RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS No. 104
Business Informatics 16 2010

Małgorzata Nycz, Mieczysław L. Owoc, Maciej Pondel
Wrocław University of Economics, Wrocław, Poland
{malgorzata.nycz,mieczyslaw.owoc,maciej.pondel}@ue.wroc.pl

BUSINESS INTELLIGENCE CONCEPTS FOR EDUCATION QUALITY MANAGEMENT

Abstract: On the one hand, there is a lack of applications within the basic areas of business intelligence applications, especially in education sector. On the other, we observe the real need of permanent improvement the education quality in modern universities. The main goal of the paper is to present our concept for improving the education quality in higher education using modern IT. To obtain this we have carried out analysis of the BI usability focused on education quality at the university level. In the context of managerial decision support, we have shown some BI features and then presented the model of data warehouse for this purpose. Next the discussion on exploration techniques in the Oracle environment as well as its adequacy has been presented. Short summary ends the paper.

Keywords: business intelligence, education quality management, data mining.

1. Introduction

Academic institutions face market challenges as well as gradual transformation of education styles and contents. One of the key factors essential for the quality of teaching improvement process is gathering data about different aspects of education from students' or teachers' points of view and collecting opinions of the scene actors. Undoubtedly, the management of quality education can be recognized as the main task assuring effective teaching and learning but it should be underlined that there is the necessity to apply modern technologies to support it. Unfortunately, there are few papers referring to this process, for example: [Cheng 2001; Christobal, Ventura, Garcia 2008; Van Dyke 2008].

The main goal of the paper is to analyze the usability of BI approach in the education quality management process. The paper consists of three sections apart from this introduction and final conclusions. The first section is devoted to the presentation of the general assumptions and concepts of business intelligence oriented on education sector. In the next one the idea of Data Warehouse model supporting the defined concepts is exposed. The last part of the paper includes discussion about data mining techniques potentially applied in the management of education quality.

2. Traditional model of education versus distance learning

In the traditional education, students and teachers are at the same place (e.g. classroom) and at the same time (e.g. lesson). Students listen to the teacher when he/she carries on the lesson/lecture. Teacher is the main source of knowledge delivered to students. He can be supported by such means as e.g. audio/video materials, slides or films. When digesting the didactic material students can use additional means such as manuals, books, etc. Teacher is the one who makes the assessments of students. The quality of education process depends on the teacher's quality, his knowledge, his personal character features and his ability to teach others.

In the distance learning system the situation looks different. The point of gravity has been moved from a teacher to the system. The main source of knowledge is not a teacher but the knowledge bases collected in the system as well as other bases and sources of information accessible mostly via Internet. Teacher is rather an instructor for students, theirs buddy but not a teacher in traditional model. He can be treated by students as an additional source of knowledge. Assessment of students is performed by both teacher and system, but the teacher as a person is responsible, in our opinion, for final results. The main differences between models of education has been shown in Table 1.

Table 1. Traditional versus distance learning model

Model / Feature	Traditional model of education	Distance education system
Main knowledge source	teacher	knowledge bases in education system, any knowledge sources accessed via Internet
Additional knowledge sources	books, manuals, audio and video materials	traditional sources, teacher
Assessment	only by teacher	system and teacher who is responsible for final assessment
Quality of education	depends on teacher's quality, his level of knowledge, his ability to share his knowledge	depends on electronic knowledge sources quality and other didactic materials

Building of knowledge in student's mind (learning) is a process of problem solving that makes student to be active, innovative and to develop his/her experiences. Using distance learning system student can be active in building knowledge within the following cooperation forms:

– common learning within the team,
– interactive process of group building of knowledge,

– active participation in generation and selection of information,
– knowledge construction in context of other students points of view.

The teacher's task is to be a supervisor of learning process and to monitor the progress in learning [Bielecki 2008].

In distance education system knowledge resources are of very various form but mainly they are in form of didactic modules. They allow student to learn particular portion of didactic material. Modules have modular structure. Passing to the next module may be dependent (or not) on whether he/she has positively completed the previous one. The same is with courses. Usually a student can access three types of modules:

– modules covering material of particular subject/course,
– tests and exercises modules, and
– help modules.

