Development of an Information System for Diploma Works Management

Tsvetanka GEORGIEVA-TRIFONOVA
Department of Mathematics and Informatics
St. Cyril and St. Methodius University of Veliko Tarnovo
3 Architect Georgi Kozarov str., Veliko Tarnovo, Bulgaria
e-mail: cv.georgieva@uni-vt.bg

Abstract. In this paper, a client/server information system for the management of data and its extraction from a database containing information for diploma works of students is proposed. The developed system provides users the possibility of accessing information about different characteristics of the diploma works, according to their specific interests. The client/server architecture of the system is described as well as the services offered. The author presents the structure of the created database that stores the necessary information. A client application ADP (*access data project*) is implemented, providing the possibility for insertion, updating and searching, as well as a client application is fulfilled with Java for discovering the constraint-based association rules.

Keywords: database; client/server system; ADP (access data project) application; diploma works; association rules.

1. Introduction

In the last years, the creation, the distribution and the usage of the training and scientific literature is performed largely by means of its digitalized form. In this manner the work of teachers, researchers, professionals in particular areas are facilitated, as well as the work of the students and mainly of the students preparing their diploma works. Providing fast access to the electronic variant of the developed diploma works and related with them materials can sufficiently support the students in their work and increase its quality.

The aim of the presented paper is to represent a client/server system for keeping information on diploma works of students. The implemented client/server system provides users the possibility of extracting information about the developed diploma works. It allows students and teachers quick access by means of a convenient interface to the data and the files, related to the diploma works of students, graduated the bachelor or the master degree of some of the specialities in Department of Mathematics and Informatics in St. Cyril and St. Methodius University of Veliko Tarnovo after the year 2002.

The storage of the data about the diploma works of the students in a database makes suitable conditions for data mining (Barsegyan *et al.*, 2008; Kantardzic, 2003), i.e., analyzing the accumulated data with the purpose of extracting the previously unknown and

potentially useful information. That motivates the utilization of a program for discovering the constraint-based association rules. Association rule mining is a form of data mining to discover interesting correlation relationships among the attributes of the analyzed data. An association rule reveals the frequent occurring patterns of the given data items in the database.

The rest of the paper is organized as follows. In Section 2, the features of the information systems for diploma works are reviewed. Section 3 contains a description of the client/server architecture and the interface of the system. In Section 4, we represent the entity-relationship model (*ER model*) of the database that stores information about the diploma works of the students. We also produce the relational tables obtained after the transformation of the created ER model into relational. These relations are implemented by using the database management system Microsoft SQL Server. Section 5 consists of the description of the client applications for updating the data and data mining.

2. Review of the Features of the Information Systems for Diploma Works

The main features of the existing storage and retrieval systems providing access to the electronic version of the diploma works of the students (HKUST Library; UM Graduate School & Fogler Library; Virginia Tech; WSU Libraries) are:

1. Insertion of the data about a given student.
2. Insertion of the data about a diploma work – topic, year of its defence, file with the electronic version of the diploma work.
3. Searching diploma works by topic, author, year.

The need of a system providing the listed features for our students is the basic motivation for the development of the system represented in the presented paper. Moreover, as a result of our research, we have established that an information system from the considered kind could propose additional features, which make it more useful for the users.

4. Insertion of the files related to the diploma works.

    Besides the files with the content of the diploma works, the corresponding presentations from their defences, the multimedia files, the programs, the databases, the programs source code and the others can be added. On one hand this supplement is helpful for the users. On the other hand, a substantial advantage is that providing the electronic sources permits the students better and more complete ways to represent their developments.
5. Applying algorithms for data mining.

    Applying different algorithms for data mining on the data collected as a result of the usage of the system, could lead to extracting useful information about the diploma works, their topics during the different years, the obtained marks from the defences, etc.

The basic features (1)–(3) are included in the developed information system and a variant for the implementation of the additional (4)–(5) is proposed.

In the presented paper, the client/server architecture of the developed system is described and the services, included in its realization. The structure of the created database

is represented, keeping the necessary information. A client application ADP (Pearson, 2004) is proposed, designed for insertion, editing and searching the data. Besides this a Java program is applied (Eck, 2006; Eckel, 2001), providing the possibility for discovering the constraint-based association rules.

