Teaching databases to hundreds of engineering students

Patrick Lambrix and Lena Strömbäck
Department of Computer and Information Science
Linköpings universitet, Sweden
{patla,lestr}@ida.liu.se

Abstract

Yearly, we teach basic database courses to about two to four hundred engineering students with different backgrounds and different requirements. We also have a limited amount of teacher resources. To deal with this situation we have organized our course topics in the form of modules. A database course is then defined by a number of modules together with a mini-project. We discuss this organization and give an evaluation.

1 Introduction

Every year the teachers in our division teach basic database courses to about two to four hundred engineering students. The students follow the following curricula: computer science, computer engineering, computer engineering short program, engineering biology, chemical biology, industrial engineering and management, and mechanical engineering. The goals and contents of the database courses are decided upon by the boards of education of the different curricula. This culminates in slightly different requirements for the courses.

The students' backgrounds with respect to computer science are also different. The students from the computer science program, the computer engineering program and computer engineering short program have a good background in computer science. The industrial engineering and management, and mechanical engineering students studied a basic programming course and a course on data structures and algorithms, while the engineering biology and chemical biology students studied a basic programming course but no other computer science related courses.

Our task is then to provide basic database courses with slightly different requirements to students with different engineering backgrounds in different periods of the year. As an extra requirement we also have to take into account the limited teacher resources.
Our solution includes defining a number of modules consisting of lectures and laboratory assignments. Most of the modules are general and are used in all the courses. Some of the modules are specialized and are used in one or some of the courses. Each of the modules is the expertise of a few teachers. A course is then defined by a number of modules together with a mini-project. Further, we schedule the courses such that there is an overlap where students from different programs participate in the same modules. We also use the same main course literature for all courses.

In the remainder of this paper we describe the organization of the courses, including descriptions of the different modules and the mini-projects. Further, we provide an evaluation.

2 Course Implementation

Every course is organized in a number of modules (see section 3) and a mini-project (see section 4). In figure 1 we summarize which modules and mini-projects (former mini-projects between parentheses) are part of which courses. We also give the number of students and the number of credits, where one credit is given for one week full-time work. Currently, the computer science and computer engineering students, and the computer engineering students short program take their courses in parallel.

The courses are divided into two parts - a practical part and a theoretical part. We start the courses with the practical part. According to our experience our students are better motivated to tackle the theory behind databases once they have learned about their use. Another advantage is that we can let the students come into contact with a database system early in the course through laboratory assignments. In parallel with the theoretical part the students perform a project using the knowledge gained in the practical part.

Some of the modules contain laboratory assignments. The students perform these laboratory assignments in groups of two. The available equipment is a SunRay per pair of students. The database that is used for the assignments is Oracle. A specific amount of computer

education programs	modules	mini-project	students	credits
computer science and computer engineering	1.1, 2-8	booking system (address book)	70-150	5
computer engineering short program	1.1, 2-6	booking system (address book)	30-60	4
mechanical engineering	1.1, 2-8	booking system (electronic catalog)	<10	5
industrial engineering and management	1.1, 2-8	booking system (address book)	70-100	5
engineering biology and chemical biology	1.2, 2-4, 6, 8-9	bioinformatics	45-70	4

Figure 1: Different database courses.

time is scheduled during day time for each assignment. However, the students have also access to the computers during evenings and weekends. There are also a number of on-line resources on the course web page, including general information about the set-up of the laboratory system, an Oracle manual and an FAQ. This on-line information is shared by all the courses.

The examination in the courses is based on the laboratory assignments, the project work and a written exam.

