INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Center of Excellence in Space Data and Information Sciences Annual Report, Year 9, July 1996 - June 1997

**B. DATE Report Downloaded From the Internet: 9 July 98**

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #: NASA Goddard Space Flight Center

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by: DTIC-OCA, Initials: __PM__ Preparation Date: 9 July 98

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.
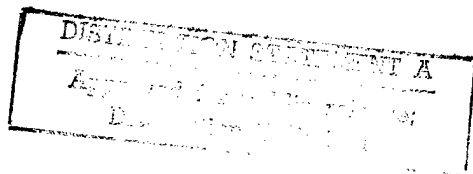
Center of Excellence in Space Data and Information Sciences

Annual Report

Year 9
July 1996 - June 1997
Dr. Yelena Yesha, Director

Table of Contents

CESDIS
NASA Goddard Space Flight Center
Code 930.5
Greenbelt, MD 20771

Phone: 301-286-4403 Fax: 301-286-1777

E-mail: cas@cesdis1.gsfc.nasa.gov

Operated by Universities Space Research Association in cooperation with the National Aeronautics and Space Administration

TABLE OF CONTENTS

Table of Contents

A. In Review. The CESDIS Newsletter

B. Seminars

C. Technical Reports

D. Personnel and Associates

FOREWARD

This report summarizes the range of computer science-related activities undertaken by CESDIS for NASA in the twelve months from July 1, 1996 through June 30, 1997. These activities address issues related to accessing, processing, and analyzing data from space observing systems through collaborative efforts with university, industry, and NASA space and Earth scientists.

The sections of this report which follow, detail the activities undertaken by the members of each of the CESDIS branches. This includes contributions from university faculty members and graduate students as well as CESDIS employees. Phone numbers and e-mail addresses appear in Appendix D (CESDIS Personnel and Associates) to facilitate interactions and new collaborations.

Table of Contents

OVERVIEW

CESDIS, the Center of Excellence in Space Data and Information Sciences, was developed jointly by the National Aeronautics and Space Administration (NASA), Universities Space Research Association (USRA), and the University of Maryland in 1988. It is operated by USRA, under a contract with NASA. The program office and a small, core staff are located on-site at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

USRA and the CESDIS Science Council

USRA is a nonprofit consortium of 80 colleges and universities, offering graduate programs in space sciences or related areas, which operates research centers and programs at several NASA centers. Most notable are the Lunar and Planetary Institute (LPI) at the Johnson Space Center in Houston, Texas, the Institute for Computer Applications in Science and Engineering (ICASE) at the Langley Research Center in Hampton, Virginia, and the Research Institute for Advanced Computer Science (RIACS) at the Ames Research Center at Moffett Field, California.

Oversight of each USRA institute or program is provided by a science council which serves as a scientific board of directors. Science council members are appointed by the USRA Board of Trustees for three-year terms. Members of the CESDIS Science Council during 1996-1997 were:

- Dr. Rama Chellappa - University of Maryland College Park
- Dr. Patricia Selinger - IBM Almaden Research Center
- Dr. Burt Edelson - George Washington University
- Dr. Harold Stone (Convener) - NEC Research Institute
- Dr. Richard Muntz - University of California, Los Angeles
- Dr. Satish Tripathi - University of Maryland College Park
- Dr. David Nicol - Dartmouth College
- Dr. Mark Weiser - Xerox PARC
- Dr. Jacob Schwartz - New York University

The CESDIS Science Council meets annually at Goddard to review ongoing CESDIS research programs and new initiatives.

The CESDIS Mission

CESDIS was formed to focus on the design of advanced computing techniques and data systems to support NASA Earth and space science research programs. The primary CESDIS mission is to increase the connection between computer science and engineering research programs at colleges and universities and NASA groups working with computer applications in Earth and space science. Research areas of primary interest at CESDIS include:

- High performance computing, especially software design and performance evaluation for massively parallel machines,
- Parallel input/output and data storage systems for high performance parallel computers,
- Parallel hardware and software systems,
- Database and intelligent data management systems for parallel computers,
- Image processing,
- Digital libraries, and
- Data compression.

CESDIS funds multiyear projects at U.S. universities and colleges. Proposals are accepted in response to calls for proposals and are selected on the basis of peer reviews. Funds are provided to support faculty

and graduate students working at their home institutions. Project personnel visit Goddard during academic recess periods to attend workshops, present seminars and collaborate with NASA scientists on research projects. Additionally, CESDIS takes on specific tasks for computer science research requested by NASA Goddard scientists.

A small, core staff is housed on-site at NASA Goddard. This staff includes USRA employees and university research personnel attached to CESDIS via subcontracts who work in one of three branches: Computational Sciences, Applied Information Technology, or Administration. The bulk of this report describes the work of each branch in detail.

Feedback and comments are encouraged electronically to: cas@cesdis.gsfc.nasa.gov

Table of Contents

DIRECTOR

DR. YELENA YESHA

(yesha@cesdis.edu)

Dr. Yelena Yesha is a tenured full professor in the Department of Computer Science and Electrical Engineering at the University of Maryland Baltimore County (UMBC), holds a joint appointment with the University of Maryland's Institute for Advanced Computer Studies (UMIACS) in College Park, and serves as the CESDIS Director through a memorandum of understanding between the University of Maryland and USRA.

Dr. Yesha received a Bachelor of Science degree in computer science from York University in Toronto, Canada in 1984, and a Master of Science and Ph.D. in computer and information science from Ohio State University in 1986 and 1989 respectively. She is a Senior Member of the IEEE Society, and a member of the ACM and New York Academy of Science. Her research interests include distributed databases, distributed systems, and performance modeling. She has authored more than 50 papers and edited six books in these areas.

Prior to joining CESDIS in December 1994, Dr. Yesha was on leave from the University to serve as the Director of the Center for Applied Information Technology at the National Institute of Standards and Technology. The Center's mission was to advance the goals of the National Information Infrastructure by identifying, developing, and demonstrating critical new technologies and their applications which could be successfully commercialized by U. S. industry.

