LIBRARIES
IN THE DIGITAL AGE
May 25-29 2004, Dubrovnik and Mljet, Croatia

Human Information Behaviour & Competences For Digital Libraries

Organised by:

Faculty of Philosophy, University J. J. Strossmayer in Osijek, Croatia

Rutgers University, New Brunswick, NJ, USA

At the:

Inter-University Center, Dubrovnik
and
Hotel Odisej, Mljet

May 25-29, 2004

The Promise of International Digital Library Collaboration for Innovative Use of Invaluable Resources[*]

Ching-chih Chen
Professor
Graduate School of Library and Information Science
Simmons College
Boston, MA 02115, USA

E-mail: chen@simmons.edu

Abstract: In this digital era, we have witness the exciting convergence of content, technology, and global collaboration in the development of digital libraries. Today's information seekers, regardless whether they are general public, school children, or those from research and higher education communities seek information for education, research, entertainment, or enrichment in very different ways from before. From the information resources point of views, the old model of "owning" a collection has given way to "sharing," and the new emphases have shifted from possessing large "physical libraries" to "virtual libraries" digitally distributed all over the world.

This author has experienced much of these transformations through her own R&D activities – from the creation of interactive videodisc and multimedia CD in the 80s and 90s to leading a current international digital library project, *Global Memory Net,* supported by the US National Science Foundation, and her involvement in the China-US Million Book Digital Library Project. In addressing the potential use and user of digital libraries, she will elaborate on how some of the new digital technology and techniques can and should be explored to enhance access, retrieval and use of global resources in addition to our traditional ways of creating and organizing materials.

I am truly delighted to be in Dubrovnik again. My sincere thanks to Tatjana Aparac-Jelušic and Tefko Saracevic for their very kind invitation! In technological terms, it has been a long time since I was in this gorgeous city in 1984 invited by my good friend, Bob Hayes, to speak at his conference. I remember vividly how difficult it was to make my interactive videodisc on *The First Emperor of China* work properly, and how excited a local Academy of Science member was when he saw the flexible and friendly way multimedia information was linked and provided instantly at the point of need with a simple click of the mouse! He came to me after my talk and said, "Professor Chen, this is unbelievable! I would not believe it if I did not see it with my own eyes! Your system is for idiots!" I was very worried for a few seconds, and then I thanked him profusely because he meant to say that my system was so easy that even the idiots can use it! This description offers the best definition of "user-friendliness." What else can be friendlier than a system that even an idiot can use without any trouble? This favorite story of mine seems to relate well to the theme of this LIDA conference on human information behavior, and I intend to expand on this later in my talk.

[*] Opening Invited speech presented on May 25, 2004 at *LIDA: Libraries in the Digital Age,* May 25-29, 2004, Dubrovnik and Mljet, Croatia.

We have truly come a very long way since then. We have moved from the workstation environment to the global networked environment. We have moved from the use of hardcopy and analog resources to digital content. We are truly living in a new period of unprecedented opportunities and challenges, fueled by enormous progress in science and technology. The information technological innovations intertwined with interdisciplinary knowledge base, which is propelling the 21st century's knowledge economy. So, in this digital era, we have witness the exciting convergence of content, technology, and global collaboration in the development of digital libraries.

CONVERGENCE OF CONTENT, TECHNOLOGY, AND GLOBAL COLLABORATION

During 1998-2002, I was privileged to serve as a member of the US President's Information Technology Advisory Committee (PITAC), and was deeply involved in the drafting of several PITAC Reports to the President. The one related to digital libraries was the 2001 Report of the PITAC Digital Library Panel, entitled *Digital Libraries: Universal Access to Human Knowledge* [1]. It offers an ambitious vision:

> *"All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge. Via the Internet, they can access knowledge in digital collections created by traditional libraries, museums, archives, universities, government agencies, specialized organizations, and even individuals around the world. These new libraries offer digital versions of traditional library, museum, and archive holdings, including text, documents, video, sound, and images. ...*
>
> *Very-high-speed networks enable groups of digital library users to work collaboratively, communicate with each other about their findings, and use simulation environments, remote scientific instruments, and streaming audio and video. ... In this vision, no class-room, group, or person is ever isolated from the world's greatest knowledge resources."* [1]

If we dissect this abbreviated vision statement, and compare the segments with several themes of this conference:

- information seeking in the environment of digital libraries;
- users and use of digital libraries;
- what do we know how do users go about the general processes of Selecting, Finding, Identifying and Obtaining materials and services in digital libraries?
- research dealing with use of features, access, outcomes, and usability.

