

A. V. Korobko, T. G. Penkova

METHOD OF CONCEPTUAL OLAP-MODELLING BASED ON FORMAL CONCEPTUAL ANALYSIS

A method of domain conceptual OLAP-modeling is suggested. The way of constructing the integral analytical model as a formal conceptual OLAP-cubes lattice is formally described. Advantages of on-line analytical processing based on conceptual OLAP-model are presented. Suggested method is implemented for integral analytical model constructing for scientific activities of the organization.

Keywords: conceptual OLAP-modeling, on-line analytical processing, formal conceptual analysis, decision support.

© Коробко А. В., Пенькова Т. Г., 2010

УДК 004.04:004.421

А. И. Ноженков

ПОСТРОЕНИЕ УПРАВЛЯЕМЫХ КОМПЛЕКСОВ OLAP-МОДЕЛЕЙ*

Представлены средства метаописания комплексов OLAP-моделей для решения сложных аналитических задач. Описаны алгоритмы управления расчетами. Полученные результаты применены в задачах территориально-планирования медицинской помощи.

Ключевые слова: OLAP-модель, комплекс OLAP-моделей, управление, многомерные данные, гибридная модель знаний, алгоритмы управления расчетами.

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. Однако традиционные OLAP-решения малоэффективны в прикладных областях организационного управления, где необходим комплексный анализ данных, связанный с реализацией сложных аналитических алгоритмов. Методы расчета аналитических показателей для решения задач оперативного управления и планирования в здравоохранении, образовании, в сфере социальной защиты населения и множестве других прикладных областей представляют собой сложные многошаговые процессы анализа многомерных данных. Потребовался новый подход в OLAP-технологии, ориентированный на решение сложных аналитических задач, использующих связные многошаговые расчеты с множеством информационных объектов.

В статье представлены результаты работ по развитию функциональности OLAP-продуктов в рамках нового подхода, названного комплексным OLAP-моделированием [1; 2]. Описаны средства автоматизации разработки OLAP-приложений для сложных аналитических задач на основе построения управляемых комплексов OLAP-моделей.

Метаописание комплекса OLAP-моделей для сложных задач. Идея технологии OLAP хорошо описана [3–5]. В основе OLAP лежит многомерная логическая модель данных [6]. OLAP-модель (далее – модель) представляет собой описание процессов расчета некоторой аналитической задачи. Структурно аналитическая модель состоит из исходных данных (витрины данных), информационного куба и операций над ним, а также способов представления результатов вычисления.

Для решения сложных аналитических задач обычно строится не одна модель, а целый комплекс OLAP-моделей, каждая из которых решает отдельную подзадачу. Например, чтобы применить эту технологию к задаче планирования медицинской помощи, требуется построить модели анализа и прогноза демографической ситуации в регионе, анализа заболеваемости, анализа соответствия необходимых объемов медицинской помощи возможностям существующих медицинских учреждений и т. п. Каждая из перечисленных подзадач имеет самостоятельное значение, однако чтобы осуществить ежегодное планирование медицинской помощи в регионе, необходимо выполнить целенаправленный связный расчет с применением всего комплекса моделей. Для этого помимо средств сохранения результатов расчетов каждой модели необходимо обеспечить управление расчетами в зависимости от состояния задачи.

Построение управляемых комплексов OLAP-моделей для сложных аналитических задач предлагается осуше-

* Работа выполнена при финансовой поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (ГК № 02.740.11.0621) и гранта ККФПН и НТД (доп. соглашение № 01/10 от 17 мая 2010 г.).

ствлять на основе метаописания, которое включает в себя описание каждой модели, множество отношений между моделями, условия применения и коррекции моделей, условия альтернативного выбора модели. Метаописание комплекса OLAP-моделей осуществляется с применением специальной гибридной модели знаний декларативно-процедурного типа:

$$U = \langle W, R, U \rangle,$$

где $W = \{M_1, M_2, \dots, M_n\}$ – описание множества моделей; R – отношения между моделями; U – условия применения и коррекции моделей.

Каждая OLAP-модель $M \in W$ представляется в виде $M = \langle X, G, Q(G), S(G) \rangle$.

