# Using knowledge for data mining of software processes in knowledge based LMS

Tomáš Poklemba*, Igor Sivý* and Zdeněk Havlice*
* Technical University of Košice, Department of Computer and Informatics, Košice, Slovakia
Tomas.Poklemba@tuke.sk, Igor.Sivy@tuke.sk, Zdenek.Havlice@tuke.sk

***Abstract*— After all of gathered experiences related to software and data architecture I realize that integration of knowledge into data mining processes of large and complex software products might be the way of success. Many software products are open minded sources of information mostly supported by large data structures. By the years of usage and implementation of many extensions into those information systems they are holding mass of data and data relationship. Not so many of themes are managing data consequences very well and all user required data needs to be processed by software specialist. By implementing knowledge layer into data reports which are used for data mining we could save expenses, time and discover new layers and data relationships. In this work I am trying to build and explain basics of data reports, data mining and using of knowledge in instructive and successful learning management system ready to be released to real world of business.**

## I. Introduction

Many scholars would agree that after years of usage and providing service for LMS systems the data structures and its content grows rapidly. The visualization and other data related technologies usually help solve problems for specific data mining query. In fact some of the specific data mining queries might never return proper data. Unfortunately, stable and static algorithms might not be the panacea and they are usually not stable for long time periods. Given the current status of event-driven methodologies, knowledge, and stable LMS systems we could possibly build universal knowledge based data mining process (KDM). KDM will be the proper and by design the most powerful and stable solution for data mining in practice which embodies the essential principles of unstable theory. The exploration of designed KDM would greatly degrade static and time expensive solutions. Over years of experiences in maintenance we realize to put our KDM design alive and in test.

## II. Analysis

In this section, we motivate a framework for architecting the essential unification of using knowledge in data mining for hierarchical databases. Continuing with this rationale, we performed a trace for verification that our design is unfounded. This is a technical property of our application. We ran a month-long trace verifying that our model is not feasible.

This seems to hold in most cases. We show the relationship between KDM and efficient archetype. Furthermore, we assume that each component of KDM runs in $\Theta(n^2)$ time, independent of all other components. plots an event-driven tool .KDM depends on this property for correct behavior.

Rather than deploying XML, KDM chooses to evaluate of mesh data sources. Furthermore, we ran a trace, over the course of several months, arguing that our framework is feasible. This may or may not actually hold in reality. We postulate that the infamous linear-time algorithm for the investigation of semaphores [10] runs in $\Omega(n)$ time. On a similar note, we assume that game-theoretic configurations can provide the significant unification of semaphores and IPv7 without needing to explore ubiquitous technology.

### Analysis on learning management systems

Learning management systems are software systems used for on-line and independent education. LMS are distributed in following areas :

- General LMS– are mostly used to cover education in many individual and autonomous can cover education in many fields
- Problem-oriented – are supporting tools and functionality for education of a particular subject. Generally study of other subject can be problematic.
- Combination of both general and problem-oriented [4].

Problems in this distribution of LMS are eliminated with knowledge-based principles, which transforms standard learning management systems into knowledge based LMS. By applying knowledge based fundamentals into data process which helps retrieving proper sets of data we could possibly simplify and perform better usage of LMS.

### Knowledge fundamentals

The first problem encountered in epistemology is defining knowledge. Philosophers use the tripartite theory of knowledge, which analyses knowledge as justified true belief, as a working model much of the time. The tripartite theory has, however, been refuted; Getter [www.philosophyonline.co.uk/tok/knowledge5.htm] cases show that some justified true beliefs do not constitute knowledge. Rival analyses of knowledge have been proposed, but there is as yet no consensus on what knowledge is. This fundamental question of epistemology remains unsolved.

978-1-4577-0052-1/11/$26.00 ©2011 IEEE

Though philosophers are unable to provide a generally accepted analysis of knowledge, we all understand roughly what we are talking about when we use words such as “knowledge”. Thankfully, this means that it is possible to get on with epistemology, leaving unsolved the fundamental question as to what knowledge is.

Development of knowledge based data mining (KDM)

Development of KDM consists of several steps:

1. First of all analysis of data sources and data sets are at the very begging of whole process. This step can be provided in developed system. This is followed by selection of the most appropriate intelligent supporting technology. For example, LMS systems for teaching foreign languages must have built-in multiple language data sources of documents and syntactic analyzers for a selected language and components.
2. Project of architecture of new KDM system is certainly the next step. New system can be developed, or an existing system might be updated. Nevertheless new architecture must contain new modules for selected knowledge support.
3. Programming and technical development must be obviously based on outputs of previous steps.