Then, we should be clear on the targets which we should be aiming for. It is clearly stated in the PITAC's vision that one should be able to find any information he/she needs whenever and wherever needed. Here, the scope and parameters are also set for "users" and "use":

- Users: ALL citizens of the world regardless age, ethnic group, religion, etc…
- Uses: For whatever they need
  - Contents – All subject areas, all types of formats, all digital collections of all types of organization;
  - Geographical areas - All parts of the world.

The PITAC's vision is so much easier said than done! It will require sustainable and long-term commitment of many. We should expect many obstacles on the long road to this "elusive" vision, as stated in the latest Report of the *US National Science Foundation's (NSF) Workshop on Research Directions for Digital Libraries*, entitled *Knowledge Lost in Information* [2]. However, the report also mentioned that we have made substantial advances in the technical area in terms of advancing capabilities, through an interagency program of integrated, interdisciplinary, project-oriented research initiatives in the last decade. Now, the convergence of content, technology, and global collaboration in the development of digital libraries should be a natural process. From content or information resources point of view, no one institution -- no matter how large it is, or one country -- no matter how abundant in resources, can possibly have everything. Thus, the old model of "owning" a collection has to give way to "sharing," and the new emphasis has to shift from possessing large "physical libraries" to "virtual libraries" digitally distributed all over the world. If we are talking about content building in "global" scale, then we must have global collaboration through global community building. Thus, our usual "user community building" is leading toward the global scale.

COMPONENTS OF INTERDISCIPLINARY DIGITAL LIBRARY RESEARCH

In order to improve productivity through universal information access, I would like to share with you the triangular relationship observed by the *DELOS-NSF Working Group in Digital Imagery for Significant Cultural and Historical Materials* as reported in the Group's Report [3, 4], Figure 1. This conceptual model attempts to illustrate the relationships among *people, content,* and *technology* in digital library research agenda. In other words, an interdisciplinary digital library research will develop technologies to enhance the way people create and access the content.

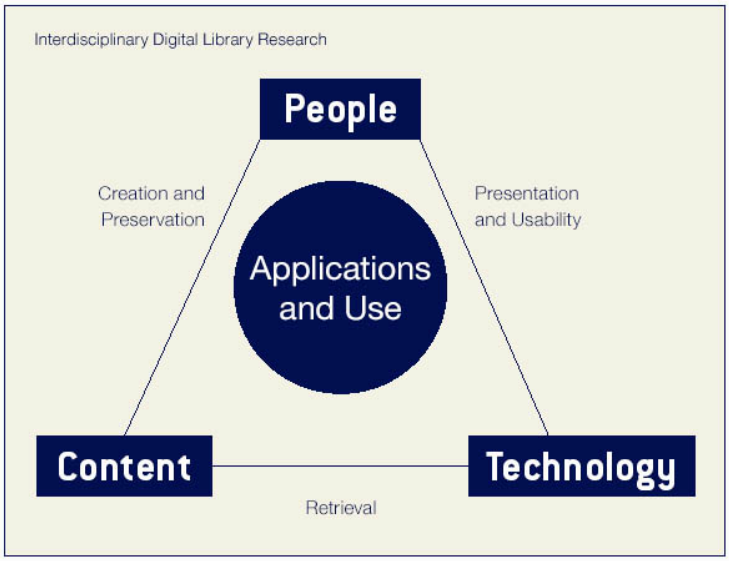

Figure 1. Interdisciplinary Digital Library Research Model ([3,4],[2], p.13)

People encompass all users, from curators and library and information scientists, to scholars, teachers, and students in all areas of the humanities, to citizens of all cultures. Content is the vast array of significant materials throughout the world. Technologies are the enabling research and development in all related technical areas such as information retrieval, image processing, artificial intelligence, and data mining.

