

# Digital Libraries: Virginia Tech Courseware

To learn about digital libraries, you may wish to visit the Self-Study course materials or the online courses listed below.

- [Self-Study Courseware](#)
- [Honors 3004](#): Digital Libraries, Fall 1997, Virginia Tech
- [CS6604](#): Digital Libraries, Fall 1997, Virginia Tech

Please send comments/suggestions to [Ed Fox](#).

## Self-study Courseware on

# Digital Libraries

### [Contents](#)

**Introduction:** This WWW site has been developed to assist those interested in learning about digital libraries. It is based upon materials tested in 2 Virginia Tech courses taught Fall 1997:

- [CS6604](#)
- [Honors 3004](#)

Students in those courses especially liked Michael Lesk's "[Practical Digital Libraries: Books, Bytes & Bucks](#)" so we refer to it as a supplemental text throughout this site.

There is a set of [quizzes](#) (to be added) to test your knowledge of the chapters in Dr. Lesk's book. We also will support discussion related to these course materials through:

- Listserv (to be added)
- Hypernews (to be added)

---

**Revisions:** This site will undergo frequent changes, so do check back. The latest revision was completed 6/27/98.

**Acknowledgements:** This WWW site was developed in part through funding from NSF grants CDA-9312611, DUE-9752408, and DUE-9752190.

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

[Booksellers](#)[Orders & Inquiries](#)[Catalog](#)[New & Noteworthy](#)[Site Index](#)

## Practical Digital Libraries: Books, Bytes, and Bucks

Michael Lesk

The Morgan Kaufmann Series in  
Multimedia Information and Systems,  
Edward Fox, Series Editor

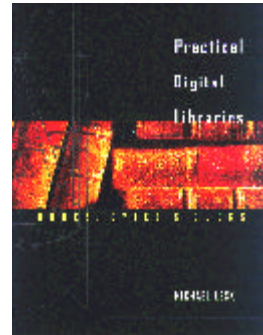
1997

320 pages

cloth

\$49.95

ISBN 1-55860-459-6



[Order This Book](#) | [Authors](#) | [Contents](#) | [Related Titles](#)

**"While others speculate on the organizational dilemmas facing libraries and universities in a digital information environment or argue the relative philosophical merits of print versus digital media, Lesk has constructed an on-the-ground picture of the various working components of the digital environment. Viewed as a whole and with an engineering sense of composition, his picture is remarkable--almost astonishing--because it reveals how advanced the digital environment has truly become."**

**--Donald J. Waters**

**Associate University Librarian, Yale University**

A digital library is not merely a collection of electronic information. It is an organized and digitized system of data that can serve as a rich resource for its user community. This authoritative and accessible guide for librarians and computer scientists explores the technologies behind digital libraries, the choices to be made in building them, and the economic and policy structures that affect them.

The most comprehensive book on the subject, *Practical Digital Libraries*

- offers the most wide-ranging overview of digital libraries currently available
- analyzes economic and intellectual issues in the emerging digital environment

- shows how text, images, audio, and video can be represented, distributed, used, and collected as forms of knowledge

**Authors:**

[Michael Lesk](#) joined the computer science research group at Bell Laboratories after receiving his Ph.D. degree in Chemical Physics in 1969. He went on to manage the computer science research group at Bellcore, where he is now a chief research scientist. He is best known for his work in electronic libraries, but has worked in document production and retrieval software, computer networks, computer languages, and human-computer interfaces as well. Past chair of the Association for Computing Machinery's special interest groups on Language Analysis and Information Retrieval, Lesk was Senior Visiting Fellow of the British Library in 1987 and is currently Visiting Professor of Computer Science at University College London.

**Table of Contents:**

- [1 Evolution of Libraries](#)
- [2 Text Access Methods](#)
- [3 Images of Pages](#)
- [4 Multimedia Storage and Access](#)
- [5 Knowledge Representation Methods](#)
- [6 Distribution](#)
- [7 Usability and Retrieval Evaluation](#)
- [8 Collections and Preservation](#)
- [9 Economics](#)
- [10 Intellectual Property Rights](#)
- [11 International Activities](#)
- [12 Future: Ubiquity, Diversity, Creativity, and Public Policy](#)

Instructors are invited to [request an examination copy](#).

**Related Titles:**

[Multimedia Information & Systems Database](#)



Copyright © 1997, Morgan Kaufmann Publishers, Inc.  
Telephone: (415) 392-2665 Email: [mkp@mkp.com](mailto:mkp@mkp.com)

## Contents :

---

- [Introduction to Digital Libraries](#): This holds general information such as definitions, glossary of digital library terms, foundations and scenarios.
  - [Topics](#): This contains information classified under various topics of/related to Digital Libraries e.g. "Metadata" etc.
  - [Resources](#): Provides other information based under more general headings such as various people involved in Digital Libraries, projects, countries and regions etc.
  - [References](#): This category contains references, links and pointers such as conferences/workshops, journals and books, and various related courses being conducted at different universities.
- 

### Pedagogy:

We recommend that beginners start with the Introduction and then proceed through the Topics, following along with the text by Dr. Lesk. The Resources provide alternate views of the contents, and the References should serve those desiring additional details.

---

[\[Main\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

# Introduction to Digital Libraries:

---

- [Definitions](#): Some of the attempts made by various people to define a digital library.
- [Foundations](#): Introductory material related to digital libraries...
- [Scenarios and Perspectives](#): Various scenarios and perspectives that arise in a Digital Library context.

---

[\[Main\]](#) [\[Contents\]](#)

---

Please send comments/suggestions to [Ed Fox](#). (c) Copyright 1998, Edward A. Fox, Rajat Gupta

## Definitions :

---

- "The generic name for federated structures that provide humans both intellectual and physical access to the huge and growing worldwide networks of information encoded in multimedia digital formats."  
([The University of Michigan Digital Library: This Is Not Your Father's Library](#), [Birmingham](#), 1994)
- "Systems providing a community of users with coherent access to a large, organized repository of information and knowledge."  
([Lynch](#), 1995)
- "Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds; static or dynamic images) and exist in distributed networks. The content of digital libraries includes data, metadata that describe various aspects of the data (e.g., representation, creator, owner, reproduction rights), and metadata that consist of links or relationships to other data or metadata, whether internal or external to the digital library."  
([UCLA-NSF Social Aspects of Digital Libraries Workshop](#))
- Digital libraries are constructed -- collected and organized -- by a community of users, and their functional capabilities support the information needs and uses of that community. They are a component of communities in which individuals and groups interact with each other, using data, information, and knowledge resources and systems. In this sense they are an extension, enhancement, and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved, and accessed in support of a user community. These information institutions include, among others, libraries, museums, archives, and schools, but digital libraries also extend and serve other community settings, including classrooms, offices, laboratories, homes, and public spaces." ([UCLA-NSF Social Aspects of Digital Libraries Workshop](#))
- "systems providing a community of users with coherent access to a large, organized repository of information and knowledge. This organization of information is characterized by the absence of prior detailed knowledge of the uses of the information. The ability of the user to access, reorganize, and utilize this repository is enriched by the capabilities of digital technology" (adapted from [Inter operability, Scaling, and the Digital Libraries Research Agenda](#))
- "Digital library is a concept that has different meanings in different communities. To the engineering and computer science community, digital library is a metaphor for the new kinds of distributed data base services that manage unstructured multimedia data. To the political and business communities, the term represents a new marketplace for the world's information resources and services. To futurist communities, digital libraries represent the manifestation of Wells' World Brain. The perspective taken here is rooted in an information science tradition."  
([Gary Marchionini](#))
- "A digital library is a distributed technology environment which dramatically reduces barriers to the creation, dissemination, manipulation, storage, integration, and reuse of information by individuals and groups."  
([Edward A. Fox](#) , editor, Source Book on Digital Libraries, pg. 65)
- "A digital library is a machine readable representation of materials which might be found in a

university library together with organizing information intended to help users find specific information. A digital library service is an assemblage of digital computing, storage, and communicate machinery together with the software needed to reprise, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, storing, cataloging, finding, and disseminating information."

([Edward A. Fox](#) , editor, Source Book on Digital Libraries, pg. 65)

- "an organized data base of digital information objects in varying formats maintained to provide unmediated ease of access to a user community, with these further characteristics:
  - an overall access tool (e.g. a catalog) provides search and retrieval capability over the entire data base;
  - organized technical procedures exist through which the library management adds objects to the data base and removes them according to a coherent and accessible collections policy."

([Peter Graham](#), Rutgers University Libraries)

- "A library that has been extended and enhanced by the application of digital technology. Important aspects of the digital library that may be extended and enhanced include :
  - Collections of the library
  - Organization and management of the collections
  - Access of the library items and the processing of the information contained in the items
  - Communication of information about the items "

([Smith](#), 1995)

## Digital Library related terms/glossary

(by [Peter Graham](#), Rutgers University Libraries):

- digital archive: a digital library which is intended to be maintained for a long time, i.e. periods longer than individual human lives and certainly longer than individual technological epochs. (Sometimes formerly also "digital research library.")
- digital preservation: preservation of artifactual information by digitizing its image (e.g. scanning a manuscript page, digitally photographing a vase, or converting a cylinder recording to digital form).
- electronic preservation: preservation of information that is in digital (that is, electronic) form, i.e. the techniques associated with refreshing, migration and assurance of integrity.

## Digital Preservation techniques:

- Refresh: to copy digital information from one long-term storage medium to another of the same type, with no change whatsoever in the bit stream (e.g. from a decaying 800 bpi tape to a new 800 bpi tape, or from an older 5 1/4" floppy to a new 5 1/4" floppy).
- "Modified refreshing" is the copying to another medium of a similar enough type hat no change is made in the bit pattern that is of concern to the application and operating system using the data, e.g. from an 800 bpi tape to a 1600 bpi tape or to a "square", cartridge, tape; or from a 5 1/4" floppy disk to a 3 1/2" floppy disk.
- Migrate: to copy data, or convert data, from one technology to another, whether hardware or

software, preserving the essential characteristics of the data; generally forward in time. (At the moment, it is recognized, this final qualifier begs many questions.) Examples: conversion of XyWrite w/p files to Microsoft Word; conversion of ClarisWorks v3 spreadsheet files to Microsoft Excel v4 files; conversion of binary tape images of survey research multi-punched tab cards to a data base format; copying an 800 bpi tape file to a sequential disk file; converting a DOS FoxPro data base to a Visual Basic database for Windows 95; converting a PICT image to a TIFF image; converting a ClarisWorks for Windows v4 w/p file to a Macintosh ClarisWorks v4 file.

Examples can be given, as here, for cases known to be required; the longer term preservation problem is to prepare for forward migrations when the future technologies are unknown.

- Emulate: (find and use better Comp SCI terms here, probably) in hardware terms, the creation of software for a computer that reproduces in all essential characteristics (as defined by the problem to be solved) the performance of another computer of a different design. Computers may emulate earlier computers in order to provide backward compatibility, or may emulate a future computer in order to provide a software development environment while the newer computer is still being fabricated.

In software preservation terms, the creation of software that analyzes the software environment of a document such that it can provide a user interface to the document that substantially reproduces the essential characteristics of the document as it was created by its originating software.

- Document: (use sense that Apple began to use, with Macintosh; anything manipulated by an application; find their definition and build on it. Note Dublin Core [and other] use of "document like object").
- Authenticate: of users, to verify that network users are in fact who they identify themselves to be; of documents, to validate the integrity of a document with respect to its original authorized creation.
- Authentication: (of a resource--i.e. of data, not people)
- Authenticity: (of a resource--i.e. of data, not people)
- Integrity: synonym of authenticity (of a resource--i.e. of data, not people)

---

[\[Main\]](#) [\[Introduction\]](#) [\[Contents\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## Foundations (see Lesk Ch. 1, 8):

---

- [As We May Think](#) by Vannevar Bush - the visionary article that helped motivate early work on digital libraries, hypertext and information retrieval
  - What is a "[digital library](#)"? (vs. a virtual library)
  - UCLA workshop (focusing on user perspectives):
    - [Introduction](#)
    - [information life cycle](#)
    - [Artists](#)
    - [Business Records as Artifacts](#)
    - [Health-Information Systems](#)
  - IITA workshop: [Definitions and Roles of Digital Libraries](#)
  - [Digital Libraries: Issues and Architectures](#)
  - [Digital Library: Gross Structure and Requirements: Report from a March 1994 Workshop](#). Also available in [PDF](#).
- 

## Pedagogy:

We recommend that the above items be skimmed to obtain a general background regarding digital library research, development, and practice. Please also read chapters 1 and 8 of Dr. Lesk's book.

---

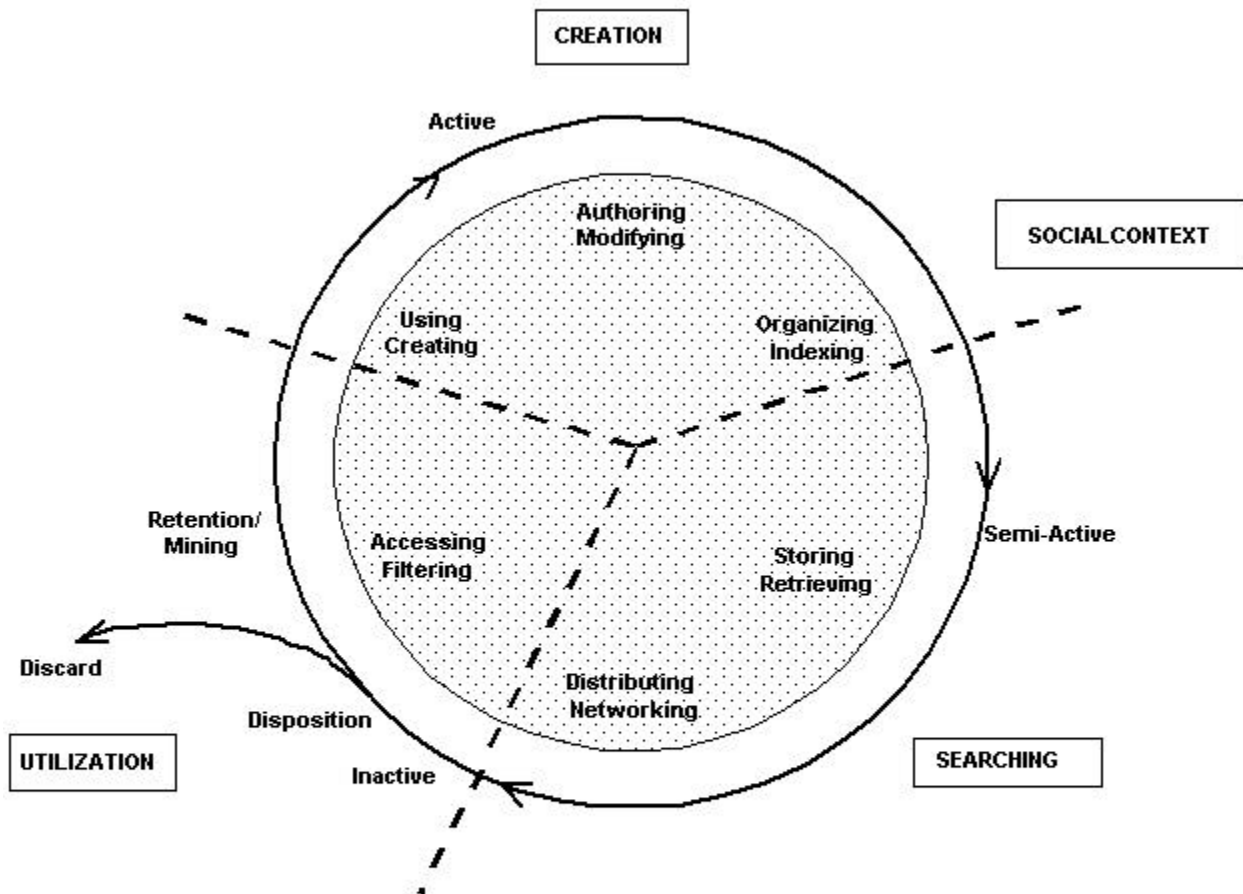
[\[Main\]](#) [\[Contents\]](#) [\[Introduction\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

### Information Life Cycle



NOTE: The outer ring indicates the life cycle stages (active, semi-active, and inactive) for a given type of information artifact (such as business records, artworks, documents, or scientific data). The stages are superimposed on six types of information uses or processes (shaded circle). The cycle has three major phases: information creation, searching, and utilization. The alignment of the cycle stages with the steps of information handling and process phases may vary according to the particular social or institutional context.

# Interoperability, Scaling, and the Digital Libraries Research Agenda:

A Report on the May 18-19, 1995

IITA Digital Libraries Workshop

August 22, 1995

Clifford Lynch ( [clifford.lynch@ucop.edu](mailto:clifford.lynch@ucop.edu) )

Hector Garcia-Molina ( [hector@db.stanford.edu](mailto:hector@db.stanford.edu) )

*Converted to HTML using GradStudentWare 2.2*

*Contact [Christian Mogensen](#) with bug reports.*

[Introduction](#)

[Definitions and Roles of Digital Libraries](#)

[Defining Interoperability in the Digital Library Environment](#)

[Infrastructure Requirements for Digital Library Research](#)

[Research Issues and Priorities](#)

[1. Interoperability](#)

[2. Description of Objects and Repositories](#)

[3. Collection Management and Organization](#)

[4. User Interfaces and Human-Computer Interaction](#)

[Conclusions](#)

[Executive Summary](#)

[Appendix 1 - List of Participants](#)

[Appendix 2 - Strawman Report](#)

[Appendix 3 - Report of the working groups](#)

[3-1 - The Publishing Perspective](#)

[3-2 - The Commercial Perspective](#)

[3-3 - The Library Perspective](#)

[3-4 - The Internet Perspective](#)

[3-5 - The Multimedia Perspective](#)

---

## Introduction

This report summarizes the results of a workshop on Digital Libraries held under the auspices of the U.S. Government's Information Infrastructure Technology and Applications (IITA) Working Group in Reston, Virginia on May 18-19, 1995. The objective of the workshop was to refine the research agenda for digital libraries with specific emphasis on issues of scaling and interoperability, and to identify the infrastructure developments needed to make progress on these issues.

While there have been a number of workshops and other meetings examining the broader questions of support for applications in the National Information Infrastructure (NII), we believe this was the first workshop that focused specifically on Digital Libraries in this context. In the past year, Digital Libraries



# Digital Libraries: Issues and Architectures

*Peter J. Nürnberg*

*Richard Furuta*

*John J. Leggett*

*Catherine C. Marshall*

*Frank M. Shipman III*

Center for the Study of Digital Libraries

Texas A&M University

College Station, TX 77843

USA

{pnuern, furuta, leggett, marshall, shipman}@bush.cs.tamu.edu

## ABSTRACT

The research field of digital libraries must be viewed as a union of subfields from a variety of domains combined with new research issues in order to realize its full potential. A clear exposition of the research issues involved has not yet been given. Most approaches to building digital library systems have thus far been limited to addressing specific digital library problems as variations of problems from other fields. This paper presents a taxonomy of digital library elements. Consideration of the elements in this taxonomy helps suggest a variety of issues. Example elements and some issues they suggest are used to populate the taxonomy. The paper continues by presenting a general digital library system architecture. Issues suggested by the taxonomy are shown to have implications at many levels of digital library system architectures for both design and implementation. This is illustrated by considering the implications of one issue (personalizing presentations) at several architectural levels and in the context of a set of current technologies.

**Keywords:** digital library issues, digital library architecture, databases, physical libraries, World Wide Web

## INTRODUCTION

The emerging field of digital libraries brings together participants from many existing areas of research. Currently, the field lacks a clear agenda independent of these other areas. It is tempting for researchers to think that the field of digital libraries is a natural outgrowth of an already known field. From a database or information retrieval perspective, digital libraries may be seen as a form of federated databases. From a hypertext perspective the field of digital libraries could seem like a particular application of hypertext technology. From a wide-area information service perspective, digital libraries could appear to be one use of the World Wide Web. From a library science perspective, digital libraries might be seen as continuing a trend toward library automation. There is some truth to these perspectives (as well as others) but none address the field as a whole and its research agenda. The field of digital libraries will be

## Defining Scenarios & Perspectives:

---

- [Publishing](#)
  - [Commercial](#)
  - [Library](#)
  - [Internet](#)
  - [Multimedia](#)
- 

### Pedagogy:

We recommend that the scenarios given be examined, especially for the group in which the reader fits.

---

[\[Main\]](#) [\[Contents\]](#) [\[Introduction\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. FOX, Rajat Gupta

## Contents :

---

- [Introduction to Digital Libraries](#): This holds general information such as definitions, glossary of digital library terms, foundations and scenarios.
  - [Topics](#): This contains information classified under various topics of/related to Digital Libraries e.g. "Metadata" etc.
  - [Resources](#): Provides other information based under more general headings such as various people involved in Digital Libraries, projects, countries and regions etc.
  - [References](#): This category contains references, links and pointers such as conferences/workshops, journals and books, and various related courses being conducted at different universities.
- 

### Pedagogy:

We recommend that beginners start with the Introduction and then proceed through the Topics, following along with the text by Dr. Lesk. The Resources provide alternate views of the contents, and the References should serve those desiring additional details.

---

[\[Main\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## Resources:

---

- [Projects](#)
- [People](#)
- [Countries and regions](#)
- [Centers, sites and organizations](#)

---

[\[Main\]](#) [\[Contents\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. fox, Rajat Gupta**

# Projects:

---

## DLI

- [home page](#)
- [information & resources](#)
- [publications](#)
- [Carnegie Mellon University](#)
- [Stanford University](#)
- [University of California at Berkeley](#)
- [University of California at Santa Barbara](#)
- [University of Illinois](#)
- [University of Michigan](#)

## [American Memory Project \(Library of Congress\)](#)

## [UK Electronic Library Programme](#)

## Virginia Tech Projects:

- [Interactive Courseware on Digital Libraries](#) (this site itself is a part of it)
- **Interactive Learning with a Digital Library in CS** <http://ei.cs.vt.edu/>
  - Interactive Learning with a Digital Library in CS arch <http://ei.cs.vt.edu/~cs5604/Adv/Adv-ILDLCS.html>
  - Courseware <http://ei.cs.vt.edu/courses.html>
  - [Project Overview \(for FIE'96, in PDF\)](#)
  - [Project Interim Report, Oct. 1996](#)
  - [Project Report for NSF EI PI Meeting, Nov. 1996](#)
- **Envision (CS literature)** <http://ei.cs.vt.edu/~cs5604/Adv/Adv-Envision.html>
  - Envision report <http://ei.cs.vt.edu/papers/ENVreport/final.html>
- **CODER** <http://ei.cs.vt.edu/~cs5604/Adv/Adv-CODER.html>
- **MARIAN**
  - overview <http://ei.cs.vt.edu/~cs5604/Adv/Adv-MARIAN.html>
  - system <http://opac3.cc.vt.edu/htbin/marian>

## Singapore Network: [SINGAREN](#)

## Some Extra Virginia Tech resources on various projects:

- Build upon existing electronic materials
  - Netlib (numerical analysis) <http://www.netlib.org/>
    - Attribute/value search [http://www.netlib.org/utk/misc/netlib\\_query.html](http://www.netlib.org/utk/misc/netlib_query.html)
- Build upon publishers collections
  - AAAS - Science Online <http://www.aaas.org/>
  - ACM DL <http://www.acm.org/dl/>
  - ACS (Chemistry) - Online <http://www.acs.org/>
    - CORE Overview <http://ei.cs.vt.edu/~cs5604/DL/DL2.html>

- D-Lib Magazine, Dec. 1995, Making a Digital Library, Chemistry Online Retrieval Experiment <http://www.dlib.org/dlib/december95/briefings/12core.html>
  - CORE at OCLC <http://www.oclc.org:5047/oclc/research/projects/core/>
- Elsevier
  - Science Direct <http://www.elsevier.nl/>
  - TULIP (material science & engineering) homepage <http://www.elsevier.nl/inca/homepage/about/resproj/tulip.shtml>
  - With universities + OCLC
- Highwire Press
- IEEE
- IEEE CS
- JSTOR
- Commercial services and systems
  - IBM <http://www.software.ibm.com/is/dig-lib/>
    - Version 2 <http://www.software.ibm.com/is/dig-lib/v2factsheet/>
    - collection treasury <http://www.software.ibm.com/is/dig-lib/treasury/>
    - images - QBIC <http://www.qbic.almaden.ibm.com/>
    - news archive <http://www.software.ibm.com/is/dig-lib/newsarchive/>
- Enhance WWW (hypertext):
  - HyperWave <http://www.hyperwave.de/>
  - HyperWave server features
  - HyperWave author <http://www2.iicm.edu/hyperwave/author>
  - HyperWave author features <http://www2.iicm.edu/hyperwave/author/features.html>
  - HyperWave author specs <http://www2.iicm.edu/hyperwave/author/specifications.html>
  - Harmony <http://www2.iicm.edu/harmony>
  - Harmony orientation
  - Harmony screens <http://ei.cs.vt.edu/~cs5604/Adv/Adv-Harmony.html>
  - Harmony information structuring
  - Harmony document viewers
  - Amsterdam model <http://ei.cs.vt.edu/~mm/gifs/Amsterdam-hm.html>
- Community network multimedia history
  - BEV <http://www.bev.net>
  - BEV History <http://history.bev.net/bevhist/>
    - Timeline <http://history.bev.net/bevhist/historyBase/mainTimeline.html>
    - Screen for 1992
    - Screen for Article
- Discipline - Greek Literature <http://www.perseus.tufts.edu/>
  - Evaluation - TOIS
- Discipline - Computer Science
  - Technical reports
    - WATERS - through 1995
    - CSTR <http://WWW.CNRI.Reston.VA.US/home/cstr.html>
    - NCSTRL <http://www.ncstrl.org/>
      - Search results, Search results abstract
      - Doc. thumbnails, Doc. page 1
  - Ptrs
    - DLs for CS <http://fox.cs.vt.edu/DLCS.html>
    - Dienst <http://researchsmp2.cc.vt.edu:8090/>
    - Results page, document page from search

- Genre - ETDs - electronic theses and dissertations
    - Virginia Tech <http://etd.vt.edu/>
      - Submission form <http://scholar.lib.vt.edu/cgi-bin/etd.cgi>
      - Approval form <http://etd.vt.edu/submit/approval.htm>
      - Letter to students <http://etd.vt.edu/submit/letter.htm>
      - Standards <http://etd.vt.edu/submit/mm.htm>
    - Collection <http://www.theses.org>
    - Project - Networked Digital Library of Theses and Dissertations <http://www.ndltd.org>
      - Brief description <http://www.ndltd.org/info/descr.htm>
      - D-Lib Magazine Overview September 1996  
<http://www.dlib.org/dlib/september96/theses/09fox.html>
      - D-Lib Magazine Update September 1997  
<http://www.dlib.org/dlib/september97/theses/09fox.html>
      - FIPSE proposal
        - abstract <http://www.ndltd.org/support/fipseabs.htm>
        - full-text <http://www.ndltd.org/support/fipse10.pdf>
  - [PDF PART 1 - PROJECTS](#)
    - Fig. 1: Timeline of Recent Information & DL Systems
    - Fig. 2: NCSTRL Architecture
    - NETLIB (numerical analysis)
    - CORE (chemistry)
    - TULIP (material science & engineering, with Elsevier, OCLC)
    - IBM digital libraries products and projects
    - Hyper-G/HyperWave (clients and servers)
    - BEV HistoryBase
    - CS technical reports (CS-TR, WATERS, NCSTRL) and related efforts
    - CS education (ACM literature, courseware on IR, multimedia, hypertext, history)
    - Digital Library Initiative (CMU, Michigan, Stanford, UC Berkeley, UC Santa Barbara, University of Illinois Urbana-Champaign)
    - ETD (electronic theses and dissertations - NDLTD)
  - [PDF PART 2 -SOURCES, RESEARCH](#)
    - Digital Library conferences
    - IITA meetings (e.g., May 1995 workshop)
    - Allerton Institutes (from U. Illinois, NSF)
    - D-Lib (research, magazine, working groups)
    - D-Lib research articles (architecture, metadata, URNs, use)
    - Virginia Tech information (DL page, Sourcebook)
    - Virginia Tech projects (Envision, ILDLCS, WWW traffic analysis)
    - Z39.50 (overview, OCLC, CNIDR)
    - Library of Congress
    - CNRI (architecture, handles)
    - UMBC agents
    - LIS: preservation, TEI, ...
-



**AVAILABLE RESEARCH**

**University of California at Berkeley**

Environmental Planning and  
Geographic Information Systems

**University of California at Santa Barbara**

The Alexandria Project:  
Spatially-referenced Map  
Information

**Carnegie Mellon University**

Informedia Digital Video Library

**University of Illinois at Urbana-Champaign**

Federating Repositories of  
Scientific Literature

**University of Michigan**

Intelligent Agents for Information  
Location

**Stanford University**

Interoperation Mechanisms  
Among Heterogeneous Services

DLI Project [Contacts](#)

[DLI Workshop Series](#)

[DLI Publications](#)

The Initiative's focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks -- all in user-friendly ways.

Digital Libraries basically store materials in electronic format and manipulate large collections of those materials effectively. Research into digital libraries is research into network information systems, concentrating on how to develop the necessary infrastructure to effectively mass-manipulate the information on the Net.. The key technological issues are how to search and display desired selections from and across large collections. Summaries of the six DLI projects from the May 1996, [Special Issue on Digital Libraries](#) in the Institute of Electrical and Electronics Engineers, IEEE Computer Magazine.

The magazine of digital library research, the [D-Lib Magazine](#), including the July/August 1996 issue [The DLI Testbeds: Today and Tomorrow.](#)

Digital Library conference information, publications, related projects and resources to the DLI, [Digital Library Related Information and Resources.](#)

[NSF Digital Libraries Contact](#)

National Synchronization for the Digital Library Initiative is being coordinated by the University of Illinois at Urbana-Champaign, and supported by a supplemental grant by the National Science

# Digital Library Information and Resources

Research on digital libraries encompasses a range of intertwined technical, social and political issues. One of the better descriptions of digital libraries comes from the Santa Fe Workshop on Distributed Knowledge Work Environments. "[T]he concept of a "digital library" is not merely equivalent to a digitized collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge." I have made my selections for this page on the basis of their breadth, depth, ingenuity and availability of content online.

## *Table of Contents:*

1. [The Digital Libraries Initiative \(DLI\)](#)
2. [Select Digital Library Related Projects](#)
3. [Upcoming Digital Library Conferences](#)
4. [Previous Digital Library Conferences](#)
5. [Previous Digital Library Related Conferences with Online Proceedings](#)
6. [Full Text of Other Digital Library Related Publications](#)
7. [Other Digital Library Related Resources](#)
8. [Digital Library Funding, Coordination and Policy Organizations](#)
9. [Intellectual Property](#)
10. [Human Computer Interaction \(HCI\)](#)
11. [Computer Supported Cooperative Work \(CSCW\)](#)

## The Digital Libraries Initiative

The [Digital Libraries Initiative](#) is comprised of six projects in the [joint initiative](#) of the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA) for digital libraries. These projects are developing the next generation of tools for information discovery, management, retrieval and analysis. A mostly comprehensive list [DLI publications](#) and the [DLI Workshop Series](#) are online.

The DLI projects are: [University of Illinois Urbana-Champaign](#), [Carnegie-Mellon University](#), [Stanford University](#), [University of California at Berkeley](#), [University of California at Santa Barbara](#) and [University of Michigan](#).

## Select Digital Library Related Projects

The [Interspace](#) is a long term information infrastructure research project which seeks to unify disparate distributed information resources in one coherent model. The Interspace, is a collection of interlinked information spaces where each component space contains the knowledge of a community or a subject domain.

The [Networked Computer Science Technical Reports Library](#) at Cornell University Department of Computer Science. NCSTRL is a distributed technical report library developed by the ARPA-sponsored Computer Science Technical Report Project. "NCSTRL (pronounced "ancestral") is an international collection of computer science technical reports from CS departments and industrial and government

## DLI - Carnegie Mellon:

---

- [Informedia](#)
- [NetBill](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta



The Informedia Digital Video Library is a research initiative at Carnegie Mellon University funded by the NSF, DARPA, NASA and others that studies how multimedia digital libraries can be established and used. Informedia is building a multimedia library that will consist of over one thousand hours of digital video, audio, images, text and other related materials.

Informedia's digital video library is populated by automatically encoding, segmenting and indexing data. Research in the areas of speech recognition, image understanding and natural language processing supports the automatic preparation of diverse media for full-content and knowledge-based search and retrieval. Informedia is one of six [Digital Libraries Initiative](#) projects.

---

[Project Description](#)

[Sponsors and Partnerships](#)

[Publications & Reports](#)

[Project Team](#)

[Project News](#)

[NSF Digital Library Projects](#)

---

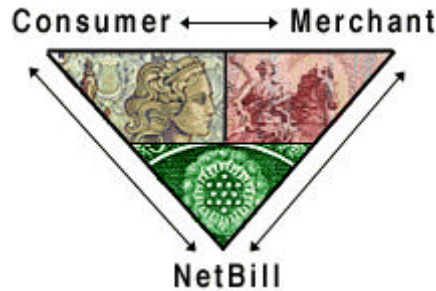
Copyright © 1997 Carnegie Mellon University  
All Rights Reserved



# The NetBill Project

- ◆ Overview
- ◆ News
- ◆ Publications
- ◆ Technical Partners
- ◆ Project Members
- ◆ Commerce Resources

*A dependable, secure, and economical payment method for purchasing digital goods and services through the Internet.*



The NetBill electronic commerce project at Carnegie Mellon's [Information Networking Institute](#) is researching design issues of highly survivable and secure distributed transaction processing systems, as well as accounting and access control for digital libraries. NetBill is addressing these issues by developing the protocols and software to support network-based payment for goods and services over the Internet.

These protocols and software have been implemented in a test system, currently in its Alpha trial, on the Carnegie Mellon campus. This system enables consumers and merchants to communicate directly with each other, using NetBill to confirm and ensure security for all transactions.

We invite you to take a look at this test system at:

<http://www.netbill.com>

NetBill is publicly available to United States residents. For those not in the US, there is plenty of information about NetBill for you to explore.

For more information about the NetBill project, please explore this web site using the links on the left of each page.

If you require further information, please contact us at [support@netbill.com](mailto:support@netbill.com)



All contents copyright © 1995,1996,1997 Carnegie Mellon University.  
All rights reserved.  
Last revision: Fri Oct 10 11:54:34 EDT 1997

## DLI - Stanford:

---

- [Home Page](#)
- [IEEE Computer article](#)
- [testbed development](#)
- [info finding](#)
- [user interfaces](#)
- [DLITE \(task env\)](#)
- [DLITE comps](#)
- [DLITE screens](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**



# Stanford University Digital Libraries Project

The Stanford Digital Libraries project is one participant in the 4-year, \$24 million Digital Library Initiative, started in 1994 and supported by the [NSF](#), [DARPA](#), and [NASA](#). In addition to the ties with the [five other universities](#) that are part of the project, Stanford also has a large number of [industrial partners](#). Each university project has a different angle of the total project, with Stanford focusing on **interoperability**.

Our collection is primarily computing literature. However, we also have a strong focus on networked information sources, meaning that the vast array of topics found on the World Wide Web are accessible through our project as well. At the heart of the project is the [testbed](#) running [the "InfoBus" protocol](#), which provides a uniform way to access a variety of services and information sources through "proxies" acting as interpreters between the InfoBus protocol and the native protocol.

With the InfoBus protocol running under the hood, a variety of user level applications provide powerful ways to [find information](#), using cutting-edge [user interfaces](#) for direct manipulation or through [Agent technology](#). A second area of focus for the Stanford Digital Library Project is the [legal and economic issues](#) of a networked environment.

---

## PROJECTS

See the entire list, or jump directly to projects related to [information finding](#), [user interfaces](#), [legal and economic issues](#), [the testbed](#), or [agents](#).

## DOCUMENTS

A collection of introductory information, and our [publications](#), our [working papers](#), our [presentations](#), our [mailing archives](#), and our [project reports](#).

## INFO RESOURCES

A collection of pointers to digital library-related resources, both at Stanford and elsewhere.

## SEMINARS

A schedule of our weekly Digital Library seminar, which meets Mondays at 4:30 in Gates B08.

## SOFTWARE

A collection of software developed for and used by the Stanford Digital Library project.

## PEOPLE

A list of the Stanford faculty, staff, and student participants and industrial partners.



**February 4, 1998**

Andreas Paepcke's [overview presentation](#) gives a good "big picture" view of the project. Gerard Rodriguez put together an [introductory page about using the InfoBus](#). There's also a [presentation](#) by Andreas Paepcke on how to build a proxy. The [slides from the most recent DLI all-project meeting](#) are online. Plus, congratulations to Martin Roscheisen, our newest Ph.D. His dissertation is entitled [A Network-Centric Design for Relationship-Based Rights Management](#).



---

Quick Tabs to Projects: [Gloss](#) -- [Query Translation](#) -- [SenseMaker](#) -- [FAB](#) -- [STARTS](#) -- [Grassroots](#) -- [SONIA](#) -- [BackRub](#) -- [Metadata Architecture](#) -- [ComMentor](#) -- [R-Manager](#) -- [InterPay](#) -- Distributed Transactions -- [InterOp Protocol](#) -- Z Server -- Proxy Generator -- [Infobus Socket Interface](#) -- [JYLU](#) -- [DLITE](#) -- [Audio HTML Access](#) -- WebWriter -- [Interbib](#) -- [SCAM](#) -- [COPS](#)

---

From *Computer* theme issue on the US Digital Library Initiative, May 1996

# Using Distributed Objects for Digital Library Interoperability

Andreas Paepcke, Steve B. Cousins, Hector Garcia-Molina, Scott W. Hassan, Steven P. Ketchpel, Martin Röscheisen, and Terry Winograd, *Stanford University*

*Distributed object technology can provide interoperability among emerging digital library services. This project uses CORBA objects as wrappers to handle differences in service interaction models.*

Information repositories are just one of many services tomorrow's digital libraries might offer. Other services include automated news summarization, trend analysis across news repositories, and copyright-related facilities. Traditional library services such as archiving and collection building will continue to be relevant as well. Archiving issues in the digital world include, for example, dangling hyperlinks and storage media obsolescence.

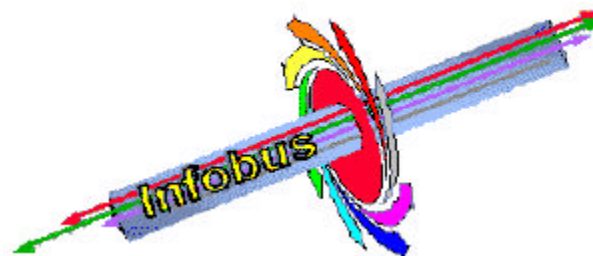
This distributed collection of services has the potential to be enormously helpful in performing information-intensive tasks. It could also turn such tasks into confusing, frustrating annoyances by forcing programmers and users to learn many interfaces and by confronting users with the bewildering details of fee-based services that were previously only accessible to professional librarians.

The Stanford Digital Library project has undertaken work to address the problem of interoperability, which is particularly important because standardization efforts are lagging behind the development of digital library services. We used CORBA,[\[1\]](#) the distributed-object standard developed by the Object Management Group, to implement information-access and payment protocols. These protocols are designed to provide the interface uniformity necessary for interoperability, while leaving implementers a large amount of leeway to optimize performance and to provide choices in service performance profiles.

We have implemented an experimental version of our information-access protocol for Knight-Ridder's Dialog information service, various World Wide Web information sources, Z39.50 servers (one of the best-known information-access protocols),[\[2\]](#) Oracle's ConText summarization tool, and others. Our implementation is based on Xerox PARC's ILU (InterLanguage Unification) facility, a public-domain implementation of CORBA. It is supported on common platforms, such as Microsoft Windows 3.1 and NT, Linux, and the Unix implementation of Sun Microsystems, IBM,

## Stanford Digital Library Testbed Development

Department of Computer Science  
Stanford University  
Stanford, CA



## What is the Stanford Digital Library Testbed?

The Stanford Digital Library testbed is our platform for experimentation with interoperation among online services. Our basic approach is to use **distributed objects** to allow integrated access to heterogeneous services across networks. The distributed approach allows the interaction of processes on different machines, with different architectures, implemented in different languages. We use **CORBA** to provide communication between remote processes. In particular, we use Xerox PARC's **ILU**, a free implementation of a CORBA superset. It offers language bindings for C++, C, CommonLisp, Python and Modula-3. We use the interpreted, object-oriented language Python for most of our development work.

For more information, see:

### CORBA

- Information from the [OMG](#), including a [Manual](#)
- [Common Object Services](#) developed at Stanford

### ILU

- [Xerox PARC's ILU Home Page](#)
- The current [ILU Manual](#)
- Information about the [Stanford installation](#)
- A [technical performance evaluation](#) of ILU, HTTP, and basic TCP

### Python

- [The Python Language Home Page](#)
- Information about the [Stanford installation](#)

---

## What Protocol does the Testbed Use?

We have developed the **Digital Library Interoperation Protocol (DLIOP)** for information access and retrieval. It is an asynchronous protocol, providing robustness in the face of network or server outages. Moreover, it also gives the programmer a high degree of control over where and when information objects are materialized, affecting tradeoffs of space and cost vs. time. This protocol has been adopted by other participants of the Digital Library Initiative, including University of Michigan and University of California at Santa Barbara.

For more information, see:



## Information Finding Projects in the Stanford Digital Library

---

One of the major research thrusts of the Stanford Digital Library project is helping users to find information. We have initiated a number of projects in this area, most related to our over-arching theme of interoperability. We have looked at ways that search tools can be used across multiple sources that use different syntaxes or languages. We have also looked at tools to provide statistical or collaborative filtering to locate relevant articles.

---

### FAB

FAB is an adaptive multi-agent information retrieval system which finds interesting pages on the web.

"[An Adaptive Agent for Automated Web Browsing](#)"

- [Marko Balabanovic](#)

---

### GLOSS

The Glossary Server of Servers (GLOSS) project is designed to locate relevant information sources for your query.

"[Generalizing GLOSS to Vector-Space Databases and Broker Hierarchies](#)"

- [Luis Gravano](#)

---

### Query Translator

Databases have different query syntax and different capabilities, even for simple Boolean queries. Translation allows a single query to be mapped into the native format appropriate for each database.

- [Chen-Chuan K. Chang](#)

---

### SenseMaker

SenseMaker helps users iteratively reformulate their information needs through multi-dimensional organizing and active gathering of search results.

"[SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests](#)"

- [Michelle Q Wang Baldonado](#)
- 

### Grassroots

Groupware for information finding, combines mail, news, and web in a single environment with distribution lists

" [Grassroots: A System Providing a Uniform Framework for Communicating, Structuring, Sharing Information, and Organizing People](#) "

- [Kenichi Kamiya](#)
  - [Martin Röscheisen](#)
- 

### The Stanford Digital Library Metadata Architecture

Services need to provide

- metadata about their offerings to help users decide when they should be invoked
- protocol metadata to figure out how they should be invoked, and
- collection metadata for what they should be invoked upon.

The metadata architecture provides a system organization to provide these metadata in a uniform, scaleable way.

### Metadata for Digital Libraries: Architecture and Design Rationale

- [Michelle Q Wang Baldonado](#)
  - [Chen-Chuan K. Chang](#)
  - [Luis Gravano](#)
  - [Andreas Paepcke](#)
- 

### STARTS: Stanford Protocol Proposal for Internet Retrieval and Search

A set of informal standards negotiated among the major search vendors and users to facilitate interoperation.

- [Chen-Chuan K. Chang](#)
  - [Hector Garcia-Molina](#)
  - [Luis Gravano](#)
  - [Andreas Paepcke](#)
- 

### Machine Learning for Information Retrieval

Statistical AI techniques allow the extraction of minimal sets of meaningful search terms

"[Toward Optimal Feature Selection](#) "

- [Mehran Sahami](#)
  - [Daphne Koller](#)
-

### [BackRub](#)

BackRub is a web crawler which is designed to store the connection graph for the web. In other words BackRub stores which pages every web page links to. Currently we are developing techniques using this link data to improve web search engines as well as understand the structure of the web.

- [Larry Page](#)
- 

### [ComMentor](#)

Third-Party Annotations on web pages provide for ways to share information, rate content, and keep notes

["A Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples"](#)

- [Martin Röscheisen](#)
  - [Christian Mogensen](#)
  - [Terry Winograd](#)
- 

### [InterOp Protocol](#)

The heart of the "InfoBus", this protocol describes access methods to search collections, acquire results, and find out about sources.

- [Steve Cousins](#)
  - [Prof. Hector Garcia-Molina](#)
  - [Scott Hassan](#)
  - [Andreas Paepcke](#)
- 

### [SCAM: The Stanford Copy Analysis Mechanism](#)

Making a perfect digital copy of a copyrighted work is easy in a networked world. How can the intellectual property rightsholders be protected? By detecting attempted distribution of illegal copies. Duplicate detection has other uses in information finding as well. An earlier, related project was known as COPS: The Copyright Protection Scheme.

["Building a Scalable and Accurate Copy Detection Mechanism"](#)

- [Prof. Hector Garcia-Molina](#)
  - [Narayanan Shivakumar](#)
- 

### [InterBib](#)

InterBib is a tool for maintaining bibliographic information. Capable of reading from and writing to many different formats, it acts as a unified, searchable repository of bibliographic records.

[Information on InterBib](#)

- [Andreas Paepcke](#)



## User Interface Projects in the Stanford Digital Library

Too often the power of a search engine goes untested because users don't know how to exploit the advanced (or even basic) features. The use of a browser front-end has eased platform independent rapid prototyping, allowing a wide variety of services such as information clustering, annotating, and re-distributing via the WWW. One project even uses a web application to help create web applications! But the web does have drawbacks, such as being largely inaccessible to blind users (hear our audio interface!) and limiting the types of possible interaction. Therefore, our DLITE interface uses a direct manipulation metaphor of iconic representations, rather than relying on CGI forms.

---

### [SenseMaker](#)

SenseMaker helps users iteratively reformulate their information needs through multi-dimensional organizing and active gathering of search results.

["SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests"](#)

- [Michelle Q Wang Baldonado](#)
- 

### [DLITE: A Digital Library Interface](#)

A direct manipulation user interface designed to support user tasks, to smoothly integrate the results of many services, to handle services of widely-varying time scales, to be extensible, and to support sharing and reuse.

["The Digital Library Integrated Task Environment \(DLITE\)"](#)

- [Steve Cousins](#)
- 

### [Grassroots](#)

Groupware for information finding, combines mail, news, and web in a single environment with distribution lists

["Grassroots: A System Providing a Uniform Framework for Communicating, Structuring, Sharing Information, and Organizing People"](#)

- [Kenichi Kamiya](#)
  - [Martin Röscheisen](#)
- 

### [ComMentor](#)

Third-Party Annotations on web pages provide for ways to share information, rate content, and keep notes

["A Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples"](#)

- [Martin Röscheisen](#)
  - [Christian Mogensen](#)
  - [Terry Winograd](#)
- 

Audio Interfaces to HyperText

The structure of a document is captured in HTML/SGML tags which most browsers map to visual display characteristics. We are seeking ways in which this structural information can be conveyed in audio format for blind users or users connecting via telephone.

