
Bibliography

1. Назаров А.Н., Симонов М.В. АТМ: технология высокоскоростных сетей. М.: ЭКО-ТРЕНДЗ, 1997.
2. Зайченко О.Ю. Оптимізація характеристик мереж з технологією АТМ // Системні дослідження та інформаційні технології. – 2002. – №3. – С.57– 73.
3. Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации – К.: Наук. думка, 1988.
4. Lourenço H. R., Martin O., Stützle T. Iterated local search // Handbook of Metaheuristics: International Series in Operations Research & Management Science, vol. 57 (Eds. F. Glover and G. Kochenberger). – Norwell: Kluwer Academic Publishers, MA, 2002. – P. 321-353.
5. Aarts E., Korst J. Simulated annealing and Boltzmann machines. – New York: John Willey and Sons, 1989.
6. Гуляницький Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – Киев: Институт кибернетики им. В.М. Глушкова НАН Украины, 2004. – №1. – С. 64-72.
7. Reeves C.R., Rowe J.E. Genetic Algorithms: Principles and Perspectives. A Guide to GA Theory. – Boston: Kluwer Academic Publishers, 2002.

Authors' Information

Leonid F. Hulianytskyi – V.M.Glushkov Institute of Cybernetics of NAS of Ukraine, Glushkova ave. 40, Kyiv, 03680, Ukraine; e-mail: lh_dar@hotmail.com

Andrii M. Baklan – V.M.Glushkov Institute of Cybernetics of NAS of Ukraine, Glushkova ave. 40, Kyiv, 03680, Ukraine; e-mail: a_baklan@yahoo.com

ANALYZING THE DATA IN OLAP DATA CUBES*

Galina Bogdanova, Tsvetanka Georgieva

Abstract: OLAP applications provide a possibility to data analysis over large collections of historical data in the data warehouses, supporting the decision-making process. This paper presents an application that creates a data cube and demonstrates the effectiveness of the applying the OLAP operations when it necessary to analyze the data and obtain the valuable information from the data. It allows the analysis of factual data that is daily downloads of folklore materials, according to dimensions of interest.

Keywords: data cube, online analytical processing, multidimensional expressions

ACM Classification Keywords: H.4.2 Information Systems Applications: Types of Systems – Decision support

1. Introduction

Decision-support functions in a data warehouse, such as online analytical processing (OLAP), involve hundreds of complex aggregate queries over large volumes of data. It is not feasible to compute these queries by scanning the data sets each time [9]. The data cubes are structures designed to provide quick access to the data in data warehouses. The cube definition is determined from the requirements, which the users analyzing the data have

* Supported partially by the Bulgarian National Science Fund under Grant MM-1405/2004

and it is based on the choice of the schema representing dimensional model of the data that consists of one fact table and several dimension tables.

Some requirements of decision-support systems (DSS) are considered in [4]. Several common considerations of designing data cubes are examined in [5]. In [11] an OLAP server supporting dimension updates is presented and view maintenance under these updates is proposed. The advantages of the aggregation ranking queries in OLAP data cubes are described in [6]. In the present paper the OLAP operations are applied to analyze the data in a WEB based client/server system containing archival fund with folklore materials of the Folklore Institute at BAS. It represents the data cube creation and the dimension hierarchies. Part of present paper is reported in [2].

The rest of the paper is organized as follows. Section 2 reviews the concepts of the data cubes, the lattice corresponding to a data cube and the OLAP operations. Section 3 presents the data cube creation and applying the OLAP operations by using the language MDX (*Multidimensional Expressions*). Section 4 gives the conclusion of this paper.

2. Data Cubes and OLAP Operations

Data cubes are popular in OLAP because they provide an intuitive way for data analysts to navigate various levels of summary information in the database [9]. In a data cube, attributes are categorized into *dimension attributes*, on which grouping may be performed, and *measures*, which are the results of aggregate functions. From a data cube with n dimension attributes can be obtained 2^n cube views. For example, a data cube with dimensions A_1, A_2, A_3 is shown in figure 1(a) as a lattice structure.

The dimensions often are organized into dimension hierarchies, which can also be represented by a lattice. For example, figure 1(b) shows the lattice for the dimension hierarchies where a_{ij} is j -th level in the hierarchy of the dimension A_i . The top element of each lattice is "all", meaning no grouping by that dimension.

