CHAPTER 5

Directory of Interested Parties

[Editor's Note: If you notice errors or omissions here, please let us know so we can keep this up-to-date.]

People

Akeroyd, Richard G., Jr.
Connecticut State Library
State Librarian
231 Capitol Avenue
Hartford, Connecticut 06106-1537
Phone: 203-566-4301
Fax: 203-566-8940
rakeroyd@eagle.wesleyan.edu

Arms, William Y.

Carnegie Mellon University
Vice President for Academic Services
Warner Hall, 304
5000 Forbes Avenue
Pittsburgh, Pennsylvania 15213
Phone: 412-268-2122
Fax: 412-268-7886
wya@andrew.cmu.edu

Anderson, Greg
MIT Libraries
Massachusetts Institute of Technology
Associate Director for Systems and Planning
Room 14S-216
Cambridge, Massachusetts 02139
Phone: 617-253-5654
Fax: 617-253-8894
ganderso@athena.mit.edu

Andre, Pamela Q. J.
National Agricultural Library
Associate Director for Automation
Information Systems Division
10301 Baltimore Boulevard
Beltsville, Maryland 20705
Phone: 301-504-6813
Fax: 301-504-7473
pandre@asrr.arsusda.gov

Atkins, Dan
University of Michigan
Dean, Library and Information Science
Chair, Working Group on Inpact of National Digital Library (NDL) on Education

Atkinson, Ross W.
Cornell University
Assistant University Librarian for Collection Development
201 Olin Library
Ithaca, New York 14853-3689
Phone: 607-255-3393

Fax: 607-255-9346
rwa@cornellc.bitnet

Badger, Robert C.
Springer-Verlag New York
Manager, Electronic Media
175 Fifth Avenue
New York, New York 10010
Phone: 212-460-1622
Fax: 212-533-5781

Becker, Herbert S.
Library of Congress
Director, Information Technology Services
Madison Building
Information Technology Services LM G51
Washington, D.C. 20540-2300
Phone: 202-707-6207
Fax: 202-707-0955

Besen, Stanley
Charles River Associates
Phone: 202-393-5840
Fax: 202-393-3432

Billick, David J.
UMI
Vice President of Document Acquisition and Imaging
300 North Zeeb Road
Ann Arbor, Michigan 48106
Phone: 313-761-4700
Fax: 313-761-1204

Billington, James
Library of Congress
Librarian of Congress
Washington, DC 20540

Borgman, Christine
Graduate School of Library and Information Science
Professor

300 Circle Drive North
University of California at Los Angeles
Los Angeles, California 90024
Phone: 310-825-6164
Fax: 310-206-4460
IIN4CLB@MVS.OAC.UCLA.EDU

Braid, James Andrew
British Library
Head of Development, Document Supply Care
Boston Spa
Wetherby, West Yorkshire
ENGLAND LS23 7BQ
Phone: +937 546030
Fax: +937 546236
andrew.braid@london.bl.uk

Brett II, George H.
Clearinghouse for Networked Information Discovery and Retrieval (CNIDR)
Director
Center for Communications - MCNC
P.O. Box 12889
Research Triangle Park, North 27709-2889
Phone: 919-248-1886
Fax: 919-248-1405
GEORGE.BRETT@CNIDR.ORG

Burke, Jane
NOTIS Systems, Inc.
President
1007 Church Street, 2nd Floor
Evanston, Illinois 60201-3622
Phone: 708-866-0170
Fax: 708-866-0178

Butler, Meredith
State Univeristy of New York at Albany
Dean & Director of Libraries
University Libraries 123

Albany, New York 12222
Voice: 518-442-3568
Fax: 518-442-3567
mB801@uacsc1.albany.edu

Caviness, Jane S.
National Science Foundation
Deputy Divison Director
Division for Networking and Communications Research
and Infrastructure
4201 Wilson Boulevard, Room 1175
Arlington, VA 22230
Phone: 202-306-1950x1117
jcavines@nsf.gov

Cerf, Vinton G.
Corporation of National Research Initiatives (CNRI)
1895 Preston White Drive, Suite 100
Reston, VA 22091
Phone: 703-620-8990
Fax: 703-620-0913
vcerf@nri.reston.va.us

Chen, Su-Shing
National Science Foundation
Intelligent Search Algorithms and Natural Language Processing
4201 Wilson Boulevard, Room 1115
Arlington, VA 22230
Phone: 703-306-1925x1178
schen@nsf.gov

Christian, Eliot
U.S. Geological Survey
Information Systems Divison
802 National Center
Reston, Virginia 22902
Phone: 703-648-7245
Fax: 703-648-7069
echristi@usgs.gov

Chien, Y. T.
National Science Foundation
IRIS (Information, Robotics, and Intelligent Systems) Division
Director
4201 Wilson Boulevard, Room 1115
Arlington, VA 22230
Phone: 703-306-1927x1173
ytchien@nsf.gov

Chesler, Steve
Apple Library of Tomorrow
Manager
10381 Bandley Dr. M58C
Cupertino, CA 95014
Phone: 408-974-3258
sac@apple.com

Crocca, William T.
Xerox/PARC
CLASS Project since 1989 with Cornell Libraries
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Conley, Laura M. H.
John Wiley & Sons, Inc.
Director of Research and Planning
605 Third Avenue
New York, New York 10158
Phone: 212-850-6219
Fax: 212-850-6154
lconley@jwiley.com

Crystal, Tom
DARPA

Deogun, Jitender S.
University of Nebraska - Lincoln
Department of Computer Science
Lincoln, NE 68588-0115

Phone: 402-472-5033
402-472-3200
Fax: 402-472-7767
deogun@csr.unl.edu

Dillon, Martin
OCLC, Inc.
Director, Office of Research
6565 Frantz Road
Dublin, Ohio 43017-0702
Phone: 614-764-4365
Fax: 614-764-2344
mjd@oclc.org

Dyson, Esther
EDventure Holding, Inc
Writer, Publisher EDITs Project

Eaton, John
Virginia Tech
Associate Dean - Graduate School
Sandy Hall
Blacksburg, Virginia 24061-0325
Phone: 703-231-5645
Fax: 703-231-3714
eaton@vtvm1.cc.vt.edu

Eisenburg, Michael
ERIC Clearinghouse on Information Resources
Director
030 Huntington Hall
Syracuse University
Syracuse, New York 13244
Phone: 315-443-3640
Fax: 315-443-5732
mike@suvm

Ellis, William
Library of Congress
well@seql.loc.gov

Engelbart, Douglas C.
Bootstrap Institute
Director
6505 Kaiser Drive
Fremont, California 94555
Phone: 510-713-3552
Fax: 510-793-2362
engelbart@bootstrap.sri.com

Ertel, Monica
Apple Library of Tomorrow
Library Manager

Faloutsos, Christos
University of Maryland
Department of Computer Science
A.V. Williams Building
College Park, MD 20742-3255

Fields, Martha
State University System of Florida
Director, Information Resource Management
325 West Gaines Street
Suite 1420
Tallahassee, Florida 32399-1950
Phone: 904-488-6030
Fax: 904-487-4568
maf@nwbr400.mail.ufl.edu

Finnigan, Georgia
The Information Store, Inc.

Fox, Edward A.
Virginia Tech
Department of Computer Science
638 McBryde Hall
Blacksburg, Virginia 24061-0106
Phone: 703-231-5113
Fax: 703-231-6075

fox@fox.cs.vt.edu

Garcia-Molina, Hector
Stanford University
Department of Computer Science
Stanford, CA 94305-2140
Phone: 415-723-0685
Fax: 415-725-7411
hector@cs.stanford.edu

Garrett, John
Corporation of National Research Initiatives (CNRI)
1895 Preston White Drive, Suite 100
Phone: 703-620-8990
Fax: 703-620-0913
jgarrett@nri.reston.va.us

Gass, Linda
Adobe Systems, Inc
1585 Charleston Road
P.O. Box 7900
Mountain View, CA 94039-7900
Phone: 415-961-4400
Fax: 415-961-3769

Graham, Margaret
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Graham, Peter S.
Rutgers University
Associate University Librarian for Technology & Networked Information Services
169 College Avenue
New Brunswick, New Jersey 08903
Phone: 908-932-7505
Fax: 908-932-7637
graham@zodiac.rutgers.edu

Grisham, Frank P.

Southeastern Library Network, Inc. (SOLINET)
Executive Director
1438 West Peachtree Street
Suite 200
Atlanta, Georgia, 30309-2955
Phone: 800-999-8558
Fax: 404-892-7879
sgrisham@mail.solinet.net

Grycz, Czeslaw Jan
Univeristy of California, Office of the President
Chair, Scholarship and Technology Study Project (CNI)
Kaiser Center, Eighth Floor
Oakland, California 94612-3550
Phone: 510-987-0561
Fax: 510-527-4512
cjg@dla.ucop.edu

Hankins, John L.
CICNet, Inc.
Assistant Director
2901 Hubbard, Pod G
Ann Arbor, Michigan 48105
Phone: 313-998-6102
Fax: 313-998-6105
hankins@cic.net

Hall, Tony
Dialog Information Services
Director of Planning

Hartzband, David
Digital Equipment Corporation (DEC)
hartzband@cals.enet.dec.com

Henry, Charles
Vassar College Libraries
Poughkeepsie, New York 12601
Phone: 914-437-5782
Fax: 914-437-5864
chhenry@vassar.edu

Heterick, Robert Jr.
EDUCOM
President
1112 16th Street, N.W., Suite #600
Washington, D.C.
Phone: 202-872-4200
Fax: 202-872-4318
heterick@educom.edu

Hoffman, Ellen S.
Merit Network, Inc.
Manager, Network Information Services
2901 Hubbard, Pod G
Ann Arbor, Michigan 48109-2016
Phone: 313-936-3000
Fax: 313-747-3185
ellen@merit.edu

Hopcroft, John
Cornell University
Department of Computer Science
4130 Upson Hall
Ithaca, NY 14853-7501

Hunter, Beverly
National Science Foundation
Former Directorate for Education and Human Resources
Former Program Director, Application of Advanced Technologies,
Division of Materials Development, Research and
Information Science

Hunter, Karen
Elsevier Science Publishers BV
Vice President and Assistant to the Chairman
655 Avenue of the Americas
New York, New York, 10010
Phone: 212-633-3787
Fax: 212-633-3764

khunter@pucc.princeton.edu

Hunter, Larry
National Library of Medicine
Building 38A, MS-54
Bethesda, MD 20894
Phone: 301-496-9300
Fax: 301-496-0673
hunter@nlm.nih.gov

Jagadish, Hosagrahar
AT&T Bell Labs
Murray Hill, NJ
jag@research.att.com

Jajko, Pam
Syntex USA Corporate Library

Kahn, Robert
Corporation of National Research Initiatives (CNRI)
1895 Preston White Drive, Suite 100
Reston, VA 22091
Phone: 703-620-8990
Fax: 703-620-0913

Kantor, Paul
Rutgers University
(SCILS) School of Communication, Information and Library Studies
4 Huntington St.
New Brunswick, NJ 08903
kantor@zodiac.rutgers.edu

Lasher, Rebecca
Stanford University
TR Effort with Garcia-Molina
Head Librarian
Math/Computer Science Library
Stanford, CA 94305-2125
Phone: 415-723-0864
RLasher@forsythe.stanford.edu

cn.mcs@forsythe.stanford.edu

Lannom, Larry
Meridian Corp. (for DARPA)
Phone: 703-552-6410
lannom@darpa.mil

Lavendel, Giuliana
Xerox/PARC
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Leggett, John
Texas A&M University
Department of Computer Science
College Station, TX 77843-3112
Phone: 409-845-0298
Fax: 409-847-8578
leggett@bush.tamu.edu

Lenzini, Rebecca T.
CARL Systems, Inc.
President
3801 East Florida Avenue
Denver, Colorado 80222
Phone: 303-758-3030
Fax: 303-758-0606
rlenzini@carl.org

Lesk, Michael
Bellcore
Phone: 201-829-4070
Fax: 201-829-5981
lesk@bellcore.com
lesk@parody.bellcore.com

Levy, David
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

DLevy.parc@xerox.com

Liberman, Mark
University of Pennsylvania
Computer Science and Linguistics
Phone: 215-898-0141
Fax: 215-573-2091
myl@unagi.cis.upenn.edu

Liston, David
Council on Library Resources
1785 Massachusetts Ave., N.W.
Washington, D.C. 20036-2117

Lucier, Richard E.
University of California at San Francisco
UCSF Health Sciences
University Librarian and
Academic Vice Chancellor for Information Management
The Library
530 Parnassus Avenue
San Francisco, California 94143-0840
Phone: 415-476-8293
Fax: 415-476-4653
REL@LIB.UCSF.EDU

Lyman, Peter
University of Southern California
University Librarian and Dean of the University Libraries
3550 Trousdale Parkway, Room 100
Los Angeles, California 90089-0182
Phone: 213-740-7317
Fax: 213-749-1221
lyman@vm.usc.edu

Lynch, Clifford A.
University of California at Berkeley
Office of the President
Director, Library Automation
300 Lakeside Drive, 8th Floor
Oakland, California 94612-3350

Phone: 510-987-0522
Fax: 510-839-3573
calur@uccmvsa.ucop.edu
lynch@postgres.berkeley.edu

Lynn, M. Stuart
Cornell University
Vice President, Information Technologies
308 Day Hall
Ithaca, New York 14853-2801
Phone: 607-255-7445
Fax: 607-254-5333
msl1@cornell.edu

Malamud, Carl
Internet Talk Radio
Founder
302 W. W. Glendale Avenue
Alexandria, Virginia 22301
Fax: 703-548-1280
carl@malamud.com

Masinter, Larry
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Mayhew, Jean
United Technologies Research Center
Responsible for library network all over but seven countries in the world

McCray, Alexa T.
National Library of Medicine
8600 Rockville Pike
Besthesda, MD 20894
Phone: 301-496-9300
Fax: 301-496-0673
mccray@nlm.nih.gov

McGill, Michael J.
University of Michigan
Director, Network Systems
1075 Beal Avenue
3208 Computing Center
Ann Arbor, Michigan 48109-2112
Phone: 313-764-4262
Fax: 313-763-4050
Michael.McGill@um.cc.umich.edu

Menell, Peter
Stanford University
School of Law (1992-93)
Stanford, CA 94305
Phone: 415-723-0886
Fax: 415-723-0253

Michaiko, James P.
The Research Libraries Group, Inc
President
1200 Villa Street
Mountain View, California 94041-1100
Phone: 415-691-2243
Fax: 415-964-0943
bl.jpm@rlg.stanford.edu

Moline, Judi
National Institute of Standards Technology (NIST)
Computer Scientist
Building 225, Room B266
Gaithersburg, Maryland 20899
Phone: 301-975-3351
Fax: 301-590-0932
moline@mml.ncsl.nist.gov

Nelson, Mike
Executive Office of the Vice President
Washington, D.C. 20500
Phone: 202-395-6175
Hearings on Electronic Libraries

"Son of Gore" Legislation
mnelson@nsf.gof

Noreault, Terry
OCLC, Inc.
6565 Frantz Road
Dublin, Ohio 43017-0702
Phone: 614-764-4392
terry@rsch.oclc.org

Nunberg, Geoffrey
Xerox/PARC
Stanford University

Okerson, Ann Lillian
Association of Research Libraries
Director, Office of Scientific and Academic Publishing
21 Dupont Circle
Washington, DC 20036
Phone: 202-296-2296
Fax: 202-872-0084
ann@cni.org

Olson, Don
Eastman Kodak
Director of New Market Development
343 State Street
Rochester, New York 14650-0811
Phone: 716-781-1520
Fax: 716-724-0826
olson@pixel.kodak.com

Palmer, Roger
Getty Art History Information Program
Technical Manager
401 Wilshire Boulevard
Suite 1100
Santa Monica, California 90401-1455
Phone: 310-395-1025
Fax: 310-451-5570

Penniman, David W.
Council on Library Resources
President
1785 Massachusetts Ave, N.W.
Washington, D.C. 20036-2117
dpenn@cni.org

Perry, Susan Lane
Stanford University
Director, Departmental Systems Group
Stanford Data Center
Stanford, California 94305-4136
Phone: 415-725-8425
Fax: 415-723-2011
gd.slp@forsythe.stanford.edu

Peters, Paul Evan
Coalition for Networked Information (CNI)
Executive Director
21 Dupont Circle
Washington, D.C. 20036
Phone: 202-296-5098
Fax: 202-872-0884
paul@cni.org

Polly, Jean
Public Library Prospective
polly@lpl.org

Prime, Eugenie
Hewlett Packard Labs Research Library
Previous Publisher of printed indexes became a database on Dialog, BRS, and Datastar

Raghavan, V. V.
Center for Advanced Computer Studies
University of Southwestern Louisiana
P.O. Box 44330
Lafayette, LA 70504-4330
raghavan@cacs.usl.edu

Reich, Vicky
Libraries at Stanford University
Fax: 415-723-4775
Stanford University
Stanford, CA 94305-2125
vicky.reich@forsythe.stanford.edu

Rodgers, David L.
American Mathematical Society
Electronic Products and Services Director
Electronic Journal Project with A. Okerson
416 Fourth Street
Ann Arbor, MI 48103
Phone: 313-995-4123
Fax: 313-996-2916
dlr@math.ams.org

Rose, Dan
Apple
rose@apple.com

Rosenberg. Lawrence C.
National Science Foundation
Deputy Divison Director
Information, Robotics & Intelligent Systems Division
4201 Wilson Boulevard, Room 1115
Arlington, VA 22230
Phone: 703-306-1927x1174
lrosenbe@nsf.gov

Ryland, Jane Norman
CAUSE
President
4840 Pearl East Circle, Suite 302E
Boulder, Colorado 80301-2454
Phone: 303-449-4430
Fax: 303-440-0461
jryland@cause.colorado.edu

Sabelli, Nora H.
National Science Foundation

Program Director, Applications of Advanced Technologies
4201 Wilson Boulevard, Room 885
Arlington, VA 22230
Phone: 703-306-1651x5888
nsabelli@nsf.gov

Saltzer, Jerome H.
MIT
Department of Electrical Engineering and Computer Science
Professor
NE43-513
Cambridge, MA 02139
Phone: 617-253-6016
saltzer@mit.edu

Sato, Andrea
Columbia University School of Law
Janus Project Director
435 West 116th Street, Room 7W1
New York, New York, 10027-7297
Phone: 212-854-2635
Fax: 212-854-3295

Schnase, John L.
School of Medicine Library and
Biomedical Communications Center
Washington University School of Medicine
Director, Advanced Technology Group
660 South Euclid Avenue (Box 8132)
St. Louis, Missouri 63110
Phone: 314-362-3117
Fax: 314-362-0190
schnase@medicine.wustl.edu

Seiden, Peggy
Skidmore College
Head of Public Services
North Broadway
Saratoga Springs, New York 12866-1632
Phone: 518-584-7400 Ext. 2126
Fax: 518-584-3023

pseiden@skidmore.edu

Shaw, Ward
CARL Systems, Inc.
Chairman and Chief Executive Officer
3801 East Florida Avenue
Building D, Suite 300
Denver, Colorado 80210
Phone: 303-758-3030
Fax: 303-758-0606
wshaw@carl.org

Simpson, Donald
Center for Research Libraries
President
6050 South Kenwood Avenue
Chicago, Illinois 60637-2804
Phone: 312-955-4545
Fax: 312-955-4339
simpson@uhuru.uchicago.edu

Smarr, Larry
NCSA
University of Illinois Library School
Professor, Director of Supercomputing Applications Division
155 Comp. Appl. Building, MC 476
Champaign, IL 61820
Phone: 217-244-0079
smarr#ncsa.uiuc.edu

Soloway, Elliot
University of Michigan
Department of Electrical Engineering and Computer Science
Associate Professor
Phone: 313-936-1562
soloway@caen.engin.umich.edu

St. George, Art
University of New Mexico
CIRT
Executive Network Services Officer

Albuquerque, New Mexico 87131
Phone: 505-277-8423
Fax: 505-277-8101
stgeorge@bootes.unm.edu

Thaker, Al
American Math Society
Electronic Journal Project with A. Okerson

Tompkins, Philip
Estrella Mountain Community College Center
Director of Library Information Services
3000 Dysart Road
Litchfield Park, Arizona 85340
Phone: 602-935-8077
Fax: 602-035-8008
tompkins@names.maricopa.edu

Turner, Judith
Chronicle of Higher Education
Director, Chronicle Information Services
1255 23rd Street, NW
Washington, D.C. 20037
Phone: 202-466-1063
Fax: 202-296-2691
JUDITH@PAGE1.com

Urbanowski, Frank
MIT Press
Director
55 Hayward Street
Cambridge, MA 02142
Phone: 800-346-0343
Fax: 617-258-6779

Vassallo, Paul
National Institute of Standards and Technology
Director, Office of Information Services
Admin. E106
NIST

Gaithersburg, Maryland 20899
Phone: 301-975-2786
Fax: 301-869-8071
pvassa11@micf.nist.gov

von Wahlde, Barbara
State Univeristy of New York at Buffalo
Associate Vice President for University Libaries
Buffalo, New York 14260
Phone: 716-636-2967
Fax: 716-636-3844
unlbvw@ubvms.cc.buffalo.edu

Watts, Carol B.
National Environmental Satellite Data and Information Service
National Oceanic & Atmospheric Administration
Chief, Library and Information Services Division
6009 Executive Boulevard
Rockville, Maryland 20852
Phone: 301-443-8287
Fax: 301-443-0237
watts@nodc2.nodc.noaa.gov

Webster, Duane E.
Association of Research Libraries (ARL)
Executive Director
21 Dupont Circle
Washington, D.C. 20036
Phone: 202-296-2296
Fax: 202-872-0884
duane@cni.org

West, Richard Paul
University of California, Office of the President
Associate Vice President, Information Systems and Administrative Services
300 Lakeside Drive, 8th Floor
Oakland, California 94612-3350
Phone: 510-987-0405
Fax: 510-839-3573

west@rio.ucop.edu

Wiederhold, Gio
DARPA/SISTO
ACM Publication Board
Phone: 703-696-2249
Fax: 703-696-2202
gio@darpa.mil

Williams, James F. II
University of Colorado at Boulder
Dean of Libraries
Norlin Library N210C
Campus Box 184
Boulder, Colorado 80309-0184
Phone: 303-492-7511
Fax: 303-492-8775
williams_j@cubldr.colorado.edu

Williams, Cynthia
MCC (Microelectronics and Computer Technology Corporation)
Manager Communications
3500 West Balcones Center Drive
Austin, TX 78759-6509
Phone: 512-338-3512
Fax: 512-338-3892
williams@mcc.com

Winograd, Terry A.
Stanford University
Professor, Department of Computer Science
Bldg 460 Room 238
Stanford, CA 94305-2140
Phone: 415-723-2780
Fax: 415-725-7411
winograd@cs.stanford.edu
Interval Research (1992-93)
Phone: 415-354-0854
Fax: 415-354-0872
winograd@interval.com

Woelk, Darrell
NCR/MCC
3500 W. Balcones Center Drive
Austin, TX 78759
Phone: 512-338-3468
woelk@mcc.com

Zemankova, Maria
On Leave 1993-4 Mitre Corporation
National Science Foundation
Program Director Database and Expert Systems
4201 Wilson Boulevard
Arlington, VA 22230
mzemanko@note2.nsf.gov

Groups and Organizations

Testimonials related to the National Electronic Library

C. Borgman

Support for the initiative will be needed from all corners of the information world. This list of organizations is anecdotal, from my personal knowledge of the field. It would be useful for the developers of the initiative to do a thorough search of the directories of professional organizations and foundations to find other groups as well. Addresses and contact names for these organizations and others can be located in standard bibliographic sources.

