Comparative Analysis of Advanced Technologies for Processing of Large Data Sets

I. Valova, M. Noirhomme-Fraiture

Key Words: *On-line Analytic Data Processing; Data Mining, Symbolic Data Analysis and Visualization of data.*

Abstract. *In this paper the applied and theoretical results achieved, as well as some existing drawbacks in technologies for processing of large data sets-OLAP (On-line Analytic Processing), DM (Data Mining) and SDA (Symbolic Data Analysis) are analyzed. A comparative analysis is proposed on different types of data processing and the pros and cons of each one of them are highlighted. Here are discussed benefits and drawbacks of using data aggregates and visualization of large data set. Some topics of interest are shown for the purposes of additional scientific study, being specifically oriented to software applications.*
The paper is related with a joint implementation project between Bulgarian Academy of Science, ICSR, Sofia and CGRI/ Institut d'Informatique, FUNDP, Namur, Belgium.

1. Introduction

The increased computer power combined with the need to analyze huge data sets, created conditions for development of new techniques and technologies. These include development of new algorithms and new approaches, e.g. use of Intuitionistic Fuzzy Logic (IFL) as well as development of new methodologies (SDA) or applying of new approaches to existing algorithms. Possession of large data base by any company is insignificant, if end users may not easily synthesize necessary information. Frequently, data has valuable additional hidden potential. This is completely new information that is displayed in the form of meaningful interrelationships between the data that are either too well hidden or too complex to be discovered just by looking at the data. The data is frequently originated from different sources. Data is retrieved, consolidated, managed and prepared for analysis. A need is identified for reliable tools for analysis of company data, which to complement the existing data management systems and to be robust enough to predict and facilitate complex analysis of business data. The goal to extract new knowledge from huge data sets has urged the need to extend standard data analysis methods (exploratory, clustering, factorial analysis, discrimination,...). A task for the research community will be also to investigate how the achievements in a specific field may be applied in new data processing technologies, e.g. how the achievements of existing solutions in the fields of heterogeneous and distributed databases could be transferred to OLAP environments.

The need for applying of new methodologies in OLAP & DM systems is aiming to satisfy requirements of fast developing information technologies, and to provide new capabilities for end-users and managers of organizations for their decision-making process. As being driven by constantly increasing business needs, technical developments in the field of OLAP products are strongly market oriented. The research problems related with OLAP&DM attract the attention of researchers, but the output is not enough as compared with the market success of respective software products. The high prices of software products and requirements for powerful hardware discourage any research strive, and on other hand software developers are deeply occupied to create software applications for large organization, i.e. no viable cooperation between science and industry is available and as end result there exist no generally accepted technology and terminology on issues related with OLAP&DM systems. It is still necessary existing concepts to be clarified, the knowledge transfer to be performed in order to have a clear understanding on future research and needs to be met by developments to come.

The paper is related with a joint implementation project between the Bulgarian Academy of Science, Sofia and the CGRI/ Institut d'Informatique, FUNDP, Namur, Belgium. The mutual interest is determined by close themes of both research groups, related with information technologies as well as by necessity of knowledge transfer and sharing of achieved results. The project in question was developed with intention to achieve specific scientific results that reflect dynamics and innovations in the field of OLAP and DM technology by means of application of some new research methodology approaches, which may be of interest to vendors of software systems.

This paper is structured as follows: In section 2 key benefits and drawbacks of different types of OLAP architectures are highlighted. In section 3 the attention is focused on DM and different data processing methods. In section 4 different visualization methods of data sets in aggregated form are described. Here are discussed benefits and drawbacks at using of data aggregates. Application of new methodologies for analytical data processing is discussed in sections 5 and 6.

2. OLAP

The term OLAP (On-Line Analytical Processing) was initially introduced by the renowned and respected database researcher E. F. Codd in 1993 [1]. In 1995 he added six new rules to the original 12 for OLAP systems and restructured them into four groups, calling them features- *table 1.*

Table 1. The Codd's rules for OLAP

B Basic Features	F1-Multidimensional Conceptual View	F2-Intuitive Data Manipulation
	F3- Accessibility: OLAP as a Mediator	F4-Batch Extraction vs Interpretive
	F5-OLAP Analysis Models	F6-Client Server Architecture
	F7-Transparency	F8- Multi-User Support
S Special Features	F9-Treatment of Non-Normalized Data	F10-Storing OLAP Results: Keeping Them Separate from Source Data
	F11- Extraction of Missing Values	F12- Treatment of Missing Values
R Reporting Features	F13-Flexible Reporting	F14-Uniform Reporting Performance
	F15-Automatic Adjustment of Physical Level	
D Dimensio n Control	F16- Generic Dimensionality	F17-Unlimited Dimensions & Aggregation Levels
	F18-Unrestricted Cross-Dimensional Operations.	

