CCDL

“数字图书馆－促进知识的有效应用”国际研讨会

The Proceedings of Digital Library - Advance the Efficiency of Knowledge Utilization

论文集

主编： 詹福瑞

Editor-in-chief: Zhan Furui

主办单位： 国家图书馆

Organizer: National Library of China

科学技术文献出版社

Scientific and Technical Documents Publishing House

Global Memory Net Offers the World Instantly:
Potentials for Universal Access to Invaluable Chinese Contents

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

I am delighted to speak here again. At the 2002 conference, I was asked to speak on "Digital Library Development in the US and Global Reach" [1], but today, I would like to specifically talk about my own *Global Memory Net*, and its potential for enabling universal access to invaluable Chinese contents.

Many of you are well aware of my early interactive videodisc R&D work of the *PROJECT EMPEROR-I* in the mid-1980s and later the multimedia R&D related to the First Emperor of China's famous terracotta warriors and horses in Xian -- supported by the US National Endowment for the Humanities [2]. Two decades have passed since then, yet in technological terms, it has been very long time! In mid-1980s, *PROJECT EMPEROR-I* demonstrated how multimedia technology could change the way we seek, demand, and use information. Almost two decade later, fueled by enormous progress in science and technology, we have come a very long way from the use of interactive multimedia technology in the workstation environment to the global networked environment. We have moved from the use of hardcopy and analog resources to digital content, which users can search, retrieve and use instantly to meet their needs over the global network with no national boundaries. We have also moved from the offering of multimedia content of one specific subject topic to the digital content of all media formats on all related subject topics to the world instantly. We are truly living in a new period of unprecedented opportunities and challenges! So, in this digital era, we have witnessed the exciting convergence of content, technology, and global collaboration in the development of digital libraries [3, 4] with great potential for providing universal information access.

Thus, 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 [4].

In the last two decades, I have experienced much of these transformations up-close and personal through my 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 International Digital Library Program of the US National Science Foundation [6]. Let me take this chance to share with you the *Global Memory Net.*

WHAT IS GLOBAL MEMORY NET?

From PROJECT EMPEROR-I to *Chinese Memory Net*

In the early 80s, the by-product of *PROJECT EMPEROR-I's* is a set of interactive videodisc, called *The First Emperor of China*, content of which later was converted to a popular multimedia CD product of the same title in 1991 and published by the Voyager Company. As my institution was not big enough to be involved in any digital library R&D activities of both the first and second phases of the Digital Libraries Initiative (DL-I and DL-II) 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 [4].

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 major funded digital library R&D projects. Thus, extensive efforts were made to develop collaborative infrastructure with collaborators in Beijing – Peking University and Tsinghua University; in Shanghai – Shanghai Xiao-tong University; Taipei – Academia Sinica, National Taiwan University, and National Tsinghua University; and in the US – Carnegie Mellon University in Pittsburgh, and Penn State University. 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 core content builds upon the large quantity of visual materials of my earlier interactive videodisc (1986) and multimedia CD (1991) products, *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.

This labor-intensive R&D activity in content and metadata building has paid off because these data have formed attractive basis for a number of exciting and productive technology-oriented collaborative works with computer scientists, such as a few listed in the following with more complete reference provided in [5, 6]:

- Open Archive Initiative (OAI) research,
- Intelligent agent and text-based image retrieval [7, 8],
- Semantic sensitive content-based image retrieval [9, 10]
- Digital video using the Informedia technologies [11], and
- Machine learning for annotation [12].

From *Chinese Memory Net* to *Global Memory Net (GMNet)*

Once it is possible to develop a multimedia digital library in one subject disciplinary or for one geographical area, it is upward scalable to include more subject topics and bigger geographical areas. This was the case with the expansion of the scope of *CMNet* to *GMNet since 2002. GMNet* developed out of the *CMNet* project which concentrates on images and video related to China's ancient culture. In recent years, *CMNet* has collaborated with several major institutions in different countries and thus has become a *Global Memory Net*. The current effort in *CMNet* has gone far beyond the Chinese area. It is being expanded to cover the 'memory' of other parts of the globe [4].