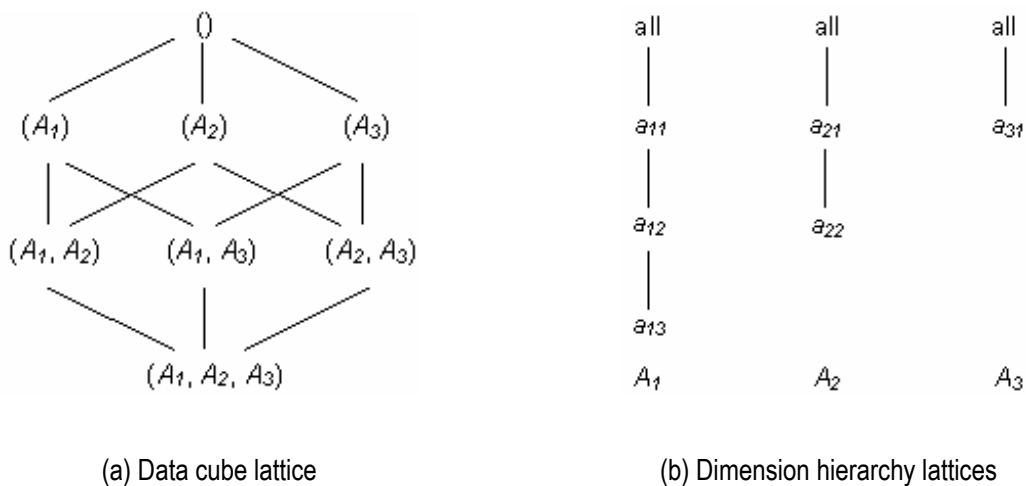


Fig. 1 Example lattices

It can be constructed a lattice representing the set of views that can be obtained by grouping on each combination of elements from the set of the dimension hierarchies. Figure 2 shows the lattice combining the data cube lattice of figure 1(a) with the dimension hierarchy lattices of figure 1(b).

OLAP includes a set of operations for manipulation of the dimensional data organized in multiple levels of abstraction. The basic OLAP operations are roll-up (increasing the level of aggregation); drill-down (decreasing the level of aggregation); slice-and-dice (selection and projection); pivot (re-orienting the multidimensional view of data) [3].

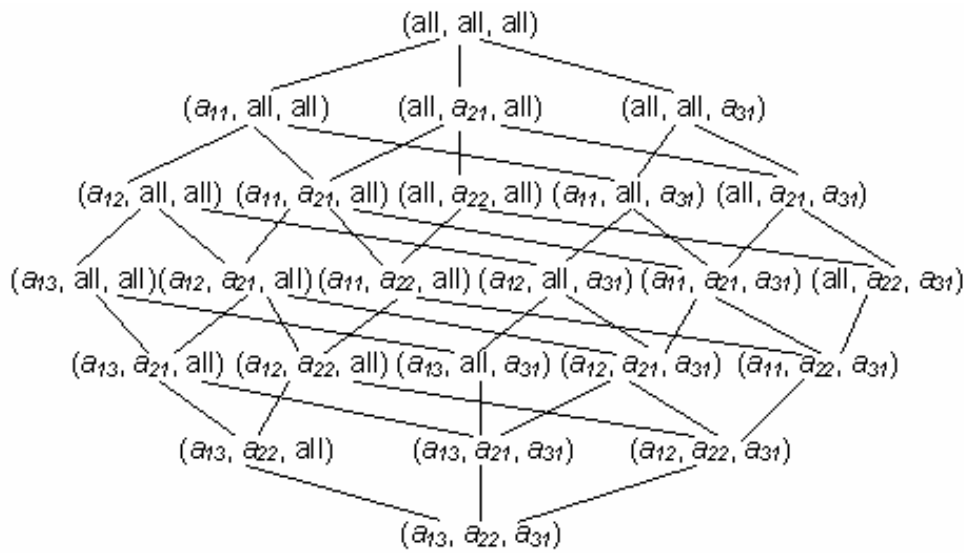


Fig. 2 Combined lattice

3. Creation of the Data Cube FolkloreCube and Analyzing the Data by Applying the OLAP Operations

The investigated archive keeps detailed information of the documents and materials, which can be downloaded by the users and contain audio, video and text information.

