

8. Токарев А.В., Якубайлик О.Э. Каталог ресурсов для ГИС мониторинга состояния окружающей природной среды в зоне действия предприятий нефтегазовой отрасли // Горный информ.-аналит. бюл. 2009. Т. 18. № 12. С. 215-219.

9. Матвеев А.Г., Якубайлик О.Э. Проектирование и разработка программно-технологического обеспечения для геопространственных веб-приложений // Фундаментальные исследования. 2013. № 10-15. С. 3358-3362.

10. Кадочников А.А., Якубайлик О.Э. Сервис-ориентированные веб-системы для обработки геопространственных данных // Вестник НГУ. Серия: Информационные технологии. 2015. Т. 13. № 1. С.37-45.

11. Матвеев А.Г., Якубайлик О.Э. Разработка веб-приложения для обработки и представления пространственных метаданных геопортала. // Вестник СибГАУ. 2012. Вып. 2(42). С. 48-54.

Development of GIS-applications based on geoportals technologies

Alexandra Konstantinovna Matuzko, Post-graduate Student

Oleg Eduardovich Yakubailik, PhD, Associate Professor, Senior Researcher

This paper discusses software tools and technologies for development of the applied web-based software systems focused on processing and presentation of geospatial data. The proposed solution is a specialized content management system for website, which takes into account the specificity of GIS web applications.

Keywords: web GIS, Internet GIS, web mapping, geoportals, spatial data, content management system, CMS, web mapping services, spatial metadata catalog, WMS service.

УДК 004.6

ФОРМАЛЬНОЕ ОПИСАНИЕ РАСЧЕТА МНОГОМЕРНЫХ АНАЛИТИЧЕСКИХ ПОКАЗАТЕЛЕЙ В ВИДЕ ПОСЛЕДОВАТЕЛЬНОСТИ ОПЕРАЦИЙ НАД OLAP-КУБОМ

Анна Михайловна Метус, аспирант

Тел: +7 391 290 7453, e-mail: metus@icm.krasn.ru

Институт вычислительного моделирования СО РАН

www.icm.krasn.ru

В работе выполнено формальное описание операций над OLAP-кубом. Представлен пример расчета многомерного аналитического показателя в виде последовательности операций над OLAP-кубом на основе предложенного формального описания для задачи гидрологического мониторинга.

Ключевые слова: оперативная аналитическая обработка данных, OLAP-куб, аналитическая операция.

Исследование выполнено при финансовой поддержке грантов РФФИ № 16-37-00014 и №16-07-01001

Введение

Технология оперативной аналитической обработки данных (On-line Analytical Processing, OLAP) широко используется для информационно-аналитической поддержки управления во многих сферах человеческой деятельности [1]. Технология OLAP пред-

ставляет современный метод анализа данных, который обеспечивает пользователя естественной, интуитивно понятной моделью данных, организуя их в виде многомерных кубов – OLAP-кубов [2].

Развитие аналитических возможностей и повышение эффективности использования инструментов оперативного анализа данных, как правило, достигается интеграцией OLAP-технологии с другими современными технологиями и методами обработки данных. Так, например, в работе [3] предложено описание многомерной модели данных с помощью нечеткой логики, позволяющее анализировать показатели, содержащие неопределенность и неоднозначность. В работе [4] на основе интеграции технологий OLAP и Semantic Web предложен подход, обеспечивающий поиск и обработку данных интернет ресурсов. Авторами [5] предложен метод обработки многомерных кубов с представлением значений аналитические показатели в виде взвешенных списков. Авторами [6] предложена реализация OLAP в среде объектно-ориентированной базы данных, позволяющая обрабатывать сложноорганизованные несбалансированные и мультииерархичные измерения. В работе [7] представлены результаты исследования аддитивных свойств показателей в реляционной модели данных, предложены правила для проверки корректности суммирования, позволяющие обнаруживать показатели, которые могут быть проанализированы совместно. В работе [8] предложен матричный подход к представлению OLAP-куба, создающий основу для развития средств параллельной аналитической обработки данных.



В настоящее время актуально описание расчета многомерного аналитического показателя в виде последовательности операций над OLAP-кубом. Описание расчета показателей в многомерном виде позволит применять многомерные нормативы и обоснованно рассчитывать многомерные оценки показателей. Описание расчета аналитических показателей в виде последовательности операций над OLAP-кубом обеспечит возможность явно задавать алгоритм вычисления оценок показателей в задаче комплексного оценивания состояния объекта контроля [9]. Формальное описание позволит описать расчет показателей на более высоком уровне абстракции, избегая, таким образом, привязки к определенному инструменту реализации OLAP.