- American Library Association (50K plus members, active lobbying office in Washington, DC, major player, very important source of support for initiative.)
- Library and Information Technology Association (a subgroup of ALA, large, represents those involved in library automation and networks)

- National Commission on Libraries and Information Services (federally funded, organizes White House Conferences on Libraries and Information Services; held hearings on NREN and libraries in July 1992).
- Computer Professionals for Social Responsibility (interested in workplace issues, responsible use of technology, privacy issues. Testified before NCLIS on NREN issues. Washington office headed by Marc Rotenberg.)
- Association for Computing Machinery (especially special interest group on Information Retrieval)
- Coalition for Networked Information (Washington, DC. Working on standards for information delivery on Internet, cooperative issues. New but very active group.)
- Association for Research Libraries (Washington, DC. Umbrella for all major university and other research libraries in the U.S.)
- American Society for Information Science (Silver Spring, MD. Scholarly and professional society for broad range of theoretical and technical issues in information and delivery. Mix of librarians, information retrieval researchers, government and private industry.)
- Council on Library Resources (Washington, DC. Primary private funding agency for library-related research in the U.S.).
- Information Industry Association (Represents small and large electronic publishers)
- Corporation for National Research Initiatives (Vinton Cerf, Robert Kahn, et al. Active in developing network infrastructure.)
- Society for Scholarly Publishing (Represents university and private scholarly presses)
- Society for the Social Studies of Science (Scholarly society, studies scholarly communication infrastructure)
- Special Libraries Association
- Medical Library Association
- Art Libraries of North America (ARLIS)

(other library associations representing special groups)

- Museum Computing Network (building databases of museum information and networking; often modeled on work of libraries)
- Professional societies of major scientific disciplines who would be served (American Chemical Society, Physics, Biology, Geology, etc.)

- American Educational Research Association (interest in classroom technologies)
- Professional education societies (teachers, university educators, etc.)

Internet Society

The Internet Society is a new international, professional membership organization to promote the use of the Internet for research and scholarly communication and collaboration.

The Society provides a forum for government, industry, educators and users to debate and recommend technical standards and procedures for the global Internet and private internets.

The Society seeks to advance open scholarship in all countries.

THE INTERNET SOCIETY ....
- began operations in January 1992
- is governed by an elected Board of Trustees
- seeks to encourage international scientific collaboration through networking

Joining now will ....
- entitle you to the quarterly member newsletter and on-line information recording important developments in the technical and operational evolution of the Internet
- help shape the international agenda for the Society
- support Internet Society objectives

Abstract

The purpose of this document is to provide a brief description of the Internet Society and its goals and objectives. It functions as a professional society to facilitate, support and promote the evolution and growth of the Internet as a global research communications infrastructure. The suggestions and recommendations of all parties interested in the Internet are solicited to assist in making the Internet Society robust,

productive and structured to meet the needs of its members.

Internet Society

The Internet is a collection of cooperating, interconnected, multiprotocol networks which supports international collaboration among thousands of organizations. Internet Society seeks to foster the voluntary interconnection of computer networks into a global research, development, and information infrastructure. The Internet Society does not operate the Internet. Internet operation continues to be a collaborative activity which the Society seeks to facilitate. The Society provides assistance and support to groups and organizations involved in the use, operation and evolution of the Internet. It provides support for forums in which technical and operational questions can be discussed and provide mechanisms through which interested parties can be informed and educated about the Internet, its function, use, operation and the interests of its constituents.

Membership

Internet Society is a professional membership organization with voting individual members and non-voting institutional members. There are several classes of institutional members. The society publishes a newsletter on a regular basis and holds an annual meeting to which all members and other interested parties are invited. The topics of the annual meeting vary, but focus on current research in networking, Internet functionality and growth, and other interests of the Society constituency.

Membership dues vary according to class of membership. The amounts of these dues and the basis on which they are set are determined by the Board of Trustees of the Society and may be revised from time to time as provided in the By-Laws.

Charter

The Society is a non-profit organization and operated for academic, educational, charitable and scientific purposes among which are:

A. To facilitate and support the technical evolution of the Internet as a research and education infrastructure and to stimulate involvement of the academic, scientific and engineering communities, among others in the evolution of the Internet.

B. To educate the academic and scientific communities and the public concerning the technology, use and application of the Internet.

C. To promote scientific and educational applications of Internet technology for the benefit of educational institutions at all grade levels, industry and the public at large.

D. To provide a forum for exploration of new Internet applications and to foster collaboration among organizations in their operation and use of the Internet.

Activities of the Society

1. Support for Internet Technical Evolution

The Internet Architecture Board (IAB) has been concerned with the development and evolution of architectures supporting the use of multiple protocols in a networked environment. The Internet Society has incorporated the IAB and its functions into the operation of the Internet Society. The Internet Society works with other interested organizations to support and assist efforts to evolve the multiprotocol Internet. The Internet Society looks to the Internet Engineering and Research Task Forces to stimulate networking research and facilitate the evolution of the TCP/IP protocol suite and the integration of new protocol suites into the Internet architecture. The Internet Society works actively with parties and organizations interested in fostering improvement in the utility of the Internet for its constituent users.

2. Meetings and Conferences

Internet Society convenes an annual meeting (“INET”) and organizes and facilitates workshops and symposia, jointly with other organizations where appropriate, on specific topics of interest to the Society membership. The annual meeting addresses issues of global and regional importance to the evolution and growth of the Internet.

3. Information and Infrastructure Services

The Internet Society publishes the quarterly Internet Society News providing members with information about the international activities of Internet constituents. In addition, the Society also provides assistance to and support for organizations responsible for maintaining the databases crucial to Internet function (e.g., the Domain Name System, X.500 Directory Services, etc.) and organizations concerned with the security of the Internet (e.g., the Software Engineering Institute Computer Emergency Response Team (CERT)). The Society assists in the development of educational, advisory and informative materials of use to Society members. Where appropriate, the Society organizes or supports activities which aid in the coordination among the organizations operating components of the Internet.

The Society refers members to appropriate parties involved in operating the various parts of the Internet where they may be helpful with specific questions. Where possible, the Society seeks to provide access to its information on-line, but also offers hard copy and, perhaps eventually, CD-ROM-based information resources.

Plans

The initial organizers of the Internet Society include the Corporation for National Research Initiatives (CNRI), EDUCOM, Reseaux Associees pour la Recherche Europeenne (RARE) and the Internet Architecture Board. Computer networking has become a critical infrastructure for the research and development community and has the potential to become the basis for world-wide collaboration and cooperation in every field of human endeavor. The Internet Society seeks to solidify, enhance and encourage further international collaborative networking. Individuals joining the Society during its formation have received special recognition as Society pioneers and have been instrumental in shaping the early agenda of Society activities. Institutional members are represented in the Advisory Council of the Internet Society where their advice and counsel is sought to refine Internet Society objectives and activities.

APPENDIX

A Brief History of the Internet and Related Networks

Introduction

In 1973, the U.S. Defense Advanced Research Projects Agency (DARPA - now ARPA) initiated a research program to investigate techniques and technologies for interlinking packet networks of various kinds. The objective was to develop communication protocols which would allow networked computers to communicate transparently across multiple, linked packet networks. This was called the Internetting project and the system of networks which emerged from the research was known as the “Internet.” The system of protocols which was developed over the course of this research effort became known as the TCP/IP Protocol Suite, after the two initial protocols developed: Transmission Control Protocol (TCP) and Internet Protocol (IP).

In 1986, the U.S. National Science Foundation (NSF) initiated the development of the NSFNET which, today, provides a major backbone communication service for the Internet. With its 45 megabit per second facilities, the NSFNET carries on the order of 12 billion packets per month between the networks it links. The National Aeronautics and Space Administration (NASA) and the U.S. Department of Energy contributed additional backbone facilities in the form of the NSINET and ESNET respectively. In Europe, major international backbones such as NORDUNET, EBONE, EuropaNet and others provide connectivity to over three hundred thousand computers on a large number of networks. Commercial network providers in the U.S. and Europe are beginning to offer Internet backbone and access support on a competitive basis to any interested parties.

“Regional” support for the Internet is provided by various consortium networks and “local” support is provided through each of the research and educational institutions. Within the United States, much of this support has come from the federal and state governments, but a considerable contribution has been made by industry. In Europe and elsewhere, support arises from cooperative international efforts and through national research organizations. During the course of its evolution, particularly after 1989, the Internet system began to integrate support for other protocol suites into its basic networking fabric. The present emphasis in the system is on multiprotocol interworking, and in particular, with the integration of the Open Systems Interconnection (OSI) protocols into the architecture.

Both public domain and commercial implementations of the roughly 100 protocols of TCP/IP protocol suite became available in the 1980’s. During the early 1990’s, OSI protocol implementations also became available and, by the end of 1992, the Internet had grown to include some 11,000 networks in over seventy countries, serving over 1,700,000 host computers used by over 5,000,000 people.

A great deal of support for the Internet community has come from the U.S. Federal Government, since the Internet was originally part of a federally-funded research program and, subsequently, has become a major part of the U.S. research infrastructure. During the late 1980’s, however, the population of Internet users and network constituents expanded internationally and began to include commercial facilities. Indeed, the bulk of the system today is made up of private networking facilities in educational and research institutions, businesses and in government organizations across the globe.

The Coordinating Committee for Intercontinental Networking (CCIRN), which was organized by the U.S. Federal Networking Council (FNC) and the European

Reseaux Associees pour la Recherche Europeenne (RARE), plays an important role in the coordination of plans for government-sponsored research networking. CCIRN efforts have been a stimulus for the support of international cooperation in the Internet environment. Reseau IP Europeenne (RIPE) has mounted a very successful grass-roots effort in Europe to extend Internet in Europe.

Internet Technical Evolution

Over its fifteen year history, the Internet has functioned as a collaboration among cooperating parties. Certain key functions have been critical for its operation, not the least of which is the specification of the protocols by which the components of the system operate. These were originally developed in the DARPA research program mentioned above, but in the last five or six years, this work has been undertaken on a wider basis with support from Government agencies in many countries, industry and the academic community. The Internet Activities Board (recently re-named the Internet Architecture Board) was created in 1983 to guide the evolution of the TCP/IP Protocol Suite and to provide research advice to the Internet community.

During the course of its existence, the IAB has reorganized several times. It now has two primary components: the Internet Engineering Task Force and the Internet Research Task Force. The former has primary responsibility for further evolution of the TCP/IP protocol suite, its standardization, and the integration of other protocols into Internet operation (e.g., the Open Systems Interconnection protocols). The Internet Research Task Force continues to organize and explore advanced concepts in networking under the guidance of the Internet Architecture Board and with support from various research sponsoring agencies.

A secretariat has been created to manage the day-to-day function of the Internet Architecture Board and Internet Engineering Task Force. IETF meets three times a year in plenary and its approximately 80 working groups convene at intermediate times by electronic mail, teleconferencing and at face-to-face meetings. The IAB meets quarterly face-to-face or by videoconference and at intervening times by telephone, electronic mail and computer-mediated conferences.

Two other functions are critical to Internet function: Publication of documents describing the Internet and the assignment and recording of various identifiers needed for protocol operation. Throughout the development of the Internet, its protocols and other aspects of its operation have been documented first in a series of documents called Internet Experiment Notes and, later, in a series of documents called Requests for Comment (RFCs). The latter were used initially to document the protocols of the first packet switching network developed by DARPA, the

ARPANET, beginning in 1969, and have become the principal archive of information about the Internet. At present, the publication function is provided by an RFC editor.

The recording of identifiers is provided by the Internet Assigned Numbers Authority (IANA) who has delegated one part of this responsibility to an Internet Registry which acts as a central repository for Internet information and which provides central allocation of network and autonomous system identifiers, in some cases to subsidiary registries located in various countries. The Internet Registry (IR) also provides central maintenance of the Domain Name System (DNS) root database which points to subsidiary distributed DNS servers replicated throughout the Internet. The DNS distributed database is used, inter alia, to associate host and network names with their Internet addresses and is critical to the operation of the higher level TCP/IP protocols including electronic mail.

There are a number of Network Information Centers (NICs) located throughout the Internet to serve its users with documentation, guidance, advice and assistance. As the Internet continues to grow internationally, the need for high quality NIC functions increases. Although the initial community of users of the Internet were drawn from the ranks of computer science and engineering, its users now comprise a wide range of disciplines in the sciences, arts, letters, business, military and government administration.

Related Networks

In 1980-81, two other networking projects, BITNET and CSNET, were initiated. BITNET adopted the IBM RSCS protocol suite and featured direct leased line connections between participating sites. Most of the original BITNET connections linked IBM mainframes in university data centers. This rapidly changed as protocol implementations became available for other machines. From the beginning, BITNET has been multi-disciplinary in nature with users in all academic areas. It has also provided a number of unique services to its users (e.g., LISTSERV). Today, BITNET and its parallel networks in other parts of the world (e.g., EARN in Europe) have several thousand participating sites. In recent years, BITNET has established a backbone which uses the TCP/IP protocols with RSCS-based applications running above TCP.

CSNET was initially funded by the National Science Foundation (NSF) to provide networking for university, industry and government computer science research groups. CSNET used the Phonenet MMDF protocol for telephone-based electronic mail relaying and, in addition, pioneered the first use of TCP/IP over X.25 using commercial public data networks. The CSNET name server provided an early example of a white pages directory service and this software is still in use at numerous sites. At its peak, CSNET had approximately 200 participating sites and international connections to approximately fifteen countries.

In 1987, BITNET and CSNET merged to form the Corporation for Research and Educational Networking (CREN). In the Fall of 1991, CSNET service was discontinued having fulfilled its important early role in the provision of academic networking service. A key feature of CREN is that its operational costs are fully met through dues paid by its member organizations.

Introducing the Current Trustees and Officers of the Internet Society

After the chartering of the Internet Society late in December 1991, the three initial trustees, Kenneth King, Juergen Harms and Robert Kahn unanimously elected an interim Board of Trustees, shown below. The interim board is responsible for initial operation of the Internet Society including approval of budgets, appointment of officers and election of an expanded Board whose members will serve for periods ranging from one to three years. At annual intervals, nominations for Board positions will be opened and elections held to refresh approximately one-third of the Board of Trustees each year. The first such election by the individual Internet Society members takes place in Spring 1993, with the elected officials taking office in July 1993.

Internet Society Board of Trustees

Frode Greisen, Treasurer
<frode.greisen@uni-c.dk>

Kees Neggers
<neggers@surfnet.nl>

Geoff Huston
<g.huston@aarnet.edu.au>

Michael Roberts, Executive Director
<roberts@educom.edu>

Anthony Rutkowski, Vice President
<amr@sprint.com>

Introducing the Charter, Founding, and Organizational Members

CHARTER MEMBERS

CORPORATION FOR NATIONAL RESEARCH INITIATIVES
EDUCOM
RESEAU ASSOCIEES POUR LA RECHERCHE EUROPEENNE

FOUNDING MEMBERS

ADVANCED NETWORK & SERVICES
APPLE COMPUTER CORPORATION
AT&T
AUSTRALIAN ACADEMIC AND RESEARCH NETWORK
BELL COMMUNICATIONS RESEARCH
BOLT BERANEK AND NEWMAN
CISCO SYSTEMS
COALITION FOR NETWORKED INFORMATION
CORPORATION FOR OPEN SYSTEMS INTERNATIONAL
CORPORATION FOR RESEARCH & EDUCATIONAL NETWORKING
DEFENSE INFORMATION SYSTEMS AGENCY
DIGITAL EQUIPMENT CORPORATION
EUROPEAN ACADEMIC RESEARCH NETWORK
EUROPEAN LABORATORY FOR PARTICLE PHYSICS
FREEPORT-MCMORAN
HUGHES AIRCRAFT COMPANY
INTERNATIONAL BUSINESS MACHINES CORPORATION
INTEROP COMPANY
ISRAELI INTER-UNIVERSITY COMPUTATION CENTER
LAWRENCE LIVERMORE NATIONAL LABORATORY

MCI COMMUNICATIONS CORPORATION
MICROSOFT CORPORATION
NATIONAL LIBRARY OF MEDICINE
NORDUNET
NOVELL, INC.
NYSERNET, INC.
PROTEON, INC.
SIEMENS AG
SOFT-SWITCH, INC.
SOFTWARE ENGINEERING INSTITUTE
SPRINT
3COM CORPORATION
UNIVERSITY OF NOTRE DAME
UNIVERSITY OF WASHINGTON
U S WEST COMMUNICATIONS, INC.
UUNET TECHNOLOGIES
WELLFLEET COMMUNICATIONS INC.

ORGANIZATIONAL MEMBERS

ELECTRONIC FRONTIER FOUNDATION
NYNEX SCIENCE & TECHNOLOGY, INC.
TENON INTERSYSTEMS
VEDA DATA SYSTEMS, INC.

Terms of Membership

Charter and Founding Organizations provide the Internet Society with vital financial support. This substantial and early support has made the founding of the Internet Society possible. For-profit Founding members commit to a total of $20,000 during the 1992 and 1993 period and $10,000 per year thereafter. Non-profit, Founding government and educational institutions commit to half that amount. Regular for-profit and non-profit organizational members commit to $10,000 and $5,000 per year respectively. There is also provision for start-ups to become regular members during their first three years at a cost of $1,000 per year. Organizations interested in participating in this program should contact Vinton Cerf at the Internet Society secretariat.

