

Электронные библиотеки и базы данных для задач экологии ¹

© Молородов Ю.И.

Институт вычислительных технологий
СО РАН
yum@ict.nsc.ru

© Федотов А.М.

Институт вычислительных технологий
СО РАН
fedotov@ict.nsc.ru

Аннотация

Технологии разработки электронных библиотек позволяют создавать информационные структуры, предназначенные для интеграции распределенных баз данных наблюдений, мониторинга, анализа и моделирования состояния экосистем. Они обеспечивают возможность для изучения и прогнозирования природных, социальных и экономических последствий, вызванных как естественными, так и антропогенными изменениями в экосистемах. Это позволит разработать рекомендации и подходы к оптимальному управлению функционированием экосистем.

1 Введение

В настоящее время на первый план выходят экологические проблемы, когда знание характеристик, описывающих состояние окружающей среды, становится чрезвычайно важным. Так измерение, контроль и прогноз различных параметров атмосферных аэрозолей (АА) 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 расширением диапазона 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 единому знаменателю и использование технологий создания электронных библиотек - 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).

Создаваемая распределенная система должна позволить оценить влияние антропогенных и естественных источников на изменение характеристик АА, решив тем самым многие вопросы и задачи экологии.

Наиболее устойчивыми к росту объемов хранимой информации оказались системы, оперирующие строго структурированными данными, например, каталоги библиотек. Но даже в таких системах в границах одной предметной области может сложиться ситуация, когда данные, вводимые разными авторами, оказываются несовместимыми между собой из-за различий в используемой ими терминологии. Вследствие чего между этими данными трудно устанавливать связи. На текущий момент эта проблема наиболее полно решена в системах, оперирующих библиографической информацией [1, 3, 4]. Для них существует несколько стандартов представления библиографических записей, например, российский коммуникативный формат RUSMARC или любой формат из группы формат MARC. Наличие таких стандартов снимает большое количество проблем связанных с систематизацией, поиском, обработкой информации. Фактически такой подход представляет собой создание словаря объектов, которые будут использоваться для задания структуры хранимой информации.

Труды 9^{ой} Всероссийской научной конференции
«Электронные библиотеки: перспективные методы и
технологии, электронные коллекции» - RCDL'2007,
Переславль-Залесский, Россия, 2007.

¹ Работа выполнена при поддержке Президентской программы НШ 2314.2003.19 и Интеграционного проекта СО РАН № 169

2 Необходимость создания информационных систем

Одним из наиболее бурно развивающихся направлений информационной индустрии последних лет стала разработка удаленно доступных информационных систем. Причинами такого бурного роста стали достижения одновременно в нескольких областях. К ним можно отнести:

- значительный рост пропускной способности каналов связи;
- рост производительности компьютеров, как по скорости, так и по объемам оперативной и внешней памяти;
- широкое проникновение компьютеров и компьютерных технологий в повседневную и производственную деятельность граждан;
- развитие сети Интернет, обеспечивающей простой и надежный доступ к огромному числу информационных ресурсов;
- развитие самих информационных технологий.

Одновременный рост производительности компьютеров, пропускных способностей каналов связи и технологий программирования дал новую жизнь некоторым старым технологиям, в частности идее интерпретируемого кода. Впервые она была предложена в 70-е годы XX века как средство машинной реализации языка Паскаль (p-code). Сегодня на этих же идеях строится язык Java с его виртуальной машиной.

Эти общие процессы, естественно, в значительной степени коснулись и науки. В этом смысле хотелось бы отметить три основных направления её информатизации.

Основным способом доступа к научной информации становится доступ через сеть. На практике это означает, что необходимо обеспечить средства предоставления результатов исследований в сеть, средства эффективного поиска необходимой информации и доступа к ней. Все это означает необходимость создания и развития разнообразных электронных библиотек и перевода традиционных библиотек на обслуживание в удаленном режиме.

