УДК 004.62:004.92

Методы управления и геоинформационного моделирования в технологии OLAP

Л.Ф. Ноженкова*, А.А. Евсюков, А.И. Ноженков

Институт вычислительного моделирования СО РАН, 660036 Россия, Красноярск, Академгородок¹

Получена 11.06.2008, окончательный вариант 15.10.2008, принята к печати 10.03.2009

Представлены методы оперативной аналитической обработки данных на основе интеллектуальных средств управления OLAP-моделированием. Предложены средства метаописания сложных комплексов OLAP-моделей, что позволило реализовать управление расчетами и автоматизировать обработку данных с применением эвристик. Описаны средства оперативного тематического геоинформационного моделирования для визуализации результатов OLAP-анализа.

Ключевые слова: обработка данных, OLAP-анализ, OLAP-модель, геоинформационное моделирование.

Введение

Технология оперативной аналитической обработки многомерных данных OLAP (On-line Analytical Processing) [1] – быстро развивающееся направление интеллектуального анализа данных. На сегодняшний день существует достаточно большое количество программных продуктов, реализующих функции OLAP-анализа. Такие крупные компании, как Hyperion Solutions Corp., IBM, Oracle, Microsoft, Sybase, Panorama Software, Cognos Inc. и другие, ведут разработки в этой области, их решения охватывают практически все существующие задачи. Из отечественных разработчиков следует отметить BaseGroup Labs, Intersoft Lab., a также Институт открытых систем при Ивановском городском энергетическом университете и Институт вычислительного моделирования СО РАН. В программных продуктах реализованы различные способы хранения многомерных структур, средства быстрого извлечения и представления информации, средства быстрого создания приложений. Тенденции развития функциональности ОLAР-продуктов в настоящий момент лежат в области создания продуктов класса data mining для постобработки данных, развития средств картографической и трехмерной визуализации данных, веб-сервисов для OLAP-приложений, а также средств, позволяющих расширить область применения OLAP за счет новых подходов к решению сложных прикладных задач.

Наибольшее применение эта технология получила в бизнес-сфере для решения оперативных задач, каждая из которых, как правило, укладывается в рамки одного куба. При этом классические OLAP-продукты малоэффективны для использования в задачах, где необходим комплексный анализ данных, связанный с реализацией сложных аналитических алгоритмов. Методы расчета аналитических показателей и решения задач планирования в здравоохранении, образовании и других областях организационного управления представляют собой сложные многошаговые процессы анализа многомерных данных. Возникает необходимость их поэтапной обработки средствами OLAP. Потребовался новый подход в

^{*} Corresponding author E-mail address: expert@icm.krasn.ru

C Siberian Federal University. All rights reserved

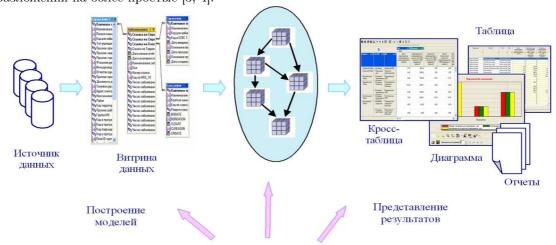
OLAP-технологии, ориентированный на решение сложных аналитических задач, использующих связные многошаговые расчеты с множеством информационных объектов.

Применение OLAP-технологии для сложных территориально ориентированных задач требует развития функций оперативного геоинформационного моделирования. Необходимо не просто визуализировать результаты OLAP-анализа, но и отображать процесс оперативной обработки, выполнять динамическое формирование тематических картографических слоев.

