

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ ДЛЯ ОТОБРАЖЕНИЯ И ИНТЕРПРЕТАЦИИ ДАННЫХ О СОСТОЯНИИ ОКРУЖАЮЩЕЙ СРЕДЫ*

Ю. И. Молородов

Институт вычислительных технологий СО РАН, Новосибирск, Россия

e-mail: yumo@ict.nsc.ru

The modern information systems intended for environmental data presentation and handling are presented. The scope of problems is described that can be solved within these systems and their applications are illustrated.

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 вырабатываемых на этой основе жестких, экологически обоснованных требований к промышленной очистке газов, чистоте промышленных помещений, контролю санитарных и жилых зон индустриальных центров, а также при оценке данных фоновых наблюдений, нацеленных на обнаружение и изучение негативных антропогенных воздействий на атмосферу и окружающую среду.

С ростом диапазона научных исследований появилась задача собрать и опубликовать информацию о характеристиках атмосферных аэрозолей в Интернете, но для этого следует 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, с этой целью требуется реализовать максимально возможную унификацию доступа к гетерогенным информационным ресурсам (состав которых со временем может изменяться).

Информационные системы могут сильно различаются по применяемым технологиям, производительности, способам доступа к информации и т. п. Это требует индивидуального

*Работа выполнена при финансовой поддержке президентской программы “Ведущие научные школы РФ” (грант № НШ-9886.2006.9), программы междисциплинарных интеграционных исследований СО РАН (гранты № 2006-34, № 2006-115).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2006.

подхода к каждому источнику. Однако существуют стандарты и технологии, позволяющие унифицировать взаимодействие с разнородными информационными системами, и их применение делает прозрачной общую архитектуру интеграции. В целом распределенная система должна позволить оценить влияние антропогенных и естественных источников на изменение характеристик атмосферных аэрозолей, решив тем самым многие вопросы и задачи экологии.

1. Необходимость создания информационных систем

Одним из наиболее бурно развивающихся направлений информационной индустрии последних лет является 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.

С этой точки зрения можно сказать, что в настоящее время программирование находится на четвертой фазе своего развития. Первой из них можно назвать “классическое” программирование (сначала в кодах, затем на ассемблере, затем на языках высокого уровня) для больших ЭВМ. Потом было “классическое” программирование для персональных ЭВМ. После него появилось программирование с использованием визуальных и CASE-средств, а затем и “сетевое” программирование.

Одновременный рост производительности компьютеров, пропускной способности каналов связи и технологий программирования дал новую жизнь некоторым старым технологиям, в частности идее интерпретируемого кода. Впервые она была предложена в 70-е годы 20 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 направления информатизации.

Основным способом доступа к научной информации становится доступ через сеть. На практике это означает, что требуется обеспечить средства предоставления результатов исследований в сеть, средства эффективного поиска информации и доступа к ней, т. е. необходимы создание и развитие разнообразных электронных библиотек, а также перевод традиционных библиотек на обслуживание в удаленном режиме.

В значительной степени деятельность ученых перемещается в область создания баз данных, содержащих результаты исследований. Часто при этом создаются не только базы данных, но и средства доступа к специфическим данным (химическим, астрономическим, физическим и др.). 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 (имеются в виду средства, предоставляемые пользователю), так и по содержанию (информационная система должна поддерживать связи по данным между различными компонентами).

2. Технологии создания информационных систем

В настоящее время к древнейшему источнику информации о состоянии экосистем — полевым наблюдениям и экспериментам добавилось дистанционное аэро- и космическое зондирование больших участков поверхности земли. Вследствие этого возникла актуальная задача — разработка методов комплексного анализа и интеграции данных многозональных космических съемок с результатами полевых исследований для создания карт экосистем окружающей среды. Создание единого информационного пространства для этого даст качественно новые возможности анализа состояния экосистем.

Данные наблюдений включают в себя числовую информацию (печатную и электронную), графический материал и словесное описание. Сведения о них и записи данных существуют на бумаге, пленке (микрофиши), магнитных дисках, в цифровой оптической среде и т. д. Однако сегодня более приемлемы данные, представленные в электронном виде.