Internet Society Advisory Council

The Organizational Members of the Internet Society provide a representative and an alternate to an Advisory Council. The Society is grateful to these individuals and their organizations for the commitments they have made to its success. The names of the Advisory Council members are listed below:

Rick Adams UUNET
Ed Albrigo (alt.) CORPORATION FOR OPEN SYSTEMS
Guy Almes ANS
Stephen An WELLFLEET COMMUNICATIONS INC.
Michael Arnold FREEPORT-MCMORAN
Cliff Bamford MICROSOFT
William Biagi CORPORATION FOR OPEN SYSTEMS
Daniele Bovio (alt.) EARN
David Brandin INTEROP COMPANY
George Buchanan HUGHES AIRCRAFT COMPANY
Lee Caldwell NOVELL
Michael Carter APPLE
Steve Cisler (alt.) APPLE
Avi Cohen IUCC
Sam Coleman (alt.) LAWRENCE LIVERMORE NATIONAL LAB
James Conklin (alt.) CREN
Michael Conn MCI COMMUNICATIONS
Hans Deckers EARN
Rashmi Doshi (alt.) NYNEX SCIENCE & TECHNOLOGY, INC.
Mathew Dovens (alt.) MCI COMMUNICATIONS
Robin Erskine AARNET
David Farber ELECTRONIC FRONTIER FOUNDATION
Francois Fluckiger CERN
Ira Fuchs CREN
Dain Gary (alt.) SOFTWARE ENGINEERING INSTITUTE
Cary Giese (alt.) U S WEST
William Grant U S WEST
Terence Gray UNIVERSITY OF WASHINGTON
Erik Grimmelmann AT&T
Phill Gross ANS
Roger Gulbranson (alt.) UNIVERSITY OF NOTRE DAME
Anita Holmgren TENON INTERSYSTEMS
Steve Holmgren (alt.) TENON INTERSYSTEMS
E.W. Bud Huber (alt.) HUGHES AIRCRAFT COMPANY

Sascha Ignjatovic VEDA DATA SYSTEMS, INC.
Ole Jacobsen (alt.) INTEROP COMPANY
Ron Johnson (alt.) UNIVERSITY OF WASHINGTON
Walter Johnston NYNEX SCIENCE & TECHNOLOGY, INC.
Mitch Kapor (alt.) ELECTRONIC FRONTIER FOUNDATION
Anthony Lauck DIGITAL
Donald Lindberg NATIONAL LIBRARY OF MEDICINE
James Luckett (alt.) NYSERNet
Richard Mandelbaum NYSERNet
Olivier Martin (alt.) CERN
Daniel Masys NATIONAL LIBRARY OF MEDICINE
Stuart Mathison SPRINT INTERNATIONAL
Jeff Mayersohn BOLT BERANEK AND NEWMAN
David McDonald (alt.) FREEPORT-MCMORAN
John Mullarvey (alt.) AARNET
Catherine Muther CISCO SYSTEMS
Dan Nessett LAWRENCE LIVERMORE NATIONAL LAB
Camillo J. Pasquariello DEFENSE INFORM. SYSTEMS AGENCY
Andrew Partan (alt.) UUNET
Janet Perry (alt.) NOVELL
Paul Evan Peters CNI
Rich Pethia SOFTWARE ENGINEERING INSTITUTE
John Pickens 3COM
Werner Sammer SIEMENS AG
Thomas Schwarcz (alt.) SIEMENS AG
Paul Severino (alt.) WELLFLEET COMMUNICATIONS INC.
Robert Shahan (alt.) IBM
David Sincoskie BELLCORE
Donald Spicer UNIVERSITY OF NOTRE DAME
Leonard Swatski (alt.) DEFENSE INFORM. SYSTEMS AGENCY
Dave Thompson (alt.) MICROSOFT
Michael Thurk (alt.) DIGITAL
Paul Toldalagi PROTEON
Nicholas R. Trio IBM
Paul Tsuchiya (alt.) BELLCORE
Peter Villemoes NORDUNET
Richard West (alt.) CNI
Michael D. Zisman SOFT-SWITCH, INC.

Introducing the Internet Society Secretariat

During its initial period of operation, the Internet Society secretariat occupies space provided by the Corporation for National Research Initiatives and by EDUCOM:

Internet Society
1895 Preston White Drive, Suite 100
Reston, VA 22091 USA
+1 703 648 9888
+1 703 620 0913 FAX
isoc@isoc.org

Britt Jackman - Secretarial Support
bjackman@cnri.reston.va.us

Cynthia Matthews - Individual Membership Services
isoc@isoc.org (preferred)
cmatthew@cnri.reston.va.us

John Stewart - Technical Support
jstewart@cnri.reston.va.us

Theresa Weigler - Organizational Membership Services
tweigler@cnri.reston.va.us

In addition to these, two EDUCOM staff members are also assisting in the operation of the Society:

EDUCOM
1112 16th Street NW, Suite 600
Washington, DC 20036
+1 202 872 4200
+1 202 872 4318 FAX

Michael Roberts - General planning and operation
roberts@educom.edu

Elizabeth Barnhart- Conference/Publication Support
barnhart@educom.edu

Library-Oriented Lists and Electronic Serials

By Charles W. Bailey, Jr.

1.0 Computer Conferences

Computer conferences are becoming an increasingly important form of communication for librarians. There are a growing number of computer conferences of interest to librarians on BITNET and Internet.

Computer conferences are commonly called “lists.” The software used to support these lists is typically referred to as the “list server.”

1.1 Lists That Use Eric Thomas’s List Server

Many BITNET lists utilize a list server that was developed by Eric Thomas.

Please note that some of these BITNET lists also have Internet addresses, which are not shown here.

(Entries that appear in this document for the first time or have been changed are marked with an asterisk.)

ADVANC-L@IDBSU Geac Advanced Library System
AFAS-L@KENTVM African American Studies and Librarianship
ALF-L@YORKVM1 Academic Librarian’s Forum
ARCHIVES@INDYCMS Archives and Archivists List
ARIE-L@IDBSU RLG Ariel Document Transmission System
ARLIS-L@UKCC Art Libraries Association of North America
ASIS-L@UVMVM American Society for Information Science
ATLAS-L@TCUBVM Data Research ATLAS Users
AUTOCAT@UVMVM Library Cataloging and Authorities Discussion Group
BI-L@BINGVMB Bibliographic Instruction
BIBSOFT@INDYCMS Discussion of Software for Citations and Bibliographies
BRS-L@USCVM BRS/Search Users
BUSLIB-L@IDBSU Business Librarians
CALL-L@UNBVM1 Canadian Academic Law Libraries List*

CARL-L@UHCCVM CARL Users
CDPLUS-L@UTORONTO CDPLUS Software User Group
CDROMLAN@IDBSU CD-ROM LANs
CDS-ISIS@HEARN UNESCO's CDS/ISIS Text Retrieval Software
CHMINF-L@IUBVM Chemical Information Sources
CIRCPLUS@IDBSU Circulation and Access Services
COLLDV-L@USCVM Library Collection Development List
CONSALD@UTXVM Committee on South Asian Libraries and Documentation
COSNDISC@BITNIC Consortium for School Networking
CWIS-L@WUVMD Campus-Wide Information Systems
ELDNET-L@UIUCVMD ASEE Engineering Libraries Division
ELEASAI@ARIZVM1 Open Library/Information Science Research Forum
ELLASBIB@GREARN Library Automation in Greece
ENDNOTE@UCSBVM EndNote/EndLink Users Forum
EXLIBRIS@RUTVM1 Rare Books and Special Collections Forum
FEMINIST@MITVMA ALA Social Responsibility Round Table Feminist Task Force
FISC-L@NDSUVM1 Fee-Based Information Service Centers in Academic Libraries
GEONET-L@IUBVM Geoscience Librarians and Information Specialists
GOVDOC-L@PSUVM Government Documents
ILL-L@UVMVM Interlibrary Loan
IMUG-L@OHSTVMA Innopac Music Users Group
INDEX-L@BINGVMB Indexer's Discussion Group
INFO+REF@INDYCMS Information + Referral List
INNOPAC@MAINE Innovative Interfaces Users
INT-LAW@UMINN1 Foreign and International Law Librarians
JESSE@ARIZVM1 Open Library/Information Science Education Forum
LABMGR@UKCC Academic Microcomputer Lab Management
LIBADMIN@UMAB Library Administration and Management
LIBEVENT@USCVM Library Events in Southern California
LIBEX-L@MAINE Exhibits and Academic Libraries Discussion List*
LIBMASTR@UOTTAWA Library Master Bibliographic Database
LIBPER-L@KSUVM Library Personnel Issues
LIBPLN-L@QUCDN Library Planning
LIBRARY@INDYCMS Libraries and Librarians
LIBREF-L@KENTVM Discussion of Library Reference Issues
LIBSUP-L@UWAVM Library Support Staff
LM_NET@SUVM School Library Media & Network Communications
MAPS-L@UGA Maps and Air Photo Forum
MEDLIB-L@UBVM Medical and Health Sciences Libraries

MLA-L@IUBVM Music Library Association
MULTILIS@ALBNYVM1 multiLIS Users
NETTRAIN@UBVM Internet/BITNET Network Trainers Discussion List
NOTIS-L@TCSVM NOTIS Users
NOTISACQ@CUVMB NOTIS Acquisitions Discussion Group
NOTMUS-L@UBVM Notis Music Library List
NOTRBCAT@INDYCMS Rare Book and Special Collections Catalogers
NYSO-L@UBVM MLA New York State/Ontario Chapter Discussion List
OFFCAMP@WAYNEST1 Off-Campus Library Services List
PACS-L@UHUPVM1 Public-Access Computer Systems Forum
PACS-P@UHUPVM1 PACS-L Publications Only
PRO-CITE@IUBVM The Personal Bibliographic Software Discussion List
RLIN-L@RUTVM1 RLIN Users*
SAVEIT-L@USCVM SAVEIT Users*
SERIALST@UVMVM Serials Users Discussion Group
SLAJOB@IUBVM Special Libraries Association Employment Opportunities*
SPILIB-L@SUVM SPIRES Users
UNICRN-L@PSUORVM SIRSI/UNICORN Automated Library Systems
USMARC-L@MAINE USMARC Advisory Group Forum
VETLIB-L@VTVM2 Veterinary Medicine Library Issues and Information*
VPIEJ-L@VTVM1 Publishing E-Journals: Publishing, Archiving, and Access
Z3950IW@NERVM Z39.50 Implementors Workshop

To subscribe to a list, send the following e-mail message to LISTSERV@NODE, where NODE is the part of the address after the “@” character:

SUBSCRIBE List First_Name Last_Name

(If you are not on BITNET, ask your computer center how to address a message to the desired BITNET node.)

For example, Jane Doe sends the following e-mail message to LISTSERV@UHUPVM1 to subscribe to PACS-L:

SUBSCRIBE PACS-L Jane Doe

You can obtain a directory of list server documentation by sending the following e-mail message to the list server: INFO ?

Once you know the name of the desired documentation file, send another INFO command to the list server to obtain the file. For example, to get a file that describes searching the message database of a list, send the following command: INFO DATABASE.

1.2 Lists That Use Anastasios Kotsikonas's List Server

Anastasios Kotsikonas's list server is being used to support a growing number of Internet lists.

To subscribe to a list, send the following e-mail message to LISTSERV@NODE, where NODE is the part of the address after the "@" character:

SUBSCRIBE List First_Name Last_Name

(If you are not on Internet, ask your computer center how to address a message to the desired Internet node.)

To obtain list server documentation, send the following e-mail message to the list server: HELP.

CNI-ANNOUNCE@CNI.ORG CNI's News Announcement Network
CNI-ARCHITECTURE@CNI.ORG CNI's Architectures and Standards Working Group
CNI-BIGIDEAS@CNI.ORG CNI's Big Ideas Forum
CNI-COPYRIGHT@CNI.ORG CNI's Copyright and Intellectual Property Forum
CNI-DIRECTORIES@CNI.ORG CNI's Directories and Resource Information Services Working Group
CNI-LEGISLATION@CNI.ORG CNI's Legislation, Codes, Policies, and Practices Working Group
CNI-MANAGEMENT@CNI.ORG CNI's Management and Professional and User Education Working Group
CNI-MODERNIZATION@CNI.ORG CNI's Modernization of Scholarly Publication Working Group
CNI-PUBINFO@CNI.ORG CNI's Access to Public Information Working Group
CNI-TEACHING@CNI.ORG CNI's Teaching and Learning Working Group
CNI-TRANSFORMATION@CNI.ORG CNI's Transformation of Scholarly Communication Working Group
PALS-L@KNUTH.MTSU.EDU PALS System
PUBLIB@NYSERNET.ORG Public Libraries and the Internet*
VISIONS@LIBRARY.SDSU.EDU Strategic Visions Steering Committee Electronic Discussion Forum on the Future of Librarianship

1.3 Other Lists

These lists use diverse software. Contact the person who sponsors the list you are interested in to get further information about how the list software works.

ACRLNY-L (Listings of Library Jobs and Events)
Send the following command to LISTSERV@NYUACF:
SUBSCRIBE ACRLNY-L First_Name Last_Name.

AGRIS-L (Agricultural Information)
Send the following message to LISTSERV@IRMFAO01:
SUBSCRIBE AGRIS-L First_Name Last_Name.

ALEPHINT (ALEPH Library System Users)
Send the following command to LISTSERV@TAUNIVM:
SUBSCRIBE ALEPHINT First_Name Last_Name.

CADUCEUS (History of Medicine Collections Forum)
Send a subscription request to Inci Bowman:
IBOWMAN@UTMBEACH.

Conservation DistList (Conservation of Archive, Library, and
Museum Materials)
Send a subscription request to Walter Henry:
WHENRY@LINDY.STANFORD.EDU.

CORMOSEA (Committee on Research Materials on Southeast Asia)
Send a subscription request to Kent Mulliner:
MULLINER@OUVAXA.CATS.OHIOU.EDU.

DYNIX_L (DYNIX Users)
Send a subscription request to:
DYNIX_L-REQUEST@SBU.EDU.

EASTLIB (Committee on East Asian Libraries)
Send the following command to LISTSERV@MENTO.OIT.UNC.EDU:
SUBSCRIBE EASTLIB First_Name Last_Name.

IAMSLIC (International Association of Aquatic and
Marine Science Libraries and Information Centers)
Send the following command to LISTSERV@UCSD.EDU:

SUBSCRIBE Your E-Mail Address IAMSLIC.

KATALIST (Discussion on Library Systems and Databases--In Hungarian)
Send the following command to LISTSERV@HUEARN:
SUBSCRIBE KATALIST First_Name Last_Name.

LAW-LIB (Law Librarians)
Send subscription requests to: LAW-REQ@UCDAVIS.EDU.
Contact Elizabeth St. Goar for technical questions:
ESTGOAR@UCDAVIS.EDU.

LIB_HYTELNET (HYTELNET Program)
Send subscription request to Peter Scott: SCOTT@SKLIB.USASK.CA.

LS2K (LS/2000 Users Group)
Send subscription request to: LS2K-REQUEST@CC.UTAH.EDU.

PAMnet (Physics, Astronomy, and Mathematics Librarians)
Send a subscription request to Joanne Goode:
JMGOODE@UKCC.UKY.EDU.

STUMPERS-L (Difficult Reference Questions)
Send the following message to ROSLIBREFRC@CRF.CUIS.EDU:
SUBSCRIBE STUMPERS-L Your E-Mail Address.

1.4 Diane Kovacs's Directory

Diane Kovacs and her colleagues have compiled a useful directory of lists in many subject areas.

To retrieve this directory, send the following e-mail message to LISTSERV@KENTVM or LISTSERV@KENTVM.KENT.EDU:

GET ACADLIST README F=MAIL
GET ACADLIST FILE1 F=MAIL
GET ACADLIST FILE2 F=MAIL
GET ACADLIST FILE3 F=MAIL
GET ACADLIST FILE4 F=MAIL
GET ACADLIST FILE5 F=MAIL
GET ACADLIST FILE6 F=MAIL

GET ACADLIST FILE7 F=MAIL

For further information about the directory, contact Diane Kovacs: DKOVACS@KENTVM.

2.0 Electronic Serials

There are a growing number of library-related electronic journals and newsletters available on BITNET and Internet.

ACQNET (The Acquisitions Librarian's Electronic Network)
Send a subscription request to Christian Boissonnas:
CRI@CORNELLC.

ALA Washington Office Newsline (ALAWON)
Send the following e-mail message to LISTSERV@UICVM:
SUBSCRIBE ALA-WO First_Name Last_Name.

ALCTS NETWORK NEWS (Association for Library Collections and
Technical Services)
Send the following message to LISTSERV@UICVM:
SUBSCRIBE ALCTS First_Name Last_Name.
For a list of back issue files, send the message:
INDEX ALCTS.

The Arachnet Electronic Journal of Virtual Culture*
Send the following message to LISTSERV@UOTTAWA:
SUBSCRIBE ARACHNET First_Name Last_Name.

Citations for Serial Literature
Send the following message to LISTSERV@MITVMA:
SUBSCRIBE SERCITES First_Name Last_Name.
For a list of back issue files, send the message:
INDEX SERCITES.

Consortium Update (SPIRES)
Send a subscription request to: HQ.CON@STANFORD.

Current Cites
TELNET MELVYL.UCOP.EDU; Enter command: SHOW CURRENT CITES.
(Also distributed on PACS-L.)

Further information: David F. W. Robison, DROBISON@UCBLIBRA.

Hot Off the Tree (HOTT) (Excerpts and Abstracts of Articles
about Information Technology)
TELNET MELVYL.UCOP.EDU; Enter command: SHOW HOTT.
Further information: Susan Jurist, SJURIST@UCSD.EDU.

Information Networking News
Sent to CDROMLAN subscribers. See above.

IRLIST Digest (Information Retrieval List Digest)
Send the following message to LISTSERV@UCCVMA:
SUBSCRIBE IR-L First_Name Last_Name.
For a list of back issue files, send the message:
INDEX IR-L.

Issues in Science and Technology Librarianship
Send a subscription request to: ACRLSTS@HAL.UNM.EDU.

Journal of Academic Media Librarianship
Send the following message to LISTSERV@UBVM:
SUBSCRIBE MCJRNL First_Name Last_Name.

LIBRES: Library and Information Science Research Electronic
Journal*
Send the following message to LISTSERV@KENTVM:
SUBSCRIBE LIBRES First_Name Last_Name.

MeckJournal
Send a subscription request to MECKLER@TIGGER.JVNC.NET.

Network News
Send the following message to LISTSERV@NDSUVM1:
SUBSCRIBE NNEWS First_Name Last_Name.

Newsletter on Serials Pricing Issues*
Send the following message to LISTSERV@GIBBS.OIT.UNC.EDU:
SUBSCRIBE PRICES First_Name Last_Name.

Public-Access Computer Systems News
Sent to PACS-L and PACS-P subscribers. See above.

For a list of back issue files, send the following message
to LISTSERV@UHUPVM1: INDEX PACS-L.

The Public-Access Computer Systems Review
Sent to PACS-L and PACS-P subscribers. See above.
For a list of article files, send the following
message to LISTSERV@UHUPVM1: INDEX PACS-L.

Michael Strangelove has compiled a helpful directory of electronic serials. To retrieve this directory, send the following e-mail message to LISTSERV@UOTTAWA or LISTSERV@ACADVM1.UOTTAWA.CA:

GET EJOURNL1 DIRECTRY F=MAIL
GET EJOURNL2 DIRECTRY F=MAIL

For further information, contact Michael Strangelove: 441495@ACADVM1.UOTTAWA.CA.

- --

Revised 1/11/93

Projects

The Clearinghouse for Networked Information Discovery and Retrieval

Several user-friendly client-server tools have been developed within the last couple of years for locating and retrieving information residing on computers reachable over the Internet. Among them, the Wide Area Information Server (WAIS), the Internet Gopher and archie have become popular, and the WorldWide Web (WWW) is poised for popularity once suitable client software is developed and distributed. WAIS, archie and Gopher present overviews of where information of interest is likely to reside and then assist the user in locating specific information objects. WWW permits a user to thread a path through the network by selecting tagged hypertext items.

While focused on the particulars of the evolution of wide area information systems like WAIS, The Clearinghouse for Networked Information Discovery and Retrieval (CNIDR) will work closely with the developers of other tools toward achieving compatibility and consistency, and, to the extent possible, convergence of the tools. Specific activities will provide a central focus and forum for networked information discovery and retrieval (NIDR) tools and will minimize the divergence of individual implementations by providing a repository for collecting, evaluating, and distributing protocol-compliant releases and enhanced versions.

CNIDR participates in standards and policy associations such as the Internet Engineering Task Force and the Coalition for Networked Information to further consensus among developers and explore appropriate uses of networked information. CNIDR also actively promotes the use of networked information discovery and retrieval tools at many national and international conferences and workshops to inform and educate end users.

CNIDR is sponsored by a three-year grant from the National Science Foundation (NSF) and by the Center for Communications at MCNC in the Research Triangle Park, NC.

For more information about the Clearinghouse contact:

George Brett, Director
Electronic mail: ghb@concert.net Telephone: 919-248-1886
Jane Smith, Assistant Director
Electronic mail: jds@concert.net Telephone: 919-248-9213

Jim Fullton, Technical Manager
Electronic mail: fullton@concert.net Telephone: 919-248-9247
Fax: 919-248-1405

As some of you may know, the U.S. National Science Foundation has funded the creation of the Clearinghouse for Networked Information Discovery and Retrieval.

One of the responsibilities of CNIDR is to promote the use and development of tools for discovering and using networked information. We have been specifically charged with the task of providing an evolutionary path for the public-domain version of WAIS, the Wide Area Information Server. Therefore, we have been paying close attention to the remarks and “wish lists” provided by the network community, and are well into a project to try to fulfill those wishes.