3. Overview of the Client/Server Architecture of the System

The client/server architecture of the developed system is created on the base of the two-layer information model (Fig. 1).

The layer for data processing is implemented by using the database management system. For the present system we use Microsoft SQL Server, which allows efficient storage of large databases and provides functionality for accessing the data (Bieniek *et al.*, 2006; Kroenke, 2003; Microsoft Corporation).

The client part consists of an ADP application, providing a convenient interface for insertion, updating and searching the data, as well as a Java program for mining the constraint-based association rules.

3.1. *SQL Server Database for Data Storage*

The `DiplomaWorksDB` database serves for storage and processing the data for the diploma works, the students and their advisors. Information on the student's faculty number, student's names, the scientific and/or educational degree, the speciality, the form of training, the topic and the annotation of the diploma work, the student's advisor, the reviewers, the date of the defence of the diploma work, the mark obtained by the student for the diploma work is maintained.

For each diploma work a possibility for insertion of an additional information is provided – files (*.doc*, *.pdf*) with its content; presentation (*.ppt*, *.pdf*) of the student for the defence of the diploma work; application implemented by the student (such as a database, a program of C, Java, etc.) and others.
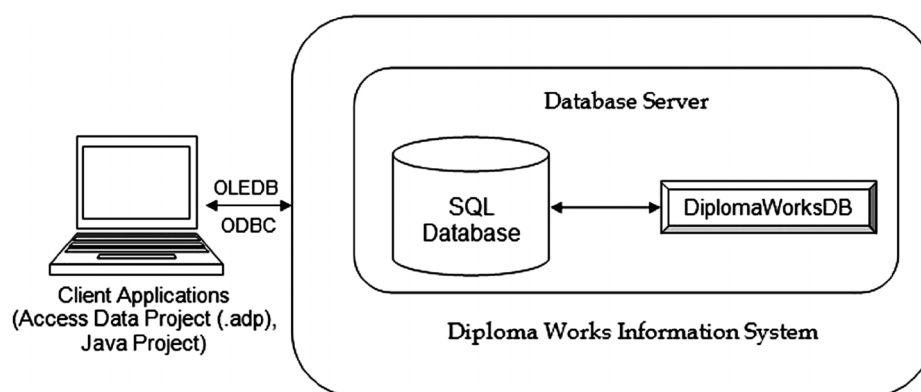

Fig. 1. The architecture of the system.

The basic functions of the database include:
- addition of a new student in the database;
- edition of the data about the students;
- deletion of students from the database;
- browsing the data for the students;
- searching by different criteria.

3.2. *Client Application ADP for Insertion and Searching the Data*

Microsoft Access allows establishing a connection between the current database and tables from other databases of Microsoft SQL Server and other data sources. ADP is connected with a database of SQL Server and provides an access to the objects created in that database (such as tables, views, stored procedures, triggers, etc.). The data are stored in the database of SQL Server. ADP does not contain any data and tables, but it can be used for easy creation of forms, reports, macros. As a result of that, the end user features opportunity for insertion, editing, and deletion of the data by means of a comfortable interface.

3.3. *Client Application for Mining the Constraint-Based Association Rules*

The goal of association rules mining (Agrawal *et al.*, 1993) is to find interesting associations or correlation relationships among a dataset, i.e., to identify the sets of attributes-items, which frequently occur together and then to formulate the rules characterizing these relationships. The constraint-based association rule mining (Fu and Han, 1995) aims to find all rules from given dataset, which satisfy the constraints required from the users. An application is implemented with Java for discovering the constraint-based association rules, which in (Trifonov and Georgieva, 2009) is utilized for the data, obtained after applying the methods of digital processing of signals for analysis of the sounds of the unique Bulgarian bells. This client application is connected with the `Diploma-WorksDB` database with the purpose of performing the association analysis of the data about the diploma works of the students.