Заметим, что данные являются критическим ресурсом, который необходимо сохранять, защищать и организовывать доступ к ним для всех пользователей в любое время суток [1]. Архив данных должен быть расширяемым, вместительным и доступным. Для адекватного восприятия данных человеком, а также осмысленной машинной обработки они должны сопровождаться документацией или метаданными, что необходимо для преодоления барьеров при коллективном использовании научных данных (решением для Интернет-ресурсов является подход Semantic Web [2, 3]). Наконец, приемлемая и эффективная стратегия архивирования данных основана на использовании распределенных архивов, управляемых теми, кто обладает предметным знанием о данных (решение лежит в плоскости подхода Grid) [4].

При коллективной работе с данными существенно задание форматов данных. В небольших научных проектах не обращают внимания на форматы данных, так как исследователи планируют и осуществляют работу со своими данными в узком кругу научного сообщества и любые семантические неувязки разрешаются на уровне личного общения. При проведении больших проектов устанавливается единый формат данных. Проблема состоит в такой подготовке структурированных научных данных, при которой их интерпретация доступна всем уровням пользователей. Структура данных должна допускать все возможные пути полного восстановления информации о получении, архивировании и обработке данных.

Продуманная многоуровневая система сбора, обработки и хранения атмосферных данных построена в США. Основной объем данных сосредоточен в центрах активных распределенных архивов — Distributed Active Archive Center (DAAC) [1]:

— Alaska Synthetic Aperture Radar (SAR) Facility (ASF) содержит данные о полярных процессах и радарные данные (<http://www.asf.alaska.edu/>);

— EROS Data Center (EDC) Land Processes описывает поверхностные процессы (<http://landcover.usgs.gov/>);

— Goddard Space Flight Center (GSFC) содержит сведения о верхней атмосфере, глобальной биосфере, атмосферной динамике и геофизике (<http://www.gsfc.nasa.gov/>) и др.

Например, распределенная информационная система “Наблюдение за Землей” (Earth Observation System/Data Information System (EOS/DIS) строится для сбора и обработки данных и их связи с данными центров активных распределенных архивов. Инструментальные и научные команды развивают алгоритмы для обработки данных и генерации продуктов данных, а центры обеспечивают сервисы пользователя. EOS/DIS устанавливает требования для создателей информационных систем и координирует работу центров.

В этом огромном потоке информации необходимо быстро ориентироваться. Имеющиеся бумажные каталоги, списки, справки, полученные с технических носителей, уже не помогают. Во-первых, информация очень быстро изменяется, массивы пополняются, корректируются и пересчитываются. Во-вторых, объем такой информации относительно большой. Так, базы сведений о характеристиках изучаемой среды того или иного района оцениваются в десятки и сотни мегабайт. В-третьих, справочная информация — это основа для перехода на безбумажную технологию обработки информации о состоянии среды. При фильтрации данных иногда требуются такие атрибуты, которых нет в исходных массивах данных. Для успешного поиска качественных данных необходимы сведения о массивах данных, форматах их обмена, программных средствах их обработки, организациях, собирающих и хранящих данные, и др. Все эти сведения называются метаданными. Кроме того, в основу создания информационных систем положено понятие метаданных. В таком контексте это формализованное описание коллекций, документов или объектов, структурированных сведений об электронных ресурсах, представляющих его свойства или атрибуты. Метаданные используются для каталогизации документов, поиска, вывода результатов поиска, для установления связей между отдельными документами, формирования тематических словарей. В набор элементов, используемых для создания метаданных, входят стандартные наборы MARC21, Dublin Core и др.

Стоит отметить значительную работу с метаданными, выполняемую National Virtual Data System (NVDS) (<http://www.nvds.noaa.gov/>), обращение к которым существенно облегчает поиск необходимых ресурсов [1].