Table of Contents

ACTIVITIES

- Attended a meeting at the White House with Thomas Kalil (Executive Director to the National Economic Council), Burt Edelson and Neil Helm (George Washington University), and Jim Johnson (GSFC Code 833.1). The purpose of the meeting was to discuss CESDIS involvement in the G-7 Information Technology Program and its potential role in coordinating projects that are part of that program.

- Gave a presentation about CESDIS technical work to the participants in the Visiting Student Enrichment Program sponsored by the USRAâs Goddard Visiting Scientist Program.

- Attended several Maryland Technology Alliance meetings at UMBC.

- Served as a guest editor (with Nabil Adam, Rutgers University) of the IEEE Transactions on Knowledge and Data Engineering special issue on digital libraries which appeared in August 1996.

- Hosted the annual meeting of USRA's CESDIS Science Council at GSFC in September 1996.

- Phil Merkey, Don Becker and I met with Mr. Koob, a program manager from ARPA, to discuss the proposal that CESDIS submitted jointly with computer science professors from Johns Hopkins University in the area of metacomputing.

- Served as one of the hosts for the Global Legal Information Network (GLIN) workshop for project directors from 12 countries.

- Attended a meeting with Abe Har'Aven, the Director of the Israel Space Agency, and Mr. Joe Rothenberg, the Director of NASA Goddard. The topic of the discussion was the scientific collaboration between the two agencies.

- Attended a workshop at Rutgers University and gave a lecture on CESDIS research.

- Visited Columbia University and met with Professor Al Aho, the Chairman of the Computer Science Department.

- Hosted a two-day meeting of the U.S.-Israel Science and Technology Commission. We presented the GSFC proposal for five projects in the area of information technology that was submitted to the Commission for possible funding.

- Met with Kathy Nado, the special assistant for outreach to the Director of Goddard, and Mr. Joe Rothenberg, the Director of Goddard. The topic of the meeting was the new presidential initiative in the Internet area.

- Visited Rome, Italy, where I attended a G-7 meeting on "The Global Marketplace for Small and Medium Enterprises" as a member of the U. S. delegation. The first three days were dedicated to meetings and the rest of the time was spent visiting Italian companies that are specializing in the area of electronic commerce.

- Hosted the Workshop on Data Matching and Mapping organized by Stanley Zdonik of Brown University and held at GSFC November 7, 1996. Presentations were made by Milt Halem (GSFC Code 930), James French (University of Virginia), Marc Postman (Space Telescope Science Institute), Susan Davidson (University of Pennsylvania), Yannis Ioannidis (University of Wisconsin), H. V. Jagadish (AT&T Laboratories), Serge Abiteboul (Stanford University), Peter Buneman (University of Pennsylvania), David Maier (Oregon Graduate Institute), Stanley Zdonik, George Lake (University of Washington), Raghu Ramakrishnan (University of Wisconsin), Peter Wegner (Brown University), Bob Grossman (University of Illinois), Mariano Consens (University of Waterloo, Ontario), Nabil Adam (Rutgers University), and Miron Livny (University of

Wisconsin). The text of the workshop report is included in the Consultants to the Director section of this report.

- Attended the annual international CASCON conference in Toronto which is sponsored by IBM. The conference was attended by 1500 scientists and engineers. I gave a tutorial on Challenges in Global Electronic Commerce and also chaired a workshop on electronic commerce.

- Met with Usia Galil, CEO of Elron Industries, several times to discuss the U. S.-Israel Information Technology Program, the involvement of Elron companies in pilot projects, and the NASA/CESDIS proposal to start a program in information technology that has been submitted to the U. S.-Israel Commission.

- Visited Shamim Naqvi, Bellcore Chief Scientist, to meet with Bellcore scientists who are working on developing technology for the Internet.

- Met with Avi Silberschatz when he visited GSFC to explore collaboration possibilities between his research group at Lucent Technologies at Bell Laboratories and CESDIS in the area of mass storage.

- Visited IBM Toronto Labs and met with scientists, developers, and Mr. Robert Leblanc, the Director of IBM Toronto Labs. The topic of the meetings was the joint research in the area of global electronic commerce.

- Traveled to France and England to establish collaborations with faculty at INSEAD in Paris, France, THESEUS in Sophia, Antibe, and the London Business School. Gave an invited presentation and chaired a panel at the conference on digital cash held at THESEUS.

- Visited the Department of Computer Science at Johns Hopkins University with Nabil Adam (Rutgers University) to meet with Yair Amir to review his research activities in the areas of networking and multi-media that are currently funded under the CESDIS contract.

- Traveled to Bonn, Germany to attend a G-7 conference on Global Marketplaces for Small and Medium Enterprises. Held numerous meetings with the representatives of business and government sectors from all over the world to discuss the role of GLIN (Global Legal Information Network) in the booming arena of global electronic commerce. Gave a presentation on the GLIN work conducted by CESDIS and the Library of Congress.

- Attended the Advances in Digital Libraries Conference held at the Library of Congress in May 1997.

- Met with Bill Howard (Director of USRA's Division of Astronomy and Space Physics) and several Goddard scientists to discuss the involvement of CESDIS in the SOFIA project.

Table of Contents

NEW INITIATIVES

Intercomparison of Automated Registration Algorithms for Multiple Source Remote Sensing Data

Proposal submitted in response to NASA Research Announcement NRA-97-MTPE-03, Satellite Remote Sensing Measurement Accuracy, Variability, and Validation Studies.

PI: Jacqueline Le Moigne (CESDIS)

Co-Investigators: James Tilton (GSFC Code 935), Samir Chettri (Global Science Technology, Inc.), Tarek El-Ghazawi (George Washington University), Emre Kaymaz, Bao-Ting Lerner, and John Pierce (KT-Tech, Inc.), Manohar Mareboyana (Bowie State University), David Mount and Nathan Netanyahu (University of Maryland College Park), and Srinivasan Raghavan and Wei Xia (Hughes STX).

