц, присутствующих в атмосфере.

Сильная пространственно-временная изменчивость свойств аэрозоля и их связь со всеми атмосферными процессами обуславливают актуальность углубленного исследования всего многообразия состояний с целью изучения роли атмосферных частиц в изменении глобального климата и их воздействия на геосферно-биосферные процессы. Также особую актуальность аэрозольных исследований определяет необходимость изучения и учета все более возрастающей антропогенной нагрузки на атмосферу Земли для корректной количественной оценки возможных негативных климатических изменений.

Науки об окружающей среде основаны в значительной степени на огромных массивах данных, и одна из ключевых задач информатизации этих наук состоит в организации коллективной работы с архивами данных. С развитием Интернет-технологий появилась возможность по-новому организовывать хранилища данных и доступ к ним.

Международным сообществом накоплен огромный и разнообразный материал о пространственно-временной изменчивости атмосферных аэрозолей [1]. Данные хранятся в различных форматах, иногда несопоставимых между собой для сравнения. Информация сильно разобщена и труднообрабатываема. Создается большое количество алгоритмов для ее обработки, представления, визуализации. В рамках решения проблемы унификации получаемого материала создаются информационные

или информационно-вычислительные системы. Авторы каждой такой системы решают для себя определенный круг задач, часто воспроизводя в рамках своей системы одни и те же алгоритмы и функциональности.

В рамках интеграционного проекта «Атмосферные аэрозоли Сибири» была создана общедоступная информационная система «Атлас “Атмосферные аэрозоли Сибири”», обладающая следующей функциональностью:

1) основная база данных «Атласа» содержит результаты мониторинга аэрозоля, собранные из разнородных источников данных, распределенных по всей территории Западной Сибири;

2) имеется возможность обработки хранимой информации некоторыми математическими алгоритмами;

3) данные представимы в графическом (графики, диаграммы) и табличном виде;

4) система доступна через Интернет по адресу: <http://web.ict.nsc.ru/aerosol/>;

5) в системе предусмотрено разграничение прав пользователей на добавление, изменение информации;

6) система поддерживает два языка (русский и английский) как на уровне интерфейса, так и на уровне самих данных.

Но в описанной системе нет возможности обрабатывать хранимую информацию алгоритмами, работающими не в рамках системы. Еще одним минусом является то, что структура базы данных четко задана (строго реляционная).

Руководствуясь этими предпосылками, пришли к необходимости создания распределенной системы, удовлетворяющей следующим требованиям:

1. Архив данных должен быть расширяемым и доступным, представлять собой набор баз данных, возможно расположенных на разных серверах и состоящих из информации разного формата.

2. Должен быть создан единый словарь основных понятий, связанных с АА, являющийся, по

сути, онтологией предметной области «исследование атмосферного аэрозоля».

3. Обработка данных возможна алгоритмами, реализованными в программных наборах, распределенных на различных серверах.

4. Для адекватного восприятия данных человеком, а также для осмысленной машинной обработки они должны сопровождаться документацией или метаданными, что необходимо для преодоления барьеров при коллективном использовании научных данных.

Для решения поставленной задачи предложено использовать технологии web-сервисов, а в качестве логически связывающего звена онтологию предметной области. Все это в совокупности позволит обеспечить виртуальную интеграцию описаний разнородных информационных ресурсов, расположенных на серверах различных организаций, в единую базу данных на основе открытых международных стандартов, а также предоставит возможность обработки данных вычислительными алгоритмами.

1. Технология построения

«Web-сервис» — это реализуемая программными средствами система для поддержки межмашинного взаимодействия через сеть. Интерфейс сервиса описывается на языке, читаемом машиной, например WSDL [2]. Другие системы взаимодействуют с web-сервисом способом, указанным в его описании, используя сообщения в стандарте SOAP, передаваемые с использованием HTTP и XML и в сочетании с другими стандартами, относящимися к web. Физически «web-сервис» представляет собой фрагмент программного обеспечения, называемый «агентом». Агент способен передавать и принимать сообщения, он реализует функциональность сервиса. Но один и тот же сервис может быть обеспечен разными агентами.

WSDL описывает функциональные возможности web-сервисов и группирует операции взаимодействия с ними в интерфейсы, задающие способы выполнения тех или иных операций и наборы входных и выходных параметров. Таким образом, появляется задача интерпретации информации, описанной в терминах WSDL, на терминологию онтологий. Это позволит расширить представление о предметной области. Например, онтология описывает web-сервис, который рассчитывает коэффициент корреляции между ионной и элементной фракциями аэрозоля. Известно, что в качестве входных параметров он использует массовую концентрацию фракций аэрозоля. Если пользователю необходимо получить информацию о входных данных, то по семантическим связям в онтологии можно будет предоставить ему всю информацию, касающуюся интересующего нас понятия (что это такое, какие значения принимает, как рассчитывается и многое другое).

