



Expanding the Information and Data Management (IDM) Research and Education Community

October 2 (8:30am) - Oct. 4 (noon), 2000

Hotel Roanoke
Roanoke, VA

**NEW! Final details about arrangements information
- all participants should read!**

- [detailed schedule, with room assignments, guest speakers](#)
- [breakout session discussion questions and topics](#)
- [list of attendees](#)
- [travel instructions](#)

Workshop Purpose

The purpose of this workshop is 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 will prepare 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

Related activities have been adopted in many diverse communities with encouraging results. For example, the [Collected Algorithms](#), [GAMS](#) and [Netlib](#) 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](#), [GenBank](#), and the [Quantum Chemistry Program Exchange](#).

IDM Issues Related to Infrastructure

The workshop will seek to explore 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 will be developed regarding mechanisms for building and maintaining infrastructure, including sources and amount of funding required. Discussion, focused on information and data management, will deal with issues such as: (a) test collections of audio/video; (b) software collections; (c) courseware collections; (d) linguistic/dictionary collections; (e) needs for interoperability (e.g., query translation); (f) metrics; (g) role of collaboratories; (h) technology for shared repositories; and (i) an IDM Web site (a portal, that should be continuously updated). Thus, in case (b), we will consider what types of software (experimental, demo, free, shareware, ...) should be emphasized; and what layers of services should support that software (e.g., a recommender, a match-maker, a collaboration facilitator).

Workshop Attendees and Program

Attendees: 30 or more attendees from the IDM community are expected to attend, including PIs and co-PIs on IDM-supported projects, both those who have developed algorithms/tools/data sets and those interested in using such resources (for research and/or educational purposes). If you are interested in participating, please send a position statement to Prof. Edward A. Fox at fox@vt.edu. In addition, partners will be invited from industry and other related domains. Partners from industry are encouraged to recommend ways that PIs could seek free/heavily discounted software, collaboration, summer jobs for students, etc. (as discussed at the [1998 Industry-Academia workshop](#)). **Format:** The format of the workshop will include plenary talks by key people in the field, and three breakout sessions, concentrating on:

1. algorithms for manipulating, extracting schema from, and querying Web data (XML, DTDs, semi-structured formats).
2. algorithms for data sets of massive dimensionality.

3. methodologies, infrastructure (system level issues) for enabling community-level testing, evaluation and computation facilities.

These areas have been chosen for their currency and immediate relevance to the IDM specialist. For example, area 1 is increasingly gathering attention in Web site management, content personalization, and the design of internet portals. Area 2 has relevance to the large-scale information retrieval, multidimensional data mining, and knowledge discovery communities. And finally, the third working group will help address the role of superstorage systems for designing testbeds, reuse methodologies, automated experiment management, and the role of recommender systems to aid in automated algorithm selection. All together, workshop activities should lead to an operational plan for establishing and maintaining information resources that will support the large IDM research and education community.

Advance Preparation: In advance of the meeting, attendees will study the [IDM program description](#), [PIs workshop reports](#), [IDM awards](#), and other relevant activities in industry and elsewhere. Based on this, attendees will prepare short reports to provide a platform for recommending the infrastructure that would enable the IDM community to be aware about relevant information.

Reporting

The recommendations of the workshop will stimulate community-wide efforts such as (i) to advance the state-of-the-art in sharing of IDM software, and (ii) to serve and extend education about IDM systems and algorithms. In addition to reporting on a WWW site (that will ultimately feed into an IDM Web site), various publication venues will be pursued. For example, with regard to aspects related to IR, there is SIGIR Forum. Also, with regard to tools that are of pedagogical value, a special issue of the new ACM Journal of Educational Resources in Computing (JERIC) will be scheduled.

Workshop Coordinators

[Edward A. Fox](#)

Department of Computer Science
Virginia Tech, VA 24061
fox@vt.edu

[Layne T. Watson](#)

Department of Computer Science
Virginia Tech, VA 24061
ltw@cs.vt.edu

[Naren Ramakrishnan](#)

Department of Computer Science
Virginia Tech, VA 24061

Steering Committee

Robert M. Akscyn, Knowledge Systems

Michael Berry, UTK

Alfonso F. Cardenas, UCLA

Don Kraft, LSU

W. Bruce Croft, U. Mass., Amherst

Edward A. Fox, Virginia Tech

Richard Furuta, Texas A & M

Alon Levy, University of Washington

Alberto Mendelzon, Univ. of Toronto

Gultekin Ozsoyoglu, CWRU

Naren Ramakrishnan, Virginia Tech

Layne T. Watson, Virginia Tech

Ellen Voorhees, NIST

Clement Yu, U. Illinois, Chicago

Monday, October 2, 2000

(All meetings on Monday morning are in the Monroe room. All breaks on all days are in the Lower Conference Center Lounge.)