Abstract: Many of the analysis techniques which will be utilized by the Mission to Planet Earth program will necessitate multiple data integration, which will require accurate registration of these data. Currently, the most common approach to registration is based on the manual extraction of a few outstanding characteristics of the data, called ground control points, from which a geometric transformation is computed. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for very large amounts of data. Also, since this interactive choice of control points is sometimes difficult, too few points, inaccurate points, or ill-distributed points might be chosen, thus leading to registration errors. For these and other reasons, automatic registration is becoming an important data analysis and production requirement.

Given the diversity of the data sources, it is unlikely that a single registration technique will satisfy all different applications. Although automated registration has been developed for a few Earth science applications, there is no general scheme which would assist users in the selection of a registration tool. In this work we propose to 1) develop an operational toolbox which consists of some of the most utilitarian registration techniques, and 2) provide a quantitative intercomparison of the different methods, which will allow a user to select for his/her application the desired registration technique based on this evaluation and the visualization of the registration results. The intercomparison will be based on accuracy, applicability, level of automation, and computational requirements criteria. These criteria will be computed for each algorithm utilizing several datasets which are relevant to the MTPE program, such as NOAA/AVHRR, Landsat/TM and MSS, GOES, and MODIS Airborne Simulator data.

The Khoros environment has been chosen as the framework for the implementation of these techniques. Since Khoros is an open software system, it will enable us to widely distribute the toolbox along with the results of our evaluation and to get feedback from a variety of users.

Integrating Environmental and Legal Information Systems

Proposal submitted in response to NASA Cooperative Agreement Notice CAN-97-MTPE-02, Extending the Use and Applications of Mission to Planet Earth (MTPE) Data and Information to the Broader User Community.

PI: Konstantinos Kalpakis, Assistant Professor, UMBC; CESDIS subcontractor on GLIN project.

Project Members: Susan Hoban (UMBC/CESDIS), Durwood Zaelke, David Hunter, and Gary Cook (Center for International Environmental Law), Steven Jamar (Howard University), Rubens Medina and Nick Kozura (Library of Congress), William Campbell (GSFC Code 935), and Pat Gary (GSFC Code 930).

Abstract: This project is a cooperative effort among the Universities Space Research Association, the University of Maryland Baltimore County, the Center for International Environmental Law (CIEL), the U. S. Library of Congress, and NASA's Goddard Space Flight Center. It is proposed to expand the use of MTPE science products to the field of environmental law, in much the same way as forensic technologies are now being applied to criminal law. One facet of the proposed effort will integrate

remotely sensed and in situ data with two existing legal systems: the Global Legal Information Network (GLIN), under development by the Library of Congress and NASA, and the Environmental Law Information Network Exchange (E-Line), constructed by CIEL. A second aspect of the proposal entails the development of a Model Environmental Legislation (MEL) system to teach law students how to craft environmental legislation which takes advantage of MTPE remotely sensed data. MEL will also be used as a teaching tool for comparative studies of existing environmental treaties and legislation.

Table of Contents

RESEARCH

Tools for Analyzing the Performance of Hierarchical Mass Storage Systems
(with Odysseas Pentakalos [NRC Research Fellow] and Daniel Menascé [George Mason University])

Hierarchical mass storage systems are becoming more complex each day and there are many possible ways of configuring them. The options range from the type and number of devices to be used to their connectivity. Furthermore, the demands placed on the mass storage systems are continuously increasing in intensity. This forces system managers to constantly monitor the system, evaluate the demand placed on it, and tune it appropriately using either heuristics based on experience or analytic models. This procedure involves two steps. First, workload characterization must be used to understand and describe in a concise manner the load imposed on the system. Then, a model of the system must be constructed and solved to detect the bottlenscks and evaluate various reconfigurations. Generating an accurate model of the workload through workload characterization is a laborious and time consuming process. Once the workload model has been obtained, constructing an accurate performance model of the system requires understanding the application, solution techniques, and limitations of analytic modeling. To automate both of these tasks we developed two tools: Pythia/WK, a tool for automated clustering-based workload characterization, and Pythia, an extensible object-oriented performance analyzer.

The main features of Pythia/WK are:

- Automatic support for peak-period determination: histograms of system activity are generated and presented to the user for peak-period determination.

- Automatic clustering analysis: the data collected from the mass storage system logs is clustered using clustering algorithms and tightness measures to limit the number of generated clusters.

- Reporting of varied file statistics: the tool computes several statistics on file sizes such as average, standard deviation, minimum, maximum, frequency, as well as average transfer time. These statistics are given on a per cluster basis.

- Portability: the tool can easily be used to characterize the workload in mass storage systems of different vendors. The user needs to specify through a simple log description language the way in which a specific log should be interpreted.

Pythia was designed and implemented to allow users to easily investigate the most cost-effective configurations for a given workload. One of the most important reasons to build such a tool is to provide a simple way through which queueing analytic models can be used for performance prediction and system sizing of mass storage systems. The tool incorporates a modeling wizard component that is capable of automatically building a queueing network model from a mass storage system representation defined through a graphic editor. Thus, the user of the tool does not need to know queueing network modeling techniques to use it.

The design of the analyzer was based on the following requirements:

- Architecture and system independence: the analyzer should not be tailored to any architecture even one as broad as the IEEE Mass Storage Systems Reference Model or be tailored to a specific system design such as the Unitree or IBM's DFSMShsm.

- Extensibility: the analyzer should be easily adaptable to new media technologies and devices.

- Graphical user interface: the analyzer should provide a graphical user interface for the specification of the particular mass storage system to be analyzed.