The didactic process can be organized in various ways. The most common one is based on didactic path. Student can be directed on a given path after, e.g., enter test. Such a test is mostly built in shape of decision tree. Recognition of student's skills and knowledge begins with simple questions, through questions based on material from the next modules. Questions with bad answers can be used when deciding about material to be taught. Directing on paths may also occur when a student has achieved poor results of tests after modules. In such a case student either has to return to a given module or some extra explanations in form of help have to be presented to him.

Within the repetition of the module, student can be asked either to repeat the whole module from the beginning or to go through additional explanations and maybe to repeat more difficult portions of material.

Realization of auxiliary, optional material covers very large range of tasks: extra tasks to be solved, manuals and other books from a given discipline, possibility to ask by e-mail colleagues or teacher/instructor for help.

The realization of help materials is possible when a student either chooses the help option or he has unsuccessfully tried several times to solve a test. In such a situation, the control – after inferring with what (which problems) he has difficulties – suggests additional explanations, exercises or maybe reading appropriate books. Realization of help option is under special supervising due to the fact that maybe not all suggested exercises have to be done. In that case student obtains the suggestion to stop the help mode and return to "normal" work. Another possibility of leaving the help mode is the student' decision to stop it [Bielecki 2008].

Obtained results are collected in the system. To enter the next course, the student has to pass the final test of the previous one. However, sometimes the system enables the student to enter a new course without requesting to pass final test of previous course. This situation takes place when the student wants only to read the material covered by the course or to check his knowledge from the course but not to be assessed by the teacher/system. What should be underlined here is that the final assessment of the student is the teacher's task. The teacher can be supported by the

system of course. This support can be realized in many different ways: either using simple mechanisms delivered by the system or using specialized external tools that carry out all information about student's progress in education process and generate for the teacher's use a piece of advice about the final assessment.

3. Knowledge as one of the most valuable assets in distance learning

When commonly speaking, knowledge is necessary in everyday life. But it is worth attention to use this term in such a sense as artificial intelligence does. Generally knowledge is connected with erudition or with data assets. What could happen if it would be unnecessary to implement the term synonymous with "data base"? But there is another aspect of knowledge: knowledge as a tool that enables drawing conclusions from premises. So the knowledge can be defined as a set of information that enables drawing conclusions from premises [Nycz 2007]. Premises can be a situation description, set of facts, input conditions in dynamic model or other sets of information. A set of information being knowledge can have various forms. This way we came to difference between knowledge base and database. Database is a set of descriptions of facts made by using a data model. Knowledge base is database plus rules of inferring from data.

In traditional model of education, teacher is this "element" of the system which possesses knowledge, and delivers it to his students during lectures or exercises. Students, except for the teacher's knowledge, can use additional knowledge sources as, e.g., manuals or audio and video materials. The teacher evaluates how well the student has learnt this material in a given time and makes assessments giving appropriate marks. The quality of education depends on the quality and level of the teacher knowledge as well as on his ability to share knowledge with students.

In the distance education system, the gravity point has been moved towards the accessibility of knowledge sources via Internet, and their quality. The teacher plays a role as the buddy who supervises the learning process and assesses the results rather than the main knowledge source. In this situation the role and importance of knowledge gathered in the system becomes one of the most important factors in distance education.

Knowledge is collected in knowledge bases. It is of different nature and can be organized in various ways. Knowledge management in the education system can be understood as in any other system and covers such areas as knowledge acquisition, its maintenance, accessing, updating and verifying. Knowledge should be actual, complete, certain, consistent as far as it is possible, etc. Knowledge comes from different sources. Among them one can distinguish: didactic courses and modules prepared by teachers, bibliography, books and manuals accessible in both traditional and electronic ways, didactic materials in form of audio and video, exchange of knowledge between students and/or institutions, etc. Knowledge covered by courses or other didactic units may quickly become outdated. The problem of knowledge

updating and acquisition of new knowledge is very difficult and complicated. The methods of automated acquisition as well as methods of data mining are used when solving this problem. New knowledge must be verified either by teacher or by special verifying procedures.

