



Proceedings of

First ACM/IEEE-CS



Joint Conference on Digital Libraries



June 24-28, 2001 • Roanoke Virginia, USA

Table of Contents

Note: Long papers are assigned a 30 minute period, while short papers (shown by an asterisk preceding the title) have a 15 minute period.

- **The Open Archives Initiative: Building a Low-Barrier Interoperability Framework** 54
Carl Lagoze, Herbert Van de Sompel (*Cornell University*)
- ***Enforcing Interoperability with the Open Archives Initiative Repository Explorer**..... 63
Hussein Suleman (*Virginia Tech*)
- ***Arc — An OAI Service Provider for Cross-Archive Searching**..... 65
Xiaoming Liu, Kurt Maly, Mohammad Zubair, Michael L. Nelson (*Old Dominion University*)

Paper Session 4B: Digital Libraries and the Web: Technology and Trust

Moderator: Jonathan Furner (*University of California at Los Angeles*)

- **Managing Change on the Web**..... 67
Luis Francisco-Revilla, Frank Shipman, Richard Furuta, Unmil Karadkar, Avital Arora (*Texas A&M University*)
- ***Measuring the Reputation of Web Sites: A Preliminary Exploration** 77
Greg Keast, Elaine G. Toms, Joan Cherry (*University of Toronto*)
- **Personalized Spiders for Web Search and Analysis** 79
Michael Chau, Daniel Zeng, Hsinchun Chen (*University of Arizona, Tucson*)
- ***Salticus: Guided Crawling for Personal Digital Libraries** 88
Robin Burke (*California State University, Fullerton*)

Panel 4C: Different Cultures Meet: Lessons Learned in Global Digital Library Development

Moderator: Ching-chih Chen (*Simmons College*)

Panelists: Wen Gao (*Chinese Academy of Sciences*), Hsueh-hua Chen (*National Taiwan University*), Li-Zhu Zhou (*Tsinghua University*), Von-Wun Soo (*National Tsinghua University*)

Paper Session 5A: Tools for Constructing and Using Digital Libraries

Moderator: Hsinchun Chen (*University of Arizona, Tucson*)

- **Power to the People: End-User Building of Digital Library Collections** 94
Ian H. Witten, David Bainbridge, Stefan J. Boddie (*University of Waikato*)
- ***Web-Based Scholarship: Annotating the Digital Library**..... 104
Bruce Rosenstock, Michael Gertz (*University of California at Davis*)
- **A Multi-View Intelligent Editor for Digital Video Libraries** 106
Brad A. Myers, Juan P. Casares, Scott Stevens, Laura Dabbish, Dan Yocum, Albert Corbett (*Carnegie Mellon University*)
- ***VideoGraph: A New Tool for Video Mining and Classification** 116
Jia-Yu Pan, Christos Faloutsos (*Carnegie Mellon University*)

Paper Session 5B: Systems Design and Evaluation for Undergraduate Learning Environments

Moderator: Traugott Koch (*Lund University & Technical Knowledge Center of Denmark*)

- ***The Alexandria Digital Earth Prototype System** 118
Terence R. Smith, Greg Janée, James Frew, Anita Coleman (*University of California at Santa Barbara*)
- ***Isclapes: Digital Libraries Environments for the Promotion of Scientific Thinking by Undergraduates in Geography** 120
Anne J. Gilliland-Swetland, Gregory L. Leazer (*University of California at Los Angeles*)
- ***Project ANGEL: An Open Virtual Learning Environment with Sophisticated Access Management**..... 122
John MacColl (*University of Edinburgh*)

PANEL — Different Cultures Meet: Lessons Learned in Global Digital Library Development

Moderator: Ching-chih Chen
Graduate School of Library and Information Science
Simmons College, Boston, MA 02115, USA
chen@simmons.edu

Panelists:
Wen Gao
Hsueh-hua Chen
Li-Zhu Zhou
Von-Wun Soo

This panel is organized to share the experience gained and lessons learned in developing cutting-edge technology applications and digital libraries when different cultures meet together. "Culture" is interpreted in different ways and different context. This include the interdisciplinary collaboration among professionals from different fields with their own cultures -- such as library/information science, computer science, humanities, social sciences, science and technology, etc; to more globally as experienced in major international collaborative projects involving R&D professionals from two or more different cultures -- the East and the West, or the North and the South.

The moderator will share her own personal perspective on the "true meaning of global and interdisciplinary collaboration," drawing upon experiences gained in conducting numerous technology related R&D activities throughout the years, starting from her award winning multimedia project on the *First Emperor of China*¹ and his 7000 magnificent terracotta warriors and horses, supported by the National Endowment for the Humanities in the mid-1980s to her recent (since May 2000) and challenging NSF's International Digital Library Project (IDL) called *Chinese Memory Net (CMNet): US-Sino Collaborative Research Toward Global Digital Library*, culminating the "community building" experiences at NIT conferences with many participants of JCDL from over 15 countries at the NIT 2001 in Beijing, China during May 29-31, 2001.