The most of OLAP software products meet the Codd's requirements for OLAP compatibility, provided that distinction is made between the rules having relation with the research approaches and application technologies. We think that some of Codd's rules may be used for improvement of the existing software technologies, and the remaining part thereof should be further developed and improved by the research community before to be proposed for practical realization.

Our brief evaluation of some of these rules is given below and it will be subject to a more detailed overview in another paper.

F1: All modern software products comply with this indication. In the field of research a special attention should be drawn on the used terminology.

F2, F3, F4, F6, F7, F8 - The experts in software technologies should exert more efforts for achievement of these requirements for on-line analytical data processing.

F5: The scientific research should be focused on issues for clarification of different *types of models*, which may be introduced. The use of mathematical calculations and definitions should be more understandable and well presented. For us the issues on terminology remain as subject to further discussion. For example: Multidimensional analysis or Analysis of multidimensional models? Multidimensional modeling or Modeling of multidimensional models? Dimensions or D-structures?

F13-F15 - There exist good achievements in the field of commercial products, such as Panorama technology, used in Microsoft Analysis Services. The product SAP BW meets these rules in sufficient extent.

F16-F18 – good rules, which will be analyzed and evaluated in other paper.

The basic core in the systems for analytical data processing is the creation of multidimensional models. The multidimensionality is also the main requirement of Dr. Codd in the formulation of the OLAP term and in determination which software products are compatible with OLAP [1].

The concept of a dimensional data model that could be represented in a relational database is described by Ralph Kimball [2]. This concept gained popularity and soon at the market appeared software products with relational OLAP (ROLAP) architecture. As opposed to multidimensional OLAP (MOLAP) systems, which use storage structures especially designed for multidimensional query processing, relational OLAP systems rely on established and traditional relational databases technologies. The innovative approach is demonstrated in ROLAP engine, which implements a transformation from a query against multidimensional data model to a query against relational storage model. As ROLAP and MOLAP technologies evolved, a large debate began within the research community whose the superior architecture for OLAP applications was. In fact, each of both options had its own benefits and drawbacks. Which was best ultimately depended upon the performance, scalability, resource utilization, underlying reporting requirements and the nature of the data? Since each set of choices (relational OLAP or multidimensional OLAP) has tradeoffs, it is imperative that the application designer understand the nature of the data that is being put into the model, as well as how the end users will be using the model.

Table 2 highlights the key benefits and drawbacks to the ROLAP and MOLAP architectures.

Table 2. Comparative analysis - MOLAP and ROLAP

Architectures	*Benefits*	*Drawbacks*
MOLAP	Fast performance	Non-scalable
	Smaller as compared with ROLAP Maintains easily D-structures with high cardinality data	Lack of common technology
	Maintains easily unbalanced hierarchical D-structures	Lack of common terminology
	MOLAP queries are very powerful and flexible within OLAP processing	More difficult navigation related with access to cardinality data
ROLAP	Scalable	Difficult maintenance
	Familiar technology	Upgrade of RDBMS
	Flexibility	Not suitable for maintenance of many unbalanced hierarchical D-structures

To combine the benefits of both technologies there are increasing efforts to integrate them in the new software architecture, so called hybrid OLAP (HOLAP) systems.

Hybrid OLAP architecture may be defined as a system which supports and integrates multidimensional and relational storage for data in an equivalent manner in order to benefit from advantages of both technologies - MOLAP and ROLAP (*figure 1*).