4. Modeling of the Data

The model of the `DiplomaWorksDB` database, in keeping with the entity-relationship model (*ER model*), introduced in Chen (1976), is shown in Fig. 2. The entity sets of the ER model are depicted as rectangles, their attributes as ellipsis, and the relationships as rhombs (Garcia-Molina *et al.*, 2002).

The database is implemented by means of the database management system Microsoft SQL Server. The relevant relational tables are shown in Fig. 3.

The structure of the database is defined to provide the best efficiency of the most frequently used operations – insertion, updating, data searching.
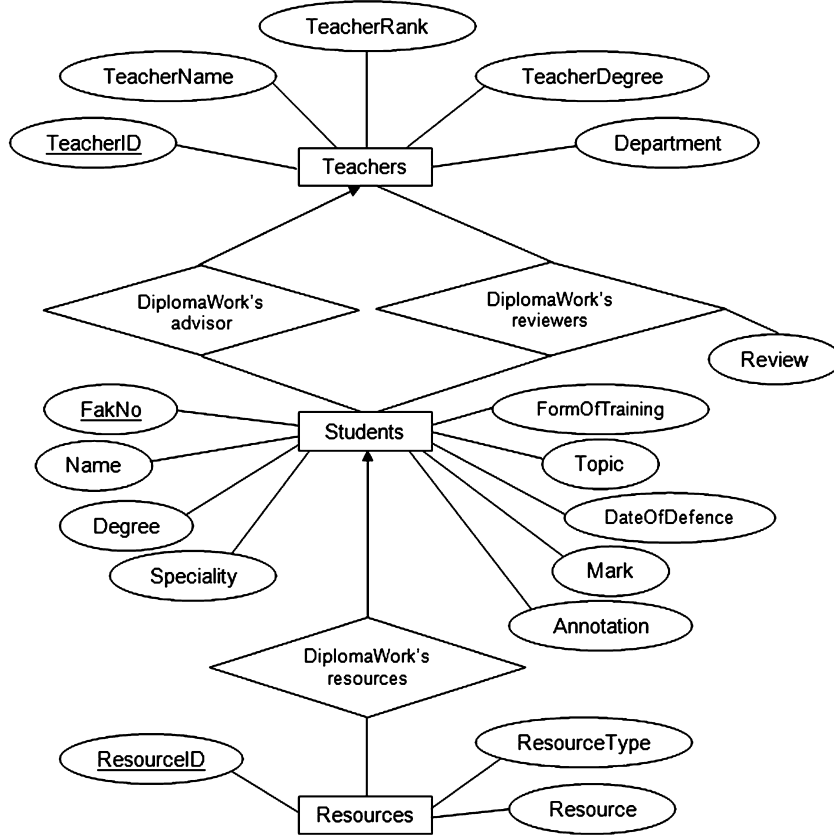
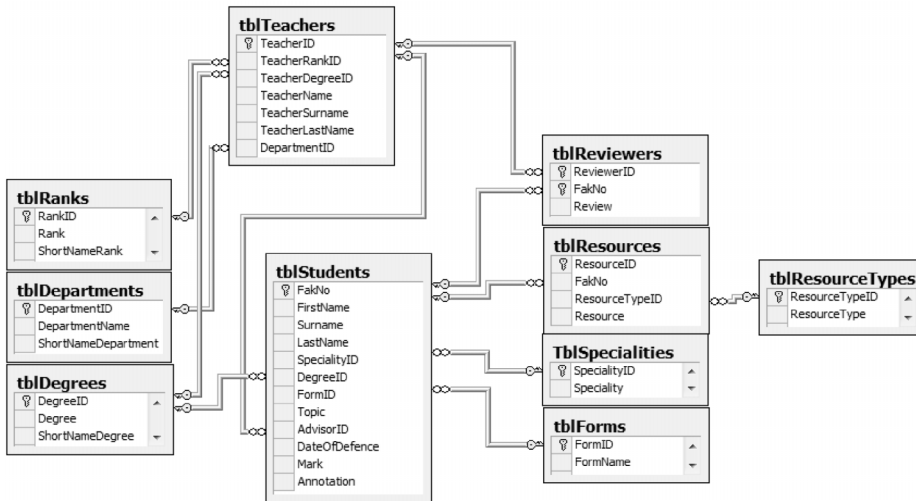Fig. 2. ER diagram of the `DiplomaWorksDB` database.