As shown in Figure 1, the tentative *GMNet* homepage*, GMNet* literally has space holder for all countries in the world although this tentative homepage has listed only a few continents and countries under each. By expanding *CMNet* to *GMNet*, this expedites the digital library collaborative development and frees the R&D activity from unnecessary logistical delays and inflexibilities. Since there are 200+ countries in the world, there are endless opportunities for digital collection development, digital partnership, and research activities. In addition, the current direction also include the possibilities of *GMNet* serving as a functional multimedia gateway or portal to world invaluable "memory" resources available in all types of resource organizations - libraries, museums, archives, academic institutions, etc.

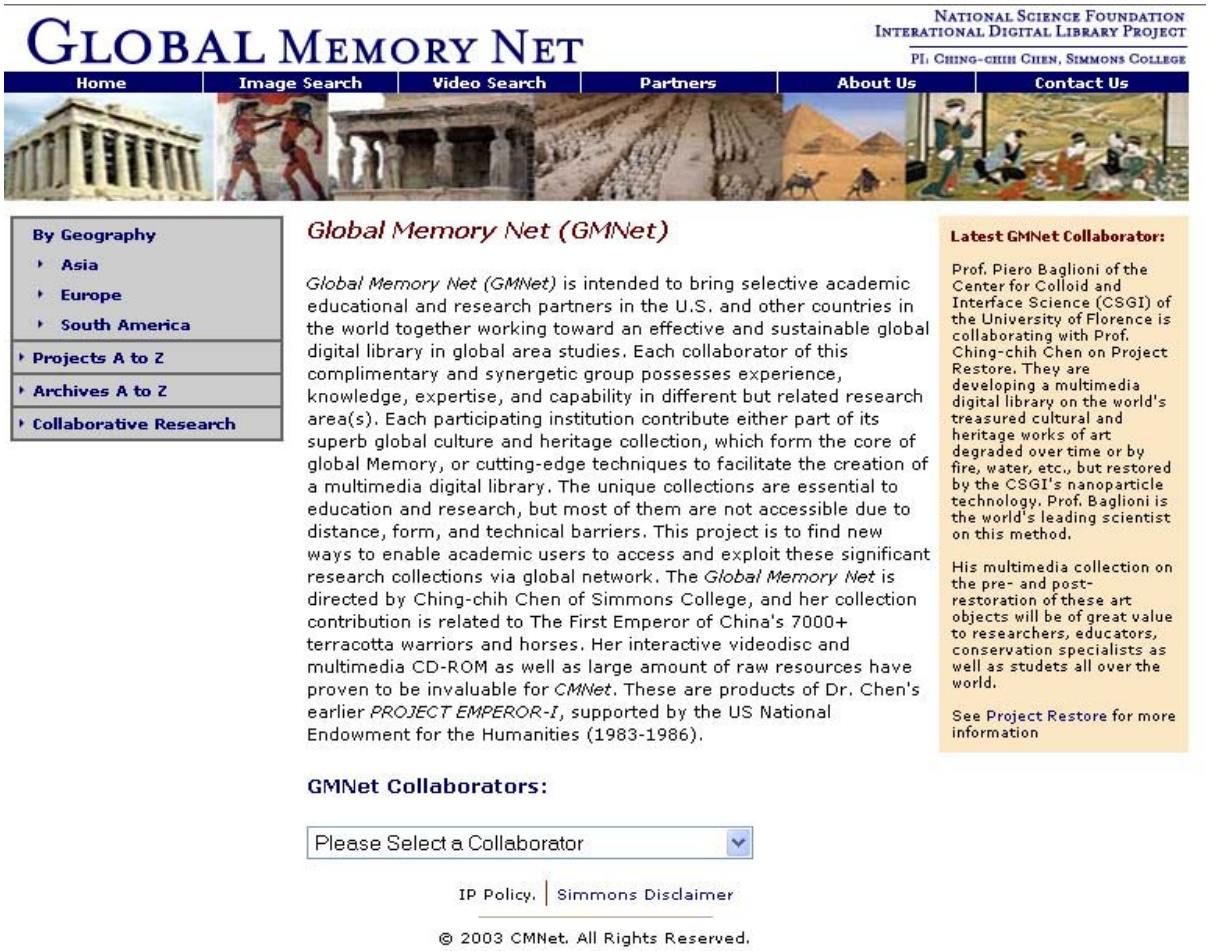

Figure 1. Home Page of the Global Memory Net

WHAT GLOBAL MEMORY NET COVERS?