3 Modules

In this section we describe the different modules. The IM symbols below refer to the Information Management topics in the Computing Curricula from IEEE and ACM [1] that we cover. We observe that according to the Computing Curricula only IM1 and IM2 are considered to be core topics.

module 1.1 This module includes the introduction lecture where databases and the people involved with databases are introduced. It also gives a motivation for why and when databases are used. Further, an introduction to database components and database architecture is given (IM1). We schedule 2 hours of lectures.

module 1.2 This module is an alternative introduction lecture, which is only for the engineering biology and chemical biology students. The focus is on introducing databases for biological data, which fits more with their curriculum than a general introduction to databases. We schedule 2 hours of lectures.

module 2 In this general module we study data modeling using Entity-Relationship (ER) and Extended Entity-Relationship (EER) diagrams. Further, the relational model and how to map EER-diagrams to the relational model are discussed (IM2). We schedule 4 hours of lectures. This module also contains a laboratory assignment. The aim of the assignment is to let the students practice data modeling using EER and database design. The assignment involves extending a given EER-diagram, translating the new EER-diagram to the relational model and implementing the changes in an Oracle database. For this assignment we schedule 2 hours of computer time, except for the engineering biology and chemical biology students, for whom we schedule 4 hours.

module 3 In this general module we study the management of data in databases, as well as the querying of relational databases. As query language we introduce SQL (IM3). We schedule 2 hours of lectures. In a laboratory assignment the students learn to query and update databases. They are given an EER-diagram and an implementation of the diagram in an Oracle database and are required to write SQL-statements involving queries, updates and views. For this assignment we schedule 2 or 4 hours of computer time, except for the engineering biology and chemical biology students, for whom we schedule 4 hours.

module 4 This general module introduces relational database design as well as functional dependencies and normalization of databases (IM4). We schedule 2 hours of lectures. Connected to the lecture is an assignment or an exercise session that allows the students to gain a better understanding of the issues of dependencies and normalization of databases. No computer time is scheduled.

module 5 This module aims to give a good understanding of how to use SQL assertion checks, triggers and stored procedures to maintain database integrity and to support applications. The module consists of a lecture (2 hours) and a laboratory assignment (4 hours of computer time).

module 6 This general module is the first theoretical module. We introduce data structures for databases including hashing, indexing and B-trees (IM8). We also look at the physical database. We schedule 2 or 4 hours of lectures.

module 7 The topics of this theoretical module are the relational algebra and query optimization (IM2). We schedule 2 hours of lectures.

module 8 In this module the issues of transactions, concurrency and database recovery are studied (IM5). We schedule 4 hours of lectures.

module 9 In this specialized module special attention is given to issues related to biological information sources. As many of today's resources are implemented using flat files and some relational and object-oriented databases, we look at notions in information retrieval, semi-structured data and object-oriented databases. We also briefly look at the issues and problems concerning providing integrated access to multiple information sources. We schedule 4 hours of lectures.

4 Projects

In this section we present the different projects that we used during the last five years. Some projects were used during the whole period, while others only for part of the time. We discuss the changes in section 5. We schedule 10 hours of computer time, but these are not supervised.

4.1 Booking system

The goal of this project is to give the students more training on how to use a relational database. The students design and implement a database. In contrast to the other projects, the implementation also involves assertions checks, triggers and stored procedure to maintain database consistency and integrity. The final implementation is then used to get a practical hands-on experience with multi-user environments and transactions. The project is done in groups of two students. The project is run on Oracle on a central server.

The task for the students is to design an airline database where customers can book tickets. The actual project is divided into three parts. The first part consists of studying the requirement specification for the airline database and based on this design an EER model and translate it to a relational database. As the database is quite advanced the students are required to hand in and get their solution accepted by the teacher before continuing.

In the second part of the project the students implement their database, i.e the tables and a set of stored procedures to manage the functionality of the database. They are also required to fill the database with data. Indexes and security issues are not implemented but need to be discussed by the students.

The third part of this project consists of using the database. This involves demonstration of their implementation. Further they need to demonstrate that the database works for parallel users. The students are required to simulate two user interactions and study whether their implementation always give a correct behavior. In most cases they will recognize that they will have problems with deadlock or double bookings and are then advised to solve this problem by using lock-table, commit and rollback.