- *8:30am - 08:50am*
Welcome and Opening Remarks
Dr. Joseph S. Merola, Associate Dean of Research and Outreach
College of Arts and Sciences, Virginia Tech
- *8:50am - 09:30am*
Introductory Remarks
Dr. Maria Zemankova, NSF
- *9:30am - 10:15am*
Plenary Talk: Dr. C. Lee Giles, Penn State University
- *10:15am - 10:45am*
Break
- *10:45am - 11:30am*
Plenary Talk: Dr. Christos Faloutsos, Carnegie Mellon University
- *11:30am - 12:15pm*
Plenary Talk: Dr. W. Bruce Croft, University of Massachusetts, Amherst
- *12:15pm - 1:45pm*
Lunch (Regency Room)
- *1:45pm - 3:15pm*
Breakout Groups
Algorithms for Web Data: Monroe Room
Algorithms for Massive Datasets: Harrison/Tyler Room
IDM Infrastructure: Madison Room
- *3:15pm - 3:45pm*
Break
- *3:45pm - 5:00pm*
Reporting from Breakout Groups (Monroe Room)
- *6:00pm - 8:00pm*
Dinner (Crystal A Room)
Banquet Speaker: Dr. Henry Bauer
Professor Emeritus of Chemistry & Science Studies
and Dean Emeritus of Arts & Sciences, Virginia Tech

Tuesday, October 3, 2000

- *8:30am - 10:00am*
Breakout Groups
Algorithms for Web Data: Monroe Room
Algorithms for Massive Datasets: Harrison/Tyler Room
IDM Infrastructure: Madison Room
- *10:00am - 10:30am*
Break
- *10:30am - 12:00pm*
Reporting from Breakout Groups (Monroe Room)
- *12:00pm - 1:30pm*
Lunch (Crystal A Room)
Luncheon Speaker: Dr. Jennifer Weller
Virginia Bioinformatics Institute
- *1:30pm - 3:00pm*
Breakout Groups
Algorithms for Web Data: Monroe Room
Algorithms for Massive Datasets: Harrison/Tyler Room
IDM Infrastructure: Madison Room
- *3:00pm - 3:30pm*
Break
- *3:30pm - 5:00pm*
Reporting from Breakout Groups (Monroe Room)

Wednesday, October 4, 2000

- *8:30am - 10:00am*
Breakout Groups
(groups may be combined based on attendance)
Algorithms for Web Data: Monroe Room
Algorithms for Massive Datasets: Harrison/Tyler Room
IDM Infrastructure: Madison Room
- *10:00am - 10:30am*
Break
- *10:30am - 12:00pm*
Reporting from Breakout Groups, Closing Discussion (Monroe Room)

Topics and Charges for the Breakout Sessions

- *Monday Afternoon, October 2, 2000*

What nonlocal infrastructure is necessary (hardware, algorithms/software, data collections) to advance science in your speciality? What national resource(s) would advance both your own work and the IDM community in general? Identify these resources in order of priority.

Remember that it is now affordable to work with gigabit networks, teraflop machines, clusters of hundreds or thousands of computers, and terabyte-scale storage servers. There are enormous multimedia, text, and data collections ready to be explored. Many research labs have rewritten monolithic systems using object-oriented methods and modern programming languages, and some make software or toolkits available, so it would not be difficult to start to assemble large portable libraries of modules. What facilities are needed to undertake really large experiments (e.g., the Bowtie analysis of the graph structure of the web at IBM) or make comprehensive analysis runs that would answer interesting open problems?

- *Tuesday Morning, October 3, 2000*

What is the IDM grand challenge?

Historically, disciplines have been galvanized by significant problems. Examples include the Erlanger program, finite group classification, and the four color problem in mathematics; the search for subatomic particles and a unified field theory in physics; and the human genome project in biology. Such ‘grand challenges’ unify a large portion of the discipline, make a case for investment of resources, and provide an identity for the field to the public. Does the IDM community have 2 or 3 such problems that could be touted as ‘grand challenges’ worthy of public support?

What solutions to IDM problems would have a revolutionary impact? Can we develop a unified theory of data, information, and knowledge? Can we transform the WWW into the ‘Semantic Web?’ Can we build a universal visualizer? Can we harness computers to increase the ‘collective IQ’ and ‘augment human intellect?’

- *Tuesday Afternoon, October 3, 2000*

What modes of collaboration can leverage IDM accomplishments?

Today the IDM community is largely scattered, with few groups of substantive size. At a time when integrated solutions are called for to many of humanity's problems, and when 'big science' often is needed to solve big problems, can we avoid being left behind or ignored?

In a variety of research domains, the notion of a virtual 'collaboratory' has facilitated nationwide efforts on focussed projects. In some fields, having a small number of national laboratories or centers has brought together a critical mass of researchers. In Germany, the Dagstuhl 'castle' retreat center is booked 50 weeks a year with groups discussing computing in workshops that keep participants away from their offices to launch new partnerships and joint efforts. In IR, the TREC competition has leveraged support flowing through NIST as well as interest in comparing approaches on constrained problems. In areas where testbeds are appropriate, diverse groups have banded together to work on a particular collection of data and application. Which of these, or other solutions, works best for IDM?

- *Wednesday Morning, October 4, 2000*

What educational resources are needed for a required CS course on IDM?

In CC2001 (the new curriculum for computing being coordinated under the auspices of ACM and IEEE-CS), there will be some required hours related to IDM. There also will be optional hours in core courses as well as separate focussed courses in this area. How can the knowledge of the relatively small IDM community be shared more broadly with educators around the nation so that these courses attract people to our field? How can the interesting work shown in demos at our conferences become visible in laboratory and homework settings for undergraduate students? How can the opportunity of NSF's support of the National Science (and Mathematics, Engineering, and Technology Education) Digital Library (NSDL - see www.smete.org) be exploited so that IDM concepts are more broadly understood by K-12 as well as undergraduate students? How can the expanding interest in data mining, bioinformatics, digital libraries, and other 'hot' fields be harnessed so the latest research becomes available to those interested?