4. Architecture of distance learning system (DLS)

As a distance learning management system we understand such a system that allows creation, storing, updating and management of didactic material used within the education process [Hayhoe 2002]. It realizes the approach that enables multi-use of didactic units or knowledge, offers collected modules and courses and should fit

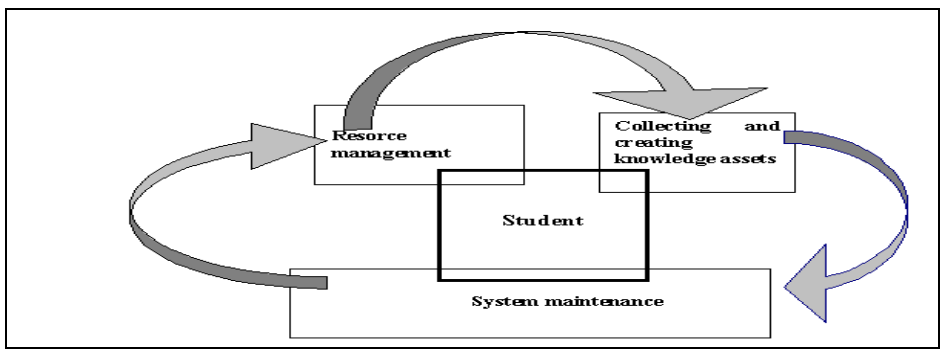

Figure 1. Idea of distance learning management system

Source: based on [Hayhoe 2002].

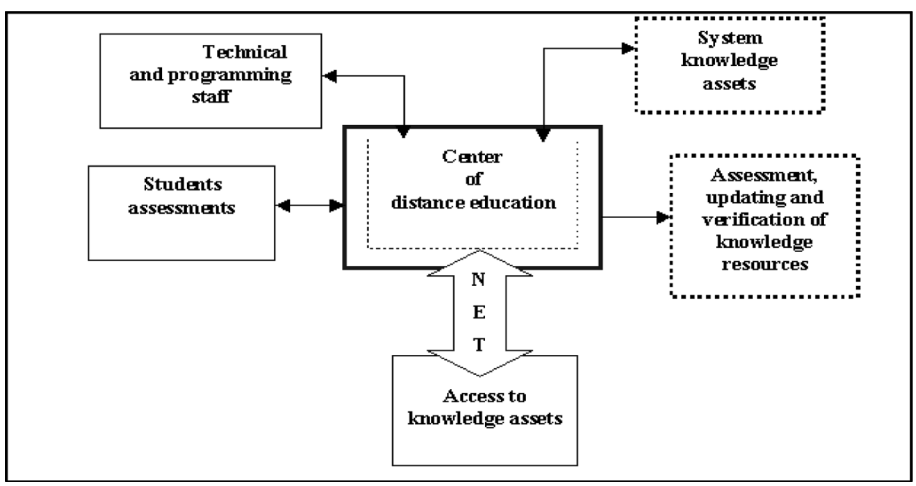

Figure 2. Organization of distance learning

Source: [Hayhoe 2002].

international standards [Peraya 2003; Singh 2002]. The education system has been required:

1) to collect only positively assessed didactic material,

2) to enable building new course/units by teachers without programming by them, only using sheets accessible in the system.

Database should be opened to be flexible and both teachers and students should accept formats of presented material [Hayhoe 2002]. The main idea of the distance learning management system is presented in Figure 1.

From the student's point of view it is not important how the education environment has been organized. He is interested only in interactive access to large resources of knowledge necessary within learning. In Figure 2 we present how such an environment can look like.

5. Business Intelligence essence and its specialty in the education sector

The term Business Intelligence (BI) has many descriptions or definitions. According to one of them, BI can be treated as a set of concepts and methods that help with improving the decision process and it is done by using the fact-based support systems. We can say that BI systems are the way of obtaining the business information and its verification. From the perspective of information systems, BI is seen as a system which – using the OLAP (On Line Analytical Processing) technology and data analyses – delivers to the manager answers on important business queries and identification of important trends and patterns. Within Business Intelligence we can exploit the data collected in various information systems regardless of the fact that they were in use in past or they are in use today.

As we see, BI is the center of attention within all information solutions in organization and can be treated as a "binder" as well as a possibility of enterprise improvement. There are some requirements of BI, for example: integration of data coming from heterogeneous sources into one consistent data store, delivery of interactive possibilities for data manipulation, offering different ways of data presentation (graphics, tables, standard and ad hoc reports, etc.), and simplicity when using.

