The Database Group at National Technical University of Athens (NTUA)

Computer Science Division, Dept. of Electrical and Computer Engineering
National Technical University of Athens, Zografou, Greece 15773
URL: http://www.dbet.ece.ntua.gr

1. Overview

The National Technical University of Athens (NTUA) is the leading Technical University in Greece. The Computer Science Division of the Electrical and Computer Engineering Department covers several fields of practical, theoretical and technical computer science and is involved in several research projects supported by the EEC, the government and industrial companies.

The Knowledge and Data Base Systems (KDBS) Laboratory was established in 1992 at the National Technical University of Athens. It is recognised internationally, evidenced by its participation as a central node in the Esprit Network of Excellence IDOMENEUS. The Information and Data on Open MEdia for NEtworks of USers, project aims to co-ordinate and improve European efforts in the development of next-generation information environments which will be capable of maintaining and communicating a largely extended class of information in an open set of media.

The KDBS Laboratory employs one full-time research engineer and several graduate students. Its infrastructure includes a LAN with several DECstation 5000/200 and 5000/240 workstations, an HP Multimedia Workstation, several PCs and software for database and multimedia applications.

The basic research interests of our Laboratory include: *Spatial Database Systems, Multimedia Database Systems* and *Active Database Systems*. Apart from the above database areas, interests of the KDBS Laboratory span several areas of Information Systems, such as Software Engineering Databases, Transactional Systems, Image Databases, Conceptual Modeling, Information System Development, Temporal Databases, Advanced Query Processing and Optimization Techniques.

The group's efforts on Spatial Database Systems, include the study of new data structures, storage techniques, retrieval mechanisms and user interfaces for large geographic data bases. In particular, we look at specialized, spatial data structures (R-Trees and their variations) which allow for the direct access of the data based on their spatial properties, and not some sort of encoded representation of the objects' coordinates. We study implementation and optimization techniques of spatial data structures and develop models that make performance estimation. Finally, we are investigating techniques for the efficient representation of relationships and reasoning in space.

The activities on Multimedia Database Systems, include the study of advanced data models, storage techniques, retrieval mechanisms and user interfaces for large multimedia data bases. The data models under study include the object-oriented model and the relational model with appropriate extensions to support multimedia data. We are also investigating content-based search techniques for image data bases. In a different direction, we are studying issues involved in the development of multimedia front-ends for conventional, relational data base systems.

In the area of Active Database Systems, we are developing new mechanisms for implementing triggers in relational databases. Among the issues involved, we address the problem of efficiently finding qualifying rules against updates in large sets of triggers. This problem is especially critical in database system implementations of triggers, where large amounts of data may have to be searched in order to find out if a particular trigger may qualify to run or not.

Continuing work that started at the Foundation for Research and Technology (FORTH), Institute of Computer Science, the group is investigating reuse-oriented approaches to information systems application development. The approaches are based on a repository that has been implemented at FORTH as a special purpose object store, with emphasis on multimodal and fast retrieval. Issues of relating and describing software artifacts (designs, code, etc.) are among the topics under investigation.

A new important research direction of the group is on Data Warehouses, which are seen as collections of materialized views captured over a period of time from a heterogeneous distributed information system. Issues such as consistent updates, data warehouse evolution, view reconciliation and data quality are being investigated.

Research in Image Databases deals with the retrieval by image content, that uses techniques from the area of Image Processing. We are currently at early stage in this direction, having collected many segmentation and edge detection algorithms, which will be used and evaluated in images of various contents.

Our work on Advanced Query Processing and Optimization Techniques includes dynamic or parametric query optimization techniques. In most database systems, the values of many important run-time parameters of the system, the data, or the query are unknown at query optimization time. Dynamic, or parametric, query optimization attempts to identify several execution plans, each one of which is optimal for a subset of all possible values of the run time parameters.

In the next sections we present in detail our research efforts on the three main research areas of the KDBS Laboratory: Spatial, Multimedia and Active Databases.