3. Информационные ресурсы, созданные за рубежом

Задачи численного прогноза о состоянии окружающей среды имеют более чем полувековую историю, но последнее десятилетие стало этапным в подходе к их решению. Прогресс достигнут за счет координации и интеграции усилий многих организаций в области анализа наблюдений и моделирования, а также в значительной мере за счет использования информационных технологий, которые позволяют несколько с иных позиций подойти к решению как климатических задач, так и других проблем фундаментальных наук. Техническая сторона прогресса связана с формированием распределенных информационных ресурсов, содержащих физические характеристики окружающей среды и описывающих связи между ними. Дело в том, что исследователям необходимо все более детальное описание физических процессов в атмосфере, океане, на поверхности суши, что требует привлечения множества организаций и специалистов разных областей наук. Для разработки компонентов математических моделей, связанных с ними задач и вычислений, проводимых на их основе, требуются все более производительные ресурсы. Исследования последних лет показали, что изменения климата Земли связаны не только с взаимодействием атмосферы, океана, поверхности суши и морских льдов, но и с их химическим и биологическим составом, более того, на него влияют антропогенные факторы (парниковые газы,

землепользование и т. д.). Управление и прогноз изменений климата и погоды возможны на пути дальнейшего уточнения физических, химических, биологических процессов и их интеграции в существующую систему знаний о поведении системы Земля.

Среди множества разномасштабных проектов, выполняемых в настоящее время иностранными учеными и направленных на решение упомянутых проблем, выделим следующие:

PRogram for Integrated Earth System Modelling (PRISM) (<http://www.enes.org>);
Earth Frontier Project (<http://www.es.jamstec.go.jp/>);
Earth System Modelling Framework (ESMF) (<http://www.esmf.ucar.edu/>);
Earth System Grid (ESG) (<http://www.earthsystemgrid.org/>).

В этих проектах решаются задачи многократного использования кода, обеспечения интероперабельности, простоты использования и компактного выполнения в приложениях (климатических, погодных) и усвоения данных [5]. Там же описаны проблемы технического и информационного обеспечения, а также подходы и проблемы, связанные с развитием аппаратного и программного обеспечения и использованием масштабируемых компьютерных архитектур.

4. Информационные ресурсы, созданные на территории Сибири

Учеными Сибирского отделения РАН также создаются распределенные базы данных (БД) по разным разделам наук об окружающей среде.

4.1. Базы данных для ГИС “Аэрозоли Алтай”

База данных для ГИС “Аэрозоли Алтай” <http://iwer.asu.ru/altai/index.php> [6] сформирована на основе многолетних наблюдений за характеристиками атмосферных аэрозолей в приземном слое воздуха, проведенных на территории Алтайского края и Республики Алтай в период 1991–2002 гг. Ее основой служит материал, полученный в ходе разовых и многосуточных наблюдений в экспедициях с помощью передвижной лаборатории.

База данных содержит информацию о микрофизических параметрах аэрозолей: спектрах размеров частиц, элементном и ионном составе, массовой концентрации, а также данные о метеопараметрах в пунктах отбора проб. Кроме того, в базу данных вошли результаты исследований качества атмосферного воздуха и заболеваемости населения г. Барнаула.

С целью обеспечения доступа к базе данных как можно большего числа специалистов в области космического мониторинга, физики атмосферных аэрозолей и пр. создана топологическая векторная модель данных ГИС “Аэрозоли Алтай”, которая представляет интерактивную векторную карту с элементами гипертекстовой разметки.

Для представления информации, содержащейся в многомерной базе данных “Аэрозоли Алтай”, разработана концептуальная модель построения ГИС (статичная реализация web-сайта). В основе построения модели базы данных лежит реляционная модель БД. На рисунке терм 1 соответствует начальной странице Интернет-проекта, на которой представлены растровые изображения Алтайского края и Республики Алтай, совмещенные посредством инструментов отображения векторной графики.