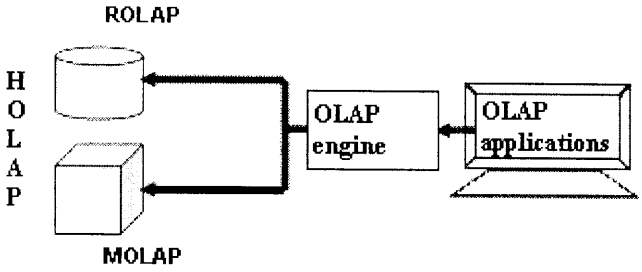


Figure 1. HOLAP architecture

A hybrid OLAP approach lets the application designer choose a combination of relational OLAP and multidimensional OLAP architectures based on the reporting requirements of the user, the system resources available, and the nature of the data. For example, the D-structures Customer and Product are sometimes problematic due to their high data cardinality. In HOLAP architecture, these D-structures can be stored in a relational OLAP schema, while all other D-structures are stored in a more manageable MOLAP architecture. The true hybrid OLAP architecture allows each piece of the model to be stored on a separate computing platform, thereby further increasing the scalability across host systems. The application designer must make choices about both what sub-tables to store in the OLAP database, and how to physically store them. Since each set of choices has tradeoffs, it is imperative that the application designer understand the nature of the data that is being put into the model, as well as how the end users will be using the multidimensional model. A data and business requirements assessment, thus, becomes one of the most critical steps in the application design process.

A task for the research community will be to investigate how the achievements of existing solutions in the fields of heterogeneous and distributed databases could be transferred to HOLAP environments.

Among popular HOLAP products are as follows: Crystal Holos, Microsoft Analysis Services, SAS CFO Vision, Speedware Media, Pilot Analysis Server, Applix iTM1, etc. The product Microsoft OLAP Services is a market leader in OLAP servers segment due to its easy installation, advance technology architecture for data storage (ROLAP/MOLAP HOLAP) and good technical support [4].

The product BW (Business Information Warehouse) of SAP supports different concepts for physical data storage: ROLAP and MOLAP. Before taking of decision whether is advisable to create MOLAP- or ROLAP-aggregates of SAP Business Information Warehouse for certain InfoCube, prerequisites must be verified for establishment of both options. Availability of such function depends on the type of database used. Contrary to ROLAP aggregates, MOLAP aggregates do not allow conditions for filtering of characteristics. BW MOLAP Version 3.5 uses Microsoft Analysis Services application and represents a specific platform method for optimization of certain query performance. BW MOLAP is available only for Microsoft SQL server. BW InfoCubes, for which MOLAP aggregates are created, basically have two conditions precedent: they may contain only pre-defined number of characteristics, navigation attributes and hierarchies and could not use specific BW objects (e.g. time-dependent objects).

The existing problem issues related with OLAP technologies may be classified as industrial and research ones.

2.1. There Exist no Generally Accepted Technology and Terminology on Issues Related with OLAP Systems

In this type of modeling the information is divided in facts and dimensions (D-structures).

Fact (F-structure) - it represents the data, subject to analysis. The fact contains numerical attributes. The fact in a multidimensional scheme is the object, which contains measures. Measures evaluate attributes of fact.

D-structure, (Dimension) - different initial viewpoints at data selection, which will be used during fact analyzing. D-structures contain mainly description attributes.

The term Dimension has been accepted by almost all research and commercial software models, but its translation (Dimensia or Izmerenie) into Bulgarian language arose indignation within the Bulgarian mathematicians and after considering the issue in-depth we adopted the term D-structure.

D-structure represents a connected directed graph, provided that each node of the graph corresponds to a given aggregation level, and its arcs reflect „part-whole" relations between the objects within the aggregation levels.

2.2. High Prices of Software Products and Requirements for Powerful Hardware

As being driven by constantly increasing business needs, technical developments in field of OLAP products are strongly market oriented. The research problems related with OLAP attract attention of researchers, but the output is not enough as compared with market success of respective software products. The high prices of software products and requirements for powerful hardware discourage any research strive, and on other hand software developers are deeply occupied to create software applications for large organization, i.□. no viable cooperation between science and industry is available and as end result there exist no generally accepted technology and terminology on issues related with OLAP systems. It is still necessary existing concepts to be clarified, the knowledge transfer to be performed in order to have a clear understanding on future research and needs to be met by developments to come [3].

2.3 HOLAP - Architectural Reality or Marketing Hype

All vendors and developers of OLAP products are working to make their products marketable as hybrid. It is critically important to closely examine the architectures of these new software applications, as their HOLAP claims may be more

marketing hype than architectural reality.

