

ка // Наукові нотатки: міжвуз. збірник. – 2014. – № 44. – С. 119–123.

8. Струтинський, В. Б. Статистична динаміка шпиндельних вузлів на гідростатичних опорах [Текст] / В. Б. Струтинський, Д. Ю. Федориненко. – Ніжин: ТОВ „Видавництво „Аспект-Поліграф”, 2011. – 464 с.

9. Белоусов, А. И. Нелинейные колебания роторов на гидростатических подшипниках [Текст] / А. И. Белоусов, В. Г. Луканенко // Исследование и проектирование гидростатических опор и уплотнений быстроходных машин. – 1977. – № 4. – С. 44–51.

10. Регульований радіальний гідростатичний підшипник. Патент України на корисну модель МПК F16C 32/06 [Текст] / Федориненко Д. Ю., Сапон С. П., Хабібулїна А. М.. – №89288; заявл. 09.12.2013; опубл. 10.04.2014, Бюл. № 7.

11. Цибуля, С. Д. Протикорозійний захист елементів гідростатичної опори шпинделя металорізального верстату [Текст]: матер. 11- й междунар. пром. конф / С. Д. Цибуля, Д. Ю. Федориненко, І. А. Костенко, Н. П. Буяльська // Эффективность реализации научного, ресурсного и промышленного потенциала в современных условиях, 2011. – С. 217–221.

References

1. Directive of the European Parliament on Energy using Products (2005). Directive 2005/32/EC.

2. Postanova KMU vid 01.03.2010 r. № 243 (2015). Pro zatverdzhennja Derzhavnoi cil'ovoї ekonomichnoi programi energoefektivnosti i rozvitku sferi virobництва energonosiv z vidnovljуваниh dzhерel energii ta al'ternativnih vidiv paliva na 2010-2015 roki.

3. Gutovski, T., Dahmus, J., Thiriez, A. (2006). Electrical Energy Requirements for Manufacturing Processes. 13th CIRP International Conference of Life Cycle Engineering. Lueven (Belgium), 1–5.

4. Pidvishennja energoefektivnosti v Ukraini: zmnshennja reguljuvannja ta stimuljuvannja energospozhivannj. Available at: http://www.iер.ua/ua/publications/consultancy_work/archive_2012/?pid=3348

5. Aspects of Energy Efficiency in Machine Tools. Available at: http://www.heidenhain.com/en_US/documentation-information/technical-information/aspects-of-energy-efficiency-in-machine-Tools

6. Abele, E., Sielaff, T., Schiffler, A., Rothenbücher, S. (2011). Analyzing Energy Consumption of Machine Tool Spindle Units and Identification of Potential for Improvements of Efficiency. Globalized Solutions for Sustainability in Manufacturing, 280–285. doi: 10.1007/978-3-642-19692-8_49

7. Zubovetska, N. T. (2014). Vusokoproduktyvna ta vusokoshvydkisna obrobka na verstatakh. Scientific note: Interuniversity collection, 44, 119–123.

8. Strutytskyj, V. B., Fedorynenko, D. Y. (2011). Statystychna dynamika shpyndel'nykh vuzliv na gidrostatychnykh oporakh [Statistical dynamics of spindle units on the hydrostatic bearings]. Nizhyn, Ukraine: Aspect-Polygraph, 464.

9. Belousov, A. I., Lukanenko, V. G. (1977). Nelinejnye kolebaniya rotorov na gidrostatychnykh podshypnikakh [Nonlinear oscillations of rotors on the hydrostatic bearings]. Research and design of hydrostatic bearings and seals of high-speed machines, 4, 44–51.

10. Fedorynenko, D. Y., Sapon, S. P., Habibulina, A. M. (2014). Regul'ovaniy radial'nyy gidrostatychnyy pidshypnik. Patent Ukraini na korisnu model' MPK F16C 32/06; №89288; zajavl. 09.12.2013; opubl. 10.04.2014, Bjul. 7.

11. Zybulya, S. D., Fedorynenko, D. Y., Kostenko I. A., Buyal'ska, N. P. (2011). Protikorozijnij zahist elementiv gidrostatychnoi opori shpyndelja metalorizal'nogo verstatu 2011. Materialy 11- j mezhdunar. prom. konf. Jeftektivnost' realizacii nauchnogo, resursnogo i promyshlennogo potentsiala v sovremennyh uslovijah, 217–221.

Дата надходження рукопису 26.05.2015

Федориненко Дмитро Юрійович, доктор технічних наук, професор, кафедра технологій машинобудування та деревообробки, Чернігівський національний технологічний університет, вул. Шевченка, 95, м. Чернігів, Україна, 14000

E-mail: fdy@mail.ru

Урліна Анастасія Андріївна, аспірант, кафедра технологій машинобудування та деревообробки, Чернігівський національний технологічний університет, вул. Шевченка, 95, м. Чернігів, Україна, 14000

E-mail: urlina.anastasia@gmail.com

Аванесов Єгор Вікторович, кафедра технологій машинобудування та деревообробки, Чернігівський національний технологічний університет, м. Чернігів, вул. Шевченка, 95, м. Чернігів, Україна, 14000

E-mail: yehoravanesov@gmail.com

УДК 004.652.5

DOI: 10.15587/2313-8416.2015.43994

ІНТЕГРАЦІЯ ТЕХНОЛОГІЙ OLAP ТА DATA MINING ПРИ ПОБУДОВІ МІЖВИМІРОВИХ АСОЦІАТИВНИХ ПРАВИЛ

© М. Т. Фісун, Г. В. Горбань

Представлено особливості пошуку асоціативних правил у багатомірних даних, а саме показані теоретичні засади знаходження асоціацій між різними вимірами у OLAP-кубах та формули обчислення характеристик їх значущості (підтримка, достовірність, ліфт та леверидж). Запропоновано метод генерації міжвимірових асоціативних правил та описано його реалізацію як складової інформаційної системи оперативного та інтелектуального аналізу даних на платформі об'єктної системи керування базами даних Caché

Ключові слова: OLAP, Data Mining, багатомірні дані, куб, асоціативні правила, предметний набір, підтримка, достовірність, ліфт, леверидж

The features of searching associative rules in multidimensional data are presented in the article, specifically theoretical basis of association searching between different dimensions in OLAP cubes and formulas of calculation of their significance characteristics (support, confidence, lift, leverage) are shown. The method of interdimensional association rules generation is proposed. The implementation of this method as a component of operative and intellectual data analysis information system on database management system Caché platform is described

Keywords: OLAP, Data Mining, multidimensional data, cube, association rules, objective set, support, confidence, lift, leverage

1. Вступ

У сучасних системах підтримки прийняття рішень (СППР) все більш часто застосовуються технології OnLine Analytical Processing (OLAP) і Data Mining. Вони дозволяють в тій чи іншій мірі ефективно проводити аналіз даних. Технологія OLAP полягає в побудові хоча б на концептуальному рівні багатовимірних гіперкубів, над якими користувачі можуть проводити різні аналітичні операції. На відміну від OLAP технологія Data Mining являє собою засоби (інструментарій) дослідження в накопичених даних якихось прихованих знань, які до цього не були відомі і можуть виявитися досить корисними в процесі підтримки прийняття рішень.