С целью совмещения позиционной и атрибутивной информации каждая из территорий представлена в виде структур, состоящих из триад термов: двух основных (3-Т-7, 2-Т-6) и трех дополнительных (типа 4-Т-5). В описываемой структуре термы 2, 3 и 4 — это интерактивные карты территорий; термы 5, 6 и 7 представляют собой наборы фреймовых структур, объединяющих детальные карты местности пунктов отбора проб, таблицы данных и диаграммы; термы Т содержат списки пунктов отбора проб для каждой из территорий.

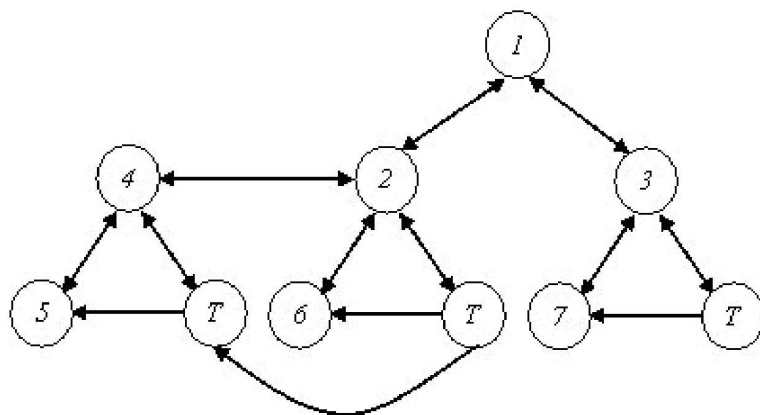
Совокупность термов 3-Т-7 образует интерактивную карту для Республики Алтай. Интернет-страница, соответствующая терму 3, представляет собой растровое изображение карты территории, на которое нанесен слой маркеров, которыми отмечены места отбора проб. Маркеры содержат гиперссылки на наборы фреймовых структур (терм 7) для представления атрибутивной информации (таблицы и диаграммы). Карта дополнена списком, содержащим таблицу мест отбора на данной территории (терм Т). Каждый список объектов связан с атрибутивной информацией (терм 7).

Совокупность термов 2-Т-6 образует интерактивную карту для Алтайского края и имеет аналогичную структуру представления данных. Отметим, что на данной карте с помощью графических маркеров, содержащих гиперссылки, сделаны переходы на карты районных уровней (дополнительные триады термов).

В соответствии с масштабом представления территорий векторную интерактивную карту можно условно разделить на три уровня, связанных между собой.

На первом уровне интерактивной карты представлен Алтайский регион: растровые изображения карт Алтайского края и Республики Алтай, совмещенные между собой и имеющие привязку в относительных координатах (см. рисунок). Пользователь может выбрать для просмотра информацию по Алтайскому краю или Республике Алтай либо вернуться на Интернет-страницу лаборатории.

На втором уровне находятся карты региональных единиц — Алтайского края и Республики Алтай. На Интернет-странице растровая карта совмещена с позиционным слоем, содержащим относительные координаты точек отбора проб воздуха и гиперссылки на атрибутивную информацию по объекту. Web-страница является интерактивной картой, на которой пользователь может выбирать объекты, отображенные на карте, и получать в интерактивном режиме атрибутивную информацию об объекте. Точки отбора проб — объекты карты — обозначены флажками красного цвета. При наведении указателя мышки на один из флажков появляется указатель на гиперссылку (атрибутивный фрейм объекта)



Граф представления базы данных для ГИС "Аэрозоли Алтай".

в виде подписи к объекту, содержащей дату отбора пробы и условное название пункта. На данной странице имеются также гиперссылки на таблицу, включающую общий список пунктов отбора проб, и таблицу данных по содержанию тяжелых элементов в пробах воздуха.

Третий уровень — растровые фрагменты карт местности отбора проб. Атрибутивная информация объектов (проб) представлена в виде фреймовых структур, содержащих многомерные таблицы спектров размеров частиц, метеопараметров, а также диаграммы, составленные на основе этих таблиц.