2. Research on Spatial Databases

Timos Sellis (Assoc. Prof.), Dimitris Papadias (PhD), Yannis Theodoridis (PhD Candidate), Emmanuel Stefanakis (PhD Candidate)

2.1. Spatial Data Structures.

The need to store and manipulate multi-dimensional objects like points, lines or polygons in Database Management Systems (DBMSs) is obvious when dealing with applications such as Geographic Information Systems (GIS), Computer-Aided Design (CAD) or Image and Multimedia Databases. Because of that, several data structures have been proposed to efficiently handle multi-dimensional (*spatial*) data. We concentrate on the R-tree family of spatial data structures because they seem to be the most efficient ones for indexing large sets of spatial data.

The main operations that are addressed by spatial data structures are *point queries* (i.e., given a point in space, find all objects that contain it) and *region queries* (i.e., given a region in space, find all objects that intersect it). Other, more application oriented, spatial relations that are also included in a spatial database system can be grouped as:
- *Topological relations:* Given an object in space, find all objects that cover (or meet, overlap etc.) it.
- *Direction relations:* Given an object in space, find all objects that lie left (or above, north etc.) of it.
- *Distance relations:* Given an object in space, find the k-nearest (or k-furthest) objects to it.

Spatial data structures are usually designed to efficiently support the first group of operations but it is obvious that support for the second group of operations is also necessary. Our research focuses on the following major tasks:
(i) A proposal for new techniques that optimize the performance of R-trees, compared to the existing variants. In particular, we have focused on several factors that affect the R-tree performance and we are studying optimum configurations of them in order to achieve more efficient indexing [3,7,9].
(ii) A proposal for analytical models that predict the performance of R-trees for the retrieval of spatial operations. In particular, we extend related work on R-tree analysis in order to support several distributions of data and combinations of spatial queries apart from the typical point or range ones.

2.1 Spatial Information Representation and Processing.

On the other hand, research on qualitative representation and processing of spatial knowledge, includes the identification of the common concepts underlying qualitative spatial knowledge representation and the computational tasks involved in relation-based spatial information processing. To preserve a set of spatial relations among distinct objects in an image, discarding information, such as shape and size of objects, and irrelevant spatial relations, *symbolic spatial indexes* are maintained. The construction of a symbolic spatial index from an input image can be thought of as a transformation that keeps only a set of *representative points* needed for the definition of the relations of the modeling space. By keeping the relative arrangements of the representative points in symbolic spatial indexes and discarding all other points, we maintain enough information to answer queries regarding spatial relations in two-dimensional space without the need to access the initial image [1,2,4,5,6,8].

[1] D. Papadias, T. Sellis, "The Semantics of Relations in 2D Space Using Representative

Points: Spatial Indexes", *Proceedings of the 1st COSIT Conference*, Springer Verlag, September 1993.
[2] D. Papadias, T. Andronikos, "Relation-Based Information Processing With Symbolic Spatial Indexes", *Proceedings of IGIS Workshop*, Springer Verlag, March 1994.
[3] Y. Theodoridis, T. Sellis, "Optimization Issues in R-Tree Construction", *Proceedings of IGIS Workshop*, Springer Verlag, March 1994.
[4] D. Papadias, M. Kavouras, "Acquiring, Representing and Processing Spatial Relations", *Proceedings of the 6th SDH Symposium*, Taylor Francis, September 1994.
[5] D. Papadias, T. Sellis, "A Pictorial Language for the Retrieval of Spatial Relations from Image Databases", *Proceedings of the 6th SDH Symposium*, Taylor Francis, September 1994.
[6] D. Papadias, T. Sellis, "On the Qualitative Representation of Spatial Knowledge in 2D Space", *VLDB Journal*, Special Issue on Spatial Databases, vol. 4(3), pp. 479-516, September 1994.
[7] D. Papadias, Y. Theodoridis, T. Sellis, "The Retrieval of Direction Relations Using R-trees", *Proceedings of the 5th DEXA Conference*, Springer Verlag, September 1994.
[8] D. Papadias, T. Sellis, "A Pictorial Query-By-Example Language", *Journal of Visual Languages and Computing*, Special Issue on Visual Query Systems, vol. 6, pp.53-72, March 1995.
[9] D. Papadias, Y. Theodoridis, T. Sellis, M. Egenhofer, "Topological Relations in the World of Minimum Bounding Rectangles: a Study with R-trees", *Proceedings of ACM SIGMOD Conference*, May 1995.

