

УДК 504.05: 004.047

И.С. СКАРГА-БАНДУРОВА

Технологический институт Восточно-украинского национального университета им. В. Даля, Северодонецк, Украина

МЕТОДЫ МНОГОМЕРНОГО МОДЕЛИРОВАНИЯ В ЗАДАЧАХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА И ИНВЕНТАРИЗАЦИИ ОТХОДОВ

Рассмотрена задача реализации формальных методов многомерного моделирования для анализа данных инвентаризации отходов в связи с важностью экологического мониторинга и управления потоком промышленных отходов. Предложен подход к решению поставленной задачи путем использования многомерных операторов. Рассмотрен пример реализации предлагаемого метода многомерного моделирования при создании систем экологического мониторинга и инвентаризации отходов. Модель основана на уникальной концепции измерений и адаптирована для многомерного анализа в системах экологического мониторинга и инвентаризации отходов.

Ключевые слова: модель данных, многомерное моделирование, экологический мониторинг, инвентаризация отходов, поддержка принятия решений.

Введение

Экологический мониторинг и управление потоком промышленных отходов являются одной из наиболее злободневных проблем современности. Мониторинг и управление опасными промышленными отходами неразрывно связан с проблемой их учета. На сегодняшний день в Украине принят ряд нормативных документов, регламентирующих деятельность, связанную с обращением отходов, разработан план мероприятий по организации ведения государственного учета и паспортизации отходов [1-4], однако информационно-аналитическая поддержка в виде автоматизированных систем анализа данных инвентаризации отходов отсутствует даже на производственном уровне. Очевидно, что решение задачи создания систем анализа данных не возможно без использования многомерных баз данных и компонентов программных решений класса Business Intelligence, реализуемых на основе централизованных систем управления ресурсами данных, называемых хранилищами данных [5].

В общем случае, многомерный анализ критически важных данных позволяет выявить ключевые показатели эффективности и отслеживать различные тенденции и закономерности [6].

Постановка задачи исследования

Основная цель данной работы – обеспечить реализацию формальных методов многомерного моделирования для эффективного анализа данных инвентаризации отходов.

Формализм позволяет решать задачи инструментальной поддержки алгоритмов преобразования данных и повышает эффективность использования CASE средств.

Модель данных

Многомерное моделирование – методология логического проектирования, используемая для хранилищ данных.

Технология многомерного моделирования и анализа данных предполагает, что данные хранятся в виде многомерного массива (куба), представляющего собой группу ячеек данных, расположенных по измерениям данных.

Измерение определяется в [7] как структурный атрибут куба, который представляет собой список элементов, все из которых имеют похожий тип с точки зрения восприятия их пользователем. Например, все месяцы, кварталы, годы составляют измерение времени; все города, регионы, страны составляют измерение географии и т.д. Измерение действует как индекс для определения значения в многомерном массиве.

В общем случае многомерная база данных представляет собой пространство в виде набора измерений D (dimension), уровней измерений и базового куба C : $\langle D, C \rangle$.

Пусть Ω – пространство измерений. Для каждого измерения D_i существует набор уровней. Измерение представляет собой решетку уровней.

Каждый путь в решетке иерархий измерения, начинающийся от наивысшего уровня и заканчи-

вающийся самой нижней границей уровня называется путь измерения. Путь измерения является линейным, полностью упорядоченным списком уровней. Полное упорядочение позволяет использовать операторы сравнения для уровней измерений. Наиболее простым примером упорядочения может служить иерархия [год, месяц, день].

Для решения задачи учета и инвентаризации отходов, в качестве примера будем рассматривать четыре измерения: «Объект», «Вещество», «Время» и «Объем», представленные в виде соответствующих иерархий на рис. 1.

