

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 7

Москва 2007

ОБЩИЙ РАЗДЕЛ

УДК [002:004.89]:55

Ю. М. Арский, А. В. Веселовский, А. Н. Шогин

Модель интегрированной информационно-аналитической системы сопровождения фундаментальных исследований в области наук о Земле

Рассматриваются вопросы проектирования интегрированных информационно-аналитических систем с использованием математической модели объединения научно-исследовательских институтов в группы и выбором центра группы портала сети Интернет. При моделировании учитываются тематическое сходство задач, решаемых в области наук о Земле, и экономические характеристики. Приводится краткое описание построенной на основе компьютерного моделирования системы тематических интернет-порталов, замыкающихся на центральный портал (в ВИНТИ РАН) в проекте Президиума РАН "Электронная Земля".

ВВЕДЕНИЕ

Создание интегрированных информационно-аналитических систем (ИАС) вызвано необходимостью внедрения современных информационных технологий в области естественных наук, информатизации науки и производства.

Выбор оптимальных стратегий обработки информационных потоков рассматривается в качестве главной задачи информатизации научных центров. Решение этой задачи требует организации системного применения современных информационных технологий и создания новых информационных технологических структур, в частности, ин-

тегрированных ИАС отдельных институтов или группы институтов, выполняющих исследования в определенных областях наук (в рассматриваемом случае — наук о Земле).

Развитие информатизации в соответствии с мировым уровнем требует изучения специфики научных исследований, экономических особенностей сопровождения междисциплинарных исследований в области наук о Земле. Для этого необходимы развитая тематическая классификация и автоматизированное индексирование информационных ресурсов, учет стоимости оборудования, программных продуктов, приобретаемых баз данных и других составляющих, обеспечивающих создание и функционирование ИАС. С учетом экономических аспектов построения ИАС для сопровождения фундаментальных исследований в области наук о Земле осуществляется моделирование процесса распределенной обработки запроса пользователя и интегрирования данных различного характера на пространственной основе.

Исследователи, работники сфер управления и производства с помощью ИАС могут получать оперативный доступ к информационным ресурсам и технологиям современных центров научной информации, глобальных информационных и вычислительных сетей [1].

Построение интегрированных ИАС позволит научным центрам Отделения наук о Земле РАН принимать равноправное участие в академических, федеральных и международных проектах информатизации, в научно-исследовательских программах, решающих задачи по использованию, развитию и поддержанию информационного пространства РАН.

В соответствии с общими тенденциями развивается телекоммуникационная среда передачи данных, наращиваются информационные ресурсы, подключаются внешние информационные системы и механизмы предоставления услуг пользователю.

ОСНОВНЫЕ ЗАДАЧИ, РЕШАЕМЫЕ ИНТЕГРИРОВАННЫМИ ИАС

Проблема создания интегрированных ИАС является ключевой для полноценного информационного обеспечения научных центров, занимающихся междисциплинарными исследованиями в области наук о Земле. Для этого следует в первую очередь решить следующие фундаментальные задачи:

- создание в Отделении наук о Земле РАН информационного ядра как единой информационной среды с использованием системы интернет-порталов, WEB-, GIS- и GRID-технологий и эффективной развитой телекоммуникационной инфраструктуры;
- организация доступа к информационным ресурсам, имеющимся в институтах, научных центрах Отделения наук о Земле РАН и в сети Интернет;
- введение стандартов представления информации о ресурсах (метаданных), учитывающих интересы пользователя и владельца информационных ресурсов, обеспечивающих доступность данных в оперативных режимах работы и управляемость потоками информации в рамках Отделения наук о Земле РАН;

- внедрение современных информационных технологий в процессы сбора, обработки и распространения научной информации, формирования новых информационных продуктов, поддержки исследований в научных центрах;
- разработка сетевых технологий обработки и анализа данных с использованием вычислительных кластеров и распределенных специализированных веб-сервисов в среде Интернет;
- создание классификационной модели информационных ресурсов интегрированных ИАС.