Однак, технології OLAP і Data Mining рухаються у різних напрямках, оскільки в OLAP звертається увагу виключно на забезпечення доступу до багатомірних даних, а методи Data Mining у більшості випадків працюють з реляційними таблицями, у яких, як правило, розміщені часові ряди. Інтеграція цих двох технологій підвищить функціональність інформаційних систем, що орієнтовані на аналіз даних та підтримку прийняття рішень. Однак питання такої інтеграції в окремих публікаціях розглядалися для реляційних баз даних [1–3] і не розглядалися для об'єктних СКБД, що є предметом дослідження даної статті.

2. Постановка задачі

Одним з найпоширеніших методів інтелектуального аналізу даних (Data Mining) є асоціація, що представляє собою виявлення закономірностей між зв'язними об'єктами, прикладом яких може бути правило, що з події X слідує подія Y. X називається умовою або антецедентом, а Y – наслідком або консеквентом. Подібні правила називаються асоціативними.

В основному методи Data Mining, включаючи і пошук асоціативних правил, працюють на даних, що представлені в табличному вигляді. Однак подібні закономірності можливо знайти і у багатомірних даних [4].

Метою дослідження є інтеграції технологій OLAP та Data Mining у єдиній системі в частині виявлення асоціативних правил у багатомірних даних шляхом запропонованого методу їх генерації.

3. Аналіз літературних даних

Основними труднощами об'єднання OLAP і Data Mining є той факт, що традиційні алгоритми Data Mining, як було сказано вище, в основному працюють із табличними даними [5], тому багато-

мірні дані для даних алгоритмів не підходять. Однак можна зустріти достатньо багато робіт, які стосуються об'єднання технології OLAP з методами Data Mining. Виділяються три основні підходи в цій області [6].

Перший підхід полягає в розширенні мови запитів основними методами Data Mining. Цей підхід використовує система Dbminer [7]. Також були спроби розширення функцій OLAP і спільного використання розподіленого сервера OLAP з інфраструктурою Data Mining, результатом чого стало виявлення асоціативних правил, що були представлені в кубах, які були названі Association Rule Cubes [8]. Ще однією спробою було узагальнення інформації в кубі даних, розширення операторів OLAP алгоритмами пошуку асоціативних правил [9].

Другий підхід інтеграції OLAP і Data Mining полягає в адаптації багатомірних даних усередині БД або поза ними й припускає застосування класичних алгоритмів Data Mining для результуючих наборів даних. Прикладами даного підходу є інтеграція багатомірної інформації в послідовності даних і подальше їх дослідження на виявлення закономірностей [10], згладжування кубів даних і витяг з них матриць для кожного виміру на кожному кроці побудови дерева рішень [11]. Також був запропонований метод пошуку асоціативних правил у сховищах даних, який заснований на організації багатомірних даних і здатний витягувати асоціативні правила з декількох вимірів на різних рівнях абстракції [12].

Третій похід інтеграції технологій OLAP і Data Mining заснований на адаптації методів Data Mining і їх застосуванні безпосередньо на багатомірних даних. Прикладом даного підходу є ідея теоретичної системи OLAP Data Mining, яка має можливість інтегрування багатомірних даних з метою дослідження окремо за кожним виміром [13], пропозиція інтеграції в OLAP-сервер модуля багатомірної регресії [14]. Також був запропонований підхід, який полягає в генерації звітів кількісного аналізу з даними кубів [15].

4. Матеріали та методи досліджень

4.1. Міжвимірні асоціативні правила у багатомірних даних та їх різновиди

Якщо замість реляційних даних розглядати OLAP-куб, то предметним набором для асоціативних правил можна представити множину значень (атрибутів) кожного з вимірів.

Дослідження представлено на прикладі бази даних деякої торговельної компанії [16], даталогічна модель якої представлена на рис. 1.

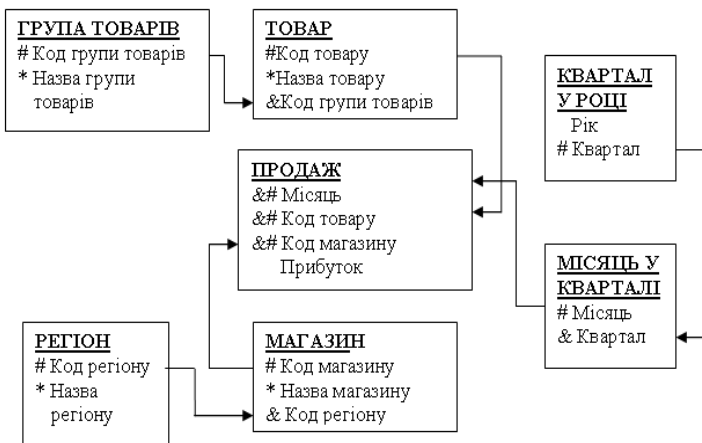


Рис. 1. Даталогічна модель бази даних торговельної мережі

Для представленої бази даних був реалізований куб у вигляді В-дерева за допомогою запропонованого комбінаторного алгоритму [17], до якого був запропонований метод генерації асоціативних правил, представлений у даній статті.

У багатомірних даних є можливим встановити залежності між даними різних вимірів кубу, які представляються міжвимірними асоціативними правилами (inter-dimensional association rules). Дані правила можна представити у наступному загальному вигляді:

$$(A_i^x \in D_i) \wedge \dots \wedge (A_j^y \in D_j) \rightarrow A_k^z \in D_k,$$

де I, J, K – відповідні індекси вимірів, що входять в асоціативне правило; $I, J, K = 1..n$; n – кількість вимірів в OLAP-кубі, D_i – i -вимір, x, y, z – відповідні індекси атрибутів вимірів, $x, y, z = 1..m_i$; m_i – кількість атрибутів i -го виміру; A_i^x – відповідний атрибут I -го виміру.

