# Final Progress Report for Incorporation of Data Files into Semantic Databases (DAAH04-96-1-0049)

## 1. STATEMENT OF THE PROBLEM STUDIED

This project supported the investigation of ways to incorporate heterogeneous data files as well as information about these files into a high performance semantic database. The incorporation of these files will allow scientists to quickly search a diverse set of information in ways that are impractical using current databases. Our methods do not logically change the data files, allowing the programs that are currently used to generate, process and access these files to remain in operation. Towards this goal, we have enabled our database engine to store raw values of attributes of arbitrary length, e.g., a datum 2GB long, transparently fragmented and load balanced among many disks comprising the database with highly efficient access to any offset within this attribute value datum. We are developing a loading program that facilitates the import of data into our Semantic Object-Oriented database system, technology that streamlines access to arbitrary WWW data, and technology that allows us to easily design semantic access methods for other databases.

## 2. SUMMARY OF THE MOST IMPORTANT RESULTS

Our efforts have focussed on providing transparent access to heterogeneous data files (including DX files) via our Semantic Object-Oriented database (Sem-ODB). This includes the development of a loading program that facilitates the import of data into our Semantic Object-Oriented database system, technology that streamlines access to arbitrary WWW data, and technology that allows us to easily design semantic access methods for other databases.

### Data Loader

SemLoader allows its user to define loading methods for arbitrary text and binary files. By reading a control file, the data is imported into a Sem-ODB database. The user is able to specify the semantic schema to be used and how the data is to be stored using that schema. Full logging and error reporting is performed during the loading process. SemLoader also allows data to be exported from a Sem-ODB database into SDL or XML format.

### Access to WWW Data

We are developing technology that allows arbitrary WWW data to be more easily accessed. The system will allow buffering and streamlining between the user and web data providers; converting visual presentation of information into data for further processing, translating one data request into a cascade of data requests and pasting results together; filtering data output; allowing a variety of presentations of data different from the original presentation; optional dataflow between the user's applications and the third-party data providers bypassing

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



1. AGENCY USE ONLY (*Leave Blank*)

2. REPORT DATE
6/13/00

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Incorporation of Data files into Semantic Databases

5. FUNDING NUMBERS
Grant Number DAAH04-96-1-0049

6. AUTHORS
Naphtali Rishe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Florida International University
School of Computer Science
University Park
Miami, FL 33199

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. BOX 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO PROPOSAL NUMBER
ARO 35536.15-MA-SAH

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be constructed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
ABSTRACT OF REPORT: This project supported the investigation of ways to incorporate heterogeneous data files as well as information about these files into a high performance semantic database. The incorporation of these files will allow scientists to quickly search a diverse set of information in ways that are impractical using current databases. Our methods do not logically change the data files, allowing the programs that are currently used to generate, process and access these files to remain in operation. Towards this goal, we have enabled our database engine to store raw values of attributes of arbitrary length, e.g., a datum 2GB long, transparently fragmented and load balanced among many disks comprising the database with highly efficient access to any offset within this attribute value datum. We are developing a loading program that facilitates the import of data into our Semantic Object-Oriented database system, technology that streamlines access to arbitrary WWW data, and technology that allows us to easily design semantic access methods for other databases.

14. SUBJECT TERMS
semantic databases, distributed databases, multimedia databases, heterogeneous databases, DX file

15. NUMBER OF PAGES
15

16. PRICE CODE

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-1
298-102

interactive interfaces. We are developing a prototype tool, Extractor, that will allow buffering and streamlining between the user and web data providers. This is an anti-GUI tool "stripping" out "sugar" from data, hiding the GUI from the batch user, and translating one data request into a cascade of data requests.

We are also developing technology that will enable users to browse a variety of spatial data via the WWW. The availability and use of remotely-sensed data has increased dramatically in the past several years. The amount and varied types of information that can be extracted from remotely-sensed data is vast and extremely useful for science education and research inquiry. These spatial data sets, however, are in many different formats. This can make it difficult and expensive to disseminate this data, particularly when the information is found in two or more different data sets that frequently require separate programs to view and extract the data. Problems are further increased when the amount of data is considered. Spatial data sets are inherently large. Storage and retrieval of spatial data, even when the desired information is of a uniform format, is often cumbersome at best. Many programs that allow access to spatial data sets are rather difficult to use. With our WWW spatial data browser technology, the user could easily access this data over the Internet. The user will not need to install any software, as our "browser" will be a dispatchable Java agent running under the user's regular browser.