3.1. The Relational Database FolkloreDB

The OLTP (online transaction processing) database FolkloreDB is created in accordance to the classification schema described in [7]. This database is realized by using the client/server relational database management system (RDBMS) Microsoft SQL Server [8, 10] and consists of the tables shown in figure 3.

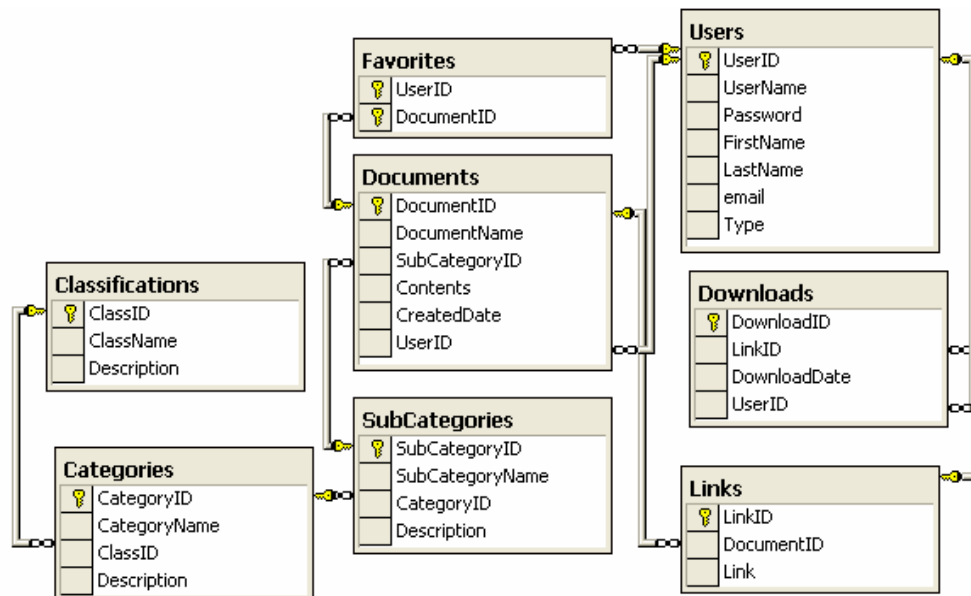


Fig. 3 The structure of the database FolkloreDB

3.2. The Data Warehouse Database FolkloreDB_DW

The database FolkloreDB_DW in the data warehouse is designed by using the dimensional model represented by the star schema. This database consists of one fact table Downloads_fact and four dimension tables Documents_DW, Links_DW, Users_DW and Dates (fig. 4).

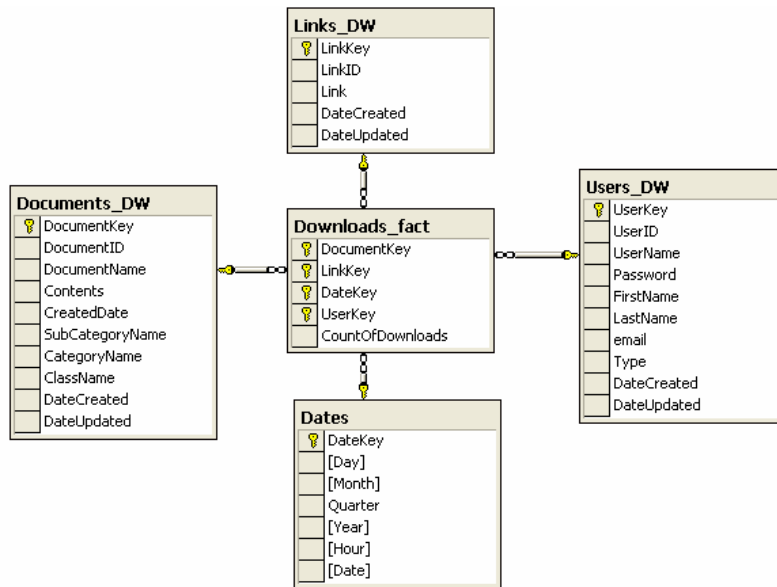


Fig. 4 The structure of the database FolkloreDB_DW in the data warehouse

3.3. Creation of the Data Cube FolkloreCube

The data cube FolkloreCube is created in correspondence with the star schema of the dimensional model of the database FolkloreDB_DW. The represented statement defines the data cube structure.