Наприклад, для наведеної в попередніх розділах БД торговельної компанії, одним з міжвимірних асоціативних правил може бути наступне правило:

$$(Місяць = "Січень 2011") \wedge (Магазин = "Промінь") \rightarrow (Товар = "Молочні продукти").$$

Дане правило висуває гіпотезу про те, що у січні 2011 року в магазині «Промінь» був отриманий великий прибуток саме від продажу молочних продуктів, а не від інших товарів. Після розрахунку наведених характеристик значущості асоціативного правила можна буде зробити висновок про корисність даного асоціативного правила.

Слід зауважити, що з операндів, які входять в дане асоціативне правило, можна скласти інші асоціативні правила, змінивши місцями знаки кон'юнкції (\wedge) та імплікації (\rightarrow) між ними. Наприклад:

$$(Місяць = "Січень 2011") \wedge (Товар = "Молочні продукти") \rightarrow (Магазин = "Промінь").$$

Дане правило має вже зовсім іншу інтерпретацію. Воно висуває гіпотезу, що в січні 2011 року від продажу молочних продуктів був отриманий великий прибуток саме в магазині «Промінь», а не у будь-якому іншому.

Ще одне асоціативне правило, отримане з множини даних атрибутів:

$$(Магазин = "Промінь") \wedge (Товар = "Молочні продукти") \rightarrow (Місяць = "Січень 2011").$$

Воно означає припущення, що великий прибуток від молочних продуктів в магазині «Промінь» був отриманий саме в січні 2011 року, а не у будь-якому іншому місяці.

У розглянутих вище асоціативних правилах можна помітити, що в антецеденті зустрічається більш ніж один операнд, а в консеквенті – один. Якщо, поставити знак імплікації між операндами в іншому місці, то також будуть отримані абсолютно інші асоціативні правила, які навпаки можуть мати у антецеденті один операнд, а в консеквенті – декілька, або в обох складових асоціативного правила по декілька операндів (у випадках, якщо кількість вимірів у кубі перевищує за три). Таким чином, асоціативні правила будуть мати наступний загальний вигляд:

$$A_i^x \in D_i \rightarrow (A_j^y \in D_j) \wedge \dots \wedge (A_k^z \in D_k),$$

а при кількості вимірів більше за три, можна отримати і наступні асоціативні правила:

$$(A_i^x \in D_i) \wedge \dots \wedge (A_j^y \in D_j) \rightarrow (A_n^v \in D_n) \wedge \dots \wedge (A_k^z \in D_k),$$

у яких в обох складових асоціативного правила по декілька операндів.

Для наведеного вище прикладу можна отримати наступні правила:

$$\begin{aligned} &(Місяць = "Січень 2011") \rightarrow (Магазин = "Промінь") \wedge (Товар = "Молочні продукти"); \\ &(Магазин = "Промінь") \rightarrow (Місяць = "Січень 2011") \wedge (Товар = "Молочні продукти"); \\ &(Товар = "Молочні продукти") \rightarrow (Місяць = "Січень 2011") \wedge (Магазин = "Промінь"); \end{aligned}$$

Дані правила вже мають дещо іншу інтерпретацію. Наприклад, перше з них визначає, що якщо конкретно розглядати січень 2011 року, можна припустити, що великий прибуток у ньому приніс продаж молочних продуктів у магазині «Промінь», а не будь-якого іншого товару в будь-якому іншому магазині.

Міжвимірне асоціативне правило не обов'язково повинно містити атрибути з всіх існуючих вимірів. Наприклад, можуть існувати і наступні асоціативні правила:

$$A_i^x \in D_i \rightarrow A_j^y \in D_j,$$

тобто перевіряється залежність між атрибутами тільки 2 вимірів багатомірного кубу. Проте дані правила можуть мати різновиди. Наприклад, правило:

$$(Товар = "Молочні продукти") \rightarrow (Місяць = "Січень 2011")$$

означає припущення, що великий прибуток від продажу молочних продуктів був отриманий в січні 2011 році, причому у всіх магазинах мережі, оскільки в даному асоціативному правилі немає чітко визначеного атрибуту виміру «Магазин».

Але можна накласти на дане правило обмеження, коли все ж дане припущення необхідно перевірити для конкретного магазину. Таким чином, отримується наступне правило:

$$\text{В контексті (Магазин = "Промінь"): (Товар = "Молочні продукти") \rightarrow (Місяць = "Січень 2011"),}$$

яке означає, що відповідна гіпотеза буде перевірена для магазину «Промінь». Іншими словами можна сказати, що вимір «Магазин» є фіксованим.

В загальному випадку такі асоціативні правила будуть виглядати наступним чином:

$$\text{In the context } (A_i^x \in D_i): A_j^y \in D_j \rightarrow A_k^z \in D_k.$$

Таким чином, з міжвимірових асоціативних правил слід виділити:

- **Повні міжвимірові асоціативні правила**, що містять атрибути всіх вимірів кубу;
- **Неповні міжвимірові асоціативні правила**, що не містять атрибути всіх вимірів кубу;
- **Контекстні міжвимірові асоціативні правила**, у яких на деякі виміри кубу накладені певні обмеження.

4. 2. Розрахунок характеристик значущості міжвимірових асоціативних правил

Підтримка асоціативного правила дорівнює підтримці частого предметного набору, за яким воно формується. Дане твердження стає очевидним завдяки тому, що певний предметний набір завжди включає у себе і антецедент, і консеквент асоціативного правила.

Достовірність асоціативного правила у багатомірних даних можна представити як відношення значення міри кубу для значень вимірів, що зазначені разом у антецеденті та консеквенті, до агрегованого значення кубу за вимірами, які зазначені тільки у антецеденті.

Наприклад, є куб з трьома вимірами, що позначаються D_1, D_2 та D_3 . Нехай у даному кубі існує комірка, значення вимірів якої дорівнюють відповідно x, y та z .

Припустимо, існує наступне асоціативне правило: $R = (D_1 = x) \wedge (D_2 = y) \rightarrow (D_3 = z)$. Тоді його достовірність дорівнюватиме:

$$\text{Conf}(R) = \frac{M_{x,y,z}}{M_{x,y,0}}.$$

Відповідно, ліфт як відношення частоти появи антецедента у транзакціях, що мають також і консеквент, до частоти появи консеквента в цілому для асоціативного правила R буде дорівнювати:

$$\text{Lift}(R) = \frac{\text{Conf}(R)}{\text{Supp}(z)} = \frac{M_{xyz}}{M_{xy0}} = \frac{M_{xyz} \cdot M_{000}}{M_{00z} \cdot M_{xy0}}.$$

У свою чергу, леверидж як різниця між частотою, з якою антецедент та консеквент з'являються спільно, і добутком частот появи антецедента та консеквента, для асоціативного правила R буде дорівнювати:

$$\text{Lev}(R) = \text{Supp}(R) - \text{Supp}(x \wedge y) \cdot \text{Supp}(z) = \frac{M_{xyz}}{M_{000}} - \frac{M_{xy0}}{M_{000}} \cdot \frac{M_{00z}}{M_{000}} = \frac{M_{000} \cdot M_{xyz} - M_{xy0} \cdot M_{00z}}{M_{000}^2}.$$

Тепер представимо формули обчислень даних характеристик значущості асоціативного правила для загального випадку.