In order to achieve the defined aims, Business Intelligence requires specialized tools. Main components of the BI infrastructure are presented in Figure 3.

More generally, principal tools for BI include ***data warehouses*** (DW), ***OLAP*** and ***data mining*** (DM) apart of data querying and reporting the results.

The *data warehouse* (as the basic element of any BI system) technology plays the fundamental role in gathering source data and storing them in multidimensional storages enabling further data querying, mining and reporting. There are specialized models applied here which allow for storing data aggregates for many purposes.

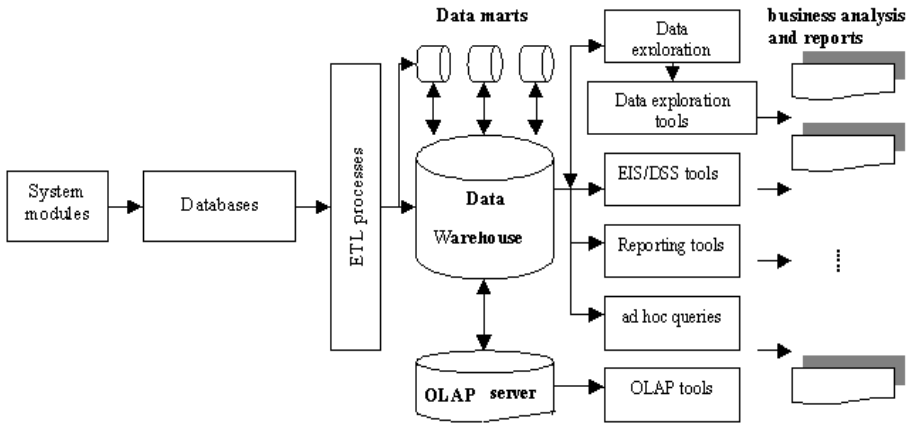

Figure 3. The general BI model

Source: based on [Business Intelligence… 2003].

An *OLAP* technology allows for quick and intuitive analysis of large amount of data adopted to user's needs in various sections, on various levels of details. OLAP can be implemented in different data models: relational (ROLAP – relational OLAP), multidimensional (MOLAP – multidimensional OLAP) or in HOLAP (hybrid OLAP). Among many techniques of multidimensional data analysis we can distinguish: drill-down, drill-up, slice-and-dice, drill-through, pivoting, rotating, ranking and aggregation.

Data mining technology reflects the process of extracting useful patterns from data resources. Therefore DM is becoming an increasingly important tool (rather discovery driven) to transform these data into information or better into knowledge pieces. There are many approaches and algorithms applied in DM (for example: classification, regression or clustering) that enable for delivering of patterns and rules.

The analysis of data allows better understanding of the phenomena that are in focus of interest in a university. It also allows to search out and to assess the relations among investigated phenomena (e.g. quality of education and university infrastructure), and to identify critical factors affecting these phenomena.

The in-depth analysis results can increase the effectiveness of undertaken activities, improve effects of the projects in progress and allow planning these activities and projects in more effective way. Basically, all these activities should be oriented on improvement of teaching quality. In fact all mentioned BI components and tools can be used in universities. The analysis of historical data collected for reporting purposes can be the basis for changes in educational systems in future.

6. A concept of data warehouse model supporting education quality improvement

A data warehouse is an advanced database that stores current and historical data from several operational databases aimed at future OLAP processing and DM. The general model for different purposes and users consists of a fact table (includes measures and keys to dimensions) and some dimension tables (including their keys and data). Particular tables are connected by keys.

There are several models elaborated for storing data warehouse schema. The star structure is a basic model of data in DW and consists of fact and dimension tables mentioned before. The structure of snowflake model is an improved star model. The most complex data structure is called a hybrid one.

The basic logical structure in any DW is the star model. It consists of one normalized table of facts and some denormalized tables of dimensions. There are the main key being the concatenation of main keys of each dimension table as well as some numerical measures in the fact table. The exemplary star structure for our purposes can be as it is shown in Figure 4.