The following are some of the principal features of our virtual flight technology over WWW:

- Main Flight Window: The main flight window displays the spatial data image and allows users to fly over the available images. The direction of flight is determined by the position of the mouse in the window.
- Varied Flight Speed: The user may vary the speed of the flight by positioning the cursor closer to the edge of the Main Flight Window to fly faster and positioning the cursor closer to the center of the window to fly slower.
- Print function: The user has the capability to print out the image or any part of it.
- Informational and Drop-down textboxes: These are textboxes and drop-down menus from which the user may select the desired information or data.
- Go-To function: This function allows the user to specify the latitude and longitude to which he or she wishes to travel. This currently loads the desired location directly.
- Sensor Band Controls: These controls allow the user to manipulate the sensor band combinations of Landsat TM data to view false color images. This provides greater flexibility and availability of information. For example, with the Landsat data, users are able to select from a list of seven possible sensors for each color band.
- RGB Intensity Control: This control allows the user to increase or decrease the intensity of the color bands.
- Smooth high-resolution flying over a variety of GIS data at varying speeds, directions, and altitudes over the Internet
- Interactive marking of an area for getting re-sampled GIS data by acquiring the original data, re-sampling it, overlaying it with other GIS data on the server and making the data available to the user via download or a CD

- The optional posing of queries via a URL (bookmarked or from a script) that can specify the desired region and resolution

**Semantic Wrapper**

We are developing technology that will allow other databases to be accessible via Sem-ODB. As a first step, we have developed a wrapper for relational database systems, which provides a semantic interface to relational databases, but our techniques are also applicable to other databases. The advantages of a semantic interface include friendlier and more intelligent generic user interfaces based on the stored meaning of the data, comprehensive enforcement of integrity constraints, greater flexibility, and substantially shorter application programs. Since SQL is the standard relational database query language that users are familiar with, we have defined Semantic SQL query language for semantic schemas.

Semantic SQL has the same syntax and extended semantics of standard relational SQL. Semantic SQL queries are interpreted over virtual tables, which span across categories in the semantic schema, rather than on static pre-defined tables in the relational schema. The virtual table(s) against which a particular query is interpreted is determined by examining the query statement. A major advantage that has been realized is that Semantic SQL queries over the semantic schema are much shorter and less complex than equivalent queries on the relational schema.

In developing our semantic wrapper, we have designed and developed four major components: Schema Loader, Knowledge Base, Translator and Knowledge Base Editor. The Schema Loader imports the relational schema into the knowledge base. It also creates an equivalent semantic schema for the relational database with derivation rules and stores it in the knowledge base using a bottom-up methodology. The Knowledge Base stores both the semantic and relational schemas along with derivation rules for query translation. We use a Semantic Database for the storage component of the knowledge base and are able to easily capture complex semantic information with the semantic schema. The Knowledge Base assists the DBA in making intelligent design decisions when creating complex semantic schemas and also keeps the meta-data consistent. The Translator translates Semantic SQL queries (based on the semantic schema) into equivalent relational SQL queries based on relational schema of the commercial RDBMS. It uses derivation rules as well as semantic and relational schema information stored in the Knowledge Base for this purpose. The relational SQL queries are transmitted to the RDBMS via an ODBC interface. The query results are converted to the appropriate format and transmitted to the user. The database administrator can use the Knowledge Base Editor to add complex features that are unavailable in relational databases (such as inheritance and m:m relations) to the semantic schema along with derivation rules.

The following is a list of our other principal accomplishments and research directions.

## 2.1. Semantic DBMS Advances and Advantages

- The semantic model gives us a view of the database that is closer to the real world. The semantic schema is easier to understand and expresses the problem domain with no restrictions.
- Many to many relations can be represented easily. In relational databases this has to be modeled by an additional table.
- Queries to semantic database (SDB) are shorter and more understandable than those to relational databases; no joins are necessary.
- User programs for a semantic database are substantially shorter than for a relational one, achieving major improvements in the application software development cycle, maintenance, and reliability.
- Data types are unlimited — strings can be of any length and we have developed techniques to represent numbers of unlimited length and precision.
- We have developed algorithms to provide very efficient full indexing, allowing fast access to every single fact in the database. Further, our algorithm guarantees optimality of the basic queries defined in our Semantic Algebra; this includes optimality of range queries.
- Objects can belong to several different categories at the same time. The operation to categorize/de-categorize objects can be performed efficiently and on-line.
- We have developed the original technique of lazy queries that allows disk accesses to retrieve facts to be delayed until they are actually needed (if they are needed at all). It also allows efficient query optimization, including lazy query intersection and subtraction.
- There is no need for NULL attributes. Sparse tables in relational databases may waste space and processing time.
- No keys are needed. Referential integrity constraints are supported automatically by the semantic database.
- We have developed and improved a semantic optimistic concurrency control algorithm supporting maximal theoretical granularity without the overhead that such precision would normally require. Further, the algorithm offers maximal safety.
- The semantic database is highly parallel. An efficient load balancing algorithm has been designed that will allow arbitrary chunks of data to be stored on different servers, optimizing the server performance.
- A Multiuser Semantic Database Engine is operational. A main goal of our work has been to achieve the quality that would make the SDB server viable as a commercial product. Additional features that have to be available in every commercial database system (integrity constraint checking, backup-recovery features, administrative tools, performance monitors, etc.) have been implemented. The theory necessary to support database versions has been developed and implemented. This allows each client to be fully isolated from all others clients and for each client to operate against a stable and consistent database. This feature is especially useful and efficient when used in Data