OLAP is a part of scope of tools, supporting the decision making process. OLAP technologies allow the managers, taking decisions to analyze in fast and interactive manner the multidimensional data models, according to their business considerations.

Traditional tools for queries and reports describe what is contained in a database. The software product OLAP answers the question why certain things are true. The user creates a hypothesis of specific relationship and checks its adequacy through series of queries to available data. In other words, OLAP-analyst is generating series of hypothetic models and relationships and uses queries to database in order to confirm or reject the same. The query against a multidimensional model determines which D- and F-structures will be used, what filter to apply and the navigation area. The user is provided with possibility to formulate individual, customized queries. The users may overview and evaluate data from different perspectives in course of runtime of task assigned. The whole set of functions, necessary for retrieval, processing and formatting of such data is provided by OLAP processor. Efficiency of resultant query is a dominant factor for general assessment of the system. For the purposes of further scientific research of great importance are the issues related with finding of most effective way of presentation of queries and balancing of the dynamic load between database where data is stored and the OLAP engine [4]. OLAP-analysis in its nature is a deductive process. Or, as Codd says, it is a process where the user „peels back one or more layers of the onion through subsequent simple queries".

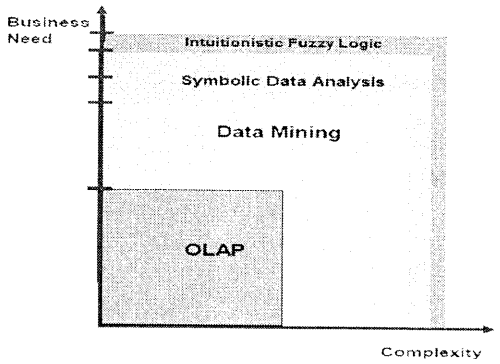


Figure 2. OLAP,DM, SDA, IFL

The processes related with DM methods differ from those in OLAP systems, as instead to verify validity of hypothetic models, they use available data to create new models. *Figure 2* shows the relation between business need in certain organization and using of OLAP&DM.

3. Data Mining (DM)

DM- deriving of valid, previously unknown information from large databases and using it at taking of critical business decisions. Data analysis is performed aiming to discover hidden or not presumed earlier connections and patterns in analyzed data. DM uses methods from different fields, such as statistics, neural networks, machine learning, etc [5].

Most frequently used DM methods, being realized in modern software products are as follows:

3.1. Decision Tree

Here data is displayed divided into categories. The resulting model is presented in the form of a tree structure. The Decision tree is one of the most popular classification algorithms in current use in Data Mining.

3.2. Clustering

Divides data up into homogeneous groups.

3.3. ABC Analysis (Pareto Analysis)

The entire production is divided into three groups of products - A, B and C. Group A products contribute to highest annual earnings, and these of Group C - to the lowest earnings.

Pareto analysis is a predecessor of ABC classification for determination of the value of inventory stock for each group of products. Pareto (1848-1923), discovered the law on nonproportional causes, where are. 80% of consequences are result of 20% of their causes. Generally, Group A products represent 80% of annual demand for products of the organization. On the other side, they account for a less than 20% of inventory stock in any given moment. Pareto diagrams may be divided in two types:

- Pareto diagrams for events serve for identification of most important issues needed for the taking of decision.
- Pareto diagrams for causes allow to discover what are major causes for occurrence of certain events in order to undertake measures for its removal.

Pareto analysis helps to highlight most important issues (these ones, called by Juran the vital few) and all efforts to be focused on its solving. For finding solution of a complex issue and removing the causes of the same, as well as for search of appropriate measures different methods may be used. An appropriate method for categorization of already identified causes is the Ishikava diagram, named after the Japanese professor Kaoru Ishikava. Such diagram has a complex branched structure and is called also fishbone diagram.

3.4. Association Analysis (Affinity Analysis or Market Basket Analysis (MBA)) is designed to determine the associations between different events. In case a customer will buy a video-player, whether he will buy another electronic device? In case a customer will buy a video-player without cassette, he would buy a DVD carrier.

The relationship will be in the form of a rule as follows:

$$If\ \{Y, \overline{X}\} Then\ \{Z\}.$$

The association analysis (MBA) is an algorithm that examines a long list of transactions in order to determine which items are most frequently purchased together. The input to a MB Analysis is normally a list of sales transactions, where each column represents a product and each row represents either a sale or a customer, depending on whether the goal of the analysis is to find which items are sold together at the same time, or to the same person.