The name of *Global Memory Net* clearly articulates both the potential coverage and scope of this project:

1. **Global** – It means global coverage. Although Figure 1 is a tentative screen which shows only three continents at the moment.

2. **Memory** – "Memory" refers to all the treasured, thus *GMNet* has the structure to cover all kinds of invaluable memories related to culture, heritage, history, art, music, science, technology, medicine, etc. However, at this initial stage, and with initial entry of the

extensive visual memory related the First Emperor of China's terracotta warriors and horses, *GMNet* is focused on the world significant cultural, historical, and heritage materials. Once this focus is well underway, *GMNet* will expand to cover other kinds of "memories."

3. **Net** – This means that *GMNet* hopes to network all significant world resources together. Instead of encouraging the development of small and fragmented digital libraries, it hopes to be a networked portal to offer needed resources instantly with the simple click of the mouse. It is clear now that *GMNet* should not try to have the world valuable "memory" collections in one digital location. Instead, it should try to bring distributed collections around the world together. Thus, from the few sample images, *GMNet* will link the users to the desired image collection instantly.

HOW GLOBAL MEMORY NET OFFERS THE WORLD!

GMNet supplements well another major digital project, *China-US Million Book Digital Library Project* (hereafter referred to as the *Million Project*), of which Dr. Raj Reddy and I are US co-PIs. The US-NSF supported *Million Project* is in collaboration with over a dozen Chinese higher education and research institutions which are supported by the Chinese Ministry of Education. It is a major digitization project which currently is mainly text-based.[1] Yet, *GMNet* concentrates on multimedia resources – images, videos, music and other audio resources. Currently, it is mainly related to digital image resources. For this reason, for this presentation, I shall mainly offer examples related to the cutting-edge content-based image retrieval with limited mention of the digital video potential. Although *GMNet* includes valuable image collections from many countries in the world (some are shown in Figure 2), this paper will concentrate on Chinese contents.

Figure 2. Some sample cultural and heritage contents in *GMNet*

In the simplest way, just imagine taking a visual tour of a country's culture, heritage, history, and world contributions, all while sitting at your computer. This soon to be available *GMNet* on the Internet will provide image retrieval capabilities with considerable textual supports in a way not possible before. For example, from the page like that shown in Figure 1, one can go to China and then Emperor Image Base quickly. Then one will be able to retrieve

[1] *Million Project*'s Chinese partners include the Phase I institutions -- Chinese Academy of Sciences (Northern center), Zhejiang University (Southern center), Peking University, Tsinghua University, Fudan University, and Nanking University; and Phase II institutions – Beijing Normal University, Jilin University, Shanghai Jiao-tong University, Chung-san University, Wuhan University, Chengdu University, and Xian Jiao-tong University.

invaluable images related to the First Emperor of China, for example, by conducting the traditional search using Google protocol if predefined specifics of the images are known. In this case, one can search literally every field of the metadata, such as creator, title, location, time period, description, keyword, reference source, etc. In this approach, keyword search is likely to be the most popular one. Thus, if "keyword" search is selected, and the search terms are types as (+tile, +end) (in the Google search terms with "+" indicates that both "tile" and "end" are required), then, almost instantly from the thousands of images in the image base, the screen of Figure 3 will show the first 10 tiles ends located with the total indicated. In this type of searches, precise retrieval of available images is made.