Используются следующие обозначения:

X – дискретное множество входных данных модели (витрина данных);

$G = \langle Z, F \rangle$ – гиперкуб – модель логического многомерного представления данных, характеризующаяся двумя наборами параметров: показателями $Z = \langle z_1, z_2, \dots, z_m \rangle$ и измерениями $F = \langle f_1, f_2, \dots, f_n \rangle$. Каждый показатель представляет собой множество значений, количественно характеризующих анализируемый процесс. Каждое измерение представляет собой упорядоченное множество значений определенного типа. Множество измерений образует оси гиперкуба. Измерения могут быть организованы в виде упорядоченной иерархической структуры;

$Q(G)$ – функции, или процедуры, описывающие построение элементов гиперкуба G : показателей Z и измерений F , исходя из множества входных данных X , а также операции над гиперкубом: агрегирование гиперкуба по иерархии атрибутов выбранного измерения, срез гиперкуба G по одному или нескольким заданным измерениям, операции поворота гиперкуба, изменяющие порядок измерений в гиперкубе;

$S(G)$ – описание результатов вычисления OLAP-модели, обычно сохраняемых в виде таблиц агрегатов. Особенностью технологии OLAP является возможность наглядной визуализации результатов обработки в виде таблиц, кросс-таблиц, диаграмм, а также динамических картограмм [2].

Для описания моделей в работе [7] предложен специальный язык и разработаны программные средства генерации в этом языке скрипта OLAP-модели, который генерируется в результате ее интерактивного формирования в программной системе. Однако работать со скриптовым представлением оказалось довольно сложно, поскольку комплекс может содержать большое число моделей. К примеру, для задачи ежегодного планирования медицинской помощи комплекс включает более сорока OLAP-моделей.

Для повышения обзорности комплекса нами введено краткое метаописание OLAP-модели M в виде кортежа:

$$M: \langle \text{витрина данных} \rangle; \langle \text{назначение модели} \rangle; \langle \text{результат} \rangle.$$

Метаописание каждой модели M в базе знаний формируется автоматически, с помощью специально разработанного программного обеспечения на основе скрипта.

Для реализации управляемых расчетов, помимо описания отдельных моделей, необходимо описать их взаи-

мосвязи. Это осуществляется путем описания отношений между моделями и условий их применения.

Основной вид отношений между моделями – отношение информационной зависимости моделей $M_1 \rightarrow M_2$, которое определяет необходимость предварительного расчета модели M_1 для последующего расчета модели M_2 . Обычно это отношение имеет место, если результаты вычисления M_1 используются как исходные данные для M_2 . Отношение информационной зависимости применяется для построения упорядоченных цепочек расчета OLAP-моделей. Можно также использовать отношение альтернативного выбора модели для случая, когда разные модели решают одну и ту же задачу разными методами.

Условия применения моделей описываются с помощью правил следующего вида:

$$R_1: \text{ЕСЛИ } P(x_1, x_2, \dots, x_n) \\ \text{ТО ВЫПОЛНИТЬ } M / \text{ИЗМЕНИТЬ } M,$$

где R_1 – уникальное в базе знаний имя правила; x_1, x_2, \dots, x_n – показатели, заданные как результаты расчета какой-либо модели, либо заданные в витрине как исходные данные. Каждый показатель представляет собой многомерный куб, описанный в витрине данных соответствующей OLAP-модели. Предикат P задается в виде логико-лингвистического выражения. Операция ВЫПОЛНИТЬ M в правой части правила R_1 интерпретируется как выполнение расчета OLAP-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 ИЗМЕНИТЬ M в правой части правила интерпретируется как переход к интерактивному процессу коррекции модели M . Пример правил такого типа приведен ниже.

Алгоритмы управления аналитическим расчетом.

Метаописание комплекса OLAP-моделей используется в управляемых расчетах следующим образом. Генерация цепочек OLAP-моделей выполняется на основе их метаописания как процедура логического вывода на гибридной базе знаний. Процедура вывода представляет собой интерактивный процесс, позволяющий производить многовариантный расчет сложных многоэтапных задач.