ACM has launched the Journal of Educational Resources in Computing (JERIC) so there is now a reward mechanism for educators willing to develop effective educational resources. What resources would be most useful for the IDM community? How can our existing efforts be leveraged to produce materials with the greatest pedagogical value?

List of Confirmed Attendees

- Robert M. Akscyn, Knowledge Systems Incorporated
 - Michael Berry, University of Tennessee
 - Lois Boggess, Mississippi State University
 - Athman Bouguettaya, Virginia Tech
 - W. Bruce Croft, University of Massachusetts, Amherst
 - Abdur Chowdhury, AOL Technologies
 - Alex Delis, Brooklyn Polytechnic
 - Indherjit Dhillon, University of Texas, Austin
 - Ahmed K. Elmagarmid, Purdue University
 - Christos Faloutsos, Carnegie Mellon University
 - Edward A. Fox, Virginia Tech
 - Richard Furuta, Texas A&M
 - Minos N. Garofalakis, Bell Labs, Murray Hill, NJ
 - Susan Gauch, University of Kansas
 - Arif Ghafoor, Purdue University
 - C. Lee Giles, Penn State University
 - Marti Hearst, SIMS, University of California, Berkeley
 - Anupam Joshi, University of Maryland Baltimore County
 - Paul Kantor, Rutgers University
 - Donald H. Kraft, Louisiana State University
 - Mann-Ho Lee, Virginia Tech
 - Weiyi Meng, SUNY Binghamton
 - Naren Ramakrishnan, Virginia Tech
 - Bill Ogden, New Mexico State University
 - Gultekin Ozsoyoglu, Case Western Reserve University
 - Shalini Urs, Virginia Tech
 - Ellen Voorhees, National Institute of Standards and Technology
 - Layne T. Watson, Virginia Tech
 - Maria Zemankova, National Science Foundation
-

Other Local Attendees from Virginia Tech

- Fernando A. Das-Neves
- Ameya Datey
- Satadip Dutta
- Geoff Filippi
- Robert K. France
- Marcos Goncalves
- Saverio Perugini
- Hussein Suleman

Arrangements for the IDM Oct Workshop

As an invited participant to the IDM workshop, please be informed of the following arrangements, complete the form at the end of the page, and mail it to: naren@cs.vt.edu

Accomodation

A block of rooms has been reserved exclusively for attendees at the workshop venue: Hotel Roanoke and Conference Center

110 Shenandoah Avenue, Roanoke, VA 24016

Phone: 540-985-5900

Fax: 540-853-8290

1-800-222-TREE

In conjunction with the Divison of Continuing Education at Virginia Tech, an integrated complete meeting package (covering food, beverage, lodging, workshop materials) has been developed and these expenses will be handled directly by Virginia Tech. Participants need only inform us about their dates of arrival and departure, room preferences etc. (see below form). Reservations will be made in their names immediately on receipt of this information. The default package includes stay for three nights (Oct 1-3). Checkin time is 4pm and Checkout time is prior to 11am.

Travel

You are requested to make your own flight/driving arrangements and will be reimbursed after the workshop. If you plan to arrive on Saturday, September 30 to avail of lower airfares, please inform us using the form below, and we will extend your reservation. The closest airport is the Roanoke Regional Airport (ROA), which is serviced by four major carriers. Courtesy shuttles run by Hotel Roanoke are available from the Airport. Alternatively, participants might wish to arrive at the Piedmont Triad International Airport, and take an hour and a half's drive up 220 to Roanoke. This would be in case you want to save on airfare and would like a scenic drive. Driving directions to Hotel Roanoke are available at the bottom of the hotel reservations page (notice that you do not need to make reservations at Hotel Roanoke yourself). Keep in mind that Hotel Roanoke charges a daily parking rate of \$5 per car (self-parking).

Sight seeing and Tourism

Roanoke provides various opportunities for recreation and outdoor activities. Depending on participant interest and arrival dates, short tours might be conducted on Sunday, October 1st of nearby places and, in particular, the new Advanced Communications and Information Technology Center at Virginia Tech in Blacksburg, which is 33 miles south of Roanoke.

General Policies

Please retain all originals of receipts. Public transportation rates must not exceed those for tourist class accommodations. Receipts are required for taxis, shuttle vans, and other forms of transportation if the claim

exceeds \$10. Reimbursements for mileage is provided at 23 cents/mile if the round trip is 100 miles or more and at 32.5 cents/mile if the round trip is less than 100 miles. When travel is by personally-owned automobile, the total transportation expenses reimbursed, including meals and lodging, should not exceed the cost of the most economical public air transportation fare available. Travel routing, whether by public transportation, privately owned vehicle, state-owned vehicle, or for-hire conveyance, shall be by the most-direct practicable route.