The Working Group recommends focused, interdisciplinary research programs along the three edges and the center of the triangle – Applications and Use – areas that traditional research program currently neglect. The research area between people and content is the area of digital imagery *creation and preservation*. The area between content and technologies is the efficient and effective retrieval of content using technologies. Research into *presentation and usability*

will enhance the ability to access the content. Effective *applications and use* of the research results, under lifecycle management, will integrate research of the three related areas.

USE AND USABILITY FOR APPLICATION DEVELOPMENT

It is not my intent to try to address the use and usability of digital libraries or resources in different types of libraries in this paper. In 2002, Denise Troll of Carnegie Mellon University stated the needs and concerns in having good methods in assessing use and usability of libraries in the digital environment in her publication [5]. Since then, many initiatives have been introduced to explore how information is made, assembled into collections, presented and used; and also how the digital collections differ from the traditional library ones. Clearly the fast developing digital environment has changed the needs and expectations of library users, libraries have to understand what are of greatest value to their users, to redefine their user communities, and to develop their digital collections and services to meet their users' needs and demands. I shall leave discussions on these topics to other presenters in this conference.

Instead, I would like to take this opportunity to approach information seeking and use of information from a more conceptual angle. When talking about seeking information to satisfy our users' information needs, libraries generally follow a linear reference model:

Fined out what one wants -> use the keyword(s) etc… to search library card catalog
-> find the book(s) or publication(s) containing the information.

Many libraries approach the use of digital resources in the same manner. In other words, taking searching the Web as an example, one determines the appropriate term(s) [keyword(s)], search the Internet, go the located web site(s), and find the information in that navigational path. Likewise, digital libraries are developed with this similar retrieval route in mind.

Nothing wrong with this traditional approach, and in fact, this has been very effective for us for many years. But in the current digital environment, we should expand our capabilities far beyond this. We should realize that "digital libraries offer unparalleled access to information for a far broader range of users than prior physical and organizational arrangement." [2] In order to benefit from this digital environment, we need to find out whether we have a scalable, interoperable infrastructure that is able to bridge context, culture, and language, and enable us to gather, organize, utilize and share the rich information resources effectively. We need to change the traditional linear reference model to a new one on use and usability, in which information resources actually "talk" to each other. It is a model which stimulates the users' thinking and learning; redefines user experiences, and gears to a much wider-range and broader-based user groups. In this direction, "information access" is taking on a very different meaning, while data, information, knowledge, users, information services and applications all have much more expanded therefore different definitions than before.

THE EXPERIENCE OF GLOBAL MEMORY NET

Now I am going to shift gear to spend the rest of the time to share with you some of my own experience during this period of dynamic technological transformations since 1983. The presentation will be quite visual, but for the benefit of the readers, I shall provide some summary descriptions.

In the last two decades, I have experienced these transformations up-close and personal. I have created the interactive videodisc called *The First Emperor of China*, in the early 80s – which I brought to Dubrovnik in 1984; then with the coming of digital technology, I have converted the contents of the analog videodisc and authored the digital multimedia CD of the same title in the early 1990s. When my institution was not big enough to be involved in any digital library activities of the first phase of the Digital Libraries Initiative (DL-I) of the US National Science Foundation (NSF), I was busy in building up contents, and develop more complete descriptive information (later known as metadata) of these resources. In 1999, when NSF first introduced its International Digital Library Program (NSF/IDLP), I proposed the *Chinese Memory Net (CMNet)*, which became one of the first NSF/IDLP Projects. Since 2002, *CMNet* has expanded its scope to *Global Memory Net*. Since 2001, I have also been involved in the *China-US Million Book Digital Library Project,* which Prof. Raj Reddy of the Carnegie Mellon University and I served as co-PIs.

Chinese Memory Net (CMNet)

The NSF's supported *CMNet* since 2000 is intended to develop a model for international collaboration with various R&D activities in digital libraries. It hopes to accomplish "more" with "less," avoid duplication efforts, and capitalize R&D results from other funded digital library projects. In the short three years, it has made a great effort in developing collaborative infrastructure for digital library development and actual multimedia digital content development. *CMNet*'s (Figure 2) core contents build upon the large quantity of visual materials of my earlier interactive videodisc (1986) and multimedia CD (1991) products, both called *The First Emperor of China*. To each image included, extensive research efforts were made to provide relevant descriptive data (metadata) with annotations, as well as links to relevant references and texts whenever possible.

