This Festschrift is dedicated to Martti Laiho, the “heart and soul” of DBTechNet on the occasion of fifteen years of DBTechNet!
Contents

Foreword........................................................................................................................................................................... 4
Motivation and Mission......................................................................................................................................................... 5
Dedication for Excellence in Database Teaching........................................................... 8
European DBTechNet Projects................................................................................................................................. 12
Concurrency Control Phenomena.......................................................................................................................... 14
Knowledge Discovery from Databases.................................................................................................................. 24
Semantics and Databases........................................................................................................................................... 42
Outlook............................................................................................................................................................................ 48
DBTech Humor............................................................................................................................................................. 51
Timeline.......................................................................................................................................................................... 52
Authors............................................................................................................................................................................. 54
References........................................................................................................................................................................ 56

DBTechNet

DBTechNet is an initiative of European universities and IT-companies. It gathered database academics, professionals, and trainers to set up a transnational collaboration scheme in order to improve database teaching by meeting the needs of industry and to prepare the European Information Society for the new millennium.

Imprint

December 2012

DBTechNet
c/o Fritz Laux
Reutlingen University
Alteburgstr. 150
D-72762 Reutlingen
GERMANY

Telephone: +49-7121-271 4001
E-Mail: fritz.laux@reutlingen-university.de
I applaud DBTechNet and its efforts toward excellence in database teaching. Over the past several years I’ve increasingly encountered the perception that “database isn’t interesting any more” or (worse) that it’s “not very important any more.” Certainly it seems to be getting harder and harder to attract good students to the subject, a state of affairs I find frustrating for several reasons. For one thing, it’s certainly not the case that all of the theoretical problems have been solved. For another, even where they have been, commercial products all too often fail to conform to the known correct solution. And for another, I believe strongly that database technology is important, and I think it’s interesting, too.

That said, perhaps I might be allowed to offer my own opinion on the subject of teaching in this field. My own mantra for many years has been that theory is practical. Thus, I sincerely hope that “excellence in database teaching” doesn’t mean sacrificing theory on the altar of commercial practice. If you don’t get the foundations right, you can forget the rest. And by “foundations” here, what I really mean is the relational model. To me, excellence in database teaching includes — but of course is not limited to — (a) recognizing that the relational model is the foundation for everything else; (b) recognizing in particular that SQL and the relational model aren’t the same thing; and (c) recognizing that if SQL is to be taught at all, then it should be done in terms of the relational model and not the other way around.
Motivation and Mission
by Fritz Laux

The increasing importance of databases has not lead to an appropriate consideration in the curricula nor did it converge to a more professional practice. The functioning and reliability of those systems are carelessly taken for granted. From time to time there are reports about malfunctioning of software that have lead to large financial and sometimes personal damage. For example, a high bay warehouse had to be emptied and reloaded for three month because the inventory software “forgot” the locations of the materials. Million Euro damage resulted from incorrect money transfer software and a German mobile phone network was paralyzed when the software crashed [2]. On November 1st, 2005, the Tokyo stock exchange had to close down in emergency for the first time since its opening due to a software crash. The software expert Les Hatton estimates the annual economic damage from malfunctioning software between 100 and 150 billion € annually [3]. All these information processing systems heavily rely on database software.

The following questions arise: How can database education meet the increasing demands in an efficient and sustainable way – despite the growing complexity of information technology? And how do we cope with the Bologna Process for standardized and modularized European university and professional degrees?

In 1997 some lecturers from former Helia Business Polytechnic (now Haaga-Helia University of Applied Sciences) first formulated the demand for high quality teaching on a European scale. They gathered pro-

DBTechNet
DBTechNet is dedicated to achieve a three-fold goal, namely:

- Develop a reference framework for database teaching
- Define and set the standards for database education
- Develop didactically proven e-learning course modules that address the needs of industry including “out-of-the-box” hands-on labs
Information and Society

Knowledge is the key success factor in our so-called “Information Society”. This finding is widely supported by scientists, politicians, and entrepreneurs. The most prominent representative may be Bill Gates who stated: „How you gather, manage, and use information will determine whether you win or lose” [1]. Database systems are at the core of every information system. Operational systems, like materials and production planning, web based e-commerce, or business-to-business applications fully depend on the reliability and high-availability of database systems. With these systems we accumulate constantly new data that is impossible to manage manually. As early as in 1985, the Yale professor Rutherford D. Rogers noted: “We’re drowning in information and starving for knowledge”. With the help of special data warehouse systems, professors, managers and lecturers from all over Europe in order to pursue intensively the questions mentioned above. This initiative, later called DBTechNet, set up a transnational collaboration scheme to improve database teaching in order to prepare the European Information Society for the new millennium. (see sidebar)

Computer science professionals are already present on a global labour market and work in international teams. To ensure high competitive qualifications in database technology the graduates need the best possible education that provide for the key competences needed in the future. Teachers and students tend to prioritize theory and sustainable conceptual knowledge in the curricula. From the industry’s perspective the database competence profile is mostly gui-
ded by the actual demands for directly applicable skills. In fact, computer science students and experts need a good foundation of database design theory, but also technical and practical skills for the development and mastery of highly complex systems. In a rapid changing technology this is an indispensable asset. Given the shorter and regulated study programs due to the Bologna Process bring teachers in conflict to choose between theory and practice. What are the consequences of this field of tension and how should the study of computer science be affected and adapted? DBTechNet is committed to address this challenge and tries to harmonize theory with practice, say the interests of our students with the demands from industry. The methods and didactical approach to reach this dual goal are grounded in constructionist learning theory and use a mixture of multimedia supported teaching and problem based learning concepts which we call “studying and verifying”. As result the DBTechNet consortium expects an increase of the market attractiveness for computer science graduates from our institutions. Feedback from our graduates confirms the expectations and report that the quality and sustainability of the DBTechNet curriculum prepares well for a professional career. And, last but not least, we believe that this initiative contributes to harmonize and improve the standard of database education in the European Union. The following articles give some examples of DBTechNet learning modules and explain how theory and practice are harmonized and how sustainable knowledge together with problem solving competency can be acquired.

these data floods can be analysed in seconds, e.g. in order to better understand customer behaviour or to assess companies in terms of their performance. Databases are essential components of these business intelligence systems. The “mining” for hidden knowledge (data mining), the aggregation and refining of information provide managers with important arguments for decision making and lead to competitive advantages for enterprises.
**Didactical Challenge for Database Teaching**

Database teaching that responds to the needs of our “Information Society” has to take in account the following requirements:

- Tailor content into modules
- Address technological reality
- Teach theoretical underpinning
- Use didactical concepts for sustainable learning
- Keep the motivation high

**Dedication for Excellence in Database Teaching by Fritz Laux**

Effective knowledge transfer at Higher Education (HE) institutions and at Vocational Educational Training (VET) centres should be tailored to the needs of its clients. Employees are highly motivated to acquire new skills but are often hindered to follow a scheduled training programme. Students face a denser curriculum with a high degree of optional courses whose schedules and prerequisites are not aligned. Therefore it is essential to provide self-study courses with small module sizes to enable the participants to learn in their spare time at an individual pace. In addition, in financially difficult times, knowledge transfer should be highly scalable in terms of costs. E-Learning offers this capability but lacks the means of keeping motivation high.