Новый подход к решению разнообразных задач с применением OLAP-технологии основан на построении комплексов так называемых OLAP-моделей [2–6]. В статье изложены результаты работ по развитию технологии OLAP-моделирования в направлении наглядного представления сложных методик, автоматизации управления расчетами и оперативного геоинформационного моделирования результатов. Представлены средства онтологического описания сложных комплексов OLAP-моделей и управления расчетами, позволяющие автоматизировать применение экспертных методов, в частности эмпирических критериев согласования и балансировки расчетных показателей [7, 8]. Описаны методы оперативного геомоделирования результатов OLAP-анализа путем динамического формирования картографических слоев, основанные на картографической привязке многомерных данных OLAP-системы к территориальным объектам ГИС [9, 10].

Построение управляемых комплексов OLAP-моделей

Новый подход в OLAP-технологии ориентирован на решение сложных аналитических задач, использующих связные многошаговые расчеты с множеством информационных объектов, представленных многомерными кубами данных. На рис. 1 схематически показан процесс решения сложной задачи путем реализации комплекса так называемых OLAP-моделей. Термин "OLAP-модель" введен для повышения наглядности описания сложных задач путем разложения на более простые [3, 4].



Построение комплекса OLAP-моделей

Рис. 1. Схема построения комплексов OLAP-моделей

 ${
m OLAP}$ -модель строится пользователем и несет в себе описательную информацию о решении некоторой аналитической задачи. ${
m OLAP}$ -модель M можно представить в виде

- $M = \langle X, G, \Psi(X,G), Q(G), I, J(G) \rangle$.
- 3десь X дискретное множество входных данных витрина данных;
- $G = \langle D, F \rangle$ гиперкуб модель логического многомерного представления данных, характеризующаяся двумя наборами параметров измерениями и показателями;
- $D = \langle d_1, d_2, \dots, d_m \rangle$ измерения гиперкуба: каждое измерение представляет собой упорядоченное множество значений определенного типа. Измерения могут быть организованы в виде составной иерархии;
- $F = \langle f_1, f_2, ..., f_n \rangle$ показатели (меры) гиперкуба: каждый показатель представляет множество значений, количественно характеризующих анализируемый процесс;
- $\Psi(X,G)$ функции, описывающие построение элементов гиперкуба G: показателей F и измерений D, исходя из множества входных данных X;
- Q(G) операции над гиперкубом: $Q = \left\langle C(G, \varphi(F, \mu(d_i))), \ V(G, d_1' \dots d_n', \ R(G)) \right\rangle$, где $C(G, \varphi(F, \mu(d_i)))$ агрегирование гиперкуба по иерархии атрибутов выбранного измерения преобразование G к гиперкубу меньшей мощности за счет агрегирования показателей $\varphi(F, \mu(d_i))$ с учетом отношения иерархической зависимости $\mu(d_i)$ атрибутов измерения d_i . В качестве функции φ агрегирования показателей могут выступать, например, min, max, сумма и др.
- $V(G,d_1',\ldots,d_n')$ срез гиперкуба G по измерениям d_1,\ldots,d_n операция получения подмножества гиперкуба в результате фиксации подмножеств значений d_1,\ldots,d_n' измерений d_1,\ldots,d_n соответственно.
- R(G) операция поворота гиперкуба, которая изменяет порядок измерений в гиперкубе. $I=\langle T,\ K,\ L,\ H,\ \gamma(T,K,L,H)\rangle$ модель логического представления результатов вычисления OLAP-модели.

Она включает формы представления результатов модели: T – таблицы, K – кросстаблицы, L – диаграммы, H – карты и $\gamma(T,K,L,H)$ – операции над ними: фильтрация, сортировка таблиц, перемещение, ротация строк и столбцов таблицы, фильтрация, сортировка кросс-таблиц, группировка данных, сортировка, разбиение на сегменты диаграмм, фильтрация карт.

- $J(G) = \langle T, \alpha(T) \rangle$ операция сохранения гиперкуба в таблицу агрегатов. Операция применяется к представлению гиперкуба в виде таблицы, или кросс-таблицы.
- $\alpha(T)$ преобразование наименований в код, наложение внешних ключей, удаление таблицы агрегатов перед вставкой данных операции по трансформации таблицы агрегатов.

