

BUSINESS INTELLIGENCE IN PROCESS CONTROL

Alena KOPČEKOVÁ, Michal KOPČEK, Pavol TANUŠKA

Abstract

The Business Intelligence technology, which represents a strong tool not only for decision making support, but also has a big potential in other fields of application, is discussed in this paper. Necessary fundamental definitions are offered and explained to better understand the basic principles and the role of this technology for company management. Article is logically divided into five main parts. In the first part, there is the definition of the technology and the list of main advantages. In the second part, an overview of the system architecture with the brief description of separate building blocks is presented. Also, the hierarchical nature of the system architecture is shown. The technology life cycle consisting of four steps, which are mutually interconnected into a ring, is described in the third part. In the fourth part, analytical methods incorporated in the online analytical processing and data mining used within the business intelligence as well as the related data mining methodologies are summarised. Also, some typical applications of the above-mentioned particular methods are introduced. In the final part, a proposal of the knowledge discovery system for hierarchical process control is outlined. The focus of this paper is to provide a comprehensive view and to familiarize the reader with the Business Intelligence technology and its utilisation.

Key words

business intelligence, system architecture, life cycle, analytical methods, data mining

INTRODUCTION

This paper deals with the Business Intelligence technology. A brief overview of its architecture and its most used analytical methods is offered. Every current company or organization has a place for data storage. However, even if the data were collected systematically and structured, they do not have the necessary information value for decision-making process. Therefore, it is necessary to use appropriate extraction tools and analytical methods to transform information into knowledge. Such knowledge plays an important role in the business decisions. The abilities of Business Intelligence could be also used in the process control with the advantage discussed below.

Ing. Alena Kopčková, Ing. Michal Kopček, PhD., doc. Ing. Pavol Tanuška, PhD. - Institute of Applied Informatics, Automation and Mathematics, Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava, Hajdóczyho 1, 917 00 Trnava, Slovak Republic
e-mail: alena.kopcekkova@stuba.sk, michal.kopcek@stuba.sk, pavol.tanuska@stuba.sk,

BUSINESS INTELLIGENCE

The Business Intelligence (BI) could be defined as a collection of mathematical models and analytical methods which are used to generate the knowledge valuable for the decision-making processes from available data. To highlight the importance of BI in companies and organisations, some fundamental advantages are listed below. The BI:

- helps to direct the organisation towards its main objectives,
- enhances the decision-making ability of analysts and managers,
- enables faster and factual decision making,
- helps to meet or to exceed customer expectations,
- helps to identify competitive advantages,
- combines multiple data sources for the decision-making process,
- ensures efficient acquisition and distribution of essential data and statistics,
- finds hidden problems using information that is not visible,
- provides immediate answers to the questions that arise during the data study.

BUSINESS INTELLIGENCE ARCHITECTURE

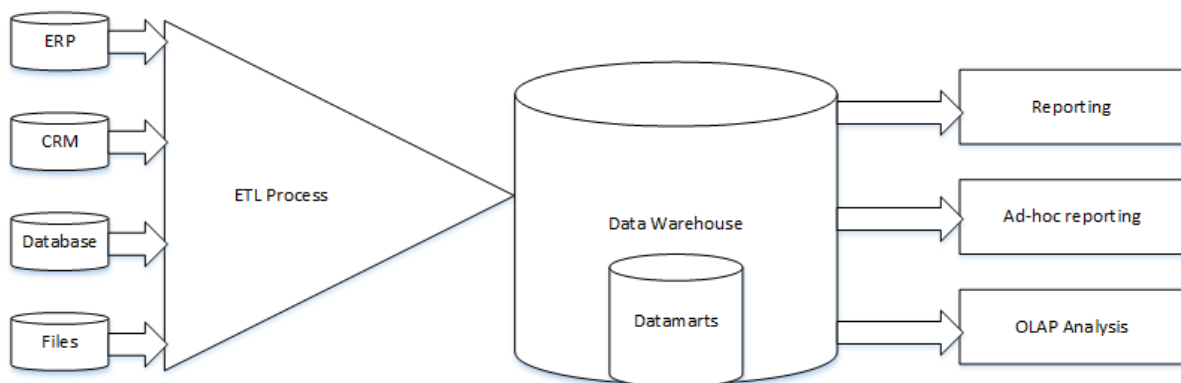


Fig. 1 Typical architecture of the BI system

The BI system architecture shown in Fig. 1 consists of three main parts.