Figure 1 shows the main screen of the tool. Pythia provides a graphic editor for describing the architecture of the hierarchical mass storage system to be analyzed. The main screen consists of five components: the menu bar, the toolbar, the mode selection button, the canvas, and the status bar. The

menu bar provides access to the major functions of the tools such as file loading and saving through the file menu, canvas editing functions through the edit menu, performance solutions through the tools menu, and performance experiment plotting through the view menu. Keeping with the extensibility design requirement, the toolbar is dynamically created based on the objects specified within the backend database. Adding support to the tool for a new type of storage media or interconnect can be done by simply adding a record to the backend database.

One of Pythia's main features is a modeling wizard that builds a queuing network from the graphical description of the system using a set of heuristics. In what follows, we give an example of the heuristic used to model a robotic tape library placed at the nearline level. First we introduce the notation used to describe the heuristics. The lower case letters o, w, and q denote elements of the sets of mass storage system objects, workloads, and queuing network objects. The expression $o ==> q$ means that object o generates a single device q in the queuing network. The expression $a.b$ evaluates to the value of attribute b of element a. The function $HitRatio(o,w)$ returns the effective hit ratio of workload w on object o. It is defined as

$$HitRatio(o,w) =$$

1	if $o.num1 = 1$
$(1-w.hr_o)$	if $o.level = $ n and $o.numl = 2$
$(1-w.hr_o) \times w.hr_n$	if $o.level = $ n and $o.numl = 3$
$(1-w.hr_o) \times (1-w.hr_n)$	if $o.level = $ f and $o.numl = 3$

The function evaluates to 1 if the number of levels in the hierarchy is 1, it evaluates to $(1-w.hr_o)$ if the number of levels is 2, and the storage object is at the nearline level, it evaluates to $(1-w.hr_o) \times w.hr_n$ if the number of levels is 3, and the storage object is at the nearline level, and evaluates to $(1-w.hr_o) \times (1-w.hr_n)$ if the number of levels is 3 and the storage object is at the offline level. The following heuristic is used by the modeling wizard to convert a robotic tape library object o into components q_{robot} and q_{tape} of the QN.

The above heuristic indicates that two devices, q_{robot} and q_{tape} should be generated. The equations for $q_{robot}.sd_w$ and $q_{tape}.sd_w$ indicate how the service demands should be computed for these two QN devices. The different workload and object parameters that appear in the equations are obtained from the user during the specification of a mass storage system.

The generated QN models are then solved using the approximate multi-class MVA algorithm along with a set of approximations developed by the authors specifically for the domain of hierarchical mass storage systems. This approximations include techniques for dealing with the fork and join synchronization found in RAID-5 devices, and for dealing with the simultaneous resource possession exhibited by requests to transfer files from a network-attached tape drive to a network-attached disk drive. These approximations were validated against discrete-event simulations and also actual measurements at the mass storage system at NASA's Center for Computational Sciences.

The tool solves the queueing network generated internally to provide the user with performance information such as the throughput per workload measured, the response time of the system, the name of the bottleneck device, the residence time at the bottleneck device per workload, and the congestion factor (defined as the ratio of residence time over service demand) per class at the bottleneck device. Figure 2 shows the screen the presents the results of solving the analytic model.

Figure 2: Sample System Solution

In addition to solving the analytic model for the specified parameters, the tool also allows the user to

perform a number of experiments. The experiments are throughput and response time versus workload intensity and throughput and response time versus hit ratio. The user may go back and modify the configuration of the mass storage system and see the effects of the change by running the experiment again. The automatic generation of the analytic model from the graphical description along with the efficiency of the approximation algorithms make Pythia a valuable tool for capacity planning of mass storage systems.

The following papers were published as a result of this research effort.

Odysseas I. Pentakalos, Daniel Menascé, and Yelena Yesha. Pythia and Pythia/WK: tools for the performance analysis of mass storage systems, to appear in *Software Practice and Experience*.

Odysseas I. Pentakalos, Daniel Menascé, and Yelena Yesha. Pythia: A Performance Analyzer of Hierarchical Mass Storage Systems, 9th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, St. Malo, France, June 2-6, 1997.

Odysseas I. Pentakalos, Daniel Menascé, and Yelena Yesha, Automated Clustering Based Workload Characterization, *Proceedings of the 5th Goddard Conference on Mass Storage Systems and Technologies*, September 1996, College Park, Maryland.

Strategies for Maximizing Seller's Profit Under Unknown Buyer's Utility Values (with Konstantinos Kalpakis and Bella Bellagradek, UMBC)

We studied a simple market model with very restricted information available to the participants. The main incenitive to our work is the lack of reliable or up-to-date information in the existing markets, especially in the electronic markets where the situation tends to change rapidly.

Our model employs the following protocol of sales: at time moment (we assume that time is divided into some units, not necessarily uniform, numbered by non-negative integers). The seller posts the current price per unit of a product together with the number of units available for sale. At the same time the buyer reads the seller price and makes a decision to buy or not to buy, based on the comparison of the posted price and the buyer's utility value unknown to the seller initially. This process repeats a given number of times or until the product supply lasts or indefinitely.

The main question is "Is there an algorithm (pricing strategy) that allows the seller to maximize the collected profit, using mainly the history of his/her own sales?" Some additional questions:

1. How do we measure the optimality of the algorithm?
2. What additional assumptions are to be made to get meaningful results?
3. What are the computational complexities of the proposed algorithms?

We obtained the following results in assumptions that the buyer's utility value is a constant in a given interval and only single unit of the product can be sold at any moment of time:

1. There are polynomial time/space dynamic programming algorithms that maximize the average case profit or the expected profit under a given distribution of buyer's utility values. This method works when both supply and time to sell are unbounded or one or both of them are finite.

2. We also studied the worst case when both supply and time to sell are unbounded. We got the lower and upper bounds for the loss of a pricing algorithm that differs by factor of order logN. The loss of an algorithm is a difference between the best achievable profit for a given input and the actual profit made. It is a convenient measure of "goodness" of an algorithm, often used in computational learning theory.