The association rule finding algorithm is trained on historical data, i.e. past transactions. The data contains checkout information and a list of products that were purchased in each transaction, perhaps along with other information (volume, sale

amount, although in many cases just the presence or absence of a product in a transaction is sufficient).

The main problems in using algorithms of MB□ occur in analysis of a product, which is sold only 1 or 2 times as reflected in a analyzed data file. Products, which do not attract commercial interest, should not be included in the file, as the algorithm will determine rules including products, which are not statistically significant. It is preferable all products to be equally represented.

Unlike the decision tree classification, in clustering and association analysis, the models are determined on the basis of the data itself.

To obtain maximum effect, users must use such methods that are most suitable for a certain organization.

The Analysis Process Designer (APD) is the application environment for the SAP Data Mining solution.

4. Visualization of Data

Tools for graphic presentation and visualization are important help engines for data preparation and their importance in terms of data analysis is not to be underestimated. Visual analysis allows the discovery of overall trends but also smaller hidden patterns. Models, links and missing values are frequently perceived easier, when displayed graphically, than if presented as a list of figures or text. Several taxonomies and surveys are available for data visualization [e.g. 6, 7, 8, 9 and 10]. Use of pie chart on *figure 3* allows us to obtain a full picture on sales of beverages within the territory of given country. (In the specific case of *figure 3* it is shown the consumption of the given stock on □G map. For development of the model shown on Fig.3 the following software components have been used: SAP_BW - Business Information Warehouse, SAP_BASIS - SAP Basis Component, SAP_ABA - Cross-Application Component, BI_CINT - Business Intelligence Content [4].

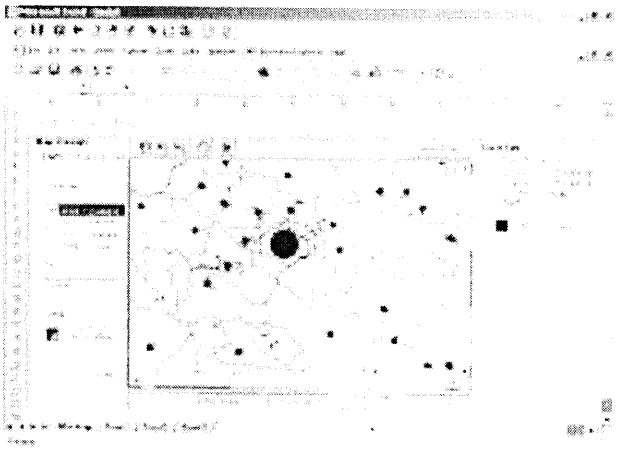


Figure 3. Map presentation using pie chart - The size of circles in individual regions shows different volume of sales of certain goods

In case we are interested in sales dynamics of several stocks within certain timeframe, *figure 4* is our ideal case. This representation has been used to visualize specific symbolic objects varying with time and is a development by BE scientists. The modules have been developed in environment Java, Open Inventor, 3D Master-Suite and Java Media Framework. Open Inventor is a 3D graphic API, using OpenGL standard. 3D Master-Suite extends and includes Open Inventor and high level 3D graphic classes. The user is provided with several means of interactivity. They concern standard visualization features such as zoom, rotation of figures and setting of color and fonts [11,12].

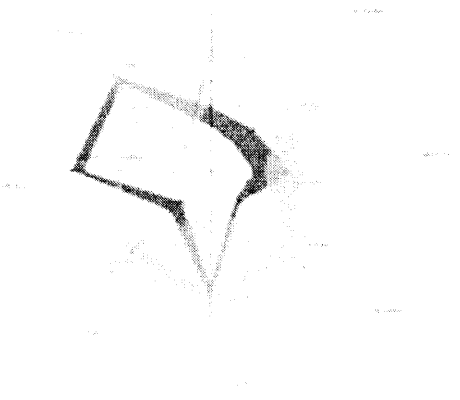


Figure 4. Example of superposition of stars. 8 stocks value for three different weeks.

This representation has been used to visualize a symbolic object varying with time. What these both methods of representation have in common is the visualization of data sets on an aggregated form. Such sets may be represented as existing data or as results of preliminary analysis.