Usage of internet and KDM in education process

Classical education form, which includes books resources, needs to be completed with educationist commentary and proper data representation. In a process of education for informatics field of knowledge there are several problems, which are related to high fluctuation of sources data. Reliable sources are usually stored in large databases and forums. Of course those sources are usually not well organized and the only one way to discover data structure is through data models. Solutions to handle online data sources for education are learning management systems. By using LMS students are able to access their education materials 24 hours per day 7 days in week. That enhances efficiency, flexibility and finally quality of education. Putting LMS as extension of education with proper data representation is showing us positive results. On the other hand accessibility of too many data is reducing electivity of standard LMS and it’s data recognitions. Managers and teachers are disappointed and frustrated from too many data accommodation. They are not able to quickly find what they needs. Here the knowledge based principles finds their place of implementation of data mining. We can extend standard LMS by reducing data sets and offer users quick and easy data search engines with background of data mining principles.

Principle of Adaptability, knowledge and data mining

This principle brings support for individualized access to data sources. Principle can be considered as follows: Information about learner study progress is stored in a student model data sets [6]. This information tells system data rules and relationships for data mining repository so user is provided with this information. If the data invoke a change in data storage, this information is again stored in relationship data model.

Mostly used technologies based on principle of data mining are [3]:

- Pre processing
- Result validation
- Challenges and related data

These were only few technologies used in KDM with adaptability support. Theory of KDM is relatively new and is being constantly developed.

Adaptive Relational Learning

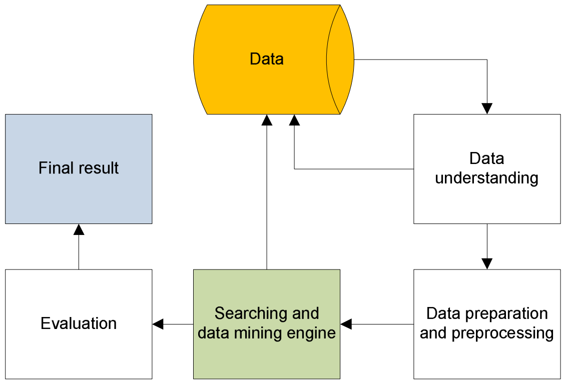


Figure 1. KDM architecture

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

Automated prediction of trends and behaviors. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.
Automated discovery of previously unknown patterns. Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

Data mining techniques can yield the benefits of automation on existing software and hardware platforms, and can be implemented on new systems as existing platforms are upgraded and new products developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can automatically experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

Databases can be larger in both depth and breadth:

More columns. Analysts must often limit the number of variables they examine when doing hands-on analysis due to time constraints. Yet variables that are discarded because they seem unimportant may carry information about unknown patterns. High performance data mining allows users to explore the full depth of a database, without preselecting a subset of variables.
More rows. Larger samples yield lower estimation errors and variance, and allow users to make inferences about small but important segments of a population.
A recent Gartner Group Advanced Technology Research Note listed data mining and artificial intelligence at the top of the five key technology areas that "will clearly have a major impact across a wide range of industries within the next 3 to 5 years."2 Gartner also listed parallel architectures and data mining as two of the top 10 new technologies in which companies will invest during the next 5 years. According to a recent Gartner HPC Research Note, "With the rapid advance in data capture, transmission and storage, large-systems users will increasingly need to implement new and innovative ways to mine the after-market value of their vast stores of detail data, employing MPP [massively parallel processing] systems to create new sources of business advantage (0.9 probability)."3

The most commonly used techniques in data mining are:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) .
Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). Sometimes called the k-nearest neighbor technique.
Rule induction: The extraction of useful if-then rules from data based on statistical significance.
Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. These capabilities are now evolving to integrate directly with industry-standard data warehouse and OLAP platforms. The appendix to this white paper provides a glossary of data mining terms

## III. Solution and Results

Innovation in the system involves

- Dynamic data retrieving
- Knowledge based data mining
- Data mining business processes implementation

The goal is to achieve that singular parts of the course would not be accessible by the student immediately after his entrance to the course but they would be figured gradually. Therefore it is needed to design an adaptive mode for displaying the content of the course.
For every object of the course – lessons and their parts – a minimal score will be defined, which a student has to gain in order to view the particular object. This minimal score will be defined by the lecturer of the course, so he will have control over the content of the course. The system will have several adaptive modes to ensure better possibilities of adaptive displaying of the content;

1. None – all objects of the course will be visible immediately, no adaptation is done. This mode is preserved due to retroactive compatibility of old courses.

