Measuring Performance in the Retail Industry
(Position Paper)

Gerasimos Marketos and Yannis Theodoridis

Department of Informatics, University of Piraeus,
80 Karaoli-Dimitriou St., GR-18534 Piraeus, Greece
{marketos, ytheod}@unipi.gr
http://isl.cs.unipi.gr/db

Abstract. Bearing in mind the changeable and complicated needs of business environment, in this paper we examine the necessity of evolution in the traditional decision support techniques. Our aim is to intensify the need for integrated performance measurement and management, as a way to ameliorate the existing tools for decision making, which are currently based on historical data. Because of the nature of challenges and trends in the retail industry, it is considered to be an appropriate application scenario. In addition to that, a framework is proposed and a case study is described as a proof of our claim.

Keywords: Performance management, business intelligence, retailing.

1 Introduction

Business Intelligence (BI) developed a few years ago as a set of applications and technologies for gathering, storing, analyzing, and providing access to corporate data to aid in decision making. BI includes, among others, decision support systems (DSS), statistical analysis, information visualization, data warehousing (DW) and online analytical processing (OLAP), and data mining (DM).

Turban and Aronson [16] argue that the decisions are taken at three levels: strategic, tactical and operational. The differences among them are related with the time scale that every decision demands and with the nature of them as well. The top management is responsible for the strategic planning of their organizations, middle managers make tactical decisions following the plans of top management and finally operational managers are responsible for the daily activity of the organization.

Obviously, performers at each level need different kind of information. The top management wants to see the "big picture" of the company situation. They usually prefer dashboards, consisting of Key Performance Indicators (KPIs), which show the trends of the organization. Middle managers want to have access in advanced, dynamic reports. They prefer aggregated instead of raw data, thus OLAP cubes and patterns extracted from data mining models look very useful for them. Operational managers need more real-time information. In fact, traditional BI can not serve them because it focuses on historical business data and thus it fits in better with strategic and tactical decision making (figure 1).

J. Eder, S. Dustdar et al. (Eds.): BPM 2006 Workshops, LNCS 4103, pp. 129 – 140, 2006.

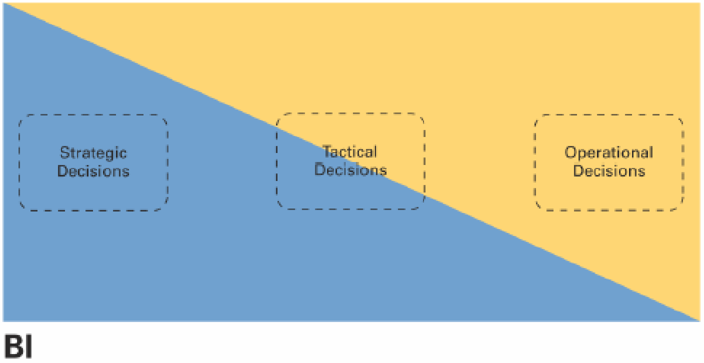

Fig. 1. Traditional BI coverage of decision making levels [15].

The above prove that modern organizations need something more than BI. Furthermore, the need for process-oriented organizations having efficient business processes that cut across organizational boundaries, raise the need for a more complete management of organizational performance. Focus on operational data is required because performance can not be measured only by trying to find patterns on historical business data. Strategic and tactical decisions are still critical, but without efficient operational decisions the real time and process oriented enterprise can not be realized.

The target is clear: decision makers, independently of level, should have the right information on the right time in order to serve efficiently and effectively the customer-centric processes in which they participate. This paper proposes a realization of the above target and its application in the retail industry.

The rest of the paper is organized as follows. In Section 2, we outline the challenges of the retail industry. In Section 3 we survey the proposed approaches for measuring and managing performance. Sections 4 and 5 present a framework for measuring performance on the retail industry and an application in a case study, respectively. Conclusions and hints for further work are drawn in Section 6.