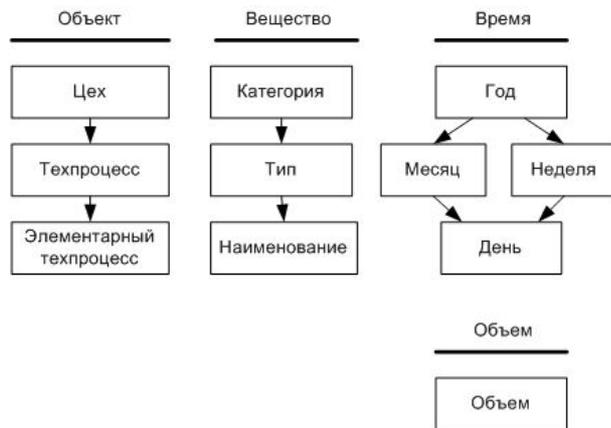


Рис. 1. Измерения и уровни измерений для учета отходов предприятия

Измерение «Объект» включает в себя сведения о цехе, технологических процессах и элементарных технологических процессах. Измерение «Время»: год, месяц (неделя), день. Структура и пример представления измерения «Вещество», представлен в табл. 1.

Таблица 1

Измерение «Вещество»

Категория		Тип	Наименование
К1	сырье и материалы	основное	M1
		основное	M2
		давальческое	M3
		давальческое	M4
К2	продукция	первичная	P1
		первичная	P2
		попутная	P3
К3	отходы	производства	W1
		производства	W2
		потребления	W3

Логическая модель данных организована в виде одного или нескольких многомерных кубов, каждый из которых содержит следующие компоненты:

- список k-измерений $D = \langle D_1, \dots, D_k \rangle$ с именем D_i и значениями из dom_i ;
- набор элементов кортежей $x = \langle x_1, \dots, x_n \rangle$ $\forall i \in [1, \dots, n] \ x_i \in dom_i(D)$.

Элементы кортежа определяются как отображение $R(C) \text{ } dom_1 \times dom_2 \times \dots \times dom_k$ и могут принимать следующие значения: 0, 1, кортеж из n-элементов [7].

Если $R(C)(d_1, \dots, d_k) = 0$, то комбинация значений измерения в базе данных не встречается.

Если $R(C)(d_1, \dots, d_k) = 1$, то существует полностью соответствующая информация.

Если $R(C)(d_1, \dots, d_k) =$ кортежу из n-элементов, то это указывает, что для данной комбинации значений измерения доступна дополнительная информация.

Рассмотрим возможности использования некоторых многомерных операторов применительно к рассматриваемой задаче.

Подробный обзор операторов представлен в работах [8 – 11].

Пусть считаются известными значения объемов потребления сырья и материалов, выпуска продукции и производства отходов (табл.2). Для объединения наименований по категориям и отдельных дат в соответствующие месяцы используется оператор слияния (merging).

Таблица 2

Объемы потребления сырья и материалов, выпуска продукции и производства отходов (куб С).

	1 янв.	31 янв.	15 февр.	28 февр.	15 март	31 март	15 апр.
M1	20				20		20
M2		40				10	10
M3		30				10	10
M4	10						10
P1	40		50		30		40
P2		40				30	10
P3	40			50	40		10
W1	20			40			40
W2		60	30				
W3				10			

Операция слияния данных:

$$\text{merge}(C, \{[D_1, f_1], \dots, [D_m, f_m]\}, f_{\text{elem}}) = C_a$$

$$dom_i(C_a) = \begin{cases} \{f_i(e) | e \in dom_i(C)\} & \text{при } 1 \leq i \leq m, \\ dom_i(C) & \text{при } m < i < l. \end{cases}$$

$$R(C_a)(d_1, \dots, d_k) = f_{\text{elem}}(\{t | t = R(C)(d'_1, \dots, d'_k)\},$$

где $f_i(d'_i) = d_i$ при $1 \leq i \leq m$, иначе $d'_i = d_i$).

Результат вычислений представлен в табл.3.

Таблица 3
Объемы потребления (производства)
по категориям измерения «Вещество» (куб С1)

	янв.	февр.	март	апр.
K1	100	0	40	50
K2	120	100	100	60
K3	80	80	100	40

Предположим далее, что в системе инвентаризации отходов суммируются значения за месяц по основным производственным показателям (потребление сырья и материалов, выпуск готовой продукции, объемы отходов и выбросов). Если необходимо оценить долю от общего объема за месяц (год), то этот результат не может быть получен прямыми вычислениями из куба С1 или С.

Для решения этой задачи используются операторы объединения и ассоциирования.