Figure 3. 10 roof-tile ends retrieved by using traditional keyword search terms - [+(roof tile), +end]

However, in most cases, one does not have any idea on what kind of images are available in *GMNet*. Most image databases do not offer the users the chance to browse. Yet, in the case of *GMNet*, one can use the cutting edge content-based image retrieval technique, SIMPLIcity, to enable her/him to browse, retrieve, enjoy, and learn in just seconds through multiple thousands of digital images. As shown in Figure 3, under the "Traditional Search" on the left panel, there are three buttoms which provide the users these possibilities. They are:

- Random – by clicking on this, images in the image base will show up randomly;
- Browse – by clicking on this, users will be able to browse images 10 or 15 at a time from page to page until they spot the desired image.

- URL – the user will be able to ask the system to find images that are similar to the one located on the given URL address.

Figure 4. Random images for user's browsing and selection

Until recently, most archival images were not available in digital form. Now we have a large quantity of invaluable digital materials from multiple countries, together in one place. One can ask the system to bring out image icons randomly, or to browse the images by displayed icons page to page until one locates the image of interest. For example, when the icons of the mages of the Emperor collection are displayed randomly in Figure 4, one spots the image related to a map of interest. In this case, one can ask the system to provide "SIMILAR" images by clicking "Similar" without typing any word, *GMNet* will display in seconds all the maps in the collection similar to the one selected. This opens up all possibilities for all related maps which are totally unknown to the user prior to the showing (see Figure 5). Once these massive numbers of images are displayed, one will be able to find instantly more textual descriptive information as well as reference sources and in some case, full-text descriptions on a chosen map by clicking "Info". If the chosen image needs to be enlarged, then click on "larger", and multiple levels of zooming will be possible to show the desired details of the map or a portion of the map. Concurrently, dynamic digital water mark will be instantly generated at any zooming level to