The data sources. In the first step, it is necessary to collect and integrate all available data from the primary and secondary sources, which differ in type and origin.

The data warehouses and data stores. Next step is to use the extraction and transformation with the following storage of data from multiple sources (ETL) into the BI supportive databases. These databases are so-called data warehouses and data marts. Bill Inmon defines the data warehouse as a collection of integrated, subject oriented databases designed in support of management decisions. The data marts represent an essential part and form a subset of the whole company data warehouses. The data for data marts is selected to meet the specific requirements of certain part of the organization.

Methodologies of BI. The extracted data serves as inputs of mathematical models and analytical methods. Several applications could be implemented as a support for the decision-making process within the BI system:

- multidimensional cube analysis,

- exploratory data analysis,
- time series analysis,
- inductive learning models for data mining),
- optimization models.

A more detailed description of the components of the BI system architecture described above is provided by a hierarchical (pyramid) model in Fig. 2. The components of the two bottom layers were described above, therefore the next part of this chapter focuses on the description of the remaining model layers.

Data exploration. The third layer of the pyramid model provides tools for the passive BI analysis. The methodologies including the query and reporting systems as well as statistical methods are considered passive, because they require prior determination of initial hypotheses and definitions of the data extraction criteria from decision-makers. Subsequently, the answers obtained by analytical means are used to verify the correctness of decision-makers' initial view.

Data mining. The data mining within the introduced model is understood as an equivalent of the term Knowledge Discovery in Databases (KDD), nevertheless, many authors describe it only as a part of the process. It represents the fourth layer of the pyramid model and provides tools for the active BI analysis, which role is to extract information and knowledge from data. Active methodologies do not need the prior definition of hypotheses to be verified; on the contrary, they serve to expand the decision-makers' knowledge.

Optimisation. It forms the last but one layer of the pyramid model and makes easier the selection of the best or optimal solution from large, often infinite, number of alternatives.

Decisions. The top of the pyramid model is represented by the selection and acceptance of specific decisions, which ends-up the decision-making process. Even if the company has effectively set BI methodologies, the final decision rests on the shoulders of decision-makers who dispose the informal and unstructured information. Therefore, they have an ability to suitably modify the recommendations and outcomes obtained using mathematical models.