Figure 2. The web-based *Chinese Memory Net*

This labor-intensive R&D activity in content and matedata building has paid off because these data has become the attractive basis for a number of exciting and productive technology-oriented collaborative works with computer scientists. I shall highlight a few in the following [8, 9]:

- **OAI research** among Chinese collaborators at Tsinghua University and Peking University in Beijing and Shanghai Jiao-tong University in Shanghai;

- **Intelligent agent and image retrieval** – Collaboration with Prof. V. Soo of National Tsinghua University in Taiwan [10, 11] has yield yielded interesting results in developing similarity matching algorithm for retrieving relevant Emperor images;

- **Semantic sensitive content-based image retrieval** – Collaboration with Prof. James Z. Wang of Penn State University has enabled us to explore the use of semantic sensitive content-based image retrieval (CBIR) technology in web-based retrieval of *CMNet's* Emperor images using the SIMPLIcity technology [12, 13, 14]. The SIMPLIcity CBIR has enriched the retrieval of images in *CMNet* and *GMNet* far beyond the usual traditional

retrieval by metadata fields as the real-time online demonstration will shown. Figures 3-5 are shown here to demonstrate how this kind of retrieval technique can help to stretch the innovative use of a large image resource collection. When a user enters the Emperor Image Database, images will show up randomly as shown in Figure 3. The user can then choose to search either by traditional retrieval method for precise images -- by title, keyword, location, date, description etc. or by CBIR. If latter, for example, one can ask the system to show all the images matching the first image on the upper left, then almost instantly, images of similar color or pattern will show up (Figure 4). On that screen, descriptive information of any chosen image can be provided if asked or several derivatives of the image can also zoomed as shown in Figure 5.

Fig. 3. Random images show when user selects the image database

Fig. 4 (left). Images of the same color, shape and pattern

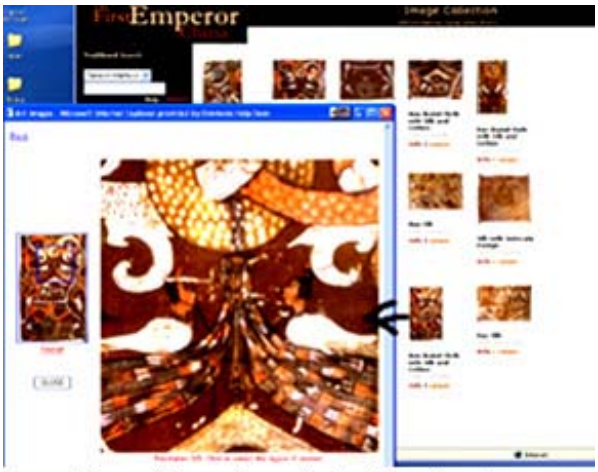

Fig. 5. Chosen image is further enlarged

To overcome the problem related to "copyright" and intellectual property issues, all images larger than the thumbnail size would be displayed with a dynamically generated digital watermark of the image owner. This has facilitated the contribution of image content resources to *GMNet*.

- **Emperor's digital video and Informedia Technologies** – Collaboration between Informedia and *CMNet* has enhanced perspectives from cultural and historical video documentaries. Its multi-lingual (English and Chinese) has also posed challenges in its speech recognition research. Collaboration between Informedia and *CMNet* has enhanced perspectives from cultural and historical video documentaries. Its multi-lingual (English and Chinese) has also posed challenges in its speech recognition research [15]. When the Informedia technology is ready for web-based use, *CMNet* will be ready to explore it. Figure 6 shows some of the screens generated from the latest collaboration (more real-time demo will be given). Upper left shows that when "emperor" is searches, 60 video segments

with that word have been identified and retrieved, these can be visualized in timeline as shown in the lower left. Map is shown in the upper right, and when one of the video is chosen, the video will play in the upper right of the lower right screen, and below that, the actual text will also be displayed with the word "emperor" highlighted in red.