-----CUT HERE-----

Name of Participant: _____

Address and Contact Info (including Email):

Intended Date and Time of Arrival: _____

Intended Date and Time of Departure: _____

As mentioned earlier, we have negotiated a complete meeting package with Hotel Roanoke. If you would like to stay at a different location, please mention the place, and your plans: _____

Room Preferences: _____

Dietary Restrictions and Special Needs: _____

Mode of Transportation to Roanoke: Air/Driving/Both

Details of Transportation (Flight numbers, Airline, Highway Route etc.): _____

Estimated Travel Cost (include tickets, car rentals, parking, and associated expenses):

-----CUT HERE-----



*The Hotel Roanoke & Conference Center
combines 19th-century
ambiance with 21st-century meeting
facilities.*

**Welcome to The Hotel Roanoke & Conference Center,
click on this button  to return to this page.**

**Please click on the picture above to enter
our lobby and enjoy your tour. [Gracious Accommodations and
Facilities](#)**



[Tomorrow's Facilities for Today's Conferences](#)



[Timeless Diversions](#)



**[Experience the Hospitality of the Past at the Meeting Place of the
Future](#)**



[Planning Guide](#)



[Reservations](#)



The Hotel Roanoke & Conference Center
[A Doubletree Hotel](#)

110 Shenandoah Avenue, Roanoke, VA 24016
Phone: 540-985-5900 Fax: 540-853-8290
1-800-222-TREE

In Partnership with [Virginia Tech](#)



VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

Carvins Cove Water Level

Join the wave to save

5.6'

Below Spillway
Oct. 1, 2000



SIGN THE GUEST-BOOK AND YOU MIGHT WIN A FREE T-SHIRT!



[Online Service Request Form](#)

(Request a Service or Make a Comment)

[Geographic Information System](#)

(Citywide Mapping and Real Estate Data)



ABOUT THE CITY OF ROANOKE



ROANOKE CITY COUNCIL



FACTS AND FIGURES



HOW TO REACH YOUR GOVT



NEWS AND ANNOUNCEMENTS



EMPLOYMENT OPPORTUNITIES



LOCATION INFO & MAPS



CITIZENS' REQUEST FOR SERVICE



DEPARTMENTS & OFFICES



OTHER LINKS



ALL-AMERICA CITY AWARD



CITY CODE & PUBLICATIONS



Index



Search



Comments



Statistics

[Text Version](#) | [Display Notes](#)



Entire site © copyright 1996-2000 by the city of Roanoke, Virginia
Please send your comments to cityweb@ci.roanoke.va.us.

Rev. 9-18-00

==



Collected Algorithms

[Background](#) . . . [Index to Algorithms](#) . . . [Contact](#)

The *Collected Algorithms* (CALGO) is part of the family of publications produced by the [Association for Computing Machinery \(ACM\)](#).

Background

Software associated with papers published in the [Transactions on Mathematical Software](#), as well as other ACM journals are incorporated in CALGO. This software is refereed for originality, accuracy, robustness, completeness, portability, and lasting value. (See the [ACM Algorithms Policy](#) for details.)

Readers may [subscribe](#) to CALGO. Subscribers receive quarterly notification of the appearance of new algorithms, as well as copies of research papers describing them in loose-leaf binder form. Back issues may also be ordered.

Use of ACM Algorithms is subject to the [ACM Software Copyright and License Agreement](#)

[Index to Algorithms](#)

All algorithms numbered 493 and above, as well as a few earlier ones, may be downloaded from this server. Many of these files are quite large. To reduce download time the files have been compressed using gzip. If you do not have gunzip, or if your Web browser is not configured to automatically uncompress such files, please see our [notes on file compression](#).

Contact

For further information about CALGO contact its Editor-in-Chief:

[Tim Hopkins](#)

Computing Laboratory

[The University of Kent](#)

Cantebury

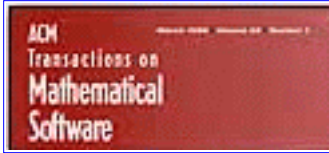
Kent CT2 7NF

United Kingdom

+44-122-776-4000 ext. 3793 (Voice)

+44-122-776-2811 (FAX)

trh@ukc.ac.uk



Last change in this page : *February 18, 1996*. [[Comments](#)].



NIST

Guide to Available Mathematical Software

A cross-index and virtual repository of mathematical and statistical software components of use in computational science and engineering.

[More Math and Statistics at NIST](#)

Do you find this service useful? [Help us justify it.](#)

Mathematical Software Cross Index

Search for software according to

- what [problem](#) it solves.
- [package name](#).
- [module name](#).
- [text in module abstracts](#).

Go straight to the [problem decision tree](#).

Try [HotGAMS](#), a Java interface to GAMS with additional search features.

If you cannot find what you need in the cross index, you might try some of these [other sources of information about mathematical software](#).

Background Information

- [Project summary](#).
- [Glossary of terms](#).
- [Repositories Indexed](#).
- [References, credits and disclaimers](#).
- [News](#).
- [Other NIST projects related to applied mathematics and statistics](#).

Address comments and questions to gams@cam.nist.gov.

The Guide to Available Mathematical Software is a service of the [Mathematical and Computational Science Division](#) within the [Information Technology Laboratory](#) of the [National Institute of Standards and Technology](#). See our [disclaimer](#) and the [NIST Privacy Policy](#).

Last change in this page : 20 June 2000.



Netlib Repository at [UTK](#) and [ORNL](#)

Netlib is a collection of mathematical software, papers, and databases.

There have been [94,239,807](#) requests to this repository as of Sun Oct 1 08:44:14 EDT 2000 .

Software, papers, etc.