Warehouse applications.

- User interfaces to this engine for C++, C, and for Java have been developed.
- We have adapted SQL, the standard relational database language, to semantic databases. The original purpose of this adaptation was to be compatible with, and be able to communicate with, relational tools. Interestingly, it turned out that the size of a typical SQL program for a semantic database is many times smaller than for an equivalent relational database. While we have previously demonstrated substantial program-size advantage for other languages, we had not anticipated an even greater advantage with SQL — a specialized language for relational databases.
- Our ODBC driver provides a standard API one level above the SQL client library. It follows the syntax definition of the MS ODBC 2.0. A subset of the ODBC 2.0 that meets the requirements of relevant projects at HPDRC has been implemented. Our ODBC driver is operational. The current ODBC driver has been tested using Microsoft Access's Query-by-Example, Microsoft Access's wizards and tools for report and form generation, and Crystal Reports.
- We have also developed our own tools for use via the ODBC SQL interface — see below. Using these tools, the number of user keystrokes required is in correlation to the size of the generated SQL program. Since the SQL programs for the semantic database are substantially shorter, the third-party query tools are much more ergonomic with the semantic database than with the relational databases for which they were originally designed.
- An embedded SQL pre-processor has been developed and is fully operational.
- Semantic databases containing significant quantities of spatial data are in constant use for testing in the following areas: ocean temperature; ozone layer thickness; reflectivity, SeaWiFS (simulated) and LandSat. To this list we have added NOAA data that we receive daily from the National Hurricane Center on FIU's campus, and an aerial photography database describing large of areas of Miami-Dade County. These databases are stored using the improved version of the semantic binary engine which uses the binary storage engine described below.
- We have completed development of an Oracle relational database derived from our large semantic schema (over 2000 relations and attributes) for environmental research activities at the South Florida (Everglades) Research Center of the National Park Service. All data provided by the authorities at the Park have been loaded. The database is now in use by the scientists at Everglades National Park.
- A semantic database is being installed for use by the NOAA to manage wind data.
- A semantic database schema has been designed to store Landsat data that would allow users to efficiently retrieve any desired segment of Landsat data with graphical selection.
- Performance analysis for three different lossless compression techniques (pkzip, gzip, and IP_compression) has been carried out using spatial images. The IP_compression method yields the highest compression ratio. For performance enhancement, we have developed main-memory based compression software.

- We have continued development of a Simple Geographic Information System (GIS) which illustrates random access to spatial data at high speed.
- The current commercial Geographical Information Systems provide limited extensions for external database interface and query support. We have analyzed ArcInfo, ERDAS, and ENVI. We have investigated the existing database support for both vector, raster and attribute data. We have designed a scheme to integrate semantic database with GIS systems for efficient storage and retrieval.
- A self contained terrestrial data browser with an elegant graphical interface has been developed and named "TerraFly." In this project, the spatial data (images) that are of interest to the user community is partitioned into tiles and stored in a semantic database after compression. The database, the semantic DBMS, and a image browser are stored on a single CD ROM for client distribution. The versatile browser enables the user to fly over the terrestrial data (LandSat, etc.) in real-time (with the support of the semantic DBMS) with audio and video effects on any standard IBM PCs. The browser forms the basis of an edutainment CD ROM which was developed under Phase I support from the NASA STTR program for schools, database customers, and other interested parties. Recently added features of TerraFly include efficient processing algorithms and data compression for improved flight speed; sensor band controls; RGB intensity controls; a goto function which allows users to enter latitude/longitude; geographic names information system (GNIS) integration with LandSat and aerial photography; data text boxes; and more comprehensive help menus.
- The semantic database engine can run on a high performance parallel processing computer consisting of off-the-shelf Intel-PC compatible hardware — a Beowulf configuration akin to the Hive parallel computer developed at the NASA GSFC. We have built a Beowulf cluster consisting of 12 Gateway 2000 Pentium-II 300Mhz PC computers connected to a 100 Mbps Ethernet switch. We have developed a parallel binary storage container for our semantic database server that can be distributed over the Beowulf nodes. We have conducted performance tests by using the parallel semantic database server to provide data to TerraFly browsers. TerraFly requires real-time and intensive binary data retrieval from the database. A single TerraFly client consumes about 8Mbps of data bandwidth. With the current version of our semantic database server software, the semantic database running on our Beowulf cluster was able to handle 7 TerraFly clients (a total of about 56 Mbps throughput) without visible delays. We plan to improve the database software performance, so that the data will be delivered even faster to the clients and allow more parallel query processing. FIU has been awarded an NSF High Performance Connection grant that allows and pays for FIU's connection to Internet-2 (via the Abilene network); N. Rishe is the P.I. of this NSF grant. We intend to provide TerraFly and other databases stored using our semantic database via the high performance connection.
- Spatial data images often overlap with adjacent images and sometimes differ in size and orientation. A few expensive commercial software packages perform image registration with high computations. We have studied the methods to extract features from the images. We have developed mosaicing software that allows the user to produce huge pictures from many pieces of spatial data stored in our database.