4.2 Address book project

The goals of this project work include letting the students implement a complete application that uses a database. It allows them to practice modeling and database development. It also teaches them issues in the implementation of an application on top of a database. Further, the students come into contact with a different database management system (DBMS) than the one they used in the laboratory assignments. Also the interface to the database is different. While a command-line interface was used in the laboratory assignments, here a graphical interface is used. Finally, the students learn the characteristics of different DBMSs and learn which characteristics fit for which kind of applications. The project is done in groups of two students. The available equipment is a PC with Microsoft Access installed for each pair of students. We provide the students with a small tutorial on the use of Access.

The project is divided into two parts. In the first part the students create a database with an interface. The application is an address book that contains information about persons, addresses, phone numbers, e-mail addresses and interest groups. The operations that need to be defined on the database deal with adding, deleting and updating as well as displaying information about these different entities. For instance, it should be possible to add and remove people from interest groups as well as to show a list of the names of the members of the interest groups together with their contact information. Further, there are a number of restrictions. For this project the students need to model the domain using an EER-diagram and translate the EER-diagram to a database schema. Then, they design and create a user interface and produce the relevant queries for the different operations. The students are required to hand in their EER-diagrams, a description of the schema and a description of the interface. Further, the solution is discussed with a teacher.

In the second part of the project the students are required to answer a set of questions which compare Oracle and Access in terms of their functionality. First, we ask the students to identify the similarities and differences between the two systems The questions need to be answered for both DBMSs. Then we ask the students to discuss which database they would use for a number of applications. For each application they are required to motivate their choice based on the requirements of the applications and the characteristics of the DBMSs. Also these solutions are discussed with a teacher.

4.3 Electronic catalog project

This project is targeted to the mechanical engineering students. The goals of this project work include letting the students implement a complete application that uses a database in a relevant setting. It allows them to practice modeling and database development as well as it teaches them implementing an application on top of a database. The available equipment is the same as during the lab assignments.

In this project the students implement an electronic product catalog for cutting tools. The students receive an electronic version of a product catalog for cutting tools that they have been using in other courses in the curriculum before. In addition to this catalog they receive a guide that describes which cutting tools should be used in which situations based on, for instance, the characteristics and the thickness of the material that needs to be cut or the shape of the cut surface. For instance, this guide may be used to find out what methods and tools to use to cut a round hole of 10 cm diameter in an aluminum plate. The students are required to model the domain using EER-diagrams, translate the EER-diagrams into a schema and implement a database using the schema and the information in the electronic product catalog. Further, the students need to define interesting questions that can be asked to the database based on the domain knowledge in the guide. These questions need to be described in SQL and embedded in functions that can be used by an end-user. The students are required to hand in their EER-diagrams, a description of the schema and a description of the SQL queries. Further, the system is demonstrated and the solution is discussed with a teacher.

4.4 Bioinformatics project

This project is targeted to the students in engineering biology and chemical biology. In the project the students study a particular topic in bioinformatics, namely modeling of biological data and implementation of biological databases. The available equipment is the same as during the lab assignments. There are many biological data sources on the world wide web that researchers use on daily basis. In the first part of the project the students study in groups of two a number of biological data sources. These sources are well-known and they are among the most used sources in bioinformatics. Links to these data sources are provided on the course home page. We have divided the task into four sub-projects: genome databases, protein databases, enzyme system databases and biological control systems databases. Each pair of students performs one sub-project. The students investigate the information in the data sources and create a data model using EER-diagrams. In addition to the databases on the web the students also obtain a minimal set of information that needs to be modeled as well as a number of queries that one should be able to answer using the modeled information. For instance, typical questions for the genome database sub-project are ‘to which organism does a particular gene belong?’ and ‘which control systems exist involving a particular gene and if there are any, are other genes involved?’. The fact that biological data is recognized to be complex makes this part of the work an excellent exercise and particularly interesting.