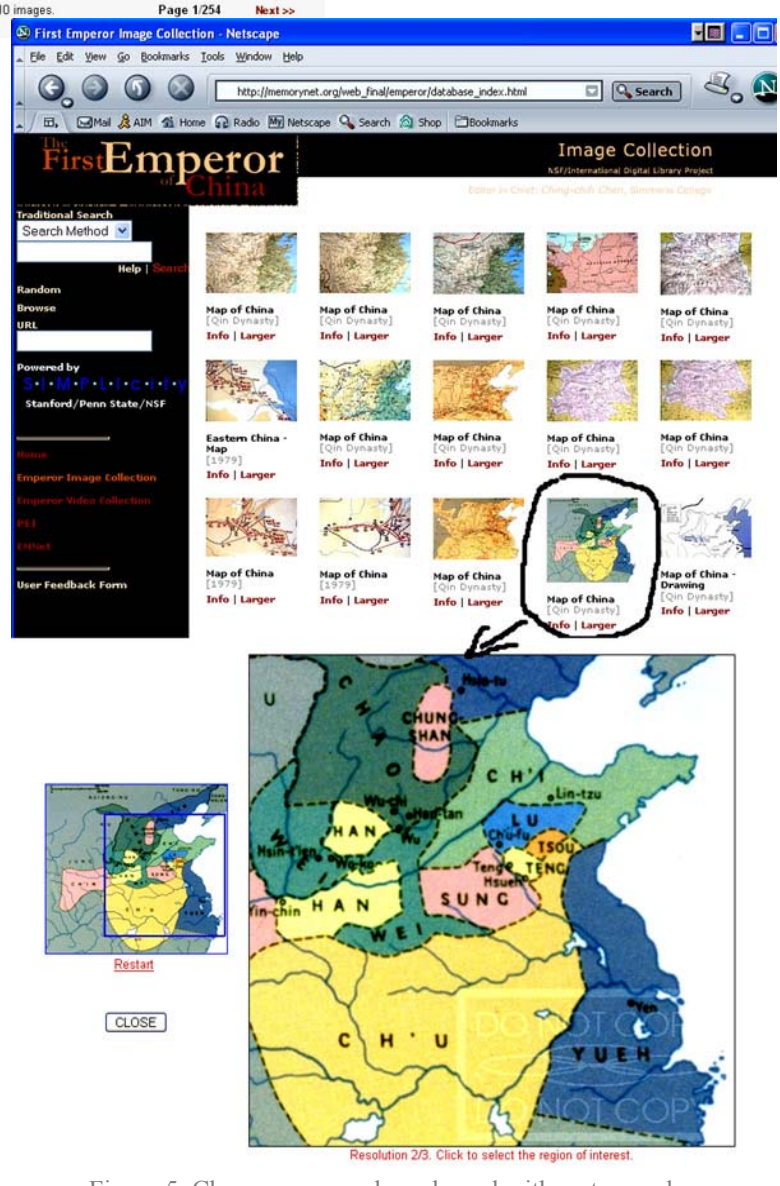

Figure 5. Chosen map can be enlarged with water mark

offer the "ownership" information of the image. This also overcomes the problem related to copyright and intellectual property.

POTENTIAL FOR DELIVERY INVALUABLE CHINESE CONTENTS

Figures 4 and 5, shown above, should reveal vividly the potential of *GMNet* for delivering of invaluable Chinese memory image contents to the world. Additional examples are giving in the following by providing content-based image retrieval from the painting collection of a contemporary Chinese artist, Mr. Hu Boxiang, for his 100th Birthday Celebration [13] as shown on the left cover. The inclusion of this art collection is possible at the courtesy of his children, particularly his daughter, Prof. H. C. Hu.

Figure 6 shows some of the random images of Hu's painting collection. Of these images, if one spotted a "fan" painting in the lower right corner and if the user chooses that image and ask for all similar images from the collection, *GMNet* will display instantly the four "fan" paintings from Hu as shown in Figure 7.

Figure 6. Hu Boxiang's painting showing in random

(Courtesy of Prof. H. C. Hu of Beijing, daughter of the artist)

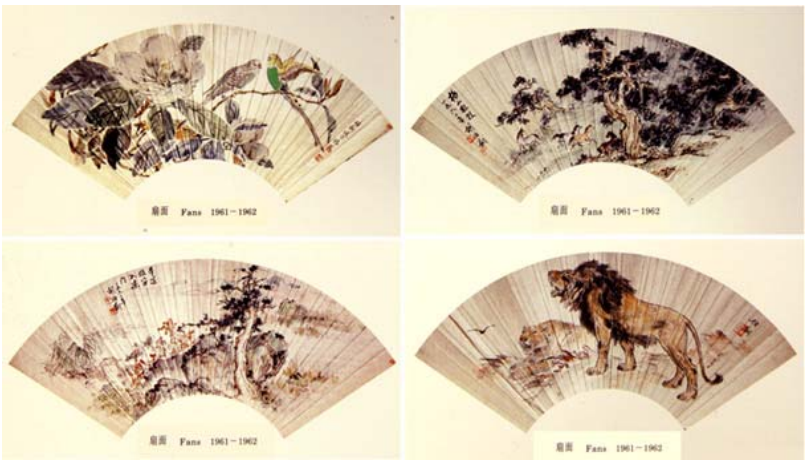

Figure 7. 4 Fan Paintings of Hu Boxiang

As one of the oldest countries in the world, China's 5000-year cultural, historical and heritage resources are not only rich but also abundant. Thus, millions and millions of cultural resources have been passed on from one generation to the other. Yet, most of these rich resources remain unknown to the world and certainly difficult for those outside China to gain easy access to them. *GMNet* can be a perfect channel to help to deliver these invaluable resources to the world. It is important to note again that the dynamic digital water marks will automatically appear when these images are shown in the size larger than the thumb nails. This not only will protect the intellectual properties of the creators and/or owners of the artifacts, but also will discourage any illegal copying of the images.

Once the information of a desired artifact is located, *GMNet* will link the information seekers to the original source of the images and information. Thus, *GMNet* essentially provide effective marketing for these invaluable resources, and enhance universal access.

FUTURE DEVELOPMENT

Currently *GMNet* is engaging in numerous exciting collaborative projects including the University of Florence on *Project Restore*. Pre- and post-restoration images and descriptive information of restored Italian artifacts will be posted on the Web, including wall paintings damaged from the 1966 Florence flood. This world renown nano-partical technology for physical restoration of damaged art objects developed by our collaborator, Prof. Piero Baglinoi and his research group at the Center for Colloid and Interface Science (CSGI) of the University of Florence should have great potential for the restoration of Chinese artifacts. Figure 8 shows how one badly damaged artifact is miraculously restored to its appreciable form with this technology [14].