- We explored integrating our semantic database technology into SeaDAS as a testbed for our Sem-ODB technology. This integration enhanced the functionality of SeaDAS by allowing users to randomly access different levels of data by performing arbitrary queries on data in addition to the main SeaDAS application and provided the other advantages of our semantic database to SeaDAS (parallelization, compression, backup and recovery, etc.).
- We have refined our prototype binary data storage engine. The developed prototype includes two levels of Binary Servers. The lower-level server efficiently retrieves the data from disk and returns it to the requester with minimal processing; its algorithms are very simple and it allows several requests to be served simultaneously. The data is stored in a way that does not necessitate any structural changes to the retrieved blocks and that does require time to be spent searching for the data. This is achieved by transferring the complexity of these algorithms to the upper-level server, which is dedicated to these particular tasks. The upper-level server does not store massive quantities of data; its functionality is to organize the data stored on the lower-level servers rather than to store objects itself. The current prototype includes several lower-level servers and one upper-level server that is responsible for maintaining the structure, taking requests from clients, finding the best algorithms for processing each request, and partitioning requests into sub-requests which are then distributed among the lower-level servers to parallelize execution. The lower-level servers are distributed over a Local Area Network and can communicate via TCP/IP or Netbios protocols. An optimization is being considered to transfer the burden of compiling the requested data together from the upper-level server to the client computer. This optimization should eliminate a communication bottleneck around the upper-level server by routing the bulk of the data from the lower-level servers directly to the clients. This, however, requires a network configuration that provides a direct interconnection between the client proxies and all levels of Binary Servers. A Beowulf cluster architecture developed at NASA is being used as the underlying parallel machine for our prototype; it uses a fast Ethernet switch that can support simultaneous connections between any computers in the cluster.
- We have developed an approach that allows us to add "stored procedure" capability to a semantic database system using Java byte-codes and Java's ability to dynamically load and execute Java code. Several steps were necessary: first we added a Java application programmer interface to the database system; then we created a database schema to hold Java executable code; then we constructed a Java class loader to allow code to be loaded from the database; then we enabled the creation of Java objects and executed the Java code for them. Our approach is not specific to our semantic database system, rather it can serve as a recipe for adding "stored procedures" to any database system.
- We integrated our semantic database technology with the NASA Regional Application Center v.1.0 software; we refer to this integrated software package as RAC-SDB. This integration was done as testbed and allowed our database to replace the commercial ObjectStore software. RAC-SDB does not require create index, create index root, os_relation commands, and query (create and bind) that the RAC software does require. RAC-SDB provides a number of advantages over the RAC v.1.0 software: a simple and high-level semantic representation of data; content-based search on all data; our Sem-ODB database provides a clean, modular, and simple programming interface; and other

advantages of our Sem-ODB database (ODBC compliance, an interactive visual-aided query tool, automatic report generators, etc.). Our integration was successful and was demonstrated to NASA.