Управление процессом расчета осуществляется на основе применения правил выполнения и выбора OLAP-моделей, а также с учетом действий пользователя. Пользователь может интерактивно изменить или заменить любую модель, вследствие чего изменяется ее системное представление. Также пользователь может изменить условия применения, изменив метаописание в базе знаний. Процедура логического вывода учитывает оба варианта. При этом автоматически исследуются возможные изменения расчета и строятся новые цепочки моделей. Кроме того, производственные правила позволяют не только указывать условия применения моделей, но и в явном виде задавать в базе знаний такие важные элементы, как критерии согласованности показателей.

Рассмотрим алгоритм логического вывода, позволяющий выполнять управляемый многовариантный расчет на примере расчета Территориальной программы

государственных гарантий оказания населению бесплатной медицинской помощи (далее – ТППГ) с применением метаописания комплекса OLAP-моделей [8].

Чтобы выполнить процесс генерации цепочек OLAP-моделей, позволяющих рассчитать показатели объемов медицинской помощи и нормативы требуемого финансирования, а также выполнить согласование этих показателей, применяется построенное метаописание комплекса моделей. Генерация цепочек OLAP-моделей для расчета ТППГ выполняется как интерактивный процесс логического вывода, позволяющий производить много-

вариантный расчет ТППГ либо в целом, либо отдельных ее фрагментов. Управление процессом расчета выполняется на основе применения правил выполнения и выбора OLAP-моделей, а также с учетом действий пользователя.

Рассмотрим алгоритм расчета с применением обратной цепочки вывода на базе знаний, представляющей метаописание комплекса моделей (рис. 1). Алгоритм условно можно разделить на три этапа: формирование цепочек OLAP-моделей, генерация скрипта цепочек и сам процесс расчета.