В ряде отделений Российской академии наук — в Отделении наук о Земле, в Отделении информационных технологий и вычислительных систем, в Отделении математических наук и в региональных отделениях — к настоящему времени при проведении фундаментальных и прикладных исследований накоплен значительный объем баз и банков научных данных (как фактографических, так и документальных) по различным отраслям знания, созданы основы межсетевых связей, разработаны методы математического моделирования, информационные технологии и системы для науки и высшей школы, производственных отраслей и органов государственной власти. Это и есть базис решения задач, поставленных перед интегрированной ИАС.

Распределенная интегрированная ИАС представляет собой целостную технологическую и программную среду взаимодействия научно-исследовательских лабораторий институтов и организаций РАН, что предусматривает разработку алгоритмов функционирования интернет-портала для использования баз данных, баз знаний, цифровых карт и аэрокосмических снимков. ИАС должна обеспечивать поддержку научных исследований в институтах, обмен сетевыми информационными ресурсами или информацией об этих ресурсах (метаданными информационных ресурсов) на основе единых принципов и открытых стандартов.

Достигнутые в настоящее время в проекте “Электронная Земля” результаты работ по созданию системы GeoSINet, содержащей шесть интернет-порталов, интегральный банк данных, развитую информационную инфраструктуру с локальными вычислительными сетями и телекоммуникациями, методические материалы по построению интегрированных ИАС, а также уровень профессиональной подготовки разработчиков системы позволяют считать реальным решение поставленных выше задач.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСПРЕДЕЛЕНИЯ И ОРГАНИЗАЦИИ ДОСТУПА ЧЕРЕЗ СЕТЬ ИНТЕРНЕТ К ИНФОРМАЦИОННЫМ РЕСУРСАМ СИСТЕМЫ ПОРТАЛОВ ИНТЕГРИРОВАННОЙ ИАС

Система интернет-порталов “Электронная Земля” построена на основе математического моделирования структуры информационных ресурсов и механизма доступа к ним пользователей. Качественную основу модели системы порталов составляют идеи формирования информационного пространства в области наук о Земле на базе метода стратифицированного моделирования [2].

Моделирование структуры интегральной ИАС позволяет оптимизировать процесс организации системы, содержащей средства распределенной обработки и интеграции разнообразных данных в широкой тематической области научных исследований.

В условиях достаточно большого количества научных центров РАН, относящихся по тематике исследований к наукам о Земле, очевидна необходимость создания диспетчерских узлов — тематических порталов в сети Интернет [3]. Их целесообразно организовать на основе имеющихся академических институтов с развитием технологического оснащения. Предлагаемая модель представляет возможность по количественным критериям осуществить выбор центрального портала и построение иерархической структуры системы серверов научно-технических центров, тематических интернет-порталов.

При моделировании возникает необходимость декомпозиции интегральной ИАС на подсистемы и элементы для оптимизации технологических и управленческих процессов. Пусть имеется конечное множество $W = \{W_i\}$ потребителей информации в области наук о Земле и множество информационных потребностей $V = \{V_i\}$. Удовлетворение этих потребностей осуществляется конечным множеством структурно организованных серверов тематически специализированных научных центров $F = \{f_i\}$, где $i = \overline{1, n}$; n — общее число серверов (центров). Структура схемы доступа пользователей к информационным ресурсам системы “Электронная Земля” содержит $J = \overline{1, m}$ уровней с центрами $q = \overline{1, r}$ статусов, причем $m < r$ (рис. 1).

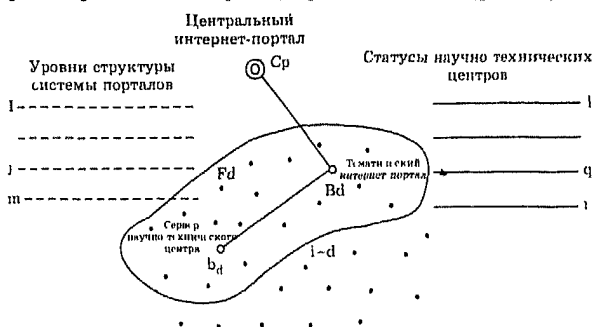


Рис 1 Схема доступа к информационным ресурсам из подмножества F_d серверов научных центров системы интернет-порталов “Электронная Земля”