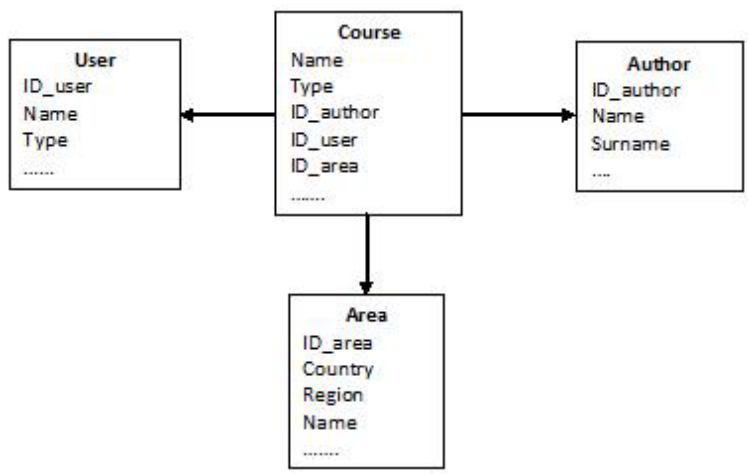

Figure 4. A star structure in the exemplary didactic data warehouse

The interpretation of this structure can be as follows. The course is the fact table. It covers a given didactic material and can be seen in three dimensions: the course author (mainly teacher), user (student/pupil) who wants either to possess knowledge from this course or to be familiar with the didactic material, area from which the student wants to be familiar with this course (e.g. country, university, enterprise, etc.)

The final model of a DW should be developed regarding main functions of the projects.

1. *Monitoring and controlling of student education*. The currently available standard data refer to student evaluation records (including student grouping), scholarships received by students (treated as educational motivation) and the organization of education (assignments to groups and majors as basic educational objects). This range of information should be extended to a synthesis of the educational process including work and diploma examinations and related directly to some model of graduates and their further professional careers.

2. *Monitoring and controlling the educational process from departments performing teaching*. The point of interest are university units (departments, institutes and faculties) teaching different courses. Current available data show the standard plans and reckoning of teaching activities for particular organizational units. The same teaching contents as well as learning hours and approaches are closely correlated with the previous area. As the valuable information can be treated the information about the activities of individual employees as well as their research topics correlated with the didactic process.

3. *Synthetic evaluation of quality education management based on decision-making areas previously defined*. In this area standard data are very limited because of the imprecise quality measures used (one of a few measures that are carried out concerning the teaching efficiency). The both previously listed decision-making areas are included in this group of the concepts.

In the proposed solution closely related data models for each of the listed areas will be elaborated in concordance to the data warehouse design rules. Basic aggregated data should include the following thematic areas:

- ***education of students*** – data represented by the following groups: assessments, reviews, achievements and scholarships. The corresponding data structuration should make it possible to make analysis with respect to time, generic and teaching subject aspects,
- ***didactics of teachers*** – data depicted by the following groups: assessment, reviews, educational plans and reckoning. Preparation of appropriate structures for analytical processing provides an opportunity for research on teaching process including time, generic, organizational groupings aspects in which the employee works and the courses orientation,
- ***domain education*** – expressed by data collection covering assessment, opinions and structural elements of delivered knowledge. Accordingly organized for Business Intelligence data groups should enable the analysis taking into account timescales, courses and majors orientation.

Two initial models of DW quality education evaluation are presented in Figures 5 and 6.

The first model is student-oriented, therefore all necessary data about student grading should be stored in this table of facts. Considered as potential dimensions (analysis directions) are: time (semesters or years), courses (their contents and way