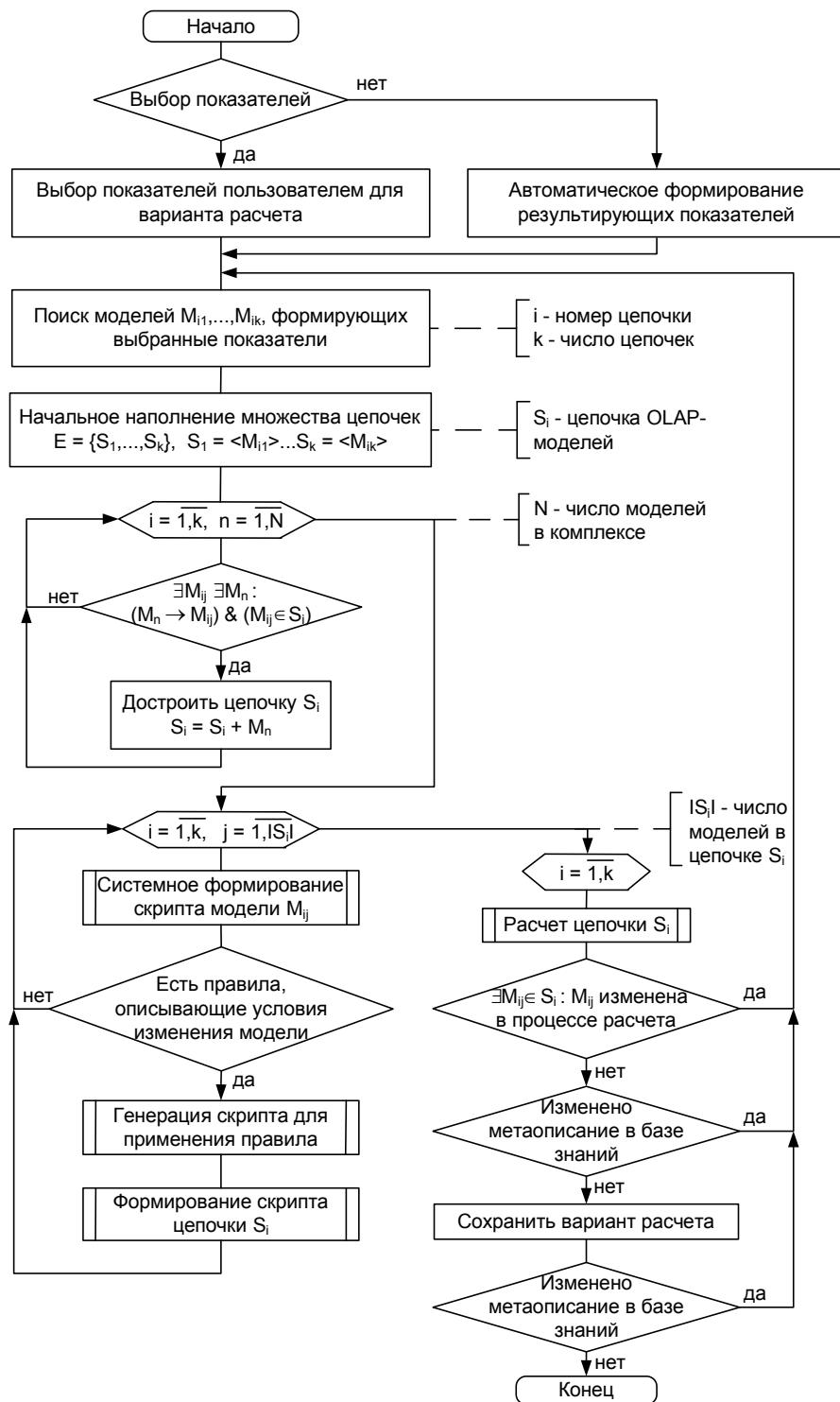


Рис. 1. Алгоритм расчета выбранных показателей

Построение цепочек предполагает определение зависимых моделей. Задаются показатели, которые необходимо рассчитать. Автоматически определяются необходимые для этого модели и порядок расчета. По описанию в базе знаний находятся модели такие, от которых информационно зависит какая-либо модель из числа помещенных в цепочки. Соответствующие цепочки дополняются.

Результирующие цепочки должны удовлетворять следующему требованию: любая зависимая модель должна располагаться в цепочке не раньше любой модели, от которой она информационно зависит; для каждой пары моделей в цепочке, если имеет место отношение $M_1 \rightarrow M_2$, то M_1 должна предшествовать M_2 в цепочке. Это требование обеспечивает корректность расчетов.

Затем формируется скрипт построенных цепочек моделей (рис. 2). Для этой цели используется системное описание моделей и правила выполнения и выбора моделей, представленные в базе знаний. В зависимости от вида правил формируются и вставляются в скрипт цепочек управляющие элементы.

Полученный вариант расчета ТПГТ можно сохранить, затем изменить методику расчета, изменив модели или условия выполнения моделей, и вновь выполнить расчет.

Таким образом, можно получить несколько вариантов расчета ТПГТ и выбрать из них наиболее рациональный.

Многовариантный расчет можно выполнить также для того, чтобы применить разные модели балансировки показателей ТПГТ. Уточнение и балансировка показателей необходимы для согласования требуемых объемов медицинской помощи и выделенного финансирования. Можно изменить порядок расчета моделей балансировки, изменив в базе знаний приоритеты правил применения этих моделей.

В случае если в результате интерпретации правила R_i , описывающего в базе знаний условия выполнения, или изменения OLAP-модели, или отдельного показателя, оказалось, что левая часть правила – условие – оказалось выполнено, то в скрипт расчетной цепочки, куда помещена модель в результате выполнения предыдущего алгоритма, добавляется скрипт, позволяющий инициировать соответствующее действие. При этом, если в результате интерактивного взаимодействия с пользователем модель изменяется, то по алгоритму (см. рис. 1), это приводит к возврату к построению последовательности расчета заново.