As consequence, e-learning has to solve a multidimensional problem:

1. The learning content needs to be chunked into „digestible“ portions while keeping the necessary context (see Figure 1)

2. Technological reality and the theoretical underpinning need to be harmonized

![Figure 1: Database Context and Subject Areas as Basis for Learning Units](image)
3. The learning should be guided by didactical concepts that ensure sustainable knowledge.

4. Keep the motivation high despite the impersonal learning environment.

The last requirement is of particular importance for e-learning as the motivation and support provided by face-to-face learning is missing. Therefore, DBTechNet aims to use the best suited learning concepts for each particular topic and to use commercial products found in industry to empower students and employees for a competitive labour market. This will stimulate the secondary motivation of the learners.

The reason for different didactical concepts for learning will be explained by the following two examples. One crucial competence within the database education is how to design a database. The adequate structuring is a key factor for a well-designed, flexible, high-performance information system. This structuring not only requires a sound knowledge about the theory of the data model behind the database management system but also needs to elicit the requirements and to understand the semantics of the data. These are analytical and creative abstraction processes which can be best learned from real world problems. The most appropriate learning theory for this case appears to be a constructivist-based approach because it naturally emphasizes abstraction and synthesis to construct new knowledge. Project based learning is a way to encompass most aspects of constructivism theory. These are cognitive, social, and emotional aspects resulting from the successful outcome of the project that generate personally meaningful and transferable knowledge. For practical implementation the students form teams of 3-6 persons. Each team presents its design in a plenary session and discusses the result with fellow students and their instructor.

The other example deals with the processing of banking data. The database has to guarantee correctness and reliability of the financial transaction under any circumstances. This includes situations such as the correct processing of concurrent transactions and resilience to hard- and software failures. Usually the commercial base technology upon which an application is built does not provide these reliably per se. It depends on the developer to work around the shortcomings and design effective applications. Such a highly specialized knowledge can neither be taught on purely theoretical level nor can it be trained based on cookbook-style scenarios. This is the use-case for our e-learning based concept with hands-on labs. In order to illustrate the consequences of possible concurrency problems it is best to let the students experiment various situations with real commercial products.
After experiencing a problem it is their task to find a solution to avoid it. When the learners find it difficult to solve the problem for all possible situations they are more open to the theory which will give the foundation for a comprehensive solution. The best supporting learning theory seems to be based on insight learning as described by the cognitivist theory. Experimental results in different situations promote the insight for a general theory-based solution.

In terms of practical implementation we provide the students with a working laboratory environment. The environment is installed on a virtual computer. The students work in pairs and each one tries to interleaver the transacational process of his opponent in such a way that they experience undesirable effects or even incorrect results.

Taking into account the before mentioned dimensions that challenge the e-learning in the realm of databases, DBTechNet developed a unique framework for teaching database concepts and technology as well as special technological skills by applying different learning models. A central role plays “problem based learning” with hands-on labs that encourage experimenting and verifying the theory. We believe it is best to decide from the learning content, which learning concept will be best suited for a specific content. The e-learning model shown in Figure 2 integrates the following learning concepts (see Issing [7]):

- Learning as behavioural modification for practical skills and verification of the theory
- Learning as active information processing using assimilation
and accommodation processes to build a mental model of the *theory*.

Learning as construction of knowledge for problem based learning as in *project work*

All these concepts are used in an integrative way in order to get the most effective results in terms of applicable knowledge and profound cognition that enable abstraction and problem solving to a large extent.

The design of the e-learning model starts with structuring the learning units (see Figure 1) guided by a taxonomy developed by DBTechNet. The learning area is sliced by minimizing the dependencies and each chunk of learning content is represented in a theory unit with examples and demonstrations of the theory (active, insight based learning). Hands-on experiments help to verify the theory (instructional learning). Given an authentic real world problem the students need to apply and practice their full knowledge and skills in order to analyse the problem, design a solution and implement it for verification (constructivist learning). The global optimization task is to put together all aspects in balance with the target learning group.

From a didactical perspective the benefit of this model may be summarized as follows:

Examples and demonstrations explain the theory, making it easier to understand. Hands-on experiments motivate and stimulate students to reflect the theory. Examples provide the students with analogous situations that could be applied and abstracted in the project work. The interrelation of all these elements as provided in a virtual lab environment together with theory units and examples is shown in Figure 2. The concrete real world problem forces the students to abstract from examples and construct a model of the problem world in order to find a solution.

---

**Learning by Doing and Verifying**

Aristotle already promoted „learning by doing“ in his eminent work on ethics, the Nicomachean Ethics [4]. The concept became known in pedagogy through the work of Comenius [5]. From the perspective of developmental biology “learning by doing” is known even from animals [6] and experimenting (the systematic learning by doing) is fundamental in the development of the homo sapiens in the sense of Jean Piaget and David Kolb. The explorative nature of humans is essential for experimental science.
European DBTech Projects

From 2000—2003 and 2008-2011 the European Commission supported within the Leonardo-da-Vinci Programme resp. the Life-long-Learning Programme two DBTech Projects. During the first project, called DBTech Pro, the partners conducted a survey on the IT needs in Europe, developed a reference curriculum reflecting the outcome of the survey, and organized conferences and workshops to prove the didactical concepts and promote its use.

In the second project, DBTech EXT, a range of teaching units has been developed. The reference framework was extended towards a taxonomy that takes into account different levels of competency as defined by Bloom's knowledge taxonomy. A couple of “out-of-the-box” virtual hands-on labs have been developed for experimenting with commercial and open source products and for verifying the theory.

Recently (July 2012) the European Commission has approved our third successor project DBTech VET. This project focuses on Vocational Educational Training (VET). The aims are to transfer the DBTech knowledge to VET teachers and transform example courses to a format that is suitable for professional training. It is a constant challenge in a rapidly developing information society that ICT employees stay up to date with technologies.

The on-going project operations are successfully organized through virtual meetings by means of internet-based collaboration tools, and face-to-face meetings at partner sites. During the projects lifetime the partners organised and hosted Work-
shops for students and participants from industry to test and validate the course modules. The workshops consisted of lectures on important topics like transaction processing, concurrency control, database modelling, database administration and tuning, data warehousing, data mining, semantic modelling, XML-, object-oriented, and distributed database systems. Valuable feedback was collected from the participants and evaluators that lead to improved versions of the modules.