[CSLI Annual Report](#)

- [Frankie James](#)
  - [Prof. Terry Winograd](#)
- 

WebWriter

WebWriter is a direct manipulation Web page editor that allows users to create new web pages, including advanced features such as tables, without knowing HTML or CGI.

["WebWriter: A Browser-Based Editor for Constructing Web Applications"](#)

- [Arturo Crespo](#)
- 

[RManage/FIRM](#)

Interoperable rights management is one of the service layers that the current Internet is still lacking. FIRM defines a platform for "smart contracts" that is based on a computational reification of contract law; it is realized as part of a novel, network-centric architecture for managing control information that generalizes previous models centered around clients or servers.

["A Network-Centric Design for Relationship-based Rights Management"](#)

- [Martin Röscheisen](#)
  - [Prof. Terry Winograd](#)
- 

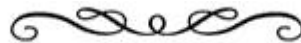




## Digital Library Integrated Task Environment



The Digital Library Integrated Task Environment (DLITE) is an experimental, direct-manipulation interface to information objects and services. Information services are accessed via the InfoBus, and are presented to the user as components in workcenters.



## For More Information...

- ▶ [Interface Details](#)
- ▶ [Interface Architecture paper](#)
- ▶ [Summary of interface goals \(CHI '96 paper\)](#)
- ▶ [List of DLITE Components](#)
- ▶ [Screen Shots](#)



DLITE is implemented as a distributed, client/server application. The server is written in Python, and clients have been written in Python/Tk and Java/AWT. DLITE makes extensive use of the Stanford InfoBus for search and query translation. We have completed a pilot study of the interface, and are continuing to test various aspects of it as well.



## Credits...

DLITE is the PhD project of Steve Cousins. The following people have helped to build or design various aspects of the system:

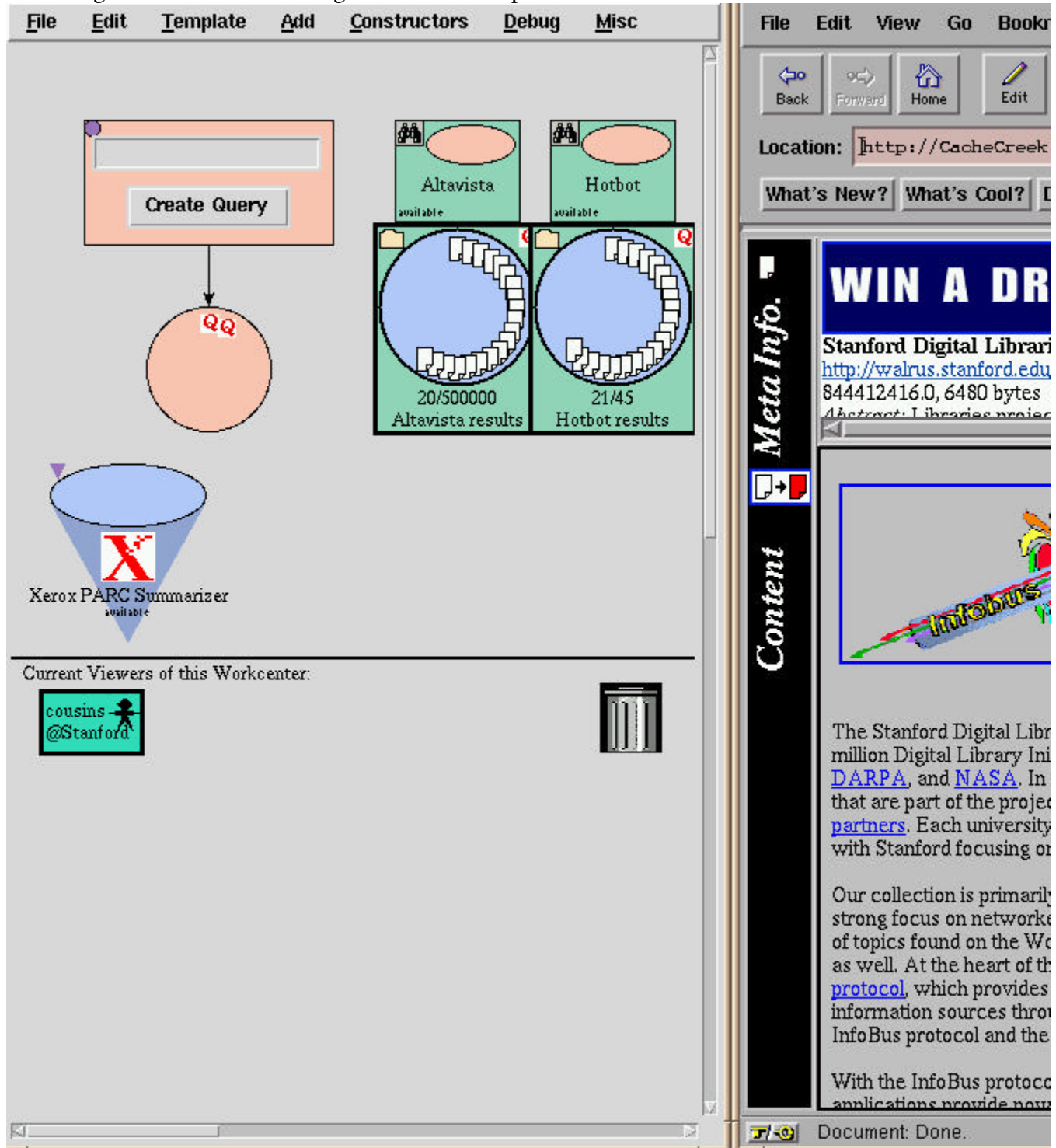
- Scott Hassan
- Alan Steremberg
- Terry Winograd
- Ken Pier
- Eric Bier
- Andreas Paepcke
- Mark Mortensen



# DLITE Screen Shots

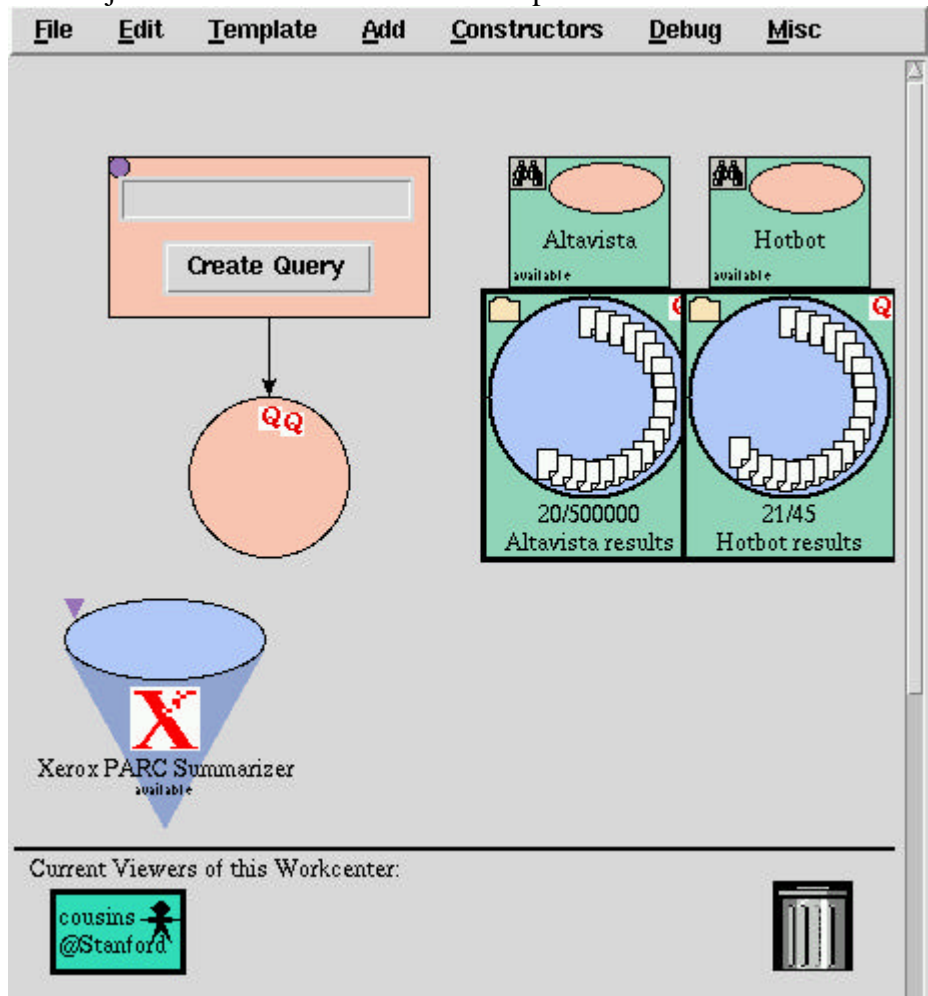
## The whole screen

This image shows DLITE running next to a Netscape browser.



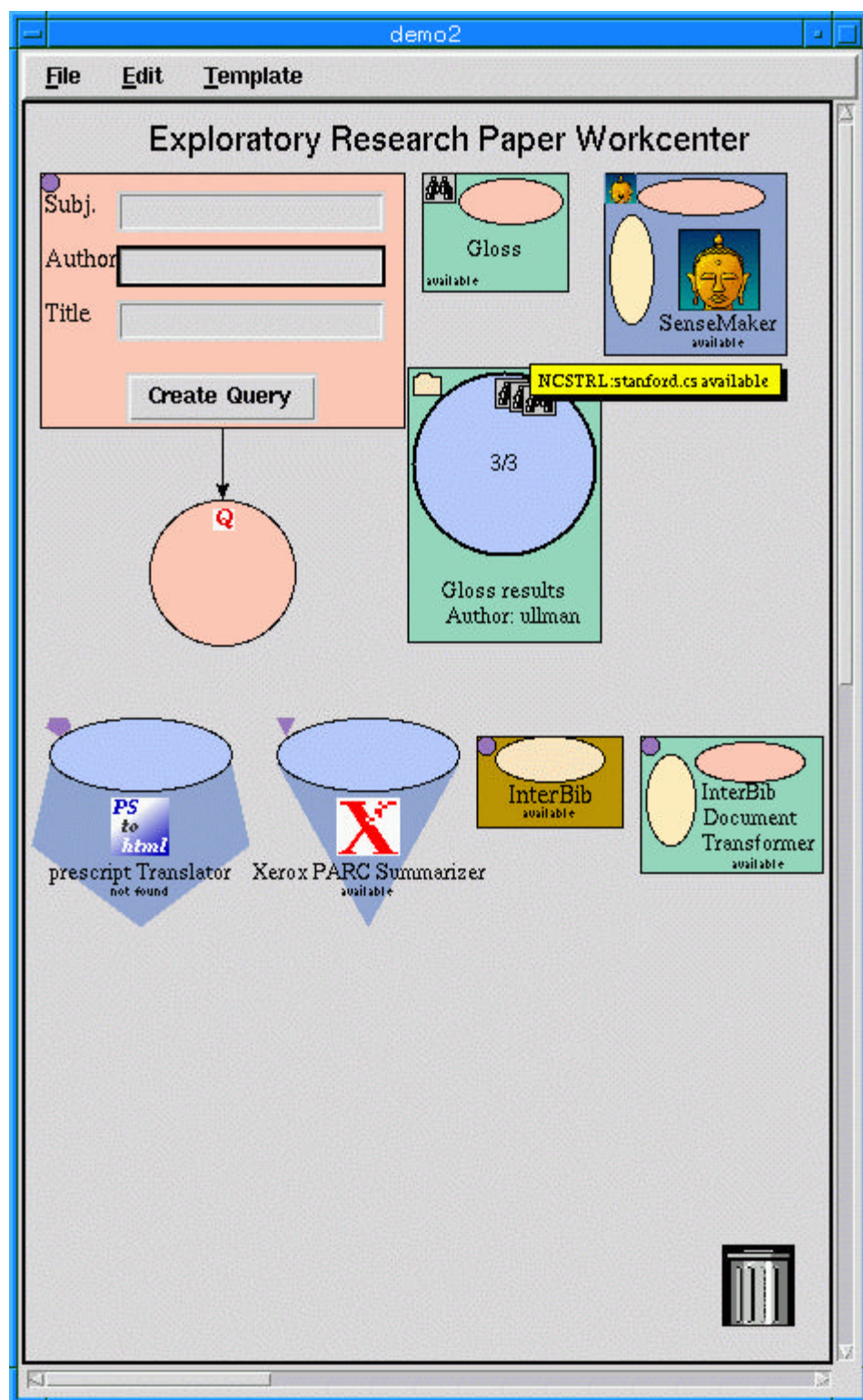
## Simple search

Here is just the DLITE window for a simple search task.



## Exploratory Research Workcenter

Here is a workcenter for doing exploratory research. Services are present for source-finding, sense-making, postscript-to-ascii translation, text summarization, and bibliography generation.



[Steve Cousins](#)

Last modified: Tue Jan 28 23:47:12 PST

## DLI - Berkeley:

---

- [Home Page](#)
  - [IEEE Computer article](#)
  - [Tours](#)
  - [Collections](#)
  - [Source Code](#)
  - [Document-specific image decoders](#)
  - [GISviewer](#) (needs latest browser)
  - [Photos](#) and demos
    - [Context-based image queries](#)
    - [Blobworld](#)
    - [Image classification](#)
  - [California Aerial Photos](#)
  - [United States Department of Agriculture PLANTS Photo Gallery](#)
- 

## Pedagogy:

We recommend that the reader study these materials as part of work to answer the following questions:

- MVD
  - How well does [MVD 0.9](#) work for you? Could you get the links on that page to work (use 2 windows of browser, one for the instructions, and one for testing)? What do you like most about it?
  - Did you use it on video or a PC or Mac with Netscape 4?
  - Did you work out Lens overlaying, such as OCR and then Magnify?
  - For the TableSort example, could you under Anno view the note?
  - Could you get the special behaviors to work: Biblio, where you Select a type of format, use the mouse to select an entry, use Edit and Copy to get a version in that format, and then paste elsewhere?
  - Could you get Doublespace in the View menu to work?
- Cheshire
  - Can you find interesting environmental documents using Cheshire II?
- TileBars
  - What happens with TileBar search of "document" and "retrieval"?
  - What happens with TileBar search of "fault" and "dam"?
  - When is TileBar searching useful on a single document?
- Collections
  - What is the name of the DBMS used?
  - What is a database "schema"? How does it relate to "metadata"?
  - How many documents and how many images are in their collection?
  - How good is the OCRing? What research is underway to improve OCRing beyond that of ScanWorX and how well does it work? What is the main idea behind it?
  - How can you find the dams for a county?
  - How does the database table information for Almond dam relate to the page about it? To the OCR output about that page?

- What is a VLURL? How do you construct it? Can you build one and show results for getting pictures of California wildflowers that have the string "rose" in their common names?
- Display a distribution map for your favorite flower in California.
- Can you tell the direction of flight from the aerial photos?
- How do layers help with managing GIS information with the [GIS viewer](#)? Can you zoom in and out and pan around?

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**



The UC Berkeley Digital Library project is part of the [NSF/ARPA/NASA](#) Digital Library Initiative and part of the California Environmental Resource Evaluation System ([CERES](#)). Research at Berkeley includes faculty, staff, and students in the Computer Science Division, the School of Information Management & Systems, and the Research Program in Environmental Planning & Geographic Information Systems, as well as participation from government agencies and industrial partners. The project's goal is to develop the technologies for intelligent access to massive, distributed collections of photographs, satellite images, maps, full text documents, and "multivalent" documents.



### **Welcome to the Berkeley Digital Library Project.**

[Click here to browse all our collections.](#)

#### **Our Collections:**

- ♦ [Environmental Documents](#)
- ♦ [Photographs](#)
- ♦ [Aerial Photos](#)
- ♦ [Geographic Data](#)
- ♦ [Botanical Datasets](#)

#### **Participants:**

- ♦ [People](#)
- ♦ [Organizations](#)
- ♦ [Other DL Sites](#)
- ♦ [Related Projects](#)

#### **About the Project:**

- ♦ [Tours](#)
- ♦ [Papers](#)
- ♦ [Presentations](#)
- ♦ [Source Code](#)
- ♦ [Database Info](#)
- ♦ [Data Statistics](#)
- ♦ [Web Statistics](#)

#### **Administrative Files:**

- ♦ [Calendar](#)
- ♦ [Mailing Lists](#)
- ♦ [Seminar Schedule Winter '98](#)

**What's New:** new pictures, ARIA web reports, DLI 98 talks

#### **Contact Us:**



[Email to: www@elib.cs.berkeley.edu](mailto:www@elib.cs.berkeley.edu)

[Sign Our Guestbook](#)

This server is powered by a [SUN Microsystems](#) Enterprise 3000 Server, backed by an [IBM](#) 7013 RS 6000 and 3494 Tape Library Dataserver running AMASS software by [EMASS](#). For additional information, see [About Our System](#).

**Digital Library Initiative University of California at Berkeley**

From *Computer* theme issue on the US Digital Library Initiative, May 1996

*Information retrieval becomes an increasing challenge as comprehensive image databases emerge alongside traditional text databases. Here, a set of digital information services offers intriguing new retrieval possibilities.*

## **Toward Work-Centered Digital Information Services**

Robert Wilensky, *University of California, Berkeley*

Work-centered digital information services are library services that address a work group's information retrieval needs. These services differ in several ways from those required of digital libraries or information systems that meet, for example, education- or entertainment-related needs.

First, work groups frequently want to retrieve information, rather than documents per se. Because the answer to a query may be in more than one document, or even in textual form, users require information systems that can perform powerful, complex retrieval and analysis of heterogeneous objects.

Second, a work group must be able to access its own collections of varying data types, including legacy documents, in addition to external data collections. Work groups also continually create new materials, which are subject to differing degrees of external access. This requires flexible authoring, structuring, and delivery mechanisms.

Third, users must be able to integrate an information system into their established work practices, even as the system augments those practices. System interoperability is thus essential and may require custom interfaces. Information system evaluation must consider the system's contribution to the work group's goals, its support of existing work group practices, and its contribution to work practice innovations.

Realizing work-centered digital information systems requires a broad technical agenda that includes

- document image analysis, natural language analysis, and computer vision analysis for effective information extraction;
- new user interface paradigms and authoring tools for better accessing of multimedia information; and
- improved protocols for client program interaction with repositories.

We need to better understand the database issues involved in managing these distributed collections so that digital information services can be used by tens



## Our Collections

*Berkeley Digital Library Project*

Other Information: [About the Database](#) | [About the Digital Library project](#) | [Data Statistics](#)

	description	query form	browse a list	static map	active map*	more information
<b>Documents</b>	Cal. environmental documents: plans, ordinances, EIRs, etc.	<a href="#">search</a> <a href="#">TileBars</a> *	<a href="#">X</a>			<a href="#">about the collection</a> <a href="#">advanced structured docs</a> <a href="#">about MVD</a> <a href="#">about TileBars</a> <a href="#">tour the documents</a>
<b>Databases</b>	California dams	<a href="#">X</a>	<a href="#">X</a>	<a href="#">X</a>	<a href="#">X</a>	<a href="#">about the dams</a>
	CalFlora	<a href="#">X</a>				<a href="#">about Calflora</a>
	CalFlora Occurrences	<a href="#">X</a>				<a href="#">about Calflora Occurrences</a>
	Fishes of the Sacramento-San Joaquin Estuary		<a href="#">X</a>			
	Bay Area Streets				<a href="#">X</a>	
	California Gazeteer				<a href="#">X</a>	
<b>Aerial Photos</b>	Sacramento R. Delta, Suisun		<a href="#">X</a>	<a href="#">X</a>		<a href="#">about air photos</a>
<b>Geographical Layers</b>	Northern California				<a href="#">X</a>	<a href="#">about GIS Viewer</a> <a href="#">tour the GIS Viewer</a>
	Russian River				<a href="#">X</a>	
	Sacramento R. Delta Fish Flow				<a href="#">X</a>	
<b>Photographs</b>	Cal. flora and habitats, Dept. of Water Resources photos, etc.	<a href="#">X</a>	<a href="#">X</a>			<a href="#">about the photos</a> <a href="#">computer vision research</a> <a href="#">tour the images</a>
<b>External Collections</b>	BIOSIS dictionary: water subdomain	<a href="#">X</a>				<a href="#">about OASIS</a>
	INSPEC dictionary	<a href="#">X</a>				<a href="#">about OASIS</a>

\* java is required

UC Berkeley Digital Library Project | [www@elib.cs.berkeley.edu](http://www@elib.cs.berkeley.edu) | last updated June 26, 1997



## GIS Viewer: North Coast of California

*Berkeley Digital Library Project*

---

*\*\*Please be patient while this Java applet loads.\*\**

For more information about the GIS Viewer check out our [tour](#) or our help pages. If you click on the help button within the GIS Viewer applet the help pages will come up in a separate browser window. If you click [here](#) the help pages will come up in this window.

---

[Berkeley Digital Library](#)

[www@elib.cs.berkeley.edu](http://www@elib.cs.berkeley.edu)



# Photographs

## *Berkeley Digital Library Project*

---

The Berkeley Digital Library collection includes a large number of digitized images from many different sources. We add new images to the collection on a weekly basis. As of April 1998 there were over 58,000 images available for online searching. For a current count of images, see [Data Statistics](#).

### Search all the images in our collection

- Fill out a [query form](#) for all the images
- Take a [tour of the image collection](#)
- View [sample queries](#) for all the pictures

### Browse individual photo collections

- [California Department of Water Resources](#) DWR film library collection of over 15,000 images.
- [Brousseau Collection of California Flora](#) Brother Alfred Brousseau (1908-1988) of St. Mary's College in Moraga, California made a collection of 35mm color slides of native wildflowers of California which consists of over 11,000 slides of over 2,000 species, as well as pictures of trees and mushrooms.
- [California Wilderness Scenes](#) Brother Alfred Brousseau's photographs taken between 1954 and 1984 of California's natural beauty.
- [California Habitats](#) A collection of 158 photographs of California habitats taken by Marc Hoshovsky of the State of California Department of Fish and Game.
- [Russian River](#) A small collection of photographs from the area around the Russian River.
- [Corel Stock Photos](#) This is a collection of images from [Corel](#) that we use for computer vision research. These images may not be downloaded or saved.

### Computer vision research

Click on the heading above to learn more about how we use these pictures for our computer vision work, including demos, papers, faculty and researchers involved in this work.

**NEW** Check out our [Blobworld](#) image retrieval system

---

UC Berkeley Digital Library Project | [www@elib.cs.berkeley.edu](http://www@elib.cs.berkeley.edu)



## Demos: Content-based Queries

*Berkeley Digital Library Project*

---

The following queries use image content information alone to retrieve pictures from a collection of 50,000 images. The database query that was generated will be shown at the bottom of each page of pictures. For more information about image analysis techniques used, see [Computer Vision Research](#). To construct your own query, see [Content-based Query on all Images](#).

---

### Finding Objects in Pictures



Horses (14)

see [Finding horses using body plans](#)

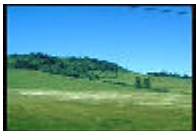
---

### Colored Blobs and Color Percentages



Sailing and Surfing (17)

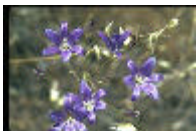
*blue-green % > 30 and very sm. yellow dots > 0 and collection = corel or DWR*



Pastoral Scenes (93)

*green % > 25 and lt. blue % > 25*

[Pastoral Scenes: non-Corel pictures only](#)



Purple Flowers (114)

*sm. purple dots > 3*



Fields of Yellow Flowers (75)

*very sm. yellow dots > 15*



Pink People (69)

*lg. or very lg. pink dots > 0 and orange % > 1 and collection = corel or DWR*



Animals (229)

*very lg. brown dots > 0 and very sm. black dots > 1 and green % > 20*

---

# Welcome to Blobworld!

## Why Blobworld?

Very large collections of images are growing ever more common. From stock photo collections to proprietary databases to the Web, these collections are diverse and often poorly indexed. Unfortunately, image retrieval systems have not kept pace with the collections they are searching. The shortcomings of these systems are due both to the image representations they use and to their methods of accessing those representations to find images:

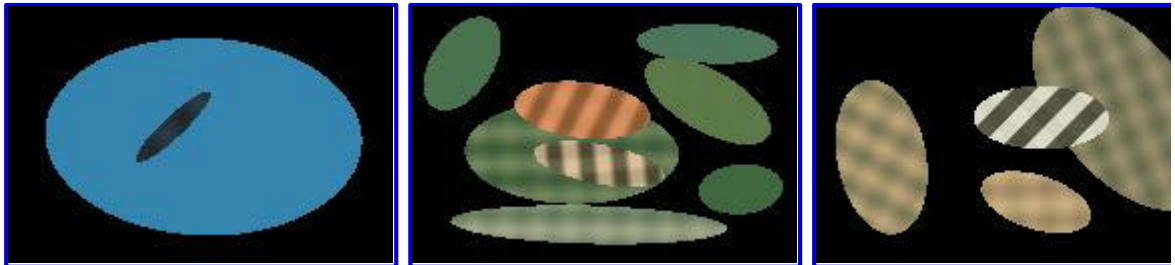
- While users often would like to find images containing particular objects ("things"), most existing image retrieval systems represent images based only on their low-level features ("stuff"), with little regard for the spatial organization of those features.
  - Systems based on user querying are often unintuitive and offer little help in understanding why certain images were returned and how to refine the query. Often the user knows only that he has submitted a query for, say, a bear and retrieved very few pictures of bears in return.
  - For general image collections, there are currently no systems that automatically classify images or recognize the objects they contain.
- 

## What is Blobworld?

We have developed a new image representation, "Blobworld," and a retrieval system based on this representation. While Blobworld does not exist completely in the "thing" domain, it recognizes the nature of images as combinations of objects, and querying and learning in Blobworld are more meaningful than they are with simple "stuff" representations.

We use the Expectation-Maximization (EM) algorithm to perform automatic segmentation based on image features. EM iteratively models the joint distribution of color and texture with a mixture of Gaussians; the resulting pixel-cluster memberships provide the segmentation of the image. After the image is segmented into regions, a description of each region's color, texture, and spatial characteristics is produced.

Here is a visualization of the Blobworld representation. We show each image region as an elliptical blob; each blob's two dominant colors are shown in the plaid patterns. The orientedness of the pattern indicates the texture's anisotropy, the orientation of the plaid indicates the orientation of the texture, and the sharpness of the plaid indicates the texture contrast.

**Original  
images:****Blobworld:**

---

## What can we use Blobworld for?

In a querying task, the user can access the regions directly in order to see the segmentation of the query image and specify which aspects of the image are central to the query. When query results are returned, the user sees the Blobworld representation of the returned images; this assists greatly in refining the query. You can see the [results](#) of several image queries using Blobworld, or [try your own query](#) on the images in the Digital Library collection.

Because Blobworld often encodes the objects in an image, we can also classify images automatically using an algorithm that learns the distributions of categories in Blobworld. You can see the [results](#) from such a system.

---

## Want to learn more?

- [Try a Blobworld query!](#)
- Check out the [query results](#) or [classification results](#).
- Read our most recent [paper about Blobworld](#) or [other papers](#).

---

*The original images are copyright [Corel](#). They are for viewing only and may not be saved or downloaded.*

---

*Last updated January 29, 1998, by Chad Carson*



# Image Classification

*Berkeley Digital Library Project*

The 14 categories shown below were chosen from the [Corel](#) image collection. About 90 pictures from each category were used for training and testing an algorithm that classifies images using [regions of coherent color and texture](#). The images used for testing are available [here](#). Use the table below to see all the images in each category and the classification of each image in a given category. For comparison, we also show the classification using color histograms.

All images in a category	Classified into a category using Blobworld	Classified into a category using color histograms
<a href="#">Airplanes</a>	<a href="#">Classified as airplanes by Blobworld</a>	<a href="#">Classified as airplanes by color histograms</a>
<a href="#">Bald eagles</a>	<a href="#">Classified as bald eagles by Blobworld</a>	<a href="#">Classified as bald eagles by color histograms</a>
<a href="#">Brown &amp; black bears</a>	<a href="#">Classified as brown &amp; black bears by Blobworld</a>	<a href="#">Classified as brown &amp; black bears by color histograms</a>
<a href="#">Cheetahs</a>	<a href="#">Classified as cheetahs by Blobworld</a>	<a href="#">Classified as cheetahs by color histograms</a>
<a href="#">Deserts</a>	<a href="#">Classified as deserts by Blobworld</a>	<a href="#">Classified as deserts by color histograms</a>
<a href="#">Elephants</a>	<a href="#">Classified as elephants by Blobworld</a>	<a href="#">Classified as elephants by color histograms</a>
<a href="#">Fields</a>	<a href="#">Classified as fields by Blobworld</a>	<a href="#">Classified as fields by color histograms</a>
<a href="#">Horses</a>	<a href="#">Classified as horses by Blobworld</a>	<a href="#">Classified as horses by color histograms</a>
<a href="#">Mountains</a>	<a href="#">Classified as mountains by Blobworld</a>	<a href="#">Classified as mountains by color histograms</a>
<a href="#">Night scenes</a>	<a href="#">Classified as night scenes by Blobworld</a>	<a href="#">Classified as night scenes by color histograms</a>
<a href="#">Polar bears</a>	<a href="#">Classified as polar bears by Blobworld</a>	<a href="#">Classified as polar bears by color histograms</a>
<a href="#">Sunsets</a>	<a href="#">Classified as sunsets by Blobworld</a>	<a href="#">Classified as sunsets by color histograms</a>
<a href="#">Tigers</a>	<a href="#">Classified as tigers by Blobworld</a>	<a href="#">Classified as tigers by color histograms</a>
<a href="#">Zebras</a>	<a href="#">Classified as zebras by Blobworld</a>	<a href="#">Classified as zebras by color histograms</a>



[Berkeley DL](#)



[AccessMatrix](#)



[Information](#)



[Photographs](#)



[Comments](#)



## California Aerial Photos

*Berkeley Digital Library Project*

Click on a **Flightline** to see thumbnail images for that flightline.

Some of the available flightlines are positioned on a map. Click on a **Description** to see the map for that area, with links to images.

Description	Contractor's ID	Elib ID	Type	Flightlines	Date	Contractor	Source
California Aqueduct: East Branch	WR-BED-C	aqd_east	color	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>   <a href="#">5</a>   <a href="#">6</a>   <a href="#">7</a>   <a href="#">8</a>   <a href="#">9</a>	Aug 03, 1994	I.K.Curtis Services, Inc.	DWR
North Bay Aqueduct	WR-AXY-C	aqd_nbay	b&w	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>	Oct 02, 1990	Radman Aerial Surveys	DWR
South Bay Aqueduct: Livermore to Terminal Facilities	WR-AXX	aqd_sbay	b&w	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>   <a href="#">5</a>   <a href="#">6</a>   <a href="#">7</a>	Oct 02, 1990	Radman Aerial Surveys	DWR
<a href="#">North Delta Flood Plain Environmental Study</a>	WR-BBG-C	delta_nflood	color	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>   <a href="#">5</a>   <a href="#">6</a>   <a href="#">7</a>	Feb 14, 1993	Radman Aerial Surveys	DWR
<a href="#">Statutory Delta</a>	WR-BCM-CIR	delta_stat	colorIR	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>   <a href="#">5</a>   <a href="#">6</a>   <a href="#">7</a>   <a href="#">8</a>   <a href="#">9</a>   <a href="#">10</a>   <a href="#">11</a>   <a href="#">12</a>   <a href="#">13</a>   <a href="#">14</a>   <a href="#">15</a>   <a href="#">16</a>   <a href="#">17</a>   <a href="#">18</a>   <a href="#">19</a>   <a href="#">20</a>   <a href="#">21</a>   <a href="#">22</a>	Jun 22-23, 1993	Radman Aerial Surveys	DWR
<a href="#">Suisun Marsh Vegetation Study Low Tide</a>	WR-BDW-C	suisun	color	<a href="#">1</a>   <a href="#">2</a>   <a href="#">3</a>   <a href="#">4</a>   <a href="#">5</a>   <a href="#">7</a>   <a href="#">9</a>   <a href="#">11</a>   <a href="#">13</a>   <a href="#">16</a>   <a href="#">19</a>   <a href="#">20</a>   <a href="#">22</a>   <a href="#">23</a>   <a href="#">24</a>   <a href="#">25</a>   <a href="#">26</a>   <a href="#">27</a>   <a href="#">28</a>   <a href="#">29</a>   <a href="#">30</a>   <a href="#">31</a>   <a href="#">32</a>   <a href="#">33</a>   <a href="#">34</a>   <a href="#">35</a>   <a href="#">36</a>   <a href="#">37</a>   <a href="#">38</a>   <a href="#">39</a>   <a href="#">40</a>   <a href="#">41</a>   <a href="#">42</a>   <a href="#">43</a>   <a href="#">44</a>   <a href="#">45</a>   <a href="#">46</a>   <a href="#">47</a>   <a href="#">48</a>	Jun 10-14, 1994	Radman Aerial Surveys	DWR

## DLI - Santa Barbara:

---

- [Home Page](#)
- [Tutorial](#)
- [World Spatial Data](#)
- [Annual Report](#)
- [H. Chen's work](#) (with "cool DL, Web, agent, visualization, and multilingual IR demos"), and [GIS work](#)

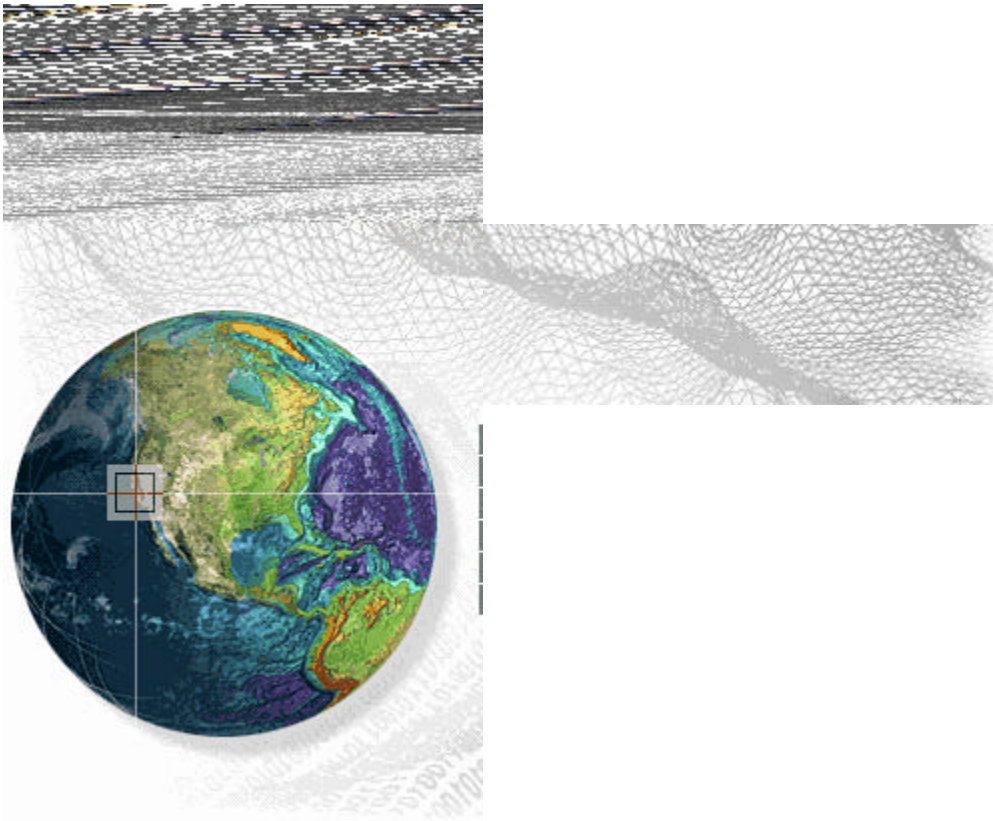
---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta



**THE ALEXANDRIA DIGITAL LIBRARY**

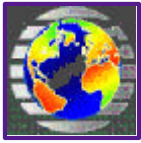
University of California, Santa Barbara  
1205 Girvetz Hall  
Santa Barbara, CA 93106, USA  
**TEL:** 805.893.7665 **FAX:** 805.893.3045  
**URL:** [www.alexandria.ucsb.edu](http://www.alexandria.ucsb.edu)



---

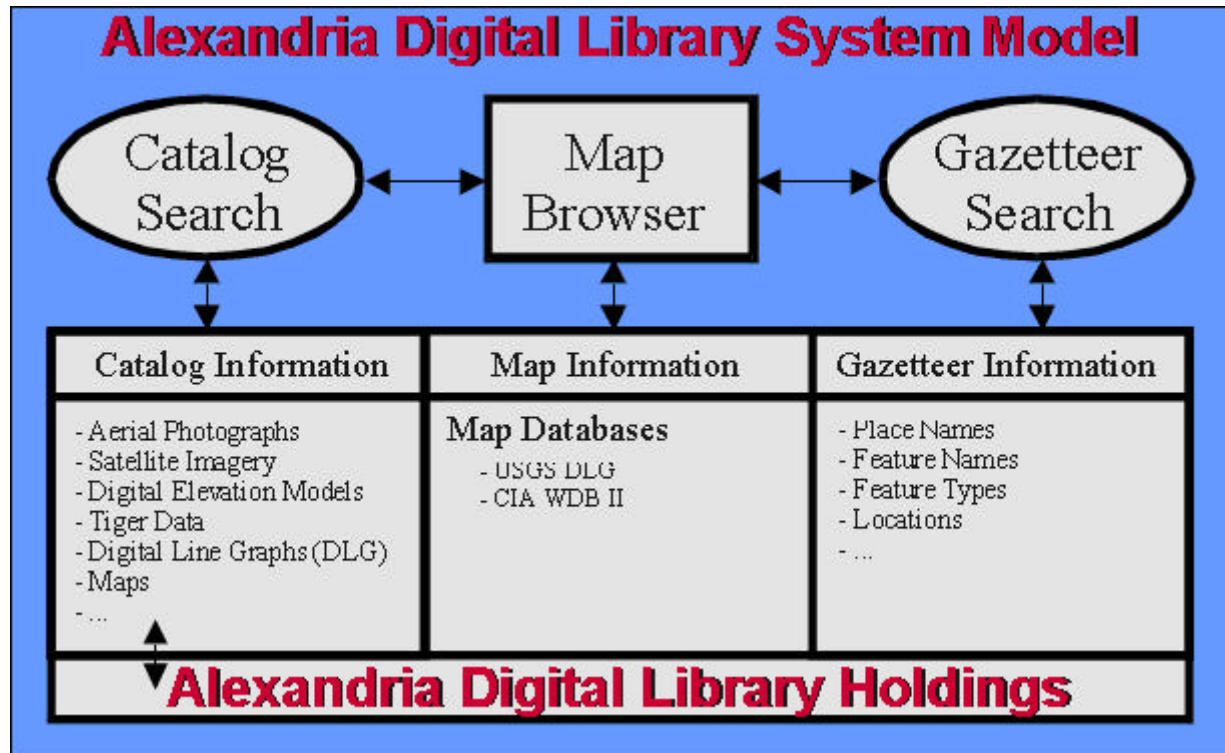
**Last Modified: May 30, 1998**

Our website is being redesigned and now uses some high-end browser features that may not be available in the browser you are currently using. We utilize frames and tables extensively in the new design and require a compatible browser, such as **Netscape Navigator 3.0** or **Microsoft Internet Explorer 3.0**.



# Tutorial Table of Contents

◀ Prev    Next ▶



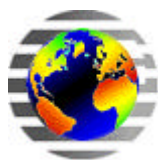
## Conceptual model of the Alexandria Web interface

- [Conventions](#)
- [Session / System Setup](#)
- [Map](#)
- [Gazetteer](#)
- [Catalog](#)
- [Overview of Current Holdings](#)
- [Walkthroughs \(Example Sessions\)](#)
- [Feedback](#)
- [Technical Reference](#)
- [Acknowledgements](#)
- [Access ADL](#)

- [World Wide Web Help](#)  
(How to use a web browser. That's what you're using right now.)

◀ Prev    Next ▶

# Universe



Alexandria Digital Library: [ADL](#)

[\[comment\]](#) [\[suggestions\]](#) [\[information\]](#) [\[add a URL\]](#)

## Universe

[\[UNIVERSE\]](#) [\[EARTH\]](#) [\[AFRICA\]](#) [\[AMERICAS\]](#) [\[ANTARCTICA\]](#) [\[ASIA\]](#) [\[EUROPE\]](#) [\[OCEANIA\]](#)  
[\[By Subject\]](#) [\[By Title\]](#)

<a href="#">Earth</a>	<a href="#">Jupiter</a>	<a href="#">Mars</a>	<a href="#">Moon</a>
<a href="#">Neptune</a>	<a href="#">Saturn</a>	<a href="#">Sun</a>	<a href="#">Uranus</a>
<a href="#">Venus</a>			

## Universe

### Aerial photographs

- [Regional Planetary Image Facility](http://ceps.nasm.edu:2020/rpif.html)::http://ceps.nasm.edu:2020/rpif.html
- [Sources of Earth and Planetary Photography](http://ceps.nasm.edu:2020/RPIF/RPIFsources.html)::http://ceps.nasm.edu:2020/RPIF/RPIFsources.html

### Artificial satellites

- [Mission and Spacecraft Library](http://leonardo.jpl.nasa.gov/msl/home.html)::http://leonardo.jpl.nasa.gov/msl/home.html

### Astronomical - Observations

- [The Web Window to the Invisible Universe](http://www.pkts.atnf.csiro.au/databases/surveys/aitoff/aitoff.html)::http://www.pkts.atnf.csiro.au/databases/surveys/aitoff/aitoff.html

### Astronomical photometry

- [JPL Public image archive](http://www.jpl.nasa.gov/archive/images.html)::http://www.jpl.nasa.gov/archive/images.html
- [Latest HST Observations. \(Hubble\)](http://www.stsci.edu/pubinfo/Latest.html)::http://www.stsci.edu/pubinfo/Latest.html
- [NASA JSC Digital Image Collection](http://images.jsc.nasa.gov/)::http://images.jsc.nasa.gov/
- [Space Image Libraries](http://www.okstate.edu/aesp/image.html)::http://www.okstate.edu/aesp/image.html
- [Stereoscopic Maps of Nearby Stars](http://www.clockwk.com/stars/index.html)::http://www.clockwk.com/stars/index.html
- [Today@NASA](http://www.hq.nasa.gov/office/pao/NewsRoom/today.html)::http://www.hq.nasa.gov/office/pao/NewsRoom/today.html

### Astronomy

- [Astronomical Data Center](http://adc.gsfc.nasa.gov/)::http://adc.gsfc.nasa.gov/
- [CyberAstronomy](http://reality.sgi.com/sambo/Oobe/CyberAstronomy/CyberAstronomy/intro.html)::http://reality.sgi.com/sambo/Oobe/CyberAstronomy/CyberAstronomy/intro.html
- [NASA/IPAC Extragalactic Database \(NED\)](http://ned.ipac.caltech.edu/)::http://ned.ipac.caltech.edu/
- [Planet Finder](http://www.calweb.com:80/~mcharvey/planet_all.html)::http://www.calweb.com:80/~mcharvey/planet\_all.html
- [SEDs Internet Headquarters](http://sed.s.lpl.arizona.edu/)::http://sed.s.lpl.arizona.edu/

### Astrophysics

- [HEASARC/GSFC Home Page](http://guinan.gsfc.nasa.gov/)::http://guinan.gsfc.nasa.gov/

**Lab Info**

Personnel  
Recognition  
Tech Reports  
Facilities

**Project Groups**

COPLINK  
GIS  
Interspace  
MedInfo

**Technologies**

Analysis & Viz  
Multilingual IR

**Public Demo**

CSQuest  
ET-Map  
GS-Map  
Internet Spider  
WormSpace

**Private Demo**

Cancer Space  
Cat-Map  
Geo-Map

**Java Agent Workshop**

BFS Spider  
GA Optimizer  
Itsy Bitsy Spider

**Artificial Intelligence Lab**

Department of MIS, University of Arizona.

Head, [Dr. Hsinchun Chen](#)

University of Arizona, Department of MIS, Artificial Intelligence Lab, McClelland Hall 430, Tucson, AZ 85721.

TEL: (520) 621-2748, FAX: (520) 621-2433, <http://ai.bpa.arizona.edu>

For comments, please send to: [Webmaster](#)

## Welcome

### Lab Information

The UA/MIS Artificial Intelligence Lab, headed by Dr. Hsinchun Chen, consists of 3-5 Ph.D.-level researchers and 15-20 research scientists and assistants. It specializes in database integration, digital libraries, knowledge discovery, internet/intranet technologies, and intelligent information retrieval.

It has received multi-million-dollar funding from various government agencies including National Science Foundation (NSF), Advanced Research Projects Agency (ARPA), National Aeronautics and Space Administration (NASA), and National Institutes of Health (NIH).

### Selected Ongoing Projects

- DARPA-funded ``The Interspace Prototype: An Analysis Environment for Semantic Interoperability" project, 1997-2000.
- NSF/NASA/ARPA-funded Digital Library Initiative (DLI) ``[Building the Interspace](#)" project, 1994-1998.
- NSF-funded ``[Internet Categorization and Search](#)" project, 1995-1998.
- NSF-funded ``Supplement to Alexandria DLI Project: [A Semantic Interoperability Experiment for Spatially-Oriented Multimedia Data](#) ," 1996-1998.
- [National Center for Supercomputing Applications](#) High-performance Computing Resources Grant ``[Information Analysis and Knowledge Discovery for Digital Libraries](#) " project, 1995-1997.
- [National Library of Medicine](#) funded ``Semantic Retrieval for Toxicology and Environmental Health Databases" project, 1996-1997.
- [National Cancer Institute](#) funded ``Information Analysis and Visualization for Cancer Literature" project, 1996-1997.

### Projects Featured In

## DLI - Illinois:

---

- [Home Page](#)
- [IEEE Computer article](#)
- [Glossary](#)
- [SGML/XML Home Page](#), [SD Unit Notes in CS5604](#), [SoftQuad Products](#)
- Collections: [Publishers](#), [Software Companies](#)
- [Interspace](#), [concept extraction](#), [concept spaces](#), [term suggestion](#)
- [UIUC DLI Spring 1997 Partners Workshop](#)
- [Social Science Home Page: Fn Reqts](#)
- [DeLiver](#)
  - Before using DeLiver you should get one of the following 2 files and install it on your Windows 95/NT system. Be sure to have any version of Netscape closed after the download, when you do the install. These files are local to VT to save you the time of downloading as per the U. Ill. instructions. The Panorama versions each take about 1.9M for the install package but less than 1M for the C: drive installed version: [for Netscape 3](#), [for Netscape 4](#)
  - Explore the DeLiver pages, and try to answer the following questions.
  - What does the Help tell you about the system?
  - What is the coverage?
  - What are unusual services not provided by similar systems?
  - What is Panorama and what does it do to enhance WWW capabilities?
  - Can you use browsing to find the IEEE-CS articles (i.e., v. 29 n. 5) we looked at for this course?
  - Can you use searching to find the IEEE-CS articles we looked at for this course?
  - How does the presentation using WWW and Panorama differ from that you are familiar with (HTML, PDF)? What benefits are there from having Panorama?
  - What other interesting articles about digital libraries did you find?
  - Is the field specific searching of help?
  - Is the interface for DeLiver easy to understand? How could it be improved?