Pros and cons in use of aggregates:

a. Aggregates improve performance at runtime of certain query, but increase loading time.

b. Aggregate must be checked regularly whether additional data is missing or not.

c. When to be compressed associated aggregates - upon entering of data or after the data was already loaded in database?

d. Aggregates allow fast access to data in reporting mode.

Data visualization is also interesting and useful theme and is worthy of further research. Integration of different representation systems by common interface is subject of future studies.

5. Symbolic Data Analysis (SDA)

The French scientist Edwin Diday defines the Symbolic Data Analysis (SDA) as the extension of standard Data Analysis. The data descriptions of the units are called symbolic when they are more complex than the standard ones due to the fact that they contain internal variation and are structured. Symbolic data happen from many sources in order to summarize huge sets of data [13]. Symbolic data analysis has been developed to solve the problem of the analysis of data known on an aggregated

form, i.e. where quantitative variables are given by intervals and where categorical variables are given by histograms.

SDA in its essence was intended as a methodology and in this regard the contributions are indisputable [13,14,15]. Articles published by Diday et al. represent a combination of philosophy (good knowledge and interpretation of works of Aristotle), mathematics and good knowledge in computer technologies. Further development of this methodology would transform SDA in a technology. For instance, by use of software intelligence tools for retrieval and transfer of logic and data, development of communication infrastructure, etc. On *figure 5* additional units are shown that may be added to existing developed modules.

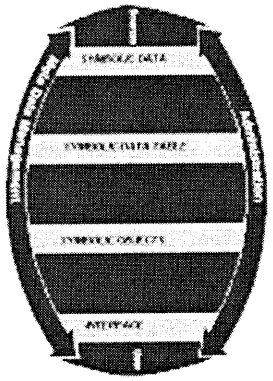


Figure 5. SDA architecture

Administrator Unit performs the following functions:

➢Provides functions for the controlling, monitoring and maintenance of all processes.

➢Performs tasks in data storage process.

➢Provides library of transformation functions that contain transaction logic for the purposes to cleanse data and to render data analyzable.

Metadata represents indispensable attribute of all modern information systems. In information systems, containing large data sets some problems occur when users need to be provided with manuals, indexes, i. e. they should have information on stored products, i. e. they need to have information on information available (metadata). The user should have available the following types of descriptions of:

- ➢ database structure;
- ➢ primary and archive data;
- ➢ attributive data;
- ➢ technology of data replication;
- ➢ validation and updating of data, etc.

In cases of change of data structure, introduction of new primary data, change or addition of new requirements to the system such metadata descriptions as above must be mandatory changed.

Recent information about new SDA modules is available on www.assoproject.be/

6. Intuitionistic Fuzzy Logic (IFL)

Intuitionistic Fuzzy (IF) Logic can be used in evaluation of the models for large data set. IF Set is defined as follows [16]:

$A=\{\langle x, \mu_A(x), \nu_A(x)\rangle / x\in E\}$,

Where E is a fixed set, functions $\mu_A:E\rightarrow[0,1]$ and $\nu_A:E\rightarrow[0,1]$ give the degree of membership and non-membership of the element $x\in E$ to set A. The set A is a subset to E and $\forall x\in E: 0\leq\mu_A(x)+\nu_A(x)\leq 1$.

The value $\pi_A(x)=1-\mu_A(x)-\nu_A(x)$ gives the degree of non-determinacy of the element x:E to the set A.

More detailed and in-depth research of IFL and its implementation in the formalization of the OLAP theory will be subject to discussion in future papers. The applicability of IFL will be especially useful in the analysis of the models, in the formalization of the relevant operators and last, but not least in the evaluation of the taken decisions

7. Conclusions

The development of information technologies, which allow storage of ever-increasing volumes of huge datasets, urged the necessity of creation and development of new methods for information analysis. Such needs has led to the rise of new and perfection of existing methodologies for data analysis, which objectives are to enhance the standard data analysis methods and to provide some clarifying results, expressed by notions of real world, and presented mathematically through symbol objects and IF logic.

All discussed technologies may be mutually supplemented.

OLAP & DM market is characterized by its higher dynamics and innovations. Research activities need to engage in continuous in-depth study of new software products aiming to provide the necessary concepts. The industry should pay adequate attention to the research community, aiming the development of new software products, satisfying requirements of fast developing information technologies. The quality of strategic and business decisions, being taken by using of the new technologies discussed in this paper, is significantly higher and they are much well-timed as compared with the decisions taken by using traditional methods.