В значительной степени деятельность ученых и коллективов ученых перемещается в создание баз данных, содержащих результаты исследований. Часто при этом создаются не просто базы данных, но и средства доступа к специфическим данным (химическим, астрономическим, физическим и т.д.). Разрабатываются средства визуализации этих данных.

Принципиально важной становится интеграция разнообразных данных и систем, поддерживающих и обеспечивающих научный процесс. Здесь в свою очередь необходимо отметить следующие две проблемы.

Первая из них: научному сотруднику необходимо, чтобы по возможности вся информация была одновременно и легко доступна, т.е. чтобы все компоненты, образующие

информационную систему поддержки научных исследований, были интегрированы как по форме (имеется в виду средства, предоставляемые пользователю), так и по содержанию (информационная система должна поддерживать связи по данным между различными компонентами).

3 Электронные коллекции

Изучением проблем изучения состояния и использования современных методов наблюдений, вычислительных и информационных технологий для оценки, моделирования и смягчения последствий изменения окружающей среды под воздействием естественных и антропогенных факторов занимаются коллективы специалистов, работающие в различных областях наук об окружающей среде. Эти коллективы успешно работают в области физики, химии, гидро- и газодинамики, биологии, экологии и т.д. Из этого следует необходимость мультидисциплинарных подходов в науках об окружающей среде, а присущее любой науке стремление к пониманию явлений и их прогнозированию приводит к широкому использованию математического моделирования в качестве вычислительного аппарата.

Развитие математических и вычислительных моделей показало явную недостаточность существующих данных о состоянии окружающей среды и стимулировало как интенсивное накопление временных рядов пространственно распределенных данных инструментальных наблюдений (локальных и дистанционных), так и разработку новых инструментальных средств. Эти обстоятельства предопределили особую роль информационных технологий в этой области.

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 [1,2]. На сегодняшний день существует ряд информационных систем, предназначенных для решения различных задач из области наук об окружающей среде и экологии. Это HYSPLIT (<http://www.arl.noaa.gov/ready/hysplit4.html>) - обширная база данных наблюдений, позволяющая определить перенос воздушных масс за определенный период времени; GISS (http://www.giss.nasa.gov/data/rsp_air/) комбинирует и сравнивает результаты набора данных космических наблюдений с глобальными моделями атмосферы, земной поверхности и океанических процессов; PANGAEA (www.pangea.de) - библиотека данных и публикаций в области геофизики; мультимедийный информационный ресурс, открывающий интерактивный доступ к собранной информации по оценке влияния

атмосферных примесей на окружающую среду (<http://www.ess.co.at/AIR-EIA>). Примером отечественной разработки является научный портал для атмосферных наук ATMOS (<http://atmos.iao.ru>), который представляет собой интегрированный набор множества распределенных, но координируемых предметных сайтов, содержащих типовую мультимедийную информацию с исследовательскими базами данных, моделями и аналитическим инструментарием для прямого использования и визуализации.

4 Данные об атмосферных аэрозолях

В последние три десятилетия в связи с интенсивным изучением проблемы глобальных изменений окружающей среды выполнялись крупные региональные проекты [3] по комплексному сбору необходимой информации по атмосферным аэрозолям (АА). Они - неотъемлемая компонента атмосферы и объяснение всех сложных процессов, происходящих в ней, невозможно без подробной информации об ионном и элементарном составе АА, спектре размеров, массовой и счетной концентрации, химическом и фазовом составе, структуре аэрозольных частиц.

Начиная с 1991 г. специалисты ИХКиГ СО РАН, ИВЭП СО РАН, ИНХ и ЛИН СО РАН проводили регулярные измерения временных характеристик атмосферы. Мониторинг проводился на территории Западной и Восточной Сибири, Алтайского и Красноярского краев. Исследовалась пространственно-временная изменчивость химического состава атмосферного аэрозоля, изменчивости массовой концентрации, многоэлементного состава, содержания органического и неорганического углерода в АА в различных регионах Сибири. Этот огромный объем экспериментальных данных неизмеримо расширил и во многом изменил существующие представления о техногенном влиянии промышленных центров на окружающую среду региона.

