Database Research Activities at The University of Vienna

A M. Tjoa, G. Vinek

Department of Information Systems
University of Vienna
Liebigg. 4/ 3-4, A-1010 Vienna, Austria
{tjoa|vinek} @ifs.univie.ac.at

Database research projects at the Department of Information Systems of the University of Vienna belong to four main categories. Research in the first category is focused on *conceptual and logical design* of databases. Activities in this area cover structural, behavioral, and object-oriented modeling and have resulted in enhancements of existing data models, formal specifications of modeling constructs, and in a set of tools supporting the design task. Research in the second category can be summarized under the term *database security*. In related projects we are concerned either with performance enhancement of multilevel secure DBMSs or with adapting mandatory access controls to better fit into commercial data processing practice. Research in the third category includes a number of efforts dealing with *internal data structure design* and *DBMS implementation aspects*. A main concern of those projects is to demonstrate the practical usability of non-traditional index structures in relational and object-oriented DBMSs under special consideration of parallel hardware architectures. Research in the fourth category is focused on *information retrieval* issues. Main attention is given to the development of an expert system for the automated document classification and in the improvement of user interfaces of information retrieval systems. There are additional research projects at the Department of Information Systems which are either just in their take-off phase or do not belong to the DB research in its common definition. New efforts focus for example on natural language interfaces and on evaluation of object-oriented DBMSs for the CIM environment. Activities not belonging to DB research include software engineering topics like man/machine interaction and software process programming. The following is a summary of our research activities in the database area.

1. Conceptual and Logical Database Design

Participating Faculty: C. Breiteneder, G. Kappel, R. Motschnig-Pitrik, T. Mueck, G. Pernul, M. Polaschek, A M. Tjoa, G. Vinek

The conceptual and logical design of a database involves the specification of some universe of discourse in terms of a data model (conceptual design) and its mapping into the model of an existing database system (logical design). If the models of conceptual and logical design are the same the two phases collapse. Research in this area is conducted in the realm of structural modeling, behavioral modeling, and object-oriented modeling. Structural modeling focuses on static aspects of the universe of discourse, behavior modeling focuses on dynamic aspects, and object-oriented modeling covers both with extensions due to the object-oriented paradigm.

Our work deals with the following interrelated topics in the areas mentioned above: enhancements to the expressive power of existing data models, development of tools to support the design task, and investigation of formal specifications to reason about modeling constructs. Within the area of data model enhancements we have developed the dynamic data models MODYN and BIER, and the object-oriented modeling technique OBD. The dynamic concepts of all three models are based on extensions to Petri nets. The motivation behind the development of MODYN (MOdeling DYNamics) was to support

executable specifications of object state transition constraints. The motivation behind BIER (Behavior Integrated Entity Relationship model) was the integration of static and dynamic modeling aspects within a uniform framework. As such, BIER has served as precursor to the development of OBD (Object/Behavior Diagrams), a uniform graphical design technique for the object-oriented design of information systems and object-oriented databases. OBD is based on semantic data modeling concepts for modeling objects, their properties, and their relationships, and on extensions to Petri net concepts for modeling the possible life cycles of and interactions between objects. Encountered problems with the object-oriented paradigm, such as the lack of object evolution and multiple type instantiation, and the intertwining of local behavior of a single object and global behavior where several objects are involved have been topic of research within OBD. To cope with the former the concept of roles, well-known from semantic data modeling, has been incorporated where several objects represent the same real world object in different contexts. To cope with the latter, 'cooperation contracts' are modeled in addition to object types where one contract represents the behavior in which several objects are involved. We have developed a set of proof-of-concept prototypes incorporating the model enhancements discussed above. Within the area of formal specifications we are building our research on sequential graph rewriting systems and evolving algebras. In order to specify and reason about various Entity Relationship (ER) dialects we are using sequential graph rewriting systems, which provide precise specifications yielding well-formed ER diagrams and transitions between them. For the formal specification of the object-oriented data model underlying OBD we have chosen the evolving algebra approach by Yuri Gurevich. The main advantages of this approach are the possibility to dynamically create and destroy objects and to easily experiment with various semantics due to the specification of an abstract machine. A further research topic within this area deals with the development of a set of cognitively grounded and representationally powerful yet canonical structuring concepts which can be used as reference model to compare the expressive power of existing data models.

Related Publications:

J. Eder, G. Kappel, A M. Tjoa, R. R. Wagner. BIER-The behavior integrated entity relationship approach. In: *Entity-Relationship Approach: Ten Years of Experience* (S. Spaccapietra, ed.) North-Holland, 1987.