К характеристикам OLAP-модели относятся: мощность измерения, мощность, размерность и объем гиперкуба. Количество элементов измерения есть мощность измерения. Мощность гиперкуба – произведение мощностей его измерений. Объемом гиперкуба называется произведение мощности куба и количества показателей.

OLAP-модель состоит из исходных данных – витрины данных, информационного куба, операций над ним и способов представления результатов вычисления. Важным моментом построения модели является возможность сохранения в источнике данных результатов расчета. Такой подход позволяет применять поэтапный процесс анализа данных, то есть анализировать ранее полученные результаты.

Для автоматизации поэтапного анализа модели объединяются в комплексы (см. рис. 2). Перед созданием комплекса моделей задачу необходимо декомпозировать на подзадачи таким образом, чтобы каждая подзадача могла быть представлена OLAP-моделью [3]. В рамках одного расчета модели образуют последовательно выполняемую цепочку, при этом данные, рассчитанные одной моделью, в дальнейшем используются другими моделями. Таким

образом, OLAP-модели образуют комплекс моделей, а их связи осуществляются через наследование информации.

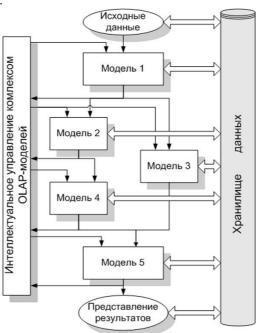


Рис. 2. Схема исполнения комплекса OLAP-моделей

Выполнение комплекса аналитических моделей сопровождается так называемым интерактивным аналитическим экспериментом, т. е. возможно вмешательство пользователя в выполнение расчета для модификации параметров и настройки отдельной модели.

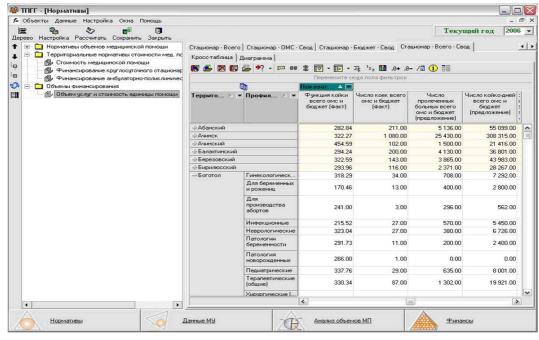


Рис. 3. OLAP-модели в задачах анализа и планирования медицинской помощи в здравоохранении

На рис. З представлен фрагмент расчета комплекса OLAP-моделей для задачи формирования территориальной программы государственных гарантий оказания бесплатной медицинской помощи жителям Красноярского края [7]. Это задача ежегодного планирования объемов и видов медицинской помощи в регионе. Для автоматизации этого процесса построено более сорока OLAP-моделей. Формирование территориальных программ требует многовариантных расчетов и балансировки показателей, рассчитываемых OLAP-моделями.

Чтобы повысить наглядность построения и расчета комплексов OLAP-моделей для сложных задач, обеспечить управление расчетами и многовариантные расчеты, разработаны средства метаописания комплекса моделей с применением онтологии [8].

Онтология комплекса OLAP-моделей имеет вид

 $U = \langle W, Y, R \rangle$, где W — описание моделей, Y — отношения между моделями, R — условия применения и коррекции моделей.

Описание OLAP-модели задается в виде кортежа:

M: <витрина данных>; <назначение модели>; <результат>.

Описание каждой модели в базе знаний формируется автоматически, с помощью специально разработанного программного обеспечения, на основе системного описания модели, которое формируется в процессе ее интерактивного создания.

Основной вид отношений между правилами – отношение информационной зависимости моделей $M_1 \to M_2$ – используется для построения цепочек расчета OLAP-моделей. Дополнительно могут использоваться отношения альтернативности моделей, информационной независимости и др.

Условия применения моделей описываются с помощью правил следующего вида.