- We have developed benchmarks for the semantic database engine. The purpose of the benchmarks is to show that while our SDB offers substantially better flexibility and logical properties, its performance is not much worse, and is often much better than that of the best other DBMS's. Specifically, we have designed and implemented a semantic benchmark, as well as the TPC-D standard benchmarks for relational databases. Our first semantic benchmark, SB1, running on a Pentium-200 computer with 128MB of RAM, showed that on certain types of queries the Semantic Database is 30 times faster than a highly-optimized fully-indexed Oracle database working on the same hardware. The Oracle database also requires about ten times more disk space than SDB. Using another variation of the schema of the Oracle database, we reduced Oracle space requirements to about 3 times more than that of SDB, but then the speed of Oracle was up to 120 times slower than SDB. Although the results are already quite favorable to SDB, we expect to further improve (reduce) the space requirements of SDB by implementing our new data compression algorithms in the next SDB release. The tests have also shown that Oracle requires a lot of fine-tuning to make it work fast on a particular benchmark. The Semantic database does not require any tuning at all, which means that the same installation of SDB will work equally well on all databases that are used on a server. Oracle, being tuned for one application, will not perform as well for another application, nor will it perform as well for the same application when users pose new types of ad-hoc queries. TPC-D results for the semantic database are comparable or better than Oracle and DB2.

## 2.2. Other Database Research Accomplishments

### 2.2.1. Efficient distributed heterogeneous databases management over bandwidth-on-demand networks (e.g. ISDN)

We have designed query decomposition and optimization strategies for queries that access heterogeneous databases over such networks. Strategies for reducing query response time as well as reducing monetary cost were developed in the framework. We have also implemented a heterogeneous database prototype using ISDN connections. ISDN bandwidth allocation algorithms for achieving minimal weighted sum of monetary cost and query response time were developed. We have leveraged this research into a grant from the Air Force Research Lab's Rome Laboratory.

### 2.2.2. Efficient multidatabase query processing

We have proposed a new multidatabase query processing technique that alleviates the long query response time that current solutions suffer from. The improvement was achieved by using fragmented joins, which are intended to shorten response time without sacrificing turnaround time. The algorithm has been implemented in a UNIX/ORACLE environment; preliminary experimental results have been encouraging.

We have performed simulations that compare the performance of two common approaches to supporting inter-database access: federated databases and loosely-coupled mulitdatabases.

The former uses a dedicated DBMS to process global queries, while the latter relies completely on the member DBMSs to execute global queries through an SQL interface. Our results have shown that the main disadvantage of the multidatabase approach — the inapplicability of pipelined processing — is outweighed by its better load balancing. The performance of both approaches is comparable. We are incorporating the fragmented join technique into the cost-effective multidatabase approach, making it even more efficient.

### 2.2.3. Disk clustering techniques supporting efficient visualization of remotely-sensed data

The problem is to assign data blocks of a large, two-dimensional, remotely-sensed database to multiple disks (or to a disk array) to support real-time navigation. We have reexamined and analyzed traditional disk declustering algorithms such as Disk Modulo, Exclusive-OR, and Hilbert Curve, which are aimed at range queries, in the new context of incremental queries exhibited by such navigation. We have devised a new declustering scheme based on Disk Modulo that has proven to be nearly optimal in the most realistic cases. We are planning to implement this scheme on a disk array server in order to evaluate its performance in a multiple-client environment.

### 2.2.4. Solution of satisfiability, implication, and equivalence problems in DBMS

We have developed a comprehensive solution to the satisfiability, implication, and equivalence problems. The problem has been widely encountered and is fundamental in database management systems. Our solution addresses the issue from a comprehensive perspective, and provides efficient solutions under various situations, which will be instrumental to the design and implementation of a database management system, and to database practitioners.

### 2.2.5. Advancement of query optimization

We have shown that queries can be optimized using semantic integrity constraints and rules, as well as by improved or more accurate cost models. We have proven that the optimization of certain types of queries is an NP-Complete problem, and have provided optimal algorithms under certain restricted cases. We have also provided more accurate cost models that would significantly improve the quality of a query optimizer in terms of its accuracy and efficiency.

### 2.2.6. Better distributed deadlock detection

We have designed a series of distributed deadlock detection and resolution algorithms that either reduce the delay of detecting a distributed generalized deadlock by half, or achieve optimal message complexity. In addition, these algorithms significantly simplify deadlock resolution because, unlike the existing algorithms, the information about the detected deadlock is collected at a single node.

### 2.2.7. Streamlined database design tools

We have developed a tool for design of relational databases, including schemas, integrity constraints, reports, and data entry forms, using semantic binary schemas. The tool is based on a top-down methodology; a conceptual description of an enterprise is designed using a semantic binary model. This description is then converted into the relational database design. The tool automates virtually all the busy work of design and can formulate the database design automatically using the semantics of the data and "rule-of-thumb" principles, or it will allow the sophisticated user to specify design details. The tool creates a turn-key database application and graphically-illustrated design reports, manuals, application glossaries, and data dictionaries, as well as an application-customized report generator. Changes in the semantic description or in the designer's instructions are propagated into the products.