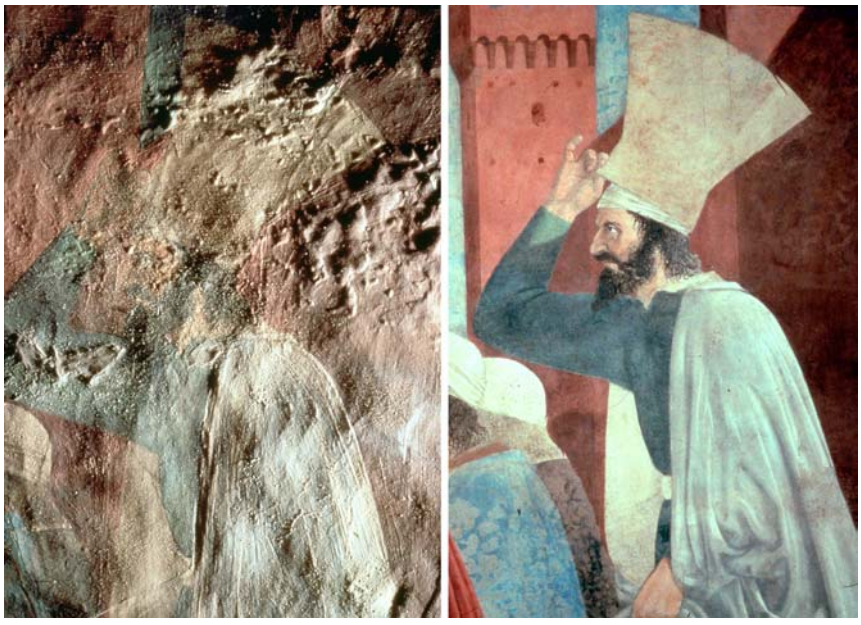

Fig. 8. Badly damaged Italian artifact restored (Courtesy of Univ. of Florence)

In addition to continuing image collection and global partnership building, future development will move more aggressively to the areas of digital video, sound and audios.

Digital Video and Informedia Technologies **[6]**

Carnegie Mellon University's well-known Informedia Project is one of the six original NSF/DLI-1 projects. It has continued its further development in digital video related technologies and tools ever since 1995. 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 [11]. When the Informedia technology is ready for web-based use, *GMNet* will be ready to explore it. Figure 10 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.