 R_1 : ЕСЛИ $P(f_1, f_2, ..., f_n)$, ТО ВЫПОЛНИТЬ М / ИЗМЕНИТЬ М

Здесь R_1 — уникальное в базе знаний имя правила. f_1, f_2, \ldots, f_n — показатели, заданные как результаты расчета какой-либо модели или заданные в витрине как исходные данные. Каждый показатель представляет собой многомерный куб, описанный в витрине данных соответствующей ОLAP-модели. Предикат P задается в виде логико-лингвистического выражения. Операция BЫПОЛНИТЬ M в правой части правила R_1 интерпретируется как выполнение расчета ОLAP-модели M. Если в процессе логического вывода правило R_1 применяется, то это приводит к расчету показателей — результатов модели M. Как следствие, должны быть пересчитаны все информационно зависимые OLAP-модели. Операция M3MEHUTb M в правой части правила интерпретируется как переход к интерактивному процессу коррекции модели M.

В качестве примера рассмотрим правило:

 $R1: ECЛИ\ (Объем\ финансирования\ расчетный < Объем\ финансирования\ выделенный)$ И (Нормативы объемов помощи на тысячу населения \leq Нормативы объемов помощи на тысячу населения $P\Phi$)

ТО ВЫПОЛНИТЬ Увеличить обеспеченность медпомощи в сельских районах

Интерпретация этого правила заключается в том, что многомерные показатели Объем финансирования расчетный, Объем финансирования выделенный, Нормативы объемов помощи на тысячу населения, Нормативы объемов помощи на тысячу населения РФ скалярно сравниваются по измерениям – в данном случае по видам и профилям медицинской помощи. Действие в правой части правила заключается в выполнении ОLAP-модели Увеличить обеспеченность медпомощи в сельских районах.

Алгоритмы формирования и расчета комплексов OLAP-моделей реализованы на основе использования представленного в базе знаний онтологического метаописания OLAP-

моделей. Управление процессом расчета выполняется на основе правил, а также с учетом действий пользователя. Пользователь может интерактивно изменить или заменить любую OLAP-модель, вследствие чего изменяется системное представление OLAP-модели. Пользователь также может изменить условия ее применения, изменив метаописание в базе знаний. Процедура логического вывода учитывает оба варианта. При этом автоматически исследуются возможные изменения расчета и строятся новые цепочки моделей для расчета показателей.

Средства оперативного географического моделирования

В основе механизма динамической связи многомерных данных ОLAP-системы к пространственной информации ГИС лежит картографическая привязка данных, позволяющая устанавливать соответствие между результатами оперативного аналитического моделирования и географическими объектами [9].

На основе таблиц источника данных OLAP-системы осуществляется построение информационного гиперкуба, формируются аналитические объекты, описывающие предметную область: показатели и измерения. Для осуществления картографической привязки в гиперкубе выделим географическое измерение \hat{d} .

Для построения тематической карты зафиксируем значения измерений $d_i \subseteq D: d_i \neq \hat{d}, i = \overline{1,m}$ и из F выберем показатель, который будет отображаться на карте: $f^* = f^*(d_1^*, d_2^*, \ldots, \hat{d}, \ldots, d_m^*)$, где $d_1^*, d_2^*, \ldots, d_m^*$ – фиксированные значения измерений (метки), кроме \hat{d} , которое не фиксировано и соответственно может принимать любое значение из своей области определения. Построение множества значений показателя f^* определим как операцию среза над гиперкубом данных G по всем фиксированным измерениям.

Для отображения значений показателя f^* определим электронную карту $H = \langle L, S \rangle$, где $L = \{L_1, L_2, \ldots, L_p\}$ – множество картографических слоев; $S = (s_1, s_2, \ldots, s_q)$ – упорядоченное множество значений свойств отображения карты.