J. Eder, G. Kappel, A M. Tjoa, R. R. Wagner. A behaviour design methodology for information systems. *Proc. 6th Int'l. Phoenix Conf. on Computers and Communications*. IEEE Computer Society Press, 1987.

T. Mueck, G. Vinek. Modelling Dynamic Constraints Using Augmented Place/Transition Networks. *Information Systems*, Vol. 14, No. 4. Pergamon Press, 1989.

G. Kappel, M. Schrefl. A Behaviour Integrated Entity Relationship Approach for the Design of Object-Oriented Databases In: *Entity-Relationship Approach: A Bridge to the User* (C. Batini, ed.). North-Holland, 1989.

R. Motschnig-Pitrik. A Framework for the Support of a Common Structural Level for Software-, Database-, and Knowledge Based Systems. *The Journal of Systems and Software*, Vol.12, No.2, North-Holland, 1990.

C. Breiteneder, T. Mueck: Performing Logical Database Design Using an ER Graph Rewriting System. *Proc. Int'l. Conf. on Databases, Parallel Architectures and Their Applications (PARBASE-90)*. IEEE Computer Society Press, 1990.

C. Breiteneder, T. Mueck. A graph grammar driven ER CASE environment. In: *Entity-Relationship Approach: The Core of Conceptual Modelling* (H. Kangassalo, ed.). North-Holland, 1991.

G. Kappel, M. Schrefl. Object/Behavior Diagrams. *Proc. 7th Int'l. Conf. on Data Engineering*. IEEE Computer Society Press, 1991.

G. Kappel, M. Schrefl. Using an Object-Oriented Diagram Technique for the Design of Information Systems. *Proc. 1st Int'l. Working Conf. on Dynamic Modelling of Information Systems* (H. G. Sol, ed.) North-Holland, 1991.

G. Kappel. Reorganizing Object Behavior by Behavior Composition - Coping with Evolving Requirements in Office Systems. *Proc. Datenbanken für Büro, Technik und Wissenschaft (BTW-91)*, (J. Appelrath, ed.), IFB 270, Springer Verlag, 1991.

M. Schrefl, G. Kappel. Cooperation Contracts. *Proc. of the 10th Int'l. Conf. on Entity-Relationship Approach*, San Mateo, CA, Oct. 1991.

G. Gottlob, G. Kappel, M. Schrefl. Semantics of Object-Oriented Data Models - The Evolving Algebra Approach. In: *Next Generation Information System Technology*. (J.W. Schmidt, A.A. Stogny, eds.), LNCS 504, Springer Verlag, 1991.

G. Pernul, K. Karlapalem, S. B. Navathe. Relational Database Organization Based on Views and Fragments. *Proc. 2nd Int'l. Conf. on Database and Expert Systems Applications (DEXA-91)*, Springer Verlag, 1991.

T. Mueck, M. Polaschek: MOG: A CASE Environment for the Specification of Transitional Constraints. *Proc. 2nd Int'l. Working Conf. on Dynamic Modelling of Information Systems* (H. G. Sol, ed.) North-Holland, 1992.

R. Motschnig-Pitrik, J. Mylopoulos. Classes and Instances. *Int. Journal of Intelligent and Cooperative Information Systems*, Vol.1, No.1, 1992.

G. Kappel, A M. Tjoa. State of Art and Open Issues on Graphical User Interfaces for Object-Oriented Database Systems. *Information and Software Technology*. Butterworth, 1992.

S. B. Navathe, G. Pernul. Conceptual and Logical Design of Relational Databases. *Advances in Computers* (M. C. Yovits, ed.), Vol. 35. Academic Press, 1992.

2. Database Security

Participating Faculty: G. Luef, G. Pernul, A M. Tjoa

The increased use of database systems technology results in the fact that public becomes more and more dependent on the proper functioning of the systems and their offered security features. This makes database security research a very important and challenging issue and subject to national and international standardization and evaluation efforts.

Database security research at the University of Vienna falls into two categories. In a first project, we are concerned with physical design aspects of multilevel secure DBMSs. Multilevel secure systems are very powerful in supporting a high degree of trust in their offered security features but due to security relevant overhead they suffer from poor performance. In this project we investigate the use of *Cross-Links* between index structures as a method to overcome the degradation of performance for range queries spanning several security levels. The method is analyzed in a multilevel secure database environment based on the kernelized architecture and supports range search and tuple insert.