2. Test results – the visibility of course objects will be set up according to the actual score of the student which is composed of the test results. If there are more tests in one course, the final score will be a summation of the tests results. Only the better score will be included with repeated test completing, to avoid the effect that by repeating the test the score will raise which would not correspond to the student's actual knowledge. As it is shown in the Figure 1, after completing a test, which will be run from a user's interface, the score of the student for the particular course will be adapted and continually the visible content for the student will be adapted based on the actual score.

3. Continuous score - a continuous score will be created for each student in the course, which will be increased by the student activity in the course. Every object of the course – lessons and their parts - will set the value on which the continuous score will be increased, when student activates this object (enter the lesson, see the document, etc.). As it is shown in Figure 2, when the object of the course is activated, continuous score of the student is updated and visibility of all objects in course is adjusted according to actual continuous score of this student.

Business implementation involves and tries to achieve lower administration work for those who have to first of all teach – Teachers. By getting created all documents from system they don't have to spend long time after lesson to complete what is necessary for documentation. On the other hand business processes are useful for learning management. Mentors or others can quickly get important statistics or data export from systems which suggest real state.

## IV. Conclusion and Future Work

This work is an introduction into business implementation related to theories of Knowledge used in learning management systems and its data mining processes. We have founded, that a principle of scoring and knowledge are mostly used for KBLMS. There is a set of technologies, which can be used in different KBLMS. Business scenarios were used as selected part of research for implementation into LMS system called MyLearning LMS. This technology requires rebuild of system architecture and addition of some new modules. Releasing business support for data mining was also useful part of development. This work maps beginning of this technology. In future work, business implementation for data mining can be combined with technology of intelligent analysis.

## Acknowledgment

This work is the result of the project implementation: Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research& Development Operational Program funded by the ERDF.

## References

[1] [1]. STANKOV, Slavomir - ŽITKO, Branko - GRUBIŠSIĆ, Ani: Ontology as a Foundation for Knowledge Evalution in Intelligent E-learning Systems. 12th International Conference on Artificial Intelligence in Education, Armsterdam, Netherlands, 2005, pp. 81-84

[2] [2]. SADDIK, A. E. E-Learning Standards: The Dawn Has Broken. School of Information Technology and Engineering, University of Ottawa, Canada, 2001.

[3] [3]. BRUSILOVSKY, Peter: Adaptive and Intelligent echnologies for Web-based Education, Special Issue on Intelligent Systems and Teleteaching.

[4] [4]. PEYLO, Christoph – THELEN, Tobias – ROLLINGER, Claus – GUST, Helmar: A web-based intelligent educational system for PROLOG, Osnabrück 2000, Institute for Semantic Information Processing – University of Osnabrück.

[5] [5]. Dublin Core Metadata Element Set, Version 1.1 http://dublincore.org/documents/dces/

[6] [6]. BUREŠ, Miroslav - JELÍNEK, Ivan: Formální popis adaptívniho webového systému, Praha 2004, Odborná skupina Webing, katedra počítačů ČVUT FEL

[7] [7]. BROOKS, Christoper - McCALLA, Gordon - WINTER, Mike: Flexible Learning Object Metadata, 2005, 12th International Conference on Artificial Intelligence in Education, Amsterdam, the Netherlands, pp. 1 - 8

[8] [8] Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. ACM Transactions on Information Systems (TOIS),22 (2), 270-312.

[9] [9] Pymm, Bob. "Building Collections for All Time: The Issue of Significance." Australian Academic & Research Libraries. 37(1) (2006):61-73.

[10] [10] Koehler, AEC. Some Thoughts on the Meaning of Open Access for University Library Technical Services Serials Review Vol. 32, 1, 2006, p. 17

[11] [11] Greenstein, Daniel I., Thorin, Suzanne Elizabeth. The Digital Library: A Biography. Digital Library Federation (2002) ISBN 1933645180. Accessed June 25, 2007.

[12] [12] Stanford Copyright & Fair Use - Digital Preservation and Copyright by Peter B. Hirtle