Conferences with leading scientists and industrial experts ensure that DBTech remains in constant touch with the latest achievements in science and technology. To mention only the latest DBTech Days in May 2012 in Helsinki the focus was on Temporal Databases complemented with lab exercises using the latest version of DB2 LUW 10.1 that supports temporal aspects. Highlight of the conference was the second seminar day with Chris Date on relational theory and practice.

DBTech members itself are frequent participants and presenters at international conferences all over the world. Nevertheless, the main concern of DBTech is the practicability and relevance of theory for real world problems. This is why DBTech academics are consulting, leading, and evaluating industrial projects as part of their academic work.

All this experience has flown into the development of course material used for face-to-face and e-learning. The examples of teaching material and lab instructions on the following pages stem from learning units that have been produced during the second project from 2008 – 2011.
Concurrency Control Phenomena
By Martti Laiho & Fritz Laux

The first example course is taken from the Concurrency Control learning unit that comprises

- Pre-teaching material to adjust the pre-knowledge of all learners properly,
- A comprehensive tutorial on concurrency control that covers the topic not only on a conceptual level but compares theory with the implementation of commercial database products,
- Review questions and model answers
- Virtual hands-on lab with self-paced exercises and detailed instructions

Concurrency Control has to coordinate concurrently executing transactions in such a way that the result is correct and each transaction does not need to care about the other transaction. The problems that may arise with concurrent transactions and how they can be avoided are the topic of this tutorial. Here is a sample describing the “Lost update problem” and the “Uncommitted dependency problem”:

<table>
<thead>
<tr>
<th>step</th>
<th>Process A</th>
<th>Accounts balance of id 12345</th>
<th>Process B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id := 12345;</td>
<td>1000 €</td>
<td>id := 12345;</td>
</tr>
<tr>
<td>2</td>
<td>SELECT balance INTO :balance FROM Accounts WHERE acctId = :id;</td>
<td></td>
<td>SELECT balance INTO :balance FROM Accounts WHERE acctId = :id;</td>
</tr>
<tr>
<td>3</td>
<td>newBalance := balance - 100;</td>
<td></td>
<td>newBalance := balance - 200;</td>
</tr>
<tr>
<td>4</td>
<td>UPDATE Accounts SET balance := newBalance WHERE acctId = :id;</td>
<td>900 €</td>
<td>UPDATE Accounts SET balance := newBalance WHERE acctId = :id;</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>800 €</td>
</tr>
</tbody>
</table>

Figure 3: A Lost Update Problem
Concurrency Control

The concurrent or interleaved execution of transactions is considered as correct if the result is the same as a sequential execution. In this case we call the execution as “serializable”.

If concurrent or interleaved operations work on the same data, incorrect results may arise. Concurrency control deals with the coordination of data access in such a way that erroneous results do not occur.

Some Problematic Concurrency Scenarios

If we don't have proper concurrency control as a service of the DBMS, the concurrent SQL transactions in a multi-user database may generate uncorrect content in the database. The following problematic behavior scenarios are widely used as examples of possible concurrency problems, for example in the textbook by Chris Date [9]

- The lost update problem
- The uncommitted dependency problem
- The inconsistent analysis problem.

The lost update problem can be described using the following kind of scenario of two concurrent processes accessing the same row (account id 12345 in table Accounts), which we present using embedded SQL and a pseudo programming language in Figure 3.

The update made by process A in step 6 gets lost in step 8 when process B overwrites it without even checking for potential concurrent updates after step 3. This kind of update is called Blind Write or Dirty Write. In this scenario the last writer wins, which is not an acceptable business rule, for example, in bank transactions, or in reserving seats to a concert or to a flight. This scenario is possible, when the processes don't use transactions or if they use auto-commit mode, in which every SQL command is committed automatically. However, if the processes use SQL transactions, then this lost update, in terms of overwriting an update, is not possible in real DBMS products as long as transaction A is active. However, after commit of the transaction, the data can be overwritten by other transactions. In our RVV Paper [10], we will cover the problem of lost committed updates and the needed programming discipline to avoid the blind writing.

The uncommitted dependency problem arises, for example, when a transaction A accesses data updated by some concurrent transaction B before transaction B gets committed like in the following scenario in Figure 4 where the transaction B actually ends with a ROLLBACK command.
The read operation by transaction A, without concurrency control protection by DBMS, is also called as Dirty Read. The application of transaction A has seen data which was never officially true, and will act depending on it, perhaps writing erroneous data to the database.

**Isolation Levels**

Without considering the implementation techniques of transaction isolation, the ISO SQL standard defines the following 4 isolation levels in terms of anomalies (badly behaving transaction types such as Dirty Read, Unrepeatable Read, and Phantom Read) against which the selected isolation level gives protection for the transaction. We refer to the basic SQL course on explanations of these and we just present the summary of effects of these transaction isolation levels in table 1. The isolation levels of ISO SQL standard don’t cover all isolation level implementations in real DBMS products, not all products implement even all these isolation levels, and for example Oracle calls its snapshot isolation (SI) as SERIALIZABLE, which is misleading. These differences with the ISO SQL standard have been discussed by Berenson et al [11].

*Lost Update* in a transaction scope means that some concurrent transaction writes over an update written by the transaction before the transaction commits. According to the ISO SQL standard, a Read Uncommitted transaction cannot write, which would imply that Lost Update is not possible.

<table>
<thead>
<tr>
<th>Anomaly:</th>
<th>Lost Update</th>
<th>Dirty Read</th>
<th>Unrepeatable Read</th>
<th>Phantom Read</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Isolation Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read Uncommitted</td>
<td>not possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>Read Committed</td>
<td>not possible</td>
<td>not possible</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>Repeatable Read</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>possible</td>
</tr>
<tr>
<td>Serializable</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
</tr>
</tbody>
</table>

Table 1: Concurrency anomalies solved by the ISO SQL transaction isolation levels
Isolation Levels

The degree of separation of concurrent transactions is called "Isolation Level". The separation is measured in terms of undesired results or anomalies. A fully "isolated" situation is called "serializable". The results for a serializable execution are the same as if all transactions have been executed one after the other (serial). Weaker isolation may result in effects as Lost Update, Uncommitted Dependency (Dirty Read), and others.