Исходное множество серверов F разбивается по тематике (рубрики классификатора) на M непересекающихся подмножеств F_d :

$$F = \bigcup_{d=1}^M F_d; F_d \cap F_m = \emptyset.$$

Требуется из каждого подмножества (объединения) элементов B_d выбрать базовый элемент $B_i \in B$, на основе которого формируется тематический интернет-портал:

$$B \cap B_1 = b_1; \dots B \cap B_M = b_m; \bigcup_{d=1}^M B_d = B,$$

где B — множество базовых элементов (порталов)

Таким образом, решается некоторое множество задач V в условиях определенной иерархической

структуры. Тематические порталы сопоставляются с подмножествами V_d задач $\bigcup_{d=1}^M V_d = V$. Формализация разбиения соответствует множеству $H = \{h\}$ различных решений. Эффективность E функционирования интегральной ИАС зависит от принятых технологий, программно-математического обеспечения, схем доступа, полноты информационных ресурсов и др.: $E = E(h)$, а оптимизация процесса принятия решения может быть представлена как $\max E(h), h \in H$.

На группы элементов множество разбивается по принципу наибольшей тематической взаимосвязанности объединяемых центров научной информации (элементов). Близость элементов может быть также истолкована по однородности информационных потоков, наличию общих материальных и финансовых ресурсов, однотипности задач информационного обслуживания, административной принадлежности и другим параметрам. Для оценки такой близости элементов вводятся коэффициенты C'_{ij} , посредством которых оценивается мера связности элементов i и j с точки зрения целей декомпозиции системы. Коэффициенты C'_{ij} в общем случае представляют собой величины, обратные затратам на обработку и передачу по каналам связи информации, и должны удовлетворять условию:

$$0 \leq C'_{ij} \leq 1; C'_{ij} = \frac{1}{1 + S_{ij}}$$

где S_{ij} — затраты на обработку и передачу информации в специализированном научном центре.

Формирование структуры доступа к информационным фондам интегрированной ИАС с учетом приведенных соображений осуществляется в два этапа: определение коэффициентов связности центров C'_{ij} (первый этап) и моделирование организационных (включая топологические) структур при заданных значениях коэффициентов C'_{ij} (второй этап).

Рассмотренный подход к декомпозиции ИАС реализуется в виде конкретных алгоритмов, построенных применительно к разного рода моделям тематической классификации.

Число подмножеств M , на которое разбивается исходное множество, заранее не задается, а определяется после реализации алгоритма:

если $F = \{B\}$ — множество научно-технических центров, $A = \{(b_i, b_j)\}$ — множество связей между центрами и $b_i, b_j \in F$, то множеству F ставится в соответствие симметрическая матрица $C = \|C'_{ij}\|$.

СИСТЕМА ТЕМАТИЧЕСКИХ ИНТЕРНЕТ-ПОРТАЛОВ ИНТЕГРИРОВАННОЙ ИАС “ЭЛЕКТРОННАЯ ЗЕМЛЯ”

Построение интегрированной ИАС учитывает специфику исследований, уровень информатизации на текущий момент научных центров, на основе которых создается система интернет-порталов, состояние информационных ресурсов, требования стандартов и нормативно-правовых документов к

системам этого класса. Реализация ИАС обуславливает необходимость анализа и интеграции информационных, коммуникационных и технологических ресурсов конкретных научных центров. Разработка интегрированной ИАС опирается также на создание классификационной модели информационных ресурсов системы. Этот подход относится к различным видам и формам представления данных.

Описываемая математическая модель реализована в системе тематических интернет-порталов интегрированной ИАС "Электронная Земля" (рис. 2) поэтапно — создание, в первую очередь, тематических интернет-порталов на основе институтов Отделения наук о Земле РАН, а затем — формирование интегральной ИАС.