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Projects\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta



## ACHIEVEMENTS

[Progress Reports](#)  
[Overview Papers & Talks](#)  
[Publications & Reports](#)  
[Workshops](#)  
[Nat'l Synchronization Effort](#)

## RESEARCH GROUPS

[Repositories \(testbed\)](#)  
[Social Science \(User Studies\)](#)  
[Semantic Research](#)  
[Interspace Prototype](#)  
[System Evaluation](#)

## PARTNERSHIPS

[Publishers](#)  
[Software Providers](#)

## PEOPLE

[Contact Information](#)

## TECHNOLOGY HIGHLIGHTS

[The 5 other DLI Projects](#)  
[UIUC Digital Libraries](#)  
[DL Related Information](#)  
[Global Cultural Memory Project](#)

[Computing the Future](#)  
[A National Research Council Report](#)



**Note:** DeLiver can be accessed by UIUC faculty, staff, and students.

**DeLiver**  
**USAGE STATISTICS: updated daily**

The Digital Libraries Initiative (DLI) project at the [University of Illinois at Urbana-Champaign](#) has developed the information infrastructure to effectively search technical documents on the Internet. We have constructed a large [testbed](#) of scientific literature, are [evaluating](#) its effectiveness under significant use, and [researching](#) enhanced search technology. We are building repositories (organized collections) of indexed multiple-source collections and federating (merging and mapping) them by searching the material via [multiple views](#) of a single virtual collection.

Our testbed of [Engineering and Physics journals](#) is based in the [Grainger Engineering Library](#). We are placing article files into the digital library on a production basis in Standard Generalized Markup Language ([SGML](#)) from engineering and science [publishers](#). The [Research](#) section of the project is using NCSA supercomputers to compute indexes for new search techniques on large collections, to simulate the future world, and to provide new technology for the Testbed section.

The UIUC DLI is a recipient of a grant in the [NSF/DARPA/NASA Digital Libraries Initiative](#).

ONLINE SUMMARIES

From *Computer* theme issue on the US Digital Library Initiative, May 1996

*A University of Illinois project is developing an infrastructure for indexing scientific literature so that multiple Internet sources can be searched as a single federated digital library.*

## Federating Diverse Collections of Scientific Literature

Bruce Schatz, William H. Mischo, Timothy W. Cole, Joseph B. Hardin, Ann P. Bishop, *University of Illinois*  
Hsinchun Chen, *University of Arizona*

The most important recorded information medium on the Internet, and in the world at large, is the document. Although text might seem prosaic in contrast to multimedia objects, it is still the major medium for communicating information. Internet document retrieval can draw upon years of research results and practical experience in on-line information access as well as from traditional physical libraries. The technology for text information retrieval is far more mature than that for other media. Therefore, documents are also the best vehicle for investigating problems specific to digital libraries, such as the federation problem of making distributed collections of heterogeneous materials appear to be a single integrated collection.

The Digital Library Initiative (DLI) project at the University of Illinois at Urbana-Champaign is developing the information infrastructure to effectively search technical documents on the Internet. We are constructing a large testbed of scientific literature, evaluating its effectiveness under significant use, and researching enhanced search technology. We are building repositories (organized collections) of indexed multiple-source collections and federating (merging and mapping) them by searching the material via multiple views of a single virtual collection.

Developing widely usable Web technology is also a key goal. Improving Web search beyond full-text retrieval will require using document structure in the short term and document semantics in the long term. Our testbed efforts concentrate on journal articles from the scientific literature, with structure specified by the Standard Generalized Markup Language (SGML). Our research efforts extract semantics from documents using the scalable technology of concept spaces based on context frequency. We then merge these efforts with traditional library indexing to provide a single Internet interface to indexes of multiple repositories.

Our project focuses on developing a large-scale infrastructure adequate for solving real-world problems. The Testbed part of the project is based in the



# Glossary

## ARPA (DARPA)

The Defense Advanced Research Projects Agency (DARPA) is the central research and development organization for the Department of Defense (DoD). It manages and directs selected basic and applied research and development projects for DoD, and pursues research and technology where risk and payoff are both very high and where success may provide dramatic advances for traditional military roles and missions and dual-use application.

## **Broad System of Ordering (BSO)**

A general subject classification scheme, commissioned by UNESCO, intended to be a switching language among existing classification schemes and thesauri to make them mutually compatible on a general level. It provides about 4,000 subdivisions.

## **Collection Interface Agent**

A program which interacts with the Collection Registry. For searchable collections (Z39.50, FTL, ...) it takes care of talking to the remote collection, submitting searches, fetching and processing results. It is also referred to as a CIA or a collection agent.

## **Collection Registry**

The database in which descriptions of collections are stored.

## **Concept Space**

Graph of terms occurring within objects linked to each other by the frequency with which they occur together.

## Corporation for National Research Initiatives (CNRI)

A non-profit organization dedicated to formulating, planning, and carrying out national-level research initiatives on the use of network-based information technology. CNRI is concentrating on research and development for the National Information Infrastructure, working collaboratively with industry, academia, and government.

## **Derived Data**

Data that was originally supplied in one form, but was converted to another form using some automated process.

## **DID**

Document Image Decoding, a methodology for document recognition founded on statistical communication theory.

## **Digital Libraries**

Digital libraries basically store materials in electronic format and manipulate large collections of those materials effectively.

## Digital Library Federation

The Federation is comprised of leaders of fifteen of the nation's largest research libraries and archives and the Commission on Preservation and Access ([CPA](#)). A primary goal of the Federation is the implementation of a distributed, open digital library accessible across the global Internet. The library will consist of collections expanding over time in number and scope to be created from the conversion of digital form of documents contained in founding member and other libraries and archives, and from the incorporation of holdings already in electronic form.

## DLI

Digital Libraries Initiative. Six research projects developing new technologies for digital libraries -- storehouses of information available through the Internet, -funded through a joint initiative of the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA). The projects' focus is to dramatically advance the means to collect, store, and organize information in digital

[Index](#)[News](#)[Software](#)[Biblio](#)[XML](#)[XSL](#)[XLL](#)

# *The SGML/XML Web Page*

Copyright (c) Robin Cover 1994-98. Last modified June 21, 1998. The SGML/XML Web Page lives at <http://www.sil.org/sgml/sgml.html>. Support for the development and maintenance of this SGML/XML Web Page is provided in part by [SoftQuad, Inc.](#) and by [the Summer Institute of Linguistics](#), to whom gratitude is acknowledged.

[[Search the entire SGML/XML database](#)] - [[Submit a Contact Address Form](#)]

*The SGML/XML Web Page* is a comprehensive online database containing reference information and software pertaining to the Standard Generalized Markup Language (SGML) and its subset, the Extensible Markup Language (XML). The database features an [SGML/XML news column "What's New?"](#) and a cumulative annotated [bibliography](#) with over 2000 entries. The collection contains over 2400 documents explaining and illustrating the application of the SGML/XML family of standards, including HyTime, DSSSL, XSL, XLL, XLink, XPointer, SPDL, CGM, ISO-HTML, and several others. These documents are accessible from the topical overview presented below, or from the [fully expanded contents listing](#) (Site Index) in a separate document.

## Overview

<b>The SGML/XML Web Page</b>	<ul style="list-style-type: none"> <li>• <a href="#">Site Index</a></li> <li>• <a href="#">Site Description</a></li> </ul>
<b>News</b>	<ul style="list-style-type: none"> <li>• <a href="#">What's New in the SGML/XML Web Page?</a></li> <li>• <a href="#">XML News Articles</a></li> <li>• <a href="#">XML Press News</a></li> <li>• Earlier News Highlights: [<a href="#">1997</a>] [<a href="#">1996</a>] [<a href="#">1995</a>]</li> </ul>
<b>Introductions</b>	<ul style="list-style-type: none"> <li>• <a href="#">General Introduction to SGML</a></li> <li>• <a href="#">General Introduction to XML</a></li> <li>• <a href="#">SGML Frequently Asked Questions (FAQs)</a></li> <li>• <a href="#">XML Frequently Asked Questions (FAQs)</a></li> </ul>
<b>XML, XSL, XLL</b>	<ul style="list-style-type: none"> <li>• <a href="#">XML (Extensible Markup Language)</a></li> <li>• <a href="#">XSL (Extensible Style Language)</a></li> <li>• <a href="#">XLL (XLink and XPointer Languages)</a></li> </ul>
<b>Related Standards</b>	<ul style="list-style-type: none"> <li>• <a href="#">DSSSL</a></li> <li>• <a href="#">HyTime</a></li> <li>• <a href="#">Other Standards Related to SGML/XML</a></li> </ul>
<b>Applications</b>	<ul style="list-style-type: none"> <li>• <a href="#">General SGML/XML Applications</a></li> <li>• <a href="#">Academic Applications</a></li> <li>• <a href="#">Government and Industry Applications</a></li> <li>• <a href="#">Proposed XML Applications</a></li> </ul>
<b>Publications</b>	<ul style="list-style-type: none"> <li>• <a href="#">Essential SGML/XML Books</a></li> <li>• <a href="#">Comprehensive SGML/XML Bibliography</a></li> <li>• <a href="#">Journals, Newsletters and other Serials</a></li> </ul>

# UNIT SD

## Course Notes on SD Unit --- SGML, Document Processing/Translation

---

### SGML and Document Processing

#### Word Processing

#### Document Management

#### Markup, OHCO

#### SGML

### Summary - SGML and Document Processing

- Word Processing - providing data
  - Document Management - bigger issue than IS&R (e.g., OIS)
  - Markup Approaches - use last 3
  - SGML - brief introduction
  - Advantages of SGML -> adoption
  - Document modeling - open problem
- 

### Document Translation

#### Electronic Publishing

#### Document Translation

&lt;PRODUCTS&gt;



## SoftQuad Product Catalogue

▶ HoTMetal PRO 4.0  
▶ HoTMetal App. Server  
▶ HoTMetal Power Pack  
▶ Panorama CDWeb  
▶ Panorama Publisher  
▶ Panorama Viewer  
▶ Author/Editor 3.5  
▶ Rules Builder  
▶ Apex  
▶ Sculptor

### Pricing

\$129 US  
\$169 CDN  
£99 UK (+VAT)  
995 FHT (France)  
DM 259 +MwSt  
(Germany)

### System Requirements

Windows 95/NT  
486/33 (Pentium Optimal)  
16MB RAM  
40MB available hard disk  
space

### Sales Inquiries

Email: [sales@sq.com](mailto:sales@sq.com)  
Phone: (416) 544-9000  
Toll-free Order Line  
(North America only):  
(800) 387-2777



SoftQuad

# HoTMetal<sup>™</sup> PRO 4.0

**Introducing HoTMetal Pro 4.0, the award-winning power tool for creating professional-quality Web pages, featuring:**

- **Easy-to-use** features including the HoTMetal Site Maker
- **Great Flexibility** with three authoring environments to choose from
- **Quick use** with hundreds of sample images and files to get you started
- **Incredible Power** with a number of top of the line tools included for Web authors
- **Guaranteed perfect syntax** with automatic rules checking

Order

[Order HoTMetal Pro 4.0](#)

Order

[Upgrade to HoTMetal Pro 4.0 for only  
US\\$49.95](#)

Free!

[Download a free Eval version of HoTMetal  
Pro 4.0](#)More  
Info[View the HoTMetal Pro 4.0 Feature sheet](#)More  
Info[Learn more about the HoTMetal Power Pack](#)



[Academic Press, Inc.](#)

[American Association for the Advancement of Science \(AAAS\)](#)

[American Astronomical Society \(AAS\)](#)

[American Chemical Society \(ACS\)](#)

[American Institute of Aeronautics and Astronautics \(AIAA\)](#)

[American Institute of Physics \(AIP\)](#)

[American Physical Society \(APS\)](#)

[American Society of Agricultural Engineers \(ASAE\)](#)

[American Society of Civil Engineers \(ASCE\)](#)

[American Society of Mechanical Engineers \(ASME\)](#)

[Institution of Electrical Engineers \(IEE\)](#)

[Institute of Electrical and Electronics Engineers \(IEEE\)](#)

[IEEE Computer Society](#)

[John Wiley & Sons](#)

---

[DLI Home](#) | [DLI National Synchronization](#) | [DL Related Information](#)  
[Glossary](#) | [Information Science](#) | [Interspace](#) | [Testbed](#) | [User Evaluation](#)

University of Illinois at Urbana-Champaign Digital Libraries Initiative  
Comments to: External Relations Coordinator, [Susan Harum](#)  
01/18/98



## KEY PAPERS

### **Information Analysis in the Net: The Interspace of the Twenty- First Century.**

Bruce R. Schatz, refereed White Paper for *America in the Age of Information: A Forum on Federal Information and Communications R & D*, July 6-7, National Library of Medicine. sponsored by CIC (Committee on Information and Communications) reporting to the Science Advisor to the President of the United States.

### **Federating Diverse Collections of Scientific Literature**

Bruce R. Schatz, William H. Mischo, Timothy W. Cole, Joseph B. Hardin, Ann P. Bishop, and Hsinchun Chen, University of Illinois DLI Paper from *IEEE Computer Magazine*, theme issue on the US DLI, May 1996

### **Information Retrieval in Digital Libraries: Bringing Search to the Net**

Bruce R. Schatz, *Science*, invited cover article, January 1997

### **Computation Cracks 'Semantic Barriers' Between Databases**

Bruce R. Schatz, *Science*, about Interspace, june 1996

[[showcase](#)] - [[proposal](#)] - [[talks](#)] - [[papers](#)] - [[architecture](#)] - [[reports: technical / quarterly](#)] - [[the team](#)]

[[interspace home](#)] - [[canis home](#)]

comments/questions to [webmaster@canis.uiuc.edu](mailto:webmaster@canis.uiuc.edu)

last updated 01-27-98



## Concept Extraction in the Interspace Prototype

Nuala A. Bennett, Qin He, Conrad Chang, Bruce R. Schatz

**Digital Library Initiative (DLI) Project**  
**CANIS : Community Systems Laboratory**  
**University of Illinois at Urbana-Champaign**  
704 S. Sixth Street, Champaign, IL 61820  
E-mail: {[nabennet](mailto:nabennet), [hqin](mailto:hqin), [t-chang2](mailto:t-chang2), [schatz](mailto:schatz)}  
<http://www.canis.uiuc.edu>

### Abstract

This paper describes the concept extraction for the Interspace Research Project. A comparison was undertaken of four parsers for noun phrase extraction - FastNPE, NPtool, Chopper, and AZ Phraser. FastNPE was found to be the fastest of the parsers, and NPtool the most correct in extracting noun phrases. Both were subsequently implemented into the Concept Extractor module of the Interspace Prototype, which is described in detail. Future work on the Concept Extractor will include image concept extraction and this is described in the final section.

**FULL-TEXT PAPERS** 

---

[[showcase](#)] - [[proposal](#)] - [[talks](#)] - [[papers](#)] - [[architecture](#)] - [reports: [technical](#) / [quarterly](#)] - [[the team](#)]

comments/questions to [webmaster@canis.uiuc.edu](mailto:webmaster@canis.uiuc.edu)  
last updated 02-04-98



## INTER SPACE ARCHITECTURE CONCEPT SPACE GENERATOR

The purpose of the **Concept Space service** is to automatically generate domain-specific thesaurus subsets which represent the concepts and their associations in the underlying information corpus. Concept Space generation is based on a statistical co-occurrence analysis which captures the similarity between each pair of concepts (1).

The greater the similarity between concepts the more relevant they are to one another. Concept Spaces are used in a retrieval environment to assist users in performing functions such as term suggestion (2,3).

(1) Chen, H. and Lynch, K. J., "**Automatic Construction of Networks of Concepts Characterizing Document Databases**", *IEEE Transactions on Systems, Man and Cybernetics*, 22(5) 885-902, Sept./Oct., 1992.

(2) Schatz, B.R., and Chen, H., "**Building Large-Scale Digital Libraries**", *IEEE Computer*, Special Issue on Building Large-scale Digital Libraries, 29(5) 22-27, May 1996.

(3) Schatz, B.R., **Information Retrieval in Digital Libraries: Bringing Search to the Net**, *Science*, 275(5298), 327-334, cover story and lead article, January 17, 1997.

[[showcase](#)] - [[proposal](#)] - [[talks](#)] - [[papers](#)] - [[architecture](#)] - [[reports](#)] - [[the team](#)] -  
[[interspace home](#)] - [[canis home](#)]

comments/questions to [webmaster@canis.uiuc.edu](mailto:webmaster@canis.uiuc.edu)  
last updated 1-5-98

# Interactive Term Suggestion for Users of Digital Libraries: Using Subject Thesauri and Co-occurrence Lists for Information Retrieval

Bruce R. Schatz  
 Eric H. Johnson, [ejohnson@uiuc.edu](mailto:ejohnson@uiuc.edu)  
 Pauline A. Cochrane  
 Digital Library Initiative  
 Grainger Engineering Library Information Center  
 University of Illinois at Urbana-Champaign  
 Urbana, IL 61801, USA

Hsinchun Chen  
[hchen@bpa.arizona.edu](mailto:hchen@bpa.arizona.edu)  
 Department of Management Information Systems  
 University of Arizona, Tucson

## Abstract

The basic problem in information retrieval is that large-scale searches can only match terms specified by the user to terms appearing in documents in the digital library collection. Intermediate sources that support term suggestion can thus enhance retrieval by providing alternative search terms for the user. Term suggestion increases the recall, while interaction enables the user to attempt to not decrease the precision.

We are building a prototype user interface that will become the Web interface for the University of Illinois Digital Library Initiative (DLI) testbed. It supports the principle of multiple views, where different kinds of term suggestors can be used to complement search and each other. This paper discusses its operation with two complementary term suggestors, subject thesauri and co-occurrence lists, and compares their utility. Thesauri are generated by human indexers and place selected terms in a subject hierarchy. Co-occurrence lists are generated by computer and place all terms in frequency order of occurrence together. This paper concludes with a discussion of how multiple views can help provide good quality Search for the Net.

This is a paper about the design of a retrieval system prototype that allows users to simultaneously combine terms offered by different suggestion techniques, not about comparing the merits of each in a systematic and controlled way. It offers no experimental results.

## Introduction to search terms

Effective information retrieval on an on-line document collection closely resembles the problem of effectively searching a library catalog by subject. As opposed to a known-item search, where you know what you want from the start and can provide precise title and/or author information, at the start of a subject search you only know that you want documents "about" something. The set of documents you come away with depends on the set of words you provide to the retrieval system and the ways in which it allows you to apply those words to the database. And even when you have a set of documents that appear relevant to your problem, you can never be sure that there are not more documents in the collection that you might find useful. This illustrates the completeness problem inherent in all information retrieval systems.

To attempt greater completeness of a set of retrieved documents, you might combine into one larger set the results of several different searches, each with a different search term. But here the problem of

# University of Illinois Digital Library Initiative

## Partners Workshop

**May 2-3, 1996**

*Fifth in the Digital Libraries Initiative Workshop Publications Series*

[Letter from NSF Program Official](#)

[DLI Project titles and homepages](#)

[Testbed Technologies for the Distributed Repository Model - Flow Chart](#)

[UIUC DL Testbed Database & Client Technologies](#)

[UIUC DLI Testbed Processing Customization](#)

[Interactive Term Suggestion for Users of Digital Libraries: Using Subject Thesauri and Co-occurrence Lists for Information Retrieval](#)

[Observations of Bibliographic Tools Use at the Grainger Engineering Library](#)

[Software Implementation and User Registration](#)

[Object Worlds and Shifting Infrastructure: Building a Digital Library for Engineers](#)

[Summary of DLI Prototype Usability Tests](#)

[Targeted Study of the First Group of DLI Users: The Physicists](#)

[Future Research: Linking Social and Technical Aspects of Digital Libraries](#)

[HPCwire: The Text-on-Demand E-zine for High Performance Computing](#)

---

[Agenda](#)

[List of Participants](#)

[Minutes](#)

[Partners Address List](#)

# Welcome to the DLI Social Science Team Home Page

[Index](#)[Diary](#)[Internal Reports](#)[Completed Papers](#)[Papers in Progress](#)[Conference  
Presentations](#)[Site Visit and  
Quarterly Reports](#)[Main DLI Page](#)[Web Client-  
DeLiver](#)

send comments or questions to:  
[l-neuma1@uiuc.edu](mailto:l-neuma1@uiuc.edu)

This page consists of links to working papers and [a brief overview of the social science team](#) projects that we have been working on as the social science team for the NSF/ ARPA/ NASA [Digital Library Initiative project](#) being conducted at the University of Illinois.

Our subgroup of the Illinois Digital Library Initiative (DLI), the Social Science Team, has a mandate to study potential and actual use of prototype systems that other subgroups of the DLI build. In addition, we study the web more generally, and how the work of engineers and other scientist will be impacted by and will impact the growth of the information infrastructure.

Our Social Science Team has articulated, from the beginning, a commitment to a three way relationship between users, designers and social scientists, following in a general way the principals of participatory design. We are especially concerned with trying to fit our formative evaluation work to the ideal of this method: close contact and communication between designers and users via a series of mutually generated, iterative prototypes. To this end, we have conducted usability studies with the emergent testbed; observations of current users of electronic systems in the traditional library and beyond; focus groups, interviews and observations with faculty and staff who are potential users; and as use of the testbed continues to grow, transaction log analyses. One of our major concerns is finding a means to fit these all together.

Members of the team include: Ann Bishop, primary investigator; [Leigh Star](#), investigator; Emily Ignacio, graduate assistant; Laura Neumann, graduate assistant; [Cecelia Merkel](#), a graduate assistant; [Bob Sandusky](#), graduate assistant; and Eric Larson, graduate assistant.

---

Toward Functional Requirements for the Digital Library

# **Toward Functional Requirements for the Digital Library (based on focus group interviews with faculty and students)**

20 December, 1994 - Draft

**\* = most important**

1. \*--DL should allow the user to follow citation links forward and backwards (preferably to full documents; otherwise to location information)
2. \*--DL should include an online meta-thesaurus that users can search and browse. The meta-thesaurus should integrated existing thesaurae across disciplines. It should also allow users to incorporate their own terms and edit existing terms. The thesaurus should allow users to type in a few letters of a word and see corresponding terms, should suggest or reference alternatives to users' terms. Users should be able to view no. and type of documents associated with terms and link automatically from thesaurus terms to documents.
3. --DL should include an acronym list to help users identify and search for terms.
4. --Users should be allowed to save a record of their searches and what each search retrieved.
5. \*--Users should be able to search and view individual components of a document (e.g., author/title, abstract, figures, references) in a dynamic manner, specifying for each search which elements should be searched and which displayed.
6. \*--Users should be able to customize their interfaces so that search options, procedures are presented in the manner they like best.
7. --Users should be able to view an overview description of the contents of the testbed.
8. \*--Display of full documents should mimic the look and feel of the article's print version in both page layout and page "flipping" (i.e., users should be able to view multiple pages at once and in quick succession)
9. \*--Users should be able to design and launch their own user profiles for any particular search session, defining what they want and how they want to get it.
10. \*--Users should be able to move easily from query to results and back, rather than moving in the linear fashion common in online systems today, revising a query upon viewing results without having to lose sight of the results or start a query over.
11. \*--Users should be able to easily create personal electronic article collections as a subset of the DL, manipulate and share that collection.

12. --Users should be able to define and set their own access points for searching personal collections derived from the DL
13. --DL should allow on-screen highlighting, bookmarking to help in reading full articles.
14. --Users should have access to DL from home and office.
15. --DL should allow printing of full documents
16. \*--Users should be able to jump to and view individual document components. They should be able to skim, open, or skip individual document components.
17. --DL should facilitate colleague networks: allow users to view list of contact info for authors, construct mailing list of colleagues to send documents to.
18. \*--DL should provide complete and intuitive online help: help balloons, full documentation, help with basic computing, gripe button, sample searches.
19. --DL should allow users to make own links to commonly used external network resources (e.g., pre-print databases, listservs)
20. --Search parameters should include physical location of material not available online
21. --DL should facilitate browsing at shelf, ToC, and article levels: users need overview and zoom capabilities.
22. \*--Interface should resemble a "natural topography" of the information landscape... with a physical layout, dynamically defined (topic, material type, author, etc.)
23. \*--DL should allow serendipitous discovery of "other books on the shelf," "other articles in the journal." Perhaps set browse mode as a purposeful search option: by call no., journal title, etc.

# University of Michigan Digital Library Activities

## DLI General Information

- [Home Page](#)
- [IEEE Computer article](#)
- [Introduction](#)
- [Current Status](#)
- [Technologies](#)
- [Agents, Ontologies](#)

## Campus Strategy

- [Strategies for DL Development](#): partnership of
  - [University Library](#)
  - [Information Technology Division](#)
  - [School of Information](#)
- combine: R&D; technology infrastructure; content access & user services; outreach
- shift to 21st century library model
  - user-centric, collaborative teams, global reach
  - distributed collections, heterogeneous access protocols, just-in-time information delivery
  - mixed funding models, value = access + services
- Registries: [SGML database](#), [subject specialist librarian created digital libraries across Internet](#)
- [Electronic Reserve Shelf](#)
- [Knowledge Navigation Center](#): develop and support teaching and learning projects
- Questions:
  - How does the infrastructure at U. Michigan compare to that at your university?
  - How does this strategy relate to previous services of libraries?

## Projects

- [JSTOR](#): Journal Storage: over 1.2M pages
- [Making of America](#): with Cornell - 5K volumes, [D-Lib article](#): scanning, OCR, SGML encoding, [tif2gif](#), interface
- [Museum Educational Site License Project](#): see also V. 5 N. 8 Oct. 1996 [Information Technology Digest](#)
- [Humanities Text Initiative](#) and [Collaboratory for the Humanities](#)
- [Pricing Electronic Access to Knowledge](#) (PEAK) - 1100 Elsevier journals with flexible pricing
- [Papryology](#)
- [Middle English Compendium Demo](#)
- [American Verse](#)
- [TULIP](#)
- [NDLF](#)
- Questions:
  - Which of these projects do you find most interesting? Why?
  - Which of these projects should your university become involved in?

## Technical Approaches

- [see especially 1996 Ann Arbor Conf. on Electronic Records R & D](#)
  - Problem scenarios (see bullet list under **The Importance of Digital Preservation**)
  - Research questions (see **The 10 Research Questions**)
  - Research results: possible, requires changes and new types of efforts (see bullet list under **Research Projects and Results**)
  - [International Council on Archives](#): see **Guide for Managing Electronic Records from an Archival Perspective**, survey, literature review
  - [Australian Council of Archives statement](#)
- [Advanced Interfaces](#)
- [Pad++](#)
- [Ontology - Concept Descriptions](#) and [May 1997 slides](#)
- [Learning Agents](#)
- [Teaching and Learning Project](#)
- [Artemis Java Interface to UMDL Production System](#)
- [SGML creation and delivery](#)
  - enormous collection: 2M pages
  - [flowchart](#)
  - [SGML Server Program](#): middleware, training
  - cross collection searching
  - multiple representations
  -
- [Leveraging rich document formats](#)
  - patterns of use
  - ease of changing delivery: new standards (HTML), new rendering/packaging
  - collection management
  - Panorama, XML support by W3C
- Questions:
  - Will the agent and ontology approach work? Soon? For production DLs?
  - What is the support needed for establishing a digital library following the UMDL approach? Training?
  - What interfaces for DLs will be usable?

THE NSF/DARPA/NASA SPONSORED  
UNIVERSITY OF MICHIGAN  
DIGITAL LIBRARY  
PROJECT



**Digital Library Initiative University of Michigan**

From *Computer* theme issue on the US Digital Library Initiative, May 1996

*In the University of Michigan Digital Library, interacting software agents cooperate and compete within a virtual information economy to provide library services to students, researchers, and educators.*

## **Toward Inquiry-Based Education Through Interacting Software Agents**

Daniel E. Atkins, William P. Birmingham, Edmund H. Durfee, Eric J. Glover, Tracy Mullen, Elke A. Rundensteiner, Elliot Soloway, José M. Vidal, Raven Wallace, and Michael P. Wellman, *University of Michigan*

Providing true access to the human record means offering relevant information without prohibitive search time or an overwhelming choice among sources. Conventional libraries provide such access through two mechanisms: information organization and librarian services. Librarians themselves often rely on services like information systems or bibliographic databases to do their jobs.

Digital libraries must likewise provide organizational schemes and a wide variety of services. Most observers focus on the vast amount of information digital libraries will offer, delivered in new and interesting ways. However, we believe it is the bounty of services that will ultimately demonstrate the potential of digital libraries.

The University of Michigan Digital Library (UMDL) project<sup>[1]</sup> is creating an infrastructure for rendering library services over a digital network. When fully developed, the UMDL will provide a wealth of information sources and library services. Of course, we cannot anticipate all the services that will eventually constitute a digital library. We therefore designed the UMDL to let third-party developers expand the library with new services and collections.

We are deploying the UMDL in three arenas: secondary-school science classrooms, the University of Michigan library, and space-science laboratories. Computer skills, information demands, and level of subject knowledge vary greatly among these user populations. Addressing the needs of high school students within a general-purpose digital library particularly stresses the flexibility of our underlying architecture. The UMDL must support services quite distinct from those that other digital libraries and the World Wide Web offer.

Many researchers and policy groups argue that students should engage in

**THE NSF/DARPA/NASA SPONSORED UNIVERSITY OF MICHIGAN PROJECT** **DIGITAL LIBRARY**

**UMDL TECHNOLOGIES**

**ARCHITECTURE: AGENTS AND ONTOLOGIES**

**ACCESS: ARTEMIS INTERFACE**

**CONTENT: COLLECTIONS**

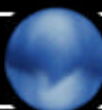
**ECONOMY: COMPUTATIONAL MARKETS**

**ADVANCED USER INTERFACE**

**CONSPECTUS & IR**

**PRODUCTION SYSTEM**

MISSION • ACCOMPLISHMENTS • IN ACTION



UMDL TECHNOLOGIES • IMPACT • TEAM

# Strategies for Digital Library Development: The University of Michigan Case Study

Wendy P. Lougee  
University of Michigan

Prepared for TICER, Second International Summer School, Tilburg, The Netherlands, August 1997

---

## Abstract

Beginning in 1993, the [University of Michigan](#) launched a partnership between the [Information Technology Division](#), [University Library](#), and [School of Information](#) to develop the campus environment for networked information resources. The collaborative effort focuses on building critical infrastructure components as well as useful information resources and services. A strategy has been employed which develops the capability, within projects, to support a variety of information formats, search engines, and applications. As the program matured, a virtual organization has been created which provides production support for digital collections and manages programs of access for University users and remote customers. Critical to the success of the joint program has been the integration of expertise and resources of the partner organizations and a commitment to address campuswide practices and policies.

---

## Introduction

There is no single strategy to employ in building a digital library. Strategies, like libraries themselves, are born out of the time, the culture, and opportunities. I've been asked to share the [University of Michigan](#) (UM) experience as a case study and hope to convey the context and opportunity-based factors which have shaped our endeavor. Our story is really a story about relationships and the critical interweaving of expertise and resources which, I believe, has been the hallmark of our success.


This overview is structured to provide: the history of our program, some perspective on the forces shaping digital libraries, a description of the [University of Michigan Digital Library](#), and finally some sense of how our organization has matured from a series of projects to a more mature production-focused service.


## History of the Digital Library Program


The roots of our campuswide digital library program can be found in a year-long committee


JOURNAL  
STORAGE

REDEFINING ACCESS TO SCHOLARLY LITERATURE

ENTER JSTOR 

 ABOUT JSTOR

NEWS & NOTES 

 DEMO

[Enter JSTOR](#) | [About JSTOR](#) | [News & Notes](#) | [Demonstration Database](#)

©1998 JSTOR  
JSTOR® and the JSTOR logo are Reg. U.S. Pat. & Tm. Off.  
All Rights Reserved  
[Contact JSTOR](#)

[BROWSE](#)[SEARCH](#)[RANDOM](#)[HELP](#)[ABOUT](#)

---

The original MESL cooperative agreement, concludes **June 1, 1998.**

**Good News!** Several of the museums have agreed to extend Michigan's use of their images. The images and searchable text will be migrated to a new system in [Digital Library Production Services](#), [Image Services](#) by the end of June 1998. In the meantime, please continue to use this site, and also come and see the new system, including the [History of Art Department Visual Resources Collections](#).

**Extended use has been granted by:** George Eastman House, Museum of Fine Arts: Houston, Fowler Museum of Cultural History: UCLA, and National Museum of American Art.  
**Still pending are:** Library of Congress, and National Gallery of Art.

**Concluded:** Harvard University Art Museums.

Please see [Background and Purpose](#) for more info.

---

Users of these images and texts agree to adhere to the [Conditions of Use](#).

The **Museum Educational Site Licensing Project** is a pilot that exists to make museum information more accessible through electronic technology. Please explore the thousands of images of museum objects that are provided for educational use at the **University of Michigan**. This material is restricted to local users under a license agreement executed by all participating institutions.

---

Please send questions and comments to the MESL Web Team at... [mesl.web@umich.edu](mailto:mesl.web@umich.edu)



*Fifteen of the nation's largest research libraries and archives have agreed to cooperate on defining what must be done to bring together--from across the nation and beyond--digitized materials that will be made accessible to students, scholars, and citizens everywhere, and that document the building and dynamics of United States heritage and cultures.*

---

## What's New NEW!

- *Name Change:* National Digital Library Federation becomes **Digital Library Federation**
- 

## About the Digital Library Federation (DLF)

- [America's Heritage: Mission and Goals](#) for a Digital Library Federation
  - [DLF Constituted as Charter Organization: Adopts Three-Point Agenda](#)
  - [DLF Policy Committee Members](#)
  - [DLF Planning Task Force Members](#)
- 

## DLF Policy Committee

Summary of Meetings:

- [June 19, 1996](#)
  - [September 18-19, 1996](#)
  - [November 22, 1996](#)
- 

## DLF Planning Task Force

- [DLF Planning Task Force: Final Report](#)

Summary of Meetings:

- [June 13, 1995](#)
- [November 14, 1995](#)
- [January 29-30, 1996](#)
- [March 18-19, 1996](#)

- [September 18-19, 1996](#)
- 

## Federation Member Web Sites and Digital Library Projects

- [Commission on Preservation and Access](#)
  - [Columbia University](#)
    - [New York State Museum Bulletins Project](#)
  - [Cornell University](#)
    - [Networked Computer Science Technical Reports Library](#)
    - [Prototype Cornell Digital Library \(includes Making of American Project\)](#)
  - [Emory University](#)
    - [Emory University Virtual Library](#)
  - [Harvard University](#)
    - [Information Infrastructure Project](#)
  - [Library of Congress](#)
    - [American Memory](#)
  - [National Archives and Records Administration](#)
  - [New York Public Library](#)
  - [Pennsylvania State University](#)
  - [Princeton University](#)
  - [Stanford University](#)
    - [Stanford Digital Libraries Project](#)
    - [Stanford University Computer Science Electronic and Technical Reports Library](#)
  - [University of California, Berkeley](#)
    - [UC Berkeley Digital Library Initiatives](#)
    - [NSF Digital Library Project](#)
    - [Computer Sciences Technical Reports Project \(NCSTRL\)](#)
  - [University of Michigan](#)
    - [Digital Library Initiatives](#)
    - [NSF Digital Library Project](#)
    - [Making of America \(MOA\)](#)
  - [University of Southern California](#)
  - [University of Tennessee](#)
  - [Yale University](#)
    - [Open Book Project](#)
- 

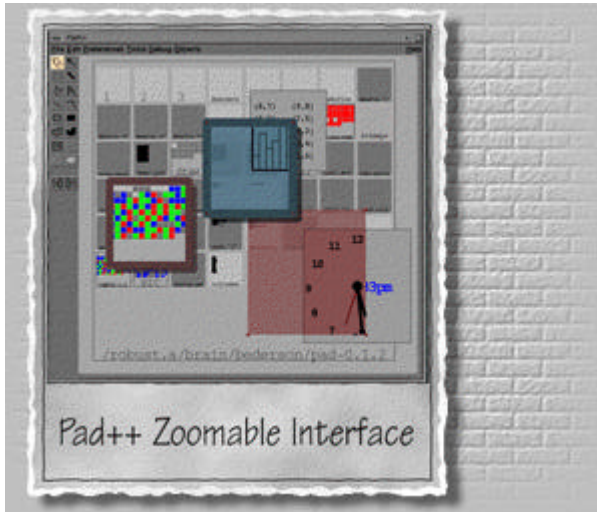
## Federation-sponsored Meetings and Conferences

- [Organizing the Global Digital Library II and Naming Conventions \(May 21-22, 1996\)](#)
- 

## Other Internet Resources on Digital Libraries

# Pad++: Zooming User Interfaces (ZUIs)

University of Maryland, College Park  
Human-Computer Interaction Lab



We are exploring Zooming User Interfaces (ZUIs) where zooming is a fundamental part of the user's interaction with the computer (also known as *multiscale interfaces*).

in collaboration with [UCSD](#), [NYU](#), and [UNM](#)

---

## You can find out about Pad++ through

- [A guided tour](#)
- [Frequently asked questions](#)
- [On-line papers](#)
- [Documentation](#)
- [Some places we've been in the press](#)
- [Download it now](#)

Zooming [Site Map](#)

Check out Ben Shneiderman's [use of Pad++](#) in his plenary talk at CHI98.

Pad++ is supported in part by DARPA grant #N660011-94-C-6039.



# Just-in-time Conversion, Just-in-case Collections

## Effectively leveraging rich document formats for the WWW

John Price-Wilkin  
Head, Digital Library Production Service  
University of Michigan  
Ann Arbor, Michigan  
[jpwilkin@umich.edu](mailto:jpwilkin@umich.edu)

D-Lib Magazine, May 1997

ISSN 1082-9873

- 
1. [Introduction](#)
  2. [Why on-the-fly?](#)
    - 2.1 [Converting for actual use, not potential use](#)
      - 2.1.1 [Case 1: Grolier's \*Encyclopedia Americana\*](#)
      - 2.1.2 [Case 2: Making of America](#)
    - 2.2 ["Re-purposing" and managing data](#)
      - 2.2.1 [Flexible delivery](#)
      - 2.2.2 [Collection management](#)
  3. [Conclusion](#)
- 

## 1. Introduction

The University of Michigan's Digital Library Production Service (DLPS) has developed substantial experience with dynamic generation of Web-specific derivatives from non-HTML sources based on several key projects and consideration of how users work with key resources. This article is based on DLPS's experience and resultant policies and practices that guide present and future projects. In a rapidly changing world where the implications of information technologies for broad yet differentiated clienteles are mysterious, we hope that our experience will contribute to a better understanding of practical strategies.

The WWW has long included the ability to offer access to documents stored in formats other than HTML. Beginning with NCSA's "htbin" mechanisms, and soon after using the now widely embraced Common Gateway Interface (CGI), managers of large document collections have been able to store materials in a variety of formats, while offering these documents to a wide consumer base. This was the model used by the author in 1993 (when NCSA introduced Mosaic) to build access to large collections of documents stored in a variety of forms of SGML.<sup>1</sup> Once CGI was

## [Library of Congress:](#)

---

- American Memory <http://lcweb2.loc.gov/>
  - Call about American Memory <http://lcweb2.loc.gov/ammem/award/>
  - Sponsors and Contributors to the National Digital Library Program  
<http://lcweb2.loc.gov/ammem/sponsors.html>
- 

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Centers\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

[\[text-only\]](#)

June 22, 1998

**Today in History**

---

*The LIBRARY of CONGRESS*

---



32 Collections Now Available

**SEARCH**[Options](#) | Simple search: **BROWSE**[Titles](#) | [Topics](#)  
[Library Divisions](#)

----- Collection Types -----

[Photos & Prints](#)[Documents](#)[Motion Pictures](#)[Maps](#)[Sound Recordings](#)**LEARN**Organized help for using the  
collections

---

**Show case**

---

---

**Sample Collection**

---

[Washington](#)**New Offerings**[Lincoln](#)[A Century  
of Lawmaking](#)[Northern  
Great  
Plains](#)[Map Collections](#)

[A Unique Public-Private Partnership](#)  
[Supporting the National Digital Library](#)

See who is helping to bring a virtual library to all Americans for the 21st Century


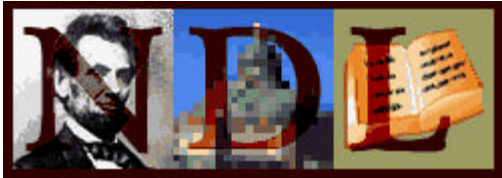

[National Science Foundation Digital Libraries Initiative](#)  
Co-sponsored by the Library of Congress

**[Previews and Future Collections](#)****[How to View These Collections](#)**[LC/Ameritech Competition](#)[Copyright and other Restrictions](#)[Background Papers and Technical  
Information](#)[About the Special Collections of  
the Library of Congress](#)Send comments about these historical collections to [ndlpcoll@loc.gov](mailto:ndlpcoll@loc.gov)

am 4-28-98

**Library of Congress**  
General Comments:  
[lcweb@loc.gov](mailto:lcweb@loc.gov)

**NOTICE**



## A Unique Public-Private Partnership

*Supporting the National Digital Library*

The Library of Congress, with the bipartisan support of the United States Congress, the Executive Branch, and America's entrepreneurial and philanthropic leadership, is bringing the National Digital Library to the nation.

The Library has proposed a public-private partnership for fiscal years 1996 - 2000 that would include a minimum of \$3 million in annual appropriations from Congress. To date, the Congress has appropriated more than \$9 million toward a total appropriation of

\$15 million by the year 2000.

To achieve the program's fundraising goal of \$60 million, the Library has embarked on a major campaign to raise the remaining \$45 million from private funds.

### **Sponsors and Contributors to the National Digital Library Program**

The Library gratefully acknowledges the generosity of the following sponsors and contributors whose support is instrumental to the success of the National Digital Library.

The United States Congress

#### **Founding Sponsors\***

Mr. John W. Kluge  
The David and Lucile Packard Foundation

\*Indicates contributions and pledges of \$5 million or more

#### **Charter Sponsors\*\***

Ameritech  
Anonymous  
AT&T Foundation  
Bell Atlantic Corporation  
Citicorp Foundation  
Discovery Communications, Inc.  
Donaldson, Lufkin & Jenrette  
Eastman Kodak Company  
H.F. Lenfest  
Jones Family Foundation  
Glenn R. Jones (Jones Intercable)  
Federal Express Corporation  
W.K. Kellogg Foundation  
Laurance S. and Mary French Rockefeller  
McCormick Tribune Foundation  
Pew Charitable Trusts  
Occidental Petroleum Corporation  
Reuters

\*\*Indicates contributions and pledges of \$1 million to \$5 million

## **Contributors**

Bankers Trust  
COMPAQ Computer Corporation  
R.R. Donnelly & Sons  
The Ford Foundation  
The Hearst Foundation, Inc.  
David H. Koch Charitable Foundation  
Mr. Carl H. Lindner  
Lucent Technologies  
NYNEX Foundation  
Shell Foundation  
Texaco Foundation

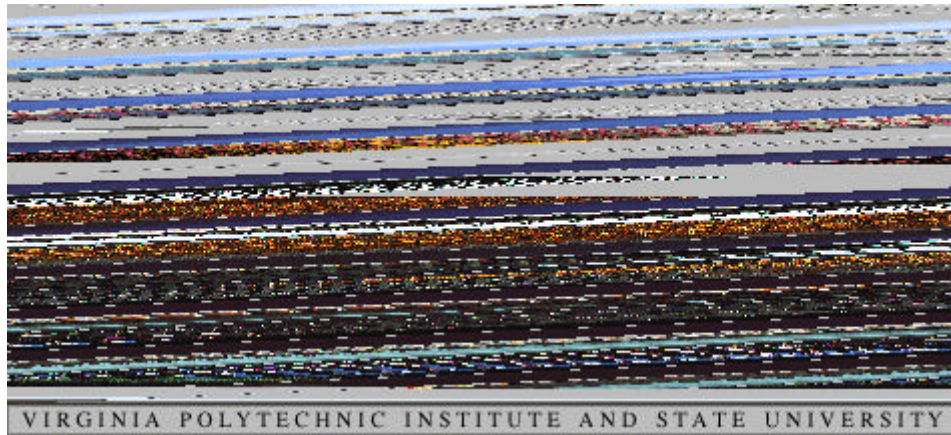
## **In-kind Contributors**

Hewlett-Packard Company  
International Business Machines Corporation  
LizardTech

---

## UK Electronic Libraries Programme ([eLib](#))

- There are online [working papers](#)
- It is funded by the Joint Information Systems Committee (JISC).
- It is based on the Libraries Review by the UK Higher Education Funding Councils, chaired by Professor Sir Brian Follett in 1993. They prepared the [Joint Funding Council's Libraries Review Group Report, Prof. Sir Brian Follett, HEFCE, 1993](#).
- As a response, JISC started eLib with 15 million pounds over 3 years, to engage the Higher Education community in developing and shaping the implementation of the electronic library.
- There have been 2 separate calls and over 60 projects in areas:
  - access to network resources
  - digitisation
  - document delivery
  - electronic journals
  - electronic short loan collections
  - images
  - on demand publishing
  - pre-prints and grey literature
  - quality assurance
  - supporting studies
  - training and awareness
- One part is the [Arts and Humanities Data Service](#) and its service for the Visual Arts.
- On preservation, a workshop was held, with report Long Term Preservation of Electronic Materials: a JISC/British Library workshop as part of the Electronic Libraries Programme, Organised by UKOLN, 27-28 November 1995, U. of Warwick, prepared by the Mark Fresko Consultancy, 1995, also available [online](#).
- [Ariadne](#) is an eLib ejournal providing current information on eLib and digital libraries in general.



# CS Dept. NSF-Supported Education Infrastructure Project / ei.cs.vt.edu

Welcome to the home page for the [NSF-supported](#) project "Interactive Learning with a Digital Library in Computer Science". We hope you find some of the results of our project useful! Please [send me comments and suggestions](#)! Thanks, Prof. E. A. Fox.