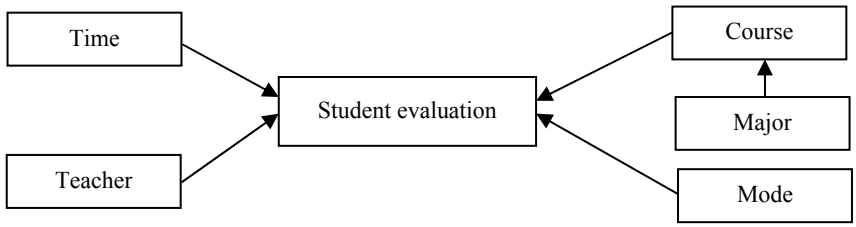

Figure 5. The data warehouse model for student evaluation

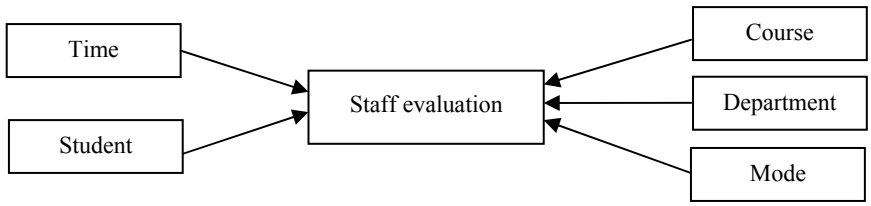

Figure 6. The data warehouse model for staff evaluation

of teaching), staff (representing different science disciplines) and modes (to include localization and forms of studying).

In the second staff-oriented model we focus on multidimensional analysis of teachers. Basically main dimensions are similarly formulated (time, mode and course). In this model student and department dimensions have been inserted.

The data gathered in these two models allow for more advanced analysis including also crossing analysis, for example course teaching (student and staff records) or diploma topics (major and departmental orientation).

7. The overview of data mining techniques useful in improving the quality of education managerial decisions

One of the most common IT tools for Data Mining is provided by the Oracle Corporation. Its name is Oracle Data Mining (ODM) and it is an option of Oracle Database Enterprise Edition starting from version 9i release 2. It is developed in every new version of Oracle Database till current version 11g Release 2 [Wikipedia 2009].

Oracle Data Mining contains such techniques as [Oracle 2009; Morzy 2005]:

- Classification,
- Regression,
- Attribute Importance,
- Anomaly Detection,
- Clustering,

– Association,
– Feature Extraction.

Classification is one of the most popular data mining techniques. Here we build a model that assigns a record to the one class that is prepared earlier. The model can be created including the experience gathered during analysis on learning and testing data sets. Data sets are the sets of records describing real object already classified. In ODM the Classification can be executed using such algorithms as:

– Logistic Regression,
– Naive Bayes,
– Support Vector Machine,
– Decision Tree.

Classification can be the basic technique supporting the process of education quality management. Basic objects taking part in a business process of education can be classified as improving education quality, spoiling it or not influencing the quality of education. Those objects can be for example: teachers, subjects, or classrooms. Introducing new elements into the learning process classification model we can predict its influence on education quality.

Classification can be also very useful in helping students making decision concerning their education [Vialardi et al. 2009]. In the education process student has to choose the field of study, the level of some subjects or even the subjects that are optional. If we had the database of the students and their decisions classified as right and wrong, we could prepare the classifier supporting new students in their choices.

We can also use classification to predict the students drop out [Dekker, Pechenizkiy, Vleeshouwers 2009]. Having the database describing the history of students who have resigned and knowing the reason of drop out we can predict if a new student is going to drop out. Students' drop out is a very vital problem especially in the first year of the study. Knowing well the mechanism of dropping out we can help students to choose their field of study better or show them on what areas they have to work harder to complete the study.

Attribute Importance is also called feature selection. It is a supervised DM technique that ranks the significance of every attribute describing business object according to the target value. In Oracle Data Mining attribute importance is implemented by Minimum Description Length algorithm. It is probably the most useful data mining technique for improving the quality of education. Having the database of objects active in education process and having them assigned to the classes representing their influence on the quality of education, Attribute Importance technique will examine every attribute's influence on education quality. In this case people making decisions have the information what features in business process have to be improved to increase the quality of the whole education.

Anomaly Detection is an unsupervised data mining technique that allows detecting rare cases in the data. In ODM anomaly detection is done by the algorithm called One-Class Support Vector Machine. Normally anomaly detection is used in

detecting frauds and some other crimes. In education it can be useful for identifying unusual cases of students that have some extraordinary skills to help them to develop their abilities. This technique can be also useful for detection of different kinds of abuse or cheating that should be eliminated. Using this algorithm we should examine student's results.

Clustering is an unsupervised data mining technique that finds in the data set the objects that are in some sense similar one to another. In ODM clustering is performed by 2 algorithms:

– Enhanced K-Means,
– Orthogonal Partitioning Clustering.