Figure 4: Uncommitted Dependency

<table>
<thead>
<tr>
<th>step</th>
<th>Process A</th>
<th>Accounts balance of id 12345</th>
<th>Process B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id := 12345;</td>
<td>1000 €</td>
<td>id := 12345;</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>UPDATE Accounts set balance = balance + 1000 WHERE acctId = :id;</td>
</tr>
<tr>
<td>3</td>
<td>SELECT balance INTO :balance FROM Accounts WHERE acctID = :id;</td>
<td>2000 €</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>ROLLBACK;</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PRINT 'Great, now I can spend 2000 €';</td>
<td>1000 €</td>
<td></td>
</tr>
</tbody>
</table>

European Project work in Málaga
However, the mainstream DBMS products, which we are studying, do allow writing, but as we will see later, the three types of concurrency control mechanisms used in modern DBMS products also prevent the lost update anomaly in scope of the current transaction:

- The locking scheme systems keep the write locks (X-locks) up to the end of transaction, thus preventing the lost update.
- In multi-versioning systems the first one of concurrent transactions to update or delete a row wins (an update action writes always a new row version), whereas competing actions by concurrent transactions will fail, so no update will be lost.
- In the genuine optimistic concurrency control the updates are written to the database only as atomic operation at the commit phase, so no updates can be lost before the transaction commits.

Depending on the concurrency control implementation also the semantics of the isolation level may differ. In locking scheme systems concurrent sessions compete and may wait their turn on accessing the single committed truth of the collected data, whereas in multi-versioning systems sessions can access without waiting for the last committed facts (on isolation level Currently Committed, called also CC) or see the content according to a snapshot from the collected history of the data as it was in the beginning of the transaction (on Snapshot Isolation, called also SI). Due to historic reasons many products, for example Oracle, call CC as Read Committed and SI as Serializable.

The default isolation level for transactions according to the ISO SQL standard is Serializable, which usually conforms to the ACID\(^1\) isolation and to ACID\(^1\) if SERIALIZABLE will be the only allowed isolation level. However, the default isolation level of the mainstream database systems is Read Committed or some almost compatible isolation level (such as Cursor Stability of DB2).

\(^1\) For the difference between ACID and ACID, see [13]
Learning by Doing: Experimenting and Verifying
According to the ISO SQL standard the isolation level is one of the three characteristics (modes) of a transaction. The modes are defined by the following command syntax:

```
SET TRANSACTION <mode> [,{<mode>}..]
```

where

```
<mode> :=
<isolation level>
| <transaction access mode>
| <diagnostics size>
```

The `<isolation level>` is declared by:

```
ISOLATION LEVEL <level>
```

where

```
<level> := READ UNCOMMITTED
| READ COMMITTED
| REPEATABLE READ
| SERIALIZABLE
```

Another transaction characteristic is the access mode, which is by default READ WRITE, but can also be declared as READ ONLY. According to the standard Read Uncommitted implies READ ONLY access mode for the transaction.

The `SET TRANSACTION` command shall be executed before start of the transaction, and according to Melton ([12], p 513) only for the next transaction. According to ISO SQL standard SQL-92, a transaction starts implicitly when the first SQL command - which can be included in a transaction - is entered after an SQL session starts or after COMMIT or ROLLBACK of a previous transaction. The later standard SQL:1999 defines also an explicit transaction start command:

```
START TRANSACTION <mode> [,{<mode>}..]
```

which can include declarations of the access mode or isolation level for the transaction.

The two concurrency phenomena (lost update, uncommitted dependency) explained above can be verified and experimented during the virtual lab. The following extract shows how the instructions and questions look like for this exercise. After a technical advice how to start the virtual system and hints on the general settings the user receives the following instructions on pages 10 and 11 of the Concurrency Lab 1 (CCLab1) document:
Basic Concurrency Anomalies

For these exercises you need to start another concurrent terminal and establish a new DB2 SQL session in it, see Figure 5. Let's call the first session as Session A and the new one as Session B.

Concurrency Anomalies

As suggested in the didactical concept (see the article “Dedication for Excellence in Database Teaching”) this laboratory exercise should be executed in pairs. In this case each student runs his own SQL command line session and competes with his partner for concurrency anomalies.
Now you can experiment with the following scenarios proceeding according to the step order in the scenario tables, and in case one of the two sessions gets blocked into waiting state, you should take notes of the situation, and then proceed with the other session until the blocked session can proceed again.

Notes:
- In the following commands we have omitted the semicolons as command separators, however, when using Oracle or DB2 Command Editor tools, you need to add semicolons at the end of every SQL command.
- Just for readability, we use lower case letters for the SQL reserved words in the textboxes below.
- We assume implicit transactions in the following scenarios, and we will usually ROLLBACK every transaction to preserve the original content in the table after every experiment. The commands in grey are not part of the essential behavior of the topic, but supplied as assisting means only.
- Let's try concurrent updates first

### Exercise 1a

<table>
<thead>
<tr>
<th>step</th>
<th>Session A</th>
<th>Session B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-- in SQL Server:</td>
<td>-- in SQL Server:</td>
</tr>
<tr>
<td></td>
<td>SET IMPLICIT_TRANSACTIONS ON</td>
<td>SET IMPLICIT_TRANSACTIONS ON</td>
</tr>
<tr>
<td></td>
<td>-- in MYSQL:</td>
<td>-- in MYSQL:</td>
</tr>
<tr>
<td></td>
<td>SET AUTOCOMMIT=0</td>
<td>SET AUTOCOMMIT=0</td>
</tr>
<tr>
<td>2</td>
<td>update T set s='value A' where id=1</td>
<td>update T set s='value B' where id=1</td>
</tr>
<tr>
<td>3</td>
<td>-- in SQL Server:</td>
<td>-- in SQL Server:</td>
</tr>
<tr>
<td></td>
<td>SET IMPLICIT_TRANSACTIONS ON</td>
<td>SET IMPLICIT_TRANSACTIONS ON</td>
</tr>
<tr>
<td></td>
<td>-- in MYSQL:</td>
<td>-- in MYSQL:</td>
</tr>
<tr>
<td></td>
<td>SET AUTOCOMMIT=0</td>
<td>SET AUTOCOMMIT=0</td>
</tr>
<tr>
<td>4</td>
<td>select * from T where id=1</td>
<td>update T set s='value B' where id=1</td>
</tr>
<tr>
<td>5</td>
<td>commit</td>
<td>select * from T where id=1</td>
</tr>
<tr>
<td>6</td>
<td>commit</td>
<td>commit</td>
</tr>
<tr>
<td>7</td>
<td>select * from T where id=1</td>
<td>select * from T where id=1</td>
</tr>
<tr>
<td>8</td>
<td>rollback</td>
<td>rollback</td>
</tr>
</tbody>
</table>

Mark here the DBMS and version you are using .................................................................

Do we need to set isolation levels for these transactions?
Which claims in the following lists are true?
1. The result sets in steps 5, 7 and 9 are equal
2. The result sets 5 and 7 are different
3. A lost update problem appears in step 4
4. A blind overwriting appears in step 4
5. A lost update problem appears in step 8
The claim numbers ............................................................... are true.
Please comment the result: .................................................................