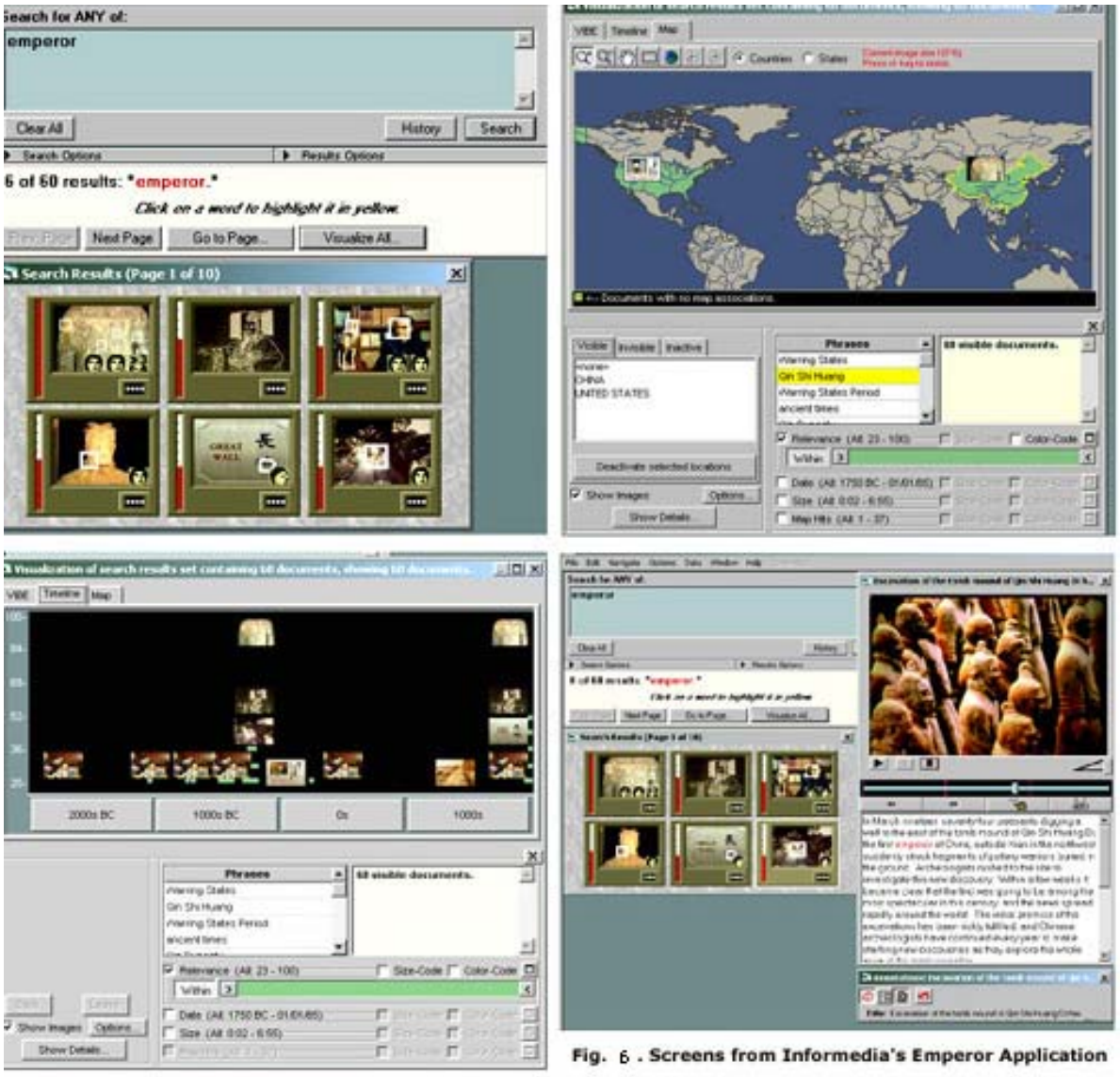

Fig. 6. Screens from Informedia's Emperor Application

Global Memory Net (GMNet)

What we have developed in the three triangular areas in *CMNet* – content building, community building re information providers and users, as well as technologies – have excited a number of researchers and content holders. Thus, from late 2002, *CMNet* has expanded its scope in content coverage to "global", and thus *Global Memory Net*. The first exciting collaborative project under *GMNet* is Project Restore with Prof. Piero Baglioni and his research group at the Center for Colloid and Interface Science (CSGI) of the University of Florence. Currently the Florence group has developed one of the most effective and sophisticated nanoparticle technology for restoring damaged or degraded cultural relics. Detailed information on this nanoparticle

chemistry technology in restoration is summarized in several recent articles [16]. What is worthwhile to mention and demonstrated here is that the invaluable *restoration images of the Baglioni group in* Florence are being organized and presented in *GMNet*, known as *Project Restore*. Figure 7 shows the pre- and post-restoration images retrieved from this project, it is clear that these images together with the associated descriptive information are of substantial significance to art historian and art specialist and students.