Слой L_i определяется как $L_i = \langle O_i, T_i, A_i(O_i, T_i), P_i \rangle$, где $O_i = (o_i^1, o_i^2, \dots, o_i^{f_i})$ – упорядоченное множество территориальных объектов заданного типа;

 $T_i = (t_i^1, t_i^2, \dots, t_i^{g_i})$ – упорядоченное множество атрибутивных свойств слоя – полей атрибутивной таблицы A_i ;

 $A_i(O_i,T_i)$ – атрибутивная таблица, элементы которой $a_i^{j,k}$ определены значениями из набора T_i для каждого объекта из O_i , здесь $j=\overline{1,f_i}$ определяет строки таблицы, а $k=\overline{1,g_i}$ – столбцы таблицы (поля);

 $P_i = \langle B_i, C_i \rangle$ — легенда слоя, здесь $B_i = (b_i^1, b_i^2, \dots, b_i^{v_i})$ — упорядоченное множество значений свойств слоя, $C_i = (c_i^1, c_i^2, \dots, c_i^{w_i})$ — упорядоченное множество классов разбиения множества объектов O_i слоя L_i .

Привязка географического измерения \hat{d} осуществляется к одному или нескольким картографическим слоям. Пусть $L'\subseteq L$ – подмножество слоев, к которым осуществляется привязка, и мощность его |L'=l|. Каждый слой содержит атрибутивную таблицу $A_i',\ i=\overline{1,l},$ в одном из полей таблицы $t_i^{k_i^*},\ k_i^*\in[1,g_i]$ хранится множество идентификаторов объектов слоя. Если значение идентификатора $a_i^{j,k_i^*}\in d,\ j=\overline{1,f_i},$ то к объекту o_i^j можно построить картографическую привязку значения показателя f^* . Используя картографическую привязку данных, устанавливается соответствие между территориальными объектами карты и географическим измерением гиперкуба \hat{d} .

Чтобы значения показателя f^* отобразить в виде тематической карты, необходимо построить легенду тех слоев, к объектам которых выполнена картографическая привязка. При построении легенды используются методы тематического картографирования, позволяющие объекты слоя разбить на классы на основе соответствующих значений анализируемого показателя. Для того чтобы на карте различать объекты разных классов, для каждого класса задается ряд настроек отображения: цвет объектов, цвет контура и тип заливки объектов для площадного слоя, условное обозначение (символ) объектов для точечного слоя, состояние (видимые, скрытые или мигающие объекты). Построенная в результате картографической привязки многомерных данных тематическая карта является отображением значений анализируемого показателя f^* из среза гиперкуба данных по всем измерениям, кроме географического измерения \hat{d} .

Помимо инструментов построения легенды для представления результатов OLAPанализа в программную реализацию средств оперативного геомоделирования включен блок управления многомерными данными. Блок управления включает в себя список аналитических показателей, доступных для построения тематических карт, а также фильтр многомерных данных. При изменении анализируемого показателя и фиксированных измерений происходит переформирование классов привязанных к многомерным данным слоев и выполняется построение тематической карты.

На рис. 4 представлен пример построения тематической карты: отображена рождаемость населения по территориям Красноярского края. Картографическая привязка данных осуществлена к картографическим слоям "города" и "районы". Территориальным объектам со светлой раскраской соответствуют более низкие показатели рождаемости, раскрашенным темнее – более высокие.

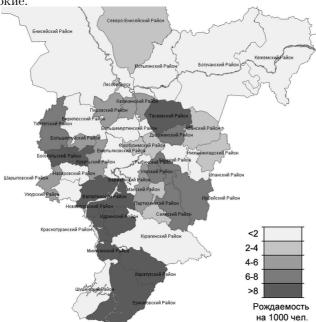


Рис. 4. Отображение рождаемости в Красноярском крае

Для реализации оперативной работы с территориальными объектами, меняющими свое местоположение, предлагается метод динамического формирования картографических слоев на основе содержимого таблиц из источников данных OLAP-системы и топографических