**Exercise 1b**
If your DBMS supports LOCK TIMEOUT repeat the test in exercise 1, setting the DBMS specific LOCK_TIMEOUT to 10000 milliseconds (10 seconds) in step 3 for session B.

How does this change the flow of the scenario?

Which claims in the following lists are true?
1. The result sets in steps 5, 7 and 9 are equal
2. The result sets 5 and 7 are different
3. A lost update problem appears in step 4
4. A blind overwriting appears in step 4
5. A lost update problem appears in step 8
Knowledge Discovery from Databases (KDD)
...its all about (gradually) progressing from the Data Information Age into the Information Age
By Dimitris Dervos, Leonidas Karamitopoulos, and G. Evangelidis

Figure 6: A sample of today's data sources, plus an indication of their contribution to world data corpora. [© E. Genard, P. Reeves, IBM DeveloperWorks]
1. Putting KDD Into Perspective

To begin with, data (variables) come in many forms and types:

*Unordered categories*, whereby one can do operations like „X is not Y“, but fails to quantify the comparison (e.g. X > Y), or subtract (Figure 7). Typical data of this sort are „apple“ and „orange“.

*Ranks*, whereby comparison is possible, but subtraction is not (Figure 8). For example people’s names (ranked alphabetically): „John“ comes after (i.e. is greater than) „Christine“, but how about „John“ - „Christine“?

*Intervals*, whereby comparison and subtraction do have meaning, but the ratio cannot be realized: „40° C“ does not imply twice as hot weather to „20° C“.

Finally, *true measures*, like the „age“: a 40-year old is older than a 20-year old person, their age difference is 20 years, and the former has lived twice as long as the latter.
Still, all data share some common characteristics:

- they represent real world facts
- they are quantifiable: the size of a given data corpus remains invariant when ported from one system to another, provided the representation technology remains the same
- they are measured in bits
- the Data Processing Age is definitely here: our everyday activities are shaped, to a great extent, by technology-assisted data modeling, storing, processing, and management operations

In fact, today’s everyday life is overwhelmed by the astronomically large amounts of data pertaining to work, education, communication, and entertainment relating activities (Figure 8). How will the transition to the Information Age feel like?

The following fictitious conversation lies in the sphere of imagination, of course, yet it makes the point with regard to information quantification: “John called me last night and in five minutes of talk he revealed five infotrons worth of information to me, whereas Mary called a bit later and in just three minutes she revealed ten infotrons worth of information on the same topic”[17]

To make possible the extraction of information from large data corpora, today’s Knowledge Discovery from Databases (KDD) technology provides the means to: (a) select/integrate data from disparate data sources, (b) select/identify the target data set of interest, (c) conduct the data cleaning stage whereby duplicate records are removed, typos are corrected, missing values are tidied up, etc., (d) transform and direct the data as input to the Data Mining (DM) processing stage, and (e) facilitate model creation and DM output interpretation (e.g. by means of a visualization end-user interface).
In this respect, the modern DBMS software platform provides a framework for data mining and machine learning algorithms: decision trees, Bayes nets, clustering, time-series analysis, etc. To utilize KDD and to benefit from this technology, one needs to first of all be fully aware of the semantics and the value of his/her data, the types of variables involved, the available DM algorithms, as well as the appropriateness of the latter to the former. Next comes the need to be able to interpret the information representation forms in the output of the DM stage, and to carry-out additional analytical and compositional information extracting (sub)tasks.
2. The DBTechNet KDD Virtual Laboratory Workshop

By completing the KDD VLW, the participant/learner:

- becomes aware of the new reality whereby the DBMS system comprises a framework for conducting data mining operations like: affinity analysis, clustering, classification, etc.
- learns to differentiate “data” from “information"
- comes to appreciate the importance of the data preparation, cleaning, transformation, and loading stages
- develops skills in the direction of planning for and conducting basic data mining tasks for the extraction of useful information from databases
- learns to interpret and evaluate the output of the data mining processing stage, and
- understands the inner workings of a typical recommender system
2.1 Getting Ready for the Workshop

The workshop addresses the topics of Knowledge Discovery from Databases at an introductory level. In this respect, the learner is advised to first familiarize himself/herself with the concepts involved by reading the corresponding section(s) from one or more textbooks of the database technology bibliography, for example:


For a concise treatise on the topics addressed, the book of M.H. Dunham, and the book of R. Roiger and M. Geatz comprise a very good source of reading material:

- Dunham M.H., Data Mining: Introductory and Advanced Topics, Prentice Hall, 2002
- Roiger R., Geatz M., Data Mining: A Tutorial Based Primer, Addison Wesley, 2002

From Data to Information

“One thing for sure: the change will be one of a qualitative as opposed to quantitative nature. One whereby technology will shape everyday human activities in a way where systems: (a) model user interests and preferences, (b) sense the current context of the user, (c) compute online information relevant to the context and to the user’s profile, and (d) pro-actively deliver ‘just-in-time’ information in a subtle, non-intrusive way [16]. Developments of this type are likely to intensify the need for modeling and quantifying information.”

Dimitris A. Dervos
Select point B from the original set of objects
Seeds = \( N(B) = \{A, B, C\} \)
| Seeds | ≥ 3

B is a core point
CL1 = \{A, B, C\} (Create a new cluster with A, B, C as its initial elements)
Seeds = \{C\} : We exclude from this list the current point B.
Also, we do not select A again, since we realize that A is a border point.
Selecting A again will not yield any additional answers

Expand the current cluster (if it is possible) with respect to the Seeds set

Select point C from Seeds
N(C) = \{B, C, D\}
| N(C) | ≥ 3
C is a core point

We evaluate the elements of N(C)
Skip B and C, since they are classified
D is unclassified
Seeds = \{C, D\} (append D to Seeds)
CL1 = \{A, B, C, D\} (add D to the current cluster)
Seeds = \{D\} (delete current point from Seeds)
**Select point D from Seeds**

N(D) = {C, D, E}

| N(D) | ≥ 3

D is a core point

- We evaluate the elements of N(D)
- Skip C and D, since they are classified
- E is unclassified

Seeds = \{D, E\}

CL1 = \{A, B, C, D, E\}

(append E to Seeds)

(append E to the current cluster)

Seeds = \{E\}

(delete current point from Seeds)

**Select point E from Seeds**

N(E) = \{D, E\}

| N(E) | ≥ 3

D is a border point

Seeds = {}  

(delete current point from Seeds)

**End of Cluster Expansion**
Supplementary to the above comes the workshop's educational and training content which is available from the DBTechNet portal at http://dbtech.uom.gr/. The latter consists of (a) tutorial video presentations of the theoretical part of the workshop, (b) video presentations on the use of the software and services used, (c) video recordings of virtual classroom sessions conducted with the use of the Adobe Connect Pro® software, (d) worked examples of exercises on the topics considered (Figure 9), (e) hands-on laboratory self-training material, etc. To obtain a feeling of "what's in there" for the learner, the interested reader is advised to visit: http://dbtech.uom.gr/mod/resource/view.php?id=372.