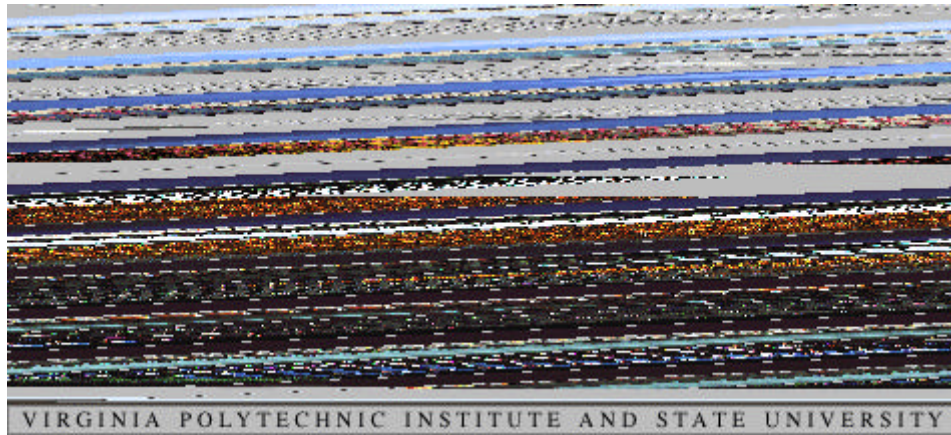
- [Courses](#) (over 40, with over 10K files)
- [Search ei.cs.vt.edu etc. \(e.g., all courses\)](#)
- [National EI Projects Home Page](#)
- [Computer Science Teaching Center](#)
- [Curriculum Resources in Interactive Multimedia](#)
- [Computer Science Education Innovation Workshop](#) June 15-21, 1997 **NEW!**
- [QUIZIT Software and thesis](#) **NEW!**

## Papers:

- [QUIZIT: An interactive online quiz system for WWW-based instruction](#) - Tinoco, L. C., Fox, E. A., Ehrich, R. W, Fuks, H. In Proceedings of the VII Brazilian Symposium on Educational Technology. Belo Horizonte, Brazil, Nov. 1996.
- Online Evaluation in WWW-based Courseware - Tinoco, L. C., Fox, E. A., Barnette, N. D. In Proceedings ACM SIGCSE'97, San Jose, Feb. 1997:  
[paper in PDF](#), [presentation in PowerPoint](#)
- [Audio and Video Tutorials on Popular Tools and Systems](#)
- [Project Overview](#)
- [ENVISION Project \(that led to development of the digital library\) Final Report](#)
- [SWAN \(algorithm vizualization system\)](#)
- [Electronic Submissions of Student Programming Assignments used in CS3204](#)
- [Faculty Development Institute](#)
- [References](#)
  - [Project Overview \(for FIE'96, in PDF\)](#)
  - [Project Interim Report, Oct. 1996](#)
  - [Project Report for NSF EI PI Meeting, Nov. 1996](#)

---

This DEC Alpha, ei.cs.vt.edu, supports the Virginia Tech/Norfolk State University project "Interactive



# CS Courses

Welcome to one of the largest (over 40 courses, over 10K files) repositories of Computer Science courseware! I hope you benefit and [send me comments and suggestions!](#)

Regards, Prof. E. A. Fox for

[Virginia Tech CS Dept.'s NSF Education Infrastructure Project](#)

## [Search ei.cs.vt.edu etc. \(e.g., all courses\)](#)

- [CS1004: Computer Literacy](#)
- [CS1014: Numerical Computational Techniques](#)
- [CS1024: Computing For Business](#)
- [MaSc1044: Computer Science: A Liberal Arts Approach](#)
- [CS1044: Programming in C](#)
- [CS1104: Introduction to Computer Science](#)
- [CS1205: Operating System Tools I](#)
- [CS1344: Introduction to C Programming](#)
- [CS1206: Operating System Tools II](#)
- [CS1604: Computers and Networked Information](#)
- [CS1704: Introduction to Data Structures & Software Engineering](#)
- [CS2304: Self Study in a Programming System \(Java\)](#)
- [CS2304: Self Study in a Programming System \(UNIX\)](#)
- [CS2604: Data Structures and File Processing](#)
- [CS2704: Object-Oriented Software Design and Construction](#)
- [CS2964: Field Studies](#)
- [Honors 3004: Digital Libraries](#)
- [Honors 3004: Multimedia Technology and Projects \(1996\)](#)
- [UH3004: High-Performance Scientific Computing](#)
- [CS3204: Operating Systems](#)
- [CS3304: Comparative Languages](#)
- [CS3304sm: Comparative Languages \(Summer, 1997, offering\)](#)
- [CS/Math 3414: Numerical Methods](#)
- [CS3604: Professionalism in Computing](#)
- [CS3724: Introduction to Human-Computer Interaction](#)
- [CS4104: Data and Algorithm Analysis](#)

- [CS4114: Formal Languages](#)
  - [CS4124: Theory of Computation](#)
  - [CS4204: Computer Graphics](#)
  - [CS4214: Simulation and Modeling](#)
  - [CS4234: Parallel and Distributed Computing](#)
  - [CS4414: Issues in Scientific Computing](#)
  - [CS4504: Computer Organization](#)
  - [CS4624: Multimedia, Hypertext and Information Access](#)
  - [CS4964: Field Studies](#)
  - [CS4984: WWW - The Underlying Technology](#)
  - [CS5014: Research Methods in Computer Science](#)
  - [CS5024: Models and Analysis](#)
  - [CS5034: Models of Computation](#)
  - [CS5104: Computability and Formal Languages](#)
  - [CS5114: Theory of Algorithms](#)
  - [CS5204: Operating Systems](#)
  - [CS5224: Systems Simulation](#)
  - [CS5314: Concepts of Programming Languages](#)
  - [CS/ECE5515: Computer Architecture](#)
  - [CS/ECE5516: Communication Networks](#)
  - [CS5604: Information Storage and Retrieval](#)
  - [CS5614: Database Management Systems](#)
  - [CS5724: Models and Theories of HCI](#)
  - [CS5734: Computer-Supported Cooperative Work](#)
  - [CS5814: Digital Picture Processing](#)
  - [CS6104: Algorithmic Number Theory](#)
  - [CS6204: The World-Wide Web: Beyond the Basics](#)
  - [CS6204: Java and the WWW](#)
  - [CS6404: Advanced Topics in Mathematical Software](#)
  - [CS6604: Digital Libraries](#)
  - [CS6604: Interactive Accessibility \(1995\)](#)
  - [CS6724: HCI of Collaborative Systems](#)
  - [Digital Libraries - self study](#)
- 

## Catalog Pages

- [Ugrad](#)
- [Grad](#)

## Class Data Archives

- [EI Archives](#)
- [CS Department Archives](#)

## Usage Statistics

All materials prepared for these [Dept. of Computer Science](#) courses are

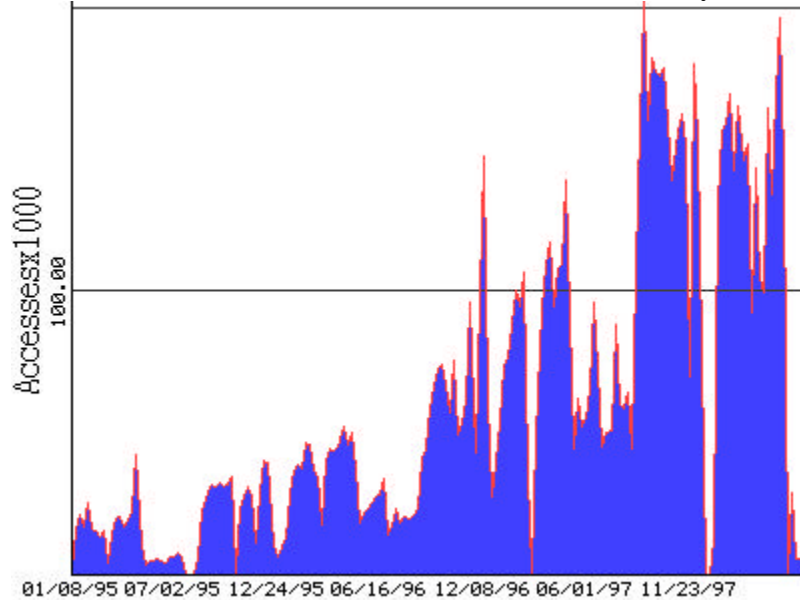
# EI Statistics

## History from 01/08/95 to 06/20/98

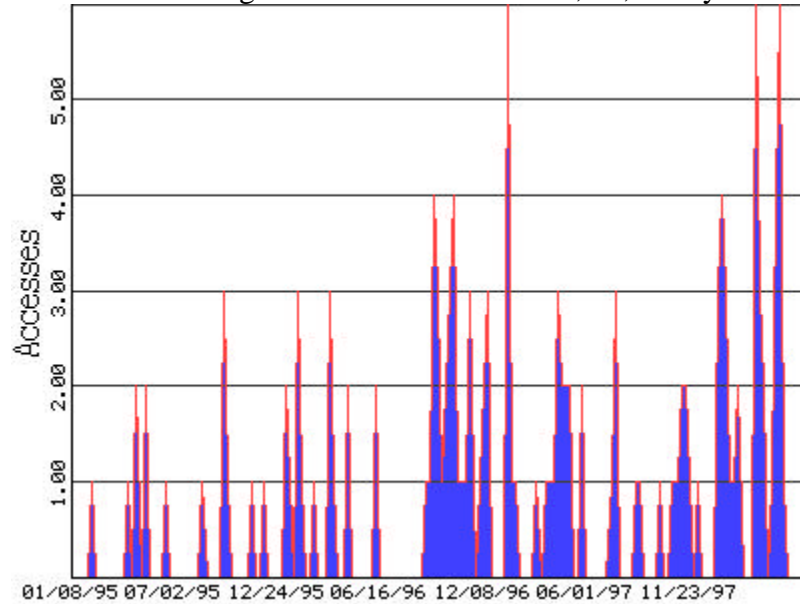
[Skip to Weekly Reports](#)

### Totals

- Overall Hits: 10,706,544 accesses, 125,705,431,180 bytes



- Courses Home Page Accesses: 136 accesses, 11,940 bytes



### Reports

- [Reports for the Year 1998](#)

D-Lib Magazine, December 1995

## Project Briefings and Updates

---

# Making a Digital Library

### *The Chemistry Online Retrieval Experiment A Summary of the CORE Project (1991-1995) December 1995*

Contributed by:

Richard Entlich, Cornell University

Lorin Garson, American Chemical Society <http://pubs.acs.org>

Michael Lesk, Bellcore <http://community.bellcore.com/lesk/home-page.html>

Lorraine Normore, Chemical Abstracts Service

Jan Olsen, Cornell University

Stuart Weibel, OCLC <http://www.oclc.org:5046/~weibel>

The CORE project was an electronic library prototype of primary journal articles in chemistry, containing about four years of twenty primary journals published by the American Chemical Society (about 400,000 pages). CORE included both scanned images and an SGML (Standard Generalized Markup Language) marked-up version for on-the-fly rendering for screen display. Each page was scanned and segmented, with graphical units isolated and linked to figure references in the articles. The original machine-readable typography was converted to SGML format and the results were used to build databases with indexes for full-text Boolean searching.

Each page image was stored as a 300 dpi bitonal image for printing, and 100 dpi greyscale for screen display. All text data and the most recent page images were available on Unix-based magnetic storage at any given time, with additional (older) page images stored on a WORM (Write Once, Read Many) jukebox.

Complex scientific material (superscripts, tables, equations, special fonts and symbols, etc.) presents substantial problems for representation and display, especially when the material is being converted from previously published information, as were these journals.

The tasks of building and maintaining electronic journal databases remains formidable (especially if conversion from older formats is involved). However, experiences with chemists in this project suggest that electronic publishing will be popular with scholars, even though there remain significant disadvantages and impediments to adoption.

Analysis of user studies and transaction logs is ongoing and will be submitted for publication in the near future.

# TULIP - The University Licensing Program

---

When you scroll further down this page you'll find

- [Introduction](#)
  - [The TULIP Final report](#)
  - [TULIP Newsletters](#)
  - The [Journal Titles](#) in TULIP
  - The [Universities](#) involved in TULIP
  - [Contact information](#)
- 

## Introduction

TULIP is a cooperative research project testing system for networked delivery and use of journals, performed by Elsevier Science and [nine Universities](#) in the USA. The participants set three objectives at the outset:

### Technical

To determine the technical feasibility of networked distribution to and across institutions with varying levels of sophistication in their technical infrastructure. "Networked distribution" means sending the information both across the national Internet and over campus networks to the desktops of students and faculty. Elsevier will deliver the journal information to participating universities in standard formats. The universities will incorporate the information in local prototype or operational systems. A wide variety of delivery alternatives, search and retrieval systems and print-on-demand options will be compared.

### Organizational and economic

To understand, through the implementation of prototypes, alternative costing, pricing, subscription and market models that may be "viable" in electronic distribution scenarios; comparing such models with existing print-then- distribute models; and understanding the role of campus organizational units under such scenarios. The overall goal is to reduce the unit cost of information delivery and retrieval. "Viable" means economically and functionally acceptable to all parties.

### User behaviour

To study reader usage patterns under different distribution (technical, organizational and economic) situations. Improvement in the functionality of the information, whether as to article structure or retrieval tools, will also be considered. Certain data will be collected uniformly at all sites for analysis in the aggregate and for comparison among different systems.

[Return](#) to top of this page

---

## The TULIP Final report



## IBM Digital Library Version 2

*An end-to-end solution for managing multimedia content.*



- Reach new markets and establish new sources of revenue through improved management and reuse of media assets
- Preserve your assets from physical deterioration
- Protect your assets with advanced rights management
- Consolidate management of text, images, audio and video with easier, faster access
- Save money with electronic delivery
- Be ready for Year 2000

Developed with a variety of key customers and business partners, IBM Digital Library has helped businesses and institutions in the areas of higher education, media and publishing, entertainment, culture, health, and commerce provide greater access to their digital assets, while enhancing their growth and new revenue opportunities. IBM Digital Library Version 2, building on the strength of these technologies, enables literally petabytes of text, images, audio and video, to be created or transformed into digital form and distributed over any network, with security, to users around the world.

### New features in Version 2 include:

- Enhanced platform support now includes Windows NT and Macintosh
- Multi-language development tools
- Enhanced rights management
- Integrated multi-search capability
- Integrated support for IBM media servers
- Java-based system administration interface
- Easier installation with graphical user interface guides
- IBM DB2 Universal Database components

### [Key Features of IBM Digital Library Version 2](#)

[IBM Digital Library Architecture](#)

[IBM Digital Library at a Glance](#)



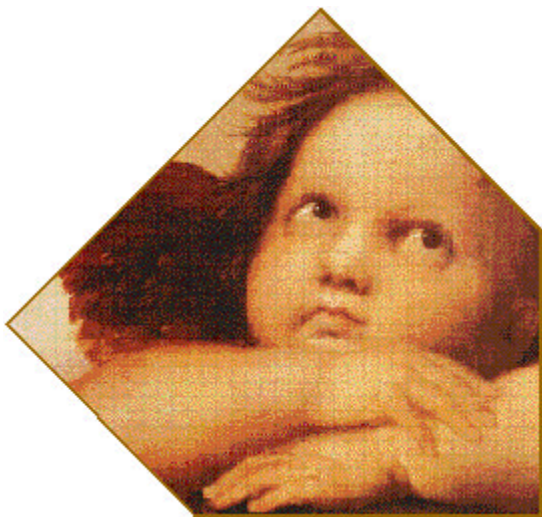
©1998 IBM Corporation



## IBM Digital Library Collection Treasury

### Extending worldwide access to special collections

Museums and libraries holding special, unique collections have two distinct missions. On the one hand, these institutions want to share their collections with as many people as possible. On the other hand, as copyright holders or caretakers for irreplaceable cultural artifacts, they want to minimize exposure to their holdings, in order to ensure their preservation.



Working with museums and libraries, IBM has developed IBM Digital Library(TM) Collection Treasury--a solution that will enable these institutions to maximize their ability to share their holdings, while substantially reducing risks to preservation. The solution, based on IBM Digital Library technology, enables institutions to provide worldwide access to their holdings via the Internet and vastly increase the potential "visitor traffic" to the institution. At the same time, none of these visitors ever touch the holdings, reducing risk of damage, loss and theft.

In effect, IBM Digital Library Collection Treasury extends the walls of the traditional library or museum, making possible "virtual libraries" and "virtual museums" that can be explored at any time, from anywhere in the world.

Sophisticated search technology enables the "virtual visitor" to sort through vast collections with ease, finding a specific document or image among thousands. This powerful search and access capability also enables the visitor to conduct lengthy ongoing research conveniently from a distance, integrating this research into a normal daily schedule.

By contrast, consider the traditional situation. Many university researchers are a plane ride away from any specific museum or library. If they wish to conduct research involving documents or images at a given institution, finding travel funds is often difficult. If they find the funds, they need to set aside at least two days to travel to and from the site, and then need to compress their research time into an arbitrary block of time reserved for the research. If they have follow-up questions, another trip must be scheduled--or the research questions go unanswered.

Now, IBM Digital Library Collection Treasury offers the potential to transform museums and libraries for the digital age, making their unique collections more relevant to more people than ever before. Thanks to IBM's rights management and watermarking technologies, institutions will continue to exercise tremendous control over their collections--minimizing illegal copying, and restricting access to any specific audience. What has changed is the breadth of access: with the tyranny of distance removed, libraries and museums can truly "serve the world."

### IBM Digital Library at work

IBM Digital Library Collection Treasury shows the power of IBM Digital Library at work. The same core IBM Digital Library technology is being used by international historical archives to digitize and manage their ancient manuscripts, and by major Hollywood studios to manage brand new production content.

With IBM Digital Library's special rights management and watermarking capabilities, valuable original holdings can be managed online with appropriate restrictions on audience access and reuse of the images. Future capabilities might include electronic commerce extensions that can provide revenue to content owners through fee-based access and reuse of content.



This site received a  
4 star rating from McKinley Group's editorial team  
and  
"Best of the Web!" by Snap! Online




## QBIC<sup>TM</sup> -- IBM's Query By Image Content

On-line collections of images are growing larger and more common, and tools are needed to efficiently manage, organize, and navigate through them. We have developed the QBIC system which lets you make queries of large image databases based on visual image content -- properties such as color percentages, color layout, and textures occurring in the images. Such queries use the visual properties of images, so you can match colors, textures and their positions without describing them in words. Content based queries are often combined with text and keyword predicates to get powerful retrieval methods for image and multimedia databases.


QBIC is available for download with a free 90 day trial license. The download package includes the image indexing and search engine (for AIX, Linux, Windows NT/Windows95, and OS/2), a Web front end, APIs for imbedding QBIC in other applications or extending QBIC with new query functions, and even a sample image collection. You can download it from [IBM software download site](http://www.ibm.com/software/awdtools/qbic/).

### News Bulletins

- 
 IBM AND [MAGNIFI](#) ANNOUNCE LICENSING AGREEMENT -- IBM Research Technology gives Magnifi the cutting edge in Visual Searching Capabilities. Click [here](#) for more information.

- IBM, VIRAGE [ANNOUNCE](#) BROAD CROSS-LICENSING AGREEMENT
- Stay tuned here for announcements about a coming new release of QBIC.

### QBIC demos on this site:

-  [A collection of all U.S. stamps before 1995, searchable by QBIC and DB2 with a Java GUI.](#)
- [A prototype trademark browsing and retrieval site.](#)
- [Our old Stock Photo demo spruced up with our Java GUI.](#)

Other sites showcasing QBIC technology:



## IBM Digital Library Media & Entertainment Solutions

### ***News Archive Solution***

***A solution for news asset management  
--spanning text, photos, graphics, transcripts, film, video and audio***

Imagine if retail stores took their cash income at the end of each day, stuffed it in a mattress, and forgot about it forever.

That may sound ludicrous. But without good management, it's essentially what happens in broadcast news when video assets aren't managed as valuable cash assets.



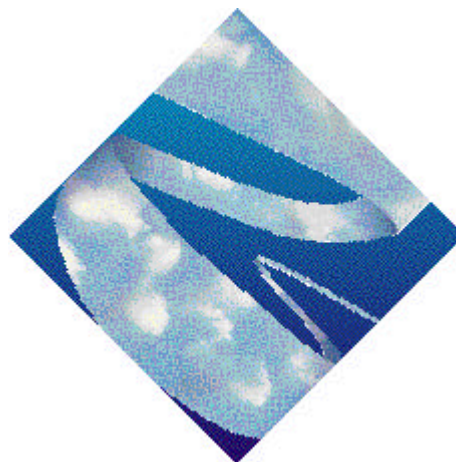
Broadcast news operations work hard every day, and spend a great deal of money, to capture the best video content possible for current news stories. Much less attention is spent on managing and reusing this content after the fact. Since video content is essentially the cash commodity of the business, that's like stuffing wads of hundred dollar bills in a mattress.

Do the math. The average finished minute of broadcast news costs the industry thousands of dollars to produce. However, once produced, that finished minute can be reused and resold again and again. It's like selling lemonade, and then putting the lemonade back in the bottle to sell again.

Often, it's more than just an economic issue. News, by definition, involves the reporting of "new" events, requiring an instant and professional response to an unpredictable world. When a major leader is overthrown in a coup, you want footage of that leader now. When two sports teams engage in a blockbuster trade, you want highlight footage of the traded players now.

The IBM News Archive Solution makes such responsiveness easily possible, helping you to break news faster, beat the competition with higher quality news stories, and end the day with more cash in your pocket than in the proverbial mattress.

Designed specifically to serve the needs of a major TV network, the IBM News Archive Solution is robust enough to support any broadcast news operation in existence today. A complete solution, it packages application software, systems software, implementation services and support. The packaged solution means that you don't have to be a multimedia database expert yourself to integrate powerful database management into your news operation. With IBM's help, you can be up and running quickly.



The News Archive Solution helps you manage all the elements that make up your news archives--including text, photos, graphics, transcripts, film, video and audio. A simple graphical user interface enables easy storing, searching, retrieval, viewing, and online ordering of your assets. This graphical user interface runs on any Intel-based PC client using the Netscape browser and/or



#### ☑ Hyperwave Information Server 4.0

Hyperwave is proud to announce the availability of Information Server 4.0 - the most powerful and scalable enterprise Information Management platform available in the market today. By combining ease-of-use with an enhanced open architecture...

#### ▶ Hyperwave forges Knowledge Management Alliances

© Copyright 1997, 1998 Hyperwave GmbH. All Rights reserved.



HYPERWAVE AUTHOR

## OVERVIEW

HyperWave Author is integrated hypermedia authoring software, specially designed to help you create and edit documents on HyperWave servers.

Special navigation tools such as collection browsers and hyperlink maps make the creation of high-quality webs easier than ever before. Integrated support for the HyperWave Server's search engines helps the user to find existing documents and reuse them. HyperWave Author will prevent you from getting "lost in hyperspace".

Distributed data management is the keyword for working with HyperWave software within large companies. The combination of HyperWave's access control features and HyperWave Author lets each department keep track of its information without having to set up its own server. This leads to a sharp decrease in software and administration costs.

- [KEY FEATURES](#)
- [TECHNICAL SPECIFICATIONS](#)
- Online documentation for [HyperWave Author for Windows \("Amadeus"\)](#)
- Online documentation for [HyperWave Author for UNIX \("Harmony"\)](#)

User: **Guest** • Owner: **gmesaric** • Last modified: **04/25/1996 15:19:31**



HYPERWAVE AUTHOR

## KEY FEATURES

### **HYPERMEDIA AUTHORIZING**

HyperWave Author is the ultimate interactive authoring tool for HyperWave Servers. Users can author remotely over network boundaries: the Internet or any other TCP/IP based network can be used. HyperWave author provides full support for the HG-CSP network protocol, special HyperWave Server features such as database and search facilities are seamlessly integrated into the interface.

### **OBJECT DATABASE**

Object orientation is one of the key concepts of HyperWave. HyperWave Author provides full support for HyperWave Server's object-oriented database system, allowing easy insertion and editing of server-side objects. HyperWave Author for Windows additionally provides a local version of the database, letting users author web applications offline which they can later easily upload to a HyperWave Server.

### **VRML AND POSTSCRIPT**

HyperWave Author software includes IICM's free VRML scene viewer and a viewer for PostScript documents. VRML is the standard 3D data format in the WWW. PostScript is the industry standard for electronic publishing. HyperWave Author provides integrated PostScript viewer software, including the facility for inserting hyperlinks in PostScript documents: annotations to non-HTML documents are possible because of HyperWave's link database approach.

### **ADVANCED NAVIGATION**

Critics of the WWW often mention the so-called "lost in hyperspace" syndrome. HyperWave Author provides advanced navigation concepts and demonstrates that there are solutions to this problem: tree-view collection browsers let you navigate easily through big web servers and dynamically generated hyperlink maps help you keep masses of interlinked information up to date.

### **DISTRIBUTED INFORMATION MANAGEMENT**

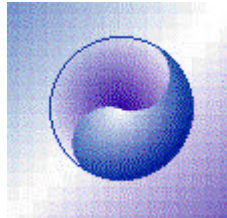
HyperWave's authoring software provides the facility of distributed information management: every logical part of a company can have a virtual web server without having the overhead of setting up its own real web server. A company can have one corporate identity on the web, running a WWW service where every department of the corporation is responsible for its own part.

### **MULTILINGUAL DOCUMENTS**

HyperWave Author supports easy creation and editing of multilingual web applications. HyperWave's support for multilingual document clusters is especially interesting if your company is located for example in Europe or Asia, or any other part of the world where more than one



# Welcome to Harmony: The Hyperwave Administrator for Unix/X11



Harmony is the Unix/X11 client for [Hyperwave](#), the first second-generation, publicly available, networked hypermedia information system running over the Internet. Hyperwave integrates hyperlinking, hierarchical structuring, sophisticated search, and access control facilities into one single system, and is interoperable with other network information tools like Gopher and WWW.

Here you can find information about

- [where to get Harmony by anonymous FTP](#)
- Harmony's special features
  - [information structuring facilities](#)
  - [orientational aids](#)
  - [multilinguality](#)
  - Harmony's [document viewers](#)
- Installation guide
  - [installation](#)
  - [user configuration](#)
- User support
  - the [Hyperwave mailing list](#)
  - the [Harmony FAQ](#)
  - further information

**Please direct any feedback** (comments, suggestions, bug reports, etc.) concerning Harmony by electronic mail to:

**[harmony@iicm.tu-graz.ac.at](mailto:harmony@iicm.tu-graz.ac.at)**

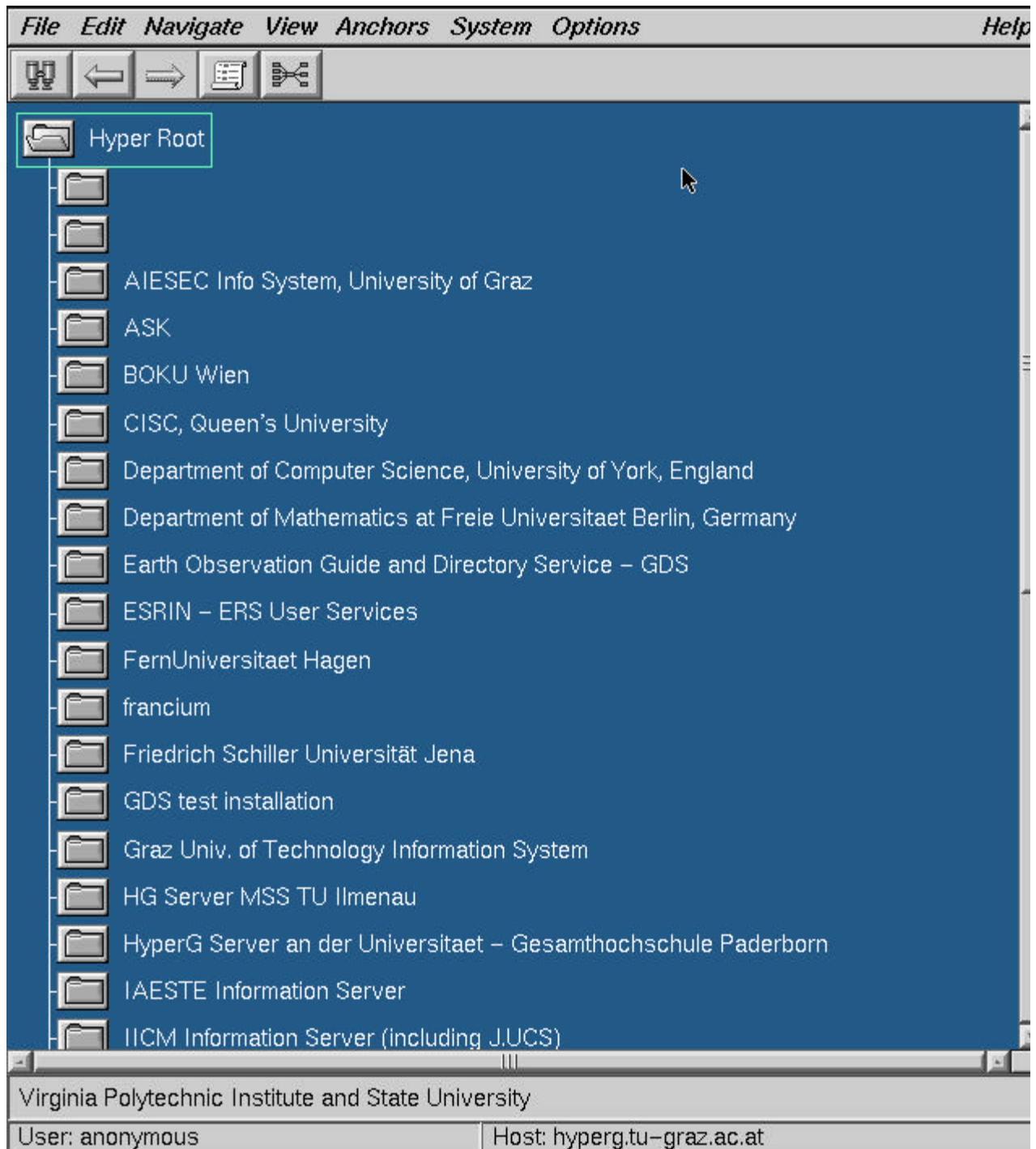
We are very interested in your feedback, even though we may not be able to respond personally to every piece of mail.

---

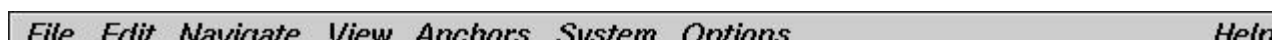
# Hyper-G --- Harmony Illustrations

Illustrations of the use of Hyper-G with the Harmony (UNIX) client include:

- connection to the global root



- expansion of the collection of nodes accessible from the root to those at the Virginia Tech server



File
Navigate
Anchors
View
Options
Help

Search
Anchors

## hginscoll (1)

### Name

hginscoll – insert a new collection

### Synopsis

```
hginscoll [-h] [-i FCollId | -n FCollName] [-N CollName] [-c] [-A Author]
[-C CDate] [-E EDate] [-O ODate] [-F][-T Title] [-R Rights] [-D
Description] [-S SortOrder] [-L Language] [-r hghost] [-d hgport]
```

### Description

hginscoll builds a collection or cluster object and insert it into the Hyper-G database.

---

hginscoll (1)

---

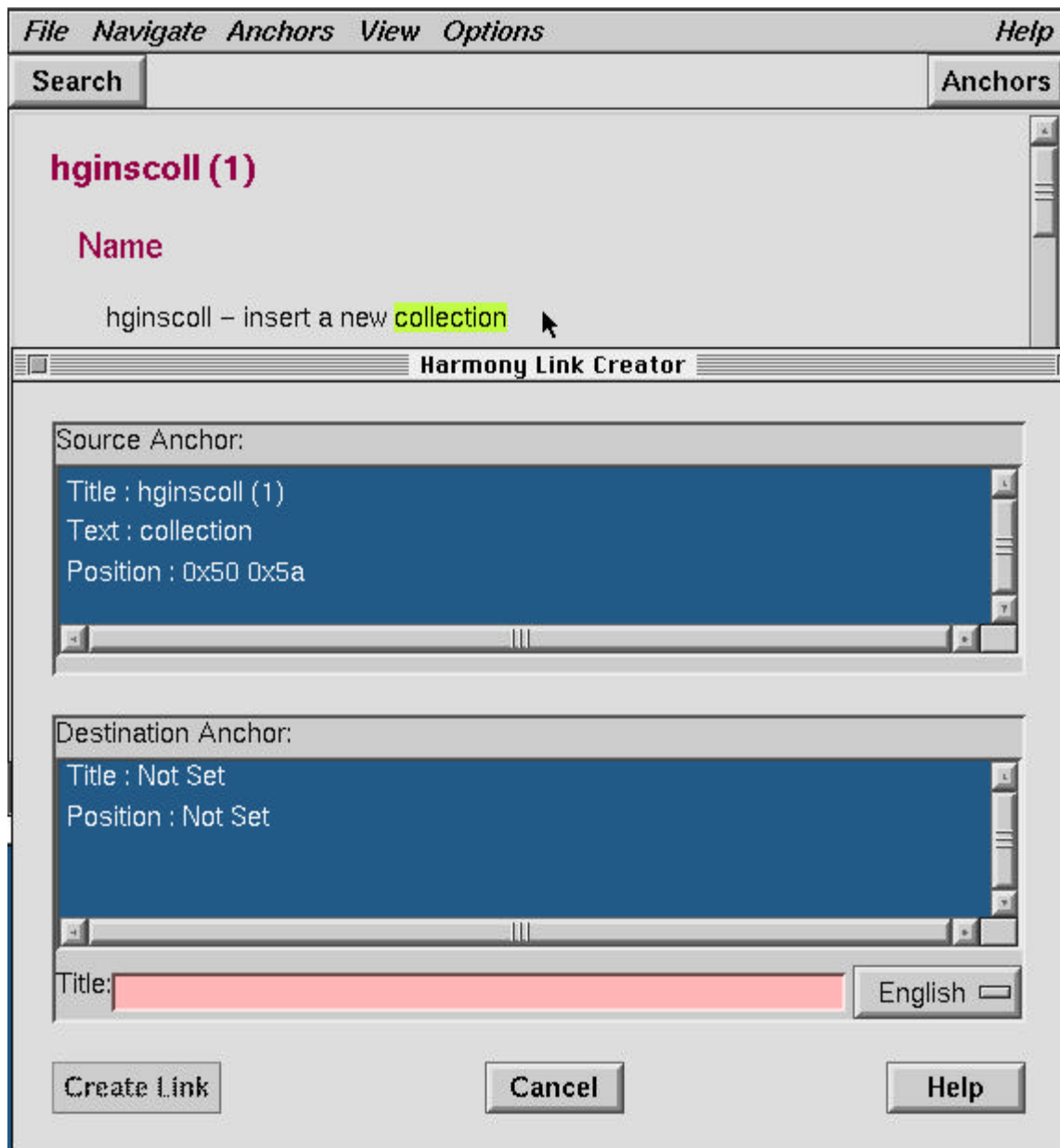
## Environment

HGAUTHOR:	Author
HGRIGHTS:	Rights
HGDESCRIPTION:	Description
HGSORTORDER:	SortOrder
HGFATHERCOLL:	FCollName
HGLANGUAGE:	Language

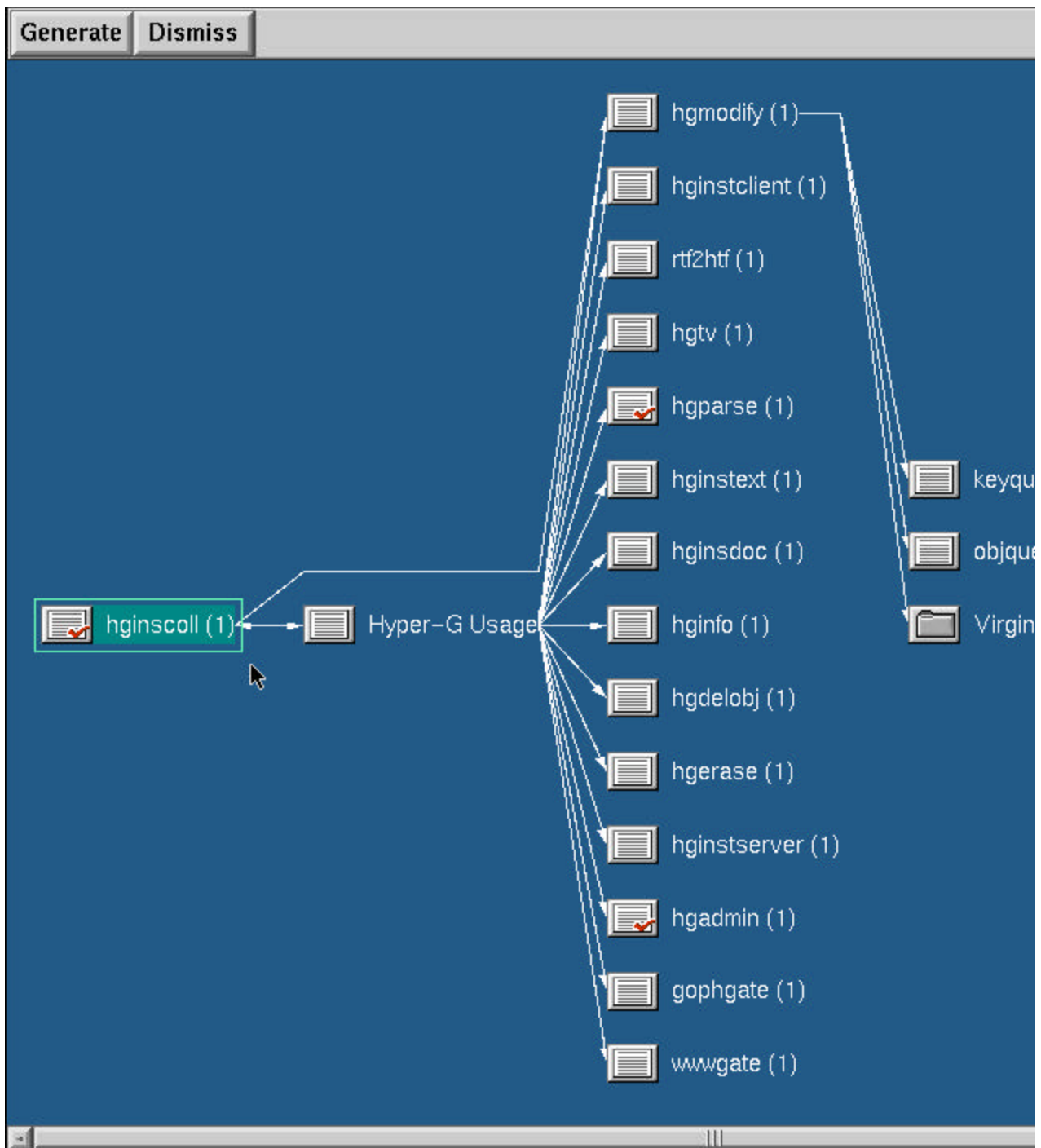
---

hginscoll (1)

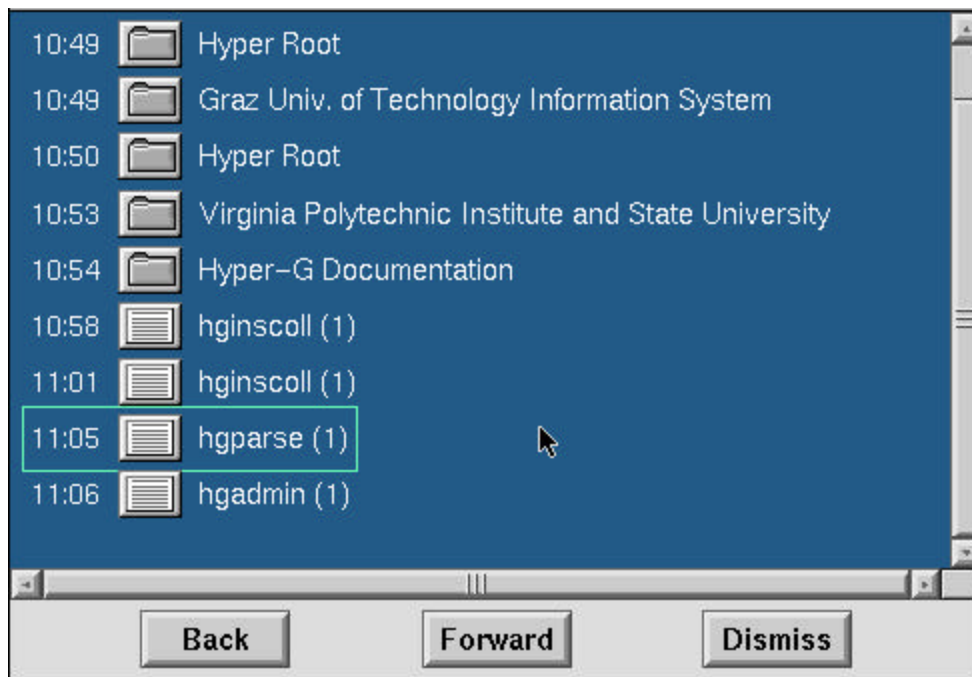
- making a link



- viewing a local map



- reviewing the history of accesses





# BLACKSBURG

## electronic village



Community  
Arts  
Organizations  
Religion  
Sports

Education  
Library  
Museums  
Schools

People  
Discussion  
Seniors

Government  
Health  
Village Mall  
Visitor's Center

Search Tools

Help Desk

▼ Full Index

### News

#### [BEV User Survey](#)

Please complete this local Internet survey to provide feedback to help shape the future of the BEV.

#### **Live in an apartment with Ethernet?**

The BEV/BA apartment Ethernet system will be transferred to the apartment owners beginning July 1st. The transition should be minor for most users, but [read the FAQ](#) to see if you will be affected.

#### **Whitewater Kayaking Classes**

Wednesday evenings we offer roll classes in the county pool, Sundays we offer river classes on the New River. Classes are instructed by Back Country Ski & Sport's ACA certified instructors, with a low student-instructor ratio. Roll class or previous experience required before you register for river class. \$15/roll class + \$5/boat rental. \$50.00/river class + \$15/boat rental. Pre-registration required - classes fill quickly! Call 382-6980 if you need more information.

#### [SEEDS Summer Field Camps](#)

Seek Education, Explore, DiDiscover - SEEDS has a few openings left for the 1998 full-day Summer field camps. Coed sessions for ages 7-9 and 10-12. Go exploring with SEEDS this Summer!

#### [Sign the Bicentennial Guestbook!](#)

Here's your chance to be a part of history.

#### **Summer Playcamp Registration**

We are a licensed day care facility offering care with a summer camp atmosphere. Arts & crafts, nature, games, pool time, free lunch and field trips are just a few of the exciting activities we offer from 7:30 - 5:30, Mon-Fri. June 15 - August 7 for \$55.00 a week. Pre-registration required, deadline is prior Wednesday.

[\[Old Messages || POST a Message\]](#)



[Weather Underground Forecast](#)  
[Blacksburg National Weather Service](#)

---

[About the BEV](#) | [Services](#) | [Training](#) | [Research](#) | [Starting a Village](#)

---



---

Welcome to the **BEV HistoryBase**, a WWW History Page for the [Blacksburg Electronic Village](http://history.bev.net/bevhist/)! Try out the [BEV History Timeline](#) to learn more about the history of our electronic community. For a non-graphical alternative, check out the [Textual BEV History Timeline](#). Both contain the same information so feel free to browse either.



[ [Main Timeline](#) | [Contribute](#) | [What's New?](#) | [Search](#) ]

[Message of the Day Listings](#)

[Blacksburg Telecommunications Advisory Committee Meeting Minutes](#)

[BEV Media Coverage Archive](#)

[BEV Group Home Pages](#)

This project is supported by NSF Grant CDA-9424506. A [copy of the grant proposal is online](#).

---

*Last updated 27 October 1995 / [schmidt@cs.vt.edu](mailto:schmidt@cs.vt.edu)*

[HistoryBase Main Page](#)

[Contribute](#)

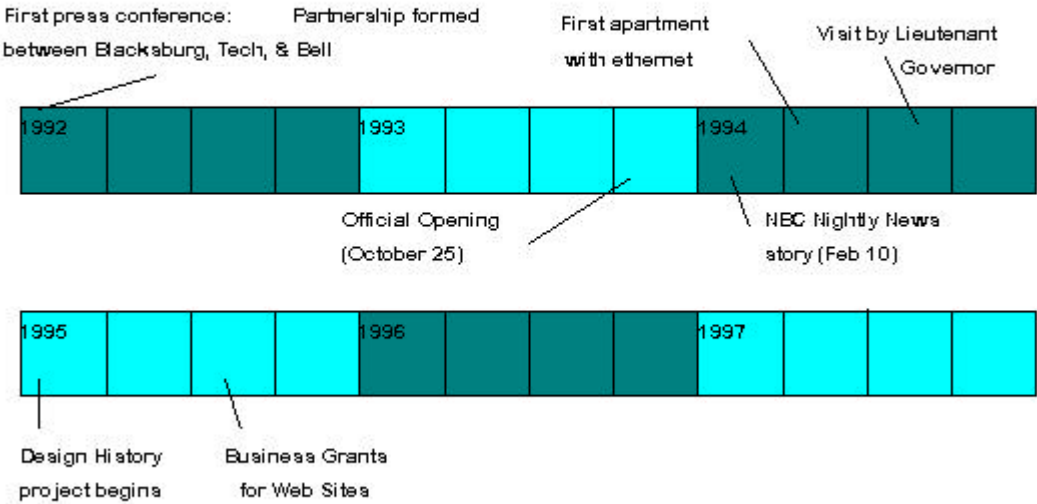
[What's New?](#)

[Search](#)

[BEV Homepage](#)

# BEV HistoryBase: Main Timeline

Click in a box to see a more detailed history for that quarter



Click in a box to see a more detailed history for that quarter

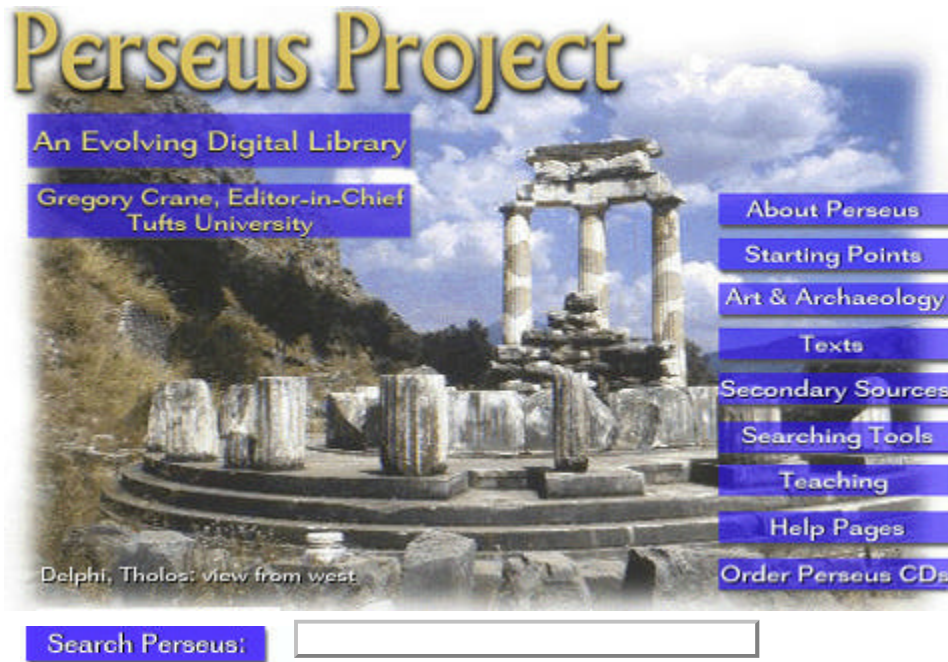
[HistoryBase Main Page](#)

[Contribute](#)

[What's New?](#)

[Search](#)

[BEV Homepage](#)



# Perseus Project

An Evolving Digital Library

Gregory Crane, Editor-in-Chief  
Tufts University

- About Perseus
- Starting Points
- Art & Archaeology
- Texts
- Secondary Sources
- Searching Tools
- Teaching
- Help Pages
- Order Perseus CDs

Delphi, Tholos: view from west

Search Perseus:

[Copyright](#) | [General Policies of this Site](#) | [FAQ](#)

---

Read the important announcement on [planned changes for access](#) to the Perseus web site.