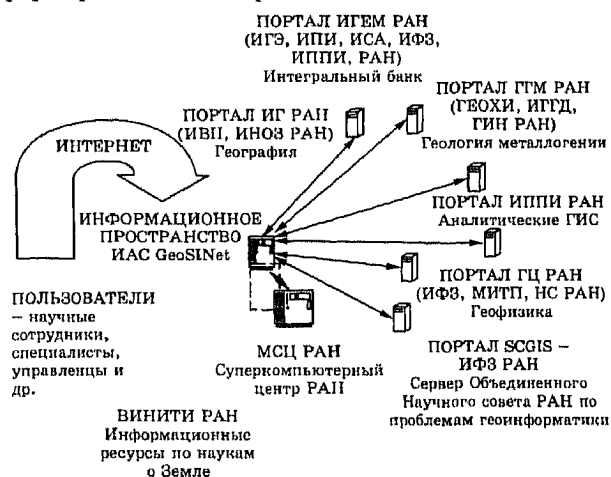


Рис. 2. Структура информационного пространства интегрированной ИАС "Электронная Земля"

Система интернет-порталов "Электронная Земля" имеет иерархическую архитектуру с многоуровневым доступом пользователей к ресурсам, что обеспечивает их оперативный поиск, обмен данными между порталами, а также предоставление пользователю соответствующего интерфейса к средствам решения определенной научной задачи и типового описания выполнения пользовательских заданий:

- Центральный портал (ВИНИТИ РАН);
- Тематические порталы участников комплексного проекта "Электронная Земля":
 - портал ГГМ РАН — совместно с ГЕОХИ, ИГГД, ГИН РАН,
 - портал ГЦ РАН — совместно с ИФЗ, МИТЦ, ИС РАН,
 - портал ИГЕМ РАН — совместно с ИГЭ, ИПИ, ИСА, ИФЗ, ИППИ РАН,
 - портал ИГ РАН — совместно с ИВП, ИГЕМ, ИНОЗ РАН,
 - портал ИППИ РАН,
 - портал Объединенного научного совета РАН по проблемам геоинформатики при Президиуме РАН.

Важная составляющая работ второго этапа — геоинформационное сопровождение научных тем, включающее подготовку информационных материалов, формирование и актуализацию БД научно-технических центров, использующих ИАС.

Предлагаемая математическая модель и ее реализация в системе GeoSINet проекта "Электронная Земля" ориентированы на использование в создаваемой системе современных информационных технологий, высокий уровень информационного обслуживания научных работников, не уступающие зарубежным аналогам.

Специфика исследований в области наук о Земле, тематическая междисциплинарность, а также необходимость организации взаимодействия с академическими, отечественными и зарубежными информационными системами определяют структурные и технологические особенности создания интегрированной ИАС.

Разработанная ИАС используется для повышения уровня информационной и технологической поддержки научных работ, а также для построения информационного пространства Отделения наук о Земле РАН.

Полученный за последние три года опыт проектирования и экспериментального функционирования системы GeoSINet показал, что предложенные принципы разработки интегральной ИАС обеспечивают экономичность в эксплуатации, междисциплинарность и, на ее основе, комплексность решения актуальных теоретических и прикладных проблем. В составе информационных ресурсов системы "Электронная Земля" в значительном объеме имеются сведения методического и фактографического характера по тематике геологии рудных месторождений, геофизики, природопользования, географии, охраны окружающей среды, геоэкологии, радиоэкологии, минералогии и другим дисциплинам наук о Земле.

СПИСОК ЛИТЕРАТУРЫ

1. Арский Ю. М., Веселовский А. В., Никитин А. Н. и др. Направления работ по организации интегрального банка данных и информационной инфраструктуры Отделения наук о Земле РАН // Проблемы окружающей среды и природных ресурсов, № 7. — М.: ВИНТИ, 2004. — С. 3-17.
2. Веселовский А. В., Красавин К. Г. Стратифицированное математическое моделирование единого информационного поля академической научной организации // Межотраслевая информационная служба. — 2003. — № 2. — С. 26-34.
3. Веселовский А. В., Стесик Г. Л. Формирование и использование информационных ресурсов в проекте "Электронная Земля" // Геоинформатика. — 2006. — № 4. — С. 34-38.

Материал поступил в редакцию 18.04.07.