картографических слоев [9]. Наполнение формируемого слоя – его пространственная и атрибутивная информация – создается на основе содержания выбранной таблицы источника данных OLAP-системы. Обозначим выбранную таблицу T. Для решения задачи динамического формирования новых картографических слоев разработаны алгоритмы получения пространственной информации, позволяющие найти координаты точек для построения объектов нового точечного, линейного и площадного слоев на основе значений координат таблицы из источника данных OLAP-системы или соответствия объектам слоя из топографической основы карты. Если географические координаты содержатся непосредственно в таблице T, то их можно использовать для формирования нового слоя. При отсутствии необходимой информации о местоположении объектов в таблицах источников данных OLAPсистемы предлагается формировать координаты точек нового слоя на основе картографической привязки к слоям из топографической основы карты. Для этого необходимо установить соответствие между полем таблицы T и полем атрибутивной таблицы слоя привязки. Для совпадающих значений формируется объект нового слоя, при этом используются координаты соответствующего ему объекта слоя карты. Атрибутивной информацией слоя является таблица, полученная из строк таблицы T, по которым сформированы объекты нового слоя. Отметим, что можно динамически сформировать и визуализировать одновременно несколько картографических слоев. ГИС позволяет производить с динамически сформированными слоями те же операции, что и со слоями из топографической основы карты, включая применение методов тематического картографирования.

Рассмотрим пример динамического формирования картографических слоев на основе оперативного аналитического моделирования. На рис. 5 изображен фрагмент карты Большемуртинского района Красноярского края. Тематическая карта дает представление о планах реорганизации сети медицинских учреждений. Для визуализации планирования реорганизации сети медицинских учреждений каждому учреждению присваивается статус: "без изменений", "модифицировано", "закрыто", "открыто". Динамически формируются слои медицинских учреждений на основе информации об их принадлежности населенным пунктам [10].

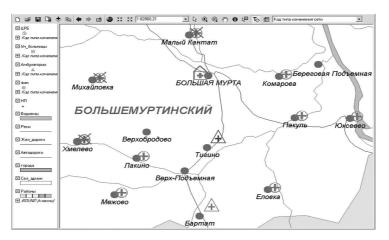


Рис. 5. Планирование реорганизации сети медицинских учреждений Обозначения:

— центральные районные больницы,
— участковые больницы,
— амбулатории,
— фельдшерско-акушерские пункты,

— населенные пункты.

Для формирования новых слоев могут быть использованы таблицы агрегатов, содержа-

щие агрегированные данные, прошедшие предварительную обработку. Это позволяет сформировать новый слой не только на основе собранных статистических данных, но и на основе аналитических результатов. Использование таблиц агрегатов для динамического формирования картографических слоев является отличительной особенностью средств оперативного геомоделирования в OLAP-системе по сравнению с традиционными геоинформационными системами.

Заключение

Средства управления OLAP-моделированием позволяют наглядно представлять процесс построения и выполнения комплекса OLAP-моделей, редактировать алгоритм расчета и интерактивно управлять процессом. Обеспечивается возможность не только автоматизировать расчетные задачи, но и реализовать эвристические и эмпирические методы, как, например, согласование результатов расчета нормативов объемов медицинской помощи и нормативов финансирования в планировании медицинской помощи.

На основе интеграции OLAP и ГИС реализована технология оперативного геомоделирования для решения территориально ориентированных задач в реальном времени. Интеграция ГИС и OLAP-системы способствует повышению наглядности представления результатов аналитической обработки данных, сочетая ее с геомоделированием. Возможности динамической связи картографических объектов с многомерными данными существенно расширяют функциональность как OLAP-системы, так и встроенной в систему ГИС. В результате интеграции OLAP-система приобретает дополнительные возможности наглядного представления многомерных данных, ГИС – инструментарий формирования аналитических запросов для построения тематических карт.