- New [interactive web atlas](#).
- Revised Catalogue of [Greek sculpture](#), including images from the [Museum of Fine Arts, Boston](#)
- Newly accessible [ancient art resources](#): over 670 vase pictures, and the landmark Caskey and Beazley vase catalog, from the Museum of Fine Arts, Boston
- Announcing our new site on [Julius Caesar](#)

Congratulations to the city of Athens,  
site of the 2004 Summer Olympic Games!

Many are the sights to be seen in Greece, and many are the  
wonders to be heard; but on nothing does Heaven bestow more  
care than on the Eleusinian rites and the Olympic games.

Pausanias, [Description of Greece 5.10.1](#)



[Awards and reviews](#) | [Text-only home page](#) | [Related sites](#)

---

The Perseus Project is supported by the [Annenberg/CPB Project](#), the [National Science Foundation](#), [Apple Computer](#) and the [National Endowment for the Arts](#), the [National Endowment for the Humanities](#), the Packard Humanities Institute, the [Getty Grant program](#), [Xerox Corporation](#), [Boston University](#), [Harvard University](#), and the [Fund to Improve Post-Secondary Education](#).

Perseus is a non-profit enterprise, located in the [Classics Department](#), [Tufts University](#).

---

*Mail problems and suggestions to:*  
[webmaster@perseus.tufts.edu](mailto:webmaster@perseus.tufts.edu)

## Search the NCSTRL Collection



Search *ALL* bibliographic fields ...

Search for:

Sort results by: rank



Search *SPECIFIC* bibliographic fields ...

*Author:*

*Title:*

*Abstract:*

(Combine fields with ☒ **AND** ☐ **OR**)

Sort results by: rank



Search

Clear fields

If you would like to view the NCSTRL collection by *year* or by *institution*, use the [browse form](#).

# Digital Libraries for CS

Here are some pointers to Digital Libraries / bibliography servers related to CS.

[ACM Digital Library Collection at Virginia Tech](#)

*Small test collection of CACM articles from those scanned in as part of the NSF-supported Envision project.*

[ACM Graphics Bib. DB](#)

*SIGGRAPH Online Bibliography Database*

[ACM Computer Graphics Courseware Repository](#)

*SIGGRAPH Computer Graphics Courseware Repository (ftp)*

[ACM HCI Bib. DB](#)

*interactions Bibliographies on Human-Computer Interaction*

[BibNet Project](#) and [TeX Users Group](#) FTP bibliographies

*bibliography collections from Nelson Beebe including HTML with extensive internal and external hypertext links. See examples: [IBM Systems Journal](#), [DEC Technical Journal](#). See [program to build these from BibTeX](#).*

[CACM Collection \(1959-1979\) using Inquiry](#)

*U. Mass. CIIR demo of Inquiry with CACM test collection*

[Collection of Computer Science Bibliographies](#)

*from Alf-Christian Achilles; updated monthly; 790 locally stored bibliographies; more than 530,000 references; 20,000 references contain URLs to an online version of the paper; more than 1600 links to other sites carrying bibliographic information; uses Glimpse*

[Databases and Logic Programming \(mirror\)](#)

*bibliography server by Michael Ley*

[Hypertext Bibliography Project](#)

*Hypertext Bibliography Project (Glimpse search of many publications)*

[NCSTRL](#)

*Networked Computer Science Technical Report Library*

[Table of Contents re LIS](#)

*Table of Contents for JASIS, IPM, etc. - may be slow*

[Univ. of Wales Cardiff CS Courseware](#)

*Courseware on Algorithms, AI, C, Graphics, Image Processing, Parallel Processing, Vision, X*

# ETD Electronic Thesis and Dissertation Initiative

---

## Welcome to the Virginia Tech Electronic Thesis and Dissertation home page!

- Browse the [ETD Library](#). Hundreds of titles!
  - Learn about our parent project, the [Networked Digital Library of Theses and Dissertations](#)
  - Learn what the Graduate School expects when you [submit your ETD](#) **HOT!**
  - Attend an [ETD Workshop](#) **NEW!**
  - Learn [how to create an ETD](#) step by step **HOT!**
  - Please complete a [survey](#)
  - What is [PDF](#) anyway? **NEW!**
  - Learn about [LaTeX](#) and [ETD-ML](#) submissions
  - Learn about [publishers and copyright](#)
  - Need help? We have installed Adobe PDF software in [computer laboratories](#) all over campus
  - Still puzzled? Try our list of [frequently asked questions!](#) **HOT!**
  - All else has failed? Contact us: [etd@vt.edu](mailto:etd@vt.edu)
- 

[Campus Labs](#) | [ETD Library](#) | [ETD-ML](#) | [FAQ](#) | [How-to](#) | [LaTeX](#) | [NDLTD](#) | [PDF](#) | [Submission Guidelines](#)

---

*etd*

*Revised: Wed Feb 18 11:12:14 EST 1998*

[index.sl](#)

*scholarly communications project*

# Virginia Tech <ETD> Submission Form

## Instructions:

Please fill out the form completely. Cut and paste, from your document and into the form, as necessary. Read the [help file](#) for help on cutting and pasting your abstract and for selecting keywords. Once you are done filling out the form read the [copyright statement](#) at the bottom of the page and if you agree to it click "Preview".

 **WARNING: There have been problems reported when using these forms with Netscape version 3.02. If you see an error message of the type "Document Contains No Data", please make sure that you are not using Netscape Navigator 3.02.** 

## Document Type:

Please select the type of document you are submitting.

Master's Thesis 

## Name:

Please enter your name just as it appears on the title page.

**First and Middle name (if desired):**

**Last Name:**

**Suffix (Jr., Sr., III, etc.):**

## Email:

Please enter your email address.

## Title:

Please enter the title just as it appears on the title page.

# Virginia Tech Graduate School

## Electronic Thesis and Dissertation (ETD) Submission

### Approval Form

Student Name: \_\_\_\_\_

ID#: \_\_\_\_\_

Department: \_\_\_\_\_

Degree: ☐ Bachelor's ☐ Master's ☐ Doctoral degree

Document Type: ☐ Project Report ☐ Thesis ☐ Dissertation

Document Title: \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

### Student Agreement:

I hereby certify that, if appropriate, I have obtained and attached hereto a written permission statement from the owner(s) of each third party copyrighted matter to be included in my thesis, dissertation, or project report, allowing distribution as specified below. I certify that the version I submitted is the same as that approved by my advisory committee.

I hereby grant to Virginia Tech and its agents the non-exclusive license to archive and make accessible, under the conditions specified below, my thesis, dissertation, or project report in whole or in part in all forms of media, now or hereafter known. I retain all other ownership rights to the copyright of the thesis, dissertation, or project report. I also retain the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report.

### Student and Committee Agreement:

**Part A.** We agree that the above mentioned document be placed in the ETD archive with the following status: (*choose one of 1, 2, 3, or 4*)

# ETD Multimedia File Formats

---

## General Comments

Including complex multimedia objects in an ETD is a relatively new possibility. Those attempting this are pioneers. You are encouraged to work with those on your committee interested in this to gain their approval and assistance. Ultimately they should check your final submission, and should be prepared and agree to do so with the multimedia part, else you may think about putting your multimedia work into some other document (e.g., report, WWW site).

There are locations on campus to help with multimedia work. One is the New Media Center in Newman Library, which supports the campus and local community. Hancock Hall houses a multimedia laboratory for Engineering and Architecture. The Center for Digital Music in Squires supports work with audio. The Information Access Laboratory in McBryde 110, supports scanning, digital audio, and digital video. Experts in digital library technology are available in the Digital Library Research Laboratory.

- [New Media Center](#)
- [Center for Digital Music](#)
- [Information Access Laboratory](#)
- [Digital Library Research Laboratory](#)

It is likely that complex multimedia objects will each reside in a different file, located in the same directory as the rest of your ETD. You may wish some icon or thumbnail or other small form of the complex multimedia object in the body of your ETD, and to have that linked to the complex multimedia object.

---

## Archiving

Be careful to consider issues of long-term archiving.

- Always include the highest resolution version of your object, not just a version suitable for today's devices, since technology may improve. You can include several versions, to help those with a variety of devices, particularly if the media itself is not scalable. For example, scan a slide at 2700dpi, but have 640x480 and 320x240 versions as well.
  - If you can, include a version using a well-accepted international standard. Thus, for video, MPEG is encouraged. If you start with QuickTime, include that, but also include MPEG if possible.
  - If you use some proprietary software, include a viewer if that is allowed by the vendor. That way people can view your object without buying that software. Realize, however, that in a few years this object may not be readily usable due to changes in versions and technology.
-

---

# ETD Digital Library

# Networked Digital Library of Theses and Dissertations

---

## Digital Library of ETDs

### Official Nodes in the NDLTD

- [North Carolina State University](#)
- [University of Virginia](#)
- [Virginia Tech](#)
- [West Virginia University](#)
- [University of Waterloo](#)

### Other Sites with ETDs

- [University of Michigan](#)
  - [Independent ETDs](#)
- 

## Federated Search (Demo) for NDLTD

Please try out the following demonstration of how federated search of NDLTD may occur. Report suggestions to James Powell at [jpowell@vt.edu](mailto:jpowell@vt.edu)

- [Federated Search Demonstration](#)
- 

## Collection Highlights - Notable ETDs

- [Notable ETDs](#)
- 

[NDTLD](#)

---

*etd*

*Revised: Mon Jun 15 16:50:00 1998*

[index.sl](#)

# NDLTD Networked Digital Library of Theses and Dissertations

---

Universities, students, publishers, other interested parties, Welcome!

- Researchers, see <http://www.theses.org/> to **search** and **browse** our library of electronic theses and dissertations (ETDs).
- Students, see <http://etd.vt.edu/> for help creating and submitting ETDs.

## What We Are

- An [initiative](#) to improve graduate education, increase sharing of knowledge, help universities build their information infrastructure, and extend the value of digital libraries
- A federation of [member universities](#)
- A project [supported by FIPSE and SURA](#)
- A [project team](#) based at [Virginia Tech](#)
- A recent topic in the [news](#)
- Led by [steering committee](#) and a [technical committee](#)

## What We Do at Virginia Tech

- Require students to develop and submit Electronic Thesis or Dissertations (ETDs)
- Provide a [web site](#) to help students
- Support a [digital library](#) of ETDs
- Develop a [workflow model](#) for submitting ETDs
- Give [talks](#)
- Write [papers](#)

## How YOU Can Participate

- Come to [organizational meetings](#)
- [Join us](#) and develop your own NDLTD member site with our help!
- Contribute to our [e-mail list\(s\)](#)

## Our Objectives

- **To improve graduate education** by allowing students to produce electronic documents, use digital libraries, and understand issues in publishing
- **To increase the availability of student research** for scholars and to preserve it electronically
- **To lower the cost** of submitting and handling theses and dissertations
- **To empower students** to convey a richer message through the use of multimedia and hypermedia technologies
- **To empower universities** to unlock their information resources
- **To advance digital library technology**

## Further Information

- Statistics on [usage](#) of Virginia Tech collection
- General and historical [information](#)
- Information for [publishers](#)
- Issues in [copyright](#)
- Doctoral students can win an [Innovation Grant](#)
- Links to [related projects](#)
- Links to [related \(meta-\)initiatives](#)

Questions? Comments? [etd@ndltd.org](mailto:etd@ndltd.org)

---

# NDLTD History, Description, and Scope

---

## Early History

The concept of electronic theses and dissertations (ETDs) was first openly discussed at a 1987 meeting in Ann Arbor arranged by UMI, and attended by representatives of Virginia Tech (Ed Fox from Computer Science and Susan Bright from the Computing Center), University of Michigan, SoftQuad, and ArborText. As followup, Virginia Tech funded development of the first SGML Document Type Definition (DTD) for this purpose, by Yuri Rubinski of SoftQuad.

Virginia Tech's Dean Gary Hooper agreed to finance further development in 1991. Ed Fox and John Eaton (Dean of the Graduate School) have collaborated on this project since that time, investigating problems associated with production, archiving and access, initially with a local faculty committee. Since 1992 they have worked with the Coalition for Networked Information (CNI), the Council of Graduate Schools (CGS), UMI and other interested organizations, helping run a series of design and discussion meetings. Additionally, the University Library's Scholarly Communications Project developed the procedures and systems for processing, archiving, and providing public access to Virginia Tech's graduate research works.

---

## SURA Support

In 1993, at the inception of the Monticello Electronic Library Project, supported by SURA and SOLINET, Professor Edward Fox of Virginia Tech became Co-Chair of its Working Group on Theses, Technical Reports and Dissertations. In 1994 SURA funded a workshop at Virginia Tech to develop plans for electronic theses and dissertations (ETDs), selecting Adobe's Portable Document Format (PDF) and the Standard Generalized Markup Language (SGML) for representation and archiving. To help implement these plans, SURA has funded a [research, development, and dissemination effort](#) based at Virginia Tech for 1996.

---

## Goals

The main goals of the ETD initiative are:

- for graduate students to learn about electronic publishing and digital libraries, applying that knowledge as they engage in their research and build and submit their own ETD,
- for universities to learn about digital libraries, as they collect, catalog, archive, and make ETDs accessible to scholars worldwide,
- for universities in the Southeast and beyond to learn how to unlock the potential of their intellectual property and productions,

## MAGAZINE

# National Digital Library of Theses and Dissertations

## A Scalable and Sustainable Approach to Unlock University Resources

Edward A. Fox, John L. Eaton, Gail McMillan  
Neill A. Kipp, Laura Weiss, Emilio Arce, and Scott Guyer  
Virginia Tech  
Blacksburg, Virginia  
<http://etd.vt.edu/etd/>  
[etd@vt.edu](mailto:etd@vt.edu)

Project Director: Edward A. Fox

**D-Lib Magazine**, September 1996

ISSN 1082-9873

---

### Table of Contents

1. [Introduction](#)
2. [Expected Benefits](#)
3. [How You Can Help](#)
4. [Since 1987](#)
5. [Pilot Efforts at Virginia Tech](#)
6. [Related Work](#)
7. [Sustainability](#)
8. [Scalability](#)
9. [Plans](#)
10. [Acknowledgments](#)

---

### 1. Introduction

As of September 1, 1996, the U.S. Department of Education provided grant support for a three-year, Virginia Tech-led project to *Improve Graduate Education with a National Digital Library of Theses and Dissertations* (NDLTD), adding to 1996 funding from the Southeastern Universities Research Association (SURA) for *Development and Beta Testing of the Monticello Electronic Library Thesis and Dissertation Program*. True success in these projects will potentially mean a permanent change in graduate education and scholarly publishing, with digital libraries playing a more dominant role in supporting and disseminating research.

M A G A Z I N E

# Networked Digital Library of Theses and Dissertations

## An International Effort Unlocking University Resources

Edward A. Fox, John L. Eaton, Gail McMillan  
Neill A. Kipp, Paul Mather, Tim McGonigle, William Schweiker, and Brian DeVane  
Virginia Tech  
Blacksburg, Virginia 24061-0106  
<http://www.ndltd.org>  
[etd@ndltd.org](mailto:etd@ndltd.org)

Project Director: Edward A. Fox

**D-Lib Magazine**, September 1997

ISSN 1082-9873

---

## Table of Contents

- [1. Introduction](#)
- [2. Progress](#)
  - [2.1. Collaboration](#)
  - [2.2. Infrastructure](#)
  - [2.3. Tools](#)
  - [2.4. Collection](#)
- [3. Controversy](#)
  - [3.1. Social Issues](#)
  - [3.2. Time, Effort, Impact, Reward, and Quality](#)
  - [3.3. Loyalties](#)
  - [3.4. Economics](#)
- [4. Future Growth](#)
  - [4.1. Digital Libraries and Education](#)
  - [4.2. Universities Working Together](#)
  - [4.3. Evolution](#)
- [5. References](#)
- [6. Acknowledgments](#)

---

## 1. Introduction

On the first anniversary of funding by the U.S. Department of Education (FIPSE) for a National Digital

## People:

---

[Dan Atkins](#) University of Michigan Digital Library Project Director.

[Edward A. Fox](#) Director of the [Digital Libraries Research Group](#) at Virginia Tech.

[Hector Garcia-Molina](#)

- [Papers](#)

[Henry Gladney](#) [Peter Graham](#)

[Michael Lesk](#)

- [Images: Quantity is not always Quality - U. KY talk](#)
- [digital libraries](#)
- [library preservation](#)
- [information retrieval](#)
- [networking, etc.](#)
- [Projections for Making Money on the Web](#)

[Gary Marchionini](#)

- [U. Md. DL Home Page](#)
- [Encyclopedia article draft](#)
- [CACM April 1995 article](#) in [that year's volume online in ACM DL](#)

[Michael Mauldin](#) (Lycos, CMU)

[Bruce Schatz](#) University of Illinois at Urbana-Champaign, DLI Principal Investigator

[Marvin Sirbu](#)

- [publications available online](#)

[Terry Smith](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

# Digital Library Research Laboratory

840 University Boulevard, Suite 8  
Mail Stop 0368  
Virginia Tech  
Blacksburg VA 24061

---

## Table of Contents

- [1](#) Projects
  - [2](#) Prototypes
  - [3](#) Research Staff
  - [4](#) Books, Papers, Publications, Presentations
- 

### 1. Projects

- [NDLTD](#) Networked Digital Library of Theses and Dissertations
  - [ETD](#) Electronic Thesis and Dissertation Initiative, Virginia Tech
  - [4S](#) Sets, Streams, Structures, and Scenarios (4S): Towards a Formal Model of Digital Libraries
- 

### 2. Prototypes

- [SAUCER: Virginia Tech Speculative Fiction \(prototype\)](#)
- 

### 3. Research Staff

- [Edward A. Fox, Director](#)
  - [Ghaleb Abdulla](#)
  - [Robert France](#)
  - [Tom Johnson](#)
  - [Neill A. Kipp](#)
  - [Binzhang Liu](#)
  - Paul Mather
- 

### 4. Books, Papers, Publications, Presentations

- [Fox, et al](#)

## Henry Gladney:

---

- Access Control Articles in D-Lib Magazine:  
Gladney et al., Safeguarding Digital Library Contents and Users:
  - [Assuring Convenient Security and Data Quality](#),
  - [Document Access Control](#)
  - [Digital Images of Treasured Antiquities](#)
  - [A Note on Universal Unique Identifiers](#)
  - [Storing, Sending, Showing, and Honoring Usage Terms and Conditions](#)
- [Gladney et al. report on DL requirements and architecture \(PostScript\)](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[People\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## STORIES

**D-Lib Magazine**  
**May 1998**

ISSN 1082-9873

# Safeguarding Digital Library Contents and Users

## Storing, Sending, Showing, and Honoring Usage Terms and Conditions

---

Henry M. Gladney and Jeff B. Lotspiech  
IBM Almaden Research Center  
San Jose, California 95120-6099  
[\(gladney, lotspeich\)@almaden.ibm.com](mailto:(gladney, lotspeich)@almaden.ibm.com)

---

### Abstract

This article knits together ideas and technologies discussed in several prior articles in the *Safeguarding ...* series in **D-Lib Magazine**.

We discuss languages for representing intellectual property usage terms and conditions in databases, for network transmission, and for presentation to and editing by human beings. Prototypes, one in each domain, can be knit together as a component of digital library services. We favor an approach based on cryptographic envelopment of document packets because this provides end-to-end protection and requires less network infrastructure and administration than alternatives. It needs protocols for enforcing information owners' rules -- protocols which govern how a user might select, request, possibly pay for, and eventually gain access to what she wants.

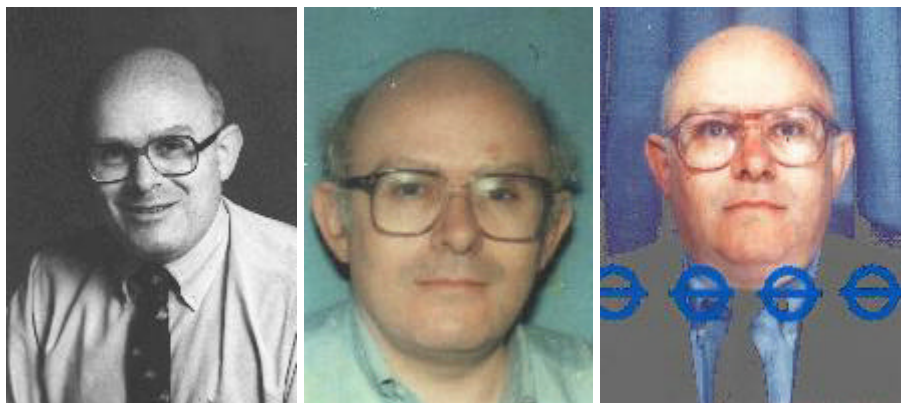
A deployed permission management and revenue collection mechanism will implement at least three system roles: a source **S** which encrypts and bundles valuable objects, an end user system **U** which manages requesting and receiving protected information, and a clearance center **C** which checks users' commitments to observe owners' conditions. We consider three alternative configurations.

We remind the readers why incomplete schemes based on much-ballyhooed "trusted systems" are fundamentally flawed, and suggest why it is unlikely that these notions will evolve to practical personal computer services. People might be less likely to be misled if this elusive objective were called "**trustworthy**

# Michael Lesk's Grade Crossing on the Information Superhighway

Please change any address/link to this page to <http://www.purl.net/NET/lesk>. The address 'purl.net' refers to 'permanent URL' and this address should survive local administrative changes. Thank you.

This page is also available from a [site in the United Kingdom](#).



Professional

Amateur

Coin-operated

---

Now out: my new book *Practical Digital Libraries: Books, Bytes and Bucks*, [Morgan Kaufmann](#), July 1997.

---

**Position:** Division Director, Information and Intelligent Systems, National Science Foundation, <http://www.cise.nsf.gov/iis>.

**Also:** Visiting Professor, University College London, Department of Computer Science.

## Biography

In the 1960's I worked for the SMART project, wrote much of their retrieval code and did many of the retrieval experiments, as well as obtaining a PhD in Chemical Physics. In the 1970's I worked in the group that built Unix and I wrote Unix tools for word processing (*tbl*, *refer*), compiling (*lex*), and networking (*uucp*). In the 1980's I worked on specific information systems applications, mostly with geography (a system for driving directions) and dictionaries (a system for disambiguating words in context), as well as running a research group at Bellcore. And in the 1990s I have worked on a large chemical information system, the CORE project, with Cornell, OCLC, ACS and CAS.

I am also Visiting Professor in computer science at University College London; I'm on the Visiting Committee for the Harvard University Library; and I've worked with the Commission on Preservation and Access addressing digital preservation issues. I received the "Flame" award for lifetime achievement from Usenix in 1994, and I am a Fellow of the ACM. You can read my [publication list](#) if you wish. The previous paragraph is available in [Japanese](#).

## Where?

Michael Lesk  
[National Science Foundation](#)  
4201 Wilson Boulevard, Room 1115  
Arlington, Virginia 22230  
703 306-1930 [Voice]  
703 306-0599 [Fax]  
[lesk@acm.org](mailto:lesk@acm.org)

---

## Interests

### [Digital Libraries](#)



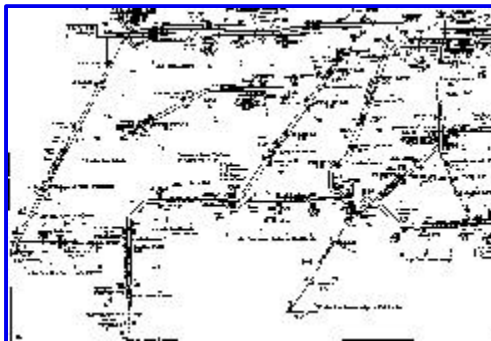
### [Library preservation](#)



### [Information Retrieval](#)



### [Networks & Misc.](#)

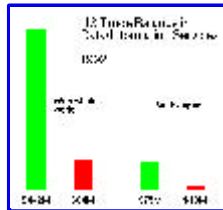


---

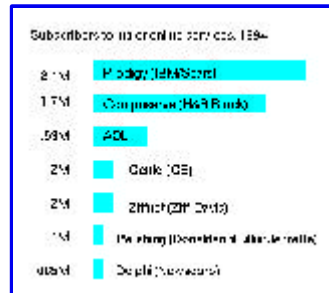
## Places I have lived

## Factoids

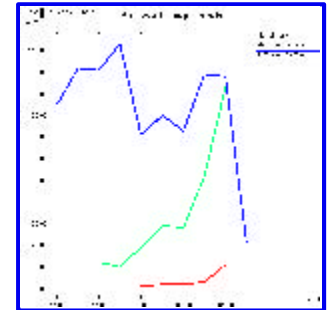
United States balance of trade in information services



Number of customers of online services



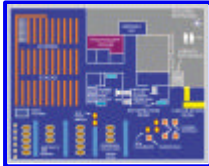
Trends in buzzwords



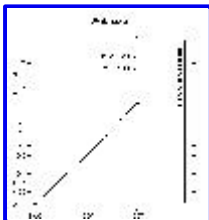
## Material on digital libraries



[\*US Digital Library Programs: What Goals?\*](#)



[\*The Organization of Digital Libraries\*](#)



[\*How Much Information Is There in the World?\*](#)



[\*Digital Libraries: A Unifying or Distributing Force?\*](#), to be presented at *Scholarly Communication and Technology*, a conference sponsored by the Andrew W. Mellon Foundation, Atlanta, Georgia (April 24, 1997).



[\*Mad Library Disease: Holes in the Stacks\*](#), Lazerow Lecture, given at University of California Los Angeles, 18 April 1996, to appear in print later in 1996



[\*Libraries and the Web\*](#), to appear *Libraries and Information World Wide*, 1996



[\*Economics of Digital Libraries\*](#), course outline for lectures given Jan-Apr 1996,



Columbia University, New York.



[\*Why Digital Libraries\*](#), Follett lecture on electronic libraries, given 19 June 1995, BBC Conference Center, London, England.



[\*The Future Value of Digital Information and Digital Libraries\*](#); lecture given 9 November 1995 at the Kanazawa Institute of Technology Roundtable on Libraries and Information Systems, Kanazawa, Japan.



[\*Making a Digital Library: The Contents of the CORE Project\*](#); draft paper, October 1994; to appear, ACM TOIS

# **The roles of digital libraries in teaching and learning**

**Gary Marchionini (University of Maryland, College Park, MD/USA,  
email: [march@umdd.umd.edu](mailto:march@umdd.umd.edu))**

**Hermann Maurer (Graz University of Technology, Graz/Austria,  
email: [hmaurer@icm.tu-graz.ac.at](mailto:hmaurer@icm.tu-graz.ac.at))**

CACM April 95-Volume 38, Number 4 pg 67-75

## **Introduction**

Libraries have long served crucial roles in learning. The first great library, in Alexandria two thousand years ago was really the first university. It consisted of a zoo and various cultural artifacts in addition to much of the ancient world's written knowledge and attracted scholars from around the Mediterranean who lived and worked in a scholarly community for years at a time. Today, the rhetoric associated with the National/Global Information Infrastructure (N/GII) always includes examples of how the vast quantities of information that global networks provide (i.e., digital libraries) will be used in educational settings [16].

This paper describes how digital libraries are evolving to meet the needs of teaching and learning and identifies issues for continued development. We distinguish formal, informal, and professional learning and argue that digital libraries will allow teachers and students to use information resources and tools that have traditionally been physically and conceptually inaccessible. We illustrate the types of information resources that digital libraries offer to teachers and learners and discuss some of the issues and challenges that digital libraries present for teaching and learning.

## **How do libraries support teaching and learning?**

A library is fundamentally an organized set of resources, which include human services as well as the entire spectrum of media (e.g., text, video, hypermedia). Libraries have physical components such as space, equipment, and storage media; intellectual components such as collection policies that determine what materials will be included and organizational schemes that determine how the collection is accessed; and people who manage the physical and intellectual components and interact with users to solve information problems.

Libraries serve at least three roles in learning. First, they serve a practical role in sharing expensive resources. Physical resources such as books and periodicals, films and videos, software and electronic databases, and specialized tools such as projectors, graphics equipment and cameras are shared by a community of users. Human resources--librarians (also called media specialists or information specialists) support instructional programs by responding to the requests of teachers and students (responsive service) and by initiating activities for teachers and students (proactive services). Responsive services include maintaining reserve materials, answering reference questions, providing bibliographic instruction, developing media packages, recommending books or films, and teaching users how to use materials. Proactive services include selective dissemination of information to faculty and students, initiating thematic events, collaborating with instructors to plan instruction, and introducing new instructional methods and tools. In these ways, libraries serve to allow instructors and students to share expensive materials and expertise.

## Countries & Regions:

---

(Chapter 11, page 245, "Books, Bytes and Bucks", Michael Lesk)

- **United States of America:** In the US, NSF, NASA and ARPA have funded six important Digital Library efforts, called the DLI (Digital Libraries Initiative). These programs each involve a large consortium of cooperating institutions but the six main ones are : University of California at Berkeley, University of Santa Barbara, University of Michigan, Carnegie Mellon University, Stanford University, and the University of Illinois.
  - University of California at Berkeley: Image content queries along with Xerox PARC, database extraction from documents, multivalent documents, NLP. Headed by Robert Wilensky.
  - University of Michigan: Scalability and Education. They are also investigating the use of agent architectures for Digital Libraries and trying to merge DLI with their other digital library efforts such as JSTOR and TULIP. Headed by Dan Atkins.
  - University of Illinois: Concentrating on using scientific journals as their base collection with diversity in both documents as well as publishers, making the transition process from SGML to HTML smoother, defining semantic spaces. Headed by Bruce Schatz.
  - Stanford University: concentration is on the infrastructure development such as base networking and databases to support digital libraries. Also concerned with interoperability between different digital library projects. Headed by Hector Garcia-Molina.
  - University of California at Santa Barbara: spatial indexing and retrieval , image processing. Headed by Terry Smith.
  - Carnegie Mellon University: digital video, image analysis, speech recognition, face recognition, natural language understanding. Headed by Michael Mauldin and Marvin Sirbu.

Other than DLI, many research projects are underway at some other universities such as Virginia Tech and Texas A&M. In the near future, extensive funds are expected to be allocated for Digital Libraries.

The Library of Congress, under James Billington is digitizing 5 million of its items in a massive \$60 million effort. Other universities involved in related projects are Georgia Tech, Cornell, MIT, University of Tennessee, Washington and California and Virginia Tech (known for the Envision system of Ed Fox). Other limited efforts include University of Virginia, University of Georgia and Columbia University.

- **United Kingdom:** Though efforts are still limited to penny-pockets, 20 million pounds have been set aside for digital library projects. The program originally called FIGIT, now known as E-LIB funded 35 projects. Work includes cataloguing of archives, digitization of documents and data sharing. Some of the more notable efforts are : Digitizing the Burney collection of pre-1800 newspapers and scanning of Batley News, the CANTERSBURY TALES project that involves scanning all pre-1500 manuscripts and some other similar projects. However, the most notable is the Electronic Beowulf project which is a US/UK collaboration between Kevin Kiernan (University of Kentucky), Paul Szarmach (Western Michigan University) and the British Library.
- **France:** Work includes some scanning of old manuscripts with the most notable being the Tresor de la Langue Francaise project at the University of Nancy. The French, along with the Japanese

are also leaders in the Group 7 project which is a museum project. Other efforts are INIST and FOUDRE (1989 to 1992) followed by EDIL and ELITE.

- **The EU:** The European Union funds a large number of international efforts in digital libraries. (Please see page 255 of Michal Lesk's book for details)
- **Japan:** Japan is involved in some digitization and cataloguing efforts and has a \$50M project on. They are also working on modern document delivery and OCR.
- **Australia:** Australia has recently made a modest effort to enter into digital library research. They are planning some digitization projects with a \$10M (Australian) digitization project on the anvil. They are also interested in digitizing Aborigine scriptures and paintings.
- **Elsewhere:** Many other countries are involved in digital library research on much smaller scales. Notable amongst them are Canada, Singapore, Korea and China.

**NOTE 1:** For detailed information on any of the above please refer to Dr. Lesk's book (recommended as supplement text for this course).

**NOTE 2:** See also the table pointing to various national digital libraries from April 1998 CACM [online pages](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

## URLs from Comm. ACM, 41(4), April 1998

This page supplements the Guest Editors' page <http://purl.lib.vt.edu/dlib/pubs/CACM199804> related to the April 1998 CACM Special Section on Digital Libraries: Global Scope, Unlimited Access. Below are all the URLs from that special section, for easy navigation. Please feel free to report any problems or changes. - eaf

### List of articles in the Special Section, with URLs

- Legally Speaking: Encoding the Law into Digital Libraries, by Pamela Samuelson
  - (p. 17) [Digital Future Coalition Web site](#)
- Toward a Worldwide Digital Library, by Edward A. Fox and Gary Marchionini
  - (p. 30) Table 1. National libraries of countries mentioned in this section
    - [Australia - National Library of Australia](#)
    - [Brazil - University of Sao Paulo Library](#)
    - [Canada - National Library of Canada / Bibliotheque Nationale du Canada](#)
    - [Denmark - Det Kongelige Bibliotek](#)
    - [Finland - Helsinki University Library / National Library of Finland](#)
    - [France - Bibliotheque Nationale de France](#)
    - [Germany - Die Deutsche Bibliothek](#)
    - [Hungary - Hungarian Electronic Library](#)
    - [Japan - National Center for Science Information Systems - may not respond](#)
    - [Korea - Five Library Consortium](#)
    - [Netherlands - Koninklijke Bibliotheek](#)
    - [New Zealand - National Library of New Zealand](#)
    - [Singapore - Multiple Agencies](#)
    - [United Kingdom - British Library](#)
    - [United States - Library of Congress](#)
  - p.32 References
    - 1. Borgman, C. [Social Aspects of Digital Libraries](#)
    - 3. Paepcke, A. [Digital libraries: Searching is not enough: What we learned on-site](#)
    - 4. Scherlis, W. [Repository Interoperability Workshop: Towards a repository reference model](#)
- Interoperability for Digital Libraries Worldwide, by Andreas Paepcke et al.
- Accessing Distributed Cultural Heritage Information, by William E. Moen
  - [Aquarelle](#)
  - [Dublin Core Metadata Element Set](#)
  - [Z39.50 Maintenance Agency](#)
- Digital Access to Antiquities, by Henry M. Gladney et al.
  - [Supporting Web pages](#)
- FedStats Promotes Statistical Literacy, by Cathryn S. Dippo
  - (p. 59) [Current Population Survey Data](#)
  - (p. 60) Reference 1. Hert and Marchionini, [Seeking Statistical Information in Federal Websites: Users, Tasks, Strategies, and Design Recommendations](#)
  - [U.S. Department of Labor: BLS Handbook of Methods, Ch. 1, Labor Force Data Derived from the Current Population Survey](#)
  - U.S. Bureau of Labor Statistics: [Labor Force Statistics from the Current Population Survey, Bureau of Labor Statistics Data, Employment Situation](#)
  - U.S. Office of Management and Budget: [Statistical Policy Working Paper 22 - Report on](#)

### [Statistical Disclosure Limitation Methodology](#)

- New Role for Community Networks, by D.D. Cowan et al.
  - [CTT Community Network \(CTTnet\)](#)
  - (p. 63) Reference 3. Hecker, [Advice for Community Network System Designers](#)
  - [Lead author's home page](#)
- Viewing Multilingual Documents on Your Local Web Browser, by A. Maeda et al.
  - [gateway server and multilingual e-text collection](#)
  - [multilingual folk tales](#)
  - (p. 65) Reference 1. Chase et al., [Web Fonts - WC3 Working Draft](#)
  - (p. 65) Reference 3. Dartois et al., [A multilingual electronic text collection of folk tales for casual users using off-the-shelf browsers](#)
- Distributed Chinese Bibliographic Searching, by M.K. Leong, L. Cao, Y. Lu
- NSF-EU Multilingual Information Access, by Judith L. Klavans and Peter Schauble
  - [Digital Library Collaboratory Working Groups information](#)
  - [Multilingual Information Access working group information](#)
- A Public Library Based on Full-text Retrieval, by Ian H. Witten et al.
  - (p. 75) Reference 7. Witten et al., [The New Zealand digital library project](#)
- Students Access Books and Journals through MeDoc, by Albert Endres and Norbert Fuhr
  - [Project home page](#)
  - (p. 77) Reference 5. Fuhr, [Optimum database selection in networked IR, in Proc. SIGIR'96 Workshop on Networked Information Retrieval](#)
- Discovery of Resources within a Distributed Library System, by Laszlo Kovacs
  - [Computer Science Technical Reports](#)
  - [DELOS Working Group](#)
  - [Digital Library Collaboratory Working Groups](#)
  - [ERCIM Digital Library Initiative](#)
  - [European Research Consortium for Informatics and Mathematics](#)
  - [MTA SZTAKI, Department of Distributed Systems](#)
  - [Networked Computer Science Technical Reports Library](#)
  - [NCSTRL overview](#)
  - [Resource Indexing and Discovery in a Globally Distributed Digital Library Workshop](#)
  - (p. 77) Reference 2. Kahn and Wilensky, [A Framework for Distributed Object Services](#)
- Initiatives That Center on Scientific Dissemination, by Marcos Andre Goncalves and Claudia Bauzer Medeiros
  - [Digital Agro-Library - EMBRAPA](#)
  - [UNICAMP Database Group page - see pointers for Geo-Library-Framework \(GeoLib\)](#)
  - [Scientific Electronic Library Online - SciELO](#)
  - [Tropical Database](#)
- R&D for a Nationwide General-Purpose System, by Sung Hyon Myaeng
  - (p. 85) Reference 1. Feedman, [WILLOW: Technical overview](#)
  - (p. 85) Reference 4. NIST, [Guide to Z39.50/PRISE 1.0: Installation, Use, and Modification](#)
- Many Projects That Depend on Collaboration, by Cliff McKnight
  - [Electronic Libraries Programme, eLib](#)
  - [Follett Report, 1993](#)
  - (p. 87) Reference 1. McKnight and Dillon, [User-centred design of library information systems: HyperLib - may not respond](#)
- Libraries' New Role in Electronic Scholarly Publishing, by Andrew E. Treloar
  - [Project home page links](#)
- Semantic Information Retrieval, by Annelise Mark Pejtersen

## Centers, sites and organisations:

---

Some major Digital Library centers and research programs, separately described:

- [Carnegie Mellon University](#)
  - [CNRI](#)
  - [Library of Congress](#)
  - [Stanford University](#)
  - [University of California at Berkeley](#)
  - [University of California at Santa Barbara](#)
  - [University of Illinois](#)
  - [University of Michigan](#)
  - [Texas A&M](#)
  - [Virginia Tech](#)
- 

### Selected other sites:

**ACM DL** : Tap into the ACM Digital Library, a vast resource of bibliographic information, citations, and full-text articles.

### IEEE

- [IEEE Computer Society Digital Library News \(DLN\) Archive](#)
- [IEEE Digital Library Task Force](#)

### IBM

- [IBM DL Home page](#)
- [IBM Renaissance Consortium Panel](#) and [workshop](#)
- [images - QBIC](#)

**OCLC** (OCLC is a nonprofit, membership, library computer service and research organization dedicated to the public purposes of furthering access to the world's information and reducing information costs.

- Research <http://www.oclc.org/oclc/menu/research.htm>  
SiteSearch <http://www.oclc.org/oclc/menu/site.htm>

### Xerox

- [Home Page](#)
- [Scientific American article](#)
- [Scatter/Gather examples](#)
- Questions:
  - Compare
    - What are the various interfaces built? How do they compare? What is the best use of

each?

- Scatter/gather
  - Explain clustering, relate it to scatter/gather.
  - What are special problems with large category systems and how can they be solved?

---

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## CNRI:

---

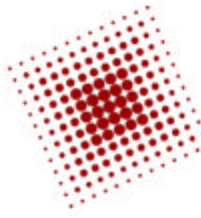
- home page (site map) [http://www.cnri.reston.va.us/site\\_map.html/site\\_map.html](http://www.cnri.reston.va.us/site_map.html/site_map.html)
  - Architecture
    - Kahn-Wilensky Framework for Distributed Digital Object Services  
<http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html>
    - key architectural issues <http://WWW.CNRI.Reston.VA.US/home/cstr/arch/slides.html>
    - architecture for information in digital libraries  
<http://www.dlib.org/dlib/february97/cnri/02arm s1.html>
    - Digital Object Architecture Project <http://www.cnri.reston.va.us/doa.html>
  - CS-TR Computer Science Technical Reports <http://www.cnri.reston.va.us/cstr.html>
- 

[\[Main\]](#) [\[Contents\]](#) [\[Resources\]](#) [\[Centers\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**



The Corporation for  
National Research Initiatives

## Site Map



web-curator@cnri.reston.va.us

### About CNRI

- [CNRI Mission](#)
- [Directions to CNRI](#)
- [Employment Opportunities](#)
- [Information Technology Infrastructure](#)
- [Officers and Directors](#)

### Publications

- [XIWT White Papers](#)
- [D-Lib Magazine](#)
- [IETF Proceedings](#)
- [Infrastructure History Series](#)
- [Recent CNRI Publications](#)
- [CNRI Publications Archive](#)

### Programs and Activities

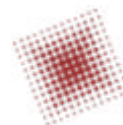
- [Application Gateway System \(AGS\)](#)
- [Cross-Industry Working Team \(XIWT\)](#)
- [D-Lib and D-Lib Magazine](#)
- [Defense Virtual Library](#)
- [Digital Object Architecture](#)
- [Digital Object Identifier System](#)
- [Electronic Payments Forum](#)
- [Grail](#) and [Python](#) and [JPython](#)
- [The Handle System](#)
- [Infrastructure History Series](#)
- [IETF Secretariat](#)
- [IOPS.ORG](#)
- [Knowbot Programs](#)
- [MAGIC](#)
- [National Digital Library Program](#)
- [The Registry](#)
- [Repository Architecture](#)
- [US Copyright Office](#)

### Recent Activities

- [Computer Science Technical Reports](#)
- [Gigabit Testbed Initiative](#)

---

Last Updated: March 26, 1998  
Corporation for National Research Initiatives



# A Framework for Distributed Digital Object Services

Robert Kahn  
Corporation for National Research Initiatives

Robert Wilensky  
University of California at Berkeley

May 13, 1995  
cnri.dlib/tn95-01

## 1. Introduction

This document describes fundamental aspects of an infrastructure that is open in its architecture and which supports a large and extensible class of distributed digital information services. Digital libraries are one example of such services; numerous other examples of such services may be found in emerging electronic commerce applications. Here we define basic entities to be found in such a system, in which information in the form of **digital objects** is stored, accessed, disseminated and managed. We provide naming conventions for identifying and locating digital objects, describe a service for using object names to locate and disseminate objects, and provide elements of an access protocol.

We use the term **digital object** here in a technical sense, to be defined precisely below. Files, databases and so forth that one may ordinarily think of as objects with a digital existence are not digital objects in the sense used here, at least not until they are made into an appropriate data structure, etc., as we will describe shortly.

Only the most basic elements of the infrastructure are described herein. These elements are intended to constitute a minimal set of requirements and services that must be in place to effect the infrastructure of a universal, open, wide-area digital information infrastructure system ("the System"). We anticipate that many other services and elaborations will come into existence as the System is further developed, either building upon or otherwise added to these elements.

This paper focuses on the network-based aspects of the infrastructure, namely those for which knowledge of the contents of digital objects is not required. Definition of the content-based aspects of the infrastructure is purposely not addressed in this paper. An important goal in limiting the description of the infrastructure in this way is not to constrain the higher level user and service level choices that, for many reasons, might be inappropriate to fix upon at this point in time. With only the most basic elements of the infrastructure in place, technological evolution would not be overly constrained. Further, the likelihood of achieving widespread interoperability of services at some early point in the future will be preserved. No doubt the resulting capability will have a greater potential for enhancement and evolution through the participation of many others in helping to define it.

## 2. Overview and Definitions

In this section, we first present an informal overview of the elements of the System, sketching its



## Digital Object Architecture Project

---

CNRI's program of research and development in digital libraries has a number of inter-related activities that overlap and build upon each other. The work includes development of core technology that is used in several testbeds and implementation projects, with funding from a variety of sources.

**The Digital Object Architecture Project** continues the architectural work of the DARPA-funded [Computer Science Technical Reports Project](#) (CSTR). This was a cooperative project, led by CNRI, with five major universities. The project developed network access to archives of technical information in the domain of computer science, and carried out related research, with the goal of evolving knowledge in the field of information storage, search, and retrieval.

The objectives of the Digital Object Architecture project are to enhance the architecture, to continue development of the core technology, and to demonstrate and evaluate them in a number of large-scale testbeds.

Support for the Digital Object Architecture project is provided by DARPA, the Library of Congress, and the Defense Technical Information Center (DTIC), through DARPA grant MDA972-92-J-1029.

### Technology

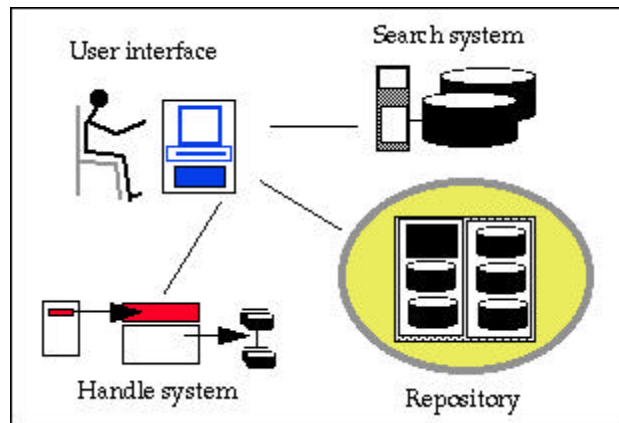


Figure 1

Figure 1 shows the principal system components. CNRI's research concentrates on the concept of digital objects, the Handle System for identifying digital objects, and the Repository for storing them and making them available over the Internet. The Registry is a specialized repository that is used to authenticate digital objects.

**The Handle System** is a system for providing persistent names for Internet resources. It is a highly reliable, high performance, distributed system.

**The Repository** Provides network based storage and access to digital objects. All access to digital objects passes uses a simple repository access protocol and is subject to access controls established by the manager of the repository.

**The Registry** is a specialized repository that provides secure registration and authentication of digital objects.

## Applications, Testbeds, and Partners

**U. S. Copyright Office (CORDS)**. This system provides copyright registration and deposit of digital materials over the Internet. When completed, it will integrate the Registry, Handle System, and Repository with the production systems at the Library of Congress.

**Library of Congress (NDLP)**. The National Digital Library Program is a large-scale program to digitize and make available over the Internet materials from the historic collections at the Library of Congress. CNRI is providing the Repository to manage the collections and the Handle System to identify the digital objects.

**Defense Virtual Library**. CNRI is working in partnership with the Defense Technical Information Center (DTIC) to design and development a digital library for DTIC's extensive collection of report literature.