Атмосферный аэрозоль является сложной структурой, состоящей из различных химических элементов и их соединений. Сбор исходного материала АА осуществляется на стационарных точках наблюдений (станциях) с периодичностью несколько раз в год. Полученные пробы разделяют на элементные, ионные и углеродные фракции, определяют их количественное и качественное содержание.

Элементный состав аэрозоля включает в себя отдельные химические элементы, такие как, например, Са, Вг, К и т.д. В ионном составе выделяют катионы NH_4^+ , K^+ , H^+ , Na^+ и анионы HCO_3^- , Cl^- , F^- , NO_3^- , так как только эти химические структуры определяемы на современном оборудовании. Для углеродного состава определяют содержание органического и неорганического углерода на фильтре, которым берут пробы. По общей массовой концентрации аэрозоля система вычисляет и другие значимые углеродные

характеристики: суммарную относительную концентрацию и соотношение органического и неорганического углеродов.

Наиболее важными для дальнейшей обработки и моделирования являются полная массовая концентрация покомпонентного состава и концентрация грубодисперсной фракции, а также субмикронная массовая концентрация и полная счетная концентрация аэрозолей, особенно отличающиеся нерегулярностью измерений. Физико-химический анализ проб осуществляется в научно-исследовательских институтах ИХКиГ, ИЯФ и ИНХ СО РАН, после чего все полученные данные экспортируются в хранилище Атласа.

5 Обработка данных

В ИВТ СО РАН совместно с ИХКиГ СО РАН была разработана информационная модель, электронная библиотека и структура метаданных на основе обобщенного подхода для формирования и заполнения файлов входных данных, включая их унификацию и связи для электронного атласа "Атмосферные аэрозоли Сибири" (<http://web.ict.nsc.ru/aerosol/>). Для построения Атласа использовался программный продукт SMART [2]. Структура Атласа представляет собой набор документов, входящих в электронные коллекции. Каждая из коллекций изначально представлена в виде RDF-описания. На выходе по запросу клиента формируется документ в HTML-формате, при этом стилевое оформление документов физически отделено от их содержательной части.

Хранимая информация представляет собой набор классов, являющихся объединением объектов. Например, класс экспедиций, станций наблюдения, статей. Классы, в свою очередь, взаимосвязаны.

Каждый объект класса для описания имеет определенный список атрибутов - метаописания. Метаописания объектов составляет отдельный класс данных. В базе данных информационной системы хранится не только информация об элементах системы, но и сведения о её структуре. Это придаёт необходимую гибкость информационной модели, позволяя ей динамично отображать структуру документа.

Система предоставляет не только средства хранения данных, визуализацию их зависимостей от времени в табличном и графическом виде (рис. 1), но и возможность обработки данных вычислительными алгоритмами для решения различных задач исследования окружающей среды. При проектировании Атласа было выделено несколько классов задач, каждый из которых порождает собственные модели хранения и обработки данных.