We are in the early beta test stages of a new generation of WAIS-like systems that build on the heritage of WAIS as released by Thinking Machines, but with quite a bit of extensibility and new features. Specifically, and with the help of many others, we’re creating a Z39.50-92 (soon to be 93) server with full backwards compatibility with current WAIS clients and indexes, features such as boolean search, formatted searching on fielded text information, thesaurus based synonym substitution, stemming, and a search engine integration toolkit.

While the server maintains compatibility with current WAIS clients, it will also allow searches and retrievals from any Z39.50 client capable of recognizing the BIB-1 or INFO-1 attribute sets. Clients developed for this system will also be able to interact with library automation systems within the constraints of the attribute sets. We also hope to provide a tool to make various Z39.50 server maintenance tasks easier.

We hope to support the following Z39.50 features:

1) BIB-1 and dynamic INFO-1 ssearches

2) ES-1 element sets, which allow even more retrieval flexibility than the current WAIS server.

3) Access and resource control - essentially, this will allow for secure servers and servers that can accept billing information from the user.

4) EXPLAIN. Smart clients can use the EXPLAIN facility in conjunction with their preferred attribute set to provide customized search options for the user.

Basic search engine features (configurable): Synonym lookup; Stemming; Formatted field searches; Boolean search logic (to be added before release)

We have not yet begun to create the search engine toolkit as we think we need more user input, and the clients are rudimentary at best. We do hope to have the server running and ready for beta-test by February, although alpha code will be available much sooner. Clients will take somewhat longer as user interface development can be quite time consuming. The entire system will, of course, be freely available for use and modification by the network community.

On the client side, work is progressing on clients for the Macintosh and Unix systems, with a PC client in the wings.

So, when will this wonderful thing be ready? We actually have a server running that answers both WAIS and Z39.50-1992 queries on the same port, but returns results in a non-standard transfer syntax. We are in the process of changing this to support INFO-1 generic data records, and hope to have that completed soon.

We have had many comments from interested folks asking us to make these plans known to the rest of the network community, and at the risk of announcing vaporware, are doing so.

How can you help? We have already had quite a bit from many people, most notably John Kunze at the University of California at Berkeley and Clifford Lynch at UCOP. John created most of the server code and designed the INFO-1 attribute set as well as the ES-1 element set structures. I have made the server modifications to support free-text searching and WAIS compatibility, as well as making the necessary changes to the WAIS search engine. Members of Joan Gargano's staff at the University of California at Davis are now working on X Windows clients.

We are looking for folks willing to work on clients and run beta-test servers with their current WAIS information bases. Once our transfer syntax code is firmed up, trial servers will be made available for this purpose.

As technical manager for the Clearinghouse (as of December 1), I will help coordinate development tasks and incorporate any of your contributions intoreleases (with attribution, of course). We also have staff who will provide documentation and support for final releases.

Jim Fullton
Clearinghouse for Networked Information Discovery and Retrieval
Phone: (919)-248-9247 E-Mail: fullton@concert.net

Networked Information Discovery and Retrieval: Tools

archie
Search & retrieve anonymous FTP
Source: ftp.sura.net:/pub/archie
 or ftp.mcgill.ca:/pub/archie
Demo: archie.sura.net, login as archie
Mailing List: Comments, bug reports to: archie-l@cs.mcgill.ca
 To add or delete a site or correct site info:
 archie-admin@cs.mcgill.ca
Contact: Alan Emtage bajan@bunyip.com or bajan@cc.mcgill.ca
 Peter Deutsch peterd@bunyip.com or peterd@cc.mcgill.ca

gopher
Hierarchical browser for numerous operating systems
Source: boombox.micro.umn.edu:/pub/gopher
Demo: consultant.micro.umn.edu, login as gopher
Mailing list: Questions or comments: gopher@boombox.micro.umn.edu
 To join gopher-news mailing list, send e-mail to:
 gopher-news-request@boombox.micro.umn.edu
Usenet: comp.infosystems.gopher
Contact: please use gopher@boombox.micro.umn.edu to communicate
 with the gopher development team

hytelnet
Hypertext interface to telnet to selected sites for numerous operating
 systems
Source: access.usask.ca:/pub/hytelnet
Demo: access.usask.ca, login as hytelnet
Mailing list: Comments: scott@sklib.usask.ca
 To join the updates mailing list, send e-mail to:
 scott@sklib.usask.ca
Contact: Peter Scott scott@sklib.usask.ca
 Earl Fogel (unix and VMS software) fogel@sask.usask.ca

libraries
Menu-driven scripted telnet connections to library OPACS and other
 selected resources
Source: sonoma.edu: /pub/libs.sh
Demo: vax.sonoma.edu, login as OPAC
Contact: Mark Resmer Mark.Resmer@sonoma.edu

libtel

Menu-driven scripted telnet connections to library OPACS and other selected resources (UNC-CH enhancements added)
Source: ftp.oit.unc.edu:/pub/doc/libtel.unix
Demo: bbs.oit.unc.edu, login as bbs
Contact: Paul Jones Paul_Jones@unc.edu

prospero
Networked file system
Source: cs.washington.edu: /pub/prospero.tar.Z
Mailing list: For more info send e-mail to: info-prospero@isi.edu
Contact: Clifford Neuman bcn@isi.edu

WAIS
Relevance feedback search and retrieval
Source: quake.think.com:/wais (some clients and servers, references to others)
sunsite.unc.edu: /pub/wais (most of the available clients and servers)
Demo: quake.think.com, login as wais
Mailing list: Moderated, digested list - send e-mail to:
wais-discussion-request@think.com
Technical discussion, unmoderated - send e-mail to:
wais-talk-request@think.com
Usenet: comp.infosystems.wais
Contact: George Brett ghb@concert.net

World-Wide Web
Hypertext interface to internet resources (links embedded in docs)
Source: info.cern.ch:/pub/www
Demo: info.cern.ch, no login or password
Note: WWW documentation is updated dynamically, and is available through the demo
Mailing list: Bugs and suggestions: www-bug@info.cern.ch
Contact: Tim Berners-Lee timbl@nxoc01.cern.ch

Project Envision: A User-Centered Database from the Computer Science Literature

NSF Project IRI-9116991 9/15/91 - 9/15/94

Principal Investigators Edward A. Fox, Lenwood S. Heath, Deborah Hix

Department of Computer Science, Computing Center Virginia Polytechnic Institute and State University (VPI&SU or Virginia Tech) Blacksburg, VA 24061-0106 Common Telephone 703/231-6931 FAX 703/231-6075 (CS Dept.) Common Internet: PI@vtopus.cs.vt.edu for PI=fox,heath,hix Direct to E. A. Fox: 703/231-5113, fox@fox.cs.vt.edu

This project involves managing the computer science literature by developing and evaluating a hypermedia database and system. Human-computer interaction and information storage and retrieval research are being applied to develop a new paradigm of user-centered database development and access. Computer science literature provided by the Association for Computing Machinery is being transformed to better serve the needs of a wide spectrum of computing practitioners and researchers. Based on interviews with computer professionals and analysis of existing publications, ‘‘objects’’ used in computing are being identified so that a tailored object-oriented database including algorithms, algorithm descriptions, bibliographic records, full texts of articles, reviews, data, page images, audio and video tapes can be stored and appropriately linked for searching and browsing. New minimal perfect hashing and graph partitioning algorithms will provide efficient access to the large, distributed graph of objects.

Information Infrastructure Project — Science, Technology and Public Policy Program

John F. Kennedy School of Government
Harvard University

Request for Information:
Policy Research on Information Infrastructure

The Information Infrastructure Project is developing a research-in-progress database on policy issues related to the emerging digital information infrastructure, especially as embodied in evolving Internet/NREN. The database will cover:

- design, management, and commercialization of the Internet and the emerging National Research and Education Network
- diffusion of scientific and technical information and knowledge
- electronic publishing and scholarly communication
- international aspects of the Internet
- national initiatives in internetworking outside the U.S.
- computer-based collaboration and cooperative work, especially across organizational boundaries
- pricing of data networks and digital information
- intellectual property
- commercial and industrial internetworking (concurrent engineering, CALS, EDI)
- access to and dissemination of government information
- network security and privacy issues
- common carrier issues and regulation of enhanced services
- network liability and First Amendment issues

- transformation of libraries and other information institutions
- public funding and procurement of elements of information infrastructure
- application of antitrust law to information resources
- economics of networked information environments
- Internet institutions and standards development processes

A major objective in developing this database is to identify potential presenters and authors for future workshops and publications. We are especially seeking rigorous work presented in a concise, lucid manner that enables policy-makers (who are typically not computer scientists, lawyers, economists, sociologists, etc.) to better understand the issues.

We plan to make the database publicly available and will consider submissions as permission to publish, unless we are notified to the contrary. We reserve the right not to list contributions and to edit those that we use. Listed contributors will have special access to early versions of the database.

If you wish to participate, please send a brief abstract of your project. You may send materials in any form, but we need the abstract if your project is to appear in the published version. Please use the following template, if possible, but we will accept submissions in your own format:

Project Title:
Institution:
Collaborating Institutions:
Principal Investigator(s):
Project Director(s):
Other Contact Person:
Address:
Address:
City:
State or Province:
Zip (Postal Code):
Country:

Telephone:
Fax:
Email:
Funding Source (optional):
Abstract:

Please send project information to:

gowlland@husc8.harvard.edu (gowlland@husc8.bitnet)

Or by U.S. Mail to:

Brian Kahin
Director, Information Infrastructure Project
Science, Technology and Public Policy Program
John F. Kennedy School of Government, Harvard University
79 John F. Kennedy St.
Cambridge, MA 02138
[kahin@hulaw1.harvard.edu
617-495-8903
Fax: 617-495-5776]

Information Infrastructure Sourcebook

The Information Infrastructure Project at Harvard's Kennedy School of Government announces the publication of the Information Infrastructure Sourcebook. The Sourcebook is designed to provide planners and policymakers a single volume reference on efforts to define and develop policy for a national information infrastructure. It includes historical policy documents, private sector vision statements and position papers, program and project descriptions (all sectors), landmark reports and pending legislation.

In assembling the Sourcebook, we have looked for documents that have had or are likely to have an impact on policy development, that are formal in nature, and that deal with information infrastructure at a general rather than topic level. In general, material is reproduced as is, although we have excerpted in the interests of space, relevance, balance, and consistency. Because of changes in technology, markets, programs and policies, the Sourcebook will of necessity require supplementation or revision on a regular basis. So we will look for new or updated contributions whenever they become available. Please direct any suggestions for additional material to:

James Keller
Project Coordinator
Information Infrastructure Project
Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
tel: (617) 496-4042
e-mail: kellerj@ksg1.harvard.edu

The Sourcebook is available for $40, including Priority Mail postage and handling. If you would like to receive a copy of the Sourcebook, please send a check for $40 payable to Harvard University to:

Graceann Todaro
Center for Science and International Affairs
Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138

The Table of Contents from the Sourcebook is provided below:

TABLE OF CONTENTS

Information Infrastructure Sourcebook

Official Documents

Federal Coordinating Council for Science, Engineering and Technology, Office of Science and Technology Policy, Executive Office of the President, FCCSET Initiatives in the FY 1994 Budget, Washington, DC, April 8, 1993. (Introduction and excerpts)

President William J. Clinton & Vice President Albert Gore, Jr., Technology for

America's Economic Growth, A New Direction to Build Economic Strength, White House, Washington, DC, February 22, 1993. (excerpts)

Allan D. Bromley, The National Research and Education Network Program: A Report to Congress, OSTP, Washington, DC, December, 1992. (Table of Contents and Executive Summary)

Committee on Physical, Mathematical and Engineering Sciences, Federal Coordinating Council for Science, Engineering and Technology, Office of Science and Technology Policy, Grand Challenges 1993: High Performance Computing and Communications, a Supplement to the President's Fiscal Year 1993 Budget, Washington, DC, 1992.

Committee on Physical, Mathematical and Engineering Sciences, Federal Coordinating Council for Science, Engineering and Technology, Office of Science and Technology Policy, Grand Challenges: High Performance Computing and Communications, a Supplement to the President's Fiscal Year 1992 Budget, Washington, DC, 1991.

U.S. Congress, Public Law 102-94 - High Performance Computing Act of 1991, Washington, DC, December 9, 1991.

U.S. Congress, Senate Committee on Commerce, Science and Transportation - Report 102-57, High Performance Computing Act of 1991, U.S. Government Printing Office, Washington, DC, May 16, 1991.

Executive Office of the President, Office of Science and Technology Policy, The Federal High Performance Computing Program, OSTP, Washington, DC, September 8, 1989. (Executive Summary)

Office of Science and Technology Policy, Executive Office of the President, A Research and Development Strategy for High Performance Computing, Washington, DC, November, 1987.

Vision Statements and Position Papers

CALS Vision: Economic Growth Through Worldwide Enterprise Integration, CALS Industry Steering Group, National Security Industrial Association, undated.

Cable Television Laboratories, Inc., Cable's Role in the "Information Superhighway", Boulder, CO, 1993.

Computer Systems Policy Project, Perspectives on the National Information Infrastructure: CSPP's Vision and Recommendations for Action, Washington, DC, 1993.

Council on Competitiveness, Vision for a 21st Century Information Infrastructure, Washington, DC, May, 1993.

Electronic Frontier Foundation, The Open Platform: A Proposal by the Electronic Frontier Foundation for a National Telecommunications Infrastructure, Cambridge, MA & Washington, DC, 1992.

Information Industry Association, Telecommunications Infrastructure Objectives and Implementation Principles, Washington, DC, April, 1993.

Mary Gardiner Jones, The Consumer Interest in Telecommunications Infrastructure Modernization, Consumer Interest Research Institute, January 21-24, 1990.

Robert Kahn, National Information Infrastructure Components, Serials Review, Spring and Summer, 1992.

National Cable Television Association, Cable Television and America's Telecommunications Infrastructure, Washington, DC, 1993.

Telecommunications industry CEOs, Policy Statement on NII, March 24, 1993.

Program & Project Descriptions - all sectors

Robert Aiken, Hans Werner Braun, Peter Ford & Kimberly Claffy, NSF Implementation Plan for Interagency Interim NREN, National Science Foundation, May 1, 1992.

The CENDI Group, an overview, undated.

EINet - Stepping into the Electronic Marketplace: a Technical Overview, Microelectronics and Computer Technology Corporation, undated.

Federal Research Internet Coordinating Committee, Department of Energy, Program Plan for the National Research and Education Network, Washington, DC, May 16, 1989.

First Cities - Interactive Multimedia Information: Where, When and How You Want It, Microelectronics and Computer Technology Corporation, undated.

The National Initiative for Product Data Exchange, NIST, an overview, November, 1992.

The National Science Foundation, Program Solicitation: Network Access Point Manager, Routing Arbiter, Regional Network Providers, and Very High Speed Backbone Network Services Provider for NSFNet and the NREN Program, Washington, DC, National Science Foundation, DC, May 6, 1993.

Technology Reinvestment Project, Program Information Package for Defense Technology Conversion, Reinvestment and Transition Assistance, Arlington, VA, March 10, 1993. (Introduction and excerpts)

Reports

Computer Science and Technology Board, Commission on Physical Sciences, Mathematics and Resources, National Research Council, The National Challenge in Computer Science and Technology, National Academy Press, Washington, DC, 1988. (Chapter 3: The Promise of Infrastructure)

Computer Science and Telecommunications Board, National Research Council, National Collaboratories, National Academy Press, Washington, DC, 1993. (Executive summary and Chapter 1)

“The Future Federal Information Infrastructure: The Responsibility of the Federal Information Resources Management Community”, excerpted from the Information Resources Plan of the Federal Government, Office of Management and Budget, November, 1992.

National Research and Education Network Review Committee of the Computer Science and Technology Board, Commission on Physical Sciences, Mathematics and Resources, National Research Council, Toward a National Research Network, National Academy Press, Washington, DC, 1988. (Executive Summary)

National Telecommunications and Information Administration, The NTIA Infrastructure Report - Telecommunications in the Age of Information, Executive Summary, U.S. Department of Commerce, October, 1991.

Office of Technology Assessment, U.S. Congress, Critical Connections: Communication for the Future, U.S. Government Printing Office, Washington, DC, January, 1990. (Chapter 1 - Summary)

Office of Technology Assessment, U.S. Congress, Global Standards - Building Blocks for the Future, U.S. Government Printing Office, Washington, DC, March, 1992. (p. 26-30)

Office of Technology Assessment, U.S. Congress, Helping America Compete - The Role of Federal Scientific and Technical Information, U.S. Government Printing Office, Washington, DC, July, 1990. (Chapter 1 - Summary)

Office of Technology Assessment, U.S. Congress, Informing the Nation, U.S. Government Printing Office, Washington, DC, October, 1988. (Chapter 1 - Summary)

Panel on Information Technology and the Conduct of Research, National Academy of Sciences, Information Technology and the Conduct of Research: The User's View, National Academy Press, Washington, DC, 1989. (Executive Summary)

Proceedings of the NREN Workshop, Monterey, California, September 16-18, 1992 (excluding appendices), Computing Research Association/EDUCOM/IEEE U.S. Activities Board, with assistance from the National Science Foundation Grant NCR-921671.

Steven R. Rivkin and Jeremy D. Rosner, Shortcut to the Information Superhighway: A Progressive Plan to Speed the Telecommunications Revolution, Progressive Policy Institute Policy Report No. 15, July 1992. (Introduction and Part four: a Progressive Strategy to Build the Fiber-Optic Network)

Proposed Legislation

U.S. Congress, Communications Competitiveness and Infrastructure Modernization Act of 1993 - H.R. 1504, Washington, DC, March 29, 1993.

U.S. Congress, To Establish a System of State-based Electronic Libraries - S.626, Washington, DC, March 22, 1993.

U.S. Congress, High Performance Computing and High Speed Networking

Applications Act of 1993 - H.R. 1757, Washington, DC, April 21, 1993.

U.S. Congress, Local Exchange Infrastructure Modernization Act of 1993 - H.R. 1312, Washington, DC, March 11, 1993.

U.S. Congress, National Competitiveness Act of 1993 - S.4, Washington, DC, January 21, 1993.

Interactive Accessibility: Breaking Barriers to the Power of Computing

Virginia Tech Research Project Announcement National Science Foundation (NSF) Institutional Infrastructure Award

Co-Principal Investigators
Dr. Roger W. Ehrich, Computer Science
Dr. Edward A. Fox, Computer Science
Dr. H. Rex Hartson, Computer Science
Dr. Deborah S. Hix, Computer Science
Dr. Robert C. Williges, Industrial and Systems Engineering

1. Project Description
The major goal of this interdiciplinary research activity is to improve interactive computing by removing two major barriers people face when using computer systems. First, usability barriers exist because the human-computer interface (HCI) is often poorly designed. Second, conceptual barriers exist because information access systems are poorly integrated. A consulting service and three inter-related research laboratories which are unique to HCI activities will be established under this grant.

THE INTERACTION CONSULTING SERVICE (DIRECTOR: DR. EHRICH) will be developed primarily as an HCI outreach program and will provide practicum experience for graduate students. This service will help the two departments, the university, and the industrial community to learn, adapt, and apply the newest HCI design, development, and evaluation technologies. This service will also help clients implement designs on diverse computing platforms. The Interaction Consulting Service provides new opportunities for practical application of HCI knowledge and skills as well as new sources of data regarding usability methods research.

THE INFORMATION ACCESS LABORATORY (DIRECTOR: DR. FOX) is an applications laboratory for studying multi-media information access for research and education. This lab will provide high volume storage, document capture, electronic publishing, imaging, and presentation services. The aim is to facilitate prototyping of new types of interactive information access systems that test our theories with real-life, realistic size collections.

THE INTERACTION TECHNOLOGY LABORATORY (DIRECTORS: DRS. HARTSON AND HIX) will house state-of-the-art equipment relating to HCI. A commitment to research in interactive systems necessitates a demonstration laboratory with representative up-to-the minute technology in computing platforms, operating systems, networked client-server technology, interactive applications software, and user interface development tools.

THE USABILITY METHODS RESEARCH LABORATORY (DIRECTOR: DR. WILLIGES) will extend and augment the existing HCI facilities in Industrial and Systems Engineering. This combined facility will be networked and will serve as the core of interdiciplinary research in HCI. An electronic conference area with telecommunications capabilities and two usability evaluation rooms will be built aroud a highly instrumented central control room containing video and computing equipment for conducting interactive experiments. Research in this laboratory will develop cost-effective and efficient methods for conducting formative and summative usability
evaluation.

2. Importance to Virginia Tech and the Commonwealth of Virginia
A. Provides advanced computer equipment and laboratory facilities.
B. Improves undergraduate and graduate education.
C. Facilitates interdiciplinary undergraduate and graduate research.
D. Encourages industrial and other outside research support.
E. Compliments other major interactive computing activities on campus.
F. Supports University educational, researcg, and service objectives.
G. Serves National needs in advancing information technology.