Further, we create project groups of eight students consisting of four sub-project groups of two students, one for each sub-project. Within these project groups the students are required to show how the information in the different EER-diagrams can be integrated. To their aid we give the students a number of complex questions using information from different sub-projects that need to be answered. One of the approaches to tackle these integration issues is the view-driven approach where the underlying schemas are integrated to form a global schema. The global schema is then queried in a high-level language. In a sense the students are doing the first step in this approach and in doing so obtain insights in some of the problems of managing multiple heterogeneous databases. At the same time they see a clear connection between what they are studying and current research in the area of bioinformatics.
At the end of this part of the project the project groups and teachers meet and discuss the EER-diagrams as well as integration issues.

In the second part of the project the students transform their EER-diagrams into database schemas and implement their schemas in an Oracle database. They do this in the sub-project groups. We require that real data that they obtain from the data sources on the web is stored in the databases. Then, we require them to turn the databases into project databases by connecting the databases from the different sub-project groups and to use the databases to give answers to our example queries. They hand in the schemas, the SQL queries and the results of the queries. Finally, they demonstrate their systems to the teachers and the other students and explain the choices they have made for the implementation. The demonstration is followed by a discussion of the solutions. Also other issues, such as often occurring mistakes or good programming practice examples from the projects, are discussed.

5 Evaluation

5.1 Method

Our discussion is based on a number of different evaluations including Muddy Cards, web evaluations, the students' own evaluations and the teachers' experiences during the last five years.

The Muddy Cards evaluation method was introduced to us by Donna Qualters, University of Massachusetts Medical School. This evaluation is done a few weeks after the start of the course. In this evaluation we give the students a small card where they can write their opinions, comments and questions about the course. They can take as much time as they want but usually not more than 15 minutes are used. We do not give any hints or questions we would like to see answered. The main reason for this kind of evaluation is to have a quick check on how the course is going. If things need to be changed, we can still do it during the running of the course. The evaluation is scheduled during lecture time. Therefore, all students have the opportunity to participate in the evaluation and it also means that we get the opinions of most students. After the evaluation the teachers read the cards, summarize and write comments that are put on the course web page. Most of the students participated in this evaluation.

For some of the education programs the students perform their own evaluation. For instance, the engineering biology students perform an evaluation immediately after the last lecture. They group themselves into smaller groups and discuss a number of aspects concerning the course including the content and quality of the lectures, laboratory assignments, project work, course literature and the goal and place of the course within the whole curriculum. Other students may have slightly different variants. After the discussion the student representatives discuss the course with the course leader and send the result of their evaluation to the education board of their curriculum. It is hard to estimate how many students participate in these evaluations. For some groups most of the students participated in this evaluation, but for other groups this may not be the case.

The web evaluation is an individual evaluation that is done up to one month after the course ends. The students can fill out a web form where a number of predefined questions are available, but where they also can comment in free text. On average, between 45% (engineering biology) and 33% (computer science and computer engineering) of the students participated in this evaluation.

5.2 Results and Discussion

General comments In general the students liked the courses. According to the student evaluations they found the courses relevant and important in their curricula. With respect to the web evaluations the engineering biology students liked the course best, while the computer engineers short program liked the course the least. To alleviate the latter problem we have changed the course by making the course more practical and replacing some modules.

In the muddy card and student evaluations it was mentioned often that the lectures were good and interesting. It was stated that the laboratory assignments were well-formulated, meaningful and on the right level. Also the information on the web (slides, a reading guide, an FAQ, some summaries etc.) was appreciated.

The examination results of the students were satisfactory in all curricula. The most positive results were obtained by the engineering biology students. All students passed the laboratory assignments and project work in time and over 85% of the students passed the written exam at the first time. In general the throughput of the students in the courses is over 80% per year.