- [Browse](#) the Netlib repository
- [Search](#) the Netlib repository

Services provided at Netlib

- [Conferences Database](#)
- [Java Version of Linpack Benchmark](#)
- [Numerical Analysis Net \(NA-Net\)](#) and [NA Digest](#)
- [Performance Database Server](#)
- [Top500 Supercomputer Sites](#)

Related efforts

- [BenchWeb](#)
- [HPC-Netlib](#), high performance branch of Netlib
- [Matrix Market](#)
- [National High-Performance Software Exchange \(NHSE\)](#)
- [Parallel Tools Library \(PTLIB\)](#)
- [Repository In a Box \(RIB\)](#)
- [StatCodes](#) at Penn State, statistical source codes and packages of use to physical scientists

Information about Netlib

- [Netlib Discussion Forum](#)
- [Frequently Asked Questions about Netlib \(FAQ\)](#)

- [Netlib Editors](#)
- [Netlib Mirror Sites](#)
- [Netlib Server Statistics](#)
- [Yesterday's 10 most accessed files at Netlib](#)

[How to contact Netlib](#)

Welcome to the PDB, the single international repository for the processing and distribution of 3-D macromolecular structure data primarily determined experimentally by [X-ray crystallography](#) and [NMR](#).

[DEPOSIT](#) Contribute structure data

[STATUS](#) Find entries awaiting release

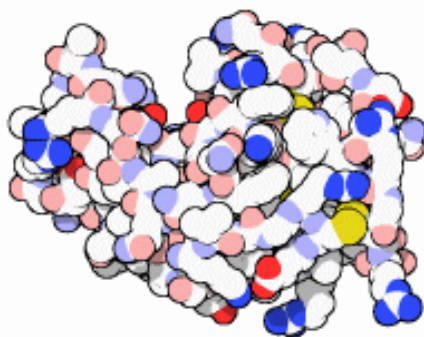
[DOWNLOAD](#) Retrieve structure files (FTP)

[LINKS](#) Browse related information

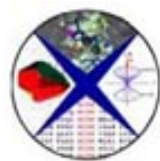
[PREVIEW](#) Beta-test new features

About the PDB

[General Information](#)
[WWW User Guides](#)
[Get Educated](#)
[File Formats & Dictionaries](#)
[News and Discussion](#)
[Press Releases](#)
[Planning](#)



Molecule of the Month: [Lysozyme](#)



The Protein Data Bank is operated by the [Research Collaboratory for Structural Bioinformatics \(RCSB\)](#) under a contract to the [U.S. National Science Foundation](#) and is supported by funds from the National Science Foundation, the [Department of Energy](#), and two units of the National Institutes Of Health: the [National Institute Of General Medical Sciences](#) and the [National Library Of Medicine](#). All inquiries can be addressed to info@rcsb.org.

In citing the PDB please refer to:

H.M.Berman, J.Westbrook, Z.Feng, G.Gilliland, T.N.Bhat, H.Weissig, I.N.Shindyalov, P.E.Bourne
[The Protein Data Bank](#). Nucleic Acids Research, 28 pp. 235-242 (2000)



Current Holdings

[13270 Structures](#)

[Last Update: 26-Sep-2000](#)

[PDB Statistics](#)

Search

Enter a PDB ID:

[SearchLite](#): simple keyword search

[SearchFields](#): advanced search

News

[Complete News](#)
[PDB Newsletter](#)
[Subscribe](#)
[Browse Mailing List](#)

26-Sept-2000

[PDB Focus: info@rcsb.org](#)

The PDB Help Service at info@rcsb.org is provided by RCSB's scientists and staff. [\[MORE...\]](#)

19-Sept-2000

[PDB Focus: The ADIT Validation Server](#)

Depositors are encouraged to use the ADIT Validation Server (<http://pdb.rutgers.edu/validate/>) prior to the deposition of a structure to the PDB. [\[MORE...\]](#)

12-Sep-2000

[New Query Features Available on the PDB Web Site](#)

The PDB is pleased to announce the availability of new features on the PDB Web site (<http://www.rcsb.org/pdb>) and

© [RCSB](#)

its worldwide mirrors
(<http://www.rcsb.org/pdb/mirrors.html>).
[\[MORE...\]](#)

5-Sep-2000

**[PDB's NMR Task Force at ICMRBS
Conference](#)**

The PDB's NMR Task Force met during the XIX International Conference on Magnetic Resonance in Biological Structures (ICMRBS), August 20-25 in Florence, Italy...
[\[MORE...\]](#)

▶ What is GenBank?

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research* 2000 Jan 1;28(1):15-8). There are approximately 9,546,000,000 bases in 8,214,000 sequence records as of August 2000 (see [GenBank growth statistics](#)). As an example, you may view the [record](#) for a *Saccharomyces cerevisiae* gene. The complete [release notes](#) for the current version of GenBank are available. A new release is made every two months. GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

▶ Submissions to GenBank

Many journals require [submission of sequence information](#) to a database prior to publication so that an accession number may appear in the paper. NCBI has a WWW form, called [BankIt](#), for convenient and quick submission of sequence data. [Sequin](#), NCBI's stand-alone submission software for MAC, PC, and UNIX platforms, is also available by FTP. When using Sequin, the output files for direct submission should be sent to GenBank by electronic mail.

There are specialized, streamlined procedures for batch submissions of sequences, such as [EST](#), [STS](#), and [HTG](#) sequences.