Нехай до множини AntDim входять виміри, які в умові певного правила мають конкретні значення, тобто за ними не здійснюється агрегування:

$$\text{AntDim} = \{D_i, \dots, D_j, \dots, D_k\},$$

де D_i – вимір, що має індекс $i; i, j, k = 1..n, n$ – загальна кількість вимірів.

Подібно до описаної вище множини також можна описати множину ConsDim, до якої відповідно будуть входити виміри, що мають конкретні значення у наслідку асоціативного правила:

$$\text{ConsDim} = \{D_l, \dots, D_m, \dots, D_p\},$$

де $l, m, p = 1..n$.

Один і той самий вимір не може одночасно входити до обох множин, оскільки не може входити разом до антецеденту і консеквенту міжвимірового асоціативного правила:

$$\text{AntDim} \cap \text{ConsDim} = \emptyset.$$

Зазначені вище множини надають інформацію тільки про виміри, що входять до відповідних частин асоціативного правила. Тобто за їх допомогою можна сформувати тільки шаблон асоціативного правила. Для того, щоб сформувати власне міжвимірове асоціативне правило, потрібні множини, які містять конкретні значення відповідних вимірів. Назвемо їх відповідно Ant та Cons.

Множина Ant матимуть наступний вигляд:

$$\text{Ant} = \{i_{ant_1}, i_{ant_2}, \dots, i_{ant_k}, \dots, i_{ant_n}\},$$

де val_{ant_k} – значення k-го виміру, яке може приймати наступні значення:

- $i_{ant_k} = x, 1 \leq x \leq t_k$, якщо k-й вимір містить фіксоване значення (t_k – кількість значень у k-му вимірі);
- $i_{ant_k} = 0$, якщо за k-м виміром здійснюється агрегування.

Подібний зміст має і множина Cons:

$$\text{Cons} = \{i_{cons_1}, i_{cons_2}, \dots, i_{cons_k}, \dots, i_{cons_n}\}.$$

Подібно до розрахунку підтримки предметного набору в багатомірних даних позначимо впорядковану множину значень відповідних вимірів певної комірки кубу як $i: i = \langle i_1, i_2, \dots, i_l, \dots, i_n \rangle$.

Тоді формула розрахунку достовірності асоціативного правила у багатомірних даних для загального випадку набуде наступного вигляду:

$$\text{Conf}(i) = \frac{M_i}{M_{Ant}}.$$

У свою чергу, якщо впорядковану множину значень вимірів у комірці кубу, в якій розташований його повний агрегат, позначити як ALL: $ALL = \langle 0, 0, \dots, 0 \rangle$, то формули розрахунку

суб'єктивних характеристик значущості асоціативного правила для загального випадку дорівнюватимуть:

$$\text{Lift}(R) = \frac{\text{Conf}(R)}{\text{Supp}(Cons)} = \frac{M_i \cdot M_{ALL}}{M_{Ant} \cdot M_{Cons}}$$

– ліфт міжвимірового асоціативного правила;

$$\begin{aligned} Lev(R) &= Supp(R) - Supp(Ant) \cdot Supp(Cons) = \\ &= \frac{M_{All} \cdot M_i - M_{Ant} \cdot M_{Cons}}{M_{All}^2} \end{aligned}$$

– леверидж міжвимірового асоціативного правила.

4.3. Метод генерації міжвимірових асоціативних правил

Алгоритми генерації асоціативних правил звичайно працюють у два етапи: на першому кроці вони знаходять часті предметні набори, а на другому – знаходять з них правила.

При постановці задачі знаходження частих предметних наборів з багатомірних даних можна виділити особливість: у OLAP-кубі можна знайти такі часті набори, які відносяться до абсолютно різних сукупностей. Це пов'язане з тим, що при розгляданні багатомірних даних оброблюються абсолютно різні виміри кубу, а потім і їх об'єднання.

У загальному випадку множиною всіх частих предметних наборів у кубі OLAP є множина S:

$$S = \{S_1, S_2, \dots, S_i, \dots, S_n\},$$

де i – кількість елементів у предметному наборі, S_i – множина частих предметних наборів з кількістю елементів i , n – загальна кількість елементів у кубі.

У свою чергу, множини S_1, \dots, S_n містять різні предметні набори за кожним з вимірів або сукупністю вимірів, якщо кількість елементів у наборі є більшою за один.

Тобто:

$$S_1 = \{s_1, s_2, \dots, s_n\},$$

де s_1 – множина частих одноелементних предметних наборів за першим виміром кубу, s_2 – за другим виміром, s_n – за n -им виміром.

У свою чергу, множину двоелементних предметних наборів можна представити у наступному вигляді:

$$S_2 = \{s_{12}, s_{13}, \dots, s_{mn}\},$$

де s_{12} – множина частих предметних наборів за сукупністю першого та другого вимірів, s_{13} – за сукупністю першого та третього вимірів, $m \neq n$.

Нехай k – кількість елементів у предметному наборі. Тоді у загальному випадку:

$$S_k = \bigcup_{i=1}^{C_n^k} \{s_{i_1, i_2, \dots, i_k}\}.$$

Пропонується створювати частий предметний набір у вигляді списку, в якому перший елемент є списком, який містить порядкові номери вимірів кубу, за якими здійснюється генерація набору (т. з. список у списку).

Всі наступні елементи списку будуть містити інформацію про певний знайдений предметний набір.

$$\begin{aligned} &<< id_1, < id_2, \dots, < id_k >> >, \\ &< val_{11}, val_{21}, \dots, val_{k1} & sup_1 >, \dots, < val_{1z}, val_{2z}, \dots, val_{kz} & sup_z >>, \end{aligned}$$

де k – кількість елементів у предметному наборі, id_i – порядковий номер i -го виміру кубу у відповідному предметному наборі, val_{ij} – значення атрибуту i -го виміру кубу у відповідному j -ому предметному наборі, sup_j – значення підтримки j -го предметного набору, z – отримана кількість частих предметних наборів.