The learner may check his/her readiness to proceed with the hands-on (virtual) laboratory part of the workshop by going through the Multiple Choice Questions self-assessing quiz provided (Figure 10).

Summarizing on the data mining algorithms addressed in the self-study/self-practicing educational and training content provided, they include the following:

- Affinity (market basket) analysis: tutorial with exercises and model answers, plus hands-on (virtual) laboratory practicing
- BIRCH: tutorial
- Decision Trees (J48): tutorial with exercises and model answers, plus hands-on (virtual) laboratory practicing
- K-Means: tutorial with exercises and model answers, plus hands-on (combined with J48, using clustering to improve the quality in the output J48)
- DBSCAN: exercises with model answers, single- and complete-linkage agglomerative clustering: exercises with model answers
It is more appropriate to construct a classification model from:

- a. the whole dataset available and tested on it
- b. a subset of the whole dataset and tested on the same set
- c. a subset of the whole dataset and tested on the whole dataset
- d. a subset of the whole dataset and tested on the remaining set

Regarding the measures of Precision and Recall for a classification model,

- a. precision can be considered as a measure of completeness, whereas recall can be considered as a measure of exactness or fidelity
- b. the measure of precision is the number of instances that are correctly classified in a given class divided by the total number of instances that are classified in the class
- c. recall is more useful measure than precision
- d. the measure of recall is the number of instances that are classified to a given class divided by the total number of instances in this class

Regarding the construction of a decision tree, the value of information gain for the attribute A

- a. represents the expected reduction in the information requirement caused by knowing the value of A
- b. should be as low as possible in order to define A as the splitting attribute at a node
- c. all of the above
- d. represents the amount of information needed to classify the instances in the resulting partitions
Figure 11: Mining flow design stage in the IBM Infosphere® platform
Figure 12.: Visualization of data mining output (clusters) in WEKA
2.2 Hands-On Laboratory Training

The hands-on laboratory training part of the KDD VLW utilizes the IBM Infosphere® platform (Figure 11) and the WEKA open source data mining software of the University of Waikato (Figure 12, [18]).

VLW participants are instructed to conduct data mining operations along the lines of two case study training sessions: (a) the IBM provided „Managing your Business in Retail“, and (b) a DBTechNet case study that utilizes usage log data from the HEAL-Link consortium’s portal in Greece (http://www.heal-link.gr/enh/).

Regarding the use of the (proprietary) IBM Infosphere® software, in cases where VLW participants are not entitled to taking advantage of the IBM Academic Initiative licensing scheme, use is made of the University of Malaga PC Virtual environment (Figure 13, [19]).

The DBTechNet KDD VLW (freely available) educational and training content at http://dbtech.uom.gr/mod/resource/view.php?id=372 includes a short introductory video on the University of Malaga PC Virtual environment service. The latter makes possible for (registered) remote workshop participants to conduct the IBM software relating part of the hands-on laboratory training part by just using their local web browser software. The PCVirtual web service provides access to the DBTechNet servers via a pool of (persistent) virtual machine clients (VM clones) acces-
Figure 14: Tom connects to the remote DBTechNet server via his UMA PCV virtual machine, over the Internet. Thanks to the University of Malaga PC Virtual service, KDD VLW participants from the industry can benefit from the hands-on laboratory training part that utilizes commercial software (Figures 15 and 16).

sible via all popular web browsers (MS Internet Explorer®, Mozilla Firefox®, Safari®, etc.): Figure 14.
Figure 15: IBM Intelligent Miner association rules output visualization
3. Benefits Summarized

When coupled with other relevant DBTechNet virtual laboratory workshops (e.g. the OLAP and Data Warehousing VLW), the KDD VLW succeeds in introducing the learner/participant to state-of-the-art technologies and concepts relating to today’s developments and IT professional skills. Introductory yet instructive, the DBTechNet KDD VLW educational and training content is subject to continuous updating and improvement. The experience of sharing it across the DBTechNet partnership comprises an invaluable asset for the IT professionals and the academic communities (teachers and students) involved.
connect to <localdbname>user<dbtechnnn>;

DROP TABLE RETAIL.PRODUCT_RECOMMENDATIONS;
CREATE TABLE RETAIL.PRODUCT_RECOMMENDATIONS (  
    TRANSID INTEGER,  
    RULEID INTEGER,  
    RECOMM_ITEM VARCHAR(30),  
    SUPPORT DOUBLE,  
    CONFIDENCE DOUBLE,  
    LIFT DOUBLE,  
);

insert into RETAIL.PRODUCT_RECOMMENDATIONS  
with  
rulebodies as  
(select * from table (IDMMX.DM_getRuleBodies((select MODEL from IDMMX.RuleModels  
where MODELNAME='RETAIL.ASSOC_RULES')))  
as rb  
),  

Figure 16: SQL script for making product recommendations (IBM case study material)
bodysize(bodyid, bodysize) as 
(select bodyid, count(*) from rulebodies group by bodyid
 ),

bodies(bodyid, bodysize) as 
(select rb.bodyid, item, bodysize from rulebodies rb, bodysize bs
where rb.bodyid = bs.bodyid
 ),

bodies_trans(transid, bodyid, bodyitem, bodysize) as
(select TRANSID, bodyid, bodyitem, bodysize
from bodies, RETAIL.TRANSACTIONS
where int(bodyitem) = ITEMID
 ),

bodies_trans_size(transid, bodyid, bodysize, bodytranssize) as
(select TRANSID, bodyid, max(bodysize), count(*) from bodies_trans
  group by bodyid, TRANSID
 ),