В настоящее время использование онтологии значительно расширяется. Своевременное оценивание свойств онтологии важно при их создании и сопровождении. В литературе отсутствует единст-

во в определении понятия онтологии и свойств, предлагаемых для ее оценивания. Приведем несколько примеров.

Онтология предметной области есть совокупность соглашений; она определяет термины предметной области, задает их толкование, содержит утверждения, которые ограничивают смысл этих терминов, а также дает толкование этих утверждений.

Онтология — это логическая теория, которая ограничивает допустимые модели логического языка [3]. Онтология в этом случае должна обеспечивать аксиомы, которые ограничивают значение нелогических символов (предикатов и функций) логического языка, используемых как «примитивы» для определенных целей представления.

Онтология предметной области есть та часть знаний предметной области, относительно которой предполагается ее неизменность. Относительно остальной части знаний предметной области предполагается, что она может изменяться, но должна оставаться согласованной с онтологией предметной области.

Одна и та же предметная область может иметь несколько онтологий. Различия между некоторыми из них могут не иметь принципиального характера. Онтологии разных предметных областей могут быть похожи друг на друга. Некоторые онтологии могут быть упрощениями других. Наконец, онтологии одних предметных областей могут быть сконструированы из онтологий нескольких других предметных областей. Например, онтология науки включает онтологию химии аэрозоля (рис. 1).

Вместе с тем можно рассматривать графовое представление базовых компонентов структуры онтологии. Графовая модель онтологии — это такой направленный мультиграф (граф, в котором могут существовать петли и повторяющиеся дуги), где каждая вершина соответствует сущностям или классам сущностей, а направленные дуги, соединяющие вершины, отражают связи разных типов между сущностями. Часть онтологии атмосферного аэрозоля в виде графовой модели представлена на рис. 2.

Здесь WS_1, \dots, WS_n — набор web-сервисов, являющихся обрабатываемыми или вычислительными программами. Онтология указывает для каждого web-сервиса группы химических структур AA, с которыми он работает. В свою очередь, для каждой структуры указаны базы данных ($БД_1, \dots, БД_n$), где хранится информация по соответствующему составу аэрозоля.

На данный момент пользователь системы указывает базу данных, с которой должен работать web-сервис, но в дальнейшем предполагается оставить такое решение онтологии. По описанию web-сервиса в онтологии необходимо принимать решение, с какой из баз предстоит работать. При этом необходим механизм отбора наиболее полной актуальной информации. Обработывая источники, надо отслеживать степень их релевантности понятиям и выдавать список наиболее подходящих для данного понятия ресурсов. Таким образом, появляется задача «принятия решения».