Генерація частих одноелементних наборів буде здійснена рівно стільки разів, скільки буде дорівнювати кількість комбінацій з одним елементом C_N^1 , двоелементних – C_N^2 і т. д. При цьому генерація частих наборів з кількістю елементів, що є більшою за один, використовує списки наборів, отримані на попередньому кроці.

$$MasOfSets = \langle \underbrace{Set_1, \dots, Set_i, \dots, Set_n}_n \rangle.$$

$MasOfSets$ – загальний список частих наборів, Set_i – список частих предметних наборів з кількістю елементів i , n – загальна кількість вимірів у кубі, $i=1..n$.

$$Set_i = \langle \underbrace{subset_{id_1, \dots, id_i}}_i, \dots, \underbrace{subset_{id_1, \dots, id_i}}_i \rangle,$$

де $subset_{id_1, \dots, id_i}$ – список частих i -елементних предметних наборів за вимірами з ідентифікаторами id_1, \dots, id_i .

Загальний список всіх частих предметних наборів у багатомірних даних стає основою для генерації міжвимірових асоціативних правил.

У подальшому метод генерації міжвимірових асоціативних правил має наступний порядок дій:

1. Почергово витягується кожний з елементів загального списку частих предметних наборів (Set_k , $k=1..n$), для якого обчислюється його довжина. Можна нагадати, що кожний Set_k представляє собою список всіх можливих частих предметних наборів за кількістю k вимірів;

2. Для кожного Set_k почергово отримується кожний з його підсписків $subset_{id_1, \dots, id_i}$ з якого відразу

витягується його перший елемент, що відповідає за порядкові номери вимірів, за якими був згенерований поточний частий предметний набір. Даний елемент можна назвати $idlist$;

3. У списку $idlist$ за кожною кількістю елементів i від 1 до $k-1$, здійснюється генерація всіх можливих сполучень за i елементами. При цьому загальна кількість таких сполучень дорівнюватиме C_k^i ;

4. Для кожного з отриманих сполучень формується множина антецеденту (AntDim), до якої покладаються всі елементи, що існують у поточному сполученні;

5. В отриманому списку з порядковими номерами вимірів поточного частого предметного набору ($idlist$) здійснюється пошук тих елементів, які не увійшли до множини антецеденту. Такі елементи будуть автоматично включені до множини консеквенту (ConsDim). Таким чином, на поточному етапі стане відомо, які виміри належатимуть умові майбутнього асоціативного правила, а які – наслідку;

6. Повертаючись до списку $subset_{id_1, \dots, id_i}$, розгля-

даються всі наступні його елементи (у них містяться дані про знайдені предметні набори за вимірами, номери яких вказані у списку *list*, який власне є першим елементом даного підсписку), окрім першого;

7. Для кожного елемента списку $subset_{id_1, \dots, id_i}$

з другого до останнього можна відразу знайти підтримку майбутнього асоціативного правила, оскільки вона буде дорівнювати підтримці частого предметного набору, яка буде отримана шляхом витягування зі значення поточного елемента підрядка після знаку «&»;

8. У подальшому для розрахунку всіх інших характеристик значущості асоціативного правила необхідно сформувати допоміжні рядки, що будуть містити відповідні значення вимірів, які входять до антецеденту, консеквенту та приситні у предметному наборі взагалі. Дані рядки будуть складатись із відповідних значень вимірів, розділених через кому, та будуть називатись відповідно *astr*, *cstr* та *acstr*. Вони будуть відповідати описаним вище множинам *Ant*, *Cons* та списку значень вимірів *I*;

9. Обчислюються достовірність, ліфт та леве-ридж асоціативного правила.

Якщо достовірність правила перевищує заданий мінімальний поріг достовірності, то нове правило можна зберегти у базі даних. Для цього створюється новий екземпляр класу, згенерованого та описаного на початковому етапі алгоритму, його властивостям присвоюються отримані значення характеристик асоціативного правила, а також вміст умови та наслідку. Після цього нове асоціативне правило зберігається у базі даних.

5. Результати досліджень

Представлений вище метод генерації міжвимірових асоціативних правил у багатомірних даних був реалізований у рамках інтелектуальної інформаційної системи, яка стала результатом інтеграції технології OLAP та Data Mining на прикладі виявленні асоціацій з об'єктними базами даних. Дана система була реалізована у середовищі об'єктної системи керування базами даних *Caché* за допомогою технології CSP (*Caché Server Pages*).

Результатом дослідження стала інформаційна система, що складається з чотирьох модулів:

- 1) модуль проектування БД (або модуль метаданих);
- 2) модуль ведення БД (або модуль даних);
- 3) модуль підсистеми OLAP;
- 4) модуль інтелектуального аналізу.

Система реалізована у середовищі СКБД *Caché* за допомогою його серверу *Caché Objects* та представляє собою сукупність взаємопов'язаних CSP-сторінок. Головною сторінкою у системі є сторінка вибору схеми даних для подальшої роботи з нею. Також з даної сторінки можна перейти до сторінок створення нового пакету, або видалення існуючого, а також подивитись список простих типів даних у системі. На головній сторінці список існуючих схем даних представляється у вигляді таблиці, при цьому назва схеми є посиланням, яка веде до сторінки опису відповідного пакету. Далі

можна працювати або з метаданими, або з даними предметної області. На рис. 2 представлений скріншот головної сторінки реалізованої інформаційної системи.

Запропонований метод генерації асоціативних правил реалізований у модулі інтелектуального аналізу даних.

За генерацію міжвимірових асоціативних правил у системі відповідає блок сторінок, до якого входять сторінка отримання контексту правила, яка використовується при побудові контекстних асоціативних правил, а також три сторінки власне генерації асоціацій незалежно від їх виду, контекстних або повних.

Спочатку розглянемо сторінку отримання контексту міжвимірового асоціативного правила. При задачі кінцевого користувача згенерувати саме такий вид правил на неї веде відповідне посилання на сторінці перегляду даних кубу. У свою чергу сторінка, що розглядається, пропонує користувачеві задати певні конкретні значення для вимірів кубу та водночас накладає до їх введення обмеження: не всі виміри повинні мати конкретні значення. Їх кількість не повинна перевищувати $N-2$, де N – загальна кількість вимірів.

Якщо для відповідного виміру було задане конкретне значення, то він увійде до контексту правила. Для генерації контекстних правил мінімум два виміри не повинні мати конкретні значення, оскільки така кількість є мінімальною для існування асоціативного правила, в якому один вимір буде знаходитись у його умові, а інший у наслідку. Скріншот сторінки отримання контексту представлений на рис. 3.