3. Level of effort
A. Five year duration (July 1993 - June 1998)
B. Funding for equipment, maintenance, facility renovations, and personnel
C. Project Total: $2,279,810 (NSF - $1,375,000 and Virginia Tech - $904,810)

Project JANUS - Large-scale Imaging and Text-Retrieval Project

Introduction

JANUS is a five-year project of Columbia Law Library to develop an electronic, or 'virtual', library that combines digital conversion and storage of document text and graphics, massively parallel supercomputing, and advanced user-friendly search and retrieval software. When completed in 1996, JANUS will be a 'library of the future' and provide unprecedented electronic access nationwide to large-scale databases comprised of full text, graphics, images, data, sound and video materials. JANUS is a joint venture of Thinking Machines Corporation (TMC) of Cambridge MA, the world's largest manufacturer of highly parallel supercomputers, and Columbia Law Library, the nation's third largest collection of legal materials. JANUS uses the world's leading parallel supercomputer, The Connection Machine System (CM-2)*, on loan from TMC and installed at the Law School Library. Columbia and TMC are jointly developing retrieval software for use on large scale databases.

A 'Virtual' Library

Advances in optical scanning allow Project JANUS to convert and store in digital form an exact copy of a document page by page. Project JANUS is developing software that correctly interprets text that may have been corrupted during the optical character recognition process. By 1996, Columbia Law Library expects to convert about 10,000 to 12,000 volumes a year to electronic storage and expects a continued increase to its collection of new documents available only in electronic form.

JANUS' optical scanning process also converts and stores an exact copy of the document into a file that can be searched and manipulated by software developed by Project JANUS. JANUS' unique and powerful ability to search for and retrieve specific information from large scale text, graphics, image, sound and video databases best demonstrates the 'library of the future.' From any computer terminal connected to the JANUS system, a user seeking specific information on a subject can type in a word, phrase or sentence in ordinary English and receive in seconds specific references listed in order of relevance. The user has access to the exact copy of complete text at the position most relevant to the query and to relevant images and sounds. The user can save retrieved items or sections of items as needed. The user can continue searching for additional references by using new words or phrases or by highlighting words, phrases or sections of text already retrieved and again receive in seconds more references listed in orders of relevance.

JANUS' Database Software

Project JANUS accomplishes free-text searches by using a non-Boolean natural language algorithm that incorporates frequency-based indexing, relational value weighting, and relevance feedback from the individual user. Using the immense "number-crunching" power of parallel computing and building on TMC's work with the Wide Area Information Server (WAIS), JANUS's unique retrieval system, WAISSeeker, allows users to search the full text of documents for arbitrary combinations of words and, unlike keyword systems, does not require users to have prior knowledge of the discipline they are searching.

JANUS is also developing a Graphical User Interface (GUI) to make all processes easy to use. In addition to being clearly instructed during the feedback and the information retrieval phases of the process, system users will also be able to incorporate retrieved products into their word processing work products.

JANUS' development of WAISSeeker will provide a storage, search and retrieval system of unprecedented power, flexibility and accuracy. Both researchers and the general user will have easy access to large scale databases of complete text, images and sounds and will retrieve items most relevant to individual user queries.

Interconnection, Standards and Copyright Considerations

JANUS will first serve scholars on Columbia's campus. Because JANUS uses the Z39.50 standard as its communication protocol in its WAIS interface, it will be accessible when operational from any remote computer using a WAIS server and, via Internet, can serve users nationally and internationally. Development of a large bandwidth network channel, such as proposed in development of a National Research and Education Network, would allow a large number of users to browse and work in the Columbia Law Library from any NREN connection in the nation.

JANUS's database control system is designed to provide a datafile tagging and logging structure to protect copyrights and intellectual property. Columbia Law School is collaborating with Simon and Shuster to set standards for copyright protections applicable to wide-scale electronic retrieval and dissemination of information.

Project JANUS 1991-1996: "The Library of the Future"

Project JANUS is nearing completion of Phase One by developing and testing its natural language algorithm and extensively testing the system's ability to retrieve corrupted text. Phase Two will see a limited use production system available to researchers at the Columbia Law Library and via the Internet. JANUS will acquire large-scale databases and develop procedures for supporting multiple users concurrently using WAISSeeker.

In its final phase, JANUS will become a full production system with staffing for the Law School 'virtual' library in place. JANUS will have established relations with publishers to allow for use of copyrighted materials directly in electronic form and will have developed programs to track and verify use of licensed materials electronically.

*Connection Machine, CM-2 and Thinking Machines are trademarks of Thinking Machines Corporation. The CM-2 is equipped with 32,000 processors and has 256 megabytes of main memory and 20 gigabytes (billions of bytes) of hard disk storage.

The Mercury Electronic Library

William Y. Arms
December 16, 1992

Carnegie Mellon has a long term project known as Mercury, to build a large-scale digital library. Mercury began in late 1988/89 and became the Libraries' production system in 1991/92. Today, Mercury is probably the most complete distributed library in the United States and is the basis for a vigorous program of research and development.

One part of Mercury to create a large collection of managed information on-line.

The Libraries are a key contributor, but Mercury is designed to support databases that are managed by independent units, such as departments and research groups, and to interoperate with similar systems elsewhere. Mercury works closely with publishers since the economic and legal issues are a key to the development of a digital library.

The software architecture uses standard protocols and formats. Version 1 uses TCP/IP networking; Kerberos authentication; Z39.58 search syntax; Z39.50 (1988) retrieval protocol; ASN.1 encoding; documents stored as ASCII text or bit-mapped images with Fax G4 compression. It has user interfaces for Unix computers with the Motif window manager and VT100 terminal emulation. A Macintosh user interface is under development. The retrieval protocols have been implemented for OCLC's Newton database system and are being imported to the Fulcrum database software.

Special strengths of the software are the distributed computing and the support for images over networks. All the software building blocks are non-proprietary except the database systems. During the fall, Carnegie Mellon will provide an unsupported distribution of the software. Digital is planning a commercially supported version.

The Mercury plan combines both research and development. The following are the principal components of the current plan: intelligent search and retrieval using methods of computational linguistics and concept based retrieval; personal organization and structuring of information by establishing associative relationships between information objects; tools and algorithms for the capture, storage, manipulation, and recognition of high resolution images over networks; scalability; and interoperation with systems that are developed and managed by others.

NSF Network Information Services Awards

For Immediate Release
January 5, 1993

In cooperation with the Internet community, the National Science Foundation developed and released, in the spring of 1992, Project Solicitation NSF92-24 for one or more Network Information Services Managers (NIS Manager(s)) to provide and/or coordinate (i) Registration Services, (ii) Directory and Database Services, and (iii) Information Services for the NSFNET. As a result of this solicitation, three separate organizations were competitively selected to receive cooperative agreements totalling over $12 million in the three areas of (i) Registration Services, (ii) Directory and Database Services, and (iii) Information Services. Together, these three awards constitute the NIS Manager(s) Project, named the INTERNIC. Network Solutions will provide registration services, AT&T will provide directory and database services, and General Atomics will provide information services. It is important that the three project participants work closely together to provide a seamless interface for users in need of services. For this reason, the three awardees, at the request of the Foundation, have developed a detailed concept and plan to provide this seamless interface called the "INTERNIC" and have agreed to the structuring of their three separate awards as one collaborative project.

Steve Wolff, Director of NSF's Division of Networking and Communications Research and Infrastructure says, "We all feel intuitively that the domestic Internet and the distributed collaboration that it facilitates are rapidly creating a national "workplace without walls". These three awards to geographically dispersed organizations for Network Information Services, which require a high degree of coordination and collaboration, will both exploit and demonstrate the success of the network in enabling such distributed collaboration."

Consistent with FNC guidelines on obtaining reasonable cost recovery from users of NREN networks, the NSF has determined that the INTERNIC Information Services provider may charge users beyond the U.S. research and education community for any services provided. Also, the INTERNIC Directory and Database Services provider may charge a fee for maintenance of special databases, for extensive directory listings and may charge users beyond the U.S.research and education community. Finally, because the registration function provided by the INTERNIC Registration Services applies to domestic and international, commercial and individual users in addition to research and educational users, it is expected that an appropriate registration fee structure will take time to develop. NSF expects to engage in an extensive discussion with the domestic and international Internet community of the motivation, strategy and tactics of imposing fees for these services during the next fifteen months. Decisions will be implemented only after they have been announced in advance and an opportunity given for additional public comment.

Network Solutions will provide registration services as the IP registrar, issue IP numbers worldwide using delegated registries under the guidance of the Internet Assigned Numbers Authority and also register domain names, and track points of contact. Applications for assignment will be accepted via email or facsimile. The information from these assignments will be provided to the directory and database services provider to be made available to the entire Internet community. As a part of the Domain registration efforts Network Solutions will periodically release the top level zone files to be used by all root Domain Name servers.

AT&T will develop and maintain a Directory of Directories, including lists of FTP (File Transfer Protocol) sites, lists of various types of servers available on the Internet, lists of white and yellow page directories, library catalogs and data archives. AT&T will also provide white and yellow pages type Directory Services. Access to these services will initially be provided through several currently popular in-use interface methods while migrating to the use of X.500 technology, the current standard specification for distributed information storage and retrieval. The database services which AT&T will provide include the establishment of Database Services to extend and supplement the resources of the NSFNET, such as databases of contributed materials of common interest to the user community. AT&T will also offer database design, management, and maintenance to institutions and groups for inclusion in the Internet.

General Atomics will provide Information Services acting as the NIC of first and last resort and the NIC of NICs. The INTERNIC information services will include a full-service Reference Desk, a database of comprehensive networking materials called the Info Source, training classes and documentation, and coordination services among all appropriate groups in the community. In keeping with the innovative spirit of the Internet, several new approaches to distributing services will be implemented. Among these innovations is NICLink, a user-friendly hypermedia interface offering access to the Info Source and all the information it contains. NICLink will be distributed on both standard computer diskettes and CD-ROM. Another is the concept of the Info Scout, an individual who will scout out new resources and innovative uses of the network for inclusion in the Info Source.

Network Solutions is a 400-person telecommunications analysis and integration company headquartered in Northern Virginia. Its mission is to support its customers in achieving their missions through the mastery and application of networking technology. Network Solutions currently operates the DDN NIC.

AT&T is the leading provider of global information movement and management products and services. AT&T offers a wide array of data communications services that includes private line, X.25, frame relay, TCP/IP, protocol conversion, and electronic mail services.

General Atomics is a San Diego-based high-technology research and development company and operates CERFnet and the San Diego Supercomputer Center. CERFnet is an Internet network service provider that operates throughout the state of California and nationally. CERFnet was launched in the spring of 1989 with a

$2.8 million grant from the National Science Foundation. The San Diego Supercomputer Center is a five year cooperative agreement funded by the National Science Foundation to support high performance computing.

National Science Foundation Contact
Don Mitchell
(202) 357-9717
dmitchel@nsf.gov

Network Solutions Contact
Mary Bloch
(703) 742-4740
maryb@netsol.com

AT&T Contact
Shelly London
(908) 221-4355
london@attmail.com

General Atomics Contact
Susan Calcari
(619) 455-3900
calcaris@cerf.net

The Open Source Processing Research Initiative

MITRE Corporation 7525 Colshire Dr, McLean, VA 22102

The United States continues to face a rapidly changing world. Issues formerly given relatively low priority, such as international trade and Third World development, are now given much more weight by present government policy makers. Additionally, the Soviet Union's dissolution has not decreased the need for information about that region instead, government officials require detailed information about the former Soviet republics and their actions. Fewer people covering a growing set of issues creates a strong role for Open Source Information, a relatively low-cost and low-risk source of breaking news, reference materials, and scientific developments.

The MITRE Corporation, recognizing the important role that Open Source material will play in the future of the IC, has created the Open Source Processing Research Initiative (OSPRI). The goal of OSPRI is to address the broad spectrum of technical issues that the government faces in attempting to take advantage of Open Source material. These area's need to be addressed due to the large volume of information from multiple sources, in multiple formats, and with a high degree of duplication. Users will quickly grow frustrated with Open Source analysis if the information cannot be provided quickly, with a high degree of relevancy, in an intuitive manner. MITRE sees these issues as a microcosm of document management and information retrieval problems faced in other domains. Therefore, the technologies developed will be applicable across the general problem area.

The underlying concept behind OSPRI is that current information retrieval techniques place a significant burden on the user to develop sophisticated queries and dissemination profiles to locate relevant documents. The OSPRI research team will attempt to relieve this burden by focusing on providing filtering assistants that will generate knowledge about the document, aiding in information retrieval. A simplistic example of this would be to have filtering assistants that understand the part-of-speech for the words in a document and at query execution ensure that a match is in the appropriate part-of-speech so that a search on the word 'cooperative' as a noun would retrieve documents where 'cooperative' occurred in the 'private cooperative' sense. These filtering assistants would encode the "meta data" (data about the data) into the document for use by the query processor. The system would be extensible to any type of filtering assistant one could implement and assistants could use the work of previous assistants, allowing the system to grow with the technology. Examples of other assistants could include ones that locate geographical references and tag the latitude and longitude for use in displaying documents on maps, identifying important individuals, or assignment of topics to components of documents. These would allow the user to make queries such as a request to show all documents that refer to locations in some part of the world (which could be done by gesturing to a map). The OSPRI is developing a proof-of-concept system which includes a range of filtering assistants, user tools, and an information schema.

The OSPRI team's research efforts to develop filtering assistants include methods such as identifying items of interest referenced directly by proper nouns or indirectly by description or allusion. Parallel research is working toward identifying other grammatical elements of sentences, with the purpose of extending these techniques to identify the overall meaning of a body of text.

The OSPRI research staff has constructed a prototype system that demonstrates the feasibility of implementing this approach. The demonstration system developed was intended to prove the feasibility of a number of concepts: using Standard Generalized Markup Language (SGML) as the encoding mechanism, the importance of both the syntactic and semantic structure of a text, and improving information retrieval over standard keyword or boolean techniques. The data set for the demonstration system consists of approximately 350 articles from the 1989 Wall Street Journal (WSJ) and four closed caption transcripts from the MacNeil-Lehrer News Hour and ABC World News Tonight. The OSPRI staff manually examined all of the WSJ articles and video transcripts for these concepts and added appropriate semantic tags where the concepts appeared. The data set contained a non-trivial number of instances of these concepts that could not be effectively retrieved by a keyword search alone.

To provide a baseline to compare enhanced information retrieval techniques from tagged text, the staff conducted an empirical study on the prototype system's data set. Specifically, the staff compared the results of searches using unmodified keyword search terms (such as 'cooperative'), modified keyword terms ('private cooperative'), and keyword modified by syntactic category ('cooperative' as a noun). The study results demonstrate that even rudimentary natural language processing (NLP) techniques available today can significantly improve the performance of keyword search systems. As the OSPRI team brings increasingly sophisticated computational linguistic techniques to bear on information retrieval, it is expected that researchers will demonstrate substantial improvement over conventional methods.

The OSPRI is an umbrella program under which several projects are grouped. Each is described below:

Open Source Research - R. D'Amore

The objective is to extend the state-of-the-art in information retrieval by improving text analysis, mark up, and search tools. The program encompasses research that will extend automated Natural Language Processing (NLP) techniques such as logical representation of semantic meaning, proper noun identification, and categorization. In this project, output from each of these tools will be encoded in the text of the processed document as semantic and syntactic SGML tags, greatly improving current information retrieval, data base generation, indexing, and intelligence dissemination techniques.

Language Processing for Intelligence Applications - M. Vilain

The objective is to provide a working natural language understanding system. The scope of our research activity towards this objective includes two major areas: Robust language processing techniques that provide, in particular, for graceful error recovery in the face of unanticipated linguistic phenomena and self extensible knowledge bases these are necessary to break the knowledge acquisition bottleneck that prevents current language systems from applying in many domains.

Message Content Analysis - E. Lusher

The objective of this research is to develop and test the utility of two methodologies: a first filter that performs broad subject classification, and a second filter that identifies documents relevant to a user request.

Contextual Analysis to Improve Optical Character Recognition Performance - D. D'Amato

The objective is to improve optical character recognition (OCR) system accuracy by developing algorithms to: combine the outputs of two or more OCR systems, use system-specific information, such as character confusion probabilities, and perform domain-specific contextual processing.

Machine Translation of Spoken Dialogues - S. Luperfoy

The objective is to advance the state-of-the-art in dialogue processing by creating discourse solutions to unsolved problems in machine translations, user interface design, and speech recognition.

Open Source Broadcast Subject Classification - T. Albina

The objective is to reduce the amount of speech data to be processed manually by automatically grouping data into topic clusters that can be analyzed more rapidly.

The final product of OSPRI at the end of the year is an Advanced Technology Demonstration of an Open Source System. This demonstration will highlight each of the research areas with a focus on the leverage one obtains by combining the capabilities.

MITRE has been evaluating and selecting sources of material for inclusion in the OSPRI testbed. The primary sources of Open Source Information currently available to MITRE and the government are commercial databases, compact discs-read only memory (CD ROMs), government publishers such as the Foreign Broadcast

Information Service (FBIS), and free or inexpensive academic sources of text. One commercial source of Open Source Information is DIALOG, a comprehensive database system that includes the full text of popular newspapers, as well as journals from many disciplines. More technical material is available through the Defense Research, Development, Testing, and Evaluation (RDT&E) On-line System (DROLS). This database system, developed by the Defense Technical Information Center (DTIC), is maintained to provide on-line access to the DTIC data collection on worldwide research and development, technology advances, and energy acquisition and exploitation. Other databases cover legal, business, and U.S. Federal Government information. Of the other commercial data resources, CD-ROMs, which can provide up to 500 megabytes (MBs) of data per CD-ROM at a price often under $250, are the most cost effective.

Future technical areas to be addressed and improved include: refining the OSPRI semantic analysis algorithms, pre-filtering the mass of available Open Source data to produce a manageable and relevant subset for each user group, integrating multimedia sources and displays, examining data security and tag encryption, and continuing work on a graphical user interface that takes advantage of the increased functionality made possible by the OSPRI NLP techniques. William (Bill) A. Ruh, war@mitre.org, The MITRE Corporation, Workstation System Engineering Center Phone - (703) 883-6529 FAX - (703) 883-3315

Project on Scholarly Communication in the Network Environment

Announcement of Computer Conferences

The Science, Technology and Public Policy Program at Harvard's John F. Kennedy School of Government and the Coalition for Networked Information have undertaken a project to address principles, policy, and practice related to new forms of scholarly communication in the network environment. Funded in part by the National Science Foundation's Program on Ethics and Values Studies in Science and Technology, the Project looks at a set of social, ethical, and legal issues raised in the communication and dissemination of research results. The goal is to help different groups ranging from research teams to academic and professional societies develop appropriate policies and practices.

A draft background paper has been prepared by project director Brian Kahin and is available on the Internet by anonymous ftp from ftp.cni.org using the following sequence of commands:

```
ftp ftp.cni.org
login anonymous
[send e-mail address as password]
cd /CNI/projects/Harvard.scp
get background.txt
exit
```

We are now ready to initiate a set of computer conferences which will bring together diverse disciplinary and service perspectives. The conferences will address six issue areas which are discussed sequentially in the background paper:

1.) JOINT AUTHORSHIP AND OWNERSHIP -- How should jointly authored research be structured and how should publication processes be handled?
list name: OWNERSHIP
moderator: Michael Strait (mstrait@linknet.com)

2.) RIGHTS IN COMPUTER CONFERENCING -- What are reasonable expectations for the handling and reuse of messages and other material posted to groups and mailing lists?
list name: REPOST
moderators: Edward Vielmetti (emv@msen.com)
Steve Cisler (sac@apple.com)

3.) DERIVATIVE AND ITERATIVE WORKS -- What practices should apply to sequenced and variant publications of the same and related work?
list name: DERIV
moderators: Ann Okerson (okerson@umdc.bitnet)
Steven Zink (stevenz@equinox.unr.edu)

4.) CONTROL OF DISSEMINATION -- To what extent should key scholarly resources be controlled by particular scholars or organizations?
list name: RESOURCES
moderator: Doug Greenberg (sdgls@cunyvm.bitnet)

5.) SITE LICENSING -- How will widespread site licensing affect access to information by unaffiliated individuals and small firms and organizations?
list name: SITE-LICENSE
moderators: John Garrett (jgarrett@nri.reston.va.us)
Steve Gilbert (gilbert@educom.edu)

6.) INTERNATIONAL ACCESS -- How should researchers and practitioners in the developing world be assured access to research results?
list name: INTERNATIONAL
moderator: Art St. George (stgeorge@bootes.unm.edu)

Persons interested in participating in one of these conferences should send a single line mail message to
LISTSERV@CNI.ORG as follows:

subscribe [name of list] [your first name] [your last name]

If you wish to participate in more than one conference, put each request on a separate line in the same format. Please also send the moderators background information so they will know who is participating. The moderators will set their own policies and some may choose limit the size of their conference.

We plan to report on these conferences at the November meeting of the Coalition for Networked Information Task Force. There will be an invitational workshop for the most active participants and representative academic organizations in Washington in early 1993. A final report is due in the Spring.