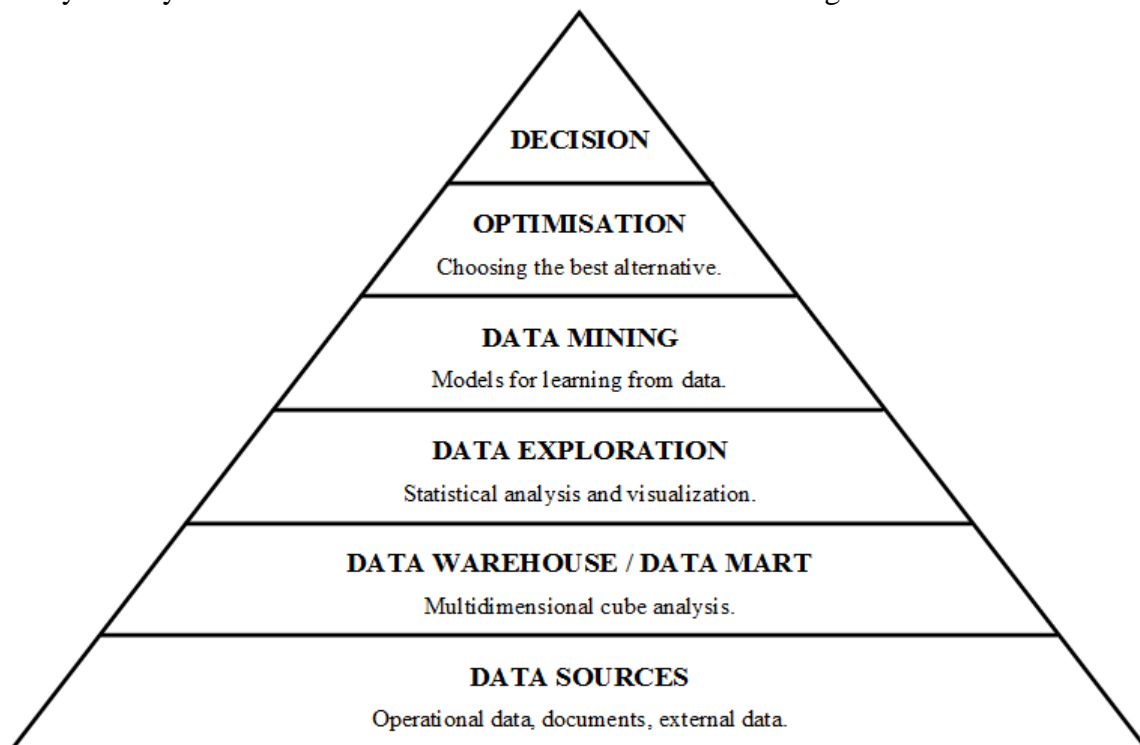


Fig. 2 Hierarchical model of the BI system (VERCELLIS 2009)

LIFE CYCLE OF BUSINESS INTELLIGENCE

The life cycle of an analysis within the BI system is depicted in Fig. 3, which reflects the above described architecture of the BI system. Successful operation of the BI system requires four steps mutually interconnected into a ring.

- Acquiring the huge amount of precise data from databases, data warehouses and another data sources.
- Analysis of the obtained data using BI methodologies, while complex elements are extracted into smaller segments for better understanding and discovering of new knowledge. Solutions to business problems are obtained using the business needs identification and reports are prepared (graphs, diagrams, annual reports, etc.).
- Trends identification using predictive analysis and identification of business threats and opportunities using complex mathematical methods.
- Simulation and acquisition of new knowledge about business problems, threats and opportunities, and application of the obtained knowledge by means of decisions. Consequently, the accuracy of the taken decisions is evaluated in a new cycle of extraction and analysis of ongoing processes.

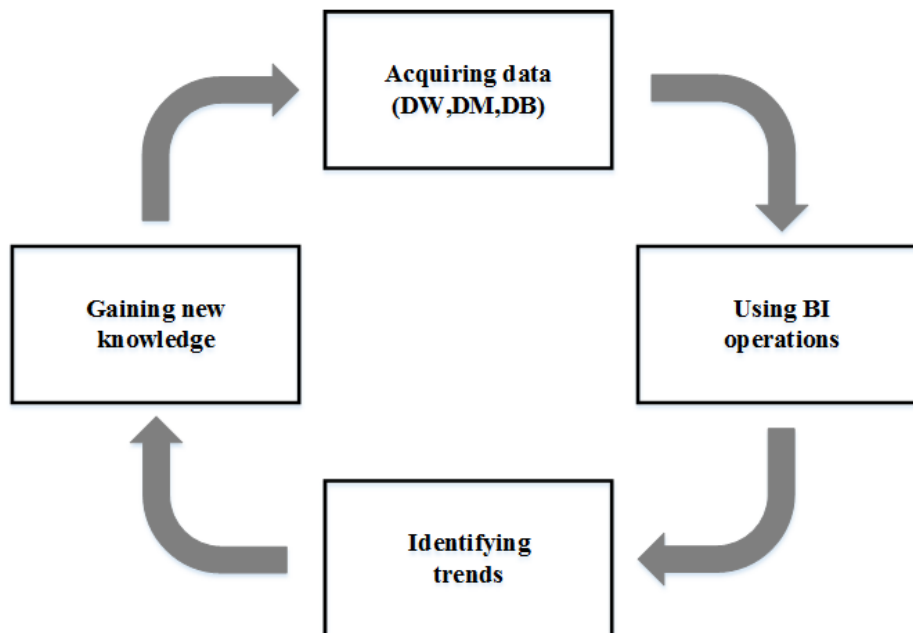


Fig. 3 Life cycle of analysis within the BI (KOCBEK 2010)

ANALYTICAL METHODS IN THE BUSINESS INTELLIGENCE SYSTEM

The basic division of analytical methods into methods, which serves for hypotheses verification and methods for knowledge expansion, was made within the BI system architecture description. Some specific methods are briefly introduced in this chapter.