Fig. 3. Relational model of the `DiplomaWorksDB` database.

The `DiplomaWorksDB` database of SQL Server contains the created views for extracting the data from several related tables, as well as the stored procedures for obtaining the information on the students, defended their diploma works during a specific month and year; the students, whose diploma works' topics comprise a specific string. The stored procedures provide a better performance of the client/server system because they make decreasing the exchange to data between the client and the server. Besides the stored procedures can accept parameters and therefore they can be executed from multiple client applications by applying different input data.

5. Client Applications for Updating, Searching and Mining the Data

A client application ADP is developed for updating and searching the data about diploma works of students, as well as a client application implemented with Java for mining the constraint-based association rules.

5.1. *ADP Client Application for Insertion, Edition and Searching the Data*

Forms for insertion and updating the data are implemented. Their purpose is to facilitate actualization of the information. The form for insertion of the data about the students and their diploma works is shown in Fig. 4.

Besides this, the application allows the execution of different queries, which perform finding the specific information, corresponding to the given searching criteria. Each user can fulfil search by filling in text boxes and/or list boxes which correspond to the chosen

Fig. 4. Form for insertion of the data about students.

characteristics of the diploma works of the students, stored in the database. The results from each query are presented in a format convenient for the end user. The forms and the reports are implemented with the record sources – views and stored procedures designed for extracting the data about:

- students from a chosen speciality;
- students with a chosen diploma work's advisor;
- students, defended their diploma works during a chosen month and year;
- students, whose diploma works contain in their topics a given string.

5.2. *Client Application for Discovering the Constraint-Based Association Rules*

The association rules are introduced in Agrawal *et al.* (1993) and they are utilized for specifying the correlation relationships among itemsets in the database.

Let $I = \{I_1, I_2, \ldots, I_n\}$ be a set of n different values of attributes. Let R be a relation, where each tuple t has a unique identifier and contains a set of items, such that $t \subseteq I$. An association rule is an implication of the form $X \rightarrow Y$, where $X, Y \subset I$ are sets of items with $X \cap Y = \varnothing$. The set X is called an antecedent, and Y – a consequent.

There are two parameters associated with a rule – support and confidence. The support s of the association rule $X \rightarrow Y$ is the proportion (in percentages) of the number of the tuples in R, which contain $X \cup Y$ to the total number of the tuples in the relation. The confidence c of the association rule $X \rightarrow Y$ is the proportion (in percentages) of the number of the tuples in R, which contain $X \cup Y$ to the number of the tuples, which contain X.

The task of association rules mining is to generate all association rules which have values of the parameters support and confidence, exceeding the previously given respectively minimal support *min_supp* and minimal confidence *min_conf*.

To increase the efficiency of existing algorithms for data mining, during the mining process constraints are applied with the goal for these association rules, of which only those interesting to the user are generated, instead of all association rules. In this way a large part of the calculations for mining those rules that are removed as unnecessary, can be avoided. Usually the constraints, provided by the users, are constraints for the data, constraints for the attributes, constraints for formation of the rule.

We have developed an application for discovering the constraint-based association rules, which needs to meet the following preliminary requirements:

- The application must give the opportunity for the user to select the attributes in the antecedent and the consequences of the searched rules.
  Usually the user is interested in a specified subset of attributes and wants to express interesting common connections between the selected attributes. Therefore a facility with a friendly interface should be provided to specify the set of attributes to be mined and exclude the set of irrelevant attributes from the examination.
- The user has to be able to define various values of the minimal support and the minimal confidence, when the items are mined.
  The support reflects the utility on given rule. The minimal support *min_supp*, which an association rule has to satisfy, means that each value, included in the study, has

to appear a significant number of times in corresponding attribute of the initial relation.

The confidence measures the reliability of the inference made by a rule.

- Reducing the number of the generated association rules must be possible by determining the criteria that the values of selected attributes have to satisfy.

  In numerous cases the algorithms generate a large number of association rules. It is almost impossible for the end users to encompass or validate such a large number of association rules, limiting the results of the data mining is therefore helpful. Besides, often the user is interested in definite values of the attributes, which are included in association rules mining.