В работе выполнено формальное описание расчета многомерных аналитических показателей в виде последовательности операций над OLAP-кубом. Представлено описание базовых аналитических операций над многомерным кубом: агрегирование, детализирование, срез и выборки. Представлен пример расчета показателя гидрологического мониторинга на основе предложенного формального описания.

Основные понятия технологии OLAP

Основные концепции и принципы организации информационной структуры технологии OLAP определены Э. Коддом [2]. Основой технологии является размещение данных предметной области в особом многомерном формате – OLAP-кубе, или гиперкубе. Основными элементами логической структуры гиперкуба являются:

Измерение – множество объектов одного типа, обеспечивающее информационный контекст показателя. Измерение принято визуализировать в качестве оси куба, на которой каждый элемент измерения имеет уникальную позицию.

Иерархия – организационная структура, определяющая вид измерения на основе связи типа «родитель-потомок» между элементами измерения. Иерархия предназначена для вычисления агрегатов показателей и навигации по кубу данных.

Показатель – аналитическая величина, количественно характеризующая некоторый параметр анализируемого объекта или процесса. Все показатели в кубе имеют одинаковую размерность.

Такая модель данных информации предметной области обеспечивает высокую скорость обработки данных, гибкое манипулирование информацией и эффективный анализ по различным аспектам.

Формализация аналитических операций над OLAP-кубом

В соответствии с приведенными выше определениями OLAP-куб можно описать с использованием теоретико-множественного подхода следующим образом:

$D = \{d_1, \dots, d_n\}$ – множество независимых измерений.

В общем случае, каждое измерение может быть представлено в виде иерархии элементов: $d_i = \{d_{ik}\}$, где $i = 1 \dots n$ – порядковый номер измерения в кубе; $l = 1 \dots L$ – порядковый номер уровня иерархии в измерении; $k = 1 \dots K$ – порядковый номер элемента измерения на l -м уровне иерархии. Уровни иерархии $d_{il} \subseteq d_i$ представляют собой непересекающиеся подмножества элементов измерения $d_{ilk} \in d_{il}$. Определим булеву функцию $parent(d_i^1, d_i^2)$ для пары элементов измерения $d_i^1, d_i^2 \in d_i$, которая принимает значение 1, если d_i^1 является родительским элементом для d_i^2 , иначе значение 0.

Заданное множество измерений D определяет многомерное пространство гиперкуба G . $M = \{m_1, \dots, m_q\}$ – множество показателей в гиперкубе G , где $m_j = \{m_{jt}\}$ множество показателей в точке (ячейке куба), определяемой декартовым произведением измерений куба $d_{1l_1} \times d_{2l_2} \times \dots \times d_{nl_n}$. Элементы измерений играют роль координат, однозначно определяющих множество показателей куба: $m_{jt} = f_t(d_1', \dots, d_n')$, где $j = 1 \dots q, t = 1 \dots T, d_i' \in d_{il_i}$.

Расчет аналитического показателя подразумевает выполнение последовательности аналитических операций, преобразующих исходный куб G к новому кубу G^* , с целью извлечения из всего объема исходной информации конкретных данных. Определим последовательность чисел для отслеживания состояние гиперкуба $l_1, \dots, l_i, \dots, l_n$, в которой l_i – это порядковый номер уровня иерархии, по которому можно судить о состоянии измерения в кубе [10]. Таким образом, состояние гиперкуба можно определить как $G(l_1, \dots, l_i, \dots, l_n)$. К базовым аналитическим операциям, осуществляющим преобразование куба, относятся: агрегация, детализация, срез, выборка, проектирование и перекрестная детализация [11].

Агрегация – операция, которая позволяет перейти от детальных данных к обобщенным путем группировки показателей куба G по иерархии i -го измерения. Оператор агрегирования *Rollup* преобразует исходный куб G к кубу меньшей размерности G^* , в котором уровень иерархии i -го измерения на единицу меньше по сравнению с исходным кубом. Агрегировать показатели можно путем суммирования, вычисления усредненного показателя, нахождения максимального, минимального показателя, числа показателей и т.д. Нулевой уровень иерархии d_{i0} представляет собой абстрактный корневой узел иерархии измерения, как правило, не имеющий соответствия реальному объекту предметной области. Агрегирование до нулевого уровня иерархии представляет собой агрегирование по всем элементам измерения.