OLAP – OnLine Analytical Processing

Online Analytical Processing or OLAP is a technology of data storage in databases, which allows organizing large amount of data so that the data is accessible and understandable to users engaged in analysing business trends and results. (LIŠKA 2008) The method of data storage differs by its focus from the more commonly used OLTP (Online Transaction

Processing), where the emphasis is primarily on the simple and secure storage of data changes in the competitive (multiuser) environment.

Basic differences between OLAP and OLTP arise from the different applications. In the OLAP, it comes to the one-time recorded data, over which complex queries are implemented. In the OLTP, data is continuously and frequently modified and inserted, and usually simultaneously by many users. (OLAP 2013)

The OLAP tools are widely used also within the data mining. The main objective of the process is to gather knowledge from the existing data set and transform it to the structures, which are easily understandable to the user for further application.

DATA MINING

There exist many different definitions of the data analysis technology known as data mining. Most of them involve a procedure of searching and discovery of useful relationships in large databases. As the personal computers became more effective and user friendly, the new data mining tools were developed to get an advantage from the growing power of computer technology. Procedures for data mining are designed in response to the new and increasingly expressed needs of management decisions. Some definitions of data mining are anchored in specific analytical methods, like neural networks, genetic algorithms, etc. Other definitions of data mining are sometimes confused with the definitions of data warehousing. Building the data warehouses and data mining are complements. The data warehouse serves as data storage, but not to transform the data into information. Data mining transforms the data into information and information into knowledge (KEBÍSEK 2010).

The most common data mining tasks comprise:

- classification,
- prediction,
- association rules.

Classification is a method used to divide data into groups according to specific criteria. If the criteria are known in advance, at least for a data sample, a model may be developed using predictive modelling methods, the output of which is a classification variable. If the resulting criteria are not known in advance and the classification task is to find them, an unguided classification occurs. The technique used in such cases is a cluster analysis. A prime example of cluster analysis is e. g. finding the groups of markets based on their turnover, product range and customer type. The found groups (clusters) are consequently used e. g. as a specification for an advertising campaign aimed at different groups of stores. (LIŠKA 2008)

The basic methods of classification are (PARALIČ 2003):

- Decision trees - Classification using the decision trees method is the most often used to determine the target attribute having discrete values. The method belongs to the inductive learning methods and has been successfully applied in many fields such as classification of patients according to diagnosis or prediction of credit and insurance frauds (classification according to the risk of credit or insurance frauds).
- Bayesian classification - Bayesian classifiers classify the given examples into different classes according to the predicted probability of attribute values. (PARALIČ 2003)
- Classifiers based on k-nearest neighbours – Classes are represented by their typical representatives in the method. A new example is assigned into the class on the basis of similarity during the qualification process (minimum distance to the representative from some class). (LIŠKA 2008)

Prediction is used to forecast a value of continuous (numeric) target attribute. A typical example is a model of behaviour of loan applicants. On the basis of the previous borrowers and their loan repayment, bank can build a classification model using data mining techniques, which is used to place new borrowers into one of the predefined categories (e.g. allocate the loan or not) based on the data contained in the loan application. (PARALIČ 2003)
The basic methods of prediction are (PARALIČ 2003):

- Regression – The linear regression is the simplest regression method which models the data using lines. The two dimensional linear regression method models the target attribute Y (predicted variable - regressand) as a linear function of another known attribute X (predictor variable - regressor). Regression coefficients α , β could be calculated by the least squares method, which minimises the error between the actual data and the approximation line.
- Model trees – In the case of prediction tasks, where the transformation of the linear model is not possible, it is useful to use e.g. M5 algorithm that generates predictive models in the form of a tree structure, which is much like decision trees.
- Predictors based on k-nearest neighbours – The final prediction is made on the basis of the values of the target attribute of k-nearest examples from the training set of the just solved example. The difference compared to classification is that the resulting predicted value is calculated as the average value of k-nearest neighbours in the training set. (PARALIČ 2003)