- Visualizing the obtained results must be represented in a tabular view with providing the opportunity for the user to order the found rules by:
  - alphabetic order of the attributes, which participate in the antecedent and the consequence of the association rules;
  - the support of the association rules in ascending or descending order;
  - the confidence of the association rules in ascending or descending order.

  In tabular view of association rules, all discovered rules are represented in a tabular table with each row corresponding to a rule and represents information about rule support and confidence. By this way the user has a clearer and complete view of the rules and can locate a specific rule more easily. The tabular view facilitates understanding the large number of rules.

An application is developed, which allows the user to set constraints for searched rules and finds constraint-based association rules. The application is used for performing the association analysis on the different characteristics of the diploma works, the information for which is kept in an information system produced for the goal.

To the user that starts the application, the following possibilities, are provided (Fig. 5):

- setting the attributes, being subject to analysis;
- setting the minimal value of the support *min_supp* and the minimal value of the confidence *min_conf*;
- setting the conditions (Boolean expression) for the values of the attributes, which can participate or not in the antecedent and the consequence of the searched rules;
- all the rules can be displayed in different order – by alphabetic order of the attributes participating in the antecedent and the consequence; by the support or confidence in ascending or descending order.

The application is utilized for discovering the constraint-based association rules in the database containing information about the diploma works of about 1000 students graduated the bachelor or the master degree of the specialities "Mathematics and informatics", "Informatics", "Computer science", "Information systems", "Information security", "Computer multimedia" in Department of Mathematics and Informatics in St. Cyril and St. Methodius University of Veliko Tarnovo after the year 2002. The attributes, which can be included in analyzing, are: student's faculty number; student's names; the scientific and/or educational degree; the speciality; the form of training; the topic; the annotation

Fig. 5. Discovering the constraint-based association rules.

of the diploma work; the year of the defence of the diploma work, the mark obtained by the student for the diploma work; the student's advisor.

Figure 5 shows an example result from the execution of the implemented program with given values of the minimal support, minimal confidence and conditions for the values of the attributes. For instance, let the following rule be generated from the database with the diploma works:

$$Speciality(\text{``Informatics''}) \rightarrow Mark(\text{``6''})$$

with values of the support $s = 0.11164$ and the confidence $c = 0.61702$. This rule means, that for the students, graduated in the speciality "Informatics", whose diploma works are in the area of the databases and the information systems, one of the most frequent marks from the defence of their diploma works is 6 (with 61.702% confidence) and the students in Informatics with diploma works in databases and information systems and marks 6 represent 11.164% from all students, included in the study.

Some other examples of association rules, which can be retrieved from the execution of the program, are:

- $Advisor(A), Speciality(S) \rightarrow Mark(M)$ with the values of the parameters support $s = 0.19477$ and confidence $c = 0.64634$, where the condition YearOfDefence = Y is given;

By means of the rules of this kind we can establish that the students with the advisor A, the speciality S and graduated in a given year Y, one of the most frequent marks from the defence of their diploma works is M (with 64.634% confidence) and they represent 19.477% from all students.

- $Mark(M) \rightarrow Speciality(S)$ with the values of the parameters support $s = 0.49644$ and confidence $c = 0.41627$.

  Such rules allow extracting information about the percentage (49.644%) of the students from a given speciality S, whose diploma works are evaluated with certain mark M. Besides, we can conclude that the students, which are evaluated with mark M, are from the speciality S with 41.627% confidence.

The user can establish the relationship between other attributes included in the study by means of the similar rules.

6. Conclusion

In this paper, the automated system is proposed. It explores a client/server based approach to managing the information on the diploma works of the students. The created database contains information about different characteristics of the diploma works and it is implemented on Microsoft SQL Server. The interface is developed by means that allow establishing a connection with the database of the ADP project. This gives users the possibility of easily accessing detailed information about the diploma works of the students.

In addition, an application is represented which provides the possibility for finding the constraint-based association rules of the data about the diploma works.