Определим оператор агрегирования *Rollup*, в качестве параметров задается куб данных, функция агрегирования A и измерение, по которому данный куб необходимо агрегировать:

$$Rollup(A, i, G(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_i^-, \dots, l_n), \text{ где } l_i^- = \max\{0, l_i - 1\}$$

Множество измерений в результирующем кубе:

$$D^* = \{d_1, \dots, d_i^*, \dots, d_n\}, \quad d_i^* \subseteq d_i$$

$$d_i^* = \begin{cases} \{d_{ilk} \mid l=1..l_i^-, k=1..K\}, & \text{если } l_i^- > 0 \\ d_{i0}, & \text{если } l_i^- = 0 \end{cases}$$

Множество показателей в результирующем кубе:

$$m_{jt}^* = f_t(d_1^*, \dots, d_i^*, \dots, d_n^*) = A(f_t(d_1^*, \dots, d_i^1, \dots, d_n^*), \dots, f_t(d_1^*, \dots, d_i^P, \dots, d_n^*)), \quad \text{где } d_i^* \in d_i^*,$$

$$d_i^p \in d_i, \quad \text{parent}(d_i^*, d_i^p) = 1, \quad p = 1..P, \quad P - \text{количество дочерних элементов для } d_i^*.$$

Детализация – операция, которая предназначена для добавления данных в куб. Оператор детализации *Drilldown* преобразует исходный куб G к кубу большей размерности G^* , в котором уровень иерархии i -го измерения на единицу больше по сравнению с исходным кубом. Ранее агрегированные по i -му измерению ячейки куба разворачиваются до меньшего уровня иерархии.

Определим оператор детализации *Drilldown*, в качестве параметров задается куб и измерение, по которому данный куб необходимо детализировать:

$$\text{Drilldown}(i, G(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_i^+, \dots, l_n), \quad \text{где } l_i^+ = \min\{l_i + 1, L_i\}$$

Множество измерений в результирующем кубе:

$$D^* = \{d_1, \dots, d_i^*, \dots, d_n\}, \quad d_i \subseteq d_i^*$$

$$d_i^* = \begin{cases} \{d_{ilk} \mid l=1..l_i, k=1..K\} \cup \{d_{ilk}^* \mid l=l_i^+, k=1..K\}, & \text{если } l_i^+ < L_i \\ d_i, & \text{если } l_i^+ = L_i \end{cases}$$

Множество показателей в результирующем кубе:

$$m_{jt}^* = f_t(d_1^*, \dots, d_i^*, \dots, d_n^*), \quad \text{где } d_i^* \in d_i^*, \quad d_i' \in d_i, \quad \text{parent}(d_i^*, d_i') = 1$$

Срез – операция, которая решает задачу фильтрации. Срез по i -му измерению выбирает показатели куба, соответствующие одному члену i -го измерения, $d_{il} = d_{ilk}$. Результирующий куб является подкубом для выбранного члена i -го измерения. Выбранное измерение в результирующем кубе не входит, таким образом, число измерений уменьшается: $|D^*| = |D| - 1$.

Определим оператор среза *Slice*, а качестве параметров задается куб и элемент измерения, по которому выполняется срез куба:

$$\text{Slice}(d_{ilk}, G(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_h), \quad \text{где } h = |D^*|$$

Множество измерений в результирующем кубе:

$$D^* = \{d_1, \dots, d_h\}, \quad d_i \notin D^*$$

Множество показателей в результирующем кубе:

$$m_{jt}^* = f_t(d_1^*, \dots, d_{ilk}, \dots, d_n^*)$$

Выборка – операция решает задачу фильтрации, но не исключает измерение из результирующего куба, а позволяет ограничить число элементов измерения $d_{il}^* \subseteq d_{il}$, $|d_{il}^*| = P$. Количество измерений в кубе не изменяется в результате операции.

Определим оператор среза *Dice*, в качестве параметров задается куб и сформированное подмножество элементов измерения, по которому делается выборка ячеек куба:

$$\text{Dice}(d_{il}^*, G(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_i, \dots, l_n)$$

Множество измерений в результирующем кубе:

$$D^* = \{d_1, \dots, d_i^*, \dots, d_n\}$$

$$d_{ik}^* = \{d_{ilk} \mid k = 1 \dots P\}$$

Множество показателей в результирующем кубе:

$$m_{jt}^* = f_t(d_1', \dots, d_i^*, \dots, d_n'), t = 1 \dots T, d_i' \in d_i$$

Проекция – операция, которая предназначена для выбора определенных данных из всего множества доступных показателей в каждой ячейке куба $m_j = \{m_{jt}\}$, $j = 1 \dots q$, $t = 1 \dots T$. $m_j^* \subseteq m_j$ – подмножество показателей для извлечения, $|m_j^*| = h$.