2 Trends and Challenges in the Retail Industry

Retailing serves the selling of goods and services to consumers for personal or household consumption. A classification of the retail industry divisions can be found in [14]: groceries, apparel, electronics, drugstores, books/music, mail-order, mixed assortments and others. Retailers are at the end of the supply chain, which may consist of various suppliers, importers, manufacturers, wholesalers and distributors, and thus they interact directly with the consumer. To serve this purpose, the majority of technological advances are quickly applied in this sector so as to facilitate trade. Data management, supply chain management and marketing strategies, among others, are combined to this aim.

From a data management perspective, the emerging trends create many opportunities for delivering more value but they also bring problems that should be faced. Radio Frequency Identification (RFID) is a new challenging technology that is coming into sight, replacing traditional barcodes. Although the adoption of this technology has raised a lot of controversy, its importance has been recognized and thus a further discussion about RFID significance is beyond the scope of this paper. Furthermore, we choose to focus on how data produced by RFID tags can be transformed into knowledge, and not on the management of the huge volume of data being researched by data streams community. What RFID can give us is:

- **a. Sequence of purchase:** It is possible to know in which order people buy things. In fact, we know the exact time of putting an item in the basket. Extracting such patterns, retailers may decide, for instance, to change the position of some items in the store in order to facilitate (or not) people in the store.
- **b. Positive/ negative preferences:** It is possible to have answers on questions such as: Are there customers that, after taking an item, change their mind and put it back on the shelf? Is there a specific pattern behind this behaviour? How much time do customers need to decide about the selection or not of a product?
- **c. Routes of customers:** By placing RFID labels on the baskets, it is possible to track the movement of customers inside the store. Thus, by placing an Indoor Positioning System (IPS), customers could be informed based on their interests and their location.

As far as *supply management* is concerned, it gives retailers a competitive advantage. The collected shopping data can be transformed to valuable information and shared throughout retailers' supply chain networks of suppliers, factories and distribution centers to predict trends on product demand for controlling inventories and stocks.

For marketing purposes, *personalized and real time offers* is a critical tool to realize the necessary customer-centricity. Customers should feel that retailers know and meet efficiently their needs. RFID technology and IPSs can provide the necessary infrastructure for collecting data and providing useful information. The issue here is how to interpret the raw data to shopping information that is valuable for each customer. Several papers have been proposed for predicting shopping lists [6] and building shopping assistants [5] but they do not take into consideration the latest technological advances and they are only based on analyzing historical data. Prediction of shopping lists can be considered as a special case of recommendation techniques which are overviewed in [2].

The above challenges show that retailers need something more than BI for supporting strategic and tactical decisions. A more complete solution is required in the retail environment in order for performance management to integrate business processes and historical data.

3 Performance Measurement and Management

Performance management is a challenging issue due to three core reasons [11]: (a) goals and objectives against which we measure companies' performance are exponentially increasing, (b) external unstructured data and events have to be encompassed and, finally, (c) acting in a timely and effective manner on the resulted imperatives is required.

The recent years, several researchers have presented their suggestions about BI evolution in order to serve performance measurement and management. We present them in order to show their common characteristics and find the set of operations that best fit in the retail industry.

In [3], performance management is concerned in terms of process execution monitoring and analysis. Authors consider that simple reports off the process execution database and OLAP-style analysis are not adequate. Business Process Intelligence is proposed as a way to explain process behavior and to predict problems in process executions by applying "process mining" algorithms. An overview of issues and approaches on workflow mining can be found in [1].