Funding: Research on Spatial Databases has been funded by the Commission of the European Union under the ESPRIT II Basic Research Actions Program project "AMUSING" (1992-1994) and by the Greek Government under the PENED Program project "Geographic Databases" (1993-1995).

3. Research on Multimedia Database Systems

Yannis Vassiliou (Professor), Timos Sellis (Assoc. Prof.), Michalis Vazirgiannis (PhD), Kyriakos Diakonikolaou (PhD Candidate), Lefteris Lykouropoulos (PhD Candidate), Andreas Maniatis (PhD Candidate)

In our lab, there is active research in the field of Multimedia Database Management Systems (MM-DBMSs). The research is backed up by a series of national and European Union research and industry projects in this field. The activities on Multimedia Database Systems, include the study of advanced data models, storage techniques, retrieval mechanisms and user interfaces for large multimedia data bases. The data models under study include the object-oriented model and the relational model with appropriate extensions to support multimedia data.

Our research is mainly oriented towards *integrated modeling* of multimedia objects and interactive multimedia applications. More specifically we have specified an integrated Object-Oriented Model which covers all the relevant aspects:

Multimedia objects as data base objects (with full temporal and spatial features). In this module multimedia objects are modeled as multimedia database (MDB) objects. In the MDB there are objects of all types: text, image, audio and video. The objects include, either explicitly or via reference, the raw media material as well as object attributes, like the spatial and temporal features of the object, and functions for storage/retrieval and presentation of the objects. The description of the class hierarchy that the model consists of is presented in [1].

Transformation and composition of multimedia objects in space and time. In this module, objects from the MDB are transformed in order to participate in multimedia applications according to the application requirements. There is a set of transformations that may be applied to a multimedia object before presenting it. The transformations manipulate spatial and/or temporal features of the media objects. The module is based on an Object-Oriented hierarchy of classes modeling the aforementioned features. The details of the class hierarchy may be found in [2,3].

Definition of simple and complex events that may be consumed by the multimedia application. An Interactive Multimedia Application (IMA) is an event based environment, thus it is critical for a relevant model to represent efficiently these entities. The general scheme is that an event triggers one or more actions. The events may be generated by the system, the user, or the application and may be simple or complex [4]. As for the actions, they may be simple (i.e. execution of a video clip) or complex (synchronized presentation of actors based on the aforementioned composition operators).

Multimedia application scenario based on events. The term "scenario" denotes the application functionality which is of two kinds: pre-orchestrated and event-based. The term pre-orchestrated implies that certain actions will take place at specific time instants while event-based implies that actions are triggered by events that occur in the application context either by the user or the system or by entities participating in the application (media objects, media compositions etc.).

The modeling of an IMA scenario must cover both pre-orchestrated and event-based cases. The scenario of an IMA consists of fundamental units, so called scenario tuples[3]. The general form of a scenario tuple should represent: the triggering event (simple or complex), stopping event, set of actions to be executed when the triggering event occurs, constraints to be fulfilled for the scenario to start its execution. An important issue is the management of the scenario tuples. Once defined there should be mechanisms for tuple storage, indexing and retrieval, consistency cross-checking etc.

There is an ongoing effort for the development of the aforementioned Object-Oriented Data Models in our lab. We design the implementation so as to exploit Object-Oriented storage managers (AT&T/BESS) in UNIX platforms using high level Toolkits (Tk) for user interfaces. An important backup in our implementation effort arises from the HP equipment grant (including advanced Multimedia Workstations) which aims at studying the incorporation of multimedia information management in OO-DBMSs.

Funding: The group participates in the following projects:

ALEXANDER
The project ALEXANDER (Esprit Project CU09/EP7521), deals with the development of a multimedia database for the cultural heritage of Northern Greece. It aims at introducing and applying multimedia information systems in cultural and historical research. Two applications are being developed in the context of this project: a multimedia database of the artifacts of the Byzantine Museum of Athens, and a CD-I title and an associated asset database of the culture and history of ancient Macedonia, based on the findings of the archaeological excavations at Northern Greece.

ARCHEOTOOL
This is a national project funded by the General Secretariat for Research & Technology. The contemporary needs regarding both the excavation methodology and systematic documentation of information lead to the accumulation of a large set of data of various types, such as texts, catalogues, pictures, drawings, measurements, etc. The recording, management, and scientific exploitation of information derived from excavations constitute a hard task, which can be significantly assisted by the use of computer technology. The project aims at the design and implementation of a system, which will adopt the capabilities offered by the computer technology, to assist the recording and exploitation of the accumulated information from the excavation procedure.