Средства управления OLAP-моделированием и оперативного геомоделирования нашли применение в решении таких важных задач здравоохранения, как анализ сети медицинских учреждений, ежегодное планирование медицинской помощи, анализ состояния социальной помощи и многих других. Средства оперативного геомоделирования применяются также в системах поддержки принятия решений по предупреждению природных и техногенных чрезвычайных ситуаций [11].

Работа выполнена при поддержке гранта Президента для ведущих научных школ № HIII-3431.2008.9.

Список литературы

- [1] Codd, E.F. Providing OLAP. On-line Analytical Processing to User-Analists: An IT Mandate / E.F. Codd, S.B. Codd. C. T. Salley, E. F. Codd & Associates, 1993.
- [2] Виноградов К.А. Информационные технологии в управлении региональным здравоохранением / К.А. Виноградов, Е.Е. Корчагин, М.И. Никитина, Л.Ф. Ноженкова. – Красноярск.: КМИАЦ, 2004. – 312 с.
- [3] Вайнштейн Ю.В. Планирование медицинской помощи с применением аналитических OLAP-моделей / Ю.В. Вайнштейн // Вестник Томского государственного университета. Серия "Математика. Кибернетика. Информатика". Приложение №8470, 9(II). Томск: ТГУ, 2004. С. 16-22.

- [4] Ишенин П.П. Инструментальные средства построения комплексов моделей и аналитических приложений в OLAP-технологии: автореф. дис. канд. техн. наук: 05.13.11. / П.П. Ишенин. Красноярск: ИВМ СО РАН, 2006. 24 с.
- [5] Ноженкова Л.Ф. Технологические решения в системах поддержки территориального управления / Л.Ф. Ноженкова // Врач и информационные технологии. 2004. №12.
- [6] Жучков Д.В. Применение технологии хранилищ данных в территориальных органах управления здравоохранением / Д.В. Жучков // Материалы IV Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям. – Красноярск: ИВМ СО РАН, 2003.
- [7] Ноженков А.И. Формирование территориальных программ медицинской помощи на основе интеллектуальных средств управления OLAP-моделированием: автореф. дис. канд. техн. наук: 05.13.01. / А.И. Ноженков. Красноярск: ИВМ СО РАН, 2007. 20 с.
- [8] Ноженков А.И. Интеллектуальные средства OLAP-моделирования в задачах планирования медицинской помощи // Открытое образование (Приложение). Красноярск: OOO "Экспресс-Офсет". 2006. С. 156-160.
- [9] Евсюков А.А. Средства оперативного геомоделирования в информационноаналитических системах: автореф. дис. канд. техн. наук: 05.13.11. / А.А. Евсюков. – Красноярск: ИВМ СО РАН, 2007. – 23 с.
- [10] Евсюков А.А. Оперативное геомоделирование сети медицинских учреждений / А.А. Евсюков, Л.Ф. Ноженкова // Вестник КрасГАУ, № 13. Красноярск: КрасГАУ, 2006. С.114-118.
- [11] Ноженкова Л.Ф. Средства построения систем поддержки принятия решений по предупреждению и ликвидации ЧС / Л.Ф. Ноженкова, С.В. Исаев, В.В. Ничепорчук, А.А. Евсюков, Р.В. Морозов, А.А. Марков// Проблемы безопасности и чрезвычайных ситуаций. №4. М.: 2008. С. 46–54.

Methods for Control and Geoinformation Modeling in OLAP-Technology

Ludmila F. Nozhenkova, Alexander A. Evsyukov and Alexander I. Nozhenkov

Institute of Computational Modeling SB RAS Akademgorodok, Krasnoyarsk, 660036 Russia

In this paper, on-line analytical processing methods based on intellectual control means for OLAP-modeling are presented. Meta-description means of complex OLAP-model systems are suggested, that allows automating of calculation control and data processing with using of heuristics. On-line geoinformation modeling facilities to visualize the results of OLAP-analysis are described.

Key words: data processing, OLAP-analysis, OLAP-model, geoinformation modeling.