### 2.2.8. Better multimedia access

We have developed:

- Novel algorithms which support data storage and retrieval in homogeneous and heterogeneous storage environments
- Data placement techniques that provide high performance in interactive video on demand environments
- Algorithms that support partitioning of compressed and uncompressed video data

# 3. LIST OF ALL PUBLICATIONS AND TECHNICAL REPORTS

## 3.1. Published

E. Alvarez, Y. Battikhi, N. Rishe, D. Barton, M. Martinez, A. Anzardo. "Semantic Database Schema Editor and Viewer Tool for Geospatial Applications." First International Conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL June 1-3, 1998, pp. I-30 - I-36.

E. Alvarez and N. Rishe "Multimedia Spatial Databases." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

E. Alvarez, N. Rishe, D. Barton, A. Anzardo. "Storage and Visualization of Ozone Layer Thickness Data." First International Conference Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL June 1-3, 1998, II-29 - II-36.

L. Amat, Jr., L. Florit, N. Rishe, D. Barton, A. Perez-Pons ''A 3-d Viewer for Superimposed Spatial Data Sets.'' Proceedings of the Twelfth International Conference and Workshops on Applied Geologic Remote Sensing, Denver, Colorado, 17-19 November 1997.

Y. Battikhi and M. Martinez "Semantic Schema Editor and Viewer Tool for Accessing, Designing and Interacting with Semantic Data Base." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

D. Beryoza, J. Uppal, P. Pardo, and N.Rishe "Interfacing Java to Semantic DBMS." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

G. Cao, N. Rishe. "A Nonblocking Consistent Checkpointing Algorithm for Distributed Systems." Proceedings of the IASTED Eighth International Conference Parallel and Distributed Computing and Systems October 16-19, 1996 - Chicago, Ill., pp 302-307.

G. Cao, M. Singhal, Y. Deng, N. Rishe, and W. Sun. "A Delay-Optimal Quorum-Based Mutual Exclusion Scheme with Fault-Tolerance Capability." 1998 IEEE International Conference on Distributed Computing Systems (ICDCS '98), Amsterdam, Netherlands, May 1998, pp. 444-451.

M. Chekmasov and M. Chekmasova "Database Design in Oracle with Srf As An Example." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

M. Chekmasov, N. Rishe, D. Barton, K. Madhyanapu, M. Chekmasova, R. Rodriguez. "Design and Implementation of the ENP Environmental Database using ORACLE." American Society of Photogrammetry and Remote Sensing/ Resource Technology Institute Annual Conference, Tampa, FL March 30-April 4, 1998. pp. 630-638.

C.M. Chen, R. Qiu, and N. Rishe "A Pipelined Query Processor for Distributed Multidatabases." Proceedings of the IEEE Southeast Conference, Apr. 24-26, 1998 Orlando FL. pp. 132-133.

C.M. Chen and N. Rishe, "Development of an Open and Scalable Web-based Information Publishing System." Proceedings of the 36th Annual ACM Southeast Conference, pp. 163-165, Apr. 1-3, 1998, Marietta, GA.

C.M. Chen and N. Rishe. "Fragmented Join: A Pipelined Multidatabase Join Method." Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, July 13-16, 1998, Las Vegas, Nevada.

C.M. Chen and R. Sinha, "Raster-Spatial Data Declustering Revisited - an Interactive Navigation Perspective", 15th Intl Conf on Data Engineering, Sydney, Australia, March 23-26, 1999. pp. 600-607.

C.M. Chen, W. Sun, N. Rishe. "Improving performance of Web-Database gateways using packed queries." IASTED 1997 Int'l Conf. on Software Engineering (SE '97).

C.M. Chen, W. Sun, N. Rishe. "Performance Comparison Of Three Alternatives Of Distributed Multidatabase Systems: A Global Query Perspective." IEEE Int'l Performance, Computing, and Communications Conference (IPCCC '98), pp. 53-59.

S.-C.Chen, M.-L. Shyu, N. Rishe. "Modeling Interactive Multimedia Presentation Systems Using Augmented Transition Networks" First International Workshop on Intelligent Multimedia Computing and Networking (IMMCN'2000), pp. 643-646, February 27-March 3, 2000, Atlantic City, NJ.

D.L. Davis-Chu "Creation of 3D Satellite Imagery." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

D.L. Davis-Chu, E. Alvarez, and N. Rishe. "The Creation of a System for 3D Satellite and Terrain Imagery." The 13th International Conference in Applied Geologic Remote Sensing, Vancouver, British Columbia, Canada, March 1-3, 1999. pp. II-329 - II-336.