For additional information, contact the list moderators at the listed email addresses. Or:

Brian Kahin
Director, Information Infrastructure Project
Science, Technology and Public Policy Program
John F. Kennedy School of Government, Harvard University
79 John F. Kennedy St.
Cambridge, MA 02138
617-495-8903
Fax: 617-495-5776
kahin@hulaw1.harvard.edu

Paul Peters
Director
Coalition for Networked Information
1527 New Hampshire Ave., NW
Washington DC 20036

202-232-2466
Fax: 202-462-7849
paul@cni.org

For technical assistance:

Craig A. Summerhill, Systems Coordinator
Coalition for Networked Information
1527 New Hampshire Ave., N.W.
Washington, DC 20036
craig@cni.org

Soviet Archive Exhibit

From: usmarc-l@RGUE.LOC.GOV

On June 17th, the Library of Congress will open an exhibit in the James Madison Memorial building of approximately 300 historically significant documents, photographs, and film clips from the Communist Party of the Soviet Union, the KGB, and other Soviet archives.

In addition, the Library will also digitize significant portions of about twenty - five documents, with translations, and distribute them as a national on-line exhibit that begins on the same day as the exhibit opening. The on-line exhibit will contain accompanying text, translations, captions, and graphic images for each of the documents included. The documents are available for personal study by scholars and students, and not for publication. Other exhibit information presented will be available for copying and subsequent use.

The on-line exhibit will be available in three services, all of which will be available on June 17, 1992. America On-line (AOL), a nationally-recognized commercial service, will provide one distribution network for the exhibit, Internet will provide another, and Sovset, an international computer network for Russian and East European studies will provide the third.

Through the use of text, message boards, libraries, "conference rooms," polls and surveys, AOL will assist the Library of Congress in its endeavor to make a major exhibit available to the public through the technology of on-line services. AOL allows for real time communications and the ability to engage in E-mail discussion of the documents. AOL's "The Library of Congress On-line" will mark the first time that a true nationwide dialog on a landmark exhibit has been made possible, and the first time that a simultaneous electronic surrogate exhibition has been presented by a museum or library.

On Thursday June 18th from 1-3pm, Dr. James H. Billington, the Librarian of Congress, Dr. Rudolph G. Pikhoia, chairman of the Committee on Archival Affairs of the Russian Federation, and General Dmitrii Volkogonov, head of the KGB Archive, will participate in an on-line computer forum on America On-line. The Library invites scholars, educators, students, and the general public across the United States to join this discussion of the exhibit and the opening of the Soviet Archives. America On-line will provide free membership software and 5 hours of on-line use without charge upon request for those who are not now members.

Those with access to Internet can obtain all of the text and graphic files that make up the on-line exhibit, but will not be able to hold real time conferences. The FTP files available from a mini-computer at the Library of Congress will require passwords that the Library will release and publicize on the opening day of the exhibit, June 17. However, the Library hopes that the exhibit will spark lively and informative exchanges of E-Mail and discussions in existing BITNET listserv forums. Library staff will monitor the discussions and seek to answer questions that may arise. The forums that LC will monitor are Sovset (which requires paid membership); Talk.politics.soviet (on BITNET from listserv@indycms under the list name TPS-L); Soc.culture.soviet (on BITNET via listserv@indycms under SCS-L); and RUSSIA (Russia and Her Neighbors List, through BITNET listserv@indycms on CREN and listserv@indycms.iupui.edu on the Internet).

Sovset is accessible over the Internet, CompuServe, and Telenet. Sovset's purpose is to encourage communication, data sharing, and collaboration on projects between research communities in the field. The three basic services offered by Sovset are computer conferencing, a data library, and electronic mail.

Those visiting the exhibit can view it in its electronic form in either the National Demonstration Lab on the first floor of the Madison building or in the Machine Readable Collections Reading Room, which is located on the ground floor of the Jefferson building (LJ G22).

TREC II Guidelines

This rather long message contains all the information needed for running TREC, broken up into convenient sections. It is important to read it all at some time, but particularly to read the introduction and guidelines sections before you make the routing queries. For those of you who participated in TREC-1, please note that the guidelines for query construction have been modified and should be re-read.

Table of Contents

Section 1 -- introduction
Section 2 -- conceptual view of TREC
Section 3 -- guidelines for building knowledge bases/indices and constructing queries
Section 4 -- formats for submitting queries and results
Section 5 -- summary form for features and timing

Section 1 -- Introduction

TREC is designed to encourage research in information retrieval using large data collections. Two types of retrieval are being examined--retrieval using an "ad-hoc" query such as a researcher might use in a library environment, and retrieval using a "routing" query such as a profile to filter some incoming document stream. The test design and evaluation procedures are based on traditional information retrieval practices and are further elaborated in section 2. There are 40+ TREC participants, including some very large commercial groups, some very small university research groups, some actual products, and some radically different experimental approaches to information retrieval. The various participants are using a wide variety of indexing/knowledge base building techniques, and are using a variety of approaches to generating search queries. Therefore it is necessary to establish clear guidelines for the TREC task and to develop some methods of standardized reporting to allow comparison. The guidelines (section 3) deal with the methods of indexing/knowledge base construction, and with the methods of generating the queries from the supplied topics. In general they are constructed to reflect an actual operational environment, and to allow as fair as possible a separation among the diverse query construction approaches.

The standardized reporting will take two complementary forms. The first will be your ranked results for each topic and a list of the actual queries used to produce those results. Section 4 gives the formats for submitting the results and some guidelines for submitting the queries. It is important that you follow the results format, and it is highly recommended that those sites new to TREC submit a sample of results to make sure your formats are correct. If you wait and submit official results in an incorrect format, there is no guarantee we will be able to get back in time to ask for corrections and your results may be rejected!

To make the actual relevance judgments, the top 100 ranked documents for each topic for each submitted set of results will be merged at NIST, randomized, and judged anonymously by relevance assessors. This ensures that the highest ranked documents from each submitted run will be evaluated and builds a large pool of sample documents for judgment. This process should also produce a quality test collection for future retrieval research, since relevance judgments will have been based on the combined output of more than 40 systems, including sites using human search expertise. Each group will get recall/precision tables for their group and the complete list of judged documents for each topic (tagged with relevance status). These can be used for failure analysis purposes, or for additional experiments before the conference. In addition to the traditional recall/precision evaluation, it is important to compare techniques by examining the amount of effort needed to produce various performance levels. The second form of reporting will be a standarized form (section 5) describing features and timing aspects of your system. This form covers many of the issues that you might be presenting in a talk or poster, but insures that everyone reports on certain items. Some of these items are for mostly statistical interest (such as how large is your common word list), but most of these issues are critical to comparing techniques as to the total amount of effort needed to perform searching. This is particularly applicable to timing issues so that we can all compare the benefit of such techniques as manually generating queries versus automatically generating queries, building extensive knowledge bases by hand versus automatic indexing, performing deep parsing of the data versus performing shallow parsing, etc. Please be as honest and consistent as possible as this form is meant to be genuinely informative to the rest of the TREC participants. The deadlines listed below are an expanded version of the ones in the call--the dates have not changed.

Schedule:

- April 1, 1993 -- Third gigabyte of data distributed via CD-ROM, after routing queries (for topics 51-100) are received at NIST. This is the earliest the test documents will be available--you may take more time to create the routing queries but it means less time to work with the new data.
- May 15, 1993 -- 50 test topics (101-150) distributed

- June 1, 1993 -- results from the 50 routing queries and the 50 test topics due at NIST
- July 30, 1993 -- relevance judgments and individual evaluation scores due back to participants
- Aug 30-Sept 1., 1993 -- TREC conference at NIST in Gaithersburg, Md.

- --

Section 2 -- Conceptual View of TREC

The Application

The TREC task is not tied to any given application, and is not concerned with interfaces or optimized response time for searching. However it is helpful to have some potential user in mind when designing or testing a retrieval system. The model for a user in TREC is a dedicated searcher, not a novice searcher, and the model for the application is one needing monitoring of data streams for information on specific topics (routing), and the ability to do ad-hoc searches on archived data for new topics. It should be assumed that the users need the ability to do both high precision and high recall searches, and are willing to look at many documents and repeatedly modify queries in order to get high recall. Obviously they would like a system that makes this as easy as possible, but this ease should be reflected in TREC as added intelligence in the system rather than as special interfaces.

Test Design

TREC has been designed to evaluate system performance both in a routing (filtering or profiling) mode, and in an ad-hoc mode. The test design is based on traditional information retrieval models, and evaluation will use traditional recall and precision measures. A diagram of the test design illustrates the various components of TREC.

(Diagram Deleted)

This diagram reflects the four data sets (2 sets of topics and 2 sets of text) that are provided to participants. These data sets (along with a set of relevance judgments for the 100 training topics) are to be used to construct three sets of queries. Q1 is the set of queries (probably multiple sets) created to help in adjusting your system to this task, create better weighting algorithms, and in general to train the system for testing. The results of this research are used to create Q2, the routing queries to be used against the test documents. Q2 should be topics 51-100, not the full set of training topics. Q3 is the set of queries created from the test topics (101-150) as ad-hoc queries for searching against the combined training documents (D1 + D2). The results from searches using Q2 and Q3 are the official test results, and these queries and their search results are the information that should be sent to NIST. For example, topic 051, one of the training topics, could be run against the training documents as part of Q1, and then developed into a routing query to be run against the test documents as part of Q2. Topic 101, one of the test topics, should be run against the combined training documents (D1 + D2) as part of Q3.

For those TREC participants using a subset of the data (Category B), your training documents would be the Wall Street Journal text on the training CD-ROMs. The subsets of the test documents will be provided later.

Evaluation

The evaluation in TREC will mainly be standard recall/precision tables generated from your results. We will be using the SMART evaluation package, which includes both recall-level and document-level averaging. The package itself (programmed in C) is available from Chris Buckley (chrisb@cs.cornell.edu) and will be very useful to have to get early results or to evaluate additional runs using the same measures. One of the evaluation "improvements" suggested at TREC-1 was to strongly urge all participants to make multiple internal runs to attempt to modularize their various techniques to separate the effects of each technique, and then present the "unofficial" evaluations at the conference in addition to their official results.

Section 3 -- TREC Guidelines

Guidelines for constructing and manipulating the system data structures

The system data structures are defined to consist of the original documents, any new structures built automatically from the documents (such as inverted files, thesaurii, conceptual networks, etc.) and any new structures built manually from the documents (such as thesaurii, synonym lists, knowledge bases, rules, etc.).

1. System data structures can be built using the initial training set (documents D1 + D2, topics 1-100, and relevance judgments). They may be modified based on the test documents D3, but not based on the test topics 101-150. In particular, the system data structures may not be based in any way on the results of retrieving documents for the test topics and having a human look at the retrieved documents (or even just their titles).

2. There are several parts of the Wall Street Journal and the Ziff material that contain manually assigned controlled or uncontrolled index terms. These fields will be delimited by SGML tags, as specified in the read_me files included with the data. Other parts of the TREC data contain no manual indexing. Since the primary focus of TREC is on retrieval and routing of naturally occurring text, these manually indexed terms should not be indiscriminately used as if they are a normal part of the text. If your group decides to use these terms, they should be part of a specific experiment that utilizes manual indexing terms, and their use should be declared.

3. Special care should be used in handling the routing topics. In a true routing situation, a single document would be indexed and "passed" against the routing queries. Since most of you will be indexing the test data set as a complete set, routing should be simulated by not using any test document collection information (such as IDF based on the test collection, total frequency based on the test collection, etc.) in the searching. It is perfectly permissible to use training-set collection information however.

Guidelines for constructing the queries

There are many possible methods for converting the topics we will supply into queries that your system can execute. We have broadly defined three generic methods, based on the kind of manual intervention used. It is expected that all TREC participants will turn in either one or two sets of results for each topic. If two sets of results are submitted, they may correspond to different methods, or if desired, can be variants within the same method. Different methods can be used for the routing tests and the ad-hoc tests. For example, a site might submit results for ad-hoc retrieval for two different runs using automatic query construction (AUTOMATIC), perhaps comparing two different ranking formulae, or may submit results for automatic initial query construction (AUTOMATIC) vs relevance feedback (FEEDBACK). However, for the routing results, they might use one AUTOMATIC technique and one MANUAL technique (see below for definitions of these categories). The intent of the categories is as follows. The first two categories below (AUTOMATIC and MANUAL) are artificial categories. No decent information retrieval system would expect the user to look at 500+ retrieved documents without interacting with the system and at least tweaking the query a bit. But such human-system interaction is difficult to evaluate. In order to evaluate and compare systems, it's helpful to put artificial constraints on the problem, thus the first two categories. The FEEDBACK category loosens those constraints, but at a cost in the ability to accurately compare systems. Please read these category definitions CAREFULLY, and contact the program chair if you have any questions. Several groups in TREC-1 submitted misleading results due to misunderstanding the category definitions. We intend to avoid this in TREC-2.

1. AUTOMATIC -- completely automatic query construction.

ad-hoc queries -- The system will automatically extract information from the topic to construct the query. The query will then be submitted to the system (with no manual modifications) and the results from the system will be the results submitted to NIST. The system is not allowed to use any software, auxiliary files, knowledge bases, etc. that were manually changed after the ad-hoc topics were received. The only exception to this is to correct for formatting problems in the topic or document collections, as officially announced by NIST.

routing queries -- the queries should be constructed automatically using the training topics, the training relevance judgments and/or the training data. The queries should then be submitted to NIST before the test data is released and should not be modified after that point. The unmodified queries should be run against the test data and the results submitted to NIST.

2. MANUAL -- manual query construction.

ad-hoc queries -- The query is constructed in some manner from the topic, either manually or using machine assistance. Once the query has been constructed, it will be submitted to the system (with no manual intervention), and the results from the system will be the results submitted to NIST. There should be no manual intervention after initial query construction that would affect the results. This mimics the performance of a manual system with the user's initial query attempt, before any interactive tweaking of the query occurs. In particular, absolutely no alteration of the query based on examining of retrieved documents is allowed!

routing queries -- the queries should be constructed in the same manner as the ad-hoc queries for MANUAL, but using the training topics, relevance judgments, and training data. They should then be submitted to NIST before the test data is released and should not be modified after that point. The unmodified queries should be run against the test data and the results submitted to NIST.

3. FEEDBACK -- automatic or manual query construction with feedback.

adhoc queries -- The system takes an automatically or manually formed query and returns some documents to the user. After looking at those documents, the query may be reformulated and more documents returned for the new query. This is the most realistic category, but is the hardest to evaluate. It is essential that every document (or document title) looked at for a query be kept track of (in the order in which they were looked at). Evaluation will be based on this ordering, as well as the final set of retrieved documents.

The general algorithm would look like:
1. Form initial query (either automatically or manually)
2. Retrieve some documents
3. Look at a number of documents (or titles), keeping track of the docid's and order in which they looked at.
4. Reformulate the query based on the original query and information from the retrieved documents in step 3
5. Retrieve more documents.
6. Either stop, or go to step 3 again if another iteration of feedback is desired.

What will finally be turned in to NIST is your final list of the top X documents, plus your list of looked-at documents. NIST will then form a new set of X documents by joining your lists, removing possible duplicates, and ranking the looked-at documents ahead of your final list, in the order that they were looked at. (This is the standard "frozen evaluation" method of feedback evaluation). Note that an implication of this approach is that you probably want to look at as few documents from your initial run as is necessary, since your reformulated query should rank documents better than your initial query.

routing queries -- FEEDBACK cannot be used for routing queries.

- --

Section 4

Results format and Query format

The results should be the top X documents retrieved for each topic. X will be determined at a later date, but will be AT LEAST 500 and might be 1000. This depends on logistics. If your group would have trouble keeping the top 1000 documents, please let us know. The format should be as follows:

030 Q0 ZF08-175-870 0 4238 prise1
030 Q0 ZF08-306-044 1 4223 prise1
030 Q0 ZF09-477-757 2 4207 prise1
030 Q0 ZF08-312-422 3 4194 prise1
030 Q0 ZF08-013-262 4 4189 prise1

etc.

The first column is the topic number. The second column is the query number within that topic. This should always be Q0 for AUTOMATIC and MANUAL, but should change in FEEDBACK as new iterations are made (i.e. the first iteration of feedback would use Q1, the second Q2, etc.). The third column is the official document number of the retrieved document and is the number found in the "docno" field. The fourth column is the rank the document is retrieved, and the fifth column shows the score (integer or floating point) that generated the ranking. This score MUST be in descending order and is important to include so that we can handle tied scores (for a given run) in a uniform fashion (the evaluation routines rank documents from these scores, not from your ranks). The sixth column should be a unique identifier for your group AND for the method used. The example shows the prise system, method 1. Please pick 5 letters followed by a 1, 2 or 3 depending on the method the run matches. If by some weird chance we get the same set of letters for two groups, we will let you know (see below). The width of the columns is not important, but it is important to include all columns and have at least one space between the columns.

Please send in a sample of results from at least two topics as soon as possible if you are new to TREC. These will not be used for anything but to make sure your formats are correct, and to help us set up our programs here to merge the official results. Because there is such a tight schedule in June, any mistakes in format must be corrected now.

- ---

Query format (by Dave Lewis)

Sites taking part in TREC are required to submit the queries they derive from the NIST-supplied topic descriptions. Routing queries are to be submitted to NIST before the test data for routing is received by the site. Ad hoc queries are submitted to NIST along with the top X documents retrieved by the ad hoc query.

The form of the queries submitted will of necessity vary from site to site. There are several somewhat contradictory goals to be met by the submitted queries:

1) The submitted queries should reflect the results of all manual intervention in the query process. For MANUAL querying without feedback the query submitted should be one for which all further system processing should be fully automated. For FEEDBACK querying with relevance feedback, we want two versions of the query---one before any relevance feedback -- i.e., one that would satisfy the MANUAL

restrictions -- and one after all relevance feedback and manual tuning is completed.

2) The submitted query should be as close as possible to the form actually used by your system in ranking documents.

3) The proprietary nature of the knowledge bases used by some systems should be respected to the extent possible.

4) The submitted query should be interpretable in isolation from the site's system to the extent possible.

There is likely to be considerable tension among desiderata 2, 4, and possibly 3. For instance, the submitted query for a system which ranked documents using a linear function might be in the form of two arrays, one of which gives unique integer term numbers for each indexing term and one of which gives real-valued coefficients for the linear function:

"POSITION: 0 1 2 3 4
ARR1: 3783 3838 4677 13999 20033
ARR2: 1.73 -3.8 13.38 2.73 0.47 "

This, however, is not a terribly useful representation from the standpoint of letting other researchers understand what your system is doing.

Now, suppose that this classifier is used to assign scores to documents based on the presence of whitespace-delimited words in the document text. We assume each indexing term corresponds to one such word, and the indexing term takes on a numeric depending on the number of occurrences of the word in a document. Then a more useful representation of the query might be:

"1.73 * Japan - 3.8 * U.S. + 13.38 * auto + 2.73 * export + 0.47 * car"

if these were the corresponding words. Even this form would only be completely interpretable in the context of additional information provided in documentation about the system, such as how it tokenizes text, and what within document term weights are used.

Some techniques, such as neural nets, factor analysis, and methods using knowledge bases, may produce queries that are difficult to represent in a comprehensible form. Sites which have questions about the form in which to submit their queries should contact the designated program committee contact for this issue, David Lewis

(lewis@research.att.com).”

- ---

Section 5
System Summary and Timing

General Comments
The timings should be the time to replicate runs from scratch, not including trial runs, etc. The times should also be reasonably accurate. This sometimes will be difficult, such as getting total time for document indexing of huge text sections, or manually building a knowledge base. Please do your best.

I. Construction of indices, knowledge bases, and other data structures
(please describe all data structures that your system needs for searching)

A. Which of the following were used to build your data structures
1. stopword list
a. how many words in list?
2. is a controlled vocabulary used?
3. stemming
a. standard stemming algorithms
which ones?
b. morphological analysis
4. term weighting
5. phrase discovery
a. what kind of phrase?
b. using statistical methods
c. using syntactic methods
6. syntactic parsing
7. word sense disambiguation
8. heuristic associations
a. short definition of these associations

9. spelling checking (with manual correction)
10. spelling correction
11. proper noun identification algorithm
12. tokenizer (recognizes dates, phone numbers, common patterns)
a. which patterns are tokenized?

13. are the manually-indexed terms used?
14. other techniques used to build data structures (brief description)

B. Statistics on data structures built from TREC text
(please fill out each applicable section)
1. inverted index
a. total amount of storage (megabytes)
b. total computer time to build (approximate number of hours)
c. is the process completely automatic?
if not, approximately how many hours of manual labor?
d. are term positions within documents stored?
f. single terms only?
2. clusters
a. total amount of storage (megabytes)
b. total computer time to build (approximate number of hours)
c. brief description of clustering method

d. is the process completely automatic?
if not, approximately how many hours of manual labor?
3. n-grams, suffix arrays, signature files
a. total amount of storage (megabytes)
b. total computer time to build (approximate number of hours)
c. brief description of methods used
d. is the process completely automatic?
if not, approximately how many hours of manual labor?
4. knowledge bases
a. total amount of storage (megabytes)
b. total number of concepts represented
c. type of representation (frames, semantic nets, rules, etc.)

d. total computer time to build (approximate number of hours)
e. total manual time to build (approximate number of hours)
f. use of manual labor
(1) mostly manually built using special interface
(2) mostly machine built with manual correction
(3) initial core manually built to "bootstrap" for
completely machine-built completion
(4) other (describe)

g. auxiliary files needed for machine use
(1) machine-readable dictionary (which one?)