CMNet's US affiliates include academic researchers from several universities, including Drexel University, Kent State University, Syracuse University, University of Kentucky, and University of Wisconsin-Milwaukee. Its collaborators in Beijing include Peking University and Tsinghua University, in Shanghai include the Shanghai Jiaotong University, and in Taipei including National Taiwan University, National Tsinghua University, and the Academia Sinica. Several collaborators from Beijing and Taiwan are represented at

this panel. Despite the challenging problems and issues related to such type of collaboration among researchers in multiple interdisciplinary fields, productive R&D activities are ongoing in areas related to:

- Multilingual and multimedia information systems,
- multilingual and multi-modal search and retrieval, and translation engine,
- interoperability and scalability technology for large global collections,
- metadata techniques and tools,
- preservation and archiving of digital information,
- intelligent agent, and
- right protection and management.

Opportunities and benefits related to these collaborative R&D activities for global digital library development when the East and West cultures meet will be highlighted. This introduction will offer a setting for exciting presentations followed by interactive discussions from leading researchers in digital library developments in both Beijing and Taipei, all of whom are current collaborators of CMNet and their current research are supported mostly with matching grants from the NSF's counter parts in those areas. They will highlight their cutting-edge R&D activities in digital library development in recent years. Thus, the panel should be of interest to all those who are interested in international digital library collaborative research.

PANEL:

Collaborative Development of a Large-Scale Global Digital Library: The China-US Million Book Project

Wen Gao, Deputy President, Graduate School, and Institute of Computing Technology, Chinese Academy of Sciences, Beijing

Content-based retrieval is a key function in the large-scale multimedia database. It is significant to many real applications, such as the application of digital libraries. In

this brief presentation, recent experience in starting to plan and implement the China-US Million Book Digital Library for education purpose will be shared. The plan in integrating both useful industry provided modules as well as those in-house developed ones will be discussed. This integration include the employment of an XML-database system to manage the multimedia content, an image-based retrieval module, a video analysis module, a multi-linguistic module to provide machine translation function, etc. Information on how these modules can be fit into the framework of content-based retrieval in this large-scale DL project, as well as the challenges and potentials of the China-US collaboration will be presented at the panel.

Taiwan's Experience of Multidisciplinary and International Collaboration in Digital Library Development

Hsueh-hua Chen, Professor and Chair, Department of Library and Information Science, National Taiwan University, Taipei; and Jieh Hsiang, Professor and Dean, College of Science and Technology, National Chi-nan University, Puli

At this panel we will share the experience of the National Taiwan University (NTU) on multidisciplinary and international collaborations, as well as outline our plan for the future.

In October 1996 NTU initiated a digital museum/library (NTU DLM) project with the goal to digitize several important historical archives of documents and artifacts of Taiwanese indigenous people. The project involved researchers from the Departments of Anthropology, Computer Science, History, Library Science, and the University Library. It was the first large-scale research collaboration between the Colleges of Humanities and Engineering of NTU. The success of the project became a catalyst for the National Digital Museum project (NDMP) of Taiwan, inaugurated in July 1998 by the National Science Council (NSC).

Like other pioneering projects, NTU DLM suffered through the initial period of trying to find a common language and a right working model among people from different disciplines. Yet, several factors contributed significantly to the success of this effort. These include:

- The bondage by a strong sense of mission and urgency to preserve the indigenous Taiwanese cultural heritage.
- The funding limitation (except for the generous support of equipment from the university) at the beginning phase made it clear that helping each other is the only way for the project to succeed. (The lack of funds added the insurance that people working on the project were not in it "for the money.")

- The creation of a prototype with search capability at the early stage of the project eliminated potential doubts on the content experts (historians and anthropologists) of the usefulness of digitization.
- The early participation in the CIMI metadata test-bed project gave the team a jump-start on the metadata development. This experience eventually led to the formation of the Resources Organization and Searching Specification (ROSS) team that developed the first comprehensive metadata system for Chinese cultural objects in Taiwan, called MICI, *Metadata Interchange for Chinese Information*.

This multidisciplinary collaboration experience was further augmented through the Taiwanese Butterfly Digital Museum project, jointly developed by the National Museum of Natural Science and National Chin-Nan University, under the sponsorship of the NDMP. In addition to the original team of computer and library specialist, this project also involved biologists, researchers on user behavior, graphic designers, and multimedia technology specialist. The interaction among these researchers eventually led to a novel approach to content-based retrieval, possibly the first in the digital museum context. It also helped the ROSS team enhanced its metadata by incorporating artifacts about biological specimens.

Last year the NTU became one of the collaborators of the CMNet project led by Professor Ching-chih Chen of Simmons College in an effort to coordinate and consolidate US-Sino digital library R&D work related to the digital Chinese cultural heritage. Like other DLM related projects, the NSC - Taiwan's the counter part of NSF - funds the IDLP-related activities. While we are still at the early stage of this collaborative work, it is expected that this project will further broaden our horizon on international collaboration, and our presence at this JCDL panel and workshop is the beginning of our digital library "community building" effort.