Рассмотрим алгоритм проверки условия правила, содержащего многомерные показатели (рис. 3). Алгоритм записывается следующим образом:

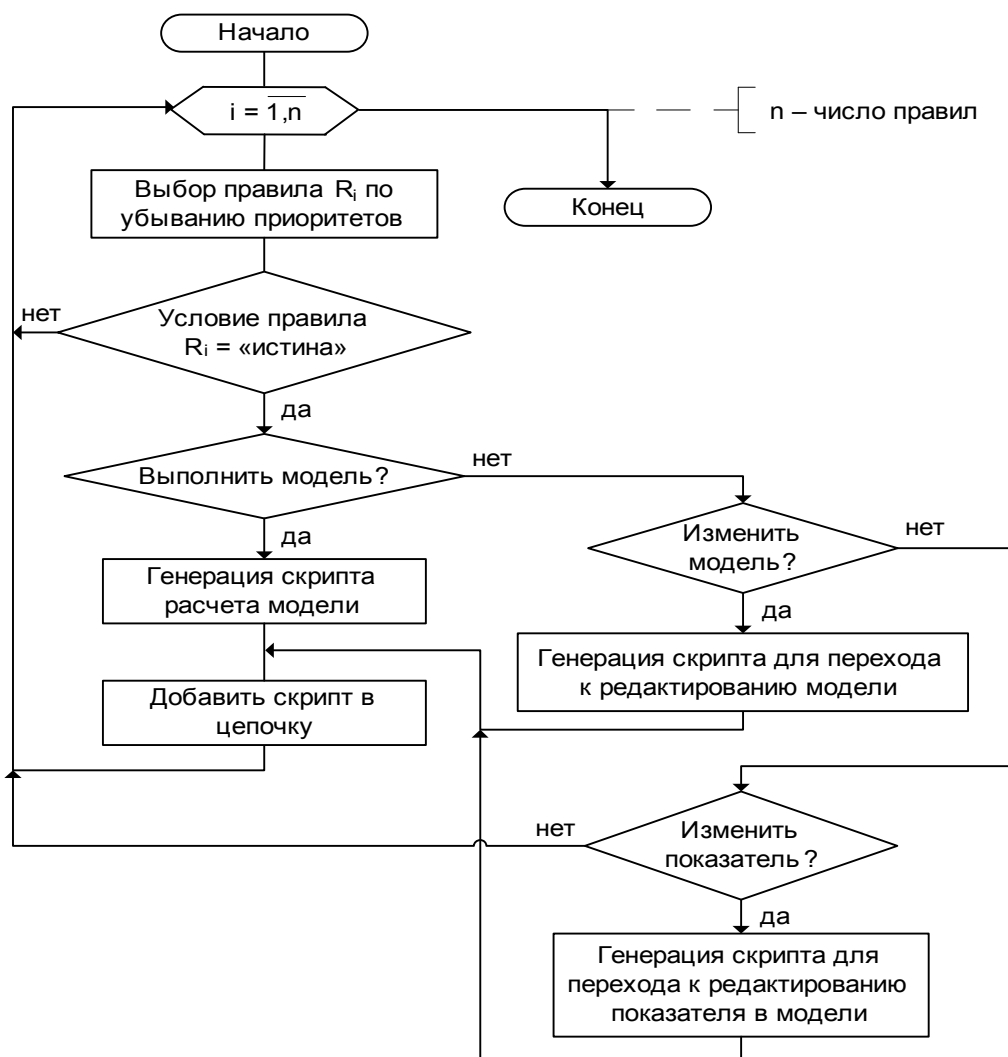


Рис. 2. Алгоритм формирования управляющих элементов в цепочке моделей

R_1 : ЕСЛИ $(x_1 \sim a_1)$ И $(x_2 \sim a_2)$
И ... И $(x_n \sim a_n)$ ТО ВЫПОЛНИТЬ M .

Суть алгоритма состоит в том, чтобы сравниваемые многомерные показатели привести к одинаковой размерности. Для этого по несовпадающим измерениям производятся срезы. Затем значения показателей, имеющие одинаковое положение в построенном кубе, попарно сравниваются, если отношение выполняется для всех пар, то условие правила истинно, в противном случае – условие ложно. Если показатели не имеют общих измерений, условие ложно.

Таким образом, рассмотренные алгоритмы позволяют выполнять расчет показателей путем реализации обратной стратегии логического вывода на знаниях, представляющих метаописание комплекса моделей.