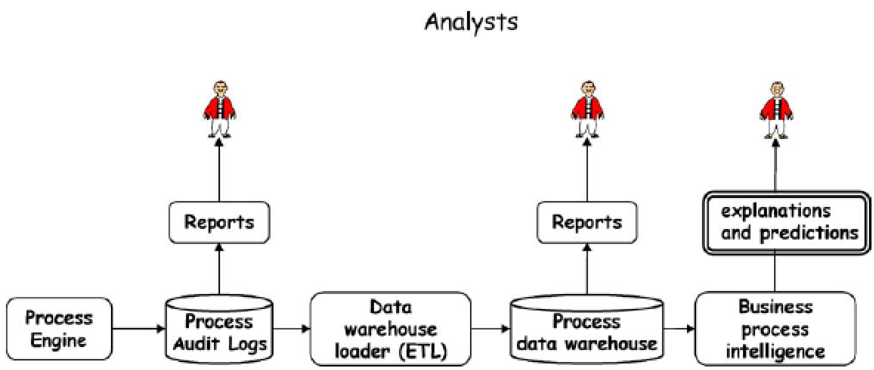

Fig. 2. Different approaches to business process analysis and management [3]

In [13], a Corporate Performance Measurement System is proposed by integrating business process performance information into a traditional data warehouse. The DW is built using operational data coming from the workflow log which provides very detailed information on the history of process instances. In fact, it is the same approach as the process data warehouse to appear in figure 2.

In [9], Business Performance Management is considered as a set of processes for optimizing performance by encouraging process effectiveness as well as efficient use of financial, human, and material resources. The main idea behind this proposal is that DW is not enough to this end since its technology is neither suitable for the grain nor for the freshness of the collected information, that should quickly flow throughout the different levels of the company.

Operational BI, Enterprise Decision Management [15], Business Activity Monitoring and Business Operations Management are other usually mentioned terms to describe the ideas presented in this section. In table 1, the differences from traditional BI are referred.

In the section that follows, we propose a framework as an application of the above mentioned and not a different approach. We try to intensify the need to evolve the traditional systems so as to satisfy the emerging needs in business environment.

Table 1. Traditional BI versus the performance management approach [15]

	Traditional BI	**The performance management approach**
Focus	Improve strategy development through insight into trends and performance	Improve strategy execution through automating decisions
Activity	After transaction	During transaction
Key methodologies	Data analysis, OLAP, reporting and query tools, data warehousing	Traditional methodologies plus KPIs, dashboards, business rules engines
Workflow	Offline, disconnected from business processes	Embedded in operational processes and systems
Analytics	Summarize past performance, group behavior, trends	Continuously measuring and managing performance

4 A Framework for the Retail Industry

In this section, we present the architecture of a framework for measuring and managing performance in the retail industry (figure 3). Combining traditional BI techniques with the technologies presented in section 3, we can have a complete solution for dealing with the challenges and trends outlined in section 2.

Our framework consists of a number of modules. In the following paragraphs, a reasonable sequence of the stages of the proposed framework is described, from the raw data to the final output. In particular:

a. **Source data:** Apart from shopping data and workflow logs, other data streams can be also input in the system. For instance, data collected from RFID tags include useful information that should be analyzed, although it is not necessary to be archived in the operational database. For a survey in data stream management see [8].
b. **Integration manager:** The role of this module is to manage the above heterogeneous data sources and to feed the appropriate analytics. Likewise, it guarantees that the feeding process happens on the right time for each analytic: ETL tools can be fed once a day as OLAP-style analysis focus on historical data while Activity Monitoring components need real time data.
c. **OLAP Cubes:** ETL tools transform raw data into aggregated information providing thus data warehousing capabilities. Instead of providing only OLAP-style analysis on shopping data (business data warehouse), the proposed architecture includes data warehouses for both business and process execution data.
d. **Activity monitoring:** This module deals with real time information. It updates and controls KPIs and triggers Business Rules Manager for verifying that corporate rules are satisfied. KPIs can be also verified for satisfying predefined Business Rules (BRs). Most Business Process Management (BPM) suites support process monitoring.

e. Data mining engine: DM engine consists of a set of algorithms and techniques for identifying patterns on data. Customer segmentation, correlations between products and prediction of product demand are typical tasks that can be applied on shopping data. We consider as important to include special process mining algorithms that are applied on workflow logs, for predicting critical situations and discovering interesting correlations. Applying mining techniques on real time information (sequence of purchase, routes of purchase) is also a challenging issue and an active research area [7].

