An Approach of Teaching Data Management

Juliana Peneva, Georgi Tuparov

Abstract: ***The paper focuses on some didactical issues of teaching courses that relate to data management for students at the undergraduate level. The method is based on more than ten years of experience gained by the authors of delivering data management courses in Bulgarian and foreign universities. The main issue discussed in this paper concerns the choice of the topics for several courses and practical seminars in database systems. We assert that students have to achieve information's systems literacy which encompasses how and why the technology is applied in organizations. So, there is a need to find the proper balance between theory and practice. This balance could be achieved through the choice of relevant topics to be covered. Finally we briefly discuss the relationship of data management courses to other courses in computer science and information systems majors.***

Key words: ***data management, DBMS, didactics.***

INTRODUCTION

In today's competitive global business environment, understanding and managing enterprise wide information is crucial for making timely decisions and responding to the changing business conditions. Information is increasingly being recognized as an organizational resource on par with financial and human resources. Many companies are realizing a business advantage by managing successfully their business data. There is a tremendous amount of data generated by day-to-day business operational applications. In addition, there is valuable data available from external sources such as market research organizations, independent surveys and quality testing labs.

Data management applies information technology to turn these huge repositories of data for meaningful information for better business decisions. Databases are a focal point of a manager's involvement with information systems. Businesses rely on database systems to store crucial business information, to maintain its accuracy, and to make it readily available. This information is analyzed and used to make operational, tactical and strategic decisions. The growth in revenues of database software companies is a testament to the economic importance of information systems.

To communicate effectively with information systems or to develop new ones professionals need a profound understanding of database theory and concepts.

That is why knowledge on data management is very important.

This paper focuses on some problems concerning didactical issues of teaching courses that relate to data management for students at the undergraduate level. More precisely we discuss the body of knowledge of two primary courses in the area of data management, namely Introduction of Database Systems and Advanced Database Systems, together with their accompanying practical seminars. These courses introduce the fundamental concepts necessary for designing, using, and implementing databases and information systems respectively. They are based on more than ten years of experience gained by the authors of delivering data management courses in Bulgarian and foreign universities.

What is the balance between theory and practice?

We have notices two basic approaches to data management courses:

1. A highly theoretical approach stressing on formalisms, algorithms for database design, etc. – Maeir's book[1], Ullmann's book[2], Abiteboul's, book[3].
2. A very practical approach consisting of teaching database systems through a concrete DBMS – for example, Access, FoxPro, MySQL, Oracle, etc.

We do not find that any of these approaches is satisfactory. The first approach is highly theoretical and the practice is missing at all. The second – vice versa, it is rather oriented towards concrete DBMS software and surely it lacks generality.

So, there is a need to find the proper balance between theory and practice. We have tried to achieve a balance between formalism and intuition. In presenting concepts and statements we leave out much of the details rather than adhere to very strict proofs. We try to achieve this balance in the first through the choice of the topics to be covered. In addition lectures are accompanied by practical seminars to illustrate some technical decisions. Our efforts were directed towards investigating proper didactical methods and means to present the learning content properly. These courses also intend students to achieve information systems literacy which encompasses how and why the technologies are applied in organizations. Both courses apply the modular approach that is the learning content has been divided into particular components called modules. Each module consists of:

- brief content description;
- educational objectives the students should achieve;
- short content of the parts in the module;
- key words;
- additional references;
- learning content;
- summary;
- references and further reading;
- questions for self assessment.

The description of the educational objectives is based on the Bloom's taxonomy: knowledge, comprehension, application, analysis, synthesis and evaluation [4].

The rest of the paper is organized as follows. In Section 2, we introduce the body of knowledge of our background course Introduction to Database Systems. Section 3 outlines the structure of our next course Advanced Database Systems. Section 4 deals with some methodological issues. The accompanying practical seminars are briefly discussed and the dependency chart of the course's modules is given. Section 5 points out the place of these courses in different major programs.