Future work lies in the direction when the buyer's utilities are varying in time rather than constant.

Publication:

Starategies for Maximizing Sellers' Profit Under Unknown Buyers' Utility Values (with Kostas Kalpakis and Bella Bellagradek), submitted to CASCON97.

Evolving Databases: An Application to Electronic Commerce (with Serge Abiteboul [Stanford University], Brad Fordham [NIST], and Konstantinos Kalpakis [UMBC])

The evolving database represents a universe of discourse by capturing arbitrarily many semantic dimensions of its constituent entities. Here a "semantic dimension" is defined by the knowledge engineer by specifying two things: 1) the syntax(es) for values which populate that semantic dimension and 2) an evolving algebra specification of the "semantics" implied by each syntax.

This renders ALL semantics of an entity into a common form, the evolving algebra, which can be executed automatically -- allowing the user to experience the interactions of the various entity semantics. In effect, the evolving database has two powers. First, it captures static entity representations and relationships. Second, it maintains known semantics over time as stimuli external to the system change the evolving database's state.

The term "semantic dimension" in very broad. This is possible because the underlying evolving algebra formalism is a Turing-complete formal specification methodology. Some examples of semantic dimensions may be:

1. DATA defined with a syntax "?attribute=?value" and instantiated as color=green, weight=1.5.

2. CONSTRAINT defined with syntaxes "?attribute BETWEEN ?min AND ?max", "?attribute < ?value" and instantiated as weight BETWEEN 1 AND 5, height < 12.

3. I/O-BEHAVIOR defined with a syntax "?inputs ==> ?outputs" and instantiated as paper, toner, power ==> hardcopy

This broadness in the semantic dimension, the key to our data modeling and manipulation approach, permits a clean integration between many traditional semantic notions found in current DBMSs including: relationships like inheritance, constraints, behaviors of various flavors, inference, and so on. Thus, the major objective of the evolving database is to provide a DBMS in which very semantically complex and dynamic entities can be fruitfully modeled and managed.

Publications:

Evolving databases: an application to electronic commerce, (with Serge Abiteboul and Brad Fordham), to appear in the *Proceedings of the International Database Engineering and Applications Symposium*, Montreal, Canada, August 25-27, 1997.

Electronic commerce: current limitations and future visions (with Kostas Kalpakis and Brad Fordham), invited for IEEE *Transactions on Knowledge and Data Engineering*.

Table of Contents

SELECTED PUBLICATIONS

Publications in refereed journals

Adam, N., Awerbuch, B., Slonim, J., Wegner, P., & Yesha, Y. (1997). Globalizing business, education, culture through the Internet. Communications of the ACM, 40(2), 115-121.

Adam, N., El-Ghazawi, T., Halem, M., Kalpakis, K., & Yesha, Y. (1996). The Global Legal Information Network. The American University Law Review. 46(2), 477-491.

Adam, N., Halem, M., Holowcazak, R., Lal, N., & Yesha, Y. (1996). Digital Libraries Task Force, IEEE Computer.

Adam, N. & Yesha, Y. (Editors). (1996). Special section on digital libraries, IEEE Transactions on Knowledge and Data Engineering.

Yesha, Y., & others. (1996). Strategic directions in electronic commerce and digital libraries. ACM Computing Surveys. 28(4), 818-835.

Articles/Papers Accepted

Kalpakis, K., Awerbuch, B., and Yesha, Y. Towards free information markets. Mathematical Modeling and Scientific Computing.

Kalpakis, K., Fordham, B., & Yesha, Y. Electronic commerce: Current limitations and future visions. IEEE Transactions in Knowledge and Data Engineering.

Pentakalos, O. I., Menascé, D., & Yesha, Y. Pythia: a performance analyzer of hierarchical mass storage systems. PNPM/Tools Conference, San Malo, France, June 3-6, 1997.

Pentakalos, O. I., Menascé, D., & Yesha, Y. Pythia and Pythia/WK: Tools for the performance analysis of mass storage systems. Software Practice and Experience.

Books

Adam, N., and Yesha, Y. (Eds.). (1996). Electronic commerce: Current research issues and applications. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag.

Adam, N., Halem, M. & Yesha, Y. (Eds.). (1996). Proceedings of the Third Forum on Research and

Technology Advances in Digital Libraries (ADL '96). Los Alamitos, CA: IEEE Computer Society Press.

Chapters in books

Adam, N., & Yesha, Y. (1996). Electronic commerce: An overview. In Adam, N., and Yesha, Y. (Eds.). Electronic commerce: Current research issues and applications. (pp. 5-12). Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag.

Publications in proceedings

Pentakalos, O. I., Menascé, D., & Yesha, Y. (1996). An analytic model of hierarchical mass storage systems with network-attached devices. In ACM SIGMETRICS `96, Philadelphia, PA.

Pentakalos, O. I., Menascé, D., & Yesha, Y. (1996). Automated clustering-based workload characterization. In Proceedings of the Fifth NASA Goddard Conference on Mass Storage Systems and Technologies, College Park, Maryland.

Table of Contents

CONSULTANTS TO THE DIRECTOR

Task 1 on the CESDIS contract (the general administrative task) allows the Director to bring to CESDIS consultants who are not funded by specific task originators. CESDIS entered into agreements with the individuals reported upon in this section for the purpose of program development.