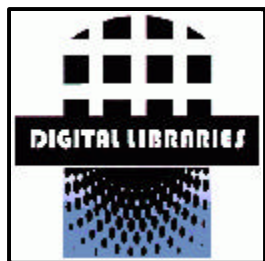
## Papers

**A Framework for Distributed Digital Object Services** by Robert Kahn and Robert Wilensky, May 1995

**Key Concepts in the Architecture of the Digital Library** by William Y. Arms, D-Lib Magazine, July 1995

"Implementation Issues in an Open Architecture Framework for Digital Object Services" by Carl Lagoze and David Ely. Cornell Computer Science Technical Report TR95-1540

"A Design for Inter-Operable Secure Object Stores (ISOS)" by Carl



[THE CENTER](#)

[FACILITIES](#)

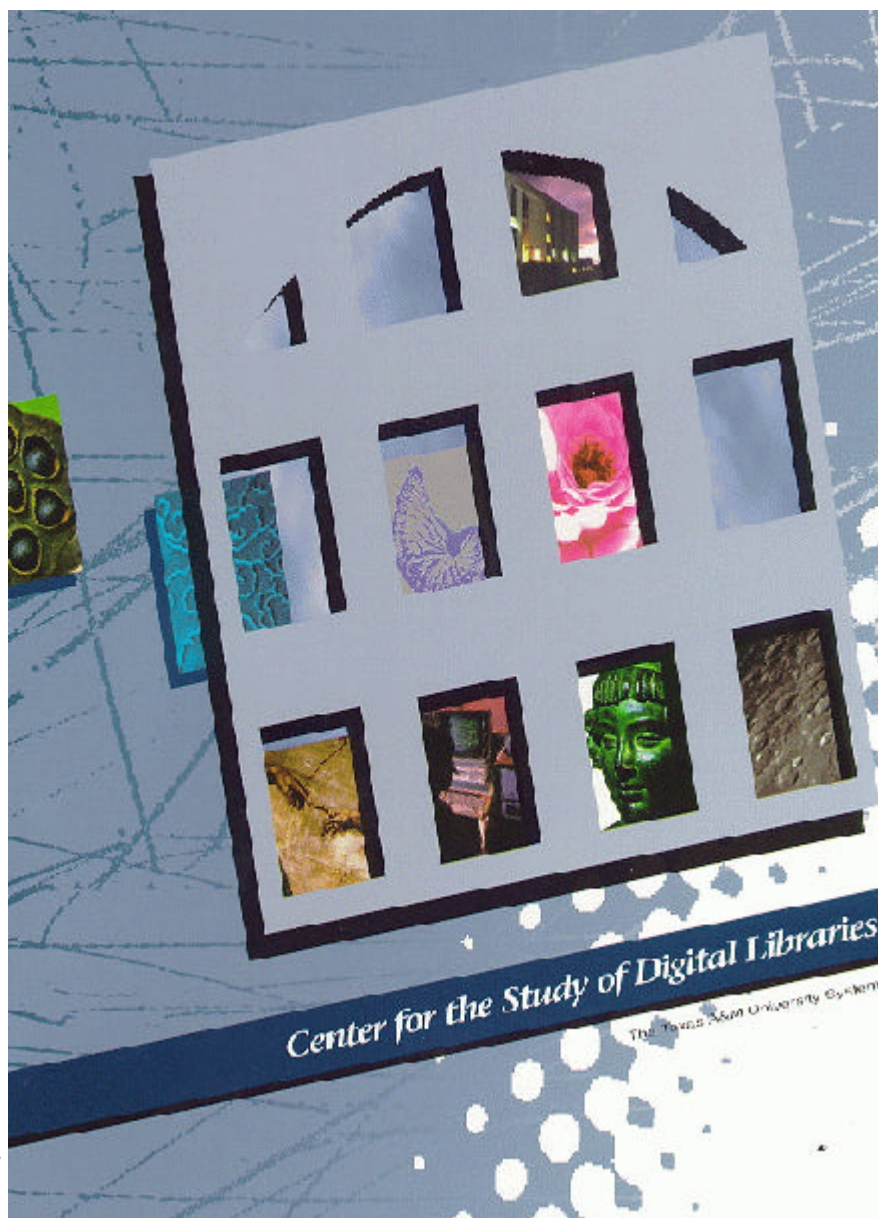
[RESEARCH](#)

[FACULTY/STAFF](#)

[COURSES](#)

[PUBLICATIONS](#)

[CONFERENCES](#)



Center for the Study  
of Digital Libraries  
Texas A&M  
University  
College Station,  
Texas, USA  
77843-3112  
Telephone:  
01-409-862-3217  
Fax:  
01-409-847-8578  
[csdl@csdl.tamu.edu](mailto:csdl@csdl.tamu.edu)



HEWLETT  
PACKARD



The Center for the Study of Digital Libraries gratefully acknowledges the corporate support of the  
**Hewlett-Packard Company; Informix Software, Inc.; and Knowledge Systems, Inc.**

[home](#)[feedback](#)[join/renew](#)[go shopping](#)[search](#)

- [Tables of Contents](#)
- [Search the Digital Library](#)
- [Content Updates](#)
- [Digital Library FAQ](#)
- [Organization of the Digital Library](#)
- [Usage of the Digital Library](#)
- [ACM Members: Create an Account](#)
- [Visitors: Register](#)
- [Download Acrobat® Reader 3.0 to read full-text articles](#)

**Tap into the ACM Digital Library, a vast resource of bibliographic information, citations, and full-text articles.** Access to full-text is by subscription only: ACM members who are Digital Library subscribers have access to all full-text articles. Members and nonmembers who subscribe to electronic publications (but not to the entire Library) have full-text access to their subscriptions only.

**If you are not yet a subscriber, you can still use the Digital Library:** As a service to the computing community, the Digital Library will continue to offer its search and bibliographic database resources to all visitors, for free. All you need to do is register with us.

### Highlights:

#### [How to subscribe to the Digital Library](#)

Rate information on joining ACM and subscribing to the Digital Library

#### [Journal & Magazine Issues](#)

##### [Not Yet in Full Text](#)

Listing of [Conference Proceedings by Sponsoring SIG](#)

#### [ACM Document Delivery Service](#)

To purchase articles published by ACM prior to 1991

---

### What's New:

- [1997 Proceedings in the Digital Library](#)
- Please Read: [Subscription-based Access](#)

---

[SEARCH THE DIGITAL LIBRARY](#) || [TABLES OF CONTENTS](#) || [REGISTER WITH US](#)  
[ABOUT THE DIGITAL LIBRARY](#) || [FEEDBACK](#) || [DIGITAL LIBRARY HOMEPAGE](#)

# Welcome to the IEEE Computer Society Digital Library News (DLN) Archive

This web site will provide access to back issues of the IEEE Computer Society Digital Library Newsletter. Details on subscribing to the newsletter are given below.

---

## Presentation

[Presentation of IEEE Computer Society DLN](#)

---

## Archives

- [Initial Announcement of IEEE Computer Society DLN](#)
  - [June/July 1997 Volume 1 Number 1](#)
  - [January 1998 Volume 1 Number 2](#)
- 

## Points of Contact

### Articles of Interest

and general comments may be directed to:

Sue Feldman  
Datasearch  
170 Lexington Dr.  
Ithaca, NY 14850  
607-257-0937 Phone/Fax  
sef2@cornell.edu

### Subscription Information

To subscribe to DLN, send an e-mail message to: [ieeedln@cimic.rutgers.edu](mailto:ieeedln@cimic.rutgers.edu) with the contents:

subscribe ieedln Your Real Name

### Technical Questions

about this web server or the mailing list can be directed to: [holowcza@cimic.rutgers.edu](mailto:holowcza@cimic.rutgers.edu)

---

Page: <http://cimic.rutgers.edu/~ieeedln/index.html> Updated: Thu Jun 26 11:58:21 EDT 1997

## Technical Activities Forum

Technical Activities Board (TAB) coordinator: Deborah Scherrer, 3061 Palomares Rd., Castro Valley, CA 94552; voice (510) 881-4489; [dscherrer@computer.org](mailto:dscherrer@computer.org)

---

# Digital Library Task Force

Nabil R. Adam and Richard Holowczak  
*Rutgers University, CIMIC*

Milton Halem and Nand Lal  
*NASA Goddard Space Flight Center*

Yelena Yesha  
*UMBC/Center for Excellence in Space Data and Information Sciences*

In the past, global networks have usually transported textual information, but there is a growing need for these networks to transport other forms of information such as images, video, and audio. Until recently, electronic information sources served mainly specialized clients, but now these sources will be accessed by a wide range of users, ranging from computer specialists, discipline experts, engineers, and the general public, including novice computer users and students at all levels.

These trends have created an emerging, important discipline: digital libraries. Several US agencies, including NASA, ARPA, and NSF, have made available over the past few years a considerable amount of money to support research in this field. Other countries, including Canada, the UK, France, Italy, and the Netherlands have also invested in digital library development:

National Library of Canada Electronic Collection: <http://www.nlc-bnc.ca/eppp/e-coll-e.h>  
Initiative for Access—British Library Board: <http://portico.bl.uk/access/overview.html>  
International Institute for Electronic Library Research (involved in several projects):  
<http://ford.mk.dmu.ac.uk>.  
Elite Project (Italy): <http://cosimo.ing.unifi.it/research/elite/elitinfo.html>

As a result of these activities, a number of recent symposia, workshops, and conferences have been recently devoted to digital library issues, and several journals have published editorial about digital libraries, including *Computer* and *IEEE Transactions on Knowledge and Data Engineering*.

## Technical challenges

Digital library development faces challenges in several areas, including the subdisciplines we summarize here.

- **Storage.** A digital library's storage system must be capable of storing a large amount of data in a variety of formats and accessing this data as quickly as possible. Text-only documents—stored in formats such as ASCII, LaTeX, HTML, SGML, and PostScript—are by far the easiest to store. Digital audio and video are more difficult to store because they require significantly more storage



Up



OCLC Home



Search



Site Map



What's New



Feedback



Site Help

[News](#)[About OCLC](#)[OCLC Services](#)[Support & User Doc.](#)[Contacts & Addresses](#)

# OCLC Research

**OCLC identifies research and technical advances** that are of value to the organization in meeting its corporate purposes. Some research projects may lead directly to the development of services for OCLC members, and others may be of general use to the library and information science community. Much of this research and technology assessment is conducted internally, but OCLC also funds or otherwise supports research conducted at universities or other research centers that furthers the corporate purposes of OCLC. OCLC is engaged in a number of research projects in the following areas:

1. to make more effective use of WorldCat (the OCLC Online Union Catalog);
2. to explore the concept of the electronic library;
3. to determine the feasibility of image processing applications; and
4. to improve interface design and human/computer interaction.

► [Mission Statement](#)

► [Projects and Publications](#)

► [Related Conferences](#)

► [Staff](#)

► [Links to Related Web Sites](#)

► [Search OCLC Research](#)



OCLC ONLINE COMPUTER LIBRARY CENTER, INC.



Up



OCLC Home



Search



Site Map



What's New



Feedback



Site Help

News

About OCLC

OCLC Services

Support &amp; User Doc.

Contacts &amp; Addresses

## OCLC Reference Services

### Considering SiteSearch?

- [Overview](#)
- [Guided Tour](#)
- [Demonstrations](#)
- [Components](#)
- [Case Studies & Current Users List](#)
- [How to Order](#)

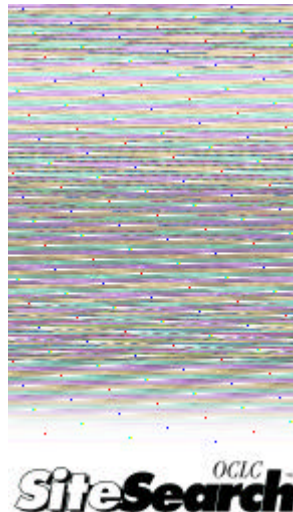
## OCLC SiteSearch

The OCLC SiteSearch suite provides a comprehensive solution for managing distributed library information resources in a World Wide Web environment. It offers tools that **integrate** electronic resources under one Web interface, control **access** to resources, and **build** text and image databases locally.

**New** OCLC SiteSearch was [featured](#) in the January/February 1998 issue of the [OCLC Newsletter](#).

### Using SiteSearch?

- [Support](#)
- [News](#)
- [Product Requirements](#)
- [Documentation & Technical Updates](#)
- [Training](#)



OCLC ONLINE COMPUTER LIBRARY CENTER, INC.

# Interfaces for Information Access

## Companion Pages for *Scientific American* Article [\*Interfaces for Searching the Web\*](#)

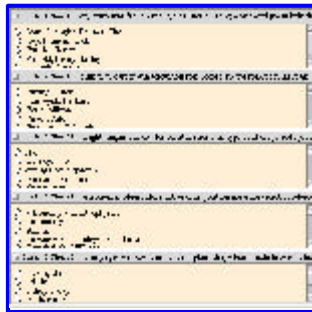
The field of Information Access concerns helping people find, use, understand, and create the information they need, often using computer systems as tools. Information can be found in many forms and media, although much of our research has been concerned with text in general, not focusing exclusively on the Web.

Text analysis and user interface technology must be combined with an understanding of how users work with information and computer tools when building systems to support information access.

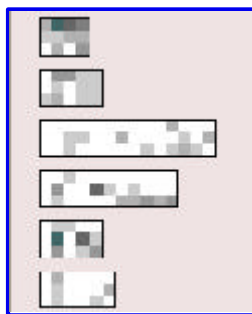
Currently, these pages provide additional information about some of the ideas discussed in the *Scientific American* article *Interfaces for Searching the Web* by [Marti Hearst](#). There is a great deal of research in Information Access at [Xerox PARC](#), of which this pages show only a small sample.

---

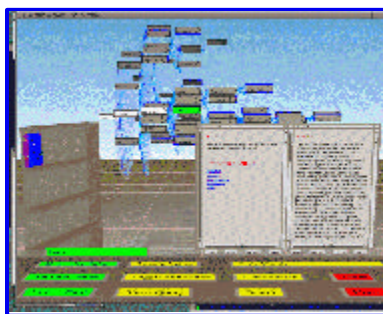
### [About Scatter/Gather](#)



### [About Tilebars](#)



### [About the Cat-a-Cone](#)



---

*Xerox PARC*  
*hearst@parc.xerox.com*  
*6/9/97*



## SPECIAL REPORT

## Interfaces for Searching the Web

**The rapid growth of the World Wide Web is outpacing current attempts to search and organize it.**

**New user interfaces may offer a better approach**

*by [Marti A. Hearst](#)*

**SUBTOPICS:**

- [The \(Slow\) Speed of Thought](#)
- [Organizing Search Results](#)

**[FURTHER READING](#)****[BACK TO THE INTRODUCTION](#)**

How does anyone find anything among the millions of pages linked together in unpredictable tangles on the World Wide Web? Retrieving certain kinds of popular and crisply defined information, such as telephone numbers and stock prices, is not hard; many Web sites offer these services. What makes the Internet so exciting is its potential to transcend geography to bring information on myriad topics directly to the desktop. Yet without any consistent organization, cyberspace is growing increasingly muddled. Using the tools now available for searching the Web to locate the document in Oregon, the catalogue in Britain or the image in Japan that is most relevant for your purposes can be slow and frustrating.

More sophisticated algorithms for ranking the relevance of search results may help, but the answer is more likely to arrive in the form of new user interfaces. Today software designed to analyze text and to manipulate large hierarchies of data can provide better ways to look at the contents of the Internet or other large text collections. True, the page metaphor used by most Web sites is familiar and simple. From the perspective of user interface design, however, the page is unnecessarily restrictive. In the future, it will be superseded by

## A Scatter/Gather Example

Here we demonstrate the use of Scatter/Gather on a collection of encyclopedia articles. Our query is very simple:

**Retrieve the top 250 documents that contain the word *star* .**

Here we show that Scatter/Gather text clustering does a reasonably good job at organizing the documents into meaningful themes or topics.

We ask Scatter/Gather to place the 250 documents into 5 groups. Here is what results. (Bear in mind that encyclopedia articles are well-written and uniform format. The [next example](#) shows the results of a more complicated query on a more unruly text collection.)

<input type="checkbox"/> Cluster 1 Size: 8	key army war francis spangle banner air song scott word poem british
<input type="radio"/> Star-Spangled Banner, The <input type="radio"/> Key, Francis Scott <input type="radio"/> Fort McHenry <input type="radio"/> Arnold, Henry Harley <input type="radio"/> ...	
<input type="checkbox"/> Cluster 2 Size: 68	film play career win television role record award york popular stage p
<input type="radio"/> Burstyn, Ellen <input type="radio"/> Stanwyck, Barbara <input type="radio"/> Berle, Milton <input type="radio"/> Zukor, Adolph <input type="radio"/> ...	
<input type="checkbox"/> Cluster 3 Size: 97	bright magnitude cluster constellation line type contain period spectr
<input type="radio"/> star <input type="radio"/> Galaxy, The <input type="radio"/> extragalactic systems <input type="radio"/> interstellar matter <input type="radio"/> ...	
<input type="checkbox"/> Cluster 4 Size: 67	astronomer observatory astronomy position measure celestial telescop
<input type="radio"/> astronomy and astrophysics <input type="radio"/> astrometry <input type="radio"/> Agena <input type="radio"/> astronomical catalogs and atlases <input type="radio"/> ...	
<input type="checkbox"/> Cluster 5 Size: 10	family specie flower animal arm plant shape leaf brittle tube foot hor
<input type="radio"/> blazing star <input type="radio"/> brittle star <input type="radio"/> bishop's-cap <input type="radio"/> feather star	

## References:

---

- [Courses](#): Digital Library and related courses being offered at various Universities.
- [Conferences/Workshops](#): Links to various conferences/workshops that have been held in the recent past or will be held in the near future.
- [Journals](#): Digital Library related journal information with links.
- [Repositories & Bibliographies](#): contains information and links to some of the repositories maintained by various organizations such as the [D-Lib Magazine](#).
- [Books](#): Some books that contain valuable information on Digital Libraries (along with links to some publishers)

---

[\[Main\]](#) [\[Contents\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

## Digital Library and related courses:

---

- [Digital Library course offered at Pittsburgh](#)
- [Michael Lesk's Digital Library course at Columbia University](#)
- Virginia Tech
  - [CS6604 \(1997\) Digital Libraries](#)
  - [UH3004 Fall 1997 Honors 3004 - Digital Libraries](#)
  - [CS5604 Information Storage and Retrieval](#)
  - [CS4624 Multimedia, Hypertext and Information Access](#)
  - [CS6604 \(1995\) Interactive Accessibility](#)
- CSEI: [NSF CS Education Innovation](#) - projects around the nation
- Furman University:
  - [Exploring the Digital Domain](#)
  - [Web site on creating WWW pages](#)
- [Cyberspace Law for Non-Lawyers](#): This is an electronic course : a "real" course in the "real world" This site includes a discussion function which will allow you, if you are so inclined, to post your own comments and reactions to the individual messages that the instructors have mailed out.
- [Digital Library \(Alexandria\) Online Tutorial at UCSB](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[References\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**



***LIS 2970: Digital Libraries***

[School of Information Sciences](#)  
[University of Pittsburgh](#)

Fall Term, August-December 1997

[Christinger Tomer](#)  
Associate Professor

## **Economics of Digital Libraries: E6998-043**

Columbia University  
Mudd Rm 327  
Tuesdays, 6:10-8pm, 16 January 1996 to 23 April 1996

### [Course Outline & References](#)

#### **Assignments**

[First assignment: Write publishers](#)

[Second assignment: Estimate course needs](#)

[Third assignment: Prepare a pre-proposal](#)

#### **Lecture viewgraphs**

[Course Summary: 16 January viewgraphs](#)

[Library & IR history: 23 January viewgraphs](#)

[Value of information: 30 January viewgraphs](#)

[Copyrights & patents: 6 February viewgraphs](#)

[The course is now running 1/2 session late]

[Buying clubs; sales on Internet: 13 February viewgraphs](#)

[Text and image storage & retrieval: 20 February viewgraphs](#)

[Formats & Image analysis: 27 February viewgraphs](#)

[Scanning resolution: 5 March viewgraphs](#)

[Well, we caught up but 12 March was spring break.]

[Interface evaluation; US Digital library projects: 19 March viewgraphs](#)

[World DL efforts: 26 March viewgraphs](#)

[Archiving & preservation: 2 April viewgraphs](#)

[Cryptography: 2 April viewgraphs](#)

[Social & research issues: 9 April viewgraphs](#)



# CS6604 - Digital Libraries - Fall 1997

## Table of Contents

- **Contacts:** [Instructor](#), [GTA](#)
- [Department and Class Policies](#)
- [DLI Articles in IEEE Computer](#)
- [Explore on WWW](#)
- [HyperNews](#)
- [Labs](#)
- **Lectures:** [970911](#), [Multimedia in Digital Libraries](#)
- [Listserv](#)
- [News / Announcements](#) (updated 971203@2am)
- **Overview:** [WWW](#), [PDF - Part 1](#), [PDF - Part 2](#)
- **Projects:** [Grading](#); [Guidelines](#), [old examples](#); [List of project ideas](#); [Projects and People](#)
- **Quizzes:** [Password request](#), [Take a quiz](#) -- Let [Patrick](#) know if you have any problems.
- [Syllabus](#)
- [Topics](#)
- [UNIX Use Hints \(for video accounts\)](#)
- [Materials on reserve in Newman Library](#)

---

## Related Courses

- UH3004 Fall 1997 [Honors 3004 - Digital Libraries](#)
- CS5604 [Information Storage and Retrieval](#)
- CS4624 [Multimedia, Hypertext and Information Access](#)
- CS6604 (1995) [Interactive Accessibility](#)

---

Please send comments and suggestions to: [fox@fox.cs.vt.edu](mailto:fox@fox.cs.vt.edu)

# Education Innovation

NSF CISE funds Education Innovation projects, like our own local [EI project](#). Since some of these relate to digital libraries, and all to CS education, it is worthwhile examining key aspects.

- [Home Page for CSEI](#)
- Using modules: [Brooklyn College](#) - Distributed Processing
- Integrating research: [Evergreen State College, OGI](#) - software engineering of scientific systems; formal methods and higher order logics; and neural networks applied to speech recognition
- Multimedia Support: [Georgia Tech](#) - including STABLE, a case library of projects in Smalltalk
- Scheme Programming: [Indiana U.](#) - including a [Scheme Repository](#)
- Parallel and Distributed Computing: [Louisiana Tech University](#) - Java Concurrency Simulator
- Intro Labs: [Oberlin College](#) - [HtX](#) tool to generate WWW pages, [labs, etc.](#)
- Networking Labs: [Ohio State U.](#) - [7 labs](#), [software](#)
- Multi-semester projects: [ODU](#) - [Computer Productivity Initiative \(CPI\)](#)
- Software Design & Development: [RPI](#) - [Design Conference Room](#), Standard Template Library (STL) since incorporated into standard C++ library
- Programming: [Rice](#) - [14 labs with Scheme and Java](#)
- HCI: [San Jose State U.](#) - [devices and distance education](#)
- High-Performance Scientific Computing: [U. Colorado Boulder](#) - [tutorials, etc.](#)
- Wireless Comm.: [VT/UMR](#) - [3 courses](#)
- TeleMentoring: [U. Penn.](#) - ATM distance education, seminars
- Literacy: [Utah State U.](#) - multimedia modules

## Conferences/Workshops:

---

- ACM DL'98: Pittsburgh, June 23-26 <http://fox.cs.vt.edu/confs/DL98.txt>
- ACM DL'97: Philadelphia, July 23-26 <http://www.lis.pitt.edu/~diglib97/>
- DL'96: Bethesda, March (1st ACM ...) <http://fox.cs.vt.edu/DL96/>
- DL'95: Austin, June <http://csdl.tamu.edu/DL95/>
- DL'94: Texas A&M University
- Santa Fe Workshop, Digital Knowledge Work Environments, March 9-11, 1997  
<http://www.si.umich.edu/SantaFe/>
- UCLA Workshop, Social Aspects of Digital Libraries, Feb. 16-17, 1996  
<http://www-lis.gseis.ucla.edu/DL/>
  - [life cycle](#)
- [IITA Digital Libraries Workshop, 1995](#)
- Allerton, 1996 <http://edfu.lis.uiuc.edu/allerton/96/> and [map](#)
- Allerton, 1995 <http://edfu.lis.uiuc.edu/allerton/95/>
- ADL 96, Forum on Research and Technology Advances in Digital Libraries May 13-15, 1996, Washington, D.C.
- KOLISS DL 96, Proc. Int'l Conf. on Digital Libraries and Information Services for the 21st Century, Sept. 10-13, 1996, Seoul, Korea
- D-Lib supported meetings, conferences and workshops <http://www.dlib.org/groups.html>

---

[\[Main\]](#) [\[Contents\]](#) [\[References\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

## Journals:

---

Selected special issues include:

- Commun. ACM
  - [April 1995](#): 38(4)
  - [April 1998](#): 41(4)
- [IEEE Computer, May 1996](#)
- J. American Society for Information Science, Sept. 1993: 44(8)
- J. of Visual Communication and Image Representation, 7(1), March 1996

There also are closely related journals like:

- J. of Digital Information (British Computer Society)

---

[\[Main\]](#) [\[Contents\]](#) [\[References\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

# Repositories & Bibliographies:

---

- **D-Lib** <http://www.dlib.org/>
  - Articles (by author) <http://www.dlib.org/author-index.html>
  - Articles (by title) <http://www.dlib.org/title-index.html>
  - Research Projects (incl. DLI) <http://www.dlib.org/projects.html>
  - D-Lib Working Groups <http://www.dlib.org/groups.html>
    - Metadata <http://www.dlib.org/metadata/overview.html>
    - Naming <http://www.dlib.org/naming/overview.html>
    - Repository Interfaces <http://www.dlib.org/repository/overview.html>
    - Social Aspects <http://www.dlib.org/social/overview.html>
  - D-Lib Magazine Articles on Key Topics
    - Agents <http://www.dlib.org/dlib/July95/07birmingham.html>
    - Architecture (incl. handles) <http://www.cnri.reston.va.us/home/dlib/July95/07arms.html>
    - Metadata <http://www.dlib.org/dlib/July95/07weibel.html>
    - Uniform Resource Names (URNs) <http://www.dlib.org/dlib/february96/02arms.html>
    - Use <http://www.dlib.org/dlib/october95/10bishop.html>
    - Informedia <http://www.dlib.org/dlib/july96/07wactlar.html>
    - Variations <http://www.dlib.org/dlib/june96/06fenske.html>
    - Access Control: [Articles by Gladney et al.](#)
- UIUC Pointers to Publications <http://dli.grainger.uiuc.edu/pubs/natsynch.htm>
- **Annotated Bibliography for Digital Libraries** - Christine Woerner, 1996 (Case Studies, DL as Place, Archive/Organization/Preservation, Librarianship, Mediation/Interaction, Authoring/Authenticity/Originality)
- Virginia Tech
  - [Digital Library Research Laboratory Publications](#)
  - [BibTeX file](#) for article: E. Fox and O. Sornil. Digital Libraries. Chapter 11 in Modern Information Retrieval, AWL England, 1999: Ricardo Baeza-Yates and Berthier Ribeiro-Neto, eds., to appear.
  - misc ptrs <http://scholar.lib.vt.edu/digilib/>

---

[\[Main\]](#) [\[Contents\]](#) [\[References\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta



► **D-Lib Magazine**

► **Current Issue**

► **Back Issues**

► **Ready Reference**



► **D-Lib Metrics**

► **Activity Hosted by D-Lib**

► **NSF SMETE-Lib Study**

Welcome to D-Lib, an evolving community of research interests in digital libraries.

From here, you may visit **D-Lib Magazine**, a monthly compilation of contributed stories, commentary, and briefings. Or you may find Ready Reference a convenient clearinghouse of pointers to other sites on the web of interest to researchers and users of digital libraries. Finally, the D-Lib Working Group on Digital Library Metrics addresses the thorny problem of developing appropriate metrics.

Research in digital libraries is an expanding horizon with applications for today and implications for tomorrow. We invite you to read, observe, and participate.



---

Last Updated 6/15/98 af:cr



Site for publications from all six Digital Libraries Initiative projects.

- ⊙ **UNIVERSITY OF CALIFORNIA  
at Berkeley**
- ⊙ **UNIVERSITY OF CALIFORNIA  
at Santa Barbara**
- ⊙ **CARNEGIE MELLON UNIVERSITY**
- ⊙ **UNIVERSITY OF ILLINOIS  
at Urbana-Champaign**
- ⊙ **UNIVERSITY OF MICHIGAN**
- ⊙ **STANFORD UNIVERSITY**

Comments to [Susan Harum](#)

Last updated 5/26/98

# An Agent-Based Architecture for Digital Libraries

William P. Birmingham  
The University of Michigan  
Electrical Engineering and Computer Science Department  
School of Information Science and Library Studies  
Ann Arbor, MI 48109  
[wpb@eecs.umich.edu](mailto:wpb@eecs.umich.edu)

D-Lib Magazine, July 1995

- 
- [Introduction](#)
  - [Agents](#)
  - [What the architecture provides](#)
  - [The Conspectus and the conspectus language](#)
  - [Status and summary](#)
  - [Acknowledgements](#)
  - [References](#)
- 

A button with the text "d-lib forum" in a stylized font, enclosed in a rounded rectangle with a double border.

A button with the text "d-lib magazine" in a stylized font, enclosed in a rounded rectangle with a double border.

---

## Introduction

One of the most exciting promises of digital libraries is access to a great variety of information and *services* that transcend what is available today through on-line services, such as the World-Wide Web (WWW). A library is more than just stacks of materials on shelves; it is also highly trained people that provide valuable services. These services include such things as *organization and cataloging*, research, notification of new publications, and so forth. Indeed, one of the greatest assets of libraries are these high-valued services. The WWW, while it probably contains more information than any single traditional library, is arguably not as useful as a traditional library because it lacks these services (particularly organization and sophisticated search support). No one is dismantling their libraries because of the WWW yet. The University of Michigan Digital Library Project (UMDL) [1,2] believes that a successful digital library needs to provide both access to a wide variety of valuable content and services.

Because the range of both content and services that are possible for a digital library are potentially large (we cannot even imagine what will be available or needed in the future), there will be no single, complete digital-library solution. Rather, we expect that as editing tools become better and access to networks becomes easier and cheaper, there will be millions of content suppliers; "everyman" can

# Key Concepts in the Architecture of the Digital Library

William Y. Arms  
Corporation for National Research Initiatives  
Reston, Virginia  
*warms@cnri.reston.va.us*

**D-Lib Magazine**, July 1995

---

## Introduction

For the past two years, the Computer Science Technical Reports project (CS-TR) has been developing an architecture for a digital library with funding from the Department of Defense's Advanced Research Projects Agency (ARPA). This is a general purpose framework for a digital library in which very large numbers of objects, comprising all types of material, are accessible over national computer networks. It is described in a paper by Robert Kahn and Robert Wilensky (cnri.dlib/tn95-01).

This introduction describes the author's view of eight general concepts that emerged from the discussions. These concepts are key issues in the transition to a true digital library from the network services that we have today. The Kahn/Wilensky paper contains a comprehensive framework for resolving the issues.

## General Principles

- [1. The technical framework exists within a legal and social framework](#)
- [2. Understanding of digital library concepts is hampered by terminology](#)
- [3. The underlying architecture should be separate from the content stored in the library](#)
- [4. Names and identifiers are the basic building block for the digital library](#)
- [5. Digital library objects are more than collections of bits](#)
- [6. The digital library object that is used is different from the stored object](#)
- [7. Repositories must look after the information they hold](#)
- [8. Users want intellectual works, not digital objects](#)
- [Reference](#)

# Metadata: The Foundations of Resource Description

Stuart Weibel

Office of Research, OCLC Online Computer Library Center, Inc.

*weibel@oclc.org*

**D-Lib Magazine, July 1995**

---

This paper is an abbreviated version of the [Summary Report of the OCLC/NCSA Metadata Workshop](#). It sets forth a proposal for the content of a simple resource description record (the Dublin Core Metadata Element Set) and outlines a series of further steps to advance the standards for the description of networked information resources.

- [Introduction](#)
  - [Underlying Assumptions](#)
  - [Implementations](#)
  - [Next Steps](#)
  - [References](#)
- 

**d-lib forum**

**d-lib magazine**

---

## Introduction

The explosive growth of interest in the Internet in recent years has created a digital extension of the academic research library for certain kinds of materials. Valuable collections of texts, images and sounds from many scholarly communities -- collections that may even be the subject of state-of-the-art discussions in these communities--now exist only in electronic form and may be accessible from the Internet. Knowledge regarding the whereabouts and status of this material is often passed on by word of mouth among members of a given community. For outsiders, however, much of this material is so difficult to locate that it is effectively unavailable.

Why is it so difficult to find items of interest on the Internet or the World Wide Web? A number of well-designed locator services, such as Lycos [\[http://lycos.cs.cmu.edu/\]](http://lycos.cs.cmu.edu/), are now available that automatically index many of the resources available on the Web and maintain up-to-date databases of locations. But indexes are most useful in small collections within a given domain. As the scope of their coverage expands, indexes succumb to problems of large retrieval sets and problems of cross disciplinary semantic drift. Richer records, created by content experts, are necessary to improve search

## MAGAZINE

# Informedia Digital Video Library

## Technology Outreach

Howard D. Wactlar  
Carnegie Mellon University  
*Howard.Wactlar@cs.cmu.edu*

**D-Lib Magazine**, July/August 1996

ISSN 1082-9873

## Background

The [Informedia Digital Video Library at Carnegie Mellon University](http://www.dlib.org/dlib/july96/07wactlar.html) is one of the NSF/DARPA/NASA jointly funded Digital Library Initiative projects, established in 1995. This particular effort focuses on search and discovery in the video medium. The Informedia project will establish a large, on-line digital video library by developing intelligent, automatic mechanisms to populate the library and allow for full-content and knowledge-based search and retrieval via desktop computer and metropolitan area networks. Initially, the library will be populated with several thousand hours of raw and edited video drawn from licensed public television documentaries and broadcast news and special events. The library is being deployed in testbeds at local area K-12 schools, at Carnegie Mellon University, and as demonstration systems at government sponsors.

The distinguishing feature of our technical approach is the integrated application of speech, language and image understanding technologies for efficient creation and exploration of the library. Using a high-quality speech recognizer, the sound track of each videotape or broadcast, combined and aligned with closed-captioning information when available, is converted to a textual transcript. A language understanding system then analyzes and organizes the transcript and stores it in a full-text information retrieval system. Likewise, image understanding techniques are used for segmenting video sequences by automatically locating boundaries of shots, scenes, and conversations. The system thus partitions video into small-sized segments and provides alternate representations and abstractions of video content to better support information retrieval and manipulation. Exploration of the library is based on these same techniques.

## Component and Content Availability

### Present

The highly modular system structure and implementation of the Informedia Digital Video Library system is itself a fertile testbed for researchers in many disciplines. Any of the component systems (e.g., speech recognition, image sequence segmentation; user interface display and control tools; text indexing, search and retrieval; video servers; network streaming protocols; dynamic pricing algorithms) can be exported for use in other research projects elsewhere. It is our intent to encourage investigation by DLI researchers

## MAGAZINE

# The VARIATIONS Project at Indiana University's Music Library

David E. Fenske  
Head, Music Library  
VARIATIONS Project Director  
Indiana University  
[fenske@indiana.edu](mailto:fenske@indiana.edu)

Jon W. Dunn  
VARIATIONS Project Technical Director  
Indiana University  
[jwd@indiana.edu](mailto:jwd@indiana.edu)

**D-Lib Magazine, June 1996**

ISSN 1082-9873

---

## History, Context and Background

The VARIATIONS Project is best known for the distribution of high-quality digital audio via an ATM network from servers and storage systems having some special characteristics to Intel-based and Macintosh clients. The evolution of this project from its beginnings in the late 1980's to its initial operational state today is inextricably connected with the design and construction of a new [School of Music Library](#) at Indiana University, and with the opportunities presented by a new design. It also addresses some pedagogical and library preservation problems. This article describes the motivation for the project and its history, its operation and experiences to date, and its future goals. Although the project is now operational, this report should not be viewed as a final one. VARIATIONS is a work in progress and represents several partnerships within Indiana University and our partnership with IBM. Information can be obtained about our internal partnerships by following links to the Indiana University [School of Music](#), the [Indiana University Libraries](#), Indiana University's [University Computing Services](#). The VARIATIONS Project, as a result of its partnership with IBM, uses many of the [IBM Digital Library](#) technologies. Information about the Indiana University School of Music Library's relationship to [IBM](#)'s plans is publicly available at the IBM Digital Library site.

Common knowledge has it that university buildings take a long time to accomplish. We can validate this observation. The first internal documents for a new music library were written in 1977. The officially endorsed proposal was first produced in 1983 with subsequent revisions in 1986 and 1989. It was with the 1989 version that the new Music Library was built.

In the earlier versions, a traditional library of the time was envisioned. The principal issue was providing twenty years of collection growth without compromising the available number of readers and listeners. The debate in 1983 was over allocating space to the listeners versus the readers.

# Annotated Bibliography - Digital Libraries

## Case Studies:

- UC Berkeley, Digital Library Project  
 Van House, N.A., Butler, M.H., Ogle, V., & Schiff, L. (1996, February). User-Centered iterative design for digital libraries: the Cypress experience. D-Lib Magazine 2 (2),(On-line). Available: <http://www.dilb.org/dlib/february96/02vanhouse.html>  
 This paper reports on the experiences and results of the project's design and evaluation process. After the image-retrieval system Cypress is introduced, the method of assessment and evaluation is described. Important findings for interface design are reported, such as reduction of cognitive load and error prevention, and a table is provided to demonstrate the changes made to the interface.
- Van House, N., Butler, M., & Schiff, L. (1996). Needs assessment and evaluation of a digital library environment: the Berkeley experience, (On-line). Available: <http://info.sims.berkeley.edu/~vanhouse/dl96.html>  
 This report describes the design approach and context and preliminary findings of the assessment and evaluation study. A number of desired uses were identified, such as locating information, analyzing data, dissemination, publishing and re-use of information.
- University of California, Santa Barbara (Alexandria Project)  
 Alexandria Digital Library Project Team (1996). 1996 annual report, (On-line). Available: <http://alexandria.sdc.ucsb.edu/public-documents/annual-report/>  
 Detailed report about the progress of the ADL until February 1996 and plans for the future. Includes sections on the WWW prototype development and the testbed components, management issues (organizational structure, staff), various research efforts (such as metadata, interface design, evaluation, image processing), collaborations with outside institutions and research projects, and finally, educational activities.
- Frew, J., Carver, L., Fischer, C., Goodchild, M., Larsgaard, M., & Smith, T. (1995). The Alexandria rapid prototype: building a digital library for spatial information. 1995 ESRI User Conference Proceedings, (On-line). Available: <http://www.esri.com/resources/userconf/proc95/to300/p255.html>  
 Describes the first testbed developed at ADL, the Alexandria Rapid Prototype. In particular, the user interface, the prototype architecture (catalog, collection, software, hardware) and future plans are introduced.
- Smith, T.R. (1996). A digital library for geographically referenced materials. Proceedings of Untangling the Web 1996, Santa Barbara, California, (On-line). Available: <http://www.library.ucsb.edu/untangle/frew.html>  
 This paper describes the different elements of the ADL, such as its general architecture, the catalog component, the user interface, and image processing. For each component technical and methodological issues are discussed.
- University of Illinois at Urbana-Champaign  
 Bishop, A.P. (1995). Working towards an understanding of digital library Use. D-Lib Magazine 1 (3),(On-line). Available: <http://www.dlib.org/dlib/october95/10bishop.html>  
 Describes the joint efforts of the user-research working groups of six DLI projects.

# DLRL Related Publications

---

## Table of Contents

- [1](#) BOOKS
  - [2](#) JOURNAL ARTICLES
  - [3](#) BOOK CHAPTERS
  - [4](#) REFEREED CONFERENCE/WORKSHOP PAPERS
  - [5](#) REFEREED POSTERS
  - [6](#) UNREFEREED INVITED CONFERENCE/WORKSHOP PAPERS
  - [7](#) KEYNOTE / BANQUET / DISTINGUISHED SPEAKER PRESENTATIONS
  - [8](#) ORAL PRESENTATIONS
  - [9](#) TECHNICAL REPORTS, OTHER PUBLICATIONS
- 

## 1. BOOKS

- E. Fox and G. Marchionini, eds., Proceedings of the First ACM International Conference on Digital Libraries, DL'96, Bethesda, MD, March 20-23, 1996.
  - E. Fox, ed. Sourcebook on Digital Libraries: Report for the National Science Foundation, TR-93-35, VPI&SU Computer Science Dept., Dec. 1993, Blacksburg, VA. Available by anonymous FTP from directory pub/DigitalLibrary on fox.cs.vt.edu, over 400 pages.
- 

## 2. JOURNAL ARTICLES

- E. Fox and G. Marchionini. Toward a Worldwide Digital Library. Guest Editors' Introduction to special section (pp. 28-98) on Digital Libraries: Global Scope, Unlimited Access. Commun. of the ACM, Apr. 1998, 41(4): 28-32. <http://purl.lib.vt.edu/dlib/pubs/CACM199804>
- E. Fox. Networked Digital Library of Theses and Dissertations: An International Collaboration Promoting Scholarship. ICSTI Forum, Quarterly Newsletter of the International Council for Scientific and Technical Information, No. 26: 8-9, Nov. 1997. <http://www.icsti.nrc.ca/icsti/>
- E. Fox, R. Hall, and N. Kipp. NDLTD: Preparing the Next Generation of Scholars for the Information Age. The New Review of Information Networking (NRIN), 3: 59-76, 1997.
- Edward A. Fox, Robert Hall, Neill A. Kipp, John L. Eaton, Gail McMillan, and Paul Mather. NDLTD: Encouraging International Collaboration in the Academy. In Special Issue on Digital Libraries, DESIDOC Bulletin of Information Technology, 17(6): 45-56, Nov. 1997.
- E. Fox, J. Eaton, G. McMillan, N. Kipp, P. Mather, T. McGonigle, W. Schweiker, and B. DeVane. Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources. D-Lib Magazine, The Magazine of Digital Library Research, ISSN 1082-9873, Sep. 1997.
- E. Fox, J. Eaton, G. McMillan, N. Kipp, L. Weiss, E. Arce, S. Guyer. National Digital Library of

## Books:

---

There is only one really good book on digital libraries:

- Michael Lesk, [Practical Digital Libraries](#), Morgan Kaufmann, 1997, San Francisco

For a history of many digital library activities through Fall 1993, including reports on key workshops, see:

- Digital Library Source Book, Edward Fox, ed., 1993 <http://fox.cs.vt.edu/DLSB.html>

In the related field of Information Retrieval the best set of readings is:

- Karen Sparck Jones and Peter Willett, [Readings in Information Retrieval](#), Morgan Kaufmann, 1997, San Francisco

Some miscellaneous related works include:

- Elsevier, [TULIP Final Report](#), 1996, New York. This booklet was distributed after completion of the TULIP digital library prototype [project](#) by [Elsevier](#), and led to their current digital library effort, [EES](#).
- Hermann Maurer, ed., Hyper-G/Hyperwave: The Next Generation Web Solution, Addison Wesley Longman, 1996, Harlow, England
- Setrag Khoshafian, A. Brad Baker, MultiMedia and Imaging Databases, Morgan Kaufmann, 1996, San Francisco
- V.S. Subrahmanian, Sushil Jajodia, eds., Multimedia Database Systems: Issues and Directions, Springer, 1996, Berlin

---

[\[Main\]](#) [\[Contents\]](#) [\[References\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## Topics:

---

- [Search, retrieval, resource discovery](#) (See Chapter 2 in Dr. Lesk's book.)
  - [Multimedia, representations](#) (See Chapter 4 in Dr. Lesk's book.)
  - [Architectures](#) (See Chapter 6 in Dr. Lesk's book.)
  - [Interfaces](#) (See Chapter 7 in Dr. Lesk's book.)
  - [Metadata](#)
  - [Electronic publishing, SGML](#)
  - [Database issues](#)
  - [Agents](#)
  - [Commerce, economics, publishers](#) (See Chapter 9 in Dr. Lesk's book.)
  - [Intellectual property rights, copyright laws & security](#) (See Chapter 10 in Dr. Lesk's book.)
  - [Social issues](#) (See Chapters 11, 12 in Dr. Lesk's book.)
- 

## Pedagogy:

We recommend that the topics be covered in the order given above, with the reader examining the material in the book by Dr. Lesk before visiting the online information. Topics that do not correspond to chapters in the book have been included as supplementary material that seemed to be of special interest to students at Virginia Tech, and/or where there is keen interest and progress by the digital library community. However, these can be skipped by novices interested in a general overview.

---

[\[Main\]](#) [\[Contents\]](#)

---

# Search, retrieval, resource discovery:

---

## Searching - LoC

- [LoC Home Page](#)
- [The WWW Virtual Library arranged by LoC standards](#)
- [UNDERSTANDING AND COMPARING WEB SEARCH TOOLS](#)
- [Matrix of WWW Indices: A comparison of Internet indexing tools](#)

## **Federated search**

- [UIUC Federation Across Heter. DBs](#)
- [STARTS](#)
- [INFOSEEK patent](#)
- [TSIMMIS](#)
- [Virginia Tech Federated Search Demonstration for NDLTD \(theses, dissertations\)](#)

## **CyberStacks (WWW, Classification, Catalogs, Reviews/Clearinghouses)**

- [Home Page](#)
- [Net Projects](#)
- [Alphabetical topics vs. LC ranges](#)
- [Call for contributions](#)
- Question: Which efforts are far along? What demonstrations can you find that are the most informative / explanatory? How well does the Library of Congress classification system fit for WWW resources?
- Related work: [OCLC's Scorpion Project](#); [DDC](#)

## **Columbia**

- [D-Lib Article on Images/Video](#)
- [WebSeek Home Page](#)

## **BioKleisli**

- [project](#)
- [demos](#)

## [Filtering](#)

### [Cross-Language Information Retrieval Resources](#)

- [Eurospider Demos](#)
- [Analogical Language Processor Demo](#)
- [Mundial](#) - English and Spanish Demo
- Questions:
  - What languages are covered?
  - How well are phrases handled?