При натисненні кнопки ОК на розглянутій сторінці буде здійснений перехід до першої зі сторінок генерації асоціативних правил, призначення якої полягає у тому, що за її допомоги кінцевий користувач може задати системі значення мінімальної підтримки та достовірності у відсотках, спираючись на власний розсуд.

Якщо користувач залишить поля для введення зазначених величин порожніми, то мінімальні підтримка та достовірність будуть автоматично прирівняні до 0 %, що буде означати генерацію максимально можливих частих предметних наборів та міжвимірових асоціативних. Можна розпочати їх генерацію, пропустивши при цьому етап отримання контексту. В даному випадку будуть згенеровані повні міжвимірові асоціативні правила. Типове зображення описаної сторінки представлено на рис. 4.

У свою чергу друга зі сторінок генерації міжвимірових асоціацій приймає за параметри введені користувачем значення мінімальних підтримки та достовірності та викликає скрипт, що складається з функцій, формуючих часті предметні набори у багатомірних даних.

У результаті дана сторінка виводить всі отримані предметні набори, що мають значення підтримки, яка є меншою або дорівнює відповідному введеному їй мінімальному значенню.

Спочатку виводяться часті предметні набори з даних за одним виміром, потім за двома і так далі до загальної кількості вимірів, якщо такі предметні набори підходять до категорії частих за значеннями їх підтримок. Приклад зображення даної сторінки у контексті

вересня 2011 року у якості значення виміру "Місяць" представлена на рис. 5.

Остання, третя сторінка відповідно відображує останній етап, який власне і полягає у остаточній гене

рації міжвимірових асоціативних правил. Дана сторінка виводить у таблиці всі правила, які пройшли поріг мінімальної достовірності.

Скріншот даної сторінки представлено на рис. 6.

