FINAL REPORT FOR AWARD # 9905026

Edward A Fox ; *VA Polytechnic Inst & St U*
1999 NSF Information and Data Management Workshop: Research Agenda for the 21st Century -- IDM 99

Participant Individuals:

Senior personnel(s) : Alfonso F Cardenas; Layne T Watson; Narendran Ramakrishnan
Technician, programmer(s) : Debra K Dudley; Karen Sewell

Participants' Detail

Partner Organizations:

University of California Los Angeles: Financial Support; In-kind Support; Facilities; Collaborative Research

Professor Alfonso F. Cardenas of UCLA was PI on the parallel grant (9904998) from NSF to run the IDM'99 workshop. That was held at UCLA and helped by many students and faculty there. Many facilities at UCLA were made available for the meeting.

Other collaborators:

Many people helped with the IDM'99 meeting and with the October 2000 IDM workshop as well, aiding in the planning for the events, speaking, giving demonstrations, joining discussions at the meetings, writing portions of reports, etc.

For names of other participants, please refer to the NSF IDM '99 web page (http://www.cs.ucla.edu/csd/IDM99/) and to the Oct. 2000 workshop home page (http://fox.cs.vt.edu/IDM/).

We also have had contact and collaboration with representatives of other governmental and industrial organizations, including the National Institute of Standards and Technology (NIST) and the United States Geological Survey (USGS).

Activities and findings:

Research and Education Activities:

The recent advances in computer, networking and storage technologies, the explosion of publishing, and the vast increase in data availability from the Internet and other sources such as satellites, have enabled the emergence of an unprecedented number of new computer applications, which present new challenges to the ways data and

information are used and managed. These challenges will shape both the research agenda as well as the technology to be developed for the 21st century.

The 1999 workshop brought together the PIs and Co-PIs funded in 1998-99 by the Information and Data Management Program (IDM) of the National Science Foundation, NSF CISE managers, and selected industry and government invitees including a member of the President's Information Technology Advisory Committee, to

* Identify (a) problems that are fundamental obstacles to making progress toward this new challenge; (b) areas where major breakthroughs appear possible; (c) needed collaborations (e.g., inter-disciplinary, academic/industry); and (d) research initiatives and facilities needed to meet these and future challenges.

* Provide demonstrations and interact with each other on the objectives, contributions and challenges of major research activities funded by the IDM and explore fruitful collaboration and synergism.

* Provide an opportunity for NSF program officers, other foundations and funding agencies, and industry representatives to learn more about
the current research efforts and successes of projects funded by IDM, and for such officers to share their program highlights and concerns.

Since expenses for IDM99 were lower than expected, remaining funds were used to extend that workshop with a smaller, more focused event, 'Expanding the Information and Data Management (IDM) Research and Education Community', October 2-4, 2000 at Hotel Roanoke and Conference Center, Roanoke, VA, see http://fox.cs.vt.edu/IDM. The purpose of this workshop was to bring together IDM practitioners, algorithm specialists, and tool developers to briefly summarize the state of the art in IDM and to map out a support infrastructure for the larger IDM research and educational community. Specifically, the workshop prepared recommendations to serve the IDM community through online resources (e.g., IDM portal, digital library, Web site) that aid research, development, and education about IDM-related fields. Related activities have been adopted in many diverse communities with encouraging results. For example, the Collected Algorithms (http://www.acm.org/calgo/), GAMS (http://gams.nist.gov) and Netlib (http://netlib2.cs.utk.edu/) facilities pioneered experimental investigations in the field of mathematical software. Repositories and testbeds at the community level have become accepted forums for disseminating experimental results. Software libraries and support for software testing are well developed in some research communities. Other related work includes the
* Protein Data Bank (http://www.rcsb.org/pdb/),
* GenBank (http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html), and the
* Quantum Chemistry Program Exchange (http://qcpe.chem.indiana.edu/).

The October 2000 workshop explored these issues in various key areas of information and data management. Important issues include modeling the experimental process of defining a population of test problems, schema management, determining problem features most relevant to algorithm analyses, data set modeling, experiment management, and analyzing the applicability of algorithms and tools in different situations. Recommendations were developed regarding mechanisms for building and maintaining infrastructure, including sources and amount

of funding required.