Определим оператор проектирования *Projection*, в качестве параметров задается куб и подмножество показателей, которое необходимо извлечь из каждой ячейки куба:

$$Projection(m_j^*, G(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_i, \dots, l_n)$$

Множество измерений в результирующем кубе остается неизменным.

Множество показателей в результирующем кубе:

$$m_{jt}^* = f_t(d_1', \dots, d_i', \dots, d_n'), t = 1 \dots h, d_i' \in d_i$$

Перекрестная детализация – предназначена для добавления в куб новых показателей. В отличие от детализации, перекрестная детализация позволяет добавлять данные для анализа из другого куба. Для того чтобы куб G^1 можно было объединить с кубом G^2 , они должны иметь одинаковые измерения и находиться в одном состоянии. $M^1 = \{m_1^1, \dots, m_q^1\}$, $m_j^1 = \{m_{jt}^1\}$, где $j = 1 \dots q$, $t = 1 \dots T$ – множество показателей куба G^1 . $M^2 = \{m_1^2, \dots, m_q^2\}$, $m_j^2 = \{m_{jz}^2\}$, где $j = 1 \dots q$, $z = 1 \dots Z$ – множество показателей куба G^2 . $D^1 = \{d_1^1, \dots, d_n^1\}$ – множество измерений куба G^1 . $D^2 = \{d_1^2, \dots, d_n^2\}$ – множество измерений куба G^2 . m_j^* – множество показателей в каждой ячейке результирующего куба, $|m_j^*| = |m_j^1| + |m_j^2|$.

Определим оператор перекрестной детализации *Drillacross*, в качестве параметров задается два куба, подлежащих объединению:

$$Drillacross(G^1(l_1, \dots, l_i, \dots, l_n), G^2(l_1, \dots, l_i, \dots, l_n)) = G^*(l_1, \dots, l_i, \dots, l_n)$$

Множество измерений в результирующем кубе:

$$D^* = \{d_1^*, \dots, d_i^*, \dots, d_n^*\}, d_i^* = d_i^1 \cap d_i^2$$

Множество показателей в результирующем кубе:

$$m_j^* = m_j^1 \cup m_j^2, j = 1 \dots q$$

Для расчета аналитического показателя необходимо провести ряд операций над OLAP-кубом. Поскольку технология OLAP не накладывает ограничений на порядок следования операций, расчет показателя можно провести несколькими путями, количество которых соответствует количеству вариантов перестановки n операций и определяется по формуле $P_n = n!$. Скорость выполнения каждой операции и объем используемых вычислительных ресурсов пропорционально зависят от количества элементов куба, которые необходимо обработать. Поэтому описание показателя в виде последовательности операций над OLAP-кубом позволяет рассчитать аналитический показатель с минимальными вычислительными затратами.

Пример расчета аналитического показателя

Рассмотрим пример расчета аналитического показателя «Уровень воды» на основе данных гидрологического мониторинга. Фрагмент исходных данных по результатам наблюдений уровня воды на водных объектах (куб данных в виде сводной таблицы) представлен на рис. 1.

Объект гидромониторинга	Время измерения	Уровень воды
▣ Ангара		
▣ Тасеева		
▣ Бирюса		
▣ Верх. Бирюса		
	14.10.2015 17:00	401
	15.10.2015 14:00	400
▣ Ниж. Тунгуска		
▣ Кислокан		
	14.10.2015 10:00	387
	14.10.2015 17:00	392
▣ Тура		
	14.10.2015 8:00	563
	14.10.2015 12:00	571
	15.10.2015 8:00	569
	15.10.2015 14:00	570
▣ Усть-Камо		
	14.10.2015 10:00	330
	15.10.2015 14:00	329

Рис. 1. Мониторинг уровня воды

Требуется сформировать показатель «Уровень воды», соответствующий максимальному суточному значению уровня воды на водных объектах речного подбассейна реки Тасеева за определенную дату «14.10.2015», учитывая данные по всем гидропостам.

Куб данных содержит два измерения: «Объект гидромониторинга» и «Дата измерения», каждый из которых имеет иерархическую структуру: d_1 – дата измерения с иерархией «час – день – месяц – год», $L_1 = 4$. d_2 – объект гидрологического мониторинга с иерархией «гидропост – водный объект – речной подбассейн – речной бассейн», $L_2 = 4$. $G(4,4)$ – множество данных, характеризующих состояние водных объектов.