(2) other (identify)
5. special routing structures (what?)

a. total amount of storage (megabytes)
b. total computer time to build (approximate number of hours)
c. is the process completely automatic?
d. brief description of methods used

6. other data structures built from TREC text (what?)

a. total amount of storage (megabytes)
b. total computer time to build (approximate number of hours)
c. is the process completely automatic?
if not, approximately how many hours of manual labor?
d. brief description of methods used

C. Data built from sources other than the input text
1. internally-built auxiliary files
a. domain independent or domain specific (if two separate files, please fill out one set of questions for each file)
b. type of file (thesaurus, knowledge base, lexicon, etc.)

c. total amount of storage (megabytes)
d. total number of concepts represented
e. type of representation (frames, semantic nets, rules, etc.)

f. total computer time to build (approximate number of hours)
(1) if already built, how much time to modify for TREC?
g. total manual time to build (approximate number of hours)
(1) if already built, how much time to modify for TREC?
h. use of manual labor
(1) mostly manually built using special interface
(2) mostly machine built with manual correction
(3) initial core manually built to “bootstrap” for completely machine-built completion
(4) other (describe)

2. externally-built auxiliary file
a. type of file (Treebank, WordNet, etc.)

b. total amount of storage (megabytes)

c. total number of concepts represented
d. type of representation (frames, semantic nets, rules, etc.)

II. Query construction
(please fill out a section for each query construction method used)

A. Automatically built queries (ad-hoc)
1. topic fields used
2. total computer time to build query (cpu seconds)
3. which of the following were used?
a. term weighting with weights based on terms in topics
b. phrase extraction from topics
c. syntactic parsing of topics
d. word sense disambiguation
e. proper noun identification algorithm
f. tokenizer (recognizes dates, phone numbers, common patterns)
(1) which patterns are tokenized?
g. heuristic associations to add terms
h. expansion of queries using previously-constructed data structure (from part I)
(1) which structure?
i. automatic addition of Boolean connectors or proximity operators
j. other (describe)

B. Manually constructed queries (ad-hoc)
1. topic fields used
2. average time to build query (minutes)
3. type of query builder
a. domain expert
b. computer system expert
4. tools used to build query
a. word frequency list
b. knowledge base browser (knowledge base described in part I)
(1) which structure from part I
c. other lexical tools (identify)

5. which of the following were used?
a. term weighting
b. Boolean connectors (AND, OR, NOT)
c. proximity operators

d. addition of terms not included in topic
(1) source of terms
e. other (describe)

C. Feedback (ad-hoc)
1. initial query built by method 1 or method 2?
2. type of person doing feedback
a. domain expert
b. system expert
3. average time to do complete feedback
a. cpu time (total cpu seconds for all iterations)
b. clock time from initial construction of query to completion of final query (minutes)
3. average number of iterations
a. average number of documents examined per iteration
4. minimum number of iterations
5. maximum number of iterations
6. what determines the end of an iteration?

6. feedback methods used
a. automatic term reweighting from relevant documents
b. automatic query expansion from relevant documents
(1) all terms in relevant documents added
(2) only top X terms added (what is X)
(3) user selected terms added
c. other automatic methods
brief description

d. manual methods
(1) using individual judgment with no set algorithm
(2) following a given algorithm (brief description)

D. Automatically built queries (routing)
1. topic fields used
2. total computer time to build query (cpu seconds)
3. which of the following were used in building the query?
a. terms selected from
(1) topic
(2) all training documents
(3) only documents with relevance judgments
b. term weighting

(1) with weights based on terms in topics
(2) with weights based on terms in all training documents
(3) with weights based on terms from documents with relevance judgments
c. phrase extraction
(1) from topics
(2) from all training documents
(3) from documents with relevance judgments
d. syntactic parsing
(1) of topics
(2) of all training documents
(3) of documents with relevance judgments
e. word sense disambiguation
(1) using topic
(2) using all training documents
(3) using documents with relevance judgments
f. proper noun identification algorithm
(1) from topics
(2) from all training documents
(3) from documents with relevance judgments
g. tokenizer (recognizes dates, phone numbers, common patterns)
(1) which patterns are tokenized?
(2) from topics
(3) from all training documents
(4) from documents with relevance judgments
h. heuristic associations to add terms
(1) from topics
(2) from all training documents
(3) from documents with relevance judgments
i. expansion of queries using previously-constructed data structure (from part I)
(1) which structure?
j. automatic addition of Boolean connectors or proximity operators
(1) using information from the topics
(2) using information from the all training documents
(3) using information from documents with relevance judgments
k. other (brief description)

E. Manually constructed queries (routing)
1. topic fields used
2. average time to build query (minutes)
3. type of query builder

a. domain expert
b. system expert
4. data used for building query
a. from training topic
b. from all training documents
c. from documents with relevance judgments
d. from other sources (what?)

5. tools used to build query
a. word frequency list
b. knowledge base browser (knowledge base described in part I)
(1) which structure from part I
c. other lexical tools (identify)

d. machine analysis of training documents
(1) describe

5. which of the following were used?
a. term weighting
b. Boolean connectors (AND, OR, NOT)
c. proximity operators
d. addition of terms not included in topic
(1) source of terms
e. other (brief description)

III. Searching
A. Total computer time to search (cpu seconds)
1. retrieval time (total cpu seconds between when a query enters the system until a list of document numbers are obtained)

2. ranking time (total cpu seconds to sort document list)

B. Which methods best describe your machine searching methods
1. vector space model
2. probabilistic model
3. cluster searching
4. n-gram matching
5. Boolean matching
6. fuzzy logic (include your definition)

7. free text scanning
8. neural networks
9. conceptual graph matching
10. other (describe)

B. What factors are included in your ranking?
1. term frequency
2. inverse document frequency
3. other term weights (where do they come from?)
4. semantic closeness (as in semantic net distance)
5. position in document
6. syntactic clues (state how)
7. proximity of terms
8. information theoretic weights
9. document length
10. completeness (what % of the query terms are present)
11. n-gram frequency
12. word specificity (i.e., animal vs. dog vs. poodle)
13. word sense frequency
14. cluster distance
15. other (specify)

IV. What machine did you conduct the TREC experiment on?
How much RAM did it have?
What was the clock rate of the CPU?

V. Some systems are research prototypes and others are commercial.
To help compare these systems:

1. How much "software engineering" went into the development of your system?

2. Given appropriate resources, could your system be made to run faster? By how much (estimate)?

3. What features is your system missing that it would benefit by if it had them?

Research Announcements

Research on Digital Libraries

,

A Joint Initiative of:

NATIONAL SCIENCE FOUNDATION
COMPUTER AND INFORMATION SCIENCE AND ENGINEERING DIRECTORATE

ADVANCED RESEARCH PROJECTS AGENCY
COMPUTING SYSTEMS TECHNOLOGY OFFICE and the
SOFTWARE AND INTELLIGENT SYSTEMS TECHNOLOGY OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

PROPOSALS MUST BE RECEIVED AT NSF NO LATER THAN FEBRUARY 4, 1994

NSF 93-141 (NEW)

Digital Library Initiative, FY 1994

INTRODUCTION
The past decade has seen a remarkable expansion in digital networks within the U.S. research and education community from a state where networking was the purview of the privileged few to one where it is considered an essential tool by millions of researcher and educators. Since the mid-1980's with the advent of NSFNET the volume of traffic, the number of interconnected networks and the functionality of the networks has grown and continues to grow exponentially. The entire assemblage of linked networks using the IP communications protocol throughout the world is now referred to as the Internet. It is a network of networks, which within the U.S., links one third of all two year and four year colleges and universities, many primary and secondary schools, public and private institutions, commercial enterprises,

individuals in their homes, and foreign institutions in sixty countries. Information sources accessed via the Internet are the ingredients of a digital library. Today, the network connects some information sources that are a mixture of publicly available (with or without charge) information and private information shared by collaborators. They include reference volumes, books, journals, newspapers, national phone directories, sound and voice recordings, images, video clips, scientific data (raw data streams from instruments and processed information), and private information services such as stock market reports and private newsletters. These information sources, when connected electronically through a network, represent important components of an emerging, universally accessable, digital library.

To explore the full benefits of such digital libraries, the problem for research and development is not merely how to connect everyone and everything together in the network. Rather, it is to achieve an economically feasible capability to digitize massive corpora of extant and new information from heterogeneous and distributed sources; then store, search, process and retrieve information from them in a user friendly way. Among other things, this will require both fundamental research and the development of "intelligent" software. It is the purpose of this announcement to support such research and development by combining the complementary strengths of the participating agencies in basic research, advanced development and applications, and academic/industry linkage.

IMPORTANT CONSIDERATIONS

1. Awards will be made by NSF as Cooperative Agreements for proposals to engage in research, prototype and testbed activities. Awards of up to $1,200,000 a year for up to four years will be made to conduct programs of research and to develop and test elements of a digital library on a significant scale in a distributed environment. It is expected that up to 6 awards will be made under this announcement, depending on the quality of proposals and the availability of funds.

2. Cost sharing of at least 25% is required for all projects submitted in response to this announcement. The proposed cost sharing will be considered in evaluating proposals and will be a condition of any resulting awards (See the Proposal Preparation section, below).

3. Successful proposals will have demonstrated that as part of the research they will digitize a significantly large and important information collection, or use an existing collection, to serve as an experimental platform to demonstrate scale-up potential and as an experimental testbed for the research proposed. These testbeds must also be made accessible for research purposes to individuals who are not part of the

proposal team (see Proposal Preparation, 3, below)

4. Each proposal should, as appropriate to the research focus, include the active participation of the following groups, as relevant (these may be separate organizations or parts of a single organization): (1) client groups (e.g., specific research communities or other users of the information encompassed in the proposal); (2) commercial enterprises that would be involved in the commercialization of a digital library system (e.g., publishers, software houses, stock exchanges, equipment manufacturers, communications companies, etc.); (3) archival establishments, either private or governmental (e.g., libraries, data repositories, clearing houses, government or private information or data services); and (4) relevant computer and other science and engineering research groups (e.g., academic departments, supercomputer centers, industrial laboratories). These groups should be involved as sub-contractors to a single primary proposing academic institution.

5. A requirement for all awardees is that all publications, reports, data and other output from awards must be prepared in digital format and meet requirements for storage, indexing, searching and retrieval in a repository to be set up to capture results of these awards. These requirements will be devised jointly by the group of all awardees and the sponsoring agencies.

RESEARCH TO BE SUPPORTED
It is the purpose of this initiative to provide the funding and leadership for research fundamental to the development of digital libraries. Applicants can propose research in any or all of the following areas:

1-capturing data (and descriptive information about such data) of all forms (text, images, sound, speech, etc.) and categorizing and organizing electronic information in a variety of formats.

2-advanced software and algorithms for browsing, searching, filtering, abstracting, summarizing and combining large volumes of data, imagery, and all kinds of information; and

3-the utilization of networked databases distributed around the nation and around the world.

The types of research related to the above three areas are shown below. The examples are not meant to be exclusive but are meant to be illustrative:

Area 1
- - - New research on systems for capturing data of all forms

For example:
* OCR page layout, segmentation and analysis software
* Speech recognition, audio segmentation and analysis software
* File conversion into editable, processable representations
* Broadcast capture & digitization (to generate multimedia data bases)
* Graphics understanding (image, drawing, graph recognition)
* Quality, fidelity maintenance

- - - New research on how to categorize and organize electronic information in a variety of formats

For example
* Indexing, interpretation, classification and cataloging
* Multi-lingual indexing on content and citations
* Hypermedia structuring and linking of documents
* Graphical interfaces for knowledge representation
* Browsing technology for large knowledge spaces, news grazing

Area 2
- - - New Research fundamental to the development of advanced software for searching, filtering, and summarizing large volumes of data, imagery, and all kinds of information

For example:
* Retrieval theories and models for data, metadata, information, knowledge bases, evaluation methods
* Formal structures of documents and texts, query languages
* Intelligent text processing and document management
* Feature-based image analysis and classification, pattern recognition
* Multi-pass retrieval algorithms (progressive refinement, feedback)
* Updatable indexing systems
* Spatial-temporal feature indexing of video
* Filtering, routing, alerting, selective dissemination of information
* Clustering, summarization, abstracting
* Natural language analysis for data extraction or representation
* Natural language generation systems
* Lexicon, thesaurus, concept space generation
* Adaptive/learning systems: connectionist, neural networks

* Context based pattern matching and retrieval of multimedia data
* Robust matching with noisy data, uncertainty, imprecision
* Fast search, query optimization

- - - Research on visualization and other interactive technology for quickly browsing large volumes of imagery

For example:
* Pictorial feature recognition, image classification
* Human perception (visual, auditory)
* Multi-scale displays, zooming
* Data visualization (e.g., airflow, human genome)
* Interactive visualization control
* Use of simulation to improve visualization/description
* Navigation, hypermedia, retaining serendipity, guides/paths/tours
* Developing metaphors, usable virtual reality environment: suites of objects (peripherals, information types, relations, properties, views) that work together for each domain
* Sustaining rapid performance with regard to moving and manipulating large digital imagery data sets.

Area 3
- - - Research on networking protocols and standards needed to insure the ability of the digital network to accommodate the high volume, bandwidth and switching requirements of a digital library.

For example:
* Network security
* Protocol design
* Data compression
* Ensuring scalability for large orders of magnitude increases in the number of simultaneous users

- - - New research leading to simplifying the utilization of networked databases distributed around the nation and around the world

For example:
* Knowbots/agents/mediators, intelligent gatekeepers
* Federated heterogeneous distributed object-oriented data and information base systems
* Personalized interactive news, magazine, and journal services

* Adaptable systems and services for disabled users, human augmentation
* Authentication, authorization
* Modeling, simulating usage, economics of access
* Collaboration technology with multimedia information interchange, multi-user editing, drawing, storage, retrieval, display, annotation, shared objects

- - - Research on individual and group behavioral, social and economic issues in digital libraries

For example:
* Intellectual property rights
* Privacy and security
* Impact of digital libraries on the conduct of science
* Publishing in a digital environment
* Charging mechanisms for copyrighted documents

PROPOSAL PREPARATION
The proposals must be marked DIGITAL LIBRARIES in the top left hand box, "Program Announcement", on the cover sheet (NSF Form 1207).

Proposals must be prepared according to the instructions given in Grants for Research and Education in Science and Engineering (GRESE: NSF 92-89) except that the following sections are also required.

1. A section following the Project Summary to be entitled, "Executive Summary". This section is limited to five pages. It will form an important first step in the review process and it should therefore be a careful abstraction of the key aspects of the proposal.

2. The Project Description section may not exceed 40 pages. It should be prepared as indicated on Page 4 of GRESE, making sure to explain the specific research and experiments to be conducted. These must be described in sufficient detail to allow merit review of the scientific content of the proposed research and its chance of success.

3. A section of the Project Description, entitled "Testbed Facility", describing the purpose and operation of the experimental testbed platform: its design; development; and management. This should include an explanation of how experiments and tests proposed by individuals or groups who are not part of the proposal will be solicited, evaluated and accommodated. (It is the intention of the

National Science Foundation to encourage applications for supplemental funding in years 2 through 4 to help fund the research of successful individual applicants.)

4. A section of the Project Description entitled, “Expected Accomplishments”. This should be a statement and enumeration of the expected accomplishments and the measures you propose for evaluating progress annually for each year of the award duration.

5. A section separate from the Project Description entitled “Organizational Roles”. This section should describe the nature of each organization’s participation in the proposal with a description of its cost sharing, and the overall management plan for the project.

6. A section entitled, “Cost Sharing”. The amount of cost sharing must be shown in the proposal in enough detail to allow NSF to determine its impact on the proposed project. Documentation of availability of cost sharing must be included in the proposal. Only items which would be allowable under the applicable cost principles, if charged to the project, may be made from any non-Federal source, including non-Federal grants or contracts. Contributions from non-Federal sources may be counted as cost
sharing toward Federal projects only once.

PROPOSAL EVALUATION
Proposals will be evaluated by panels of experts. Supplementary mail reviews will be solicited as feasible and necessary to achieve a fair and accurate review of all proposals. Some potentially successful submissions may receive site visits if the panel deems this desirable in order to properly evaluate the proposals. Selection of awards will be made jointly by the sponsoring agencies from the group of proposals receiving the highest ratings from the panel of experts.

Evaluation criteria which are applied to all NSF proposals are those listed in Grants for Research and Education in Science and Engineering, NSF 92-89, October, 1992. Additional criteria for this announcement are:

- inter-institutional aspects which would provide a positive stimulus to the design, development and commercialization of digital libraries;

- a commitment to and plan for operating an experimental testbed platform (see Proposal Preparation, 3, above);

- the extent to which the proposed research is likely to produce results generalizable to other digital libraries.

- cost sharing of at least 25% will be a specific evaluation criterion.

SCHEDULE
Organizations interested in submitting proposals should send an electronic mail message with the following statement: "I am interested in submitting a proposal to the Digital Libraries Initiative in Areas
___(1,2,3)."

Name
Title
Organization
Mailing address

email address

Address the email to:
gbarber@nsf.gov
cc: lrosenbe@nsf.gov

If email on the internet is not available to you then send a postcard with the same information to:

Up to October 25, 1993
Gwendolyn Barber
Room 310
National Science Foundation
1800 G Street, NW
Washington, D.C. 20550

As of October 25, 1993
Gwendolyn Barber
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

This information will be used to notify you of special regional briefings that will be held for this announcement. These briefings will be designed to answer your questions and to help you find potential partners for research consortia under this announcement. It is anticipated that these briefings will occur during November/December, 1993.

IN ORDER TO BE CONSIDERED UNDER THIS ANNOUNCEMENT, PROPOSALS MUST BE RECEIVED BY NSF NO LATER THAN FEBRUARY 4, 1994

SUCCESSFUL PROPOSERS WILL RECEIVE NOTIFICATION IN JUNE/JULY, 1994.

OTHER
The National Science Foundation (NSF) provides awards for research in the sciences and engineering. The awardee is wholly responsible for the conduct of such research and preparation of the results for publication. The Foundation, therefore, does not assume responsibility for such findings or their interpretation.

The Foundation welcomes proposals on behalf of all qualified scientists and engineers, and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and research related programs described in this document.

In accordance with Federal statutes and regulations and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, denied the benefits of, or be subjected to discrimination under any program or activity receiving financial assistance from the National Science Foundation.

Facilitation Awards for Scientists and Engineers with Disabilities provides funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on an NSF project. Contact the program coordinator in the Directorate for Education and Human Resources. The telephone number is (202) 357-7562.

The Foundation has TDD (Telephonic Device for the Deaf) capability, which enables individuals with hearing impairment to communicate with the NSF Information Center about NSF programs, employment, or general information.
The telephone number is (202) 357-7492.

This program is described in the Catalog of Federal Domestic Assistance category 47.070

Copies of NSF publication, Grants for Research and Education in Science and Engineering (GRESE: NSF 92-89) are available at no cost from:

National Science Foundation
Forms and Publications, Room 232
Washington, DC 20550

Telephone number: (202) 357-7963 or Fax number: (703) 644-4278.

Publications may also be ordered electronically using the Science and Technology Information System (STIS). The full text can be searched on-line, and copied from the system. Instructions for the use of the system are in NSF 91-10 "STIS Flyer". The printed copy is available from the Forms and Publications Unit. An electronic copy may be requested by sending a message to "stis@nsf" (bitnet) or "stis@nsf.gov" (Internet).

Technical questions about this announcement may be addressed to:

Laurence C. Rosenberg
Deputy Division Director
Information, Robotics & Intelligent Systems Division
National Science Foundation
1800 G Street, NW
Washington, D.C. 20550

Starting October 25, 1993
4201 Wilson Boulevard
Arlington, VA 22230

Electronic Mail address: lrosenbe@nsf.gov.
Telephone: (202) 357-9592
Fax: (202) 357-0320

Privacy Act and Public Burden Statements
The information requested on the application materials is solicited under the authority of the National Science Foundation Act of 1950, as amended. It will be used in connection with the selection of qualified proposals and may be used and disclosed to qualified reviewers and staff assistants as part of the review process and to other government agencies. See System of Records, NSF-50, Principal Investigator/Principal File and Associated Records and NSF-51, Reviewer/ Proposals File and Associated Records, 56 Federal Register 54907 (October 23,

1991). Submission of the information is voluntary. Failure to provide full and complete information, however, may reduce the possibility of your receiving an award.