The second project is concerned with the design and implementation of the design environment *AMAC* (Adapted Mandatory Access Controls). The main goal of this project is to fine-tune mandatory access controls to better fit into commercial data processing practice. *AMAC* consists of high-level data and security modeling using ER techniques, a transformation policy from ER schemas into multilevel secure relations, a decomposition approach for the construction of single level fragments from multilevel base relations, a supporting policy for the automated security labeling of security objects and subjects, and finally it supports security enforcement by using hidden trigger rules. Since labeled data items are not available in most security critical civil database applications, the approach serves well as the underlying basis for the construction of a computerized design tool that assists a human database or security administrator during the different phases of the construction of a multilevel secure database. Further research is directed towards applying *AMAC* for the reverse engineering of existing databases that do no longer meet increased security requirements and using *AMAC* as a canonical security model for a federated information base consisting of a set of tightly coupled component databases.

Related Publications:

G. Pernul, G. Luef. A Multilevel Secure Relational Data Model Based on Views. *Proc. 7th Annual Computer Security Applications Conference*. IEEE Computer Society Press, 1991.

G. Pernul, A M. Tjoa. A View Integration Approach for the Design of Multilevel Secure Databases. *Proc. 10th Int'l. Conf. on the Entity-Relationship Approach*, San Mateo, CA, Oct. 1991.

G. Pernul, G. Luef. A Bibliography on Database Security. *ACM SIGMOD Record*, Vol. 21, Nr. 1, March 1992.

G. Luef, G. Pernul. Supporting Range Queries in Multilevel-Secure Databases. In: *Database Security: Status and Prospects V*, (S. Jajodia, C. E. Landwehr, eds.) North-Holland, 1992.

G. Pernul, A M. Tjoa. Database Security Policies. *Proc. Int'l. Conf. on Safety, Security and Reliability of Computers (SAFECOMP-92)*. Pergamon Press, 1992.

S. Eichinger, G. Pernul. Design Environment for a Hospital Information System: Meeting the Data Security Challenge. *Proc. 7th World Congress on Medical Informatics (MEDINFO-92)*. North-Holland, 1992.

G. Pernul. Security Constraint Processing in Multilevel Secure *AMAC* Schemata. *Submitted for publication.*

3. Internal Data Structure Design and Implementation Aspects

Participating Faculty: G. Luef, T. Mueck, E. Schikuta, G. Vinek

Providing reasonable performance for query execution in relational and object-oriented database systems is still a challenge. In particular, research in internal data structure design and evaluation has been driven to a certain extent by the needs of interactive users launching ad-hoc queries against large data sets. At the University of Vienna, the research efforts in internal data structure design and evaluation focus on the efficient utilization of multi-attribute search structures (mainly grid files and grid file derivates) not only for special-purpose systems like spatial databases but also for traditional business oriented DBMSs with ad-hoc query interfaces. Within this field, there are three main streams of research:

- Adapting and evaluating multi-attribute search structures in such a way that a substitution of traditional B+-tree index maintenance systems becomes an attractive option for DBMS designers. Currently, multi-attribute search structures are well established for applications handling spatial objects as found in geographical databases. However, considering the vast amount of manpower already put into the implementation of traditional B+-tree concepts as well as the positive aspects of those concepts demonstrated in existing systems, a paradigm shift from traditional indexing strategies to multi-attribute search concepts will only occur if forthcoming data structures outperform even sophisticated B+-tree systems for most requirement profiles.
- Adapting and evaluating concurrency control algorithms, error recovery algorithms and other algorithms needed to support internal data structures as known from B+-tree systems for the special needs of multi-attribute structures. At the moment, B+-tree systems profit not only from the positive features of the tree data structure itself, but also from very elaborated supporting algorithms for various subproblems like concurrency control, error recovery and so on. Consequently, any multi-attribute search structure meant to challenge B+-tree superiority has to be supplied by comparable or even better algorithms with respect to those DBMS subtasks.
- Utilizing parallel processing power as well as disk arrays for DBMSs which are based on multi-attribute search structures in order to take full advantage of certain structural properties favoring parallel query processing. An attractive property of the grid file approach and of some follow-up approaches is the stepwise partitioning of the overall data space into disjoint subspaces (i.e. hyper-rectangles). In particular, this property gives an extremely useful hook for parallel query processing strategies. Two research groups are engaged in the development of highly parallel DBMSs on shared nothing MIMD architectures, considering a multi-level approach towards query processing.