Расчет показателя задается следующей последовательностью аналитических операций:

1. Агрегация: усреднение суточных данных, $Rollup(avg, i, G(4,4)) = G^*(3,4)$
2. Агрегация: выбор наивысшего уровня воды на каждом водном объекте, $Rollup(max, i, G(3,4)) = G^*(3,3)$
3. Срез: выбор подкуба за определенную дату $d_1 = '14.10.15'$, $Slice('14.10.15', G(3,3)) = G^*(3)$

Показатель «Уровень воды» на водных объектах речного подбассейна реки Тасеева выполнения последовательности аналитических операций представлен на рис. 2.

Объект гидромониторинга	Уровень воды
▣ Ангара	
▣ Тасеева	
▣ Бирюса	401
▣ Ниж. Тунгуска	567

Рис. 2. Уровень воды в речном подбассейне Тасеева за 14.10.15 г.

Заключение

В работе выполнено формальное описание расчета многомерных аналитических показателей в виде последовательности операций над OLAP-кубом. Предложено фор-

мальное описание базовых аналитических операций над многомерным кубом: агрегирование, детализирование, срез и выборки. Представлен пример расчета показателя гидрологического мониторинга на основе предложенного формального описания.

Литература

1. *Kimball R., Ross M.* The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition: John Wiley & Sons, Inc. 2013. 447 p.
2. *Codd, E.F., Codd S.B., Salley C.T.* Providing OLAP to User-Analysts: An IT Mandate, E.F. Codd & Associates. 1993. 24 p.
3. *Gondaled C., Tineo L., Urrutia A.* Fuzzy OLAP: A Formal Definition: Advances in Computational Intell, AISC 61. 2009. PP. 189-198.
4. *Abell A., Romero O., Pedersen T.B., et al.* Using Semantic Web Technologies for Exploratory OLAP: A Survey: IEEE transactions on knowledge and data engineering. 2015. N. 27(2). PP. 571-588.
5. *Aouiche K., Lemire D., Godin R.* Collaborative OLAP with tag clouds. Web 2.0 OLAP Formalism and Experimental Evaluation: WEBIST. 2008.
6. *Nguyen T.B., Tjoa A.M., Wagner R.R.* An object oriented multidimensional data model for OLAP-Web-Age Information Management: Springer. 2000.
7. *Niemi T., Niinimäki M., Thanisch P., Nummenmaa J.* Detecting summarizability in OLAP: Data & Knowledge Engineering. 2014. N. 89(2). PP. 1-20.
8. *Macedo H.D., Oliveira J.N.* A linear algebra approach to OLAP: Formal Aspects of Computing. 2015. N. 27. PP. 283-307.
9. *Пенькова Т.Г., Метус А.М.* Концептуальное моделирование процессов интегрального аналитического оценивания природно-техногенной безопасности территории// Материалы XIV Всероссийской научно-практической конференции «Проблемы информатизации региона» ПИР-2015. – Красноярск: ИВМ СО РАН, 2015. С 181-188.
10. *Хрусталева Е.М.* Агрегация данных в OLAP-кубах // Алеф Консалтинг & Софт. – 2003 [Электронный ресурс]. URL <http://www.olap.ru/home.asp?artId=84>
11. *Abelló A., Romero O.* On-line analytical processing. Encyclopedia of Database Systems: Springer US. 2009. PP. 1949-1954.

Formal definition of multidimensional analytical indicators calculation as a sequence of OLAP-cube operations

Anna Mihajlovna Metus, PhD student, Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Science

This paper presents a formal definition of OLAP-cube operations. There is multidimensional analytical indicators calculation as a sequence of OLAP-cube operations based on proposed formal definition using hydrological monitoring task as an example.

Keywords: on-line analytical processing, OLAP-cube, analytical operation.

УДК 621.9

МОДЕЛИРОВАНИЕ СВОЙСТВ ИНСТРУМЕНТАЛЬНЫХ МАТЕРИАЛОВ С ПОКРЫТИЕМ

*Елена Борисовна Мокрицкая, к.т.н., доцент
Тел.: 8 914 177 2621, e-mail: mokritskaya@mail.ru
Дальневосточный федеральный университет
www.dvfu.ru*

*Лариса Ивановна Прудникова, к.т.н., доцент
Тел.: 8 902 485 9543, e-mail: prudnikova430460@yandex.ru
Дальневосточный федеральный университет
www.dvfu.ru*