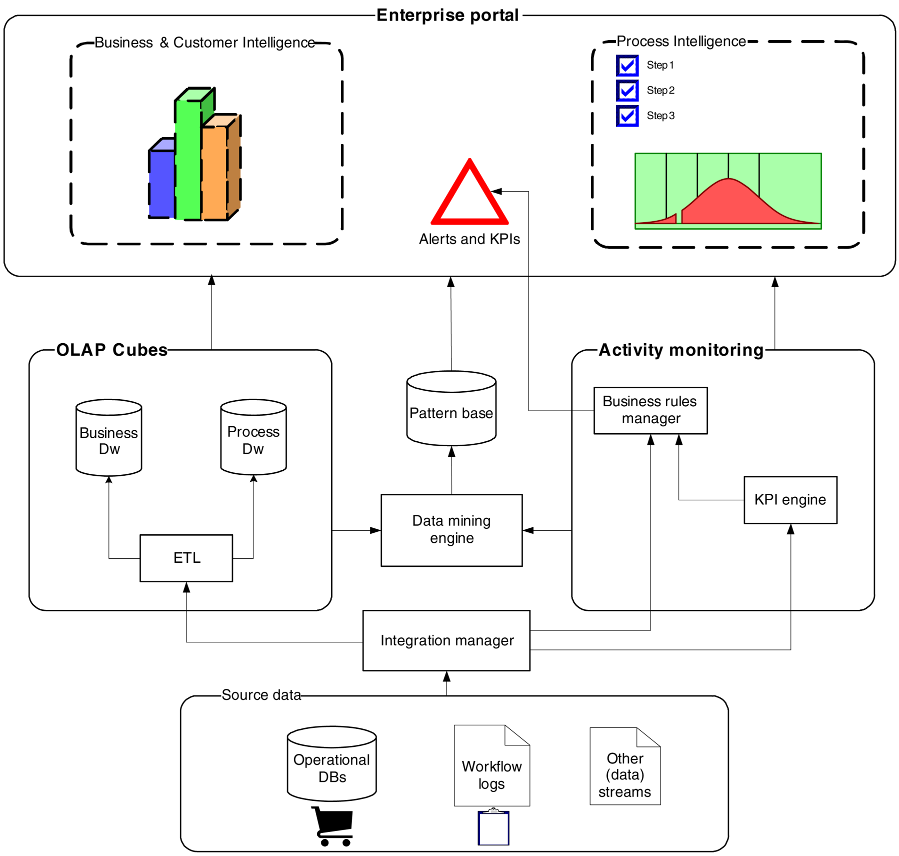

Fig. 3. The proposed architecture

f. Pattern base: A Pattern Base Management System (PBMS) provides pattern management functionality as a Database Management System (DBMS) does for data. Patterns are extracted from various data sources applying the data mining algorithms and techniques included in the data mining engine. Our framework proposes the integration of the pattern base with the data warehouses and the operational databases. Thus, data can be viewed from three quite different but useful perspectives: raw data (operational databases), aggregated data (data warehouses), and analyzed data-extracted knowledge (pattern base). Pattern management is an active research area in which various approaches have been introduced [4], [12].

g. **Enterprise portal:** This is the output (portal) of the proposed system. It includes a role-based architecture which provides users in various positions with the appropriate information. For instance, operational managers may be interested in process intelligence reports while tactical decision makers may find useful the business and customer intelligence reports. Generally speaking, users will be able to build their own dashboards by subscribing to the services the role based systems has allocated to them.

5 Applying Our Framework: A Case Study

In the following paragraphs, we outline the application of our framework addressing design issues. In particular, we present the data sources involved, the data warehouses (cubes) built for OLAP purposes, and some indicative KPIs and BRs materialized to address top management needs.