Related Publications:

E. Schikuta: A Data Base Kernel For Parallel Computer Architectures. *Proc. 3rd Annual Parallel Processing Symposium.* IEEE Computer Society Press, 1989.

E. Schikuta: The Grid File Data Base System. *Proc. 10th SCCC Int'l. Conf. on Comp. Science.* Santiago de Chile, Oct. 1990.

E. Schikuta: A Grid File Based Highly Parallel Relational Data Base System. Proc. 4th ISMM Int'l. Conf. Parallel and Distributed Computing and System. Washington D. C., Oct. 1991.

T. Mueck, M. Schauer: Sorting in the BANG File. *Tech.Rep. TR-ISI/InfoSys-109*, Dept. of Inf. Sys., Univ. of Vienna, Dec. 1991.

G. Luef, T. Mueck: Concurrent Operations in Balanced and Nested Grid Files. *Proc. of the 6th Int'l. Working Conference on Scientific and Statistical Database Management.* Ascona, June 1992.

M. Kollingbaum, T. Mueck, G. Vinek: PARABASE - A federated DBMS architecture supported by a massively parallel shared nothing database server. *Proc. of the OpenForum92 Technical Conference - Distributed Computing: Practice and Experience.* Utrecht, Nov. 1992.

4. Information Retrieval

Participating Faculty: M. Höfferer, D. Merkl, A M. Tjoa, S. Vieweg, W. Winiwarter.

In general the research in the area of information retrieval (IR) can be described as the development of procedures for representation, storage, organization, and retrieval of textual information. IR research at the University of Vienna focuses on the areas of organization and representation of documents. Our investigations cover the problem that both document representation and the description of the user's need for information described in a query are uncertain. The two major goals are the *development of an expert system* to semantically classify documents and the *improvement of user interfaces* of information retrieval systems.

The expert system for document classification attempts to solve the problems arising from classical document clustering which is based on statistical measures of co-occurrences of index terms in documents. The main drawback of this approach is that only the document's syntax is taken as the basis for document classification whereas the contents (semantics) of the document is not considered. The prototype developed overcomes this shortcoming by using a hierarchically organized knowledge base where index terms are related to subject areas. The task of knowledge based document classification in this environment is defined as finding a plausible subject area for a document built up from automatically extracted document descriptors. The inference process is governed by three probabilistic parameters: estimating the importance of an index term with respect to a given subject, the occurrence frequency of an index term in a subject, and the importance of an index term with respect to a document. The result of the inference process is the mapping of a document to a subject area. So far, the expert system has been tested using a database consisting of legal documents (decisions of the Austrian Supreme Court in Civil Law). Each of these documents is represented by words and automatically extracted citations of legal statutes.

A further objective of IR research at the Department of Information Systems is investigating the problems of user interfaces in information retrieval systems. As the semantic relationship between documents is not considered in document representation of classical information retrieval systems this relationship is not visible in the user interface of such systems. Our investigations are directed towards the application of the hypertext metaphor for document presentation. The results of knowledge based document classification are the basis for establishing a set of hypertext links between documents covering the same or similar subject areas.

Related Publications:

D. Merkl, S. Vieweg, A. Karapetjan. KELP - A Hypertext oriented User-Interface for an Intelligent Legal Fulltext Information Retrieval System. *Proc. 1st Int'l. Conf. on Database and Expert Systems Applications (DEXA-90)*. Springer Verlag, 1990.

D. Merkl, A M. Tjoa, S. Vieweg. Hypertext as a tool for humanized interface in the information retrieval area. *Proc. of the IFIP WG 9.1 Working Conference on Human Jobs and Computer Interfaces*. Tampere, Finland, June 1991.

D. Merkl, A M. Tjoa, S. Vieweg. A Hypertext User Interface for a legal Full Text Information Retrieval System. In: *Information Retrieval.* (N. Fuhr, ed.). IFB 289, Springer Verlag, 1991, (in German).

D. Merkl, A M. Tjoa, S. Vieweg, M. Höfferer. Knowledge Based Document Classification in the Information Retrieval Domain. *Proc. of the 2nd Golden West Int'l Conference on Intelligent Systems.* Reno, Nev, July 1992.

D. Merkl, A M. Tjoa, S. Vieweg: BRANT - An Approach for Knowledge Based Document Classification in the Information Retrieval Domain. *Proc. 3rd Int'l. Conf. on Database and Expert Systems Applications (DEXA-92)*. Springer Verlag, 1992.