R. Ege. "Storing Java Objects in any Database." Proceedings of Technology of Object-Oriented Languages and Systems, August 1-5, 1999, Santa Barbara, California. pp. 313-321.

R. Ghersgorin, A. Anzardo, and N. Rishe "Western Cultures Semantic Database Application." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

A. Gonzalez "Creating a Semantic Database to hold SeaWiFS Data, That Can Be Used By Scientists Studying Ocean Color." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

M. Gutierrez "Applications of Remote Sensing Data." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

F. Haayen. "GOES 8 GVAR Ingest System." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

T. Ho, E. Alvarez, N. Prabhakaran, D. Barton, N. Rishe, S. Graham. "Integration of a GIS and a Semantic Database System." First International Conference Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL June 1-3, 1998, pp. I-629 - I-636.

T. Ho, C. Pesantes, and V. Maddineni "Integration of a GIS and a Semantic Database System." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

O. Joukov, N. Rishe "Effective computer algebra for some applications" Proceedings of World Congress on Scientific Computation, Modelling and Applied Mathematics, Berlin, 1997, Vol. 2, pp. 761-765.

A. Kirienko, N. Rishe, J. Obando, H. Liu, J. Lu, W. Meng, and M. Zborovskiy "Techniques of Software Quality Assurance at the FIU High Performance Database Research Center." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

Y. Ling, W. Sun, N. Rishe, X. Xiang. "A Hybrid Estimator For Selectivity Estimation", IEEE Transactions on Knowledge and Data Engineering. pp. 338-354.

K. Madhyanapu and P. Pardo "Reverse Engineering: Relational to Semantic Database System." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

M. Martinez "GIS and Internet Access to Spatial Data." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

W. Meng, K. Liu, C. Yu, X. Wang, Y. Chang, N. Rishe, "Determining Text Databases to Search in the Internet", Published in the proceedings of VLDB '98, August 24-27, 1998, New York, NY, p. 14-25.

W. Meng, K. Liu, C. Yu, W. Wu, and N. Rishe. "Estimating the Usefulness of Search Engines." Proceedings of the 1999 International Conference on Data Engineering. pp. 146-153.

W. Meng, C. Yu, W. Wang and N. Rishe. "Performance Analysis of Three Text-join Algorithms." IEEE Transactions on Knowledge and Data Engineering. pp. 477-492.

W. Ng "Western Cultures Oracle Database." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

J. Pontillo "A Priori Selection of Ellipsometry Angles and Wavelengths for In Situ Process Control." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

G. Pozo, E. Perez, and R. Athauda "Landsat Viewer: Create Color Composite Images of Landsat Thematic Mapper Data." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

N. Prabhakaran, V. Maddineni, and N. Rishe. "Spatial Overlay of Vector Data on Raster Data in a Semantic Object- Oriented Database Environment." International Conference on Imaging Science, Systems, and Technology (CISST '99), June 28 - July 1, 1999, Las Vegas, Nevada, pp. 100-104.

N. Prabhakaran, N. Rishe, R. Athauda. "Tracking Hurricane Paths." Proceedings of the Image Registration Workshop, NASA GSFC. November 20-21, Maryland, 1997.

N. Prabhakaran, S. Sridhar, and N. Rishe. "A Two Phase Digital Ortho Photo Mosaicking System." International Conference on Imaging Science, Systems, and Technology (CISST '99), June 28 - July 1, 1999, Las Vegas, Nevada, pp. 151-154.

N. Rishe, D. Barton, M. Chekmasov, K. Madhyanapu, S. Graham, M. Chekmasova. "Everglades Data Integration using a Semantic Database System."International Conference Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL June 1-3, 1998, pp. I-567 - I-573.

N. Rishe, D. Barton, N. Prabhakaran, M. Gutierrez, E. Alvarez, R. Athauda, J. Tola-Rodriguez, A. Gonzalez "Landsat Data Visualization via the Internet" International Symposium on Spectral Sensing Research (ISSSR), San Diego, CA. December 13-19, 1997.

N. Rishe, D. Barton, N. Prabhakaran, M. Gutierrez, M. Martinez, R. Athauda, A. Gonzalez, S. Graham. "Landsat Viewer: A Tool to Create Color Composite Images of Landsat Thematic Mapper Data." International Conference Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL June 1-3, 1998, pp.I-529 - I-536.

N. Rishe, D. Barton, F. Urban, M. Chekmasov, M. Martinez, E. Alvarez, M. Gutierrez, P. Pardo. "High Performance Database Management for Earth Sciences." NASA University Research Center Technical Conference, Huntsville, AL. Feb 22-25, 1998, pp. 539-544.