Association rules are used to describe the categories or classes and to reflect the connection between objects. The rules very often use the IF-THEN construction and logical operators AND, OR, and their reliability is determined using probability. (MAGYAROVA 2008)
The association rules are (PARALIČ 2003):

- simple association rules,
- hierarchical association rules,
- quantitative association rules.

Methodologies of knowledge discovery in databases

There are methodologies the aim of which is to provide a uniform framework for the solution of data mining tasks. Some were developed by software producers e.g. methodologies 5A and SEMMA, others arisen from the cooperation of research and commercial institutions e.g. CRISP-DM, which is shown in Fig. 5.

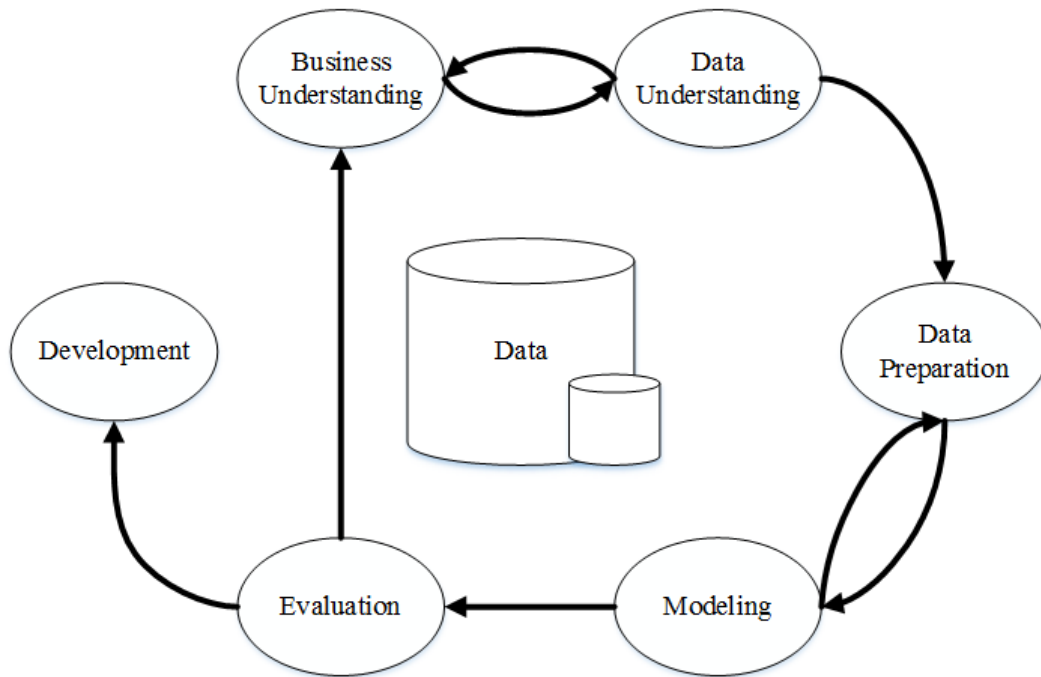


Fig. 5 Phases of the CRISP-DM reference model (MIRABADI 2010)

The common essence of all methodologies is the following steps:

- Business/practical - the formulation of the task and understanding of the problem.
- Data - search and preparation of data for analysis.
- Analytical - finding information in the data and creating statistical models.
- Application - acquired knowledge and models are necessary to e.g. launch an advertising campaign or reorganize Website.
- Revisory - the need to ensure feedback and for long-term deployed models to check, whether the model has not become obsolete and still retains its effectiveness.

Applications of data mining

- Relational marketing:
 - identification of customer segments that are most likely to respond to the targeted marketing campaigns, such as cross-selling and up-selling,
 - prediction of the rate of positive responses to marketing campaigns,
 - interpretation and understanding of the buying behaviour of the customers,
 - market basket analysis,
- fraud detection,
- risk evaluation,
- image recognition,
- medical diagnosis.