4.2. Web-портал ATMOS — основа выполнения исследований по окружающей среде Сибири

Двухязычный научный web-портал ATMOS (<http://atmos.iao.ru/>) разработан специалистами Института оптики атмосферы СО РАН как интегрированное множество распределенных, но согласованных тематических web-сайтов, комбинирующих стандартную мультимедийную информацию с исследовательскими базами данных, моделями и аналитическими инструментами для использования и визуализации online [7]. Основной темой является физика и химия атмосферы, включая спектроскопию атмосферы, аэрозоли, радиацию, а также прикладные области оценки и управления качеством воздуха, моделирования и оценки воздействия на окружающую среду.

Портал 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. Каждый предметный сайт является отображением информационной системы средствами Интернет-технологий.

Портал разделен на две группы сайтов.

К первой группе <http://atmos.iao.ru/> относятся сайты аналитического направления, большая часть из которых включает в себя информационно-вычислительные системы. В эту группу входят сайты “Атмосферная химия” <http://atchem.atmos.iao.ru/>, “Атмосферная спектроскопия” <http://saga.atmos.iao.ru/> и “Атмосферный аэрозоль” <http://aerosol.atmos.iao.ru/>. Они содержат данные измерений оптических характеристик аэрозоля, химические свойства компонентов атмосферы, микрофизику аэрозоля, описание элементарных химических и физических процессов в атмосфере и т. д.

Сайт “Данные измерений” <http://solter.atmos.iao.ru/> содержит данные измерений, производимых в Иркутске и Томске.

Сайт “Атмосферная радиация” <http://atrad.atmos.iao.ru/> используется для вычисления радиационных потоков и изучения влияния аэрозоля, облаков и малых составляющих атмосферы на радиационный режим.

Сайт “Климат” <http://climate.atmos.iao.ru/> ориентирован на рассмотрение климатических моделей и моделей описания окружающей среды.

Вторая группа сайтов содержит описание подхода к задачам оценки и управления качеством воздуха, физико-географическую информацию о двух географических объектах Сибири. К этой группе относятся три сайта:

- сайт “Оценка и управление качеством воздуха” <http://air.atmos.scert.ru/>;
- сайт “Озеро Байкал” <http://baikal.atmos.scert.ru/>;
- сайт “Западно-Сибирская низменность” <http://west-sib.atmos.scert.ru/>.

4.3. Атлас “Атмосферные аэрозоли Сибири”

Основной темой информационно-вычислительного атласа, представленного на сайте Института вычислительных технологий СО РАН (<http://web.ict.nsc.ru/aerosol/>), являются атмосферные аэрозоли Западной Сибири [8]. В атлас вошли результаты регулярных измерений временных характеристик атмосферных аэрозолей, полученные специалистами институтов СО РАН: Института химической кинетики и горения, Института неорганической химии, Лимнологического института, Института водных и экологических проблем, Кемеровского, Красноярского и Томского научных центров. Мониторинг проводился на территории Западной и Восточной Сибири, Алтайского и Красноярского краев. Наблюдения осуществлялись в различных почвенно-климатических зонах: в лесостепной и степной, лесотундровой, тундровой, а также в Арктическом бассейне России. Кроме того, в рамках совместных исследований с Сибирским отделением РАН Институт космических исследований РАН (г. Москва) проводит сбор данных со станций приема космической информации, полученной с низкоорбитальных спутников серии NOAA, Метеор с многоканальной аппаратурой низкого, среднего и высокого разрешения, представляющих собой спутниковые снимки различных участков Земли. При этом используются возможности бортовых радиометров систем наблюдения Earth Observing System (EOS) на MODIS/Terra, MODIS/Aqua-, AMSR-E/Aqua-платформах, позволяющих определять атмосферные и поверхностные характеристики, а также свойства подстилающей поверхности.

Разработанные сервисы позволяют пользователю заносить данные измерений на сервер ИВТ СО РАН, предоставлять результаты обработки в виде документов, не загружая вычислениями его персональный компьютер.