Математически процедура объединения может быть представлена следующими положениями:

$$\text{join}(C, C1, [f_{m-k+1}, \dots, f_m, f'_{m-k+1}, \dots, f'_m], f_{\text{elem}}) = C_a$$

$$\text{dom}_i(C_a) = \begin{cases} \text{dom}_i(C) & \text{при } 1 \leq i \leq m-k, \\ \text{dom}_i(C1) & \text{при } m+1 \leq i \leq n, \\ \{d^a \mid d^a \in f_i(d), d \in \text{dom}_i(C) \text{ или} \\ d^a \in f'_i(d'), d' \in \text{dom}_i(C1)\}. \end{cases}$$

$$R(C_a)(d_1, \dots, d_{m-k}, \dots, d_m, \dots, d_n) = f_{\text{elem}}(\{t1\}, \{t2\})$$

такие, что $t1 = R(C)(d_1, \dots, d_{m-k}, \dots, d_m)$,

$$t2 = R(C1)(d'_{m-k}, \dots, d'_m, d_{m+1}, \dots, d_n),$$

$$d_i^a \in f_i(d_i) \text{ или } d_i^a \in f'_i(d'_i).$$

Ассоциирование является ассиметричной процедурой в которой каждое измерение С1 соединяется с некоторыми измерениями в С. Результат выполнения процедур объединения и ассоциирования для данных табл. 2, 3 представлен в табл. 4.

Таблица 4

Относительные объемы потребления сырья и материалов, выпуска продукции и производства отходов (ассоциативный куб)

	1 янв.	31 янв.	15 февр.	28 февр.	15 март	31 март	15 апр.
M1	0,2				0,5		0,4
M2		0,4				0,25	0,2
M3		0,3				0,25	0,2
M4	0,1						0,2
P1	0,3		0,5		0,3		0,66
P2		0,3				0,3	0,17
P3	0,3			0,5	0,4		0,17
W1	0,25			0,5			1
W2		0,75	0,38				
W3				0,12			

На сегодняшний день для числовых значений, большинство операторов успешно реализовано в системах OLAP. Однако для полного покрытия задач мониторинга и инвентаризации отходов этого не достаточно, поскольку только 20% данных в информационно-аналитической системе инвентаризации отходов являются числовыми. Оставшиеся 80% - за рамками возможностей OLAP технологии, поскольку являются текстовыми документами (приказы, лимиты на размещения отходов, ДСТУ, справочники и т.д.) или полуструктурированными документами (XML документы) [11].

Так как эти типы данных не аддитивны, традиционные операторы (sum, average и др.) работать не будут, поэтому необходимо осуществлять адаптацию операторов агрегирования.

Системы мониторинга, обеспечивающие постоянный сбор данных также содержат в себе нестандартные типы данных (поток данных), для эффективного оперирования которыми необходимо решить ряд дополнительных задач.

Таким образом, стоит согласиться с [12], что исследования в области многомерного моделирования являются открытыми и нуждаются в дополнительной формализации.

Заключение

Рассмотрена задача реализации формальных методов многомерного моделирования для анализа данных инвентаризации отходов. Предложен подход к решению поставленной задачи путем использования многомерных операторов.

Рассмотрен пример реализации предлагаемого метода многомерного моделирования при создании систем экологического мониторинга и инвентаризации отходов. Модель основана на уникальной концепции измерений и адаптирована для многомерного анализа в системах экологического мониторинга и инвентаризации отходов.

Предложенный подход позволяет обеспечить инструментальную поддержку алгоритмов преобразования данных и открывает несколько новых проблем исследования.

Литература

1. Постанова Кабінету Міністрів від 1 листопада 1999 р. N 2034 «Про затвердження Порядку ведення державного обліку та паспортизації відходів».

2. Распоряжение Главы областной государственной администрации № 1323 от 2.11.2010 г. «Об утверждении мероприятий по организации ведения государственного учета и паспортизации отходов».

3. Приказ № 45 Государственного управления охраны окружающей природной среды в Луганской области от 16.11.2010 г.