KNOWLEDGE DISCOVERY IN HIERARCHICAL PROCESS CONTROL

Before submitting a proposal for the knowledge discovery for hierarchical process control, the following stages should be implemented:

- Analysis of data at all levels of the pyramid model of process control and subsequent design of the functions and structure of specialised data storage.
- Verification and validation of data storage design in accordance to the existing international standards and guidelines.

- Proposal of transformation subsystem of heterogeneous data structures into a single structure, which is given by the proposal of data storage.

Mastering the three stages is a prerequisite for achieving the defined objective, i.e. to use the data warehouses for process control by using the system with hierarchical (multilevel) structure.

The aim can be formulated as a comprehensive approach to solving problems which relate to the processing of extreme amount of data for the control purposes of complex systems.

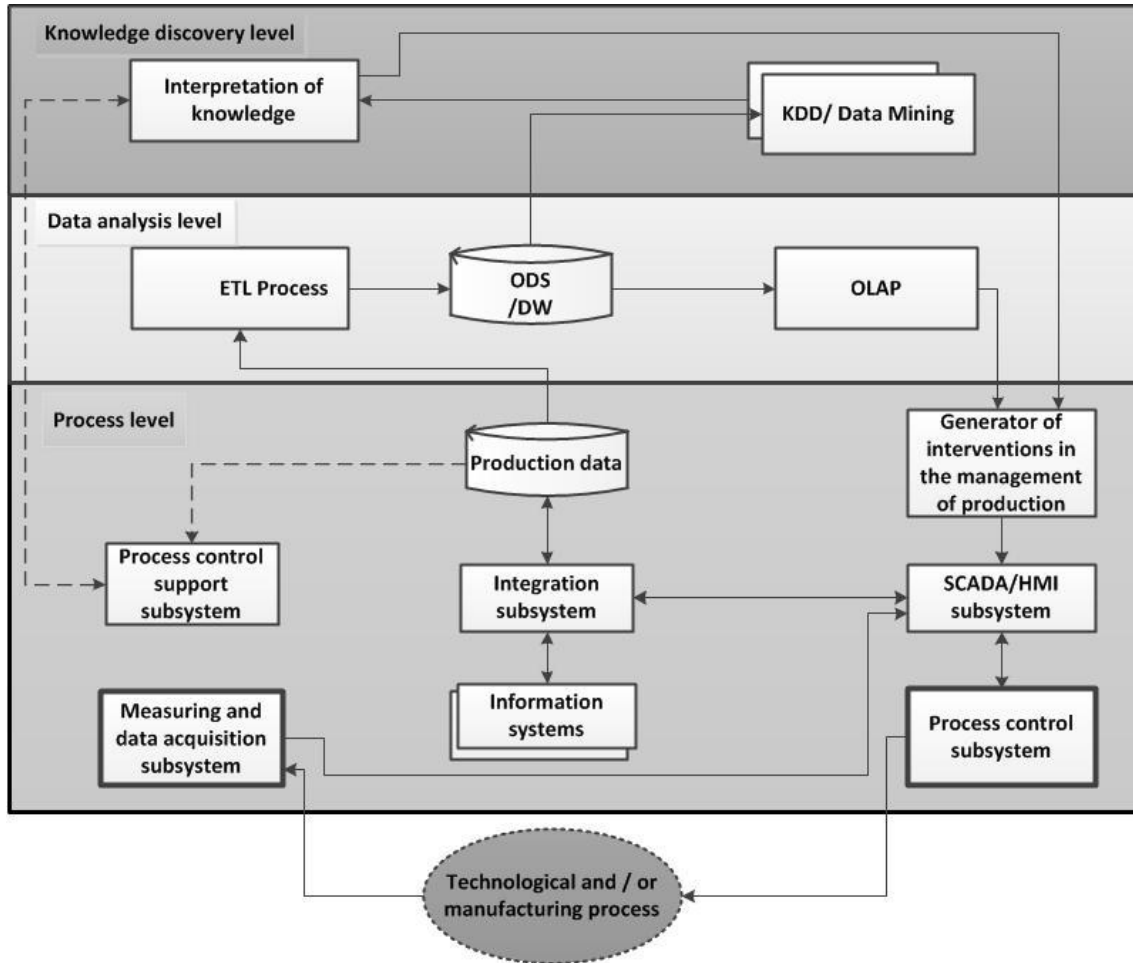


Fig. 6 The conceptual scheme of proposal of the knowledge discovery system

The overall solution is divided into three levels as shown in Fig. 6. The basic level, so called process level, is the information related to the technological and/or manufacturing process. It is based on the standard pyramid model.

It contains all the relevant elements and subsystems which are essential for the process control of any industrial enterprise and provide, among others, the functions such as measurement of the instantaneous values of variables, monitoring parameters (e.g. thresholds, trends, etc.), actuating of the actuators, manual operation, setting the setpoint of controllers and logic control, data visualisation, archiving the process data, the transfer of information to the superior level, etc.

The next level is the level of data analysis and it should be understood as the superior one. It includes subsystems for the acquisition, extraction, transformation and integration of data, including the data warehouse itself or operational data store. The system also includes OLAP tools and technologies.

The fundamental requirement for the extraction and processing of data from production databases is to maintain data integrity. It is in many cases difficult to solve and extremely complicated requirement due to the transformation and processing of data into different data structures within the data storage. Potential knowledge hidden in a huge amount of data, is necessary to prepare in several steps for processing and yet not to violate the relations and linkages in these data.