▶ Updating or Revising a Sequence

Revisions or updates to GenBank entries can be made at any time and can be accepted as [BankIt](#) or [Sequin](#) files or as the text of an e-mail message. Be sure to give the accession number of the sequence to be updated in the subject line. Send it to:

`update@ncbi.nlm.nih.gov`

▶ Access to GenBank

GenBank is available for [searching](#) via several methods.

▶ New Developments

NCBI is continuously developing new tools and enhancing existing ones to improve both submission and access to GenBank. The easiest way to keep posted of these and other developments is to check the "What's New" section of the NCBI Web page and to read the [NCBI News](#), which is also available by free subscription.

Revised September 5, 2000

[About QCPE](#)[QCPE Catalog](#)[Order Programs](#)[Download Programs](#)[Deposit Programs](#)[QCPE Bulletin](#)[QCPE Documents](#)[QCPE Staff](#)[Other Links](#)[Search QCPE Site](#)[QCPE Site Index](#)[Contact QCPE](#)

QCPE: Quantum Chemistry Program Exchange

The QCPE web site has new and updated features:

- The web-accessible [QCPE Catalog](#) now has an entry for all QCPE programs.
- Customers can search the [QCPE Catalog](#) using Case-Sensitive or case-insensitive information.
- Customers can now [Download Programs](#) directly from the QCPE Web Site.
- An archive of the [QCPE Bulletin](#) is now available. More issues of the QCPE Bulletin will be made available in the near future.
- [Other Links](#) to other computational chemistry resources is now available.
- Customers can [Search the entire QCPE Web Site](#).
- A [QCPE Web Site Index](#) is now available.

The numbering for QCPE programs for the PC and Mac platforms, and for the Linux operating system, has been augmented:

- Each QCPE program for the PC platform, historically annotated as a "QCMP" program, now also has a "QCPE" program number. Customers can refer to either the "QCMP" or "QCPE" number, and the QCPE Catalog lists both numbers.
- Each QCPE program for the Mac platform, historically annotated as a "QMAC" program, now also has a "QCPE" program number. Customers can refer to either the "QMAC" or "QCPE" number, and the QCPE Catalog lists both numbers.
- Each QCPE program for the LINUX operating system, historically annotated as a "LINUX" program, now also has a "QCPE" program number. Customers can refer to either the "LINUX" or "QCPE" number, and the QCPE Catalog lists both numbers.

Please [contact QCPE](#) if you have questions or suggestions regarding this new web site.

Recent News

- [Kristine Schleede joins QCPE Staff](#)
- [New Program QCMP189](#)
- [Dr. Margaret Edwards Leaves QCPE](#)
- [QCPE "Returns" to Chemistry](#)
- [New QCPE Web Site now available](#)
- [Richard Counts Retires from IU](#)

QCPE

Department of Chemistry
Indiana University
800 East Kirkwood Ave.
Bloomington, IN 47405-7102

E-MAIL: qcpe@indiana.edu

Phone: (812)-855-4784

FAX: 812-855-8300

Last updated: December 6, 1999

URL: <http://qcpe.chem.indiana.edu>

Copyright 1999, The Trustees of Indiana University

NSF Workshop on Industrial/Academic Cooperation in Database Systems

Los Gatos, California
October 29-30, 1998

Organizing Committee

Betty Salzberg

Northeastern University

salzberg@ccs.neu.edu

Mike Carey

IBM Almaden

carey@almaden.ibm.com

Joseph Hellerstein

University of California, Berkeley

jmh@postgres.berkeley.edu

Ashish Gupta

Junglee Corporation

ashish@junglee.com

Many academic database researchers in computer science want their research to be more relevant to the computer industry. But how do they select the right topics? It is difficult to formulate problems that are commercially important and are likely to yield the algorithmic or mathematical content required for publication in top journals and conferences. Industrial researchers could help by describing projects and problems with significant intellectual, possibly mathematical or scientific, content. But they have little incentive to do so. Their main goal is to solve problems, not brainstorm research agendas.

The objective of the workshop was to address this problem. Several top database architects from industry were invited to to share with academics some of the interesting intellectual problems which have arisen while implementing large-scale DBMSs. A representative of the NSF discussed the relevant NSF programs for industrial-academic cooperation and a patent attorney discussed legal issues of intellectual property. In addition, there were discussions about how to find problems of mutual interest for possible collaboration, how DBMS companies can enable academics to make their research reflect advances in current practice, and how to put in place mechanisms for longer term exchanges of information.

Many of the industrial presentations discussed the management of complexity. For example, there are a number of applications that need serious help with managing metadata, including scientific data management, CAD/CAM, warehouse design, websites, workflow, documents, heterogeneous database integration, configuration management and application development tools. Another unmanageably complex area is in understanding the meaning of SQL queries generated by high-level tools. Some of these queries can involved hundreds of joins, outer joins and views. Academic research usually only analyzes queries with one or two joins.

One reason that industry people can't use academic results is that introducing new features often makes

the performance degrade unacceptably. So the new features have to be dropped. Academics need to explain how to reliably implement their results so as not to cause performance degradation or additional errors.

Another theme that occurred was that what industry really wanted from academics was well-trained students. These students should not only know the usual algorithms, but should also know about error handling and testing. They should, for example, understand that gathering information about performance should be part of any system's original design. Access to well-trained students was the chief benefit many industries cited when they participated in NSF-sponsored industrial/academic programs.