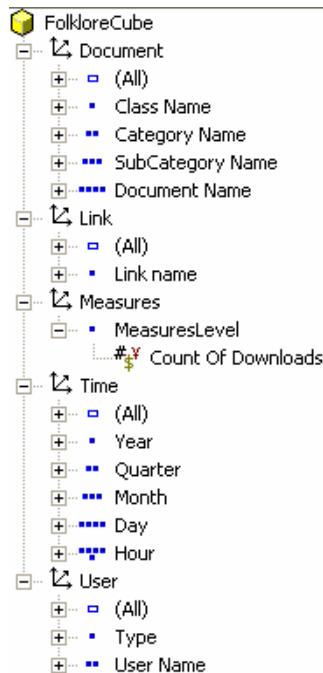


Fig. 5 The structure of the data cube FolkloreCube

```

CREATE CUBE FolkloreCube(
    DIMENSION Document,
        LEVEL [All Documents] TYPE ALL,
        LEVEL [Class name],
        LEVEL [Category name],
        LEVEL [SubCategory name],
        LEVEL [Document name],
    DIMENSION [Link],
        LEVEL [All Links] TYPE ALL,
        LEVEL [Link name],
    DIMENSION [User],
        LEVEL [All Users] TYPE ALL,
        LEVEL [Type],
        LEVEL [User name],
    DIMENSION [Time] TYPE TIME,
        LEVEL [All Time] TYPE ALL,
        LEVEL [Year] TYPE YEAR,
        LEVEL [Quarter] TYPE QUARTER,
        LEVEL [Month] TYPE MONTH,
        LEVEL [Day] TYPE DAY,
        LEVEL [Hour] TYPE HOUR,
    MEASURE [Count of downloads]
        Function Sum)

```

It is executed from the Visual Basic project [1, 12] and creates the local cube (.*cube*) by using Microsoft ActiveX Data Objects and ActiveX Data Objects Multidimensional (ADO MD).

This statement adds data into already defined cube:

```

INSERT INTO FolkloreCube
(Document.[Class name], Document.[Category name],
 Document.[Subcategory name], Document.[Document name],
 Link.[Link name], User.[Type], User.[User name], Time.[Year],
 Time.[Quarter], Time.[Month], Time.[Day], Time.[Hour],
 Measures.[Count of downloads])
SELECT d.ClassName, d.CategoryName, d.SubCategoryName,
       d.DocumentName, l.Link, u.Type, u.UserName, t.Year,
       t.Quarter, t.Month, t.Day, t.Hour, f.CountOfDownloads
FROM Documents_DW d, Downloads_fact f, Links_DW l, Users_DW u,
     Dates t
WHERE d.DocumentKey = f.DocumentKey AND f.LinkKey = l.LinkKey
      AND u.UserKey = f.UserKey AND t.DateKey = f.DateKey

```

The dimension hierarchies of the data cube FolkloreCube are shown in figure 5.

3.4. MDX Queries

MDX is Microsoft OLAP query language [13, 14, 15]. MDX queries are applied to the data cube FolkloreCube providing the dimensional view of summarized data.

3.4.1. Top 5 documents from which are downloaded materials (fig. 6);

The execution of the following MDX query provides the result represented in figure 6:

```

SELECT {Measures.[Count of downloads]} ON COLUMNS,
       TOPCOUNT(Document.[Document name].Members, 5,
       Measures.[Count of downloads]) ON ROWS
FROM FolkloreCube