The last and the highest level is the knowledge discovery level which includes KDD subsystem with the subsystem of the knowledge interpretation. Only properly obtained (using methods and techniques of data mining) and interpreted knowledge could be effectively used to improve the quality of process control. Therefore, the most important part of the system is the module for knowledge interpretation. Properly interpreted knowledge is subsequently necessary to be used and ensure its transport back into the production process. This is done through the "Generator of interventions in the management of production" module. The module should be seen as an interface between the level of knowledge discovery and the control level.

The module itself is directly connected to the SCADA/HMI and the interventions to the production process are directly implemented through it. It must be said that the interventions using the acquired and correctly interpreted knowledge may be carried out in the manual or automatic modes.

The information flow from the level of knowledge discovery to the process level itself may include:

- control algorithms parameters,
- balance calculations values,
- static and dynamic parameters of models,
- parameters useful for equipment diagnosis - maintenance support,
- documents to ensure quality of production, etc.

Benefits of the proposed solution

The proposed system of knowledge discovery for the hierarchical process control could help in solving the following issues:

- Prediction of emergency situations in controlled process based on the principle of finding the analogous situations by the processing of large amounts of data in real time.
- Prediction of preventive production equipment checks associated with maintenance.
- Identification of the impact of process parameters on the production process.
- Identification of slightly incorrect information sources (sensors), where conventional techniques for estimation of alarms limits fail.
- Diagnostics of manufacturing systems with respect to the overall service life of these systems.
- Identification and optimisation of relevant control parameters, that affect the safety improvement of process control.
- Defective operations of actuators, such as insufficient implementation of the calculated action.

- Refinement of nonlinear dynamical models of controlled processes focused on optimising the parameters.
- Continuous monitoring of the quality of process of control on the basis of the quality evaluation of the online obtained data.
- Detecting error conditions of production facilities as well as individual products - detection of rejected pieces.
- Identification of various non-standard states, that affect the production process and which the production operator must solve the most commonly by unplanned shutdown of a machine or technology.
- Solution to the problems using the obtained knowledge without pre-specified objectives.
- Prediction for the needs of enterprise management, various ad hoc reports.
- Effective implementation and especially innovation of control systems at all levels.