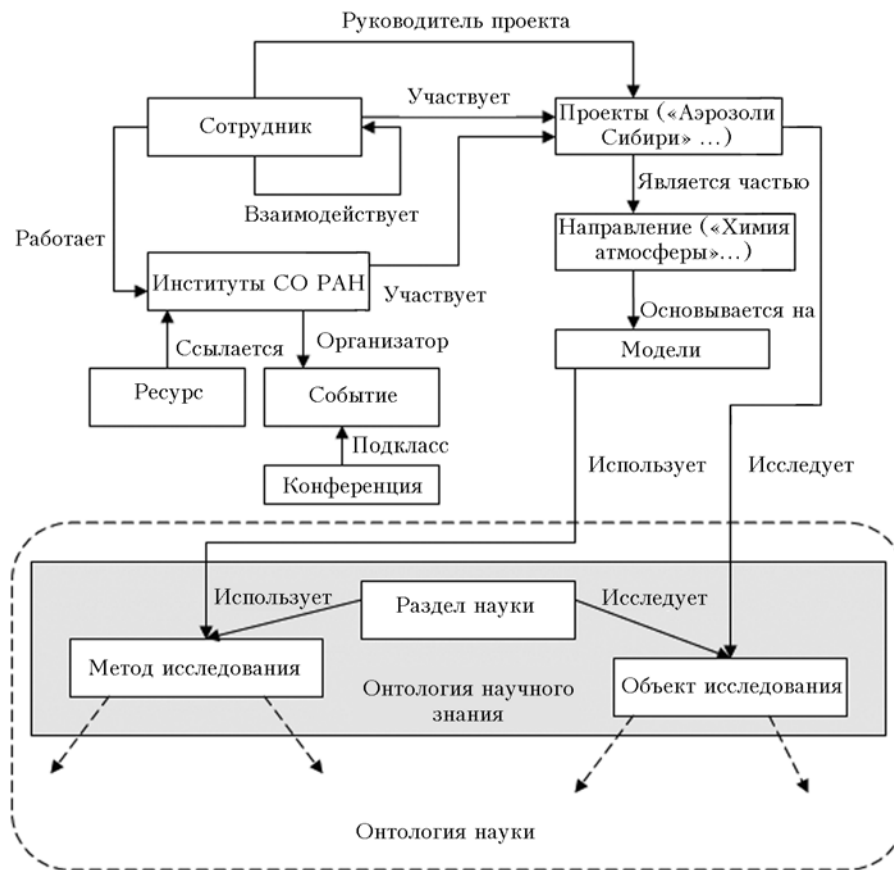


Рис. 1. Пример схемы отношений в онтологии

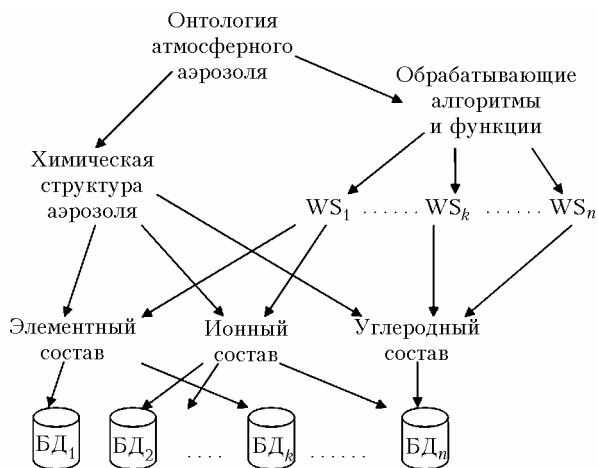


Рис. 2. Графовая модель онтологии атмосферного аэрозоля

Онтология должна содержать как описание предметной области (ПО), так и описание релевантных ей ресурсов. Описание предметной области включает совокупность терминов и отношений, семантически значимых для данной предметной области, а также правил, согласно которым можно строить утверждения об элементах ПО. Описание ресурса сети включает список ссылок на страницы

и сайты, а также описание связей между ними. Для фиксации значимых отношений между терминами выделяются основные связи между ними, которые можно графически отобразить с помощью диаграмм бинарных отношений. Такие связи в дальнейшем могут послужить основой для интеграции различных онтологий. В нашем случае онтология будет содержать полную информацию обо всех «классах» (лабораториях, экспедициях, научных коллективах, химических элементах и т.д.), с которыми может работать портал, содержать ссылки на аналогичные понятия в других базах данных, содержать свойства и ограничения, связанные с этими классами, и полные описания web-сервисов, которые могут работать с ними.

Таким образом, в онтологии портала будут описаны все возможности любого web-сервиса; понятия, с которыми он работает, ограничения на входные и полученные данные (какого типа входные и рассчитанные данные), описание исполняемых функций, место его расположения и т.д. На основе такой онтологии можно будет построить «гибкий» интерфейс пользователя. Например, выбирая интересующие понятия в меню (которое генерируется по онтологии), пользователю будет предоставлена также информация о web-сервисах, которые могут работать с этими данными.

2. Реализация

Реализацию распределенной системы можно разбить на два принципиально отличающихся случая. В первом случае web-сервис в своей работе использует большой массив данных, и тогда не имеет смысла закачивать всю используемую в расчетах информацию на основной сервер, содержащий пользовательский интерфейс, поэтому web-сервис устанавливается на том сервере, где располагается и сама база данных. По такому примеру в «Атласе» уже работает web-сервис получения коэффициента альbedo (albedo – коэффициент отражения от земной поверхности световых лучей). Схема взаимодействия пользовательского интерфейса с web-сервисом представлена на рис. 3.

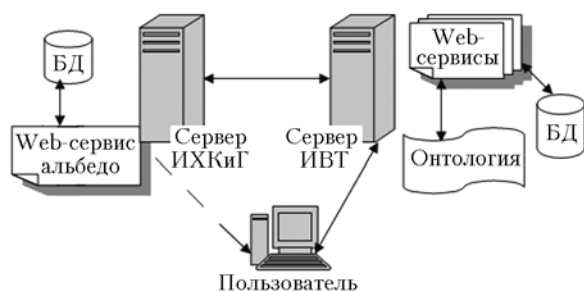


Рис. 3. Распределение данных и web-сервисов на серверах СО РАН

Пользователь через интерфейс системы указывает географические координаты и момент времени, web-сервис по входным данным определяет значение альbedo в точке земной поверхности в указанный момент времени. Вычисления производятся по результатам обработки снимков космических наблюдений, которые хранятся в файловом каталоге, расположенном на сервере ИХКиГ СО РАН. Каждый файл каталога содержит массив значений альbedo, вычисленных в некоторый момент времени для точек земной поверхности. Размер такого файла после разархивирования составляет 300 Мбайт.