Используются предметные модули, в основу которых положены данные, характеризующие предметную область и математические модели оперирования с ними. Информация и эмпирические данные хранятся преимущественно в файловых системах или в базе данных. Алгоритмы формируют базис для моделей, используемых в вычислениях. Инструментарий для представления данных включает аппаратное и программное обеспечение для передачи и отображения информации.

Реализованы интерфейсы доступа к данным космических наблюдений на основе ГИС-технологий. Усовершенствована и формализована схема данных, для того чтобы максимально облегчить подключение гетерогенных информационно-вычислительных компонентов. Выработана стратегия перехода на сервисную архитектуру на базе web-сервисов и Grid-сервисов с возможностью интеграции с другими информационно-вычислительными системами из области физики атмосферы.

Заключение

Трудно представить организацию коллективной работы без существования сети Интернет, а следовательно, без использования Интернет-технологий. Заметим, что на выбор такого пути развития оказывают влияние существенные экономические и технические факторы. Дело в том, что скорость технологических изменений такова, что производительность устройств удваивается, а цена падает в два раза каждые девять месяцев для сетевых устройств, каждые 12 — для устройств хранения данных и каждые 18 — для процессоров. Экономически выгодным становится вкладывать деньги в технологии, использующие сетевые устройства. Появление фундаментальных стандартов платформы XML (RDF, RDFS, OWL, SOAP, WSDL и т. д.), а также подходов Semantic Web и GRID создало возможность

использовать семантические особенности информационных ресурсов и переходить к глобальной компьютерной обработке типовых информационных ресурсов. Именно на этом пути стоит ожидать очередного прорыва в исследованиях окружающей среды.

Список литературы

- [1] ФАЗЛИЕВ А.З. Информационные ресурсы и Интернет-технологии для наук об окружающей среде // Вычисл. технологии. 2004. Т. 9. Спецвыпуск: Тр. Междунар. конф. “Вычислительно-информационные технологии для наук об окружающей среде” (CITES 2003). Ч. 1. С. 11–21.
- [2] BERNERS-LEE T., BRICKLEY D., CONNOLLY D. ET AL. The Semantic Web. (<http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/slide1-0.html>).
- [3] HERMAN I. Introduction to the Semantic Web. (<http://www.w3.org/2003/Talks/1112-BeijingSW-1H>).
- [4] DE ROURE D., JENNINGS N., SHADBOLT N. A Future e-Science Infrastructure. Report commissioned for EPSRC/DTI Core e-Science Programme. 2001. 78 p.
- [5] ГОРДОВ Е.П., ФАЗЛИЕВ А.З. Научные информационные ресурсы для поддержки исследований об атмосфере в сети Интернет // Вычисл. технологии. 2004. Т. 9. Спецвыпуск: Тр. Междунар. конф. “Вычислительно-информационные технологии для наук об окружающей среде” (CITES 2003). Ч. 1. С. 123–136.
- [6] ДМИТРИЕВ Б.Н., СУТОРИХИН И.А. Базы данных для ГИС “Аэрозоли Сибири” // Тр. Междунар. конф. “ENVIRONMIS-2002” / Под ред. Е.П. Гордова. Томск: Изд-во ГУ “Томский ЦНТИ”, 2002. Т. 1. С. 70–77.
- [7] ГОРДОВ Е.П., DE RUDDER A.D., ЛЫКОСОВ В.Н. и др. Веб-портал АТМОС как основа для выполнения интегрированных исследований по окружающей среде Сибири // Вычисл. технологии. Т. 9. 2004. Спецвыпуск: Тр. Междунар. конф. “Вычислительно-информационные технологии для наук об окружающей среде” (CITES 2003). Ч. 2. С. 3–13.
- [8] МОЛОРОДОВ Ю.И., КУЦЕНОГИЙ К.П. Атлас по атмосферным аэрозолям Сибири как основа обеспечения мониторинга Сибирского региона // Мониторинг окружающей среды, геоэкология, дистанционные методы зондирования земли: Сб. науч. тр. Новосибирск: СГГА, 2005. Т. 5. С. 142–146.

Поступила в редакцию 19 октября 2006 г.