Serge Abiteboul
Stanford University
Department of Computer Science

Nabil Adam
Rutgers University
Center for Information Management, Integration, and Connectivity

Ian Akyildiz
Georgia Institute of Technology
Broadband and Wireless Networking Laboratory

Data Mapping and Matching Group
Serge Abiteboul (Stanford University), Peter Buneman (University of Pennsylvania), David Maier (Oregon Graduate Institute), Stanley Zdonik (Brown University)

Burt Edelson
George Washington University
Department of Electrical Engineering and Computer Science

Miron Livny
University of Wisconsin
Department of Computer Science

Daniel Menascé
George Mason University
Department of Computer Science

Mukesh Singhal
Ohio State University
Department of Computer and Information Science

Jacob Slonim
University of Toronto(Ontario)
Computer Science Department

Peter Wegner
Brown University
Department of Computer Science

Table of Contents

Serge Abiteboul

Stanford University, Department of Computer Science
(abitebou@db.stanford.edu)

Statement of Work

Dr. Abiteboul has worked with the CESDIS Director to model the state of an electronic commerce transaction as an active database with an emphasis on communication with the external world. Rules and constraints are used to describe the agreed upon laws that govern the transaction and the protocols that describe how participants interact with the database. The goal of the effort is to determine which portion of database technology is well suited in this context, which portions have to be extended, and which features (such as process modeling tools) are missing. Providing tools with formal semantics to describe electronic commerce transactions has also been considered.

Results

Most of this work was performed while at Stanford. I also visited CESDIS for two weeks in July 1996 and for shorter visits. The work on electronic commerce (EC) was performed with Prof. Yelena Yesha (CESDIS) and Brad Fordham (NIST).

The goal was to investigate a formal approach to electronic commerce modeling based on Gurevich evolving algebras.

There is a real need for more formal foundations for definitions in that domain are often unclear and this generates confusion and a lot of redundancy in efforts. Evolving algebra is a formalism originally proposed to formally specify program semantics. Some work was needed to make it suited for describing EC applications. We developed the needed concepts. A system based on these ideas was implemented and the approach validated with some particular applications. Although more work is needed, we can already say that the work proved to be quite successful.

What has been achieved:

1. Evolving algebras are well-suited for capturing EC applications. However, they do not provide a user-friendly specification language. We developed ``evolving databases'', a general purpose model convenient for specifying such applications and in particular suited to describe their active component (via rules). The general approach is described in [1].

2. We developed a first prototype to validate the ideas. The developer and prime architect is Brad Fordham. Some applications were implemented.

3. We set up the basis for a formal study of customizable EC models based on simple rules (extending DATALOG) [2]. The main focus is on adding ``active features'', while keeping the simplicity of the language. This is in order to provide automatic help (e.g., what the user should do next) and customization checking. The work was initiated from discussion with Al Aho (Columbia University) and Alberto Mendelzohn (University of Toronto) in the summer.

The following future efforts are considered:

- Develop a complex application that would in particular highlight the possibilities of the model with respect to customization, and on-line modification of the EC model.

- Understand better how the model can integrate standards such as EDI or integrate tools such as available payment mechanisms.

- The integration of Behavioral Datalog [3] in the prototype.

Efforts in related themes:

- semistructured data: how to query data that is very irregular [3,4,5,6,7]

- distributed query optimization: how to optimize and restructure data [5,8,9]

- theory of the Web: trying to capture the essential aspects of Web computation and complexity vs. more standard kinds of computations. [10,11]

References:

1. Fordham, B., Abiteboul, S., & Yesha, Y. (1997), Evolving Databases: An Application to Electronic Commerce, International Database Engineering and Applications Symposium (IDEAS), Montreal.

2. Abiteboul, S., Fordham, B., & Yesha, Y., Behavioral Datalog, in preparation.

3. Abiteboul, S. (1997) Querying semistructured data, Proceedings of the International Conference on Database Theory, Delphi, Greece. (invited paper)

4. Abiteboul, S., Quass, D., McHugh, J., Widom, J., & Wiener, J. (April 1997) The lorel query language for semistructured data, International Journal on Digital Libraries, 1(1):68-88.

5. Abiteboul, S., Cluet, S., & Milo, T. (1997) Correspondence and translation for heterogeneous data, Proceedings of the International Conference on Database Theory, Delphi, Greece.

6. Abiteboul, S., Goldman, R., McHugh, J., Vassalos, V., & Zhuge, Y. (1997) Views for semistructured data, International Workshop on Management of Semistructured Data, Tucson.

7. Nestorov, S., Abiteboul, S., & Motwani, R. (1997) Inferring Structure in Semistructured Data, International Workshop on Management of Semistructured Data, Tucson.

8. Abiteboul, S., Cluet, S., Christophides, V., Milo, T., Moerkotte, G., & Simeon, J. (April 1997) Querying Documents in Object Databases, International Journal on Digital Libraries, 1(1):5-19.

9. Abiteboul, S. & Vianu, V. (1997) Regular Path Queries with Constraints, Proceedings of ACM Principle of Database Systems, Tucson.

10. Abiteboul, S. & Vianu, V. (1997) Queries and Computation on the Web, Proceedings of the International Conference on Database Theory, Delphi, Greece

11. Papakonstantinou, Y., Abiteboul, S., Garcia-Molina, H. (1996) Object Fusion in Mediator Systems, Proceedings of the International Conference on Very Large Data Bases, Bombay.

Consultants to the Director

Table of Contents

Nabil Adam

Rutgers University,
Center for Information Management, Integration, and Connectivity
(adam@adam.rutgers.edu)

Statement of Work

Dr. Adam worked as a CESDIS subcontractor to do the following:

- Provide technical management of CESDIS research projects and help nurture an environment of interactive supervision of the CESDIS branch heads.

- Assist the CESDIS Director and technical staff members in developing a proposal in response to the next CESDIS contract procurement notice.

- Assist the CESDIS Director and technical staff in developing new initiatives.

- Help increase the visibility of the CESDIS research staff by developing stronger ties to and improving communication with the NASA scientific community at Headquarters, GSFC, and the other NASA centers.

- Help increase the visibility of CESDIS within the scientific community by encouraging and facilitating the acceptance of articles describing CESDIS-supported research by journals and papers by conference committees.