```

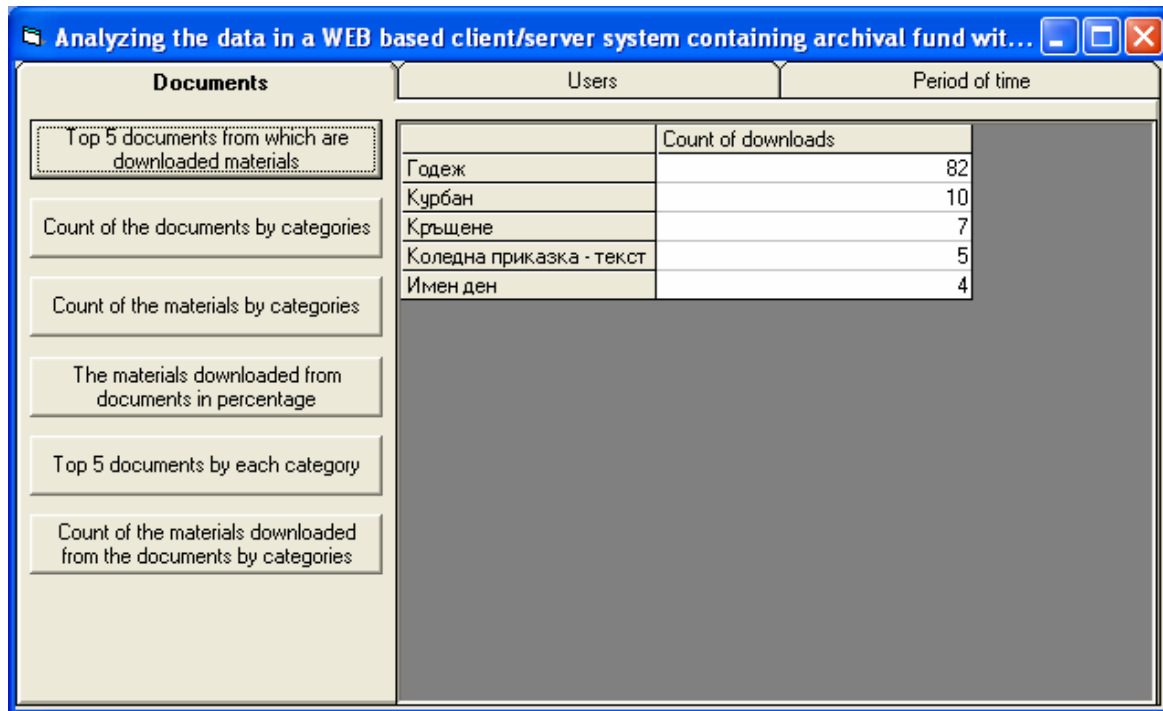


Fig. 6 Exemplary result from applying OLAP operations in FolkloreCube

3.4.2. The documents from which are downloaded the most of materials from the top 10 users (fig. 7);
 To obtain the result show in figure 7 the represented application performs the MDX query:

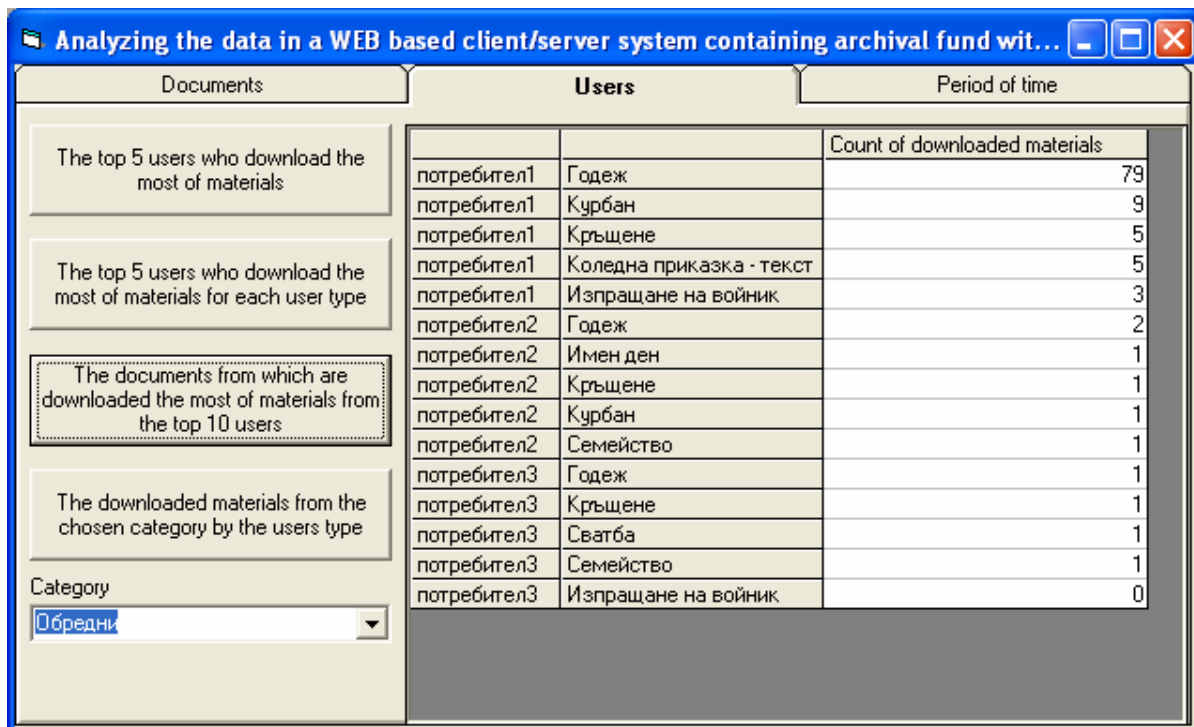


Fig. 7 Example for 3-dimensional view of the data

```

WITH SET Top10Users AS 'TOPCOUNT(User.[User name].Members, 10,
                        Measures.[Count of downloads])'
MEMBER Measures.[Count of downloaded materials] AS
  'CoalesceEmpty(Measures.[Count of downloads], 0)'
SELECT {Measures.[Count of downloaded materials]} ON COLUMNS,
       {GENERATE(Top10Users, CROSSJOIN({User.CURRENTMEMBER},
                                       TOPCOUNT(Document.[Document name].Members, 5,
                                       Measures.[Count of downloads])))} ON ROWS
FROM FolkloreCube

```

3.4.3. Count of the materials downloaded from documents by the hours of the chosen date and the difference with the previous hour (fig. 8).