The October 2000 workshop focused on:
1. Algorithms for Web Data
2. Algorithms for Massive Datasets
3. IDM Infrastructure

Findings:

Five major discussion groups gathered throughout the 1999 workshop, each led by two co-chairs. The IDM 99 Workshop Discussion Group Reports present the findings to help formulate future directions of research in information and data management focusing in five areas:

1. Next Generation Information Access
2. Information Presentation and Visualization
3. Languages, Data Models and Application Models
4. Open and Autonomous Data on the Net
5. IR and DM in the WWW, Internet and Wireless Era

For each major area covered we present the mission, issues and recommendations in one page, followed by a further brief on the subject. Documents of the findings can be found for each of the major areas at the following URL:
http://www.cs.ucla.edu/csd/IDM99/FinalReports/

These workshops led to revised information about the IDM program on its web site, helping clarify its mission and refocusing the IDM community in newly identified directions. Various proposals were submitted to the IDM program in accord with the meetings' identification of problems and promising approaches. These meetings provided the participants involved, and those who read the findings of the project, with a clear explanation of the state of the art and future directions of information and data management. The web sites for the two meetings allow the public to acquaint themselves with some of the current endeavors and future directions in information and data management.

Training and Development:

This project has provided the participants involved, and those who read the findings of the project, the state of the art and future directions of information and data management.

Outreach Activities:

We have provided 2 web sites (http://www.cs.ucla.edu/csd/IDM99/ and http://fox.cs.vt.edu/IDM) that allow the public to acquaint themselves with some of the current endeavors and future directions in information and data management.

Of particular import is for the IDM community to develop infrastructure to more effectively support research and education. To advance science in this field, it is necessary that studies be enabled at different locations that look at important problems from different perspectives. Sharing data sets, software, and comparing approaches

with related work is necessary for serious progress. Efforts funded by others than NSF, such as TREC, MUC, and CLEF, have had dramatic and positive impact; NSF should facilitate similar efforts as the field moves more toward componentized solutions. Also, the IDM program should cooperate with NSF DUE, especially in the NSDL activities, so that education in the IDM area is reinforced by having pedagogically effective demonstrations, simulations, and related software and data.

Journal Publications:

Book(s) of other one-time publications(s):

Alfonso F. Cárdenas, Wesley W. Chu, Edward A. Fox, "Proceedings of the NSF Infomation and Data Management Workshop IDM99 : Research Agenda for the 21st Century" , bibl. National Science Foundation (Washington, D. C.), Los Angeles, CA, March 7-9, 1999, 249 pages, (1999). *Book* Published

Edward A. Fox, Layne T. Watson, and Naren Ramakrishnan, "Expanding the Information and Data Management (IDM) Research and Education Community" , bibl. http://fox.cs.vt.edu/IDM/early.pdf, (2000). *Early proceedings* Web publication, handed out at meeting

Other Specific Products:

Internet Dissemination:

http://www.cs.ucla.edu/csd/IDM99/, http://fox.cs.vt.edu/IDM/

http://www.cs.ucla.edu/csd/IDM99/ is the WWW site for IDM"99

http://fox.cs.vt.edu/IDM/ is the WWW site for the Oct. 2000 IDM mtg

Contributions:

Contributions within Discipline:

We have:
* helped set the research agenda for the Information and Data Management area; and
* facilitated the sharing of results from current and recent NSF funded projects in the Information and Data Management area.

Contributions to Other Disciplines:

We have helped:
* identify ways in which Information and Data Management can contribute to work in other areas of science, engineering and technology; and
* shared knowledge about other disciplines (regarding their involvement in IDM funded projects) among those working in Information and Data Management so they may be able to contribute directly or indirectly (through advice, students' efforts) to them.

Contributions to Education and Human Resources:

We have contributed to many areas of human resources development.
* Students and others besides PIs attended the workshops and learned from the event.
* Information and demos shared at the meeting have been further disseminated through classes taught by the PIs and other attending the meeting.
* Many around the nation have accessed and learned from our WWW sites.

Contributions to Resources for Science and Technology:

There has been some involvement of attendees with the Computer Science Teaching Center (www.cstc.org) and early efforts to contribute to the new ACM Journal of Educational Resources in Computing (JERIC).

Contributions Beyond Science and Engineering:

In the Information Age, applications of Information and Data Management are found in all walks of life. Many of the tools, services, techniques, and other contributions discussed or proposed at the workshops have broad impact, in commercial products (e.g., in the WWW) and in activities of other organizations.

Categories for which nothing is reported:
Products: Journal Publications
Products: Other Specific Product

Submit Return