BODY OF KNOWLEDGE OF THE COURSE INTRODUCTION TO DATABASE SYSTEMS

This course introduces the fundamental concepts for data management [5]. It helps understanding issues concerning the design, use and implementation of database systems. The stress is on data modelling as it represents an important activity during the development of any information system. Different data models are covered and their modelling power is studied and compared. Several topics concerning database design are considered. These include formalisms, theory, and algorithms developed for relational database design by normalization. Functional and other dependencies as well as normal forms for relations are emphasized.

The goals of this course are:

- to present the basic concepts concerning database systems;
- to introduce the principles of database design and management;
- to present the approaches for data management;
- to emphasize on data modelling;
- to teach SQL;
- to provide a background for specific database design decisions;

The basic modules of this course are:

1. Organizing Data and Information: files and databases
2. Data Modeling. The Database Environment
3. The Relational Model
4. Data Processing with SQL
5. Normalization for Relational Databases

6. Conceptual Modeling
7. Database Design and Administration

Module1 examines the two main approaches for data management: the traditional file-based approach and the database approach. After the module completion students become able to distinguish the features, advantages and shortcomings of both approaches. Module 2 focuses on data modelling as the process of mapping the application into a proper representation by taking into account its features and requirements. Data modelling is critical to the overall success of any system. We introduce the concepts of data models and conceptual modelling. We also present a classification of the data models proposed in the literature. The three level pieces architecture of most commercial DBMS and the way this architecture benefits the data independence are outlined. Students learn about the purpose and importance of data modeling and the meaning of logical and physical data independence. In Module 3 we discuss the basic principles of the relational data model, the underlying concepts and terminology. Relational DBMS has become the dominant data-processing software in use today. Students become aware about the properties of database relations, the meaning of entity integrity and referential integrity and how to formulate queries in relational algebra. Module 4 gives an overview of data processing via SQL. Over the last few years the SQL language has become the standard relational database language. It may be considered as one of the major reasons for the success of relational databases in the commercial world. Our explanations conform to the SQL92 standard. The undertaken approach is "illustration by examples". Students become skilled at building basic SQL queries and subqueries, using aggregate functions and grouping, and performing database updates. Module 5 deals with the theory of normalization. The design of a database for a relational system involves the creation of an accurate representation of the data, their relationships and constraints. The formal concepts that allow performing relational design in a top-down fashion by analysing relations independently are presented. The three most commonly used in practice normal forms are described. Students gain knowledge of the purpose of normalization, the problems associated with redundant data and how to qualify the design of relations and normalize a database. In Module 6 conceptual modelling as an important stage in designing successful database applications is discussed. The main purpose for developing a high-level data model is to support the user's perception of the data, thus shielding the more technical aspects associated with database design. Moreover, a conceptual data model is independent of the particular DBMS and hardware platform that is going to be used to implement the database. The basic concepts of the E-R model, developed to facilitate the database design are introduced. Students realize the use of high-level conceptual data models for database design. Module 7 moves from the theory to practical database design as it is taken in large organizations. The basic stages of the database application lifecycle are overviewed and the main activities, associated with each of them are determined. Students learn about the main phases of database design and issues about database administration. This last module and the course of Software Engineering are interrelated.

BODY OF KNOWLEDGE OF THE COURSE ADVANCED DATABASE SYSTEMS

This course aims at introducing the student to the "broad picture" of database technology and enables the students to estimate the current state of this technology. Whereas the first course "Introduction to database systems" have contributed a wealth of concepts in database modeling, database design, normalization theory, SQL and storage management with indexing, all of these areas are being reworked in the light of emerging applications [6].

Distributed data management and the client-server architecture as well as the database integration in WWW are covered. Several topics concerning distributed query

and transaction processing are considered. These include data fragmentation, replication, allocation, schedules and transaction serialization.

New applications and the characteristics of their data, in order to understand the shortcomings of current DBMSs and the need for modern data management, are examined. The emphasis is on non-traditional data modelling. Semantic, nested relational, object and object relational data models are discussed. Their modelling power is studied and compared. The concepts presented include subclasses, specialization, generalization, aggregation. The main peculiarities of object database systems are outlined giving examples from commercial systems. The course ends with a discussion of some advanced topics in the database management area, namely Data warehousing and XML.