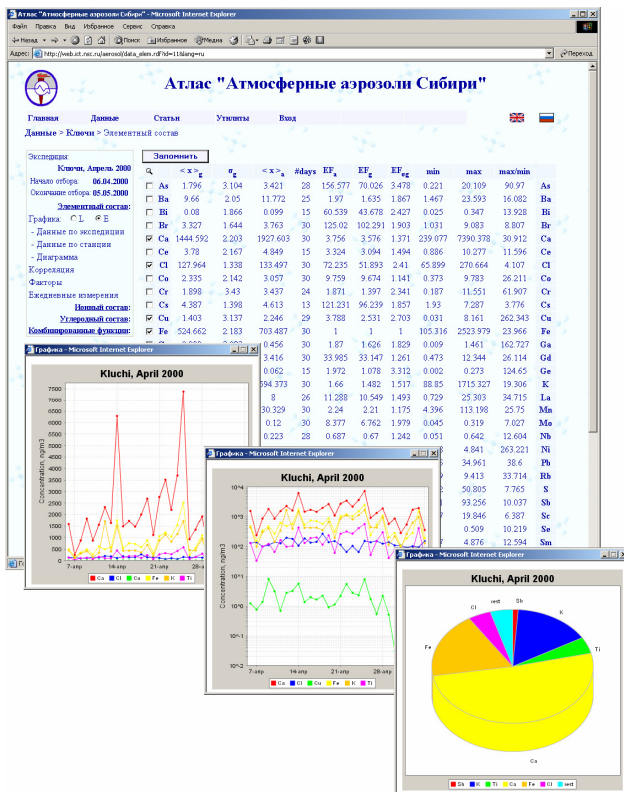


Рис. 1 Представление данных в информационной Системе

3. Заключение

Создание региональной базы данных по атмосферным аэрозолям (АА) на основе электронных библиотек повысит эффективность комплексного мониторинга окружающей среды Сибирского региона. Анализ эмпирического материала даст возможность понять механизм образования АА естественного и антропогенного происхождения. Это позволит обосновать математические модели для описания различных процессов в атмосфере и оценки влияния АА на естественные атмосферные процессы, изменчивость погоды и эволюцию климатических изменений качества окружающей среды. Появляется возможность оценить влияние АА на растительные и животные биоценозы, биохимические циклы в различных почвенно - климатических зонах Сибири и биосферы в целом, состояние здоровья населения.

Созданный информационно-вычислительный комплекс предоставляет возможность не только объединить различные информационные ресурсы в концептуально одну информационную среду, оперативно управлять и актуализировать информацию, хранящуюся в разнородных базах данных и источниках, и, что самое главное, обрабатывать данные с целью их дальнейшего использования для климатического моделирования. Результаты обработок являются доступными для общего просмотра.

Важным преимуществом предлагаемой технологии создания информационных систем и электронных архивов является значительное

расширение аудитории, которая может ознакомиться с экспонатами за счет его виртуального присутствия в сети Интернет и на различных носителях (CD). Организовано хранение огромного эмпирического материала, всех описаний и взаимосвязей экспонатов в единой базе данных, что позволит сохранить огромный объем научных результатов для следующего поколения исследователей.

Современный уровень развития информационных технологий, математического моделирования, методов вычислительной математики и наукоемкого прикладного программного обеспечения делают возможным создание нового поколения информационно-вычислительного инструментария для решения задач мониторинга и охраны окружающей среды..

Литература

- [1] Golikov N.V., Molorodov Yu.I., Sudarikova I.A. Web system for storing and processing data of atmospheric aerosols // Proceedings of SPIE Vol. 6160 Part I, 2005.-P. 66 -70
- [2] Guskov A., Shraibman V., Molorodov Y. SMART - an approach for information systems development on the basis of RDF-technology // Proc. of the VIII International Conference on Electronic Publishing. - Brasilia, Brazil. - 2004. - P. 293-305.
- [3] Молородов Ю.И., Куценогий П.К., Сударикова И.А. Содержательное наполнение атласа "Атмосферные аэрозоли Сибири". Журнал "География и природные ресурсы". Специальный выпуск. Новосибирск: Изд-во СО РАН, филиал ГЕО, 2004, с. 86-90.
- [4] Молородов Ю.И., Федотов А.М. Информационные системы для интеграции данных о состоянии окружающей среды // Вестник ИрГТУ.-2006.- №2 (26), т.3.- С.46-52.
- [5] Федотов А.М., Гуськов А.Е. Информация в Интернете: публикация, поиск, анализ // Информационные технологии в высшем образовании. – 2004, № 4. – 17-35

Database and electronic libraries for the ecologies problems

Molorodov Yu., Fedotov A.

Technologies of the development of the electronic libraries allows to formulate the recommendations and approaches to optimum control of the ecological systems.