References

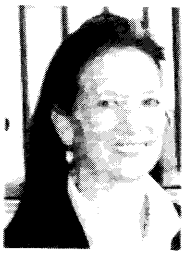
1. Codd, E., S. B. Codd, C. T. Salley. Providing OLAP to User-analysts: An IT Mandate. Technical Report, E. F. Codd & Associates, 1993.
2. Kimball, R. Data Warehouse Tool Kits. John Wiley & Sons, Toronto, 1996.
3. Valov, V., I. Valova. Some Research Problems Related with OLAP Systems, MEPhI-2007, Moscow, 2007, 133-134.
4. Valova, I. Multidimensional Object-Oriented Models of Data for Analytical Data Processing. PhD Thesis, Institute of Computer and Communication Systems-BAS, Bulgaria, 2004.
5. Weidner Jens -Integrierte Data Mining Lösung mit SAP BI, (SAP BW3.5 / SAP CRM4.0), CRM Marketing & Analytics SAP.
6. Cleveland, W. S. Visualizing Data. AT & Bell Laboratories, Murray Hill, New Jersey, 1993.
7. Grinstein, G., M. Trutschl, and U. Cvek. High-Dimensional Visualizations. Workshop on Visual Data Mining, ed. by Keim, A. and Eick, St., San Francisco, KDD-2001.
8. Iztueta, A. and P. Calvo. Comparing European Labour Force Survey Results from the Basque Country and Portugal, in Bock, H.-H. and Diday, E., Analysis of Symbolic Data. Springer, 2000, 374-382.
9. Wong, P. C. and R. D. Bergeron. 30 years of Multidimensional Multivariate Visualization. Proc. Workshop on Scientific Visualization. IEEE, 1995.

10. Wille, R. Knowledge Acquisition by Methods of Formal Concepts Analysis. Data Analysis, Learning Symbolic and Numeric Knowledge. Diday Edit. Nova Science Publishers, 1989.
11. Noirhomme-Fraiture, M., M. Rouard. Visualising and Editing Symbolic Objects. Hans-Hermann Bock, Edwin Diday (eds.): Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data. Springer Verlag, Heidelberg, ISBN 3-540-66619-2. 2000.
12. Noirhomme-Fraiture, M. Visualization of Large Data Sets: the Zoom Star Solution.– *Journal of Symbolic Data Analysis*, 1, July 2002, http://www.jsda.unina2.it.
13. Diday, E. Orders and Overlapping Clusters by Pyramids. In: Multidimensional Data Analysis (eds. J. De Leeuw, W. J. Heisen, J. J. Meulman and F. Critchley), DSWO Press, Leiden, 1986201-234.
14. Diday, E. Introduction a e'Approche Symbolique en Analyse des Donnees. Premiere Jouneles Symbolique-Numerique, CEREMADE, Universite Paris – Dauphine, 1987, 21-56.
15. Diday, E., R. Emilion, and Y. Hillali. Symbolic Data Analysis of Probabilistic Objects by Capacities and Credibilities. Atti Della XXXVIII Riunione Scientifia, 1996, 5-22.
16. Atanassov, K. Intuitionistic Fuzzy Sets. Springer, Heidelberg, 1999.

Manuscript received on 14.02.2008

Research Fellow Dr. Ivanka Valova *received M.Sc in the Faculty of Communication Techniques and Technology from the Technical University of Sofia in 1994 and PhD in field of OLAP technologies from the Institute of Computer and Communication Systems at the Bulgarian Academy of Sciences. Her research interests are in the field of computer and communication technologies*

Contacts:
Institute of Control and Systems Research
Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: vania@icsr.bas.bg



Prof. Monique Noirhomme-Fraiture *is „Licencée Mathèmatiques" from the University of Liège. She got her PhD at this University in 1973. She has taught Statistics, Probability Theory, Data Analysis, Stochastic Processes, Simulation and Performance System Measurement at the University of Namur since 1970. She has supervised many works and theses in the field of data analyses. She is the Belgian member of the IFIP Technical Committee.*

Contacts:
Institut d'Informatique, FUNDP,
Namur, Belgium