4. Методическое руководство для проведения национальной инвентаризации опасных отходов в рамках Базельской конвенции [Электронный ресурс] / Серия / СБК No: 99/009 (E) Женева, Май 2000. – Режим доступа:

<http://www.basel.int/Portals/4/Basel%20Convention/docs/pub/metologicalguide-r.pdf> - 10.02.2013.

5. Kimball, R. *The data warehouse toolkit*. [Текст] / R. Kimball. – 2nd ed. – John Wiley and Sons, 2003. – 436 p.

6. *Dimensional Analysis of Data Architecture*, [Online] / Available: <http://www.tdan.com/view-articles/10569-12.12.2012> у.

7. *OLAP AND OLAP Server Definitions* [Online] / Режим доступа: <http://olapcouncil.org/research/resrchly.htm> -14.03.2013.

8. Vassiliadis, P. *Modeling Multidimensional Da-*

tabases, Cubes and Cube Operations [Текст] / P. Vassiliadis // *Proc. of the 10th SSDBM Conference*, 1998.

9. Cabbibo L. *A logical approach to multidimensional databases* [Текст] / L. Cabbibo, R. Torlone // *EDBT Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database Technology Pages*, 1998. – P. 183 – 197.

10. Haan, L. *Applied Mathematics for Database Professionals* [Текст] / L. de Haan, T. Koppelaars. – Apress, 2007. – 408 p.

11. Tseng, F.S.C. *The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence*. [Текст] / F.S.C Tseng, A.Y.H. Chou // *J. of Decision Support Systems (DSS)*. – vol.42(2), Elsevier, 2006. – P. 727 – 744.

12. *Research in Data Warehouse Modelling and Design: Dead or Alive?* [Текст] / S. Rizzi, A. Abello, J. Lechtenborger, J. Trujillo // *Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, 2006. – P. 3 – 10.

Поступила в редакцію 11.02.2013, рассмотрена на редколлегии 13.03.2013

Рецензент: д-р техн. наук, проф., зав. каф. компьютерных систем и сетей В.С. Харченко, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина.

МЕТОДИ БАГАТОВИМІРНОГО МОДЕЛЮВАННЯ В ЗАДАЧАХ ЕКОЛОГІЧНОГО МОНІТОРИНГУ ТА ІНВЕНТАРИЗАЦІЇ ВІДХОДІВ

I.S. Skarga-Bandurova

Розглянуто задачу реалізації формальних методів багатовимірного моделювання для аналізу даних інвентаризації відходів у зв'язку з важливістю екологічного моніторингу та управління потоком промислових відходів. Запропоновано підхід до вирішення поставленої задачі шляхом використання багатовимірних операторів. Розглянуто приклад реалізації запропонованого методу багатовимірного моделювання при створенні систем екологічного моніторингу та інвентаризації відходів. Модель заснована на унікальній концепції вимірювань і адаптована для багатовимірного аналізу в системах екологічного моніторингу та інвентаризації відходів. Запропонований підхід дозволяє забезпечити інструментальну підтримку алгоритмів перетворень даних і відкриває кілька нових проблем досліджень.

Ключові слова: модель даних, багатовимірне моделювання, екологічний моніторинг, інвентаризація відходів, підтримка прийняття рішень.

A MULTIDIMENSIONAL APPROACH TOWARDS ENVIRONMENTAL MONITORING AND HAZARDOUS MATERIALS INVENTORY

I.S. Skarga-Bandurova

The problem implementing formal methods for the analysis of multi-dimensional modeling toward data waste inventory is considered because of the importance of environmental monitoring and controlling the flow of industrial waste. An approach to solving the problem by using multi-dimensional operators is proposed. An example of the implementation of multidimensional modeling for environmental monitoring systems, and inventory waste is shown. The model is based on unique approach of dimension and is adapted for multidimensional analysis in environmental monitoring and hazard materials information system. A set of manipulation operations was provided.

Keywords: data model, multidimensional modeling, environmental monitoring, waste inventory, decision support.

Скарга-Бандурова Інна Сергеевна – канд. техн. наук, доцент, доцент каф. компьютерной инженерии Технологического Института ВНУ им. В.Даля, Северодонецк, Украина, e-mail: skarga_bandurova@ukr.net.