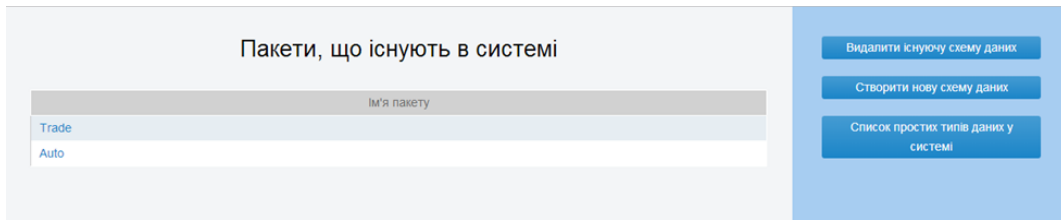


Рис. 2. Головна сторінка системи

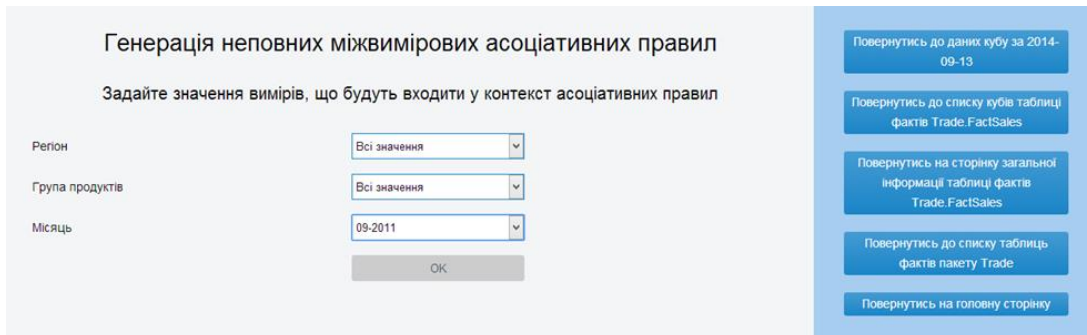


Рис. 3. Сторінка отримання контексту міжвимірового асоціативного правила

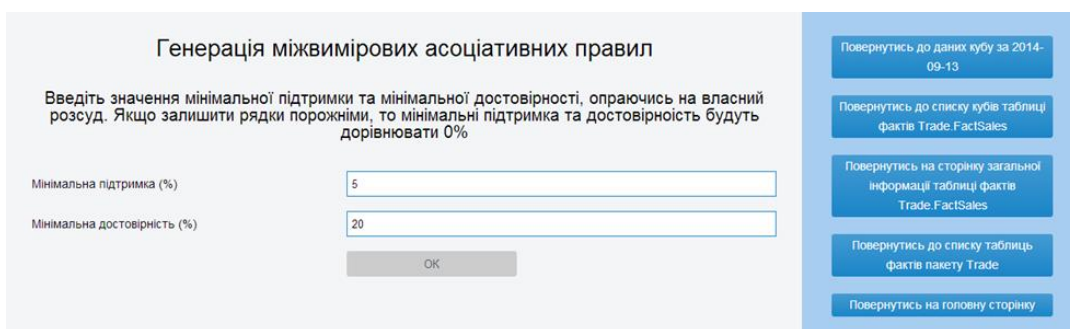


Рис. 4. Перша сторінка системи генерації міжвимірових асоціативних правил

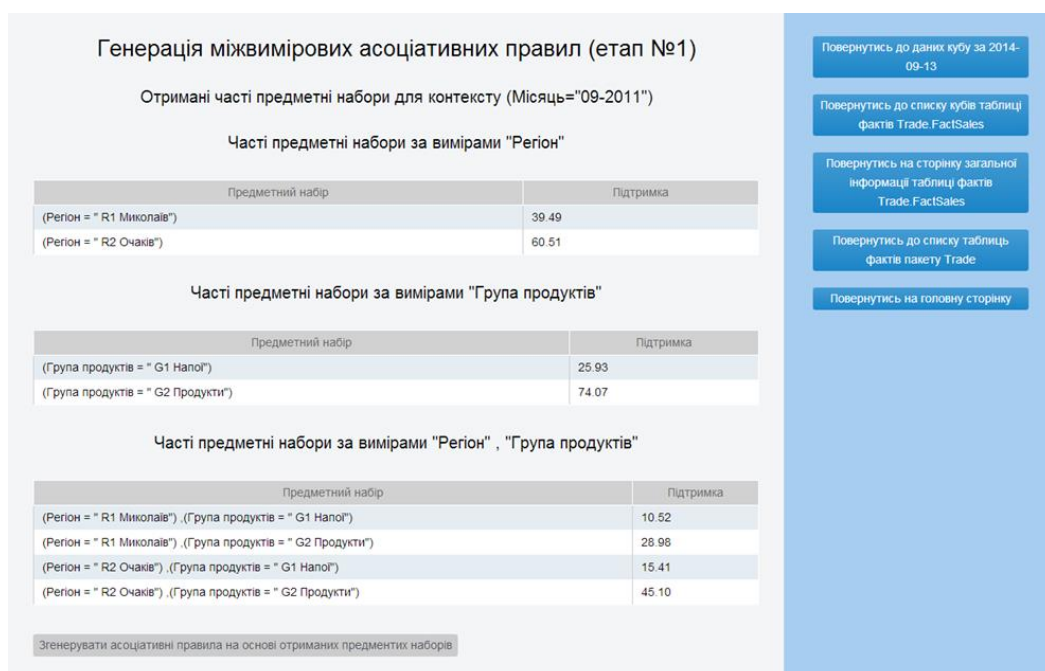


Рис. 5. Друга сторінка системи генерації міжвимірових асоціативних правил

Генерація міжвимірних асоціативних правил (етап №2)						
Отримані міжвимірні асоціативні правила						
Контекст правила	Шаблон правила	Правило	Підтримка	Достовірність	Ліфт	Лeverидж
(Місяць="09-2011")	Якщо (Реґіон), то (Група продуктів)	Якщо (Реґіон="R1 Миколаїв"), то (Група продуктів="G1 Напої")	10.52%	26.63%	1.0272	0.0028
(Місяць="09-2011")	Якщо (Реґіон), то (Група продуктів)	Якщо (Реґіон="R1 Миколаїв"), то (Група продуктів="G2 Продукти")	28.98%	73.37%	0.9905	-0.0028
(Місяць="09-2011")	Якщо (Реґіон), то (Група продуктів)	Якщо (Реґіон="R2 Очаків"), то (Група продуктів="G1 Напої")	15.41%	25.47%	0.9823	-0.0028
(Місяць="09-2011")	Якщо (Реґіон), то (Група продуктів)	Якщо (Реґіон="R2 Очаків"), то (Група продуктів="G2 Продукти")	45.10%	74.53%	1.0062	0.0028
(Місяць="09-2011")	Якщо (Група продуктів), то (Реґіон)	Якщо (Група продуктів="G1 Напої"), то (Реґіон="R1 Миколаїв")	10.52%	40.57%	1.0272	0.0028
(Місяць="09-2011")	Якщо (Група продуктів), то (Реґіон)	Якщо (Група продуктів="G2 Продукти"), то (Реґіон="R1 Миколаїв")	28.98%	39.12%	0.9905	-0.0028
(Місяць="09-2011")	Якщо (Група продуктів), то (Реґіон)	Якщо (Група продуктів="G1 Напої"), то (Реґіон="R2 Очаків")	15.41%	59.43%	0.9823	-0.0028
(Місяць="09-2011")	Якщо (Група продуктів), то (Реґіон)	Якщо (Група продуктів="G2 Продукти"), то (Реґіон="R2 Очаків")	45.10%	60.88%	1.0062	0.0028

Рис. 6. Третя сторінка системи генерації міжвимірних асоціативних правил

Слід звернути увагу на те, що правила будуть збережені у БД до того моменту, поки кінцевий користувач не визначить новий контекст правил та міру кубу, за якою слід їх генерувати, а також нові значення мінімальних підтримки та достовірності. У даному випадку старі асоціативні правила, що зберігались у БД, будуть видалені і замінені новими, згенерованими згідно заданих умов. Асоціативні правила зберігаються у БД не довгостроково, а тимчасово. Але у системі існує опція збереження асоціативних правил у текстовий файл, завдяки чому їх можна розглянути у будь-який момент часу незалежно від того, збережені вони на даний момент у БД або вже були видалені.

6. Висновки

У роботі досліджено асоціативні правила, що можуть мати місце між вимірами багатомірних кубів, виведені формули обчислення їх характеристик значущості та запропоновано метод їх генерації.

Результатом роботи стала інтелектуальна інформаційна система, що реалізована у постреляційній СКБД Caché (вона підтримує, зокрема, її об'єктну модель) за допомогою технології CSP (Cache Server Pages).

У подальшому планується дослідження інших видів асоціативних правил у багатомірних даних: правил в межах одного виміру, які визначають залежності між атрибутами одного й того самого виміру, та гібридних правил, що представляють залежності між вимірами, однак деякі операнди можуть бути атрибутами одного й того самого виміру.

Також планується дослідження застосування інших методів Data Mining (класифікація, кластеризація, прогнозування і т. д.) на багатомірних даних з подальшою реалізацією.

Література

1. Chaudhuri, S. Data Mining and Database Systems: Where is the Intersection? [Text] / S. Chaudhuri // Data Engineering Bulletin. – 1998. – Vol. 21, Issue 1. – P. 4–8.
2. Chaudhuri, S. Scalable Classification over SQL Databases. [Text] / S. Chaudhuri, U. Fayyad, J. Bernhardt // In Proceedings of the 15th International Conference on Data Engineering (ICDE'1999), 1999. – P. 470–479. doi: 10.1109/icde.1999.754963

3. Meo, R. A New SQL-like Operator for Mining Association Rules. In Proceedings of the 22nd International Conference on Very Large Data Bases Conference (VLDB'1996) [Text] / R. Meo, G. Psaila, S. Ceri. – Bombay, India, 1996. – P. 122–133.

4. Zhu, H. Online analytical mining of association rules [Text] / H. Zhu // Master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada, 1998.

5. Fayyad, U. Advances in Knowledge Discovery and Data Mining [Text] / U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy // AAAI/MIT Press, 1996.

6. Messaoud, R. B. A data mining-based OLAP aggregation of complex data: Application on XML documents [Text] / R. B. Messaoud, O. Boussaid, S. L. Rabaséda, // International Journal of Data Warehousing and Mining. – 2006. – Vol. 2, Issue 4. – 1–26. doi: 10.4018/jdwm.2006100101

7. Han, J. Toward On-Line Analytical Mining in Large Databases [Text] / J. Han // SIGMOD Record. – 1998. – Vol. 21, Issue 27. – P. 97–107. doi: 10.1145/273244.273273

8. Chen, Q. An Olap-based Scalable Web Access Analysis Engine [Text] / Q. Chen, U. Dayal, M. Hsu // In Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery (DAWAK'2000), 2000. – P. 210–223. doi: 10.1007/3-540-44466-1_21

9. Goil, S. High performance multidimensional analysis of large datasets [Text] / S. Goil, A. Choudhary // DOLAP '98. Proceedings of the 1st ACM international workshop on Data warehousing and OLAP, 1998. – P. 34–39. doi: 10.1145/294260.294269

10. Pinto, H. Multi-dimensional sequential pattern mining [Text] / H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen, U. Dayal // In CIKM '01: Proceedings of the tenth international conference on Information and knowledge management, 2001. – P. 81–88.