### [Stanford DL info finding projects](#)

[Berkeley documents and queries](#) (please study carefully, answering questions)

### [UCSB spatial indexing and retrieval](#)

---

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

## UNDERSTANDING AND COMPARING WEB SEARCH TOOLS

### [Beyond Surfing: Tools and Techniques for Searching the Web](#)

by Kathleen Webster & Kathryn Paul, January 1996

### [General Internet Resource Finding Tools:](#)

A Review and List of Those Used to Build INFOMINE, March 1996

### [How to Search the Web - A Guide to Search Tools](#)

by Terry A. Gray

### [Jacob Hausauer's Page for Search Engines](#)

March, 1996

### [Just the Answers, Please](#)

©, Susan Feldman

Searcher Magazine, 1997

*note: this link will expire in May, 1998*

### [Librarians' Index to the Internet](#)

Lots of useful links about searching. Be sure to check "about"

### [Literature about search services](#)

by Traugott Koch

January, 1996; updated Nov., 1996

### [Precision among World Wide Web Search Services \(Search Engines\): Alta Vista, Excite, Hotbot, Infoseek, Lycos](#)

By H. Vernon Leighton and Dr. Jaideep Srivastava,  
June 1997

### [Reviews of Search Engines](#) from the Search Page.

June 1996; updated, November 1996

### [Search the Net: Top Internet Searching Resources Reviewed](#)

by Tracy Marks

February, 1997; updated October, 1997

### [Signal Detection Analysis of WWW Search Engines](#)

by Carsten Schlichting & Erik Nilsen, Lewis & Clark College  
1996

### [Top keyword Resources of the Web](#)

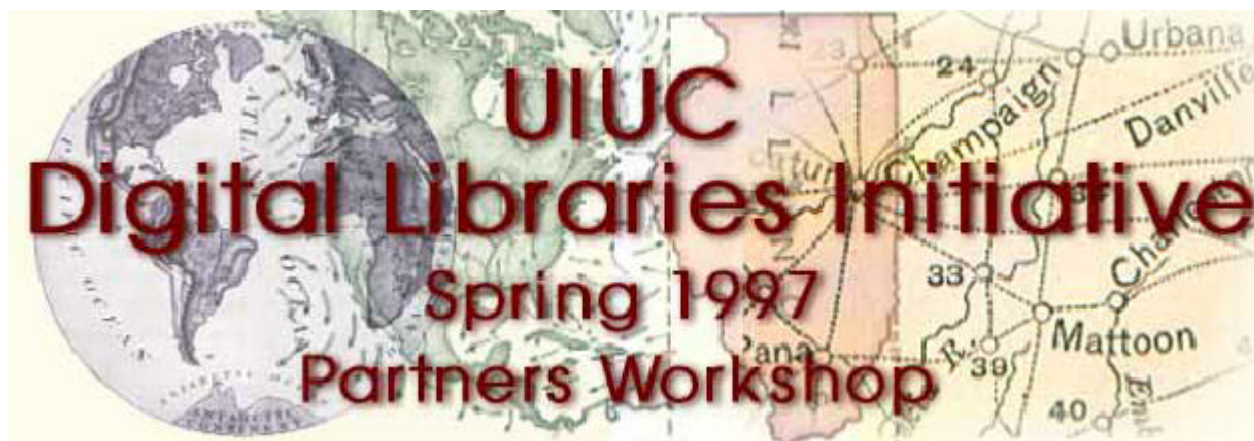
by John December, November, 1996

### [Tips on Popular Search Engines](#)

by Karen Campbell, March 1997

### [Top keyword Resources of the Web](#)

by John December, November, 1996



## CONTENTS



[Minutes](#)



[Presentations](#)



[AGENDA](#)



[ATTENDEES](#)

## FEDERATION ACROSS HETEROGENEOUS DATABASES

April 3-4, 1997

Grainger Engineering Library Information Center

University of Illinois at Urbana-Champaign  
1301 W. Springfield Ave., Urbana, IL

Welcome to the official site for the UIUC Digital Library  
Initiative Spring '97 Partners Workshop.

Please contact Susan Harum [dli@uiuc.edu](mailto:dli@uiuc.edu) for any questions  
or comments about the workshop.

[Go back to the DLI workshop page](#)

# ***STARTS***

## ***Stanford Protocol Proposal for Internet Search and Retrieval***

*STARTS* is the result of an informal "standards" effort that we ([Luis Gravano](#), [Kevin Chang](#), [Hector Garcia-Molina](#), [Carl Lagoze](#), and [Andreas Paepcke](#)) are coordinating at Stanford. This project developed a simple protocol that text search engines should follow to facilitate searching and indexing multiple collections of text documents.

[Final writeup](#) of the *STARTS* protocol ([PostScript version](#))

[A reference-implementation](#) of *STARTS* by Carl Lagoze

[A more readable description](#) of the *STARTS* protocol that appeared in Sigmod'97

[List of participants](#) of the *STARTS* Workshop, Stanford, August 1st, 1996

Slides of the talk that Prof. Hector Garcia-Molina gave at the *STARTS* workshop ([Powerpoint Version](#))

Slides of the talk that Luis Gravano gave at the *STARTS* workshop ([Powerpoint Version](#))

---

[Luis Gravano](#)  
[gravano@cs.stanford.edu](mailto:gravano@cs.stanford.edu)



# Distributed Search Patent

## [Patents](#)

The Infoseek Distributed Search patent is a novel technique for performing full-text searches over distributed databases. The technique is directly applicable to searching web sites on the Internet, as well as geographically distributed databases within corporate Intranets. The patent, US Patent Number 5,659,732, entitled "Document Retrieval over networks wherein ranking and relevance scores are computed at the client for multiple database documents," was issued on August 19, 1997.

## [Press release](#)

The official corporate press release announcing the patent

## [Background information](#)

An expanded press release containing more technical information and additional background information

## [News Articles](#)

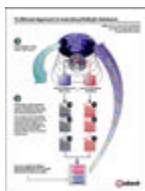
News articles appeared in [The New York Times](#), [Inter@ctive Week](#), and [CNET](#).

## [Patent text](#)

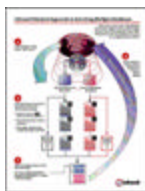
Text of the patent

## [Graphic](#)

Illustration of traditional vs. Infoseek patent approach. This file contains two pages: the first page depicts the traditional approach, while the second page portrays the Infoseek patented approach. Available in GIF (left) or PDF (below) format.



[Traditional Approach](#)  
(GIF)



[Infoseek Approach](#)  
(GIF)



[PDF format](#)



## Why the name?

As an acronym, TSIMMIS stands for "*The Stanford-[IBM](#) Manager of Multiple Information Sources.*" In addition, TSIMMIS is a Yiddish word for a stew with "heterogeneous" fruits and vegetables integrated into a surprisingly tasty whole.

## Short Project Description

The goal of the TSIMMIS Project is to develop tools that facilitate the rapid integration of heterogeneous information sources that may include both structured and semistructured data. TSIMMIS has components that:

- translate queries and information (source wrappers);
- extract data from World Wide Web sites;
- combine information from several sources (mediator);
- allow browsing of data sources over the Web.

The TSIMMIS project is funded by [DARPA](#).

## TSIMMIS Links

- TSIMMIS [publications](#)
- [People](#) in the TSIMMIS project
- [Developer's page](#) (restricted access)

## TSIMMIS Related Links

- [LORE](#), an OEM repository
- [I3 Initiative Projects Home Page](#)
- [DARPA Progress Reports](#)
- [Garlic](#), our sister project at IBM

## Demo And Source Code

An overview of [MOBIE](#) used for the demo.

- Run a [Stock mediator](#) demo
- Run a [Other sources\(weather source, bibliographic sources\)](#) demo
- [Download source code](#)

# ETD Digital Library

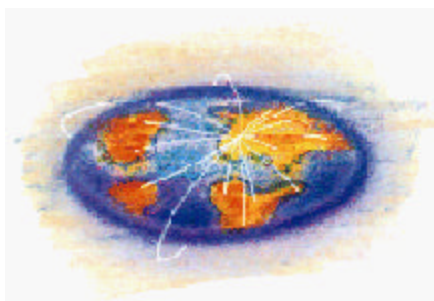
---

[About ETD Federated Search](#)

[Search or Browse the Catalog](#)

---

---



## CyberStacks(sm)

### Welcome To CyberStacks(sm)!

CyberStacks(sm) is a *centralized, integrated, and unified* collection of significant World Wide Web (WWW) and other Internet resources categorized using the Library of Congress classification scheme. Resources are organized under one or more relevant Library of Congress class numbers and an associated publication format and subject description. The majority of resources incorporated within its collection are monographic or serial works, files, databases or search services. All of the selected resources in *CyberStacks(sm)* are *full-text, hypertext, or, hypermedia*, and of a research or scholarly nature.

Using an abridged Library of Congress call number, *Cyberstacks(sm)* allows users to browse through a virtual library stacks to identify potentially relevant information resources. Resources are categorized:

- \* first within a broad classification,
- \* then within narrower subclasses,
- \* and then finally listed under a specific classification range and associated subject description that best characterize the content and coverage of the resource.

For each resource, a brief summary is provided, and when necessary, specific instructions on using the resource are also included. Where appropriate, the mode of access to the resource is noted, as is the subject coverage and scope; notable features, where applicable, are also included. At present, *CyberStacks(sm)* is a prototype demonstration service and is limited to significant WWW and other Internet resources in selected fields of Science and Technology.

A systematic effort is now underway to also identify and review resources that relate to the missions of the Center for Indigenous Knowledge for Agriculture and Rural Development (CIKARD) and the International Institute of Theoretical and Applied Physics (IITAP), two international research centers based at Iowa State University. While most of the current collection consists of Reference works, a number of scholarly journal Tables of Contents were recently added to its Title Index. Selected full-text serial titles and non-Reference monographic works, with subject coverage relevant to the interests of IITAP and CIKARD, have also been included.



# The Scorpion Project



[Scorpion](#) is a project of the [OCLC Office of Research](#) exploring the indexing and cataloging of electronic resources. Since subject information is key to advanced retrieval, browsing, and clustering, the primary focus of Scorpion is the building of tools for automatic subject recognition based on well known schemes like the [Dewey Decimal System](#).

---

## Scorpion Documentation

- [A brief introduction to Scorpion](#)
- [Evaluating Dewey Concepts](#)
- [Evaluating Scorpion Results](#)
- [Measures for Evaluating ...](#) **NEW**
- [Clustering](#) **NEW**
- [AMIGOS 97](#) (full image [version](#))
- [Scorpion helps catalog the Web](#)
- [Dewey Database Design](#)
- [ESS Field Label Descriptions](#)
- [Example ESS Record](#)
- [SMART Weighting Schemes](#)

## Automatic Subject Assignment

- [Change User Password](#)
- [Simple URL Input Form](#) **NEW**
- [Simple Text Input Form](#) **NEW**
- [Advanced Input Form](#) **NEW**

## Related Work

- [Online Classification: Implications for Classifying and Document\[-like Object\] Retrieval](#)
  - [Using Library Classification Schemes for Internet Resources](#)
  - [Dewey 2000](#)
  - [Cataloguing Rules and Conceptual Models](#)
  - [The Dublin Core](#)
  - [Prototype Dublin Core Metadata System](#)
  - [Electronic classification schemes](#)
  - Pharos ([demo](#), [publications](#))
- 



Comments/suggestions to [shafer@oclc.org](mailto:shafer@oclc.org)  
Scorpion [contributors](#)

 Dewey Home Search Site Map What's New Feedback**Content Areas**[About the DDC](#)[News](#)[DDC Updates](#)[DDC Users](#)[Products](#)[Research](#)[Dewey for Windows](#)

OCLC Forest Press  
**Dewey**  
Decimal Classification

**Popular Pages**

- [Dewey Screen Saver](#)
- [DDC Summaries](#)
- [LCSH/DDC](#)
- [New & Changed Entries](#)
- [DFW Guide](#)

 Search Dewey Web Site

For more advanced searching, select the Search icon in the toolbar above

Last updated [June 15, 1998](#). ©1997, 1998 OCLC  
[View Text Only](#)

**OCLC Forest Press**

6565 Frantz Road, Dublin, OH  
43017-3395, USA  
+1-614-764-6000  
[oclc@oclc.org](mailto:oclc@oclc.org)

**[OCLC Online Computer  
Library Center, Inc.](#)**

6565 Frantz Road, Dublin, OH  
43017-3395, USA  
+1-614-764-6000  
[oclc@oclc.org](mailto:oclc@oclc.org)

For more contact information, see [Contacts & Addresses](#).

## MAGAZINE

# Finding Images/Video in Large Archives

## Columbia's Content-Based Visual Query Project

Shih-Fu Chang, John R. Smith  
Horace J. Meng, Hualu Wang, and Di Zhong  
Department of Electrical Engineering and  
Center for Telecommunications Research  
Columbia University

*{sfchang,jrsmith,jmeng,hwang,dzhong}@ctr.columbia.edu*

**D-Lib Magazine**, February 1997

ISSN 1082-9873

---

### Table of Contents

- [An Application Driven Problem](#)
  - [State of the Art](#)
  - [Research Strategies](#)
  - [Prototype Systems](#)
  - [Testbed Support and User Evaluation](#)
  - [Open Issues](#)
  - [References](#)
- 

## An Application Driven Problem

How do we find a photograph from a large archive which contains thousands or millions of pictures? How does a CNN video journalist find a specific clip from the myriad of video tapes, ranging from historical to contemporary, from sports to humanities? How do people organize and search the content of personal video tapes of family events, travel scenes, or social gatherings?

The era of "the information explosion" has brought about the wide dissemination and use of visual information, particularly, digital images and video, which we are also seeing in combination with text, audio, and graphics. The development of tools and systems that enhance image functionalities, such as searching and authoring, is critical to the effective use of visual information in the new media applications.

The current research and development of images and video search tools is driven by practical applications. We are seeing the establishment of large digital image and video archives, such as the Corbis catalog, which includes the Bettman Archive; the Picture Exchange, which is a joint venture between Kodak and Sprint; and many digital video libraries in various domains (e.g., environment,



at Columbia University

# A Content-Based Image and Video Search and Catalog Tool for the Web

( press here to Browse all subjects )

## Animals

birds, dinosaurs,  
monkeys, fishes

## Cats

leopards, lions,  
kittens, cheetahs

## Horror

godzilla, aliens,  
skeletons, monsters

## Nature

sunsets, flowers,  
weather, mountains

## Architecture

bridges, lighting, domes  
heating

## Celebrities

bullock, aniston, monroe,  
keanu

## Humour

simpsons, beavis, dilbert,  
ren/stimpy

## Sports

baseball, basketball,  
swimming, hockey,  
olympics, surfing

## Art

painting, illustr, sketching  
cezanne, monet, vangogh

## Dogs

bulldogs, puppies, coyotes,  
wolves

## Movies

batman, starwars, jurassic,  
python, blade runner, actresses

## Transportation

cars, planes, titanic,  
motorcycles, porsches

## Astronomy

nasa, planets, eclipses,  
space

## Food

apples, beer, pizza, cakes,  
fruits, veges

## Music

beatles, metal, rock, cure,  
zeppelin, guitars

## Travel

asia, europe, newyork,  
paris, australia, mexico

Image/Video Topic

(single word)

Search

☒ all    ☐ videos    ☐ color photos    ☐ gray images    ☐ graphics

[\[WebSEEk\]](#)

[\[browse\]](#)

[\[add urls\]](#)

[\[postcards\]](#)

[\[info\]](#)

[\[credits\]](#)

WebSEEk has catalogued **665115** images and videos

---

All licensing inquiries may be directed to [Dr. Joseph R. Flicek](#) of the Columbia Innovative Enterprise.

# BioInformatics Centre

## Research Unit

### Hosted at [Kent Ridge Digital Labs](#)

**BioKleisli** is a collaboration with Kent Ridge Digital Labs (previously known as the Institute of Systems Science) and the University of Pennsylvania. It is a tool for the broad scale integration of databanks. It offers flexible access to biological information sources that are highly heterogeneous, geographically scattered, highly complex, constantly evolving and high in volume.

---

- BioKleisli offers high-level flexible access to human genome and other molecular biological sources that are
    - Highly heterogeneous
    - Geographically scattered
    - Highly complex
    - Constantly evolving
    - High in volume
  - Flexible Access = Migrate + Integrate + Restructure
  - Deployed at the HGP Philadelphia Center for Chromosome 22 for
    - Integration of private and public data banks
    - Integration of data banks, analysis software, and visualization tools
- Major benefit: Typical query implementation time is reduced from weeks to days (some times, hours).**
- Why is BioKleisli Special?
    - Self-describing data model for complex structured data. Beyond the reach of relational DBs.
    - High-level query language for data transformation. ``Internet SQL" as opposed to simple navigational IR.
    - Flexible yet precise control in ad-hoc queries.
- 

For a recent poster on BioKleisli, click [here](#).

For more information on the underlying technology, click [here](#).

For some examples queries powered by BioKleisli, click [here](#).

For list of some bioinformatics sources BioKleisli talks to, click [here](#).

For the history of the project, click [here](#).

# BioInformatics Centre

## Research Unit

### Hosted at [Kent Ridge Digital Labs](#)

Here some links to see demonstrations of some of our work. (Some machines are only accessible from within KRDL.)

---

These are demo machines. Not for heavy duty use. If you want to do large-scale queries, please contact us. (We kill external processes left-right-and-center when they interfere with our daily work.)

- **BioKleisli** ([@adenine](#) [@cytosine](#) [@guanine](#) [@thymine](#) [@uracil](#) [@alanine](#) [@arginine](#))  
A collaboration with Kent Ridge Digital Labs and the University of Pennsylvania.
- **SeqIndex** ([@adenine](#) [@cytosine](#) [@guanine](#) [@thymine](#) [@uracil](#) [@alanine](#) [@arginine](#))  
A collaboration with Kent Ridge Digital Labs.
- **ViewBLAST** ([@adenine](#) [@cytosine](#) [@guanine](#) [@thymine](#) [@uracil](#) [@alanine](#) [@arginine](#))  
A collaboration with Kent Ridge Digital Labs and Institute of Molecular and Cell Biology.
- **Web PHYLP** ([@adenine](#) [@cytosine](#) [@thymine](#) [@uracil](#) [@alanine](#) [@arginine](#))  
A collaboration with Kent Ridge Digital Labs.
- **DNA Chip** ([@adenine](#) [@cytosine](#) [@thymine](#) [@uracil](#) [@alanine](#) [@arginine](#))

---

[Limsoon Wong](#) / BioInformatics Center and [Kent Ridge Digital Labs](#), 21 Heng Mui Keng Terrace, Singapore 119613 / [Limsoon@Saul.CIS.UPenn.EDU](mailto:Limsoon@Saul.CIS.UPenn.EDU), [Limsoon@KRDL.Org.SG](mailto:Limsoon@KRDL.Org.SG)

# Information Filtering

## U. Md. Information Filtering

- [Defn](#)
- Fast Data Finder: [Genetic sequence analysis](#)

### Questions:

- What is *information filtering*? How does it differ from information retrieval?

# Information Filtering Resources

This page is designed as a resource for people conducting research in [information filtering](#). It was developed as part of the [Information Filtering Project](#) at the University of Maryland, and is maintained as a part of my ongoing [research program](#). The first section may be of interest to a wider audience, though, since it contains links to working systems for a variety of operating systems which are freely available on the net, and the third section should become more valuable to a wide audience as links to more commercial systems are added.

This page lists all known internet-accessible information filtering resources. If you are aware of resources which do not appear here, please [send mail to Doug Oard](#). Since it's impossible to look at everything here, I'm often asked where to start. I've tried to answer that in my recent [paper](#). Although I know of no comprehensive bibliography of papers, theses and dissertations on information filtering which exist only in printed form, scanning the bibliographies in the first two sections will reveal almost every source that I am aware of. There is also a good start at a comprehensive bibliography at the University of North Carolina in the second section.

## [Freely Available Information Filtering Systems](#)

Working information filtering systems which are publicly available. In most cases, papers describing the theory and/or implementation details are also available.

## [Commercial Information Filtering Systems](#)

An incomplete sample of commercial systems that might be of interest to researchers in the field.

## [Information Filtering Papers and Project Descriptions](#)

Descriptions of experimental or proprietary systems for which the software is not being distributed. If you're looking for papers you should also check the previous section because research results included there are not repeated in this section.

## [Other People Interested in Information Filtering](#)

Home pages of people who have either published work on information filtering that is not presently available on the net or who are actively working in the field. This is not a good place to look for papers, but a great way to learn how to contact people who don't appear in either of the previous two sections.

## [Related Resource Pages](#)

Web pages which collect links to resources that may be of interest to information filtering researchers.

---

Last modified: Wed Jan 28 20:10:19 1998

[Doug Oard](#) [oard@glue.umd.edu](mailto:oard@glue.umd.edu)

# Information Filtering Defined

A universally accepted definition of information filtering is, unfortunately, still lacking. So here is my personal definition, which I have used to build the Information Filtering Resources [web page](#). Generally, the goal of an information filtering system is to sort through large volumes of dynamically generated information and present to the user those which are likely to satisfy his or her information requirement.

In order to sharpen this definition, a distinction should be drawn between information collection and information filtering. In some domains (e.g. USENET News) the collection effort is minimal because the information comes to you. In other domains (e.g. the World Wide Web) the collection effort can be considerable because no mechanism exists to draw new information to the attention of a filtering system. The point to be made here, though, is that information collection is an interesting area in its own right, but I do not propose to include it in my definition of information filtering. In my view, the information filtering problem begins only after you have gained access to the new information.

Information filtering has been applied to a several domains using a variety of technical approaches. The original methods were manual alerting services that brought new information to the attention of users of research and special libraries. At the time this was referred to as Selective Dissemination of Information (SDI), a name which fell from favor about the time the Strategic Defense Initiative (SDI) was introduced in the United States :-). A few modern systems have adopted this remarkably descriptive name for the filtering process, however, and the interest in information filtering that has resulted from the present research thrusts in digital libraries arises at least in part from this tradition.

With the growth of the internet and other networked information, research in automatic filtering of networked information has exploded in recent years. Because of their low cost, large volume, and ease of recognizing new information, the most popular domains for research systems have been USENET News and electronic mail. The recent explosive growth of the World Wide Web has made this an interesting domain which has attracted some good research, although the information collection problem appears to make this a more difficult domain in which to conduct basic research on information filtering techniques. Another domain which has attracted considerable research interest is the annual Text REtrieval Conference (TREC) in which a standard text collection is used and a carefully controlled evaluation methodology is enforced. In TREC the information filtering task is referred to as "routing," adding somewhat to the confusion of terminology in this field. In fact, TREC recently adopted a special interest "filtering" track which adopts a different evaluation methodology but which conforms to the definition of filtering presented above. Commercial systems which filter newswire articles and other specialized information sources are becoming available as well. Filtering techniques will likely be applied to other domains such as images, sound and video in the future.

The distinction between information filtering and the more established field of information retrieval has proven to be the source of some confusion as well. Information retrieval broadly deals with the selection of information, and many of the features of information retrieval system design (e.g. representation, similarity measures or boolean selection, document space visualization) are present in information filtering systems as well. If one considers information retrieval from a very general "information selection" viewpoint, information filtering is simply a special case in which the information space is very dynamic. If, on the other hand, your personal definition of information retrieval involves selection of relatively static information in response to relatively dynamic queries, then information filtering is best viewed as the dual problem to information retrieval. Regardless of which viewpoint you take, though, it is clear that researchers in information filtering will likely benefit from familiarity with the legacy of

# Cross-Language Information Retrieval Resources

This page is designed as a resource for people conducting research in [cross-language information retrieval](#). It is intended to collect references to all information on information retrieval systems which can accept queries in one language and return documents in another. It is maintained by the [Digital Library Research Group](#) of the [College of Library and Information Services](#) at the University of Maryland. If you are aware of resources that are within the scope of this page but do not appear here, please [send mail to Doug Oard](#).

## [December 1997 D-lib Magazine Article](#)

A recently written introduction to cross-language information retrieval.

## [Conferences](#)

The best single source for information in the field. This page includes links to the full proceedings of every major cross-language information retrieval workshop as well as to a fairly complete list of upcoming conferences and workshops that include some treatment of cross-language information retrieval.

## [Cross-Language Information Retrieval Papers and Project Descriptions](#)

Another excellent place to look for information. Here you will find descriptions of experimental work on cross-language text retrieval that may not have been presented at one of the major workshops

## [Working Systems](#)

Here you will find links to experimental and commercial cross-language information retrieval systems that you can either obtain or use over the net. Some carry a fairly hefty price tag, others are free.

## [Bibliography](#)

A fairly comprehensive bibliography of published work on cross-language information retrieval in BibTeX form, last updated on July 3, 1997. The bibliography is also available in [postscript](#). Most of the references are described in at least one of my survey [papers](#) on cross-language information retrieval.

## [Related Resource Pages](#)

Web pages which collect links to resources that may be of interest to cross-language information retrieval researchers. None of these pages are devoted solely to cross-language information retrieval.

---

Last modified: Wed Dec 24 00:49:55 1997

[Doug Oard](#) [oard@glue.umd.edu](mailto:oard@glue.umd.edu)



## Information Finding Projects in the Stanford Digital Library

---

One of the major research thrusts of the Stanford Digital Library project is helping users to find information. We have initiated a number of projects in this area, most related to our over-arching theme of interoperability. We have looked at ways that search tools can be used across multiple sources that use different syntaxes or languages. We have also looked at tools to provide statistical or collaborative filtering to locate relevant articles.

---

### [FAB](#)

FAB is an adaptive multi-agent information retrieval system which finds interesting pages on the web.

"[An Adaptive Agent for Automated Web Browsing](#)"

- [Marko Balabanovic](#)

---

### [GLOSS](#)

The Glossary Server of Servers (GLOSS) project is designed to locate relevant information sources for your query.

"[Generalizing GLOSS to Vector-Space Databases and Broker Hierarchies](#)"

- [Luis Gravano](#)

---

### [Query Translator](#)

Databases have different query syntax and different capabilities, even for simple Boolean queries. Translation allows a single query to be mapped into the native format appropriate for each database.

- [Chen-Chuan K. Chang](#)

---

### [SenseMaker](#)

SenseMaker helps users iteratively reformulate their information needs through multi-dimensional organizing and active gathering of search results.

"[SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests](#)"

# Multimedia, Representations:

---

## The Basics:

- [text file formats](#)
- [graphic file formats](#)
- [hypermedia & multimedia](#)

ACM DL'97 Tutorial: [Multimedia Information and Systems](#)

## Digital Video

- [An iconic visual language for video annotation](#) (Table of Contents. Segmented Annotation: hierarchical annotation structure on top of physical video stream)
- [Jacob: GUI iteratively refine query till satisfying result](#) (Note: Query specification can be direct, by example, or iteratively.)
- [CNN uses Quicktime for WWW daily news clips](#)

## MHIA Courseware and Curricula

- [MHIA Home Page](#)
- [SIGIR 96 Workshop](#)
- [Drexel 96 Workshop](#)
- [IR Courses](#)
- [Information Engineering - European Commission: Work Program, Pilots](#)
- [MM 96 Workshop](#)
- [Lisbon Workshop](#)
- Questions:
  - What is the need for education related to information? What jobs?
  - What subjects should be covered in such education programs?
  - How should those subjects be ordered into each specific program?

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

# ACM DL'97 Tutorial: Multimedia Information and Systems

Edward A. Fox

---

## Content

- [Outline with References: Image and Video Processing, Retrieval](#)
- [CS4624: Multimedia, Hypertext, Information Access](#)
- **CS4624 highlights**
  - [Outline](#)
  - [Lab Sessions](#)
  - [Lecture Notes](#)
  - [Trips and Special Events](#)
  - [Computers and Tools](#)
  - [Figures](#)
  - [Figures with Captions](#)
  - **Glossary:** [Local](#), [from IMA](#)
  - [Index](#)
  - [Link Sets](#)
  - [Readings and References](#)
  - Syllabus
    - [Calendar](#)
    - [Department and Class Policies](#)
    - [Instructor and GTA](#)
    - [Syllabus Details: Grading, etc.](#)

## References

- [Links for ACM DL'97 Tutorial](#)
- [Printed References on Image and Video Processing, Retrieval](#)
- [References from NSF EI Project](#)

## [PDFs](#)

---

Copyright 1997 Edward A. Fox, all rights reserved

# MHIA: Multimedia, Hypertext and Information Access Curriculum & Courseware

Welcome to WWW pages for a multi-institution, multi-association effort in the broad area of Multimedia, Hypertext and Information Access to develop curriculum and courseware, and related materials, services, and initiatives to promote education and training in this important field.

For an idea of the potential benefits and required efforts, see our [pre-proposal](#) or the [excerpt](#) of key parts from the 10/8/96 full proposal.

For more information, or if you would like to offer to help, send mail to [fox@vt.edu](mailto:fox@vt.edu) to reach [Ed Fox](#).

## Important Events:

- *Courseware, Education and Curriculum in Information Retrieval*  
Aug. 22, 1996 - [workshop](#), part of the [workshop program](#) at [ACM SIGIR'96](#).
- *Courseware, Education and Curriculum in Multimedia*  
Nov. 19, 1996, Room 301 - [ACM MM'96 Workshop](#) and [Conf. Info.](#)  
For discussion, send email to [mm96wk@fox.cs.vt.edu](mailto:mm96wk@fox.cs.vt.edu)
- Workshop for ACM DL'97 and SIGIR'97
- Proposals to NSF:
  1. Multimedia Curricula and Repositories
  2. Computer Science Lab Repository and Registry
- National Academy of Science / NSF Workshop Aug. 1997 on National Digital Library for SME&T
- [Lisbon Workshop](#)

## References:

- E. Fox and L. Kieffer. *Multimedia Curricula, Courses and Knowledge Modules*, *ACM Computing Surveys*, Dec. 1995, 27(4): 549-551.
- [ACM SIGIR WWW pages](#)
- [Multimedia Educational Materials](#) from [ACM SIGMM](#)'s education committee (thanks to Brian Smith at Cornell!)
- E. Fox, [presentation](#) for: *Information Retrieval 2000 --- Workplace Needs and Curricular Implications*  
*A Workshop/Symposium sponsored by the W.K. Kellogg Foundation*  
*Hosted by Drexel University*  
*May 24, 1996*  
*Marriott Hotel, Philadelphia PA*

# Architectures:

---

Core topics include:

- [D-Lib article on architecture](#)
- [Other CNRI activities](#)
- **Naming**
  - [PURL](#)
  - [Handles](#)
- [Networks](#): online notes of Dr. Lesk

Other topics of general interest, that are being studied by the [D-Lib Metrics Group](#) include:

- **Distributed processing (client/server)**
- **Interoperability** (see [IITA workshop on Interoperability ...](#))
- **Performance**

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta




---

## Handles and Web Browsers

This note describes how to use a World Wide Web browser to access items identified by handles and other Uniform Resource Names (URNs).

### Adding Handle Resolution to a Browser

The  [CNRI Handle System Resolver](#) is a web browser extension, that enables Netscape and Internet Explorer (version 3+) for the Windows and Macintosh platforms, to recognize the handle protocol and communicate directly with the Handle System to resolve handles into their associated URLs.

### Resolving Handles Using a Proxy Server

Without the extension, web browsers must be directed to a Proxy Server, which understands the handle protocol, in order to resolve the handle to a URL. CNRI maintains two public proxy servers addressed by "hdl.handle.net" and "dx.doi.org", and one private proxy server, "hdl.loc.gov". Embedding a handle in a URL that uses one of the proxy URLs will permit any browser to resolve that handle.

For example, the February 1998 issue of *D-Lib Magazine* has an article by William Y. Arms with the handle (URN) "cnri.dlib/february98-arms". Currently the article is stored on three web servers, one maintained by CNRI, the others a mirror site at UKOLN in England and a mirror site at the Australian National University Sunsite. The story is formatted in html and accessible using the http protocol. The handle is the permanent name of the article. It will not be changed if *D-Lib Magazine* moves to different computers, the story is moved from a web server to some other type of storage, the magazine changes publisher, or even if CNRI and D-Lib change name or disappear. For these reasons, the best long-term way to cite the article is by the handle.

To read this article from a web browser which does not have the CNRI Handle Resolver extension, direct the browser to open the following URL, which combines a proxy server address with the handle "cnri.dlib/february98-arms", as shown in the example below:

<http://hdl.handle.net/cnri.dlib/february98-arms>


The article will then be displayed by the browser. The browser's "Address" or "Location" panel shows the location from which the article was read.

### Handle Resolution Using a Form

You may also use the web form at

 <http://www.handle.net/docs/gethdl.html>

to see what URL a handle resolves to.

The  [Digital Object Identifier System](#) web site uses a proxy server to resolve Digital Object Identifiers (DOIs) into URLs.

---

**NOTE:** To demonstrate the Handle System, the hyperlinks on this site use handles. The small red arrows accompanying the text are hyperlinks containing handles; these require the [CNRI Handle System Resolver](#) to be installed on your computer. Underlined text contains handles embedded in URLs that are resolved using a [proxy server](#). These work with any browser. Links to sites not controlled by CNRI are conventional URLs.

---

[ [home](#) | [introduction](#) | [documentation](#) | [download](#) | [administration](#) | [CNRI home page](#) ]

Updated: 11 May 98



---

## D-Lib Working Group on Digital Library Metrics

This Working Group is aimed at developing a consensus on an appropriate set of metrics to evaluate and compare the effectiveness of digital libraries and component technologies in a distributed environment. Initial emphasis will be on (a) information discovery with a human in the loop, and (b) retrieval in a heterogeneous world.

[Working Group Charter](#)  
[Other Working Group Documents](#)  
[Working Group Private Area](#)

This is an open working group, and anyone interested in the subject and in contributing to the work of the group is encouraged to join. For further information or to join the group, contact Barry Leiner <[BLEiner@cnri.reston.va.us](mailto:BLEiner@cnri.reston.va.us)>.

The next meeting of the Working Group will be held the morning of 24 June 1998, just prior to the [DL'98 conference](#). The meeting will be open but those planning on attending are asked to contact [Jeanette Bartolomeo](#) to assure there will be sufficient space.

### [Meeting Information](#)

The Working Group is also sponsoring a [Workshop on Digital Library Metrics](#), to be held 27 June 1998, just after the [DL'98 conference](#). Interested parties should contact the workshop organizers: [Bill Pottenger](#) and [Bob McGrath](#).

---

*prepared by [Barry Leiner](#)  
last modified 5/26/98*

# Interfaces:

---

## [Stanford DL user interface projects](#)

### **Xerox Interfaces for Information Access**

- [Home Page](#)
- [Scientific American article](#)
- [Cat-a-Cone figures](#)
- [Scatter/Gather examples](#)
- Questions:
  - Compare
    - What are the various interfaces built? How do they compare? What is the best use of each?
  - Scatter/gather
    - Explain clustering, relate it to scatter/gather.
    - What are special problems with large category systems and how can they be solved?

[Envision](#) : ENVISION project at Virginia Tech ...

[Berkeley:](#) TileBars, Multivalent documents

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

# Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy

Marti A. Hearst

Xerox PARC

3333 Coyote Hill Rd

Palo Alto, CA 94304

[hears@parc.xerox.com](mailto:hears@parc.xerox.com)

Chandu Karadi

School of Medicine, M121

Stanford University

Stanford, CA 94305

[karadi@leland.stanford.edu](mailto:karadi@leland.stanford.edu)

**This paper appears in the proceedings of 20th Annual International ACM/SIGIR Conference, Philadelphia, PA, July 1997.**

An unpublished [Appendix](#) to this paper contains additional figures which were omitted from the proceedings due to space limitations.

## Abstract:

This paper introduces a novel user interface that integrates search and browsing of very large category hierarchies with their associated text collections. A key component is the separate but simultaneous display of the representations of the categories and the retrieved documents. Another key component is the display of *multiple* selected categories simultaneously, complete with their hierarchical context. The prototype implementation uses animation and a three-dimensional graphical workspace to accommodate the category hierarchy and to store intermediate search results. Query specification in this 3D environment is accomplished via a novel method for painting Boolean queries over a combination of category labels and free text. Examples are shown on a collection of medical text.

In the *Proceedings of the Twentieth Annual International ACM SIGIR Conference*, Philadelphia, PA, July 1997, to appear.

© Copyright 1997 by ACM, Inc.

## INTRODUCTION

There exist today many large online text collections to which category labels have been assigned. MEDLINE, a huge collection of biomedical articles, has associated with it Medical Subject Headings (MeSH) consisting of approximately 16,000 categories [ [Lowe & Barnett1994](#)]. The Association for Computing Machinery (ACM) has developed a hierarchy of approximately 1200 category (keyword) labels ( <http://www.acm.org/class/1991/cr91.html>). Yahoo!, one of the most popular search sites on the World Wide Web, organizes web pages into a hierarchy consisting of thousands of category labels ( <http://www.yahoo.com>). And traditional online bibliographic systems have for decades assigned subject

# Metadata:

---

- [IMS Metadata spec.](#) and [tool](#)
- [Metadata: the Foundations of Resource Description](#)
- [OCLC/NCSA Metadata Workshop Report](#)
- [RFC-1807](#)
- [TEI](#)
- [BASIS article](#)
- [D-Lib Working Group on Metadata](#)
- [STARTS](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998. Edward A. Fox, Rajat Gupta**




The Instructional Management Systems Project, an Educom NLII initiative, is developing a specification and software for managing online learning resources. Learning resources can include people, educational service companies, content, tools, and activities.

The pages listed below will introduce the concept of meta-data and the specification for its use. Visitors to this site can explore the potential of meta-data by experimenting with the [Meta-data Tool](#). To join the dialogue regarding the development of the IMS Meta-data Specification, please subscribe to the [IMS Meta-data Listserv](#).



 **EXECUTIVE SUMMARY** *A brief description of the information and the organization of this site*

 **INTRODUCTION** *An overview of meta-data: its purpose and development*

- What is Meta-data?
- How is Meta-data Organized?
- How is Meta-data Used?
- Evolution of Meta-data

 **USING META-DATA** *A description of using IMS meta-data for learning modules*

- Searching
- Creating

 **MANAGING IMS META-DATA** *A description of how to manage and enhance meta-data*


 **IMS META-DATA SPECIFICATION** *Technical documents defining meta-data fields and values and their application to different learning resources*

- IMS Meta-data Dictionary
- IMS Meta-data Sets

 **CREDITS AND REFERENCES**

 **META-DATA PRESS RELEASE AND QUOTES**

- Sept. 8, 1997
- March 23, 1998

 **FORUM** *A space for dialogue about IMS meta-data fields and values. This is a section of the larger IMS Public Forum.*

 **IMS Meta-data Tool** *Access the IMS Content Server and Meta-data Tool for experimenting with IMS meta-data. This is a new*

# Metadata: The Foundations of Resource Description

Stuart Weibel

Office of Research, OCLC Online Computer Library Center, Inc.

*weibel@oclc.org*

**D-Lib Magazine, July 1995**

---

This paper is an abbreviated version of the [Summary Report of the OCLC/NCSA Metadata Workshop](#). It sets forth a proposal for the content of a simple resource description record (the Dublin Core Metadata Element Set) and outlines a series of further steps to advance the standards for the description of networked information resources.

- [Introduction](#)
  - [Underlying Assumptions](#)
  - [Implementations](#)
  - [Next Steps](#)
  - [References](#)
- 

**d-lib forum**

**d-lib magazine**

---

## Introduction

The explosive growth of interest in the Internet in recent years has created a digital extension of the academic research library for certain kinds of materials. Valuable collections of texts, images and sounds from many scholarly communities -- collections that may even be the subject of state-of-the-art discussions in these communities--now exist only in electronic form and may be accessible from the Internet. Knowledge regarding the whereabouts and status of this material is often passed on by word of mouth among members of a given community. For outsiders, however, much of this material is so difficult to locate that it is effectively unavailable.

Why is it so difficult to find items of interest on the Internet or the World Wide Web? A number of well-designed locator services, such as Lycos [\[http://lycos.cs.cmu.edu/\]](http://lycos.cs.cmu.edu/), are now available that automatically index many of the resources available on the Web and maintain up-to-date databases of locations. But indexes are most useful in small collections within a given domain. As the scope of their coverage expands, indexes succumb to problems of large retrieval sets and problems of cross disciplinary semantic drift. Richer records, created by content experts, are necessary to improve search

# OCLC/NCSA Metadata Workshop Report

Stuart Weibel, Jean Godby, Eric Miller  
*Office of Research, OCLC Online Computer Library Center, Inc.*  
Ron Daniel  
*Advanced Computing Lab, Los Alamos National Laboratory*

## 1.0 Executive Summary

The March 1995 Metadata Workshop, sponsored by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), convened 52 selected researchers and professionals from librarianship, computer science, text encoding, and related areas, to advance the state of the art in the development of resource description (or metadata) records for networked electronic information objects.

### 1.1 Goals

Goals of the workshop included (1) fostering a common understanding of the needs, strengths, shortcomings, and solutions of the stakeholders; and (2) reaching consensus on a core set of metadata elements to describe networked resources.

### 1.2 Scope

The size and complexity of the resource description problem required limiting the scope of deliberations. Given that the majority of current networked information objects are recognizably "documents", and that the metadata records are immediately needed to facilitate resource discovery on the Internet, the proposed set of metadata elements (The Dublin Core) is intended to describe the essential features of electronic documents that support resource discovery. Other important metadata elements, such as those describing cost accounting or archiving information, were excluded from consideration. It was recognized that these elements might be included in a more complete record that could be derived from the Dublin Core by a well-defined extension.

### 1.3 The Intended Niche

The Dublin Core is not intended to supplant other resource descriptions, but rather to complement them. There are currently two types of resource descriptions for networked electronic documents: automatically generated indexes used by locator services such as Lycos and WebCrawler; and cataloging records, such as MARC, created by professional information providers. Automatically generated records often contain too little information to be useful, while manually generated records are too costly to create and maintain for the large number of electronic documents currently available on the Internet. Records created from the Dublin Core are intended to mediate these extremes, affording a simple structured record that may be enhanced or mapped to more complex records as called for, either by direct extension or by a link to a more elaborate record.

### 1.4 Next Steps

The work of the 1995 workshop is one of a series of steps being taken to improve the description of

# Notes on Metadata and the Web

For an overview paper on related areas, read about the [Warwick Framework](#), a container architecture for aggregating metadata.

These notes are based on the articles that appear in the Oct./Nov. 1997 issue (v. 24 no. 1) of the *Bulletin of the American Society for Information Science* (ASIS). The issue title is *Organizing Internet Resources: Metadata and the Web*.

Some of the key topics considered are:

- Dublin Core, its evolution, its adaptations
- Cataloging, MARC, and their extension to Internet
- Automatic classification: Scorpion
- Naming: URL, URN, URI, URC, DOI

## Useful Links by Topic - Alphabetical

The following links are either taken from the articles in the *Bulletin* issue or relate closely and fill in helpful information.

- [InterCat Catalog](#)- proof-of-concept database, made of records extracted from OCLC's WorldCat, demonstrating catalog services plus Web access to resources of the Internet
- [International Conf. on Principles and Future Development of AACR](#)- related papers, on Anglo-American Cataloging Rules, and their revision
- [Persistent URLs](#)- PURLs
- [Dublin Core Home Page](#)
- [Dublin Core Elements](#)
- [Dublin Core element Coverage](#) - proposed standard
- [Center for Electronic Text in the Humanities](#)
- [EAD \(Encoded Archival Description\): SGML for Archival Finding Aids](#)
- [UC Berkeley Finding Aids](#)
- [Cataloging Internet Resources: Manual and Practical Guide, by Nancy B. Olson](#)
- OCLC and its [Research Department](#)
- [Stuart Weibel](#)- senior research scientist at OCLC, leader of Dublin Core efforts
- [Workshop on Metadata for Networked Images](#)
- [RDF Home Page](#)- Resource Description Framework, on metadata architecture on the Web
- [UKOLN Metadata Home Page](#)- summary of pubs, projects, metadata resources from UK and beyond, definitions
- [metadata element sets crosswalks](#)- mappings and relationships between various metadata sets, including Dublin Core
- [Resource Discovery project in Australia](#)
- [Dublin Core Workshop, 4th, official report](#) - held at National Library of Australia - and a [light-hearted account](#)
- [National Library of Australia PANDORA Project](#) (Preserving and Accessing Networked Documentary Resources of Australia)
- [In the Company of Strangers: Challenges and Opportunities in Metadata Implementation](#) paper by Maxine Brodie, policy level issues which impact on metadata implementation at the State Library of New South Wales, Sydney, Australia; also [Implications for Metadata Implementation](#)

- [Architecture for Access to Government Information](#) : report, Australia, 1996
- [ERIN - Environmental Resources Information Network](#), Australia - also runs a metadata listserv
- [Core Data Elements for Land and Geographic Directories in Australia and New Zealand](#)
- [Dataset Publishing - A Means to Motivate Metadata Entry, by S.D. Callahan, B.D. Johnson, and E.P. Shelley](#) - Australian Resources, NPI Theory (choice behavior)
- [meta-searcher called HotOIL that accesses both HTTP and Z39.50 servers - demo](#) - translates user requests, merges results, displays summary
- [MetaWeb project](#) - develop and disseminate metadata tools
- [GEM](#) - educational resources - which calls for adding elements like Resource Needed, Standard, Audience, Pedagogy, Quality - see [elements](#)
- [NetFirst](#) - database/directory, cataloging of Internet (uses Dewey)
- [Canadian Information by Subject](#) - info on Canada in Internet (uses Dewey)
- [BUBL Information Service, Scotland, higher education, with subject tree](#) (uses Dewey)
- [Internet Public Library Youth Division](#) (uses Dewey)
- [Blue Web'n, by Pacific Bell, to organize Web sites for students, educators, ...](#) (uses Dewey)
- [Enhancing the indexing vocabulary of DDC by C.J. Godby](#)
- [Scorpion project at OCLC](#)

### Acknowledgements

Thanks are given to the authors of the respective articles, from whose contributions the notes above are derived. All distortions of their content and intention are the fault of E. Fox, who apologizes for any misrepresentation inadvertently resulting from this attempt to summarize a valuable set of interesting articles.

- Guest editors' intro. to Special Section, by Efthimis N. Efthimiadis and Allyson Carlyle
- Cataloging Internet Resources: Survey and Prospectus, by Erik Jul
- The Dublin Core: A Simple Content Description Model for Electronic Resources, by Stuart Weibel
- Uniform Resource Identifiers and the Effort to Bring "Bibliographic" Control" to the Web: An Overview of Current Progress, by Ray Schwartz
- Options for Organizing Electronic Resources: The Coexistence of Metadata, by Sherry L. Vellucci
- Metadata in Australia, by Carmel Maguire
- GEM: Using Metadata to Enhance Internet Retrieval by K-12 Teachers, by Stuart Sutton and Sam G. Oh
- From Book Classification to Knowledge Organization: Improving Internet Resource Description and Discovery, by Diane Vizine-Goetz
- Scorpion Helps Catalog the Web, by Keith Shafer

---

Please follow the above mentioned links to find answers to the following questions:

- What is metadata?
- How many elements are in the Dublin Core?
- What are some new elements added for educators in GEM?
- Describe TEI briefly and explain how it relates to Dublin Core work.
- Explain *finding aid*.
- Describe EAD briefly and explain how it relates to cataloging archival collections.
- Where are their detailed instructions on how to catalog the internet?

- What is RDF?
- What is happening in UK re metadata?
- What mappings are there between metadata representations?
- What is the Resource Discovery project in Australia?
- What happened at the Australian metadata meeting?
- What is covered by the Dublin Core *coverage* element?
- What metadata is needed for geographic information?
- When you search on "digital library" with HotOIL, what refinements are suggested? What are the results of the default processing of your query and what sources were used? Can you find the abstract of a talk on archiving the Internet?
- What WWW search/browse services use Dewey?
- What systems are available to automatically catalog WWW pages?

## Electronic Publishing:

---

- [The SGML/XML Web Page](#)
- [CS5604 unit on SGML](#): check out the related course notes offered at Virginia Tech.
- [Elsevier](#)  
[TULIP](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

## Database Issues:

---

- [UCB database management](#)

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

**(c) Copyright 1998, Edward A. Fox, Rajat Gupta**

[\[Berkeley\]](#) [\[EECS Dept.\]](#) [\[CS Division\]](#) [\[Database Research\]](#)



# Database Research Group papers

Most technical reports for which we have [PostScript](#) and [PDF](#) are also available from the [Berkeley CS-TR server](#) as TIFF images and/or OCR'd text.

Items marked *image PDF* are copyrighted by the Institute of Electrical and Electronic Engineers ([IEEE](#)) or the Association for Computing Machinery ([ACM](#)) and are restricted to use within the University of California. They are here solely as a convenience to local users, for use when [MELVYL](#) is unavailable. If you're not a UC user, don't even try asking for access.