Fig. 7. Retrieved Pre- & Post-restoration Images

CONCLUSION

In one of the recent speech given by the Acting Director of the US National Science Foundation, Dr. Arden L. Bement, Jr., entitled "Diminishing dimensions and vanishing boundaries: Endless possibilities," [17] he said:

> *"Our world viewed from a global dimension really is much smaller than in the past. ... International partnerships may be the only way to fund cutting-edge facilities too costly for any single nation, and many disciplines require access to sites in other nations. Data is another realm in which we need to make global exchange the rule rather than the exception."*

He was talking to a group of engineers. But, it has the same meaning to us as library and information professionals. In February, I went to Vietnam, India and Thailand with special "global" lens focused on the potential sharing of cultural and heritage resources on the cyberspace, I was struck with the great potential for meaningful collaborative activities. Now, I see the similar rich opportunities for us here in Croatia and Eastern Europe. *Global Memory Net* has a space holder for *everyone*, and I hope that my presence here can stimulate interests and excitements to some institutions with quality visual resources for meaningful collaboration. Instead of building localized digital libraries, let's work on building up large-scale digital contents together beyond national borders for sharing with the world, and let's also explore together the fertile future for distributed cross-disciplinary collaboration [18].

REFERENCES

[1]. US. President's Information Technology Advisory Committee (PITAC). Digital Library Panel. *Digital Libraries: Universal Access to Human Knowledge.* February 2001. www.hpcc.gov/pubs/pitac/pitac-dk-9feb01.pdf

[2]. *Knowledge Lost in Information: Report of the NSF Workshop on Research Directions for Digital Libraries,* June 15-17, 2003, Chatham, MA. Ron Larsen and Howard Wactlar, eds. Pittsburgh, PA: School of Information Science, University of Pittsburgh, 2004. http://www.sis.pitt.edu/%7Edlwkshop/report.pdf

[3]. Chen, Ching-chih and Kevin Kiernan, ed. *Report of the DELOS-NSF Working Group on digital Imagery for Significant Cultural and Historical Materials*. December 2002.
http://www.delosnsf-imagewg.unifi.it/index.htm
http://dli2.nsf.gov/internationalprojects/working_group_reports/digital_imagery.html/

[4]. Chen, Ching-chih, Howard Wactlar, James Z. Wang, Kevin Kiernan, et al, "Digital imagery for significant cultural and historical materials: An emerging research field bridging people, culture, and technologies," *International Journal on Digital Libraries – Special Issue on EU-NSF Working Group Reports*. 2004 (forthcoming).
http://wang.ist.psu.edu/wang/jodl_img.pdf

[5]. Troll Covery, Denise. *Usage and Usability Assessment: Library Practices and Concerns*. Washington DC: Digital Library Federation and Council on Library and Information Resources, January 2002.

[6]. Chen, Ching-chih, "*Chinese Memory Net (CMNet)*: A model for collaborative global digital library development," *Global Digital Library in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration.* C. C. Chen, ed. Beijing, China: Tsinghua University Press, 2001. pp. 21-32.

[7]. Chen, Ching-chih, "Designing multimedia digital libraries with content development and management: Technology potential and challenges," in *Online Proceedings of DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.

[8]. Chen, Ching-chih, "Global Memory Net: Potential and challenges for archiving and sharing cultural and heritage resources," *Proceedings of the ICDL (International Conference on Digital Libraries) 2004*, Delhi, India, February 25-27, 2004. Delhi, India: TERI, 2004. pp. 3-10.

[9]. Chen, Ching-chih, "Past forward - Digital media for cultural heritage: The case of the Global Memory Net," Invited lecture given at the *10th Annual Lecture of Informatics*, sponsored by Informatics, Bangalore, February 29, 2004.