5.1 Data Sources

The operational database includes tables for storing customer demographics and details about their transactions. Each transaction takes place in a specific store and embodies details of the products that were purchased by the customer. The system may also include details about the supplier of each product. A sample diagram is depicted in figure 4.

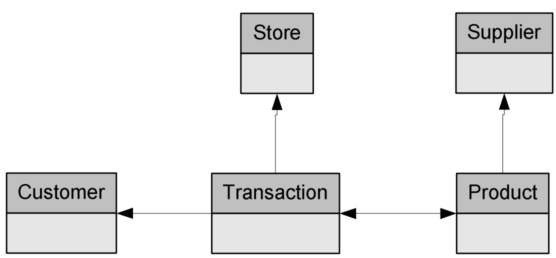

Fig. 4. A sample E-R diagram for operational data

In figure 4, the relationships between the core entities of the system are represented. As it is shown, a transaction takes place in a specific store and a single customer participates in it. Namely, a transaction represents a customer's basket and may consist of one or more products (purchased in the same basket). Moreover, the supplier of each product (and other information regarding supplying processes) is recorded.

As far as the workflow logs are concerned, the simplest format is presented in table 2 where only the execution time of each activity (step) is recorded. Workflow logs are tightly related both to the workflow engine that is used to manage business processes and to the nature of them. Typical business processes in the retail industry are orders to suppliers and distribution centers.

Table 2. A sample workflow log

Process_Instance	Activity	Timestamp
Order supplies of product A	Send order	1/4/2006 21:05:14
Order supplies of product A	Receive products	3/4/2006 10:15:04
Order supplies of product A	Inventory products	3/4/2006 12:55:09

5.2 OLAP Cubes

We have proposed the coexistence of business and process data warehouse for providing OLAP-style analysis to business and process data. In figure 5, we present the dimensions of a business data warehouse. Apart from geography, product category and time dimensions, we have included a number of customer hierarchies (age, sex, profession, marital status, education level and number of children). Measures are related to quantities and values of products. Figure 5 presents only some of them and takes into consideration the case of private label products that the retailer may sell.

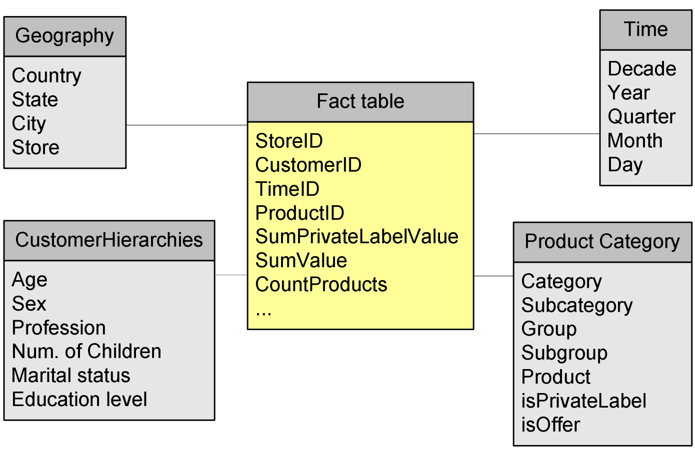

Fig. 5. A sample data warehouse for business data

Furthermore, in figure 6, a general case process warehouse is presented. The data warehouse depends on retailer's organizational structure and needs, and thus the proposed schema cannot be suitable at every case. We consider that processes and activities can be classified according to their types. For instance, both human-triggered and automatic activities should be considered. Geographical hierarchy provides analysis capabilities in many different levels. Suppliers dimension classifies them to various types according to the nature of products they provide.

The defined performance measures should help in analyzing raw process execution data in a multidimensional way. Interesting measures are the total execution time and the time lags of each process.