The full report from the workshop is available from the website below, with lists of people from both camps willing to give talks and make visits to their counterparts, full powerpoint slide presentations from the invited industrial speakers and references to NSF programs such as GOALI that would enable closer cooperation between industry and academia. It is hoped that this website can become part of the solution to the problems of closer cooperation between industry and academics.

Further Information:

The full report and supporting material can be found at the NSF Industrial/Academic Workshop website: <http://www.ccs.neu.edu/groups/IEEE/ind-acad/>

The workshop was funded by the National Science Foundation, the Information and Data Management Program, under the grant IIS-9815933 (PI: Betty Salzberg). All opinions, findings, conclusions and recommendations in any material resulting from this workshop are those of the workshop participants, and do not necessarily reflect the views of the National Science Foundation.

**CISE DIVISIONS**Advanced
Computational
Infrastructure
and ResearchAdvanced
Networking
Infrastructure
and ResearchComputer-
Communications
ResearchExperimental
and Integrative
ActivitiesInformation
and Intelligent
Systems

Information and Data Management Program Description

NATIONAL SCIENCE FOUNDATION
Information and Intelligent Systems Division
4201 Wilson Blvd., #1115
Arlington, VA 22230

Program Director: [Dr. Maria Zemankova](#)

Phone: 703-306-1926

Fax: 703-306-0599

Internet: mzemanko@nsf.gov

<http://www.cise.nsf.gov/iis/idm/>

SCIENTIFIC SCOPE

The Information and Data Management (IDM) Program in the Division of Information and Intelligent Systems (IIS) supports research fundamental to the design, implementation, development, management, and use of databases, information retrieval, and knowledge-based systems.

The aim is to build a "new generation" of distributed, interoperable, multi-media, intelligent, dynamic, evolvable information systems capable of sophisticated and efficient information processing. Research pertinent to this goal involves investigating novel concepts, or combining and extending conventional concepts and systems. It hinges on basic research in databases, knowledge-based systems and information retrieval systems, and includes a wide scope of related areas, ranging from artificial intelligence methodologies to techniques that exploit hardware technologies.

Projects supported by the IDM Program can be divided into interrelated areas: (1) data, information and knowledge modeling; (2) information access and knowledge discovery; (3) system architecture and implementation; and (4) system development and administration.

Data, Information, and Knowledge Modeling

This area provides foundations for new, more expressive models of data, information, and knowledge. Topics include object-oriented systems; temporal, spatial, pictorial, multi-media databases; special-purpose scientific databases, including metadata representation; full-text systems; constraint-based systems; active systems; process and enterprise modeling; warehouses, workflows; and knowledge-based systems. Issues considered include type systems; declarative extensions; database and knowledge-base evolution, integrity and validation;

inheritance and exceptions; and management of uncertainty arising from imprecision in data or knowledge. Basic research in formal models of knowledge and information is supported in the [Knowledge and Cognitive Systems Program](#), IIS. Related research in programming languages is considered in the [Division of Computing-Communications Research \(C-CR\)](#).

Information Access and Knowledge Discovery

The aim of research in this area is to design more intelligent and efficient information access methods and novel approaches to knowledge discovery or datamining. Research topics include query language design; data analysis tools, including automated and user-aided knowledge discovery/datamining from databases and information visualization; enhanced query processing (e.g., statistical sampling, approximate queries, cooperative answering, employment of feedback, thesaurus or semantic nets in information retrieval); information organization (e.g., categorization, summarization); browsing and navigation; user-centered information retrieval and dissemination; relevant information source discovery; knowledge-based query optimization; and resolution of incompleteness and inconsistency in heterogeneous systems. Related research, placing emphasis on human interfaces, e.g. user modeling, visualization or natural-language or visual interfaces, is jointly considered with the [Human-Computer Interaction Program](#), IIS. Research into principles for building multi-user collaborative information systems is covered in the [Computation and Social Systems Program](#), IIS.

System Architecture and Implementation

The objective of this area is building high performance systems by addressing issues in algorithms, reliable storage, access, and manipulation of actual data. Research topics include indexing and hashing algorithms; persistent object storage; main-memory systems; associative memory; cache-memory; utilization of optical storage; tertiary storage management (terabyte mass storage); distributed and heterogeneous systems, including mobile environments; extensible systems; real-time or constrained-time/space query processing; parallel processing; concurrency control; long duration transaction processing; fault-tolerant systems; backup and recovery. Related research is also supported in [Computer Systems Architecture Program](#) and Software Systems Program in the [Division of Computing-Communications Research \(C-CR\)](#), and in the [Division of Advanced Networking Infrastructure and Research \(ANIR\)](#).

System Development and Administration

This category involves development of methodologies for specification, implementation, verification, maintenance and management of information systems. Research includes work in specification languages, including areas of CAD/CAM, workflows, and scientific databases; extension and development of design tools and environments; reconfigurable systems; fast prototyping; modularization of large heterogeneous systems; evolvable information systems; information systems metrics (e.g., satisfaction of information requirements, performance, technology insertion costs); security issues; audit trails; and providing integrated interfaces for related manipulations or tasks. Related issues are supported in the [Computation and Social Systems Program](#), IIS and in the [Division of Computing-Communications Research](#)

[\(C-CR\)](#).