Во втором случае web-сервис при обработке использует небольшой массив данных. Тогда имеет смысл распределять базы данных и web-сервисы на отдельные серверы. Структура распределенной системы представлена на рис. 4.

В этом случае на главном сервере системы помимо основных программных средств расположена онтология предметной области, содержание которой участвует в создании интерфейса системы (структура онтологии представлена на рис. 2). При запросе пользователя обрабатывающих алгоритмов, являющихся web-сервисами, по онтологии предоставляется список химических структур, с которыми работает сервис, и баз данных, где имеется информация, используемая в расчетах. Как уже упоминалось выше, на данный момент выбор остается за человеком, но в дальнейшем планируется решить задачу «принятия решения машиной».

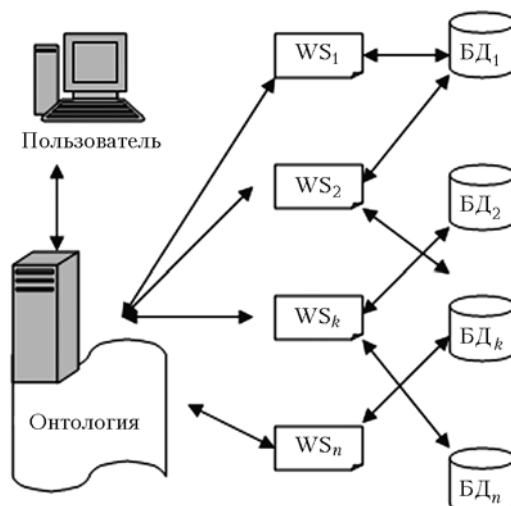


Рис. 4. Схема взаимодействия web-сервисов с распределенными базами данных

Заключение

В последние годы выходят на первый план экологические проблемы, где знание аэрозольных характеристик чрезвычайно важно. Измерение, контроль и прогноз различных параметров аэрозоля необходимы при оценке качества окружающей среды и выработываемых на этой основе жестких, экологически обоснованных требований к промышленной очистке газов, чистоте промышленных помещений, контролю санитарных и жилых зон индустриальных центров, а также фоновых наблюдений, нацеленных на обнаружение и изучение негативных антропогенных воздействий на атмосферу и окружающую среду.

С ростом научных исследований появилась задача собрать и опубликовать информацию о характеристиках АА в Интернете, но для этого необходимо решить несколько серьезных проблем. Информация, хранящаяся по частям в различных системах, с одной стороны, неизбежно дублируется, а с другой – бывает недостаточно полной. Подходы к ее структуризации могут быть различными в разных системах. Приведение всех источников к некоему единому знаменателю – важнейший шаг на пути к общей интеграции. Это объясняет необходимость обеспечить систематизацию и структуризацию исходных данных. Для этого нужно реализовать максимально возможную унификацию доступа к гетерогенным информационным ресурсам (состав которых может со временем изменяться).

Информационные системы, как правило, очень сильно различаются по применяемым технологиям, производительности, способам доступа к информации и т.п. Это требует индивидуального подхода к каждому источнику. Однако существуют стандарты и технологии, позволяющие унифицировать взаимодействие с разнородными ИС, и их применение делает прозрачной общую архитектуру интеграции. В целом, распределенная система должна позволить оценить влияние антропогенных и есте-

ственных источников на изменение характеристик АА, решив тем самым многие вопросы и задачи экологии.

1. Куценогий К.П., Куценогий П.К., Молородов Ю.И., Федотов А.М. Разработка структуры метаданных по атмосферным аэрозолям на основе информационной

модели // Тр. Междунар. конф. «Математические методы в геофизике». Ч. II. Новосибирск, 2003. С. 438–444.

2. *Web Services* Description Language: <http://www.w3.org/2002/ws/desc/>
3. *Genesereth M.R., Nilsson N.J.* Logical Foundation of Artificial Intelligence. Morgan Kaufmann, Los Altos, California, 1987.

N.V. Golikov, I.A. Sudarikova, Yu.I. Molorodov. **Distributed web-system on atmospheric aerosols.**

Brief description of the information and calculating Internet-system «Atlas of Atmospheric aerosols in Siberia» is presented. The technologies and methods of creating the distributed web-system are considered. This system provides systematization and structurization of initial data of atmospheric aerosol, and gives an opportunity of maximum possible unification of access to heterogeneous information resources, the composition of which can change in time.