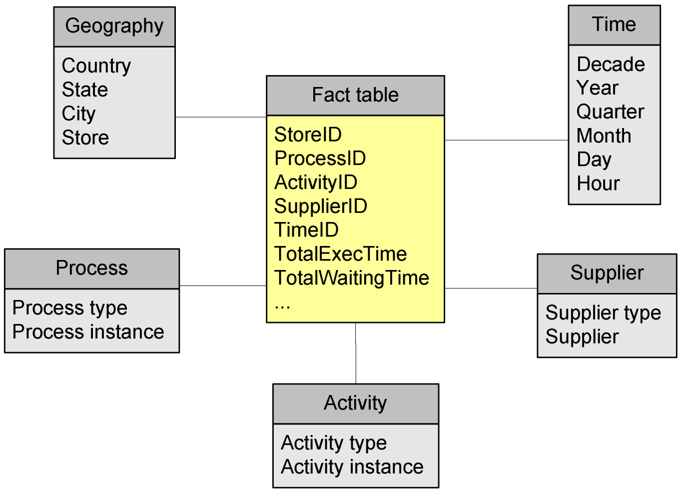

Fig. 6. A sample data warehouse for process data

In both cubes classic OLAP operations can be applied. We illustrate the benefits obtained by such an approach with two examples of operations supported by our proposed business and process data warehouses and OLAP technologies:

a. A user may ask to view the average value of private label products in baskets in the past quarter, over Athens, and, moreover, he/she can easily view the same information for specific stores in Athens (more detailed view, formally a *drill-down* operation) or over Greece (more summarized view, formally a *roll-up* operation).

b. Similarly, the total waiting time for processes of a specific process type, over Athens, in the past quarter may be requested. The user can also easily view the same information both for each month of the selected quarter and also for the whole year.

Further to roll-up and drill-down operations described above, typical data cube operations include *slice* and *dice*, for selecting parts of a data cube by imposing conditions on a single or multiple cube dimensions, respectively and *pivot*, which provides the user with alternative presentations of the cube.

5.3 Activity Monitoring

As already mentioned, BPM suites support process monitoring consisting of BRs and measured by KPIs. We argue that this can be generalized to include monitoring of business related activities. Tables 3 and 4 present some interesting KPIs and BRs that could be continuously calculated and checked respectively. Violation of BRs could trigger alerts in the enterprise portal for assuring human attention.

Table 3. Sample Key Performance Indicators

KPI	Target value
Average waiting time for check out	5 sec. per product
Average value of customer baskets	20 € (weekdays) 50 € (weekends)
Average value of private-label products in customer baskets	30%
Average weekly sales per state	500K €

Table 4. Sample Business Rules

BR	Expression
Waiting time for check out	≤ 20 min.
Execution time for process type: "Order dairy products"	≤ 1 day

KPIs provide instant and continuous view of the running activities. However, the target values are based on past experience. In other words, the definition of target values for complex KPIs can be assisted by using multidimensional analysis (OLAP) and the aggregated information stored in the two data warehouses (business and process cubes). Furthermore, it is useful to define BRs on KPIs ensuring alerting if target values are violated.

5.4 Pattern Base

The execution of data mining models results in patterns that are stored in a PBMS. Pattern management issues such us update, merging, comparing, querying, evaluation etc. are active research problems but a detailed discussion about them is beyond the scope of this paper. Frequent itemsets (FIs), association rules and clusters extracted from warehouse data are typical examples of patterns.

FIs represent sets of products that are often purchased together (e.g. {milk, bread, frozen food}) while association rules introduce correlation between items (e.g. {bread, frozen food} ⇒ {milk}). Quite more interesting is the discovery of the temporal sequence of purchasing in such itemsets (e.g. milk, then bread, then frozen food). On the other hand, clustering algorithms could result in clusters showing correlations between demographics and shopping preferences. For example a profile of shoppers (women, aged 30-35) appears to have a specific preference (buy dairy twice a week) and so on.