[10]. Soo, V. W., C. Y. Lee, C. C. Yeh, and C. C. Chen, "Using sharable ontology to retrieve historical images," accepted for presentation at the ACM/IEEE Joint Conference of Digital Libraries, Portland, OR, July 15-18, 2002. *Proceedings of the ACM/IEEE JCDL Conference, 2002.* pp. 197-198.

[11]. Soo, V. W., C. Y. Lee, C. C. Lin, S. L. Chen, and C. C. Chen, "Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques," full-length paper accepted for present at the ACM/IEEE Joint Conference of Digital Libraries, Houston, TX, May 29, 2003. *Proceedings of the ACM/IEEE JCDL Conference*, 2003.

[12]. Chen, C. C. and J. Z. Wang, "Large-scale emperor digital library and semantics-sensitive region-based retrieval," Proceedings of the *Digital Library - IT Opportunities and Challenges in the New Millennium: An International Conference*, Beijing, July 9-11, 2002. pp. 454-462.

[13]. Wang, J. Z., J. Li and C. C. Chen, "Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritages," *Proc. ACM Multimedia, Workshop on Multimedia Information Retrieval*, Juan Les Pins, France, December 2002. 6 pp.

[14]. Wang, J. Z., J. Li and C. C. Chen, "Machine annotation for digital imagery of historical materials using the ALIP system," *DELOS/NSF Workshop on Multimedia Contents in Digital Libraries*, Crete, Greece, June 2-3, 2003.

[15]. Wactlar, H. D. and C. C. Chen, "Enhanced perspectives for historical and cultural documentaries using Informedia technologies," *Proceedings of the ACM/IEEE Joint Conference of Digital Libraries,* Portland, OR, July 15-18, *2002*. pp. 338 – 339.

[16]. Baglioni, P., R. Giorgi and C. C. Chen, "Nanoparticle technology saves cultural relics, and potential for a multimedia digital library," *DELOS/NSF Workshop on Multimedia Contents in Digital Libraries,* Crete, Greece, June 2-3, 2003.

[17]. Bement, Arden L., Jr., "Diminishing dimensions and vanishing boundaries: Endless possibilities," Opening keynote speech delivered at the *Engineering Deans Institute of the American Society for Engineering Education* 2004, New Orleans, Louisians, March 29, 2004. http://www.nsf.gov/od/lpa/forum/bement/alb040329asee_engdean.htm

[18]. Chen, Ching-chih, "Sharing the abundant human knowledge when old cultures meet in cyberspace," *U & I* (Newsletter of Informatics, Bangalore, India), May 5, 2004.

Dr. Ching-chih Chen is Professor and former Associate Dean of the Graduate School of Library and Information Science, Simmons College, Boston, USA. Currently she directs a major US National Science Foundation/International Digital Library Project, *Chinese Memory Net.* Using that as a model for archiving, content building in specifically image and video areas, organization and retrieval, and international collaboration, the project has expanded to "*Global Memory Net*" ready to include valuable cultural and historical resources of the world.

A sought-after international consultant and speaker in over 40 countries, she is an author and editor of over 35 books and more than 150 scholarly journal articles. She was the Chief Conference Organizer of a series of 12 *International conferences on New Information Technology (NIT)* from 1986-2001 in different parts of the world. *NIT '91* was held in Budapest, Hungary, and *NIT '95* in Riga, Latvia.

A Fellow of the American Association for the Advancement of Science, she has received many awards and honors, including the Best Information Science Teacher Award of the American Society for Information Science, the Library and Information Technology Association's LITA/Library Hi Tech Award, the LITA/Gaylord Award for the Advancement in Library and Information Technology, and many others.

Served on the US President's Information Technology Advisory Committee from 1998-2002, she is also the Co-PI of the *China-US Million Book Digital Library Project* with Prof. Raj Reddy of Carnegie Mellon University. Active in the Digital Library area, she is the co-Chair of the 4th ACM/IEEE *Joint Conference on Digital Libraries (JCDL)* of 2004 in Tucson, Arizona. She is on the Advisory Board of DELOS (the European Digital Network of Excellence), US Co-Chair of the *NSF/DELOS Working Group in Digital Imagery for Significant Cultural, Historical and Heritage Materials,* and the Co-editor for the *Journal of Digital Library*'s Special issue on *Multimedia Contents in Digital Libraries*. With these, she is eminently qualified to address the issues involved with potential use and users of digital libraries.