The goals of this course are:

- to present the basic concepts concerning distributed processing;
- to introduce the principles of transaction management;
- to present emerging applications, standards, and challenges;
- to give a thorough understanding of the advances in data modeling;
- to emphasize the object-oriented modeling approach that supports the new generation of database applications;
- to provide a background for Data warehousing and OLAP as application and research areas;
- to discuss the advent of XML as a new approach for data interchange, standardization and integration

The basic modules of this course are:

1. Distributed Database Systems and the Client-Server Architecture
2. Transaction Management
3. Semantic Modeling – EER and UML class diagrams
4. Object Data Management
5. The Database Technology and WWW
6. XML
7. Data warehousing
8. An overview of Oracle DBMS

Module1 examines the underlying principles of the distributed DBMSs. Their extended functionality is presented and possible architectures are summarized. Detailed issues of distributed database design, involving fragmentation of data and distribution over multiple sites with possible replication are discussed. After it completion students become able to distinguish the differences among distributed databases, distributed processing, and parallel database systems as well as the features of client-server architecture. Module2 focuses on transaction management as the process of handling the multi-user access to data and recovering from failure. Transaction processing is critical to the overall database consistency. We introduce the concepts of schedule and serializability. Methods that guarantee serializability of schedules, e.g. the two-phase locking are described. Potential problems such as cascading rollback, deadlock and starvation are briefly discussed. Students learn about the purpose and importance of concurrency control and recovery. In Module3 we introduce fundamental abstractions used as a basis of many semantic models. The main concepts associated with the EER model called specialization, generalization, and categorization, are presented. The Unified Modelling Language (UML) notation and concepts for class diagrams are briefly compared to EER notation and concepts. Students enhance their knowledge about different techniques for conceptual object modelling. Module 4 gives an overview of the object database systems. Over the last ten years there is a trend to model together data and processes, i.e. a third generation of DBMS software is launched. We discuss the main features of object- oriented and object- relational DBMS and the emerging OMG, ODMG and SQL3 standards. Students

become skilled at developing databases for engineering applications. Module 5 deals with the direct access to World Wide Web (WWW) databases and databases over the Internet. Various approaches for integrating databases into Web, e.g. scripting languages, Common gateway Interface (CGI), HTTP Cookies, etc. are analyzed. The technologies of Sun comprising Java and JDBC, SQLJ, Servlets, and JSP are presented. Microsoft Web Solution Platform: ASP and ADO together with Oracle Internet Platform are briefly outlined. Students gain knowledge of the different ways to integrate databases and Web. In Module 6 semistructured data and the main rules to create XML documents are discussed. The basic language elements of XML and the difference between well-formed and valid XML documents are given. The way XML documents are organized and stored in a database as well as the data manipulation within XML documents are studied. Students realize why relational, object-oriented and object-relational DBMSs do not handle well semistructured data. Module 7 presents the main concepts and benefits associated with data warehousing. Data Warehousing has emerged as an increasingly popular and powerful concept of applying information technology to turn the huge repositories of data into meaningful information for better business decisions. The data warehouse architecture is introduced and the special characteristics and functionalities are described. OLAP tools and data mining are introduced. Students learn about the main tools and technologies associated with data warehousing. Module 8 moves from the theory to practical issues of corporative DBMS. Students are made acquainted with Oracle DBMS, its architecture and basic characteristics.

ORGANIZATION OF THE EDUCATIONAL PROCESS

The students are required to attend all the classes nevertheless it appears not obligatory. It is expected that they will be present on every meeting of the course. Additionally, they have to read the assigned part of the text in advance. The students are expected to show active participation during the lectures and the practical seminars. Exams consist of multiple choice and short essay questions as well problems that require applying specific techniques.

To illustrate the basic notions concerning the first course Computer Lab Classes are based on Microsoft Access DBMS and include quizzes and a course project. For the second course Computer Lab Classes are based on Oracle 10g Database Server (or MySQL and PostgreSQL, instead) and include design and implementation of a client-server database application. Assignments may be performed individually or as a team work of up to three members. A course project description is delivered at the end of any Computer Lab Class. All members of a team get the same grade.