Comments on the organization The fact that different teachers are specialized in (overlapping) areas, results in the fact that the students come into contact with several lecturers. Most students find this to be positive and inspiring. Some commented on the fact that this also means that there may be different teaching styles which may lead to a preference of some teachers over others. A disadvantage of having many different teachers may be that the students may have a difficulty in understanding the connections between the different lectures. To alleviate this problem we now use the same 'course map' in the beginning of each lecture. This is used to introduce the new topic in the context of the course and relationships with other lectures are taken up. This has helped us in connecting the lectures together in a better way and allowed the students to orient themselves in the course. It is also important that the teachers communicate with each other about what happened during the lectures and for instance, which questions were asked.

An advantage for the course organizers is that the courses are not that dependent on specific persons. In the case of absence of a teacher, another teacher is usually well-prepared to take over a lecture.

The combination of laboratory assignments, project and written examination as a way to test the understanding of the course was considered to be good by many students.

Other comments in the student and muddy card evaluations included requests by some students for more supervised time during the laboratory assignments and projects. Some students also would like to have exercise sessions in smaller groups. Further, several students asked for more information about the differences between Oracle SQL and standard SQL. A small document on this concerning the part of Oracle SQL that is used during the labs, was produced.

Comments on the projects The Access project was by most students considered good and educating. However, despite the addition of a tutorial session and extra material teaching Access, there were also many students that had problems in managing Access in this short time. Some opinions were raised that more time was spent on Access details than understanding databases. Also the assistants felt that it was hard to help the students in learning Access in a good way.

This led to the development of the new booking system project in Oracle. The change allowed more advanced modeling, implementation and SQL and also gives a practical demonstration of transactions. The new project demands more work by the students than the old, but since they work on a system they already know the project has been very positively received by the students. Also the teachers feel that the new project is easier to supervise than the old one. We also see the connection to the theoretical lectures as a benefit for the course.

The electronic catalog project was not considered to be very challenging by the mechanical engineering students, although it was relevant to their area and based on existing catalogs. Also the teachers found that the resulting EER-diagrams and schemas were too simple and a number of advanced constructs were not exemplified in the project. This, together with the small number of students has led to the fact that the mechanical engineering students have joined the industrial engineering and management students' course.

The bioinformatics project was considered to be a success by both teachers and students. The students found the project very interesting and appreciated the fact that it had clear connections to their previous courses in biochemistry and gene technology. They wrote that it was fun to see the connection to biology and biological applications of databases. The size of the project and the amount of work required was also considered adequate. It was stated that the students had also learned new things in gene technology. They wrote that there were issues in gene technology that they thought they had understood before, but the modeling in the project required them to discuss and study the issues in more depth than they had done before. It was also realized that modeling biological data is a hard task which conformed to our expectations.

For the courses with over 60 students, comments on the EER-diagrams were given via the documents the students had handed in. In the other courses special sessions were organized to discuss the EER-diagrams in group. This second group of students found the discussions very valuable. They found that discussing each other's diagrams allowed them to understand several issues that were not clear before. Later also the students in courses with over 60 students received the possibility to discuss their EER-diagrams in more details with the teachers.

Some background knowledge is required to define relevant projects in the areas of the students' curricula. The bioinformatics project was developed by former engineering biology students with knowledge of databases under the guidance of the database teacher. Also a researcher in cell biology was asked to give comments on the result. The electronic catalog was developed by a database teacher after discussions with a teacher from the department of mechanical engineering. Also for the teaching connected to the specialized projects some extra work needs to be done by the teachers to become familiar with the most common terms in the areas. The main situations where this background is needed is when discussing the EER-diagrams with the project groups. However, according to our experience, it is possible to rely on the domain knowledge of the students to a large extent and it is rather easy for an experienced database teacher to ask the right data modeling questions for the parts for which the domain knowledge is lacking.

Acknowledgments

Several teachers have contributed to the development and further development of the courses, as well as to the implementation of the module system. In addition to the authors, the teachers during the later years were Almut Herzog, Cécile Åberg, Juha Takinnen, Vaida Jakonienė, He Tan, José Pena and David Hall.

References

[1] The Joint Task Force on Computing Curricula, IEEE Computer Society, Association for Computing Machinery, 'Computing Curricula 2001', Draft March 2000.