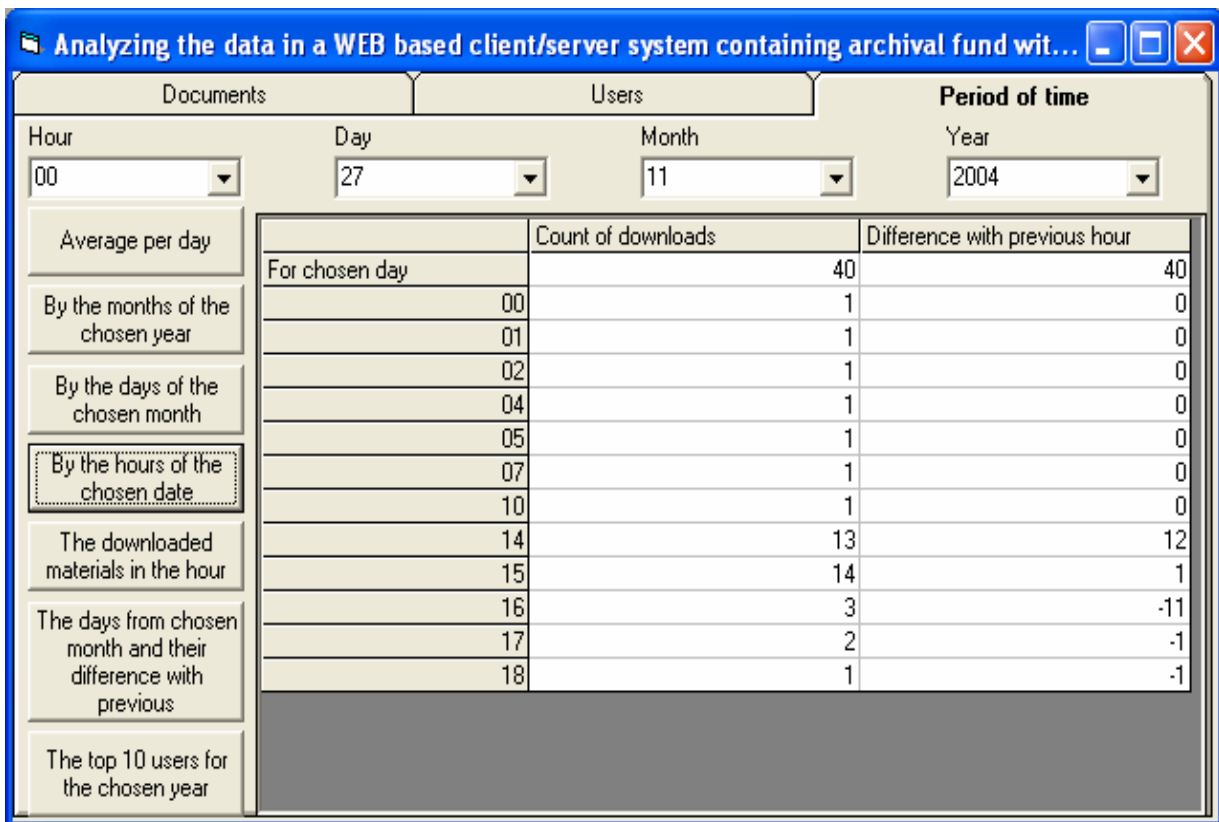


Fig. 8 Analyzing the data in FolkloreCube for a given period of time

```

WITH SET M AS '{Time.[2004].[Q4].[11].[27]}'
MEMBER Time.[For chosen day] AS 'M.Item(0)'
MEMBER Measures.[Difference with previous hour] AS
  '(Measures.[Count of downloads], Time.CURRENTMEMBER) -
  (Measures.[Count of downloads],
   Time.CURRENTMEMBER.PrevMember)'
SELECT {Measures.[Count of downloads],
       Measures.[Difference with previous hour]} ON COLUMNS,
       {Time.[For chosen day],
       GENERATE(M, {Time.CURRENTMEMBER.children})} ON ROWS
FROM FolkloreCube

```

4. Conclusion

With rapid developments of data warehouses and OLAP technologies and with enormous amount of data stored in databases in result of daily activities of the organizations it is increasingly important to develop the database applications that convert huge volumes of data into meaningful information. This assists in the decision-making process in different area by providing feedback on past actions of the users and helped to guide future decisions. Represented application gives different views of the data collected in a WEB based client/server system that contains archival fund with folklore materials.

Bibliography

- [1] Bekuit, B., VB.NET: A Beginner's Guide, AlexSoft, 2002, pages 330.
- [2] Bogdanova, G., Tsv. Georgieva, Applying the OLAP Operations to Analyzing the Data in a WEB based Client/Server System Containing Archival Fund with Folklore Materials, National Workshop on Coding Theory, Bankya, 9-12.12.2004.
- [3] Garcia-Molina, H., J. D. Ullman, J. Widom, Database Systems: The Complete Book, Williams, 2002, pages 1083.
- [4] Georgieva, Tsv., Data Warehousing and OLAP Technology, In Proc. of the International Conference on Computer Systems and Information Technologies, Veliko Tarnovo, 3-5.10.2003, pages 58-65 (in bulgarian).
- [5] Georgieva, Tsv., Designing OLAP Data Cubes, In Proc. of the International Conference on Computer Systems and Information Technologies, Veliko Tarnovo, 3-5.10.2003, pages 106-112 (in bulgarian).
- [6] Li, H., H. Yu, D. Agrawal, A. Abbadi, Ranking Aggregates, Technical Report 2004-07, Department of Computer Science, University of California, Santa Barbara, 2004, pages 14.
- [7] Mateeva, V., I. Stanoeva, Classification Scheme of the Typological Catalogue in the Folklore Institute, Bulgarian Folklore, v. 2-3, 2001, pages 96-109 (in bulgarian).
- [8] Microsoft Corporation, MCSE Training: Microsoft SQL Server 2000 Database Design and Implementation, SoftPress Ltd., 2001, pages 800.
- [9] Mumick, I., D. Quass, B. Mumick, Maintenance of Data Cubes and Summary Tables in a Warehouse, In Proc. ACM SIGMOD Conf. on Management of Data, Tuscon, Arizona, 1997, pages 100-111.
- [10] Soukup, R., K. Delaney, Inside Microsoft SQL Server 7.0, SoftPress Ltd., 2000, pages 952.
- [11] Vaisman, A., A. Mendelzon, W. Ruaro, S. G. Cymerman, Supporting Dimension Updates, In Proc. of the 14th International Conference on Advanced Information Systems Engineering, 2002, pages 67-82.
- [12] Wang, W., Visual Basic 6: A Beginner's Guide, AlexSoft, 2002, pages 562.
- [13] <http://www.georgehernandez.com/xDatabases/MD/MDX.htm>
- [14] <http://www.microsoft.com/data/oledb/olap>
- [15] <http://www.microsoft.com/sql>

Authors' Information

Galina Bogdanova – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Veliko Tarnovo, P.O.Box: 323, e-mail: galina@moi.math.bas.bg

Tsvetanka Georgieva – University of V. Tarnovo "St. St. Cyril and Methodius", Department of Information Technologies; e-mail: cv.georgieva@uni-vt.bg