11. Goil, S. PARSIMONY: An Infrastructure for parallel Multidimensional Analysis and Data Mining [Text] / S. Goil, A. Choudhary // Journal of Parallel and Distributed Computing. – 2011. – Vol. 61, Issue 3. – P. 285–321. doi: 10.1006/jpdc.2000.1691

12. Tjioe, H. C. Mining Association Rules in Data Warehouses [Text] / H. C. Tjioe, D. Taniar // International Journal of Data Warehousing and Mining, Idea Group Inc. – 2005. – Vol. 1, Issue 3. – P. 28–62.

13. Parsaye, K. OLAP and Data Mining: Bridging the Gap [Text] / K. Parsaye // Database Programming and Design. – 1997. – Vol. 10. – P. 30–37.

14. Sarawgi, S. Discovery-driven Exploration of OLAP Data Cubes. [Text] / S. Sarawgi, R. Agrawal, N. Megiddo // In Proceedings of the 6th International Conference on Extending Database Technology (EDBT'1998), Valencia, Spain. – 1998. – P. 168–182. doi: 10.1007/bfb0100984

15. Robin, J. HYSSOP: Natural Language Generation Meets Knowledge Discovery in Databases [Text] / J. Robin, E. Favero // In Proceedings of the 3rd International Conference on Information Integration and Web-based Applications and Services (iiWAS'2001), 2001.

16. Фісун, М. Т. Аналіз особливостей об'єктної та багатовимірної моделей даних в СКБД [Текст] / М. Т. Фісун, Г. В. Горбань // Вестник Херсонского национального технического университета. – 2011. – Вып. 2, № 41. – С. 116–124.

17. Горбань, Г. В. Застосування B*-дерев для створення та обчислення OLAP-кубів з використанням комбінаторного алгоритму [Текст] / Г. В. Горбань // Технологический аудит и резервы производства. – 2013. – Т. 5, № 4 (13). – С. 10–12. – Режим доступа: <http://journals.uran.ua/tarp/article/view/18216/15955>

References

1. Chaudhuri, S. (1998). Data Mining and Database Systems: Where is the Intersection? Data Engineering Bulletin, 21 (1), 4–8.

2. Chaudhuri, S., Fayyad, U., Bernhardt, J. (1999). Scalable classification over SQL databases. Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337). doi: 10.1109/icde.1999.754963

3. Meo, R., Psaila, G., Ceri, S. (1996). A New SQL-like Operator for Mining Association Rules. In Proceedings of the 22nd International Conference on Very Large Data Bases Conference (VLDB'1996), Bombay, India, 122–133.

4. Zhu, H. (1998). Online analytical mining of association rules. Master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

5. Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press.

6. Messaoud, R. B., Boussaid, O., Rabaséda, S. L. (2006). A Data Mining-Based OLAP Aggregation of Complex Data. International Journal of Data Warehousing and Mining, 2 (4), 1–26. doi: 10.4018/jdwm.2006100101

7. Han, J. (1998). Towards on-line analytical mining in large databases. SIGMOD Rec., 27 (1), 97–107. doi: 10.1145/273244.273273

8. Chen, Q., Dayal, U., Hsu, M. (2000). An OLAP-based Scalable Web Access Analysis Engine. Lecture Notes in Computer Science, 210–223. doi: 10.1007/3-540-44466-1_21

9. Goil, S., Choudhary, A. (1998). High performance multidimensional analysis of large datasets. Proceedings of the 1st ACM International Workshop on Data Warehousing and OLAP - DOLAP '98. doi: 10.1145/294260.294269

10. Pinto H., Han J., Pei J., Wang K., Chen Q., Dayal U. (2001). Multi-dimensional sequential pattern mining. In CIKM '01: Proceedings of the tenth international conference on Information and knowledge management, New York, NY, USA., ACM Press, 81–88.

11. Goil, S., Choudhary, A. (2001). PARSIMONY: An Infrastructure for Parallel Multidimensional Analysis and Data Mining. Journal of Parallel and Distributed Computing, 61 (3), 285–321. doi: 10.1006/jpdc.2000.1691

12. Tjioe, H.C., Taniar, D. (2005). Mining Association Rules in Data Warehouses. International Journal of Data Warehousing and Mining, Idea Group Inc., 1 (3), 28–62.

13. Parsaye, K. (1997). OLAP and Data Mining: Bridging the Gap. Database Programming and Design, 10, 30–37.

14. Sarawagi, S., Agrawal, R., Megiddo, N. (1998). Discovery-driven exploration of OLAP data cubes. Lecture Notes in Computer Science, 168–182. doi: 10.1007/bfb0100984

15. Robin, J., Favero, E. (2001). HYSSOP: Natural Language Generation Meets Knowledge Discovery in Databases. In Proceedings of the 3rd International Conference on Information Integration and Web-based Applications and Services (iiWAS'2001).

16. Fisun, M., Horban, H. (2011). Analysis of specific features of the objective and multidimensional data models in DBMS Caché. Bulletin of Kherson National Technical University, 2 (41), 116–124.

17. Gorban', G. V. (2013). Zastosuvannja B*-derez dlja stvorennja ta obchislennja OLAP-kubiv z vikoristannjam kombinatornogo algoritmu. Har'kov, Tehnologicheskij audit i rezervy proizvodstva, 5/4 (13), 10–12. Available at: <http://journals.uran.ua/tarp/article/view/18216/15955>

Дата надходження рукопису 20.05.2015

Фісун Микола Тихонович, доктор технічних наук, професор, кафедра інтелектуальних інформаційних систем, Чорноморський державний університет ім. Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54003

E-mail: mykola.fisun@gmail.com

Горбань Гліб Валентинович, кафедра інтелектуальних інформаційних систем, Чорноморський державний університет ім. Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54003

E-mail: gleb.gorban@gmail.com

УДК 693.6:69.003

DOI: 10.15587/2313-8416.2015.43907

НОРМИРОВАНИЕ УСТРОЙСТВА СЛОЖНЫХ КОНСТРУКТИВНЫХ ФОРМ С ИСПОЛЬЗОВАНИЕМ КОМПЛЕКТНЫХ СИСТЕМ КНАУФ

© Д. А. Хохрякова

Научная работа предусматривает создание нормы времени для устройства потолка-оболочки размером в плане 9300х6399 мм, высотой 2600 мм. Нормативные наблюдения устройства потолка-оболочки производились методом смешанного фотоучета. Произведен сопоставительный анализ результатов исследований с существующими сметными нормами

Ключевые слова: системы КНАУФ, потолок-оболочка, норма времени, технологический процесс, индивидуальная расценка