We illustrate the benefits obtained by such an approach with two examples of queries supported by the pattern base:

a. A user may request to view demographic details about a specific cluster of customers, for instance: costumers who visit any store once a week and buy private label products valued 0-5 euros, and the value of their baskets in total is between 20-30 euros.

b. A user may have the information that milk is in high demand in a specific store and he/she requires knowing what products are purchased supplementary to milk in order to check their supplies. The proposed pattern base can provide a list of these products which are associated with milk, and moreover, to give a hierarchical order of their association with milk combined with information about stores supplies.

6 Conclusions – Further Work

Taking into account current challenges and trends in the demanding and complicated area of the retail industry, it is evident that decision making analysis should be based on real time information and not only on historical data, as the traditional methods have used so far. In this paper, we propose a framework for extending traditional BI to an integrated environment for measuring and managing performance. As it was highlighted in [10] the importance of measurement in controlling, managing and improving the processes is vital. The framework consists of several modules that enable both business and process intelligence capabilities. It includes tools so as to give the proper information on the right time to each decision making level.

Our research is at early stage; future steps include the evaluation and incorporation of process mining algorithms in the data mining engine and the development of a prototype following the proposed architecture. The application of our framework in other industries could also be a task for future work.

Acknowledgements. Research supported by the General Secretariat for Research and Technology of the Greek Ministry of Development under a PENED'2003 grant.

References

[1] van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M.: Workflow mining: a survey of issues and approaches, Data & Knowledge Engineering 47 (2003) 237-267

[2] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin A.: Incorporating contextual information in recommender systems using a multidimensional approach. ACM Transactions on Information Systems 23 (2005) 103–145

[3] Castellanos, M., Casati, F., Dayal, U., Shan, M.C.: A Comprehensive and Automated Approach to Intelligent Business Processes Execution Analysis. Int. J. Distributed and Parallel Databases 16 (2004) 239-273

[4] Catania, B., Maddalena, A.: Flexible Pattern Management within PSYCHO. Proc. PaRMa'06, Munich, Germany (2006)

[5] Cumby, C., Fano, A., Ghani, R., Krema, M.: Building Intelligent Shopping Assistants Using Individual Consumer Models. Proc. IUI'05, San Diego, CA, USA (2005)

[6] Cumby, C., Fano, A., Ghani, R., Krema, M.: Predicting Customer Shopping Lists from Point of Sale Purchase Data. Proc. KDD '04, Seattle, WA, USA (2004)
[7] Gaber, M.M., Zaslavsky, A., Krishnaswamy, S.: Mining Data Streams: a review. ACM SIGMOD Record 34 (2005) 18-26
[8] Golab, L., Ozsu, M.T.: Issues in Data Stream Management. ACM SIGMOD Record 32 (2003) 5-14
[9] Golfarelli, M., Rizzi, S., Cella, I.: Beyond Data Warehousing: What's next in Business Intelligence? Proc. DOLAP'04, Washington, DC, USA (2004)
[10] Harrington, J.H.: Business Process Improvement – The breakthrough strategy for total quality, productivity, and competitiveness. McGraw-Hill, New York, USA (1991)
[11] Keziere, R.: Are we there yet? Three challenges for BPM. Cutter IT Journal 18 (2005)
[12] Kotsifakos, E., Ntoutsi, I., Theodoridis, Y.: Database Support for Data Mining Patterns. Proc. PCI'05, Volos, Greece (2005)
[13] List, B., Machaczek, K.: Towards a Corporate Performance Measurement System. Proc. ACM SAC'04, Nicosia, Cyprus (2004)
[14] Madlberger, M.: Strategies and Business Models in Electronic Retailing: Indications from the U.S. and the UK. Proc. ICEC'04, Delft, Netherlands (2004)
[15] Taylor, J.: Beyond BI: Building intelligence into your operational decisions. Fair Isaac white paper (2005)
[16] Turban E., Aronson J.E.: Decision Support Systems and Intelligent Systems. Prentice Hall (1998)