N. Rishe and K. Naboulsi. "An Efficient Web-based Semantic SQL Query Generator." IEEE ICDCS'99 Web Based Applications Workshop. pp. 23-30.

N. Rishe, K. Naboulsi, and M. Olivero "Report Generator Web Interfaces to Semantic Database Cover." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

N. Rishe, O. Wolfson, and K. Naboulsi. "Report Generators." Encyclopedia of Electrical and Electronics Engineering, John Wiley & Sons, 1999. V. 18. pp. 500-513.

S. Sridhar and A. Gonzalez "Aerial Photography Internet Access Tool." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

O. Wolfson, L. Jiang, A. Sistla, S. Chamberlain, N. Rishe, M. Deng "Databases for Tracking Mobile Units in Real Time" 7th International conference on Database Theory (ICDT99), Jerusalem, Israel, January 10-12, 1999, pp.169-186.

L. Yan, W. Sun, N. Prabhakaran, S. Guo, Y. Deng, N. Rishe. "A Dynamic Hypermedia Model for Interactive Video." Proceedings of 13th International Conference on Advanced Science and Technology in conjunction with the 2nd International Conference on Multimedia Information Systems. April 97, Illinois.

Q. Yang, S. Dao, C. Yu, N. Rishe. "A parallel scheme using the divide-and-conquer method" International Journal of Distributed and Parallel Databases.

X. Ye and M. Gutierrez "Storage and Visualization of Ozone Layer Thickness Data." Proceedings of the Workshop on Next Generation Database Design and Applications, April 30 - May 1, 1998, Miami, FL.

C. Yu, K. Liu, W. Wu, W. Meng, and N. Rishe. "Finding the Most Similar Documents across Multiple Text Databases." Proc. of the IEEE Conference on Advances in Digital Libraries (ADL'99), Baltimore, Maryland, May 1999. pp. 150-162.

C. Yu, W. Meng, K. Liu, W. Wu, and N. Rishe. "Efficient and Effective Metasearch for a Large Number of Text Databases." ACM Conference on Information and Knowledge Management. pp. 217-224.

## 3.2. Accepted for Publication

S.-C. Chen, N.Rishe, X. Wang, M. A. Weiss. A User-Friendly Multimedia System for Querying and Visualizing of Geographic Data. The 4th World Multiconference on Systemics, Cybernetics and Informatics, Invited Session on Multimedia Systems Development, July 23-26, 2000, Orlando, Florida.

D. Davis-Chu, N. Prabakar, N. Rishe. Client-Server Based Real-Time Integration Of Remotely Sensed And Digital Data. CISST'2000: June 26-29, 2000, Las Vegas, Nevada.

N. Prabakar, Q. Jin, N. Rishe. Variation Assessment of Large Entities Using Spatial Data. The 2000 Int'l Conference on Imaging Science, Systems and Technology, CISST'2000: June 26-29, 2000, Las Vegas, Nevada.

E. Vaschillo, P. Lehrer, N. Rishe, M. Konstantinov. "Frequency Characteristics of Voluntarily-Induces Heart Rhythm and Blood Pressure Oscillations". Circulation, in press.

# 4. PARTICIPATING SCIENTIFIC PERSONNEL

**Faculty:** Naphtali Rishe, Maxim Chekmasov, Chung-Min Chen

**Other Personnel:** R. Aggarwal, R. Athauda, K. Beznosov, D. Chkliaev, M. Drobintsev, D. Eidelman, S. Fedorishin, J. Feng, O. Galina, R. Ghersgorin, P. Gupta, R. Hernandez, S. Hu, J. Hyams, A. Kadiyala, V. Lebedev, Z. Ma, K. Madhyanapu, R. Malayappan, V. Mallampati, M. Monga, K. Naboulsi, B. Parenteau, R. Qiu, M. Reddy, M. Sanchez, A. Shaposhnikov, Alexander Vaschillo, Alexandra Vaschillo, D. Vasilevsky, T. Williams, C. Wong, M. Woon Choy, H. Yang

**Degrees Awarded:** Ph.D.: K. Naboulsi, B. Parenteau, A. Shaposhnikov, Alexander Vaschillo; MS: R. Athauda, K. Beznosov, R. Ghersgorin, A. Kadiyala, V. Lebedev, V. Mallampati, M. Monga, R. Qiu, M. Reddy, Alexandra Vaschillo, T. Williams, H. Yang; BS: R. Athauda, J. Hyams, C. Wong

## 5. REPORT OF INVENTIONS (BY TITLE ONLY)

- On July 6, 1999, a patent covering the principles of our semantic database engine was awarded. (patent number 5920957, "Efficient Optimistic Concurrency Control and Lazy Queries for Databases and B-Trees).