Results

I held meetings with and provided technical oversight and guidance to Drs. Yair Amir (Johns Hopkins University) and Aya Soffer (CESDIS/UMBC) regarding the digital libraries project. I accompanied Dr. Yesha on a visit to the facilities of the Johns Hopkins Computer Science Department. In addition, I explored with Dr. Soffer a possible collaboration between CESDIS and Rutgers University on the Regional Validation Center project. I also met with Dr. Susan Hoban (CESDIS/UMBC) to discuss increasing the visibility of the digital libraries program through the Advances in Digital Libraries `97 conference.

I met with Jim Fischer (Code 930) to discuss CESDIS's role in the HPCC program. I also met with Karen Moe (Code 522) to discuss CESDIS's role in the Mission to Planet Earth (MTPE) Program. I attended Goddard reorganization meetings and studied the new organization at GSFC as well as the strategic direction of NASA. I identified CESDIS in-house expertise as well as outside expertise that complement that of CESDIS which will enable us to build a world class team that is needed to meet GSFC's new environment and NASA's long term strategic direction. At this point it seems that MTPE and Regional Validation Centers have strong potentials and CESDIS should increase its future involvement in these areas. By their very nature, problems related to these areas are complex and require a team that is multidisciplinary in nature. For CESDIS to be able to effectively compete for the new contract, it is critical to build on its current strength. This includes the work by Yelena Yesha, Jacqueline Le Moigne, Phil Merkey, Don Becker, and Rick Lyon.

In an attempt to explore opportunities within GSFC, I held meetings with the SEWP team to identify possible CESDIS involvement including the development of innovative algorithms and software. Furthermore, Dr. Yesha and I met with Douglas Norton of NASA HQ (Director, Program Integration Division) to explore creating a pilot project for on-line NASA awards. I also began and continue to work on the following initiatives.

1. Maximum Entropy and Maximum Likelihood Restoration of Atmospherically Degraded Imagery - Rick Lyon (CESDIS/UMBC) and N. S. Kopeika (Ben-Gurion University, Israel)

Both Mr. Lyon and Dr. Kopeika have much to offer on this rich and exciting problem. A collaborative effort between them would be fruitful for both and potentially beneficial to NASA and the general scientific community for optimal information extraction from both satellite-based Earth sensing systems and ground-based remote sensing systems. Dr. Kopeika brings to the effort an understanding of atmospheric contributions and a model of the atmospheric blurring process; Mr. Lyon brings the contributions due to the telescope, detector, and noise statistics as well as a suite of high fidelity maximum entropy and maximum likelihood image restoration algorithms developed, coded, and implemented on a MasPar MP2. Mr. Lyon has a number of publications in both sensor modeling and algorithm development.

2. Possible CESDIS involvement in the NASA/USRA SOFIA Project - Joe Bredekamp, NASA HQ Office of Space Science, Research Program Management Division

   The Stratospheric Observatory for Infrared Astronomy (SOFIA) will be a 2.5 meter, optical/infrared/submillimeter telescope mounted in a Boeing 747, to be used for advanced astronomical observations performed at stratospheric altitudes. More than 160 science flights per year are planned with data collected 60 times faster than by the preceding flying observatory, the Kuiper Airborne Observatory. I believe that CESDIS involvement in managing the data from the SOFIA project is a good idea. SOFIA is expected to generate massive amounts of data. CESDIS's ability to handle massive amounts of data and its position at Goddard make it a strong candidate.

3. Held several meetings and discussions with Karen Moe (Goddard Mission Operations and Data Systems Directorate, Data Systems Technology Division, Software and Automation Systems Branch) about possible collaboration with CESDIS in FY98. Potential projects include:

   o Video conferencing: Dr. Yair Amir's work is a natural fit.

   o Risk management: Dr. Al Aho (Columbia University) was a CESDIS visitor in the summer of 1996. His research interests and practical experience in software engineering for large systems would be valuable for such a project.

   o CORBA (Common Object Request Broker Architecture): CORBA is not only capable of handling platform heterogeneity of underlying information systems, but, more importantly, the heterogeneity of applications used at different information systems. With CORBA-standard products, an application can communicate with another totally different application through a common interface such as the Interface Definition Language (IDL). Each of the application programs has to maintain an IDL interface (for client and server). A change in the server application program requires all the IDL interfaces (of each client as well as the server) to be updated (recompiled) accordingly. The work at Rutgers related to integrating heterogeneous and autonomous information sources would be of relevance here.

- I held regularly scheduled (every 2-3 weeks) technical staff meetings and individual meetings withthe technical staff. The following is a summary of some of these meetings.

   1. Dan Ridge made a brief presentation on Linux which generated a very lively discussion with a number of interesting questions. One of the questions raised was whether we could explore ways for combining the current CESDIS efforts in the ACTS, Linux, and Beowulf projects. Dan, Tarek El-Ghazawi, and Burt Edelson will explore further.

   2. Phil Merkey made a presentation on Beowulf.

   3. Yair Amir presented his work on videoconferencing and multicasting.

   4. Kostas Kalpakis (CESDIS/UMBC) discussed some of the technical challenges facing the GLIN team. Commonalities between some of the GLIN-related search concepts and some of Nathan Netanyahu's (CESDIS/University of Maryland College Park) work in the context of image registration was discussed.

- I served as the General Chair of the Forum on Research and Technology Advances in Digital Libraries (ADL '97) sponsored by the IEEE Computer Society that was held May 7-9, 1997 at the Library of Congress in Washington, D. C.

  1. I helped secure the sponsorship of the IEEE Computer Society, the National Library of Medicine, and the Library of Congress in addition to CESDIS and NASA Goddard Space Flight Center.

  2. I helped secure the participation of Robert Price (Director of the Mission to Planet Earth Program Office at Goddard), Bruce Lehman (Assistant Secretary of Commerce and Commissioner of Patents and Trademarks), as well as Larry Irving (Assistant Secretary of Commerce for Communications and Information).

  3. Karen Moe chaired a panel on "Very Large Digital Libraries" and John Dalton (Deputy Associate Director, Goddard Mission Operations and Data Systems Directorate, Earth Science Data and Information System Project) was one of the participants of the panel.