The basic advantages from the usage of the implemented system can be summarized by the following way:

- The system provides fast and easy access to the developed diploma works and the related materials. The user can receive an extract about the existing works of the graduated students on a given topic. Besides, browsing concrete diploma works and their reviews allows acquiring a clearer notion about the requirements to the final view of a diploma work, for the eventual notes and recommendations.
- An additional motivation of the students is provided for better and fuller representation of their developments.
- Analyzing the accumulated data with the constraint-based association rules reveals an interesting information about the characteristics of the student's diploma works.

Our future work includes development of an application for mining the constraint-based association rules in the text of the diploma works (*text mining*), which allows performing the association analysis of the different words from their contents.

References

Agrawal, R., Imielinski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington, 207–216.

Barsegyan, A.A., Kupriyanov, M.S., Stepanenko, V.V., Holod I.I. (2008). *Technologies for Data Analysis: Data Mining, Visual Mining, Text Mining, OLAP*. BHV-Peterburg.

Bieniek, D., Dyess, R., Hotek, M., Loria, J., Machanic, A., Soto, A., Wiernik, A. (2006). *Microsoft SQL Server 2005 Implementation and Maintenance – Training Kit*. Microsoft Press.

Chen, P. (1976). The entity-relationship model: toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), 9–36.

Eck, D.J. (2006). *Introduction to Programming Using Java*.
http://math.hws.edu/javanotes.

Eckel, B. (2001). *Thinking in Java*. SoftPress.

Fu, Y., Han, J. (1995). Meta-rule-guided mining of association rules in relational databases. In: *Proceedings of the International Workshop on Integration of Knowledge Discovery with Deductive and Object-Oriented Databases*, 39–46.

Garcia-Molina, H., Ullman, J.D., Widom, J. (2002). *Database Systems: The Complete Book*. Williams.

HKUST Library. *The Hong Kong University of Science and Technology Electronic Theses*.
http://lbxml.ust.hk/th.

Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.

Kroenke, D.M. (2003). *Database Processing*. Piter.

Microsoft Corporation. *Transact SQL*. http://www.microsoft.com/sql.

Pearson, W. (2004). *MS Access for the Business Environment: Stored Procedures from the MS Access Client*.
http://www.databasejournal.com/features/msaccess/article.php/3363511.

Trifonov, T., Georgieva, T. (2009). Application for discovering the constraint-based association rules in an archive for unique Bulgarian bells. *European Journal of Scientific Research*, 31(3), 366–371.

UM Graduate School & Fogler Library. *Electronic Theses and Dissertations at the University of Maine*.
http://www.library.umaine.edu/theses.

Virginia Tech. *Electronic Theses and Dissertations at Virginia Tech*.
http://scholar.lib.vt.edu/theses.

WSU Libraries. *The Washington State University Dissertations Web Site*.
http://www.dissertations.wsu.edu.

T. Georgieva-Trifonova received her MSc degree in mathematics and informatics in 1997 and her PhD degree in computer science in 2009 from University of Veliko Tarnovo, Bulgaria. Currently she is assistant professor at the University of Veliko Tarnovo, Bulgaria and teaches databases and information systems modeling. She has published over 30 papers in refereed journals and international conferences, mainly in the fields of databases, collaborative filtering, and data warehousing. She joined several national research projects on the above areas. Her current research interests include multidimensional modeling, data mining, collaborative filtering, and information systems.

Diplominių darbų tvarkymo informacinės sistemos plėtra

Tsvetanka GEORGIEVA-TRIFONOVA

Šiame straipsnyje pristatoma informacinė sistema, atliekanti duomenų tvarkymą ir jų gavimą iš duomenų bazės, kurioje pateikiama informacija apie studentų diplominius darbus. Sukurta informacinė sistema suteikia vartotojams galimybę gauti prieigą prie informacijos apie įvairias diplominių darbų savybes, priklausomai nuo konkrečių vartotojų interesų. Straipsnio autorė pristato sukurtos duomenų bazės, kuri kaupia būtiną informaciją, struktūrą. Pateikiama kliento taikomoji programa, realizuota Java programavimo kalba, teikianti įterpimo, atnaujinimo ir paieškos galimybes.