CMIS
The Corporate Multimedia Information Systems (CMIS) project is an International project funded by the European Union under the framework for “Exploratory Actions in Multimedia Technical Documentation”. The objectives of the project are to determine the potential, using a pilot application, to address the needs of SMEs in the area of multimedia technical documentation. The aims of the pilot application are to develop various Information Engineering solutions meeting the multimedia technical documentation needs of SMEs and to undertake user evaluations of the proposed solutions. These solutions and their evaluation address pre-specified aspects of information engineering as applied to technical documentation. The technological, organizational, legal and methodological framework within which SMEs throughout Europe can produce multimedia technical documents in an efficient and cost effective way are studied.

TSAI
TSAI (Telematic Systems and Services for the Advertising Industry) is an International project funded by the European Union under the framework for “Exploratory Actions in Multimedia Publishing”. The project aims at studying the feasibility of a pilot application that will improve the content and value of electronic information in the production, distribution and retrieval areas of the information chain as it appears in the Advertising Industry.

[1] M. Vazirgiannis, C. Mourlas, "An object Oriented Model for Interactive Multimedia Applications", *The Computer Journal*, British Computer Society, vol. 36(1), 1/1993.
[2] M. Vazirgiannis, M. Hatzopoulos, "A Script Based Approach for Interactive Multimedia Applications", *Proceedings of the IEEE - MMM (International Conference on Multi-Media Modelling) Conference,* Singapore, (November 9-12, 1993).
[3] M. Vazirgiannis, M. Hatzopoulos, "Integrated Multimedia Object And Application Modeling based on events and scenarios ", *proceedings of 1st International Workshop for MMDBMSs (IEEE), 8/1995*, Blue Mountain Lake, NY (The Adirondacks).
[4] M. Vazirgiannis, T. Sellis, "Event And Action Representation And Composition For Multimedia Application Scenario Modelling", *Technical Report*, Department of Electrical Engineering, Division of Computer Science, National Technical University of Athens, June 1995.
[5] M. Vazirgiannis, K.Petrou, A. Tsompanidis, M.Hatzopoulos, "An Object Oriented Framework for Knowledge representation based on Fuzzy sets", *Intelligent and Fuzzy Systems Journal*, Wiley, vol. 1., pp. 265-278.

4. Research on Active Databases

Timos Sellis (Assoc. Prof.), Manolis Koubarakis (PhD)

4.1 Active DBMS Implementation.

In the past, NTUA's research staff has been investigating methods for storing and manipulating large rule bases using a relational DBMS [1]. In this work an original approach to decompose and store conditions of rules such as those used in Production Systems (PS) is presented. A set-oriented approach uses a special data structure that is implemented using relations. Based on these structures, a matching algorithm, similar to the Rete match algorithm of OPS5 was designed to allow for the efficient identification of rules whose conditions are satisfied. An advantage of the proposed method is that it is fully parallelizable, thus making it attractive for parallel computing environments as well.

In companion work [2], research was presented on exploiting concurrency in executing a set of applicable rules, in relational active database systems, and for the particular implementation suggested in [1]. A locking protocol is designed to handle the conflicts among actions that compete for the same data; this protocol is shown to be correct and a large simulation study is performed to show the effectiveness of the proposed method under various rule base characteristics.

Recently, we have focused on the implementation of the above ideas and mechanisms. A mini relational database system developed at NTUA, MINIREL, has been extended with active capabilities: a rule syntax has been adopted and implementation of the basic mechanisms as described in [1] have been fully implemented [3].

In a different direction, the group has initiated research on the representation of multimedia applications based on the scenario concept [4]. The scenario is described in terms of events and actions. The proposed model represents all the events that may occur in a multimedia application (originated from the system, the application or the user). As regards actions a scheme has been proposed for spatial and temporal composition representation based on operators that cover all the spatial and temporal relationships among multimedia objects. Using the above, the scenario of an application is represented as a set of scenario tuples that correspond to fundamental sets of actions originated by the same event (or events).