select bts.transid, ar.id, varchar(ar.headname, 30), ar.support, ar.confidence, ar.lift
from bodies_trans_size bts, RETAIL.ASSOC_RULES ar
where ar.bodyid = bts.bodyid and bts.bodysize = bts.bodytranssize;
Over the last few years the internet has become a large information repository that is accessed manually in the vast majority of cases. However, the flexibility and openness of the internet in computer systems is lacking when software applications are connected. The main aim of the Semantic Web is to automatically perform tasks in the current content of the Web. This will be done by making explicit the semantics of the contents, thereby providing unambiguous knowledge from Web documents and applications. For any field of knowledge, research on Semantic Web infrastructures and applications can be especially helpful to improve efficiency in finding, collecting and organizing the data stored in a growing number of resources which make their semantics explicit. To address the problem of information management and interchange, the semantic Web community led by the W3C community proposed a set of standards such as the RDF [20] and OWL [21]. Since 2007, there has been a lot of effort made to provide data sets from different areas using Semantic Web technologies. In this context, a set of best practices has been proposed for sharing, publishing and connecting data, information and knowledge by using RDF and URIs. These practices are known as Linked Data principles [22]. The movement towards the publication and linking of data has been continuously growing and the number of triples stored in the Linked Data Cloud (see Figure 17) has grown from 2 billion in 2007 to 31 billion in 2011 [23].
Figure 17: Data sets that have been published and interlinked by the Linking Open Data project so far. Collectively, the 295 data sets consist of over 31 billion RDF triples, which are interlinked by around 504 million RDF links (September 2011). The raw data that was used to draw the diagram is maintained by the LOD community on CKAN and can be accessed via the CKAN API. Source: [23]
The Khaos Research group (see picture on page 47) has focused its interests on the intersection of Semantics and Databases. Thus, these research activities have evolved from Deductive Databases and Query Optimization to the management of large and complex amounts of data (Big Data). One of the problems that the Semantic Web must address is the data integration problem. In this sense, Khaos has advanced with the proposal of solutions for Data Integration and its application to different domains. KOMF [24], is a system developed by this group that is based on the use of an ontology-based integration solution. This can be done over mediation or data warehouse approaches. The increasing interest in Linked Data is promoting the publication of many databases in RDF. This leads the research to reuse traditional solutions in the database community, such as the federated databases, but they have to be updated in the current context. Additionally, many of the integration problems remain still unsolved and new ones are appearing such as the problems of the data quality and the need for data curation. Individual users can make single improvements, but collectively, a community of users could provide highly curated data based on the community’s experience. In this sense, social networking is a new issue that is attracting a lot of interest in many domains.

Semantic Web applications, such as biological tools, use large ontologies, that is, ontologies with a large number (millions) of instances. Description logic based tools, like Pellet or RACER allow us to manage OWL ontologies, but not very large ones. Reasoning algorithms are not scalable and are usually main memory oriented. These reasoners are highly optimized for reasoning on the ontology structure (Tbox reasoning in Description Logic nomenclature), but have problems when dealing with reasoning on
Figure 18: Research topics addressed by the KHAOS research group

Data Integration
Data Analysis
Application Domains (Life Sciences, Tourism, etc.)
Optimization Techniques
Text Mining

Data Management and Analysis

Databases
Big Data
Semantic Web
Linked Data

Query Optimization
Deductive Databases
Ontology Alignment
Scalable Reasoning

DBTechNet – 15 Years of Excellence in Database Teaching
instances (Abox reasoning). Several alternative approaches using relational technology have been presented. However, these proposals have some problems. On the one hand, they are not complete with respect to OWL-DL reasoning. On the other hand, the performance and scalability of these tools is not satisfactory in some cases, particularly those which implement reasoning by means of datalog rules. Finally, the best tools are commercial tools, and users must pay to use them. The solution to these problems must be in using the full potential of databases to provide persistent and scalable reasoners for very large OWL-DL ontologies, like DBOWL [25].

However, these solutions require a commitment of the ontologies used for the representation of the data. In the absence of such an agreement, ontology alignment provides a solution to find how existing ontologies relate and to try to solve semantic differences. An approach followed by our group has been the combination of simple algorithms to build high-quality matching tools by means of the MaF framework [26]. This and similar problems can be solved using Optimization Techniques, which can be also applied in the analysis of the extreme scenarios in which Big Data is found. Finally, data is not only in databases as in many domains, relevant information can be found in scientific documents. In this context the analysis of texts (Text Mining) is used to retrieve such information. The combination of Data Integration, Ontologies and Natural Language Processing are key issues to reach high quality results.

**Tbox and Abox**

Description logics are formal languages to describe knowledge. Language elements can describe conceptual and factual knowledge. Tbox (terminological box) contains the conceptual knowledge of a domain and Abox (assertional box) contains the factual knowledge about entities (instances of concepts).

Example:

Tbox: All students are persons (STUDENT is subtype of PERSON)

Abox: Jim is a person (Jim is an instance of PERSON)
The Khaos Research Group belongs to the Department of Computer Languages and Computing Sciences from the University of Málaga.
What Next?
A Dozen Information-Technology Research Goals

By Jim Gray
(1998 Turing Award)

12 long range research goals

1. Architecture that scales up by a factor of 10^6

2. Win the Turing Test at least 30% of time

3. Speech to text: Hear as well as a native speaker

4. Text to speech: Speak as well as a native speaker

5. See as well as a person: recognize objects and behavior.

6. Personal Memex: Record everything a person sees and hears

Outlook
By Fritz Laux

„Prediction is very difficult, especially if it’s about the future“. This saying attributed to Mark Twain or Niels Bohr brings the problem of forecasting to the point. Nevertheless, one can try to extrapolate history which will most likely miss the emergence of a new technology. This is probably why Alan Kay stated: „The best way to predict the future is to invent the future“. In this sense we believe that it is best to be part of research and innovation. Being at the edge of science and development allows us to watch new scientific achievements, technological trends and to contribute if not even to influence the future of database technology. Therefore, it is of utmost importance for DBTechNet to be an active part of the scientific community and its research discussions. We aim to drive new development of concepts, models, and theory in the area of database and information systems. Beside the Grand Research Goals that Jim Gray depicted in 1999 (see sidebar) there are some short term challenges that will be important to solve in the future:

- Data storage models for high availability of huge amount of internet data
- Algorithms for real-time analytics
- Models for storing knowledge rather than information
- Automatic structuring of information and knowledge
- Solve the data integration problem
- User intentional query

We provide some arguments why these issues are important. Traditional databases use the relational model to structure data. This is perfect for data with a predefined structure, but not for
unknown or dynamic structures like text documents. Moreover, the huge amount of data is difficult to organize and manage in a central way. This is why so called NoSQL databases offer a more flexible and scalable data model which is used more and more for web applications. For e-Commerce it is essential to record the web browsing of a client, trace and analyze its behavior in real-time for a better service. The semantic web needs to store knowledge rather than information and, with the realization of the “internet of things”, we face the need for data integration even more urgent.