Пример сводного расчета объемов медицинской помощи по круглосуточным стационарам представлен на рис. 4.

Таким образом, предложенный способ метаописания связанных комплексов OLAP-моделей, основанный на декларативно-процедурной модели знаний, позволяет осуществлять управление многовариантными аналитическими расчетами путем построения упорядоченных цепочек OLAP-моделей. Особенностью является возможность применения многомерных условий в правилах,

описывающих порядок вычисления моделей. Представленные алгоритмы управления аналитическим расчетом в комплексе применены для расчета показателей территориальной программы бесплатной медицинской помощи. Выполнена программная реализация полученных результатов. Программное обеспечение внедрено в агентстве здравоохранения Красноярского края.

Библиографические ссылки

1. Ноженкова Л. Ф., Шайдулов В. В. OLAP-технологии оперативной информационно-аналитической поддержки организационного управления // Информ. технологии и вычислит. системы. 2010. № 2. С. 15–27.
2. Ноженкова Л. Ф., Евсюков А. А., Ноженков А. И. Методы управления и геоинформационного моделирования в технологии OLAP // J. of Siberian Federal University. Engineering & Technologies. 2009. № 2(1). С. 49–58.
3. Арсеньев С. Б., Бритков В. Б., Маленкова Н. А. Использование технологии анализа данных в интеллектуальных информационных системах // Управление информ. потоками : сб. тр. Ин-та системного анализа РАН. М. : Эдиториал УРСС, 2002. С. 47–68.