4.2 Temporal Aspects

Temporal aspects of active databases have also been investigated within the context of Temporal Constraint Databases. Usually the term "constraints" in databases refers to integrity constraints i.e., statements, expressed in some formal language, which must be true in each state of the database (not taking dynamic constraints into account). Recently, attention has also focused on using constraints for the representation of data (e.g., see the proceedings of PODS94 or the proceedings of the 1st and 2nd Workshop on Principles and Practice of Constraint Programming). Our work in this area has concentrated on temporal constraint databases.

Our work [5,6,7,8] is motivated by the fact that an important requirement of advanced temporal applications is the ability to deal with definite, indefinite, finite and infinite temporal information. There is currently no database model which offers this functionality in a single unified framework. In [5] we have argued that the combination of relational databases and temporal constraints offers a powerful framework which addresses these needs. We have developed the foundations of a theory of temporal constraint databases and indefinite temporal constraint databases. We have studied a hierarchy of parameterized database models: M-relational databases, L-constraint databases and indefinite L-constraint databases. The language L, the parameter, defines the constraint vocabulary and M is the structure over which L-constraints will be interpreted. The models of temporal constraint databases and indefinite temporal constraint databases can then be studied as instances of the last two of the above parameterized models.

[1] Sellis, T., Lin, C-C., and Raschid, L. ``Coupling Production Systems and Database Systems: A Homogeneous Approach", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 2, April 1993.
[2] Raschid, L., Sellis, T., Delis, A. ``A Simulation-Based Study on the Concurrent Execution of Rules in a Database Environment", Journal of Par. and Distributed Comp., Vol. 20, Jan. 1994.
[3] Tsois, A. ``An Implementation of Production Rules in a Relational Database Management System", Diploma Thesis, Dept. of Electical and Computer Engineering, National Technical University of Athens, February 1995 (in Greek).
[4] M. Vazirgiannis, T. Sellis, "Event And Action Representation And Composition For Multimedia Application Scenario Modelling", *Technical Report*, Department of Electrical Engineering, Division of Computer Science, National Technical University of Athens, June 1995.
[5] M. Koubarakis, ``Foundations of Temporal Constraint Databases", Ph.D. thesis, National Technical University of Athens, February 1994.
[6] M. Koubarakis, ``Database Models for Infinite and Indefinite Temporal Information", Information Systems, Vol. 19, No. 2, March 1994, pages 141-173.
[7] M. Koubarakis, ``Complexity Results for First-Order Theories of Temporal Constraints", Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94), pages 379-390, May 1994.
[8] M. Koubarakis, ``Foundations of Indefinite Constraint Databases" Proceedings of the Workshop on Principles and Practice of Constraint Programming, Orcas Island, Washington, May 1994.

Funding: Research on Active Databases has been funded by the Commission of the European Union under the Human Capital and Mobility Program project "ACTNET" (1993-1996).

RIDS '95

Finally, NTUA organized the Second International Workshop on Rules in Database Systems (RIDS '95) in Athens on September 25-27, 1995. The aim of this workshop, which was the second organized by ACTNET members after a very successful first workshop in Edinburgh (Scotland) in September 1993, was to bring together researchers working on both theoretical and practical aspects of rules in database systems, to examine the current state of the art, and to identify complementary areas for further development, as well as interesting applications.

In response to the call for papers for RIDS '95, the program committee received 47, out of which 22 papers were selected for inclusion in the proceedings and presentation in the workshop. These papers covered several aspects, including: Semantics for Active Database Systems, Active Behaviour, Rule Base Organization and Modelling, Rule Analysis, Deductive Databases, Implementation and Benchmarking of Active Database Systems, and Cooperative System Support. In addition, there were two panel discussions, one addressing the "The Active Database Management Systems Manifesto", which is an attempt to set the scene on what an active database management system should be, and one on Implementation & Application Experiences.

The workshop attracted 52 participants who provided a lively and high quality forum for discussion of various subjects. In recognition of the importance of the field, Springer Verlag has published the proceedings in its Lecture Notes in Computer Science series.

"Rules in Database Systems", T. Sellis (Ed.) Proceedings of the Second International Workshop, RIDS '95, Glyfada, Athens, Greece, September 25 - 27, 1995. Lecture Notes in Computer Science. Eds.: G. Goos, J. Hartmanis, and J. van Leeuwen, Vol. 985.