Public reporting burden for this collection of information is estimated to average 120 hours per response, including the time for reviewing instructions. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

Herman G. Fleming
Reports Clearance Officer
Division of Contracts Policy and Oversight
National Science Foundation
Washington, DC 20550

and to:

Office of Management and Budget
Paperwork Reduction Project (3145-0058)

Research on Scientific Databases

Special Announcement

Division of Information, Robotics and Intelligent Systems
Directorate for Computer and Information
Science and Engineering

NATIONAL SCIENCE FOUNDATION

PROPOSAL DEADLINE: September 15, 1992

The Division of Information, Robotics and Intelligent Systems (IRIS) in the Directorate for Computer and Information Science and Engineering (CISE) announces a special interest in supporting interdisciplinary research efforts in scientific and engineering databases.

BACKGROUND

Research on the design, development, management, and use of databases has traditionally focused on concepts and requirements critical to business-like environments. However, current database technology falls short of supporting the diverse needs of scientific and engineering applications. New advances in data storage/access technology, knowledge-based systems, and networked computing have brought about the promise to greatly increase the productivity in science and engineering research in this final decade of the 20th century. Scientific databases can be viewed as critical repositories of knowledge, both existing and yet to be discovered. Global change studies, astronomy, human genome mapping, social and economic studies, and engineering design are a few examples of research areas that generate and require access to extraordinarily large amounts of multi-media data forms: numbers, symbols, texts, images and others. Addressing the special characteristics and requirements of scientific and engineering databases will potentially further database technologies and enable a wide range of scientists and engineers to better utilize their data and other computing resources.

The research needed to make progress in scientific and engineering databases must be drawn from specialists in many disciplines. Therefore, collaboration of CISE researchers with researchers in other scientific or engineering disciplines is strongly encouraged.

Depending on the content of the received proposals, joint evaluation and funding by the relevant scientific and engineering NSF programs outside CISE is planned.

OBJECTIVES

The objective of this announcement is to foster coupling between database technology and scientific or engineering research for the advancement of both. Specifically, the aim is to promote:

1. Stimulation of multi-disciplinary research in scientific and engineering databases that addresses significant, real requirements of an application domain. Understanding of the requirements should be derived through collaboration with the domain scientists or engineers.

2. Expansion of general database technology through addressing the generic requirements of the application domains. Methodologies and tools developed for

scientific or engineering databases should be at a high conceptual level with an aim to enhance capabilities of the next generation of general information systems.

3. Enhancement of scientific infrastructure by making the databases, software, and other sharable resources produced under this initiative available to the research community.

AREAS OF INTEREST

Research is sought on methodologies and tools for the representation and manipulation of very large volumes of scientific or engineering data in highly distributed heterogeneous environments. In this context, research in three interrelated areas is encouraged:

1. Scientific Database Models and Systems. Theoretical foundations for the representation and manipulation of new data types (e.g., temporal, spatial and image data, textual data, spectrum data, engineering design data, materials data, chemical compounds, sequences, graphs, user-defined objects with inheritance and encapsulation, or declarative extensions); metadata management; data/knowledge calibration and validation; and uncertainty handling. System issues include system extensibility; rapid prototyping support; development of user-transparent, multi-level storage management (main memory through tertiary storage); multimedia data indexing; partial match retrieval algorithms; long/parallel/concurrent transaction processing; archiving; and version control. Research in this area must consider the special data characteristics associated with a scientific or engineering discipline.

2. Knowledge Discovery in Scientific Databases. Innovative methods, techniques and tools that provide seamless integration between database management and scientific analytic or engineering design tools. Topics span computing environment transparency; event finding, data examination, selection, analysis and manipulation of temporally or spatially related data (images); data analysis algorithms; scientific visualization; parallel model execution and cross-validation on large volumes of data; automated knowledge acquisition; incorporation of new knowledge into the system; and audit trail provisions.

3. Resource Sharing Environments. Improvement of data access and resource sharing in distributed, networked environments to support collaboration among scientists or engineers. Research in interoperability is critical to browsing, resource (data, bibliographic references or tools) location, access, and joint processing on systems ranging from personal workstations to supercomputers. Topics include

heterogeneous database systems management; effective methods for transport, combination and manipulation of data subsets, images or literature references; development of domain-specific and interdisciplinary lexicons and directories for databases and software; assessment of the data relevance and quality; and evolutionary establishment of self-describing, extensible standards for data exchange.

The topics listed above are not intended to represent the complete set of issues comprising the areas; they are suggestive rather than limiting. Other relevant information may be found in NSF workshop reports, "Scientific Data Management", "Database Systems: Achievements and Opportunities", and "Heterogeneous Databases Systems", published in SIGMOD Record, Vol. 19, No. 4, Dec. 1990, and in the report "Grand Challenges 1993: High Performance Computing and Communications" (available by calling NSF, Division of Information, Robotics and Intelligent Systems, 202-357-9572).

INQUIRIES

Inquiries about this research initiative announcement are welcome and may be directed to:

Dr. Maria Zemankova
Program Director, Database and Expert Systems
National Science Foundation
1800 G Street, N.W. -- Room 310
Washington, D.C. 20550
Telephone: 202-357-9570 Fax: 202-357-0320
Email: mzemanko@nsf.gov / mzemanko@nsf.bitnet

[Editor's Note: Dr. Zemankova left NSF on leave as of Fall 1993.]

Library of Congress

Excerpts from: Workshop On Electronic Texts Proceedings

Edited by James Daly

9-10 June 1992

Library of Congress Washington, D.C.

Supported by a Grant from the David and Lucile Packard Foundation

Acknowledgements

I would like to thank Carl Fleischhauer and Prosser Gifford for the opportunity to learn about areas of human activity unknown to me a scant ten months ago, and the David and Lucile Packard Foundation for supporting that opportunity. The help given by others is acknowledged on a separate page.

19 October 1992

INTRODUCTION

The Workshop on Electronic Texts (1) drew together representatives of various projects and interest groups to compare ideas, beliefs, experiences, and, in particular, methods of placing and presenting historical textual materials in computerized form. Most attendees gained much in insight and outlook from the event. But the assembly did not form a new nation, or, to put it another way, the diversity of projects and interests was too great to draw the representatives into a cohesive, action-oriented body.(2)

Everyone attending the Workshop shared an interest in preserving and providing access to historical texts. But within this broad field the attendees represented a variety of formal, informal, figurative, and literal groups, with many individuals belonging to more than one. These groups may be defined roughly according to the following topics or activities:

- Imaging
- Searchable coded texts
- National and international computer networks

- CD-ROM production and dissemination
- Methods and technology for converting older paper materials into electronic form
- Study of the use of digital materials by scholars and others

This summary is arranged thematically and does not follow the actual sequence of presentations.

NOTES:
(1) In this document, the phrase electronic text is used to mean any computerized reproduction or version of a document, book, article, or manuscript (including images), and not merely a machine-readable or machine-searchable text.

(2) The Workshop was held at the Library of Congress on 9-10 June 1992, with funding from the David and Lucile Packard Foundation. The document that follows represents a summary of the presentations made at the Workshop and was compiled by James DALY. This introduction was written by DALY and Carl FLEISCHHAUER.

PRESERVATION AND IMAGING

Preservation, as that term is used by archivists,(3) was most explicitly discussed in the context of imaging. Anne KENNEY and Lynne PERSONIUS explained how the concept of a faithful copy and the user-friendliness of the traditional book have guided their project at Cornell University.(4) Although interested in computerized dissemination, participants in the Cornell project are creating digital image sets of older books in the public domain as a source for a fresh paper facsimile or, in a future phase, microfilm. The books returned to the library shelves are high-quality and useful replacements on acid-free paper that should last a long time. To date, the Cornell project has placed little or no emphasis on creating searchable texts; one would not be surprised to find that the project participants view such texts as new editions, and thus not as faithful reproductions.

In her talk on preservation, Patricia BATTIN struck an ecumenical and flexible note as she endorsed the creation and dissemination of a variety of types of digital copies. Do not be too narrow in defining what counts as a preservation element, BATTIN counseled; for the present, at least, digital copies made with preservation in mind cannot be as narrowly standardized as, say, microfilm copies with the same objective. Setting standards precipitously can inhibit creativity, but delay can result in chaos, she advised.

In part, BATTIN's position reflected the unsettled nature of image-format standards, and attendees could hear echoes of this unsettledness in the comments of various speakers. For example, Jean BARONAS reviewed the status of several formal standards moving through committees of experts; and Clifford LYNCH encouraged the use of a new guideline for transmitting document images on Internet. Testimony from participants in the National Agricultural Library's (NAL) Text Digitization Program and LC's American Memory project highlighted some of the challenges to the actual creation or interchange of images, including difficulties in converting preservation microfilm to digital form. Donald WATERS reported on the progress of a master plan for a project at Yale University to convert books on microfilm to digital image sets, Project Open Book (POB).

The Workshop offered rather less of an imaging practicum than planned, but "how-to" hints emerge at various points, for example, throughout KENNEY's presentation and in the discussion of arcana such as thresholding and dithering offered by George THOMA and FLEISCHHAUER.

NOTES:
(3) Although there is a sense in which any reproductions of historical materials preserve the human record, specialists in the field have developed particular guidelines for the creation of acceptable preservation copies.

(4) Titles and affiliations of presenters are given at the beginning of their respective talks and in the Directory of Participants (Appendix III).

THE MACHINE-READABLE TEXT: MARKUP AND USE

The sections of the Workshop that dealt with machine-readable text tended to be more concerned with access and use than with preservation, at least in the narrow technical sense. Michael SPERBERG-McQUEEN made a forceful presentation on the Text Encoding Initiative's (TEI) implementation of the Standard Generalized Markup Language (SGML). His ideas were echoed by Susan HOCKEY, Elli MYLONAS, and Stuart WEIBEL. While the presentations made by the TEI advocates contained no practicum, their discussion focused on the value of the finished product, what the European Community calls reusability, but what may also be termed durability. They argued that marking up--that is, coding--a text in a well-conceived way will permit it to be moved from one computer environment to another, as well as to be used by various users. Two kinds of markup were distinguished: 1) procedural markup, which describes the features of a text (e.g., dots on a page), and 2) descriptive markup, which describes the structure or elements of a document (e.g., chapters, paragraphs, and front matter).

The TEI proponents emphasized the importance of texts to scholarship. They explained how heavily coded (and thus analyzed and annotated) texts can underlie research, play a role in scholarly communication, and facilitate classroom teaching. SPERBERG-McQUEEN reminded listeners that a written or printed item (e.g., a particular edition of a book) is merely a representation of the abstraction we call a text. To concern ourselves with faithfully reproducing a printed instance of the text, SPERBERG-McQUEEN argued, is to concern ourselves with the representation of a representation (“images as simulacra for the text”). The TEI proponents’ interest in images tends to focus on corollary materials for use in teaching, for example, photographs of the Acropolis to accompany a Greek text.

By the end of the Workshop, SPERBERG-McQUEEN confessed to having been converted to a limited extent to the view that electronic images constitute a promising alternative to microfilming; indeed, an alternative probably superior to microfilming. But he was not convinced that electronic images constitute a serious attempt to represent text in electronic form. HOCKEY and MYLONAS also conceded that their experience at the Pierce Symposium the previous week at Georgetown University and the present conference at the Library of Congress had compelled them to reevaluate their perspective on the usefulness of text as images. Attendees could see that the text and image advocates were in constructive tension, so to say.

Three nonTEI presentations described approaches to preparing machine-readable text that are less rigorous and thus less expensive. In the case of the Papers of George Washington, Dorothy TWOHIG explained that the digital version will provide a not-quite-perfect rendering of the transcribed text--some 135,000 documents, available for research during the decades while the perfect or print version is completed.
Members of the American Memory team and the staff of NAL’s Text Digitization Program (see below) also outlined a middle ground concerning searchable texts. In the case of American Memory, contractors produce texts with about 99-percent accuracy that serve as “browse” or “reference” versions of written or printed originals. End users who need faithful copies or perfect renditions must refer to accompanying sets of digital facsimile images or consult copies of the originals in a nearby library or archive. American Memory staff argued that the high cost of producing 100-percent accurate copies would prevent LC from offering access to large parts of its collections.

THE MACHINE-READABLE TEXT: METHODS OF CONVERSION

Although the Workshop did not include a systematic examination of the methods for converting texts from paper (or from facsimile images) into machine-readable form, nevertheless, various speakers touched upon this matter. For example, WEIBEL reported that OCLC has experimented with a merging of multiple optical character recognition systems that will reduce errors from an unacceptable rate of 5 characters out of every l,000 to an unacceptable rate of 2 characters out of every l,000.

Pamela ANDRE presented an overview of NAL's Text Digitization Program and Judith ZIDAR discussed the technical details. ZIDAR explained how NAL purchased hardware and software capable of performing optical character recognition (OCR) and text conversion and used its own staff to convert texts. The process, ZIDAR said, required extensive editing and project staff found themselves considering alternatives, including rekeying and/or creating abstracts or summaries of texts. NAL reckoned costs at $7 per page. By way of contrast, Ricky ERWAY explained that American Memory had decided from the start to contract out conversion to external service bureaus. The criteria used to select these contractors were cost and quality of results, as opposed to methods of conversion. ERWAY noted that historical documents or books often do not lend themselves to OCR. Bound materials represent a special problem. In her experience, quality control--inspecting incoming materials, counting errors in samples--posed the most time-consuming aspect of contracting out conversion. ERWAY reckoned American Memory's costs at $4 per page, but cautioned that fewer cost-elements had been included than in NAL's figure.

OPTIONS FOR DISSEMINATION

The topic of dissemination proper emerged at various points during the Workshop. At the session devoted to national and international computer networks, Clifford LYNCH, Howard BESSER, Ronald LARSEN, and Edwin BROWNRIGG highlighted the virtues of Internet today and of the network that will evolve from Internet. Listeners could discern in these narratives a vision of an information democracy in which millions of citizens freely find and use what they need. LYNCH noted that a lack of standards inhibits disseminating multimedia on the network, a topic also discussed by BESSER. LARSEN addressed the issues of network scalability and modularity and commented upon the difficulty of anticipating the effects of growth in orders of magnitude. BROWNRIGG talked about the ability of packet radio to provide certain links in a network without the need for wiring.

However, the presenters also called attention to the shortcomings and incongruities of present-day computer networks. For example: 1) Network use is growing dramatically, but much network traffic consists of personal communication (E-mail). 2) Large bodies of information are available, but a user's ability to search across their entirety is limited. 3) There are significant resources for science and technology, but few network sources provide content in the humanities. 4) Machine-readable texts are commonplace, but the capability of the system to deal with images (let alone other media formats) lags behind. A glimpse of a multimedia future for networks, however, was provided by
Maria LEBRON in her overview of the On-line Journal of Current Clinical Trials (OJCCT), and the process of scholarly publishing on-line.

The contrasting form of the CD-ROM disk was never systematically analyzed, but attendees could glean an impression from several of the show-and-tell presentations. The Perseus and American Memory examples demonstrated recently published disks, while the descriptions of the IBYCUS version of the Papers of George Washington and Chadwyck-Healey's Patrologia Latina Database (PLD) told of disks to come. According to Eric CALALUCA, PLD's principal focus has been on converting Jacques-Paul Migne's definitive collection of Latin texts to machine-readable form. Although everyone could share the network advocates' enthusiasm for an on-line future, the possibility of rolling up one's sleeves for a session with a CD-ROM containing both textual materials and a powerful retrieval engine made the disk seem an appealing vessel indeed. The overall discussion suggested that the transition from CD-ROM to on-line networked access may prove far slower and more difficult than has been anticipated.

WHO ARE THE USERS AND WHAT DO THEY DO?

Although concerned with the technicalities of production, the Workshop never lost sight of the purposes and uses of electronic versions of textual materials. As noted above, those interested in imaging discussed the problematical matter of digital preservation, while the TEI proponents described how machine-readable texts can be used in research. This latter topic received thorough treatment in the paper read by Avra MICHELSON. She placed the phenomenon of electronic texts within the context of broader trends in information technology and scholarly communication.

Among other things, MICHELSON described on-line conferences that represent a vigorous and important intellectual forum for certain disciplines. Internet now carries more than 700 conferences, with about 80 percent of these devoted to topics in the social sciences and the humanities. Other scholars use on-line networks for "distance learning." Meanwhile, there has been a tremendous growth in end-user computing; professors today are less likely than their predecessors to ask the campus computer center to process their data. Electronic texts are one key to these sophisticated applications, MICHELSON reported, and more and more scholars in the humanities now work in an on-line environment. Toward the end of the Workshop, Michael LESK presented a corollary to MICHELSON's talk, reporting the results of an experiment that compared the work of one group of chemistry students using traditional printed texts and two groups using electronic sources. The experiment demonstrated that in the event one does not know what to read, one needs the electronic systems; the electronic systems hold no advantage at the moment if one knows what to read, but neither do they impose a penalty.

DALY provided an anecdotal account of the revolutionizing impact of the new technology on his previous methods of research in the field of classics. His account, by extrapolation, served to illustrate in part the arguments made by MICHELSON concerning the positive effects of the sudden and radical transformation being wrought in the ways scholars work.

Susan VECCIA and Joanne FREEMAN delineated the use of electronic materials outside the university. The most interesting aspect of their use, FREEMAN said, could be seen as a paradox: teachers in elementary and secondary schools requested access to primary source materials but, at the same time, found that "primariness" itself made these materials difficult for their students to use.

OTHER TOPICS

Marybeth PETERS reviewed copyright law in the United States and offered advice during a lively discussion of this subject. But uncertainty remains concerning the price of copyright in a digital medium, because a solution remains to be worked out concerning management and synthesis of copyrighted and out-of-copyright pieces of a database.

As moderator of the final session of the Workshop, Prosser GIFFORD directed discussion to future courses of action and the potential role of LC in advancing them. Among the recommendations that emerged were the following:

* Workshop participants should 1) begin to think about working with image material, but structure and digitize it in such a way that at a later stage it can be interpreted into text, and 2) find a common way to build text and images together so

that they can be used jointly at some stage in the future, with appropriate network support, because that is how users will want to access these materials. The Library might encourage attempts to bring together people who are working on texts and images.

* A network version of American Memory should be developed or consideration should be given to making the data in it available to people interested in doing network multimedia. Given the current dearth of digital data that is appealing and unencumbered by extremely complex rights problems, developing a network version of American Memory could do much to help make network multimedia a reality.

* Concerning the thorny issue of electronic deposit, LC should initiate a catalytic process in terms of distributed responsibility, that is, bring together the distributed organizations and set up a study group to look at all the issues related to electronic deposit and see where we as a nation should move. For example, LC might attempt to persuade one major library in each state to deal with its state equivalent publisher, which might produce a cooperative project that would be equitably distributed around the country, and one in which LC would be dealing with a minimal number of publishers and minimal copyright problems. LC must also deal with the concept of on-line publishing, determining, among other things, how serials such as OJCCT might be deposited for copyright.

* Since a number of projects are planning to carry out preservation by creating digital images that will end up in on-line or near-line storage at some institution, LC might play a helpful role, at least in the near term, by accelerating how to catalog that information into the Research Library Information Network (RLIN) and then into OCLC, so that it would be accessible. This would reduce the possibility of multiple institutions digitizing the same work.

CONCLUSION

The Workshop was valuable because it brought together partisans from various groups and provided an occasion to compare goals and methods. The more committed partisans frequently communicate with others in their groups, but less often across group boundaries. The Workshop was also valuable to attendees--including those involved with American Memory--who came less committed to particular approaches or concepts. These attendees learned a great deal, and plan to select and employ elements of imaging, text-coding, and networked distribution that suit their respective projects and purposes.

Still, reality rears its ugly head: no breakthrough has been achieved. On the imaging side, one confronts a proliferation of competing data-interchange standards and a lack of consensus on the role of digital facsimiles in preservation. In the realm of machine-readable texts, one encounters a reasonably mature standard but methodological difficulties and high costs. These latter problems, of course, represent a special impediment to the desire, as it is sometimes expressed in the popular press, "to put the [contents of the] Library of Congress on line." In the words of one participant, there was "no solution to the economic problems--the projects that are out there are surviving, but it is going to be a lot of work to transform the information industry, and so far the investment to do that is not forthcoming" (LESK, per litteras).

The Library of Congress Information System (LOCIS) is now available over the Internet.

The telnet address is:

locis.loc.gov
140.147.254.3

LOCIS accepts both telnet 3270 and line mode.

LOCIS includes over 15 million catalog records and over 10 million records for other types of information: federal legislation, copyright registrations, Braille and audio, organizations, and selected foreign legal materials.

Searching hours are (all times USA eastern; closed national holidays):

Monday - Friday: 6:30am - 9:30pm
Saturday: 8:00am - 5:00pm
Sunday: 1:00pm - 5:00pm

Printed manuals will be available for sale later this summer and very soon via FTP (ftp seq1.loc.gov /pub/LC.On-line). There will be a LOCIS Quick Search Guide and a LOCIS Reference Manual. LC On-line Internet: lcon-line@seq1.loc.gov Library of Congress

Stan Horwitz Internet: STAN@VM.TEMPLE.EDU Bitnet: STAN@TEMPLEVM

Temple University's Sr. Mainframe Consultant; Manager of the Help-Net and Suggest lists; Listserv Postmaster (Standard disclaimers apply)