PROGRAM HIGHLIGHTS

Many IDM researchers are involved in the NSF-wide initiative "[Information Technology Research](#)" (ITR) which provides additional funding for the IDM community and an opportunity for developing large-scale research efforts and interdisciplinary projects. IDM Program plays a role in the interagency and international initiatives on "[Digital Libraries](#)" and continues its support of research on scientific and engineering databases. In particular, several IDM researchers are collaborating with biocomplexity researchers and are involved in the NSF-wide "[Biocomplexity](#)" initiative. These highly interdisciplinary initiatives foster collaboration in information, networking and high performance technologies in close cooperation with researchers in scientific or engineering domains. Innovative regular proposals in these areas are welcome at any time. Proposals for development of courses for bridging the gap between domain scientists or engineers and information systems scientists in an effort to build effective information systems are of interest.

Continued emphasis is being given to research addressing issues in representation, storage, distribution, interoperability, evolution, access and delivery of different information forms -- text, image, sound, videos, large databases, software tools, educational aids, scientific or engineering instruments, etc., in order to satisfy a wide variety of users with specific and changing needs is encouraged.

Innovative, revolutionary, risky research ideas with a potential great impact are especially encouraged. IDM Program can support such research efforts through the Small Grants for Exploratory Research (SGER) program. You are encouraged to contact the program director to discuss your ideas.

Proposals for planning workshops intended to stimulate innovative/interdisciplinary research are welcome. On-line reports from recently held planning workshops can be found at <http://www.cise.nsf.gov/iis/idm/>.

The IDM Program will continue supporting new research ideas and will play an active role in fostering collaboration with other NSF programs, other agencies and industry in order to achieve the goal of building widely accessible intelligent information systems to augment human intelligence in a broad spectrum of endeavors.

PROPOSAL SUBMISSION

There are two proposal submission target dates annually, September 15 and February 15. For details, please see <http://www.cise.nsf.gov/iis/idm/>.

| [CISE Home](#) | [NSF Home](#) | [Site Map](#) | [Search](#) | [How to navigate](#) |

This page was last updated on 02/16/2000. Technical comments should be sent to [CISE Webmaster](#). Also see [Statement of Responsibility](#).



CISE DIVISIONS

Advanced
Computational
Infrastructure
and Research

Advanced
Networking
Infrastructure
and Research

Computer-
Communications
Research

Experimental
and Integrative
Activities

Information
and Intelligent
Systems

Information and Data Management PI Workshops

[Search all CISE Related Workshops and Conferences](#)

Workshop Titles ▲

[2000 Information and Data Management Program PI Workshop: Research Agenda into the Future](#)

Chicago, Illinois, 03/05/2000 -- 03/07/2000

[1999 Information and Data Management Program PI Workshop: Research Agenda for the 21st Century](#)

Los Angeles, California, 03/07/99 -- 03/09/99

[1998 Information and Data Management Program PI Workshop](#)

Washington, DC, 03/29/98 -- 03/31/98

| [CISE Home](#) | [NSF Home](#) | [CISE Home](#) | [CISE Events](#) | [Site Map](#) | [Search](#) | [Help](#) |

This page was last updated on 10/01/2000 and comments should be sent to [CISE Webmaster](#). Also see [Statement of Responsibility](#).



Award Search Results

Found **438** awards looking for **awdNSFProgram <CONTAINS> '6855'**
Showing awards **1 - 10** of the top .

No Previous Page | [[Next Page](#)]

1. [CAREER: The Biological Digital Library: Information Retrieval Meets Bioinformatics](#)
#9986085 -- Score: [100%](#) -- [[Text version](#)]
 2. [CAREER: Information Visualization Research and Education](#)
#9985136 -- Score: [100%](#) -- [[Text version](#)]
 3. [CAREER: Generating Provably Correct Query Optimizers](#)
#9984960 -- Score: [100%](#) -- [[Text version](#)]
 4. [CAREER: Efficient I/O for Modern Database Applications](#)
#9985019 -- Score: [100%](#) -- [[Text version](#)]
 5. [CAREER: Data Mining Techniques for Geospatial Applications](#)
#9984729 -- Score: [100%](#) -- [[Text version](#)]
 6. [CAREER: Flexible Construction and Management of Web Sites](#)
#9985114 -- Score: [100%](#) -- [[Text version](#)]
 7. [CAREER: Consistent and Robust Retrieval, Transmission and Presentation of Multimedia Data](#)
#9733730 -- Score: [100%](#) -- [[Text version](#)]
 8. [Collaborative Research: Data Mining: Theory and Algorithms](#)
#9907483 -- Score: [100%](#) -- [[Text version](#)]
 9. [Collaborative Research: Data Mining: Theory and Algorithms](#)
#0002356 -- Score: [100%](#) -- [[Text version](#)]
 10. [Transactional Threads for Reliable Persistent Application Systems](#)
#9988637 -- Score: [100%](#) -- [[Text version](#)]
-

No Previous Page | [[Next Page](#)]

Also available:

- [A text file containing the concatenated abstracts of the top awards](#)
- [A list of the top awards in comma-separated value \(CSV\) format](#) , suitable for loading into a spreadsheet or database application.

▶ [nsf.gov](#) ▶ [Search](#) ▶ [Help](#) ▶ [Grants](#) ▶ [Custom News](#) ▶ [Documents](#) [E M A I L](#)
▶ [Webmaster](#) ▶ [FastLane](#) ▶ [Info](#)

The National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230, USA
Tel: 703-292-5111
FIRS: 800-877-8339 ~ TDD: 703-292-5090