Items marked *Word PS* were originally generated using Microsoft Word. You may have difficulty previewing the resulting PostScript files; if so, try using the regular PS files or the PDF files. (We make the Word PS originals available because they may be better for hardcopy.)

- [UCB CSD Technical Reports](#)
- [UCB ERL Technical Reports](#)
- [LBL Technical Reports](#)
- [Sequoia 2000 Technical Reports](#)
- [UCB Graduate Theses/Reports](#)
- [Wisconsin Technical Reports](#)
- [Other Published Papers](#)
- [Related Papers from Other Groups on Campus](#)

We also have a [Mike Stonebraker bibliography](#). This is a work in progress; additional contributions are [welcome](#), especially for papers and technical reports from the early 1970s.

## COPYRIGHT NOTICE:

The documents in these directories have been submitted by their authors to scholarly technical reports series whose purpose is the non-commercial dissemination of scientific work. They are put on-line to facilitate this purpose. These reports are copyrighted by the authors, and their existence in electronic format does not imply that the authors have relinquished any rights. You may copy a report provided that you agree to respect the author's copyright. In particular, a report may be copied for scholarly, non-commercial purposes, such as research or instruction. Reports may not be excerpted unless due acknowledgement is given the author.

Moreover, we do not know what additional arrangements authors may have made concerning these reports. Therefore, in copying a report, you are assuming whatever legal responsibilities copying any document might entail.

# Ontologies and Agents in Digital Libraries

Key topics about *Ontology* adapted from *AI Magazine*, Fall 1997, 18(3), include:

- Defn
- Comparison criteria
- Top level categories, taxonomy. categories, realtions, axioms
- Comparison chart

URLs related include:

- [Ontologies](#)
  - [Indented list diagrams of important ontologies](#)
  - [CYC Home Page](#) and [ontology](#) and [table of contents](#)
  - [WordNet Home Page](#) and [online demo](#)
  - Generalized Upper Model: [model](#), [overall organization](#), [concept hierarchy](#), [relational hierarchy](#)
  - [UMLS Home Page](#) and [fact sheets](#), [MeSH](#), [Grateful Med](#) and [demo](#)
  - [TOVE - Toronto Virtual Enterprise](#)
  - [KIF](#) - Knowledge Interchange Format and [brief intro](#)
  - [Stanford KSL Network Services](#) and [Ontology Editor](#)
  - [Guided Tour to Developing Ontologies Using Ontolingua](#); Precursor pages: [Ontolingua](#) with its [Library of Ontologies](#)
  - [EUROKNOWLEDGE Glossary etc.](#)
  - [Stanford DLI](#) and [agents](#), especially for Web browsing
    - Recommend web pages using [Fab](#) (no demos now but several [papers](#))
    - [LIRA](#), a much older system showing key principles (series of screen dumps)
    - A newer [test system](#) without much help (but see [message explaining](#)).
    - [InterPay](#) : [Shopping Models](#), [Secure Electronic Marketplace for Europe](#)
  - [ILU](#) and [Stanford testbed use](#)
  - [Agents '97 Conf.](#)
  - [CHI '97 Software Agents Tutorial](#) by Pattie Maes and her [Software Agents Group](#)
  - [Firefly for music filtering](#) (successor to HOMR from MIT)
  - [My Yahoo](#) (successor to Webdoggie from MIT)
  - [IBM Agent Building Environment \(ABE\): A toolkit for building intelligent agent applications](#)
  - [NEWS WEEDER](#) - naive Bayes classifier - see *AI Magazine* Fall 1997 p. 18
  - [IBM DL: QBIC](#), [agents](#), [WBI tour](#), [watermarking](#)
  - Hal Berghel: [CACM Nov. 1997 40\(11\): Watermarking Cyberspace](#), and [IEEE Computer 29:7 article](#)
  - [DigiCash](#)
  - First Virtual Holdings Inc.: [Green Commerce Model](#)
  - [SSL protocol](#)
- 
- Agents: people and places
    - iimam@site.gmu.edu adaptatation, intelligence
    - yves.Kodratoff@Iri.Iri.fr
    - Brian Gaines, U. Calgary: society of agents
    - Haynes, Sen : U. Tulsa: cases
    - Rus, Dartmouth: gather info
    - Decker, Sycara, Williamson: CMU: multiagent society, planning, matchmaker info agent

Questions:

- Try WordNet on "library" and look for coordinate terms on senses 1,2,3
- Try Grateful Med and find MeSH / Meta Terms for "diabetes"

# What is an Ontology?

[Tom Gruber <gruber@ksl.stanford.edu>](mailto:gruber@ksl.stanford.edu)

Short answer:

An ontology is a specification of a conceptualization.

The word "ontology" seems to generate a lot of controversy in discussions about AI. It has a long history in philosophy, in which it refers to the subject of existence. It is also often confused with epistemology, which is about knowledge and knowing.

In the context of knowledge sharing, I use the term ontology to mean a *specification of a conceptualization*. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy.

What is important is what an ontology is *for*. My colleagues and I have been designing ontologies for the purpose of enabling knowledge sharing and reuse. In that context, an ontology is a specification used for making ontological commitments. The formal definition of ontological commitment is given below. For pragmatic reasons, we choose to write an ontology as a set of definitions of formal vocabulary. Although this isn't the only way to specify a conceptualization, it has some nice properties for knowledge sharing among AI software (e.g., semantics independent of reader and context). Practically, an ontological commitment is an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent (but not complete) with respect to the theory specified by an ontology. We build agents that commit to ontologies. We design ontologies so we can share knowledge with and among these agents.

This definition is given in the article:

T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993. [Available on line](#).

A more detailed description is given in

T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Presented at the Padua workshop on Formal Ontology, March 1993, to appear in an edited collection by Nicola Guarino. [Available online](#).

With an excerpt attached.

---

## Ontologies as a specification mechanism

A body of formally represented knowledge is based on a *conceptualization*: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (Genesereth & Nilsson, 1987). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

# Ontologies as Indented Lists

- CYC (general ontology for commonsense knowledge, 10K concept types): Thing
  - IndividualObject
    - Event
    - Stuff (parent too of IntangibleStuff)
      - Process (child of Event too)
        - SomethingExisting
          - Intelligence
          - CompositeTangible&IntangibleObject
          - TangibleObject
            - TangibleStuff
  - Intangible
    - IntangibleObject
      - IntangibleStuff (also child of Stuff)
      - InternalMachineThing
      - AttributeValue
      - Relationship (also child of RepresentedThing)
        - Slot
          - Attribute
      - Collection (also child of RepresentedThing)
    - RepresentedThing (parent too of Collection, Relationship)
- WordNet (lexical reference system, 70K synsets): thing, entity
  - living thing, organism
    - plant, flora
    - person, human being
    - animal, fauna
  - non-living thing, object
    - natural object
    - artifact
    - food
    - substance
- Generalized Upper Model (250 concepts, for NLP): Um-thing
  - Configuration
    - Doing&Happening
    - Saying&Sensing
    - Being&Having
  - Element
    - Simple-Quality
    - Simple-Thing
    - Participant
    - Circumstance
    - Process
  - Sequence
    - Expanding-Sequence
    - Projecting-Sequence
- Sowa (based on distinctions, combinations, constraints): T
  - Concrete
  - Object

- Process
- Abstract
- (Level 2)
  - PhysicalObject (child of Concrete, Object)
  - PhysicalProcess (child of Concrete, Process)
  - InformationObject (child of Object, Abstract)
  - InformationProcess (child of Process, Abstract)
- UMLS (153 medical concepts): Entity
  - Physical Object
    - Organism
    - Substance
    - Anatomical Structure
    - Manufactured Object
  - Conceptual Entity
    - Language
    - Occupation or Discipline
    - Organization
    - Group Attribute
    - Group
    - Idea or Concept
    - Finding
    - Organism Attribute
    - Intellectual Product
- TOVE (enterprise modeling): Organization-Entity
  - Organization-Individual
    - Employee
    - Contractor
  - Organization-Group
    - Board of Directors
    - Department
    - Division
- GENSIM (generic simulation): Thing
  - Bacteria
    - E.coli
  - Experiments
  - Nucleic Acid Segments
    - DNA segments
      - Genes
    - RNA segments
      - RNA Genes
    - Protein
  - Active Sites
  - Media



[Company Overview](#)

[The CYC® Technology](#)

[Applications](#)

[Platforms](#)

[Availability](#)

[The Upper CYC® Ontology](#)

[Documentation & Publications](#)

[Staff](#)

★ [Jobs](#) ★

[Where's Cycorp?](#)

[Links](#)

[What's New?](#)

---

★ [Cyc is Year 2000 compliant.](#) ★

---

Cycorp has just received an [Advanced Technology Program \(ATP\)](#) award by [NIST](#) (US Commerce Dept) for [1997-2000!](#)

★ [Click here for details.](#) ★

---

Cycorp is a major participant in the DARPA [High Performance Knowledge Base \(HPKB\) Project](#).

Click [here](#) for more information.

Click [here](#) to download a subset of the constant terms and assertions from the HPKB IKB (Integrated Knowledge Base).

---

Cycorp, Inc., based in Austin, Texas, is the world leader in commercializing software with common sense.

---

Cycorp, Inc.  
3721 Executive Center Drive, Suite 100  
Austin, TX 78731  
Internet: [info@cyc.com](mailto:info@cyc.com)  
Telephone: +1 (512) 342-4000  
Fax: +1 (512) 342-4040

---

*Copyright © 1995, 1996, 1997, 1998 Cycorp. All rights reserved.*  
Last updated: December 8, 1997.

# WordNet - a Lexical Database for English

*Cognitive Science Laboratory*

*Princeton University*

*221 Nassau St.*

*Princeton, NJ 08542*

---

WordNet® is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets.

WordNet was developed by the [Cognitive Science Laboratory](#) at [Princeton University](#) under the direction of [Professor George A. Miller](#) (Principal Investigator). Ongoing development of WordNet is supported by DARPA/ITO (Information Technology Office).

Many people have contributed to the success of WordNet. At the present time, the following individuals at Princeton work on the development of WordNet and research using it:

- Dr. Martin Chodorow
- Dr. Christiane Fellbaum
- Dr. Patricia Gildea
- [Professor Philip Johnson-Laird](#)
- [Shari Landes](#)
- [Professor George A. Miller](#)
- [Randee Teng](#)
- Pamela Wakefield
- Joshua Schechter
- [David Slomin](#)



# Unified Medical Language S (UMLS)



---

NLM's Unified Medical Language System (UMLS) project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs. System developers can use the UMLS products to enhance their applications -- in systems focused on patient data, digital libraries, Web and bibliographic retrieval, natural language processing, and decision support. Researchers will find the UMLS products useful in investigating knowledge representation and retrieval questions.

---

- [UMLS Knowledge Source Server](#) is available to those who have signed the UMLS license agreement.
- [License Agreement for use of the UMLS Knowledge Sources](#) includes a list of vocabularies in the UMLS Metathesaurus. The UMLS products are available free of charge to U.S. and international users. Use of the UMLS Metathesaurus may require additional agreements (which may involve fees) with producers of the individual vocabularies it contains.
- [Obtaining Access to UMLS Resources](#)
- [UMLS Applications](#)
- [Unified Medical Language System Fact Sheet](#)
  - [UMLS Metathesaurus Fact Sheet](#)
  - [SPECIALIST Lexicon Fact Sheet](#)
  - [UMLS Semantic Network Fact Sheet](#)
  - [UMLS Information Sources Map Fact Sheet](#)
- [UMLS Documentation](#) contains complete description of the Knowledge Sources and their distribution formats.
- [Comprehensive Bibliography 1986-96](#) For more recent articles search Unified Medical Language System in MEDLINE.

---

Send questions, comments about the UMLS project to: [custserv@nlm.nih.gov](mailto:custserv@nlm.nih.gov) or call 1-888-FINDNLM.

---

[U.S. National Library of Medicine \(NLM\)](#)

<http://www.nlm.nih.gov/>

Last updated: 12 February 1998

# TOVE Manual

## Table Of Contents

---

### Chapter 1

[A Common-Sense Model of the Enterprise](#)

- [1.1 Introduction](#)
  - [1.2 Enterprise Modeling Efforts](#)
  - [1.3 Evaluation Criteria](#)
  - [1.4 The TOVE Project](#)
  - [1.5 Acknowledgments](#)
  - [1.6 References](#)
- 

### Chapter 2

[Ontologies for Enterprise Modelling: Preliminaries](#)

- [2.1 Ontologies and Microtheories](#)
  - [2.2 Time and Action](#)
  - [2.3 References](#)
- 

### Chapter 3

[An Activity Ontology for Enterprise Modelling](#)

- [3.1 Introduction](#)
  - [3.2 Activities and States](#)
    - 3.2.1 Activity Classes and Instances
    - 3.2.2 Successor Axioms for Status of Terminal States
    - 3.2.3 Status of Non-Terminal States
    - 3.2.4 Status of Activities
    - 3.2.5 Duration
  - [3.3 Aggregation of Activities](#)
  - [3.4 Symbology](#) A
  - [3.5 Summary](#)
  - [3.6 References](#)
-

# UMBC AgentWeb

[UMBC LAIT](#) | [AgentWeb](#) | [AgentNews](#) | [agents@cs.umbc.edu](mailto:agents@cs.umbc.edu) | [KQML](#) | [Knowledge Sharing](#) | [NEW!](#) | [Search](#) | [Help](#)

## KIF Knowledge Interchange Format

Knowledge Interchange Format (KIF) is a computer-oriented language for the interchange of knowledge among disparate programs. It has declarative semantics (i.e. the meaning of expressions in the representation can be understood without appeal to an interpreter for manipulating those expressions); it is logically comprehensive (i.e. it provides for the expression of arbitrary sentences in the first-order predicate calculus); it provides for the representation of knowledge about the representation of knowledge; it provides for the representation of nonmonotonic reasoning rules; and it provides for the definition of objects, functions, and relations.

### KIF 101

[KIF101 - a brief non-technical introduction to KIF](#)

### ANSI KIF

- [KIF Specification](#)
- [Model Theoretic Semantics](#) in TeX
- [Standard Ontologies](#)
- [Open Issues](#), [Tabled Issues](#), and [Decisions](#)
- [KIF Electronic Forum](#) and [ANSI KIF Ad Hoc Group](#)

### KIF Version 3

The [Manual for Version 3](#) ( [postscript version](#) )

### KIF related software

- [Prologic](#) a common lisp knowledge representation and reasoning system compatible with KIF
- [EPILOG](#) a common lisp inference system compatible with KIF
- [JKP](#) -- a Java Kif Parser which can parse ascii strings representing sentences in a subset of KIF into a Java representation which encodes the logical structure and is ready for further manipulation
- a C [parser](#) for the Knowledge Interchange Format
- IBM [Agent Building Environment](#) -- A toolkit for building intelligent agent applications

---

AgentWeb is maintained at the UMBC [Lab for Advanced Information Technology](#) by [Tim Finin](#) ([finin@umbc.edu](mailto:finin@umbc.edu)).

---



# A Guided Tour to Developing Ontologies Using Ontolingua

This tour is for people who would like an introduction both to developing ontologies and to using the Ontolingua Ontology Editor provided by the [Stanford KSL Network Services](http://www-ksl-svc.stanford.edu:5915/doc/frame-editor/index.html) for creating and modifying ontologies. We suggest that people who have never created an ontology complete this entire tour before using the editor.

**Please note:** This tour contains a number of screen snapshots. This will mean that the tour is rather more bandwidth intensive than the ontology editor in normal operation, so things will be abnormally slow if you have a low bandwidth connection.

This tour will provide guidance on:

- [How to create an ontology](#)
- [How to create a class](#)
- [How to create a slot](#)
- [How to add a slot to a class](#)
- [How to add a facet to a slot](#)
- [How to create a function](#)
- [How to create an instance](#)
- [How to add an axiom to a class](#)
- [How to create a named axiom](#)

A [glossary of terms](#) is also available.

## [Start the tour](#)

---

*[ontolingua-librarian@ksl.stanford.edu](mailto:ontolingua-librarian@ksl.stanford.edu)*



## Agent Technology Projects in the Stanford Digital Library

Agents are a helpful programming paradigm for object-oriented systems, especially where the task is ill-defined and complex or the agents are making decisions designed to accommodate the user's preferences. The agent paradigm has been used in the Stanford Digital Library Project for information finding (collaborative filtering and distributed gathering) and for economic matters (such as payments).

---

### FAB

FAB is an adaptive multi-agent information retrieval system which finds interesting pages on the web.

"An Adaptive Agent for Automated Web Browsing"

- [Marko Balabanovic](#)
- 

### InterPay

Designed to permit interoperation between different payment services, InterPay provides levels of abstraction which allow applications to be independent of payment mechanism-specific details.

"InterPay: Managing Multiple Payment Mechanisms in Digital Libraries"

- [Steve Cousins](#)
  - [Prof. Hector Garcia-Molina](#)
  - [Scott Hassan](#)
  - [Steven Ketchpel](#)
  - [Andreas Paepcke](#)
  - [Martin Röscheisen](#)
- 

### Distributed Commerce Transactions

Gathering information from multiple, self-interested sources that distrust each other requires certain types of structuring to ensure that a multi-stage exchange has the atomicity property.

"A Sound and Complete Distributed Algorithm for Distributed Commerce Transactions"

- [Prof. Hector Garcia-Molina](#)
  - [Steven Ketchpel](#)
-

# A Demonstration of the LIRA System

---

[Marko Balabanovic](#)

*Department of Computer Science, Stanford University*

*February 1995*

## 1. Overview

The LIRA system was designed to help users keep abreast of new and interesting information appearing on the World-Wide Web. Rather than supporting the *searching* task, where the user has a good idea of what they are looking for and can formulate a search query, we are supporting the *browsing* task, often referred to as *surfing*.

Every day the system presents a selection of interesting web pages. The user evaluates each page, and given this feedback the system adapts and attempts to produce better pages the following day. The system starts with completely random pages, and over time attempts to build a profile of the users interests.

## 2. This Experiment: Finding Music Pages

In order to demonstrate the effectiveness of the LIRA system to people other than the actual user, an objective "interestingness" criterion is required. In this experiment I gave maximum +5 scores to pages related to music, minimum -5 scores to pages unrelated to music, and a +3 score to pages which looked like they might lead to information about music.

The following extracts show the output of LIRA on various days. The first day shows the completely random links from which we start. By day 5 we are starting to get music-related pages. By days 12 and 13 every page suggested is music-related.

- [Day 0](#)
- [Day 5](#)
- [Day 7](#)
- [Day 10](#)
- [Day 13](#)
- [Summary Graph](#)

## 3. Further Information

*N.B. LIRA has been superseded by*





## InterPay: A Project in the Stanford Digital Library

Public libraries have set many expectations by providing free access to high quality information. However, as budgets are slashed and prices on books and journals continue to rise, this expectation is harder and harder to support. Many libraries now charge a fee for extra services, and of course, the for-profit services and sources need to charge as well. The need for electronic currency was clear, and many rose to the challenge of providing the technical mechanisms to transfer money over the network.

However, none of the newly established vendors was able to become the dominant market player, and several competing standards co-exist, each jockeying for a position in the customer's electronic wallet. InterPay (developed in late 1994 -- early 1995) introduced three layers of abstraction which were designed to insulate the application programmer from the details of a payment mechanism. At the *application layer*, the only difference between a for-pay application and a for-free one was an additional parameter, the *payment agent* that was passed from customer to merchant. The merchant made the transition to the *payment policy layer*, asking an object known as a *Collection Agent* to collect the amount of the invoice. The collection agent dealt with the customer's payment agent, which implemented his payment policy--e.g., small amounts should be automatically approved, while larger ones required explicit user approval. The payment agent would then select one of its *payment capabilities*, such as DigiCash, First Virtual, or NetCheque. At this lowest level, called the *payment mechanism layer*, the payment capability interacted with the *collection capability* to effect the transfer and notify higher levels of the outcome.

The implementation of the InterPay architecture showed payments made by the First Virtual system co-existing with payments made through DigiCash and account-based mechanisms.

Improvements to the InterPay architecture led to UPAI, a *Universal Payment Application Interface*. In addition to a cleaner separation of the payment process from the rest of the application, UPAI specifies an asynchronous process, so multiple payments may proceed in parallel, or an initiated but not yet completed payment may be canceled.

However payment is only one part of the shopping experience, and therefore, we inaugurated a project on "**shopping models**" which broadened the scope to increase its coverage. The basic architecture seeks two objectives:

1. Interoperation of existing mechanisms for payment and delivery, and
2. Flexible specification of customer/merchant interaction without requiring significant software development for each new interaction model (like subscription, pay-per-view, gift certificates, auctions, etc)

The first goal was addressed by using UPAI and a related protocol for delivery called U-DEL to interface with existing payment and delivery mechanisms.

The second goal resulted in the specification of a shopping model, which is divided into three parts, for



# Stanford Digital Library Project



## SIDL-WP-1996-0052

### Shopping Models: A Flexible Architecture for Information Commerce

Steven P. Ketchpel, Hector Garcia-Molina, Andreas Paepcke

[ketchpel@cs.stanford.edu](mailto:ketchpel@cs.stanford.edu)

**Abstract:** In a digital library, there are many different interaction models between customers and information providers or merchants. Subscriptions, sessions, pay-per-view, shareware, and pre-paid vouchers are different models that each have different properties. A single merchant may use several of them. Yet if a merchant wants to support multiple models, there is a substantial amount of work to implement each one. In this paper, we formalize the shopping models which represent these different modes of consumer to merchant interaction. In addition to developing the overall architecture, we define the application program interfaces (API) to interact with the models. We show how a small number of primitives can be used to construct a wide range of shopping models that a digital library can support, and provide examples of the shopping models in operation, demonstrating their flexibility.

---

**Note:** Papers in this series are in development and are not in a final form for publication or general dissemination. They are subject to change. Please do not quote or further distribute them without explicit permission from the authors.

---

This paper was created on: 11/09/96 and last revised on: 6/15/1997

**Author's Comments:** The final version that appears in the published proceedings.

**Status:** PUBLIC

[Click here to see the full text of SIDL-WP-1996-0052](#) (PS)

[Click here for the full text of SIDL-WP-1996-0052](#) (PDF)

## Revision History

Version	Format	Date	Comments
<a href="#">4</a>	PS	1/14/1997	The version submitted to DL '97. A major revision of the earlier work.
<a href="#">3</a>	PS	1/14/1997	The version submitted to DL '97. A major revision of the earlier work.
<a href="#">2</a>	PS	11/23/1996	Adds some pictures, plus a better explanation of "roles".
<a href="#">1</a>	PS	11/9/1996	A first, very-preliminary discussion of how InterPay2 enables the support of different shopping models.



# Secure Electronic Marketplace for Europe

## ACTS Project AC026

---

*SEMPER* is a European R&D project in the area of secure electronic commerce over open networks, especially the Internet. It is executed by an interdisciplinary [consortium](#), combining experts from social sciences, finance, retail, publishing, IT and telecommunications, and has established [liaisons](#) with several related efforts.

*SEMPER* is part of the [European Commission's ACTS Programme](#) (Advanced Communications Technologies and Services), executing [Task 503](#). Funding is provided by the partner organisations, the European Union and the Swiss Federal Department for Education and Science.

For more information, see

- **Project Synopsis in English** ([PDF](#)) and in **Français** ([PDF](#))
- [Public Project Reports and Deliverables](#)
- [Mailing Lists](#)
- [... or contact us directly!](#)

For pointers to information on secure electronic commerce outside *SEMPER* click [here](#) (collected by *SIRENE*).

---

## History

29 May 98

New publications on [risks in implementation of digital signatures](#) and [dispute handling in payment systems](#).

31 Mar 98

Press release: ["SEMPER" Security on the Internet: Advanced European E-Business Prototype Goes Online](#) (also in [German](#))

22 Mar 98

[Poster on SEMPER](#)

5 Mar 98

[Slideshow](#) presented at TEN-Telecom Workshop, Brussels, 4 March 1998

10 Dec 97

Public Deliverable on [New Payment Instruments Prototype](#)

24 Sep 97

## Stanford Digital Library Testbed Development

Department of Computer Science  
Stanford University  
Stanford, CA



## Inter-Language Unification or ILU:

- [The ILU Project](#)
- [More ILU Information](#)
- [ILU 2.0 alpha 7 Reference Manual](#) **NEW**
- [ILU 2.0 Reference Manual](#)
- [Postscript ILU 2.0 Release Reference Manual](#) **NEW**
- [Postscript ILU 1.8 Release Reference Manual](#)
- [Postscript ILU 1.7 Release Reference Manual](#)
- [ILU Mailing list Archives](#)

---

### Stanford ILU Installation:

ILU currently supports c, c++, lisp, modula 3, and [python](#).

---

### Configuration Script:

In order for us to update and change the ilu installation, please add the following lines to your .cshrc script:

```
setenv OS_SYSTEM `uname -s | tr -d ' / ' `  
setenv OS_RELEASE `uname -r | tr -d ' / ' | sed 's/\\...*//`
```



The First International Conference on

## Autonomous Agents

Marina Beach Marriott Hotel, Marina del Rey, California

February 5-8, 1997

---

- [Order Agents '97 Conference Materials](#)
- [Read About What Happened at Agents '97](#)
- [Get Ready for Agents '98](#)
- [About the Conference](#)
- [Visit our European Mirror Site](#)
- [Our Sponsors](#)
- [Conference Program](#)
  - [Tutorials](#)
  - [Invited Talks](#)
  - [Panels](#)
  - [Papers](#)
  - [Posters](#)
  - [Robot Demonstrations](#)
  - [Software Demonstrations](#)
  - [Videos](#)
  - [Banquet](#)

# CHI97 Software Agents Tutorial

4/2/97

[Click here to start](#)

## **Table of Contents**

[Software Agents](#)

[Agenda](#)

[Agent Visionaries](#)

[Video Knowledge Navigator](#)

[What is an Agent?](#)

[Common Issues Studied](#)

[Types of Agents](#)

[What is a Software Agent?](#)

[How is an Agent different from other Software?](#)

[Why do we need Software Agents?](#)

[Why do we need Software Agents?](#)

[Direct Manipulation](#)

[Indirect Management/Agents](#)

[Criticisms of Software Agents \(Lanier, Schneiderman\)](#)

[Software Agent =? Expert System](#)

[Agents vs Expert Systems \(cont.\)](#)

[Types of Software Agents](#)

[Types of Software Agents](#)

**Author:** Pattie Maes

**Email:** [pattie@media.mit.edu](mailto:pattie@media.mit.edu)

**Home Page:**

<http://www.media.mit.edu/~pattie>

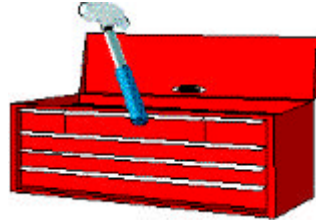
**Other information:**

Software Agents Group MIT Media Laboratory



[What's New?](#)  
[What are Agents?](#)  
[Services](#)  
[Technologies](#)  
[Download Software](#)  
**ABE**  
[WBI](#)  
[White Papers](#)  
[Resources](#)

## Agent Building Environment (ABE)



### A Toolkit for Building Intelligent Agent Applications

- [ABE Introduction](#)
- [Technical Overview](#)
- [Latest level information](#)
- [Software requirements](#)
- [Frequently asked questions](#)
- [Download](#)
- [Questions or Problems](#)

### ABE Introduction

IBM's Agent Building Environment is a toolkit for software developers that makes it easy to build an application based on intelligent agents or to add agents to an existing application. In this version, the intelligent agent watches for a certain condition, decides what to do based on the rules you've given it, and triggers an action as a result. For example, you can ask your agent to frequently check a stock quote over the internet and if the price drops below a certain point, the agent can alert you by paging you. Or, an agent can check on the inventory of a product and if stock is running low, automatically e-mail a refill order to a supplier. The conditions are based on your areas of interest and the rules for agent behavior under those conditions are based on your preferences.

This developer kit comes with a number of pre-built parts which make it easy for you to add agent technology to applications. The "central intelligence" brain for the agent is based on reasoning engine and adapter technologies from IBM's T.J. Watson Research Lab. "Adapters" or interfaces allow the agent to interact with the rest of the world. For instance, the HTTP adapter provided with ABE interfaces with the world-wide-web. The NNTP adapter interfaces with internet USENET news services, and the timer adapter allows events to be triggered based on time. You can write your own adapters as well and guidelines and a sample adapter are also provided. Custom adapters can be written in either C++ or Java \*\*.

A simple full-screen interface is also provided to allow you to specify the rules for the agent's behavior.

A [Technical Overview](#) with more details is available.

---

## Online [Machine Learning](#) software and datasets

---

Each of the following provide source code and data to accompany examples discussed in the textbook [Machine Learning](#).

- [Neural network learning to recognize faces](#) (example from Chapter 4)
- [Bayesian learning for classifying netnews text articles](#) (example from Chapter 6)
- [Decision tree code](#) (to accompany Chapter 3)

*This code and data is made available free of charge for non-commercial use. Please cite this web page in any publications that make use of this data or software.*



[What's New?](#)  
[What are Agents?](#)  
[Services](#)  
[Technologies](#)  
[Download Software](#)  
[White Papers](#)  
[Resources](#)



**What's New?**



[GINKGO Knowledge  
Management Tools](#)

Welcome to IBM's Intelligent Agents Consulting and Programming Services Group. We bring together a broad portfolio of intelligent agent and knowledge management technologies to build solutions that help businesses handle information in smarter ways, become more productive, more innovative and more competitive.

### ***What's New?***



## **"Ginkgo" Knowledge Capture and Virtual Consultation**

Are you looking for better ways for the people in your organization to share information, easily learn from experts who are always busy, eliminate re-work and duplication, and lead to new innovations? IBM now offers:

- ***Ginkgo Knowledge Capture*** learns what people know and builds a knowledge base incrementally while people do their normal jobs
- ***Ginkgo Virtual Consultation*** lets people consult the knowledge of others without needing to speak to them in person.

Turnkey solutions built just for your organization are available using these tools, which are based on new intelligent agent learning technology from IBM. Consulting, prototyping, and complete solutions are some of the service offerings available from [IBM's Knowledge Management Services group](#).

Details about the new Ginkgo technology are available in the [Ginkgo White Paper](#).

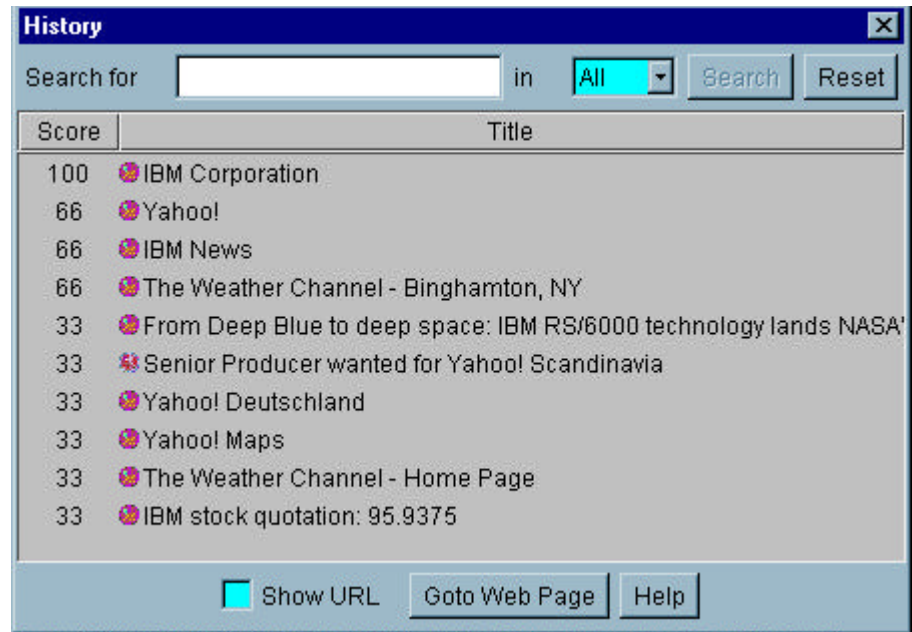


## WBI Personal Web Agent Quick Tour

Take a quick tour through some of the services your WBI agent can provide: Personal History, Watch, Path, Shortcuts, and Offline Browsing.

### Personal History

- Where have I been on the web?
- Where was that word I saw?
- Your WBI agent keeps track of web page titles, content, and url's
- Your personal history is fully searchable by keyword
- Searches return a list ranked by relevancy (how applicable is this hit to my search?)
- Save time by searching only *your* browsed pages, not the whole web



[Next Feature](#)

Return to the [Intelligent Agent Home Page](#)

[ [Home](#) | [Order](#) | [Privacy](#) | [Legal](#) | [Contact IBM](#) ]

# Commerce, Economics, Publishers:

---

## NetBill

- [Home Page](#)
- [Demo](#)
- [Overview article on payment systems from IEEE Spectrum](#)
- [Set of PowerPoint slides discussing Internet Commerce, Payment Systems and NetBill](#)
- Questions: How would this work with ETDs? What are the advantages and disadvantages relative to other approaches?

## Commerce part of CS6604 lecture

- [Workshop on Tech. of Terms and Conditions](#) and [Final Report to NSF](#) - including Breakout Group Reports
- [EC98, International IFIP Working Conference on Distributed Systems for Electronic Commerce](#), Hamburg, Germany, June 4-5, 1998
- [Stanford U. work on electronic commerce, legal pointers](#)

[Projections for Making Money on the Web](#) (Michael Lesk, Harvard Infrastructure Conference, 23-25 January 1997)

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta



**What's  
New**



**About  
NetBill**



**Consumer  
Services**



**Shopping**



**Merchant  
Info**



**Index**



**Help**

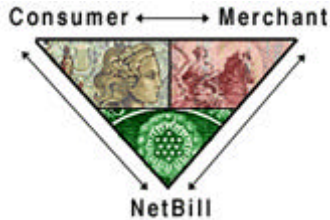


**Send  
Feedback**



**Search**

**...a dependable, secure and economical payment method for purchasing digital goods and services through the Internet.**



NetBill enables consumers and merchants to communicate directly with each other, using NetBill to confirm and ensure security for all transactions.

NetBill lets you use the Internet to order, pay for and receive information goods easily and securely. And NetBill makes it possible for merchants to sell images, articles (even a paragraph of information), and other goods at low cost.

NetBill is only a trial system, however we invite you to explore this site, learn more about NetBill, and give it a try.

[Demo](#) | [Introduction](#) | [Using this site](#) | [FAQs](#) | [Open an Account](#)



Money Tool

**NetBill is a service of:**



**Mellon**



**CyberCash**



All contents copyright © 1995, 1996, 1997 Carnegie Mellon University.

All rights reserved.

Last revision: Tue Oct 28 14:17:12 EST 1997

Version: R1.4.10

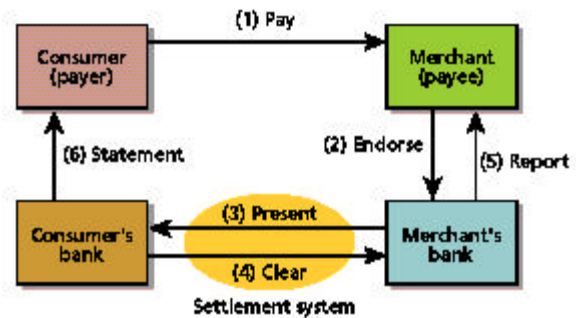
**electronic payments**by Marvin A. Sirbu,  
Carnegie Mellon University

# CREDITS AND DEBITS ON THE INTERNET

A plethora of technologies and business models are in development to enable electronic payments

Since the advent of banking in the Middle Ages, bank customers have used paper-based instruments to move money between accounts. In the past 25 years, electronic messages moving through private networks have replaced paper for most of the value exchanged among banks each day. With the arrival of the Internet as a mass market data network, new technologies and business models are being developed to facilitate electronic credit and debit transfers by ordinary consumers.

These new systems include CyberCash (which is a gateway between the Internet and the authorization networks of the major credit cards) and the Secure Electronic Transactions protocol (a standard for presenting credit card transactions on the Internet), as well as First Virtual (a way of using e-mail to secure approval for credit card purchases of information), GC Tech (a payment system that can use credit or debit via an intermediation server), and NetBill (a public-private-key encryption system for purchasing information).



## Conventional checking

In today's banking world, money consists of ledger entries on the books of banks or other financial institutions. A checking account, also known as a demand deposit account (DDA), records deposits by the consumer and can be used, via the consumer's instructions in the form of a check, to make payments to third parties. Typically, a check is written by a consumer, authenticated by signature, and presented to a merchant, who may endorse it with a signature before presenting it to a bank for payment. If the merchant's bank and the consumer's bank are the same, it can simply transfer the funds on its ledgers from the consumer's account to the merchant's. If the payer and the payee keep accounts at different banks, the payee bank presents the check for settlement to the payer's bank and receives the funds in return through a settlement system. Several private check clearinghouse systems, as well as the Federal Reserve system, provide settlement services in the United States [Fig. 1].

When checks are sent to banks for deposit, merchants do not yet know if consumers have adequate funds and therefore need to find out whether the checks cleared. Similarly, consumers receive statements from their banks showing which checks have been paid. Any discrepancy between bank records and those of the payers may indicate that forged checks were presented against consumers' accounts.

This model works equally well when there is a negative balance in consumers' accounts, at least if the consumers' banks are willing to extend credit—that is, to lend the consumers funds needed to pay off the checks. Many banks in the United States and Europe provide such credit facilities, sometimes referred to

# Towards A Formalism for Terms and Conditions

## Workshop Homepage September 24 - 26, 1996

A major obstacle to the further development of digital libraries, and the national information infrastructure as a whole, is the lack of adequate means of providing digital objects and information on any basis other than free, unrestricted access. Authors are increasingly taking the path of self-publishing using assorted home-grown schemes to seek payment and to impose terms and conditions on use. Publishers wish to specify terms of use and ensure those terms are enforced (optionally collecting payment), before providing valuable materials on the net. While payment and related topics are the subject of much commercial activity, mechanisms for the specification of terms of use seem to have been largely neglected.

Accordingly, a workshop was held on developing a formalism for terms and conditions for the use of digital objects and information. The Workshop organizers were [James R. Davis](#) (Xerox) and Judith L. Klavans (Columbia University, [Center for Research on Information Access](#) and Department of Computer Science).

The workshop took place September 24 - 26, 1996 at the Columbia University Conference Center at Arden Homestead, north of New York City. Now that the workshop is complete, we'll use this page as a reference source for further work on terms and conditions.

- [Workshop schedule](#)
- [List of attendees](#)
- [Readings from the workshop](#)

### Reports and presentations about the workshop

- [Workshop Summary: Technology Issues for Terms and Conditions](#) (a brief summary from [D-Lib magazine](#), October 1996)
- [Presentation](#) given at Conference on "Digital Content", Center for Law and Technology, University of California at Berkeley, California, November 8, 1996, by Judith L. Klavans
- Presentation at 1996 NSF Digital Libraries Initiative Meeting, Stanford University, Stanford, California, December 17, 1996. Slides from [David Millman](#), [Vicky Reich](#), and [Judith L. Klavans](#).
- [Final report to the NSF](#) by Judith Klavans. (Added May 27, 1997)

### Related links

- [Economics of Digital Information and Intellectual Property](#) (draft papers from a conference held at Harvard, January 23-25, 1997)
- [Bridging Digital Technologies and Regulatory Paradigms](#) Conference June 27-28, 1997, Haas School of Business, University of California, Berkeley.

# Projections for Making Money on the Web

**Michael Lesk**

Harvard Infrastructure Conference, 23-25 January 1997

## **Abstract**

Numerous groups will sell you advice on getting rich on the Web; they discuss online sales of information, retail catalog shopping, advertising, consulting, and connectivity. What will actually pay for the Web? What is the 'killer ap'?

This paper contains a great many conflicting numbers. The different predictors of future revenues differ; even the measures of current success differ. No effort is made to resolve the conflicts, since knowing the spread in estimates may be of value to the reader.

My personal projection for getting rich: connectivity.

## **1. Introduction.**

What is going to pay for the Web? Why should Web site providers continue to do the work of writing, drawing, and coding, plus bear the cost of equipment and communication lines? Justification for Web sites includes many reasons which involve no direct financial gain, such as self-promotion. However, many site builders are hoping to get rich, despite costs that may run over \$1M for a large professionally designed set of corporate web pages.

Among the possible models of finding wealth on the Internet are:

1. Selling objects via the Web. In this dream, the L. L. Bean catalog is replaced by a set of web pages, and calling the 800 number changes into web clicks. Delivery would still be via a parcel carrier.
2. Selling information via the Web. Again, people look at web pages and buy things, but the result can be sent to them directly since it is electronic access. People may pay item by item, or for continuous or regular access to a particular information service.
3. Selling advice. In this case, the basic information is free; and what is being sold is some kind of selection, editing, or consulting related to it.
4. Selling advertising. Again, the information is free, and is supported by advertising in the same way as broadcast TV. For the last year this has been perhaps the most hyped possibility for paying for the web.
5. Selling connectivity. One service that most people do pay for is access to the web itself. Will it be possible to fund the services provided out of such connectivity charges?

This paper discusses some of the estimates for the amount of money that people might pay for each of these services. The traditional business motivations are greed and fear. So far, a great deal of Web activity could be said to be motivated by fear: 'our competitors have a web page.' Is it likely that greed will take over? And will it actually result in riches?

## **2. Selling objects.**

# Intellectual property rights, copyright laws and legal issues:

---

(Chapter 10, page 223, "Books, Bucks and Bytes", Michael Lesk)

- [Cyberspace Law for Non-Lawyers](#): This is an electronic course : a "real" course in the "real world" This site includes a discussion function which will allow you, if you are so inclined, to post your own comments and reactions to the individual messages that the instructors have mailed out.
- [Overview of Copyright Laws in the Digital Domain](#) and [References](#) : Check out the references for some very good links and information on copyright laws and related issues.
- [Pamela Samuelson](#) and pointers based on her pages and recommendations
- [Electronic Commerce](#)
- [Workshop on Tech. of Terms and Conditions](#) and [Final Report to NSF](#) - including Breakout Group Reports
- [EC98, International IFIP Working Conference on Distributed Systems for Electronic Commerce](#), Hamburg, Germany, June 4-5, 1998
- [Stanford U. work on electronic commerce, legal pointers](#)

## Other related references:

- Digital Copyright Protection - Peter Wayner - AP Professional - Boston, 1997
- Scholarly Publishing: The Electronic Frontier - ed. Robin P. Peek and Gregory B. Newby - The MIT Press, Cambridge, MA, 1996
- The Network Nation - Starr Roxanne Hiltz and Murray Turoff - The MIT Press, Cambridge, MA, 1994
- Ubiquitous Email ...

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

# Pamela Samuelson Plus Recommendations on Law and Digital Libraries

[Professor Pamela Samuelson](#) is one of the leading authorities on legal issues in the area of intellectual property rights (IPR). A new [MacArthur Fellow](#), a Fellow of the [Electronic Frontier Foundation](#), a Fellow of the [Cyberspace Law Institute](#), she is a Professor at the [University of California at Berkeley](#) with a joint appointment in the [School of Information Management and Systems](#) and the [School of Law](#).

For more information on this and related topics, see

- [Selected Papers by Pamela Samuelson](#)
- [Law 276: Cyberlaw](#) - by Pamela Samuelson, University CA, Berkeley
- [Infosys 296A: Future of the Information Society, Copyright & Community](#) - by Peter Lyman and Pamela Samuelson, University CA, Berkeley
- [Cyberspace Law for Non-Lawyers](#), which attracted over 20,000 subscribers, by [David Post](#), [Temple U. School of Law](#); Lawrence Lessig, [Harvard Law School](#); [Eugene Volokh](#), [UCLA School of Law](#)
- [Crash Course in Copyright](#) from UT system, including the [Digital Library](#)
- [Copyright Management Center](#) of IUPUI, directed by [Kenneth Crews](#)
- [The ILTguide to Copyright](#) at Columbia, for educators
- [Copyright Law Materials](#) at Cornell Legal Info. Institute
- [Copyright & Fair Use](#) site of Stanford University Libraries
- [Copyright Basics Circular from the U.S. Copyright Office](#)
- [Copyright Clearance Center \(CCC\)Online](#)
- [Digital Future Coalition \(DFC\)](#)
- [IIP Policy Gateway, Harvard Information Infrastructure Project](#)
  - [Bibliography](#)
  - [Policy resources in the area of Internet governance](#), supplement to MIT Press [book](#)
  - [The Impact of the Internet on Communications Policy conference](#)
- [ALAWON](#) - ALA (American Library Association) Washington Office Newslines providing urgent and late breaking news
- [ARL Federal Relations and Information Policy Program](#), Prue Adler

## Social Issues:

---

- Social Aspects [D-Lib Working Group](#)
- UCLA Workshop, Social Aspects of Digital Libraries, Feb. 16-17, 1996  
<http://www-lis.gseis.ucla.edu/DL/>
  - Life Cycle [http://www-lis.gseis.ucla.edu/DL/UCLA\\_DL\\_model.gif](http://www-lis.gseis.ucla.edu/DL/UCLA_DL_model.gif)

---

[\[Main\]](#) [\[Contents\]](#) [\[Topics\]](#)

---

Please send comments/suggestions to [Ed Fox](#).

(c) Copyright 1998, Edward A. Fox, Rajat Gupta

**WORKING GROUPS**

## **D-Lib Group on Social Aspects of Digital Libraries**

### **I. UCLA-NSF Workshop on Social Aspects of Digital Libraries**

An invitational workshop was held at UCLA, February 15-17, 1996; 32 researchers, developers, and practitioners, 9 UCLA faculty facilitators, and 6 UCLA graduate research assistants participated. All materials from the workshop, including schedule and agenda, list of participants, participants' discussion papers and biographical statements, and summary reports presented at the meeting are available on the web site (<http://www.gslis.ucla.edu/DL/>).

We selected two research areas, each with three sub-topics, as focal points for a two-day workshop:

Information Needs: Identifying real information needs and developing digital libraries to meet those needs.

- Social context and culture
- Information needs and information seeking
- Linking user-learner needs and behavior to digital library design

End user searching and filtering: Designing digital libraries in which it is possible to find the right information in a glut of information.

- Organization, description and representation of information
- Search capabilities for users
- Interface design for information retrieval

### **II. Results of the workshop**

While we bounded the scope of the workshop to provide a starting point for discussion and a set of criteria for selecting participants, our participants quickly expanded those boundaries.

The boundaries expanded in several directions:

- Level of analysis: Our scope, as stated in the background paper (see web site), focused on the needs and activities of the individual user. While important, we must recognize that individuals do not work with information resources in isolation from their communities. They perform individual tasks in the context of their work teams, classroom, and other social organizations. Many tasks are performed in group contexts; we must consider CSCW and collaboratory environments as well. Multiple levels of analysis are required.
- Scope of analysis: Our scope addressed information searching and retrieval processes. While important, we must set searching in the context of the cycle of information creation and utilization. People will create information in digitized form that becomes part of digital libraries and need tools and functional capabilities for doing so. They will search for information created by other people, and for purposes other than those intended by the creators, requiring a variety of searching functions. Once located, they will incorporate new information into other products and processes that become part of the life-cycle. We need consistent means to organize, describe,