Experience of Multidisciplinary and International Collaboration in Digital Library Development at Tsinghua University

Li-Zhu Zhou, Chair, Department of Computer Science, Tsinghua University, Beijing

At this panel, we will share our experience in the development of the Tsinghua University's Architectural Digital Library (TACADL) project in support of user-centered information search, retrieval, navigation, storage, and knowledge management on Chinese ancient architecture study. This project brings together three research teams from different disciplines including computer science, architecture science, and library information management. The project requires substantial

¹ The *First Emperor of China* videodisc (1991) and multimedia CD (1994) have been published by The Voyager Company. (<http://voyager.learntech.com>).

cooperation among those computer researchers, librarians, and subject specialists involved.

Similar to any software engineering project, the building of TACADL is a typical process of requirement analysis, functional specification, system design, implementation and testing. During this process every team member has a role to play. However, the development of a digital library presents some new challenges. In TACADL, a good example is the cataloguing and metadata registration of digitized materials. When the volume of digital content grows, the bottleneck problem becomes more evident due to the lack of subject domain knowledge experts. Another problem is the development of an effective digital library system that can provide content-based indexing tools to enable users to navigate and retrieve relevant information easily. These open issues require joint effort from all teams.

Since the second half of 2000, researchers in computer science, and library and information science from Tsinghua University, Peiking University, and Shanghai Jiaotong University become collaborators of CMNet, led by Professor Ching-chih Chen. As stated by Dr. Chen, R&D projects in areas many exciting areas such as multi-lingual, multi-modal search and retrieval, and the like, have been proposed to the Natural Science Foundation of China, Chinese counter part of NSF, for funding. This collaboration has opened doors for more international interaction, and our participation at this JCDL meeting is a good example.

In celebrating Tsinghua University's 90th anniversary, Prof. Chen brought her *NIT 2001: The 12th International Conference on New Information Technology* to Beijing last month (May 29-31, 2001). The conference's theme was *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. It had a very rich program, and we at Tsinghua were delighted to welcome over 70 distinguished participants from over 15 countries as well as over 150 Chinese participants. Many of them are at this JCDL meeting. *NIT 2001* was a great learning and sharing experience, and was in the true spirit of international collaboration. It is a wonderful learning and sharing experience.²

Multiagent Coordination, Negotiation, and Applications in Digital Libraries: A Case Study of International and Multidisciplinary Collaboration

Von-Wun Soo, Professor of Computer Science, National Tsinghua University, Hsin-chu

At this panel, I shall share our experience as one of the IDLP partner of Dr. Ching-chih Chen's CMNet.

Our IDLP project started with extending the legacy information retrieval system at the National Central Library (NCL) in Taipei for electronic government official gazettes of Taiwan. The NCL documents are basically in plain Chinese text. To enhance the recall and precision of current retrieval system, we developed a multi-information agent architecture that consists of 4 types of agents:

- 1) The librarian agent is a coordinator/manager who makes decisions and allocate tasks to other agents.
- 2) The ontology agent analyzes retrieved items to produce indexed grouping of items as well as other contextual hints.
- 3) The thesaurus agent provides term manipulation capabilities, such as word segmentation, semantic association and expansion. Linguistic knowledge of the thesaurus agent is based on some well-established thesauruses with hierarchical concept structures. This allows agents to do intelligent query expansion and semantic keyword association.
- 4) Finally, a set of information gathering agents are responsible for accessing outside information sources to retrieve information items not located in the main sources, when necessary.

We have developed a prototype system and obtained satisfactory information retrieval results on several test query examples.

Recently, we have extended the system design to a multimedia information retrieval system by capitalizing the multimedia data of Dr. Ching-chih Chen's CD product, *The First Emperor of China*, as a test-bed in the IDLP project. This CD-ROM consists of numerous pictures, videos, sounds, and text on the Emperor and Qin dynasty. This shift from plain-text to multimedia historic information retrieval demands the system to allow information agents to utilize the lexical knowledge about the historic figures, systems, events, cultural objects, and so on. To allow agents to conduct semantic association of keywords and retrieve relevant multimedia literature, we are designing a multimedia annotation system to describe multimedia documents in terms of keywords coded in semantic hierarchy of a Chinese thesaurus. The annotation system can easily be extended from the current components of the thesaurus agent.

In addition, we need to develop historic event schema for the ontology agent to reason on various relations among historic events. We need to address issues on the integration of ontology and thesaurus by the semantic web approach of XML/RDF and XML/Schema. This could require tremendous expertise intensive tasks to develop generic as well as domain-specific historic ontology. Our ultimate goal in IDLP project is to develop an agent-mediated retrieval system for Chinese historic literature and media. We believe it is a challenging, interesting and worthwhile international and multi-cultural collaborating research project that deserves the study of many research issues in the near future.

From this brief presentation, it is clear that through the international collaboration with Dr. Chen's CMNet, our respective research becomes richer. In addition, from the R&D point of view of a computer scientist, to be able to benefit from the enormously diversified and rich information resources and digital data without having to spend time and effort in creating them for our research is most attractive. Dr. Chen's Emperor digital contents are particularly valuable since it includes the much coveted multimedia digital sources and data of all types of media.

² Chen, Ching-chih, ed. *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing: Tsinghua University Press, 2001.