Clustering technique can be very useful in segmenting the students according to their skills, to assign them to the best matching education program. By the matching education program we can understand the field of study fitting students' skills best, learning styles, the level of some of the subjects or any other aspects of education process.

8. Conclusions

Research findings can be expressed in the following way.

– Business Intelligence represented by three components: data warehouse, OLAP and data mining, is a very useful technology in the improvement of education quality management purpose.
– Data warehouse model for the defined problem should cover at least three areas: student education, didactic staff and the domain education.
– Data mining is considered to be useful tool for detecting the aspects of education process that should be straightened to improve education quality.

The usage of all BI components (DW, OLAP and DM) seems to be necessary in order to support properly education quality management. Not big number of research in the defined problem should encourage further investigation.

References

Business Intelligence – system, technologie czy kultura, Computerworld Custom Publ., *Raport Specjalny: Strategie*, luty 2003.

Bielecki W.T. (2008), *Założenia dla systemów e-learning*, http://republika.pl/webmarketing/materialy/eLearning/e-learning_tekst.htm (May, 2008).

Cheng Y.Ch. (2001), *Paradigm Shifts in Quality Improvement in Education: Three Ways for the Future*, http://home.ied.edu.hk/~yccheng/doc/speeches/12-15jun01.pdf.

Christobal R., Ventura S., Garcia E. (2008), Data mining in course management systems: Moodle case study and tutorial, *Computers & Education*.

Dekker G., Pechenizkiy M., Vleeshouwers J. (2009), Predicting students drop out: A case study, [in:] *Proceedings of the Second International Conference on Educational Data Mining (EDM), Cordoba, Spain, 1-3 July, 2009*, Eds. T. Barnes et al.

Hayhoe G.F. (2002), *Evaluating Distance Learning in Graduated Programs: Ensuring Rigorous, Rewarding Professional Education*, http://www.puw.pl/elerning.html (October 25, 2002).
Morzy M. (2005), *Oracle Data Mining – odkrywanie wiedzy w dużych wolumenach danych*, XI Konferencja PLOUG, Kościelisko, październik 2005.
Nycz M. (2007), Knowledge in distance learning system, [in:] *Pozyskiwanie wiedzy i zarządzanie wiedzą*, Eds. M. Nycz, M. Owoc, Prace Naukowe Akademii Ekonomicznej, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
Oracle (2009), www.oracle.com/technology/products/bi/odm/odm_techniques_algorithms.html.
Peraya D. (2003), *Distance Education and the WWW*, http://www.puw.pl/elerning.html (January 18, 2003).
Singh H. (2002), *Demystifying e-learning Standards*, http://www.puw.pl/elening (November 7, 2002).
Van Dyke L. (2008), Data warehouse model for micro-level decision making in higher education, *Electronic Journal of e-learning*, Vol. 6, No. 3, www.ejel.org.
Vialardi C., Bravo J., Shafti L., Ortigosa A. (2009), Recommendation in higher education using data mining techniques, [in:] *Proceedings of the Second International Conference on Educational Data Mining (EDM), Cordoba, Spain, 1-3 July, 2009*, Eds. T. Barnes et al.
Wikipedia (2009), http://en.wikipedia.org/wiki/Oracle_Data_Mining.

KONCEPCJA WYKORZYSTANIA INTELIGENCJI BIZNESOWEJ W OBSZARZE ZARZĄDZANIA JAKOŚCIĄ KSZTAŁCENIA

Streszczenie: wśród podstawowych obszarów zastosowań inteligencji biznesowej (IB) brakuje aplikacji w sektorze edukacyjnym. Jednocześnie istnieje realna potrzeba ustawicznego doskonalenia jakości nauczania na współczesnych uniwersytetach. Celem artykułu jest dokonanie analizy użyteczności inteligencji biznesowej wykorzystywanej do poprawy jakości edukacji w szkolnictwie wyższym. Oprócz wskazania wybranych własności IB w kontekście wspomagania decyzji menedżerskich szkolnictwa wyższego został zaprezentowany opracowany w tym celu model hurtowni danych. Ponadto została przedstawiona dyskusja dotycząca technik eksploracji danych w środowisku bazy Oracle i ich stosowalność w omawianym problemie.