The purpose of the first course project is to analyze, specify, design, implement, document and demonstrate an information system application using a relational DBMS. To develop the database schema, students are required to use Conceptual Modeling based on the E-R model. All tables in the relational schema are to be in 3NF. It is expected the system to be implemented with MS Access and demonstrated in the computer lab. With a professor's approval, any other relational DBMS, for example MySQL, may be used instead of MS Access.

The purpose of the second course project is to analyze, specify, design, implement, document and demonstrate an information system application based on relational or object-relational DBMS. To develop database schema, students are required to use Semantic Modeling based on the EER model and UML Class diagrams. The system is to be implemented with Oracle (or MySQL and PostgreSQL) and demonstrated in the computer lab.

Suggested learning ways through course's modules are presented in Fig. 1.

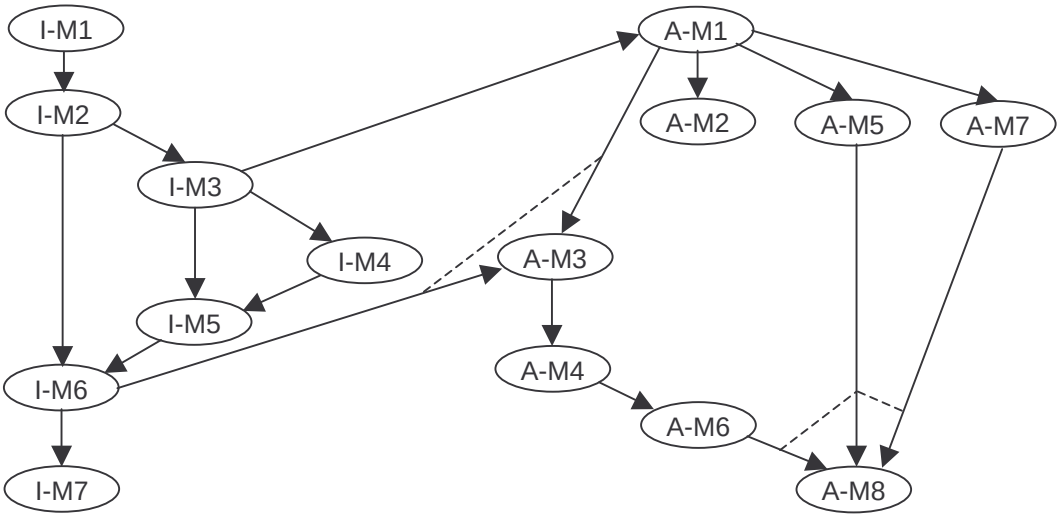

Fig.1. Dependency chart of the course's modules. (I- Introductory, A- Advance, doted lines mark "AND" condition for prerequisites modules)

CONCLUSIONS

Our experience shows that the two courses discussed above are good prerequisites for further learning both in area of information systems and data management. They represent a basis for the following courses: Object-oriented modeling, Client-server information systems, Emerging database technologies and applications, Development of information systems. Knowledge in data management is also required for advanced courses related to concrete DBMS software, e.g. Oracle, MS SQL Server, etc. in masters degree programs.

REFERENCES

[1] Maier D. The Theory of Relational Databases, Computer Science Press, 1983
[2] Ullman J.et al. Database Systems:The Complete Book, Prentice Hall, 2001.
[3] Abiteboul S. et al. Foundations of Databases, Addison Wesley, 1995.
[4] Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.) A Taxonomy of Educational Objectives: Handbook I. Cognitive Domain. New York: David McKay, 1956.
[5] Peneva J. Databases – I part, Regalia, 2004 (in Bulgarian).
[6] Peneva J., Tuparov G. Databases – II part, Regalia 2004 (in Bulgarian).

ABOUT THE AUTORS

Assoc. Prof. Juliana Peneva, PhD, Institute of Mathematics and Informatics – BAS, Dept. of Software Engineering, Acad. G. Bontchev St, bl.8, 1113 Sofia, Bulgaria, Phone ++359 2 979 2879, E-mail: july_peneva@abv.bg

Georgi Tuparov, PhD, Dept. of Computer Science, South West University, 66 Ivan Michailov str., 2700 Blagoevgrad, Bulgaria, Phone ++359 73 8889 132, E-mail: georgett@aix.swu.bg