CONCLUSION AND DISCUSSION

Business Intelligence represents the top of the hierarchical model of the business administration and information system and a strong tool which helps to increase quality, reliability, safety and efficiency of business processes. Knowledge is gained through the Business Intelligence, thus providing the analysts and managers with important information about the activities of the company and the insight into the processes within the company. Although the Business intelligence technology was originally developed as a support of the decision-making in enterprises, given its nature and the diversity of its tools, it is clear that it has a very big potential and wide application in the areas outside business management. Business intelligence tools could find application e.g. in control parameters optimisation and complex nonlinear technological processes control, in early warning systems, or other systems that uses mass storages for historical data. It is obvious that using the knowledge discovery for the need of hierarchical process control is definitely such type of application.

SUMMARY

The main aim is to give a comprehensive view and to familiarise the reader with the Business Intelligence technology and its utilization, as well as to suggest a new field of application of the progressive technology. The paper points out that the Business Intelligence technology represents a strong tool not only for decision making support, but also has a big potential in other fields of application e.g. medical diagnosis, image recognition, control parameters optimisation and complex nonlinear technological processes control, early warning systems, or other systems that uses mass storages for historical data. Fundamental definitions are offered and explained to better understand the principles and the role of the technology for a company management. Some main identified advantages of the Business Intelligence utilisation are enhancing the decision-making ability, finding and predicting hidden or future problems, etc. The system architecture is hierarchical in nature and consists of six layers from data sources to decision. The life cycle of an analysis within the Business Intelligence system reflects the described system architecture. Successful operation of the Business Intelligence system requires four steps mutually interconnected into a ring to apply Business Intelligence operations, identify trends and finally gain new knowledge. The utilised analytical methods are in general incorporated into online analytical processing and data mining. The data mining methodologies such as 5A, SEMMA and CRISP-DM could be used to simplify the application of Business Intelligence operations. The proposed knowledge discovery system for the need of hierarchical process control was designed in accordance to

the mentioned technology architecture and existing methodologies. Thanks to this a few key benefits were identified.

References:

1. VERCELLIS, C. 2009. *Business intelligence: data mining and optimization for decision making*. Chichester: John Wiley and Sons, 436 p. ISBN 978-0-470-51139-8
2. KOCBEK, A., JURIC, M. B. 2010. Using advanced business intelligence methods in business process management. In: *BOHANEK, M. (ed.) Proceedings of the 13th International Multiconference Information Society - IS 2010*. Ljubljana, Slovenia: Institut Jožef Stefan, Vol. A, pp. 184-188. ISSN 1581-9973.
3. MAGYAROVA, L. 2008. Using techniques of Data Mining in control of production process. *Journal of Information Technologies*, Vol. 2, pp. 1-12. ISSN 1337-7469 Available on: <<http://ki.fpv.ucm.sk/casopis/clanky/03-magyariova.pdf>>
4. MIRABADI, A., SHARIFIAN, S. 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. In: *Safety Science*, **48**(10), pp. 1427–1435. ISSN: 0925-7535
5. MALE, 2012. SEMMA process [online]. In: *Knowledge seeker's blog* [cit. 18.6.2013]. Available on: <<http://timkienthuc.blogspot.sk/2012/04/crm-and-data-mining-day-08.html>>
6. LIŠKA, M. 2008. *Dobývání znalostí z databází (Data mining from databases)*. [online]. Ostrava: Ostrava University in Ostrava, 86 p. [cit. 18.6.2013]. Available on: <http://www1.osu.cz/~klimesc/public/files/DOZNA/Dobyvani_znalosti_z_databazi.pdf>
7. OLAP, 2013. Online Analytical Processing [online]. [cit. 18.6.2013]. Available on: <<http://sk.wikipedia.org/wiki/OLAP>>
8. PARALIČ, J. 2003. *Objavovanie znalostí v databázach (Discovering knowledge in databases)*. Košice: Elfa s.r.o., 80 p. ISBN 80-89066-60-7
9. KEBÍSEK, M. 2010. *Využitie dolovania dát vo výrobných procesoch (The utilization of data mining in the industrial process control)*. Dissertation thesis. Trnava: MTF STU, 116 p.

Reviewers:

doc. Ing. German Michal'čonok, CSc. – Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava

visiting Prof. Ing. Miroslav Božík, PhD. – JAVYS, a.s., Bratislava