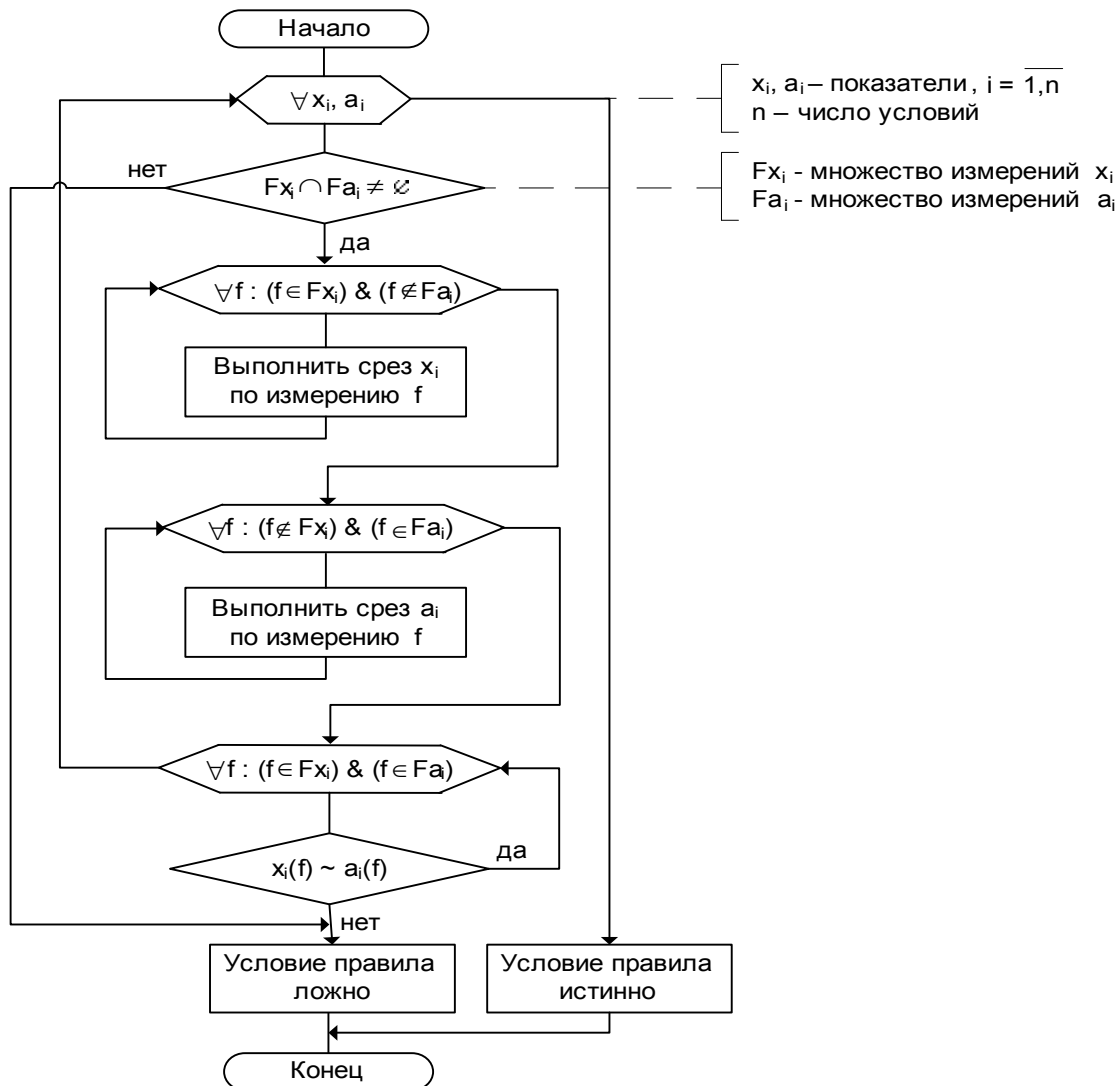


Рис. 3. Алгоритм проверки многомерного условия правила

4. Методы и модели анализа данных: OLAP и Data Mining. / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. СПб. : БХВ-Петербург, 2004.

5. Манди Д. К вопросу об OLAP [Электронный ресурс] / под ред. Р. Кимбалла // Intelligent Enterprise. 2003. № 18. URL: <http://iemag.ru/?ID=473963> (дата обращения: 23.11.2010).

6. Висков А. В., Фомин М. Б. Средства описания многомерных моделей данных // Вестник РУДН. Сер. Прикл. и компьютер. математика. 2003. Т. 2. № 1. С. 128–139.

7. Ишенин П. П. Инструментальные средства построения комплексов моделей и аналитических приложений в OLAP-технологии : автореф. дис. ... канд. техн. наук. Красноярск, 2006.

8. Ноженков А. И. Интеллектуальные средства OLAP-моделирования в задачах планирования медицинской помощи // Открытое образование (Приложение). Красноярск : Экспресс-Офсет, 2006. С. 156–160.

Террито...	Профна...	Функция койки всего оис и бюджет (факт)	Число коек всего оис и бюджет (факт)	Число примененных койкал всего оис и бюджет (предложение)	Число койко-дней всего оис и бюджет (предложение)
Абенский		282.94	211.00	5 136.00	55 099.00
Амныский		322.27	1 080.00	25 430.00	309 315.00
Амныский		454.59	102.00	1 500.00	21 416.00
Балактинский		294.24	200.00	4 130.00	36 801.00
Березовский		322.59	143.00	3 865.00	43 983.00
Бирюковский		293.96	116.00	2 371.00	28 267.00
Боготол	Гинекологическ...	318.29	34.00	708.00	7 292.00
	Для беременных и роженял	170.46	13.00	400.00	2 600.00
	Для производства абортюв	241.00	3.00	296.00	562.00
	Инфекционные	215.52	27.00	570.00	5 450.00
	Неврологические	323.04	27.00	390.00	6 726.00
	Павологин беременностл	291.73	11.00	200.00	2 400.00
	Павологин новорожденныл	266.00	1.00	0.00	0.00
	Педиатрические	337.76	29.00	635.00	8 001.00
	Терапевтические (общел)	330.34	87.00	1 302.00	19 921.00
	Хирургические г...				

Рис. 4. Сводный расчет объемов медицинской помощи по круглосуточным стационарам

A. I. Nozhenkov

OLAP-MODELS CONTROLLED COMPLEX CONSTRUCTION

OLAP-models complex meta-description means for solving complicated analytical problems are presented. Control calculation algorithms are described. The results were applied for territorial planning of medical assistance.

Keywords: OLAP-model, OLAP-model complex, control, multidimensional data, hybrid knowledge model, control calculation algorithms.

© Ноженков А. И., 2010