- Publication and Presentations

  1. Published/submitted the following papers which include acknowledgment of CESDIS support:

     Adam, N., Awerbuch, B., Slonim, J., Wegner, P., & Yesha, Y. (1977). Globalization and the future of the Internet. Communications of the ACM. February.

     Adam, N., Atluri, V., & McKeown, K. (1997). Clinical information systems: making use of research in digital libraries. NSF Workshop on Research and Development Opportunities in Federal Information Services. May.

     Adam, N., Atluri, V., & Huang, W. Modeling and analysis of workflows using Petri Nets. Submitted April 1997.

     Adam, N., Adiwijaya, I., Atluri, V., & Yesha, Y. EDI Through A Distributed Information Systems Approach.

  2. Published the following book/proceedings which include acknowledgment of CESDIS support.

     Adam, N., & Gangopadhyay, A. (1997). Database Issues in Geographic Information Systems. Kluwer Academic Publishers. June.

     Adam, N., & Aho, A. (Eds). (1997). Proceedings of the IEEE International Forum on Research and Technology Advances in Digital Libraries (ADL'97). Los Alamitos, CA: IEEE Computer Science Press. May.

  3. Made several presentations on CESDIS-supported work including the following:

     IEEE Computer Society Board meeting, May 1997.

     NSF Workshop on Research and Development Opportunities in Federal Information Services, May 1997.

     Department of Computer Science, SUNY at Buffalo, April 1997.

     GCDIS, March 1997.

- Other

    1. Next Generation Information Technologies and Systems '97 conference: Co-organized and co-chaired a panel on digital libraries with Dr. Yesha as part of the NGITS'97 conference that was held June 30-July 3, 1997 in Neve Ilan, Israel. Panel participants included researchers from Bell Laboratories, Carnegie Mellon University, University of Maryland College Park, and George Mason University.

    2. Chairing the IEEE Computer Society Task Force on Digital Libraries which is a collaboration among academia, industry, and NASA. Members of the task force include Dr. Milton Halem (Code 930), Dr. Harold Stone (NEC Research Institute), and Dr. Yesha.

    3. Lead a CESDIS-supported research project that resulted in a Ph.D. dissertation in the area of digital libraries, May 1997.

    4. Served as Editor-in-Chief with Dr. Yesha for the International Journal of Digital Libraries. The first issue was published in May 1997.

Consultants to the Director

Table of Contents

Ian F. Akyildiz

Georgia Institute of Technology,
Broadband and Wireless Networking Laboratory
(ian@ee.gatech.edu)
http://www.ee.gatech.edu/users/ian

Statement of Work

Dr. Akyildiz was tasked with advising the CESDIS Director on any or all of three proposed areas: 1) an efficient traffic/congestion control mechanism in ATM over satellite enviroment, 2) A LANs/MANs interconnection architecture using ATM over satellite, or 3) mobility management for multi-tier personal communication systems.

Results

1. Introduction

During my very productive stay at CESDIS during the summer of 1996, I prepared survey reports, wrote

proposals, participated in several meetings, seminars, and group research activities. The detailed activities are summarized below:

2. High Performance Networking

The purpose of this report [1] was to point out the state-of-the-art problems and short-term and long-term research needs. The areas covered are ATM networks, satellite ATM networks, wireless networks, mobile computing. This report should serve NASA CESDIS a basic framework to start a new research direction.

3. Wireless Networks

I also led the effort to develop a research proposal [2]. The proposal had the objective to research highly mobile wireless multimedia architectures for the needs of the digital battlefield with capability for intelligent survivability and adaptive connectivity in a hostile environment. The project will contribute to the foundation of intelligent highly mobile wireless multimedia network design, in particular, network architecture, intelligent distributed database management, routing, multicasting, location management, channel allocation, information security, digital information storage, and power control.

4. Satellite Networks

I investigated satellite networks in two directions: satellite ATM networks and mobility management.

4.1. Satellite ATM Networks: A Survey

The survey in [3] points out the key issues for interconnecting satellite and ATM networks. ATM technology is useful in multiplexing various traffic types such as data, voice, video and still images. There are several open issues which need to be overcome in order to achieve a seamless integration of ATM and satellite networks. First the requirements for the interconnection are described in [3], followed by a discussion of recent research issues, challenges, and possible solutions. Finally, an overview of current projects about ATM over satellite is given and future directions are discussed.

4.2. Handover Management over LEO Satellite Networks

Low Earth Orbit (LEO) satellite systems have been proposed in recent years to provide global coverage to a more diverse user population. In contrast to geostationary (GEO) satellites, LEO satellites circulate the Earth at a constant speed. Because of this non-stationary characteristic, the coverage area of a LEO

satellite changes continuously. The serving satellite for a particular connection may change over time resulting in a handover. Thus, LEO satellite networks require a reliable handover protocol that is critical for connections with multihop intersatellite links (ISLs). I introduced the Footprint Handover Re-route Protocol (FHRP) in [4] that maintains the optimality of the initial connection route without performing the routing algorithm after satellite handovers. Furthermore, the FHRP handles the inter-orbit handover problem which has been neglected in the existing literature.

References:

[1] Akyildiz, I. F. (September 1997), A Handover Management Protocol for LEO Satellite Networks,

Technical Report, August 1996. Also to appear in ACM/IEEE Mobicom'97 Conference.

[2] Akyildiz, I. F., & Jeong, S. H. (July 1997), Satellite ATM Networks: A Survey, IEEEE Communications.

[3] Akyildiz, I. F., Gelenbe, E., Singhal, M., Wiederhold, G., Wolfson, O., & Yesha, Y. (September 1996).

Intelligent Agents for Adaptable Wireless Networks. (Technical Report)

[4] Akyildiz, I. A., El-Ghazawi, T., & Yesha, Y. (August 1996). Research Challenges in High Performance

Networking Area. (Technical Report)

Table of Contents