Querying the web is still essentially based on string matching, but a query should know what the users means. If a query asks for the “population of Helsinki” the user probably wants the number of inhabitants and not a document with a sentence like “The population of Helsinki is very friendly”. Triggered by new findings in database science and technology we have to constantly adapt and modify our reference curriculum. This will justify our claim to provide a standard curriculum for accreditation of database curricula. The reference syllabi for each learning unit need to be defined and
maintained according to target competences or professional roles found through our DBTech surveys. A software system based on our taxonomy should support this task. Derived from this we should develop an online assessment and e-Examination for database professionals and issue a certificate for successful participants. New topics and learning content require new didactic concepts in order to best impart knowledge. It will be our constant task to improve the didactic of our learning units. It is still a research field how to effectively use multi-media and virtual reality with different learning theories. In the not so far future we can count on virtual spaces and worlds to provide improved learning environments that can simulate face-to-face team work or allow the learner to immerse into virtual reality in order to experience abstract situations like a model instance or procedure.

Discussion at the International DBTech Pro Workshop in Malaga, May 2004
DBTechNet Humor
by Dimitris A. Dervos, A.T.E..I.

SELECT U.Academics, I.Professionals, V.Teachers_Trainers
FROM EU_Universities U, EU_Industry I, EU_VET_Centers V
WHERE (Motivation = High)
  AND (Enjoy (Teaching, Training))
  AND (European Conscience = High)
  AND (NOT Represent (EU.Brand Name))
  AND (Specialize_in_Database_Technology)
  AND (Focus_on_Professional_Education_Training)
  AND (Rate_High_Pedagogical_Correctness)
GROUP BY Workgroups
HAVING ('Hands-On' Educational/Training/Workshop/Virtual Labs Content, Professional Certification Framework, & more...) AS Deliverables
Timeline

- **1996** First ideas by Tiina Junkkari evangelized to Martti Laiho and Kari Silpiö
- **1997** Jaakko Rantanen and Kari Silpiö visited the T.E.I. and UoM in Thessaloniki
- **1998** Guest lectures of Dimitris Dervos at Helia and Martti Laiho at Reutlingen
- **1998 & 2000** SQL for Smarties, Workshop with Joe Celko at Helia
- **2001** Guest lectures at T.E.I. by Kari Silpiö from Helia and Jaakko Rantanen from HAMK
- **2001 May**, 1st International DBTechNet Workshop in Reutlingen, Germany, Topics: Web Architectures, XML, and Database Technologies
- **2001 December**, 2nd International DBTechNet Workshop in Thessaloniki, Greece

DBTech Highlights

1998 International DBTech Days at University of Karlskrona/Ronneby, Sweden, with Joe Celko, Thomas Connolly, David Jordan, Jim Melton, and Ken North

SQL 2002 Technologies Seminar at Helia, Finland, with Peter Gulutzan and Trudy Pelzer

2004 International DBTech Pro Workshop and Conference at Reutlingen University, Germany, with Avi Silberschatz

2012 DBTech Days at Haaga-Helia in Finland with Chris Date

• 2003 November, International DBTech Pro Workshop in Thessaloniki, Greece. Topics: Database Administration & Database Access

• 2004 May, International DBTech Pro Workshop in Malaga, Spain. Topics: Database Design

• 2004 November, International DBTech Pro Workshop and Conference in Reutlingen, Germany. Topics: Data Warehousing & Data Mining Keynote talk, “Next Generation Information Systems” by Avi Silberschatz

• 2005 April, International DBTech Pro Workshop in Helsinki, Finland. Topics: XML and Databases

• 2009 January – 2011 January: DBTech EXT project, partly funded by EC LLP Transversal Programme

• 2009 September, International DBTech EXT Workshop in Thessaloniki, Greece. Topics: Database Modeling and Semantics

• 2010 May, International DBTech EXT Workshop in Malaga, Spain. Topics: Database Access Patterns and Object-Relational mapping

• 2010 November, International DBTech EXT Workshop in Helsinki, Finland. Topics: Distributed and Replicated Databases, Mobile and Embedded Databases

• 2010 November, DBTech Symposium in Helsinki, Finland. Topics: New Technology Trends and Challenges for the Database Professional, Challenges for Database Education

• 2012 May 28 – 29, DBTech Days in Helsinki, Finland. Topics: Brush Up of Database Technologies with IBM, A Day with Chris Date

• 2012 October – 2014 October: DBTech VET Project, partly funded by EC LLP on Knowledge transfer to VET teachers, and professional training
About the Authors

**Chris J. Date**
is an independent author, lecturer, researcher, and consultant, specializing in relational database technology. He is best known for his book *An Introduction to Database Systems* which has sold some 850,000 copies. He was involved in Edgar F. Codd’s relational model and in the design for SQL/DS and DB2.

**Dimitris A. Dervos**
Dimitris (M.A. Physics, MSc. Computer Engineering, USCA, USA), (PhD Computer Science, Aristotle University of Thessaloniki, Greece) is a full professor at the Information Technology Department of the Alexander Technology Educational Institute, Thessaloniki, Greece. He is a founding member of DBTechNet, and a DBTechNet steering committee member. Personal homepage address: http://aetos.it.teithe.gr/~dad/

**Georgios Evangelidis**
is full Professor at the University of Macedonia, Thessaloniki, Greece. He has published many international journal and conference papers and coordinated/participated in many EU and national projects. He is interested in Databases, Knowledge Discovery, and Software Engineering (http://users.uom.gr/~gevan/).
Martti Laiho
Is a Senior Lecturer (MSc) at HAAGA-HELIA University of Applied Sciences and its predecessor institutions since 1987. He is a DBTechNet founding and steering committee member. As an expert on database systems his teaching covers database administration, data access technologies and transaction programming, J2EE and .NET architectures, and SAP system management. He is author/co-author of some Finnish textbooks on databases.

Fritz Laux
Is full professor for database & information systems at Reutlingen University, DBTech founding member, IARIA Fellow and advisory board member. He has more than 30 years of industrial and academic experience. He holds PhD and diploma in Mathematics and is an Electrical Engineer.

José Aldana Montes
A full Professor at the University of Málaga and leader of the Khaos Research Group. He has 20 years of expertise in researching and teaching in the databases field. His particular fields of interest are Semantic Middleware, Semantic Web, Semantic Data and Application Integration and Database Extension using formal semantics.

Ismael Navas-Delgado
Is a researcher and member of Khaos group. He received a PhD in Computer Science from the University of Malaga in 2009. His main interests are: Systems Biology and Semantic Technologies
References


[18] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009), The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1
This Festschrift is dedicated to Martti Laiho, the "heart and soul" of DBTechNet on the occasion of fifteen years of DBTechNet.
DBTechNet – 15 Years of Excellence in Database Teaching

DBTechNet Members

- 6 EU member states
- HE academic institutions
  - Haaga-Helia (Finland)
  - University of the West of Scotland
  - Reutlingen University
  - IIAAP/ESIAAP
  - Alexander TEI, Thessaloniki
    - University of Macedonia
    - University of Káraí
- VET centers
  - ETS Polytechnico Jesús María
  - Omnias
  - Finnish Business College
- IT company
  - Tieta Public Industry