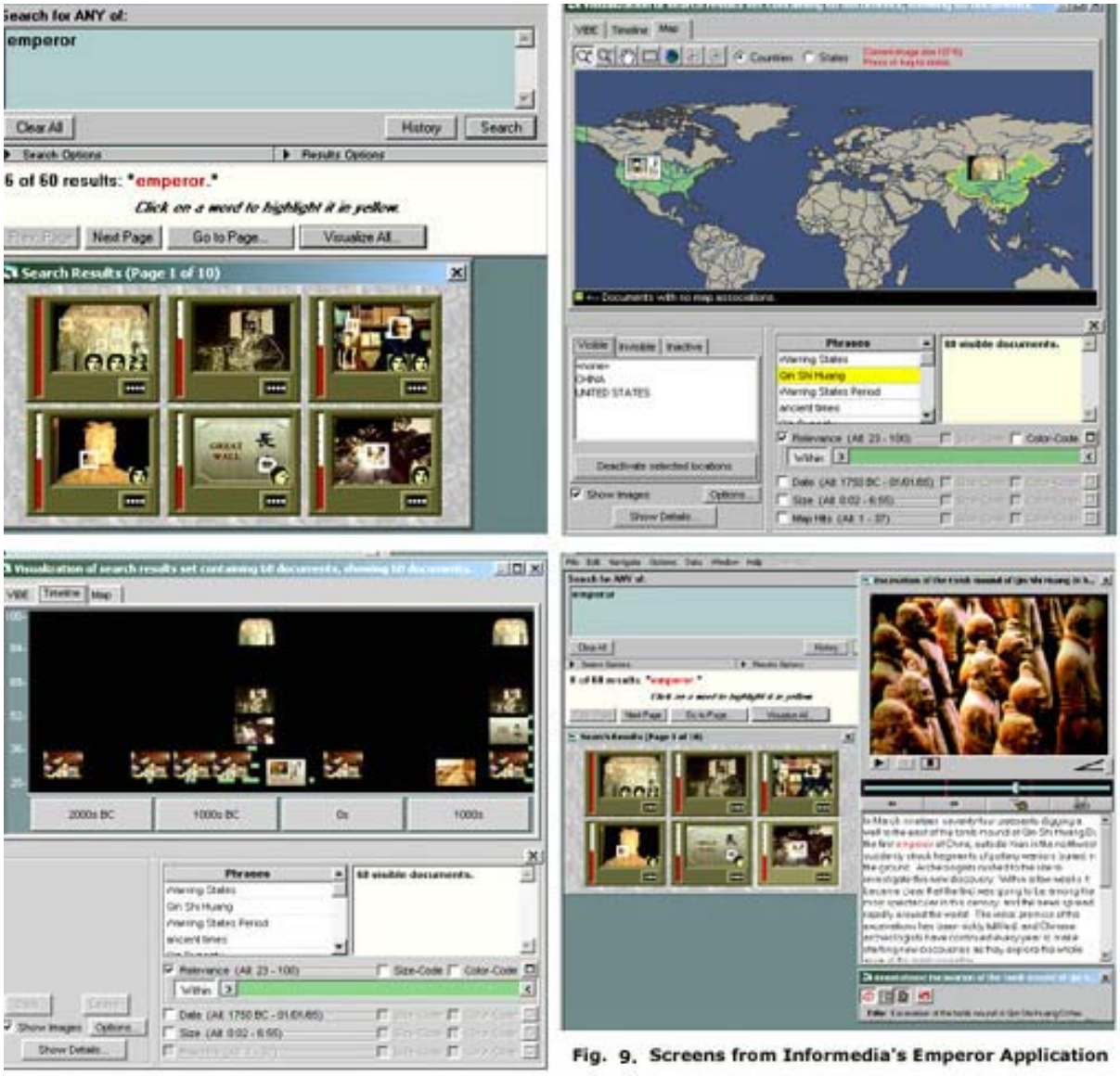

Fig. 9. Screens from Informedia's Emperor Application

In addition to this cutting-edge application, more traditional retrieval of digital videos and audios will be explored and provided. Many other R&D activities including multi-lingual research will also be explored.

For more information on the project before the website is available for public use, visit www.memorynet.org

CONCLUSION

During 1997-2002, I was privileged to serve on the US President's Information Technology Advisory Committee (PITAC). Our PITAC's Digital Library Panel's Report, *Digital Libraries: Universal Access to Human Knowledge,* has a vision for digital libraries:

> *"All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge. ... In this vision, no class-room, group, or person is ever isolated from the world's greatest knowledge resources."* [15]

This is a vision easily said than done! There are many obstacles on the road, thus we are a long way from approaching this "elusive" vision. Yet, we have to keep on working toward this ultimate goal. This is what both the *Million Project* and *Global Memory Net* have inspired to reach. We need first to have our invaluable resources available in digital form (by doing so, we also preserve these sources digitally), then we need to think of ways to make these resources known to the world, to provide quick and easy access to them, and to share them. *Global Memory Net* has demonstrated how international collaboration and community building in promoting large-scale content building, coupled with new technological tool and method development, can indeed offer users the world in a way not possible before. The potential for delivering and marketing invaluable Chinese resources should also be clear.

ACKNOWLEDGMENTS

Chinese Memory Net and *Global Memory Net* are supported by the US NSF/IDLP under grant no. IIS-9905833. PROJECT EMPEROR-I was supported by the Humanities in Libraries Program of the US National Endowment for the Humanities. Examples given on the use of SIMPLIcity is in collaboration with James Z. Wang of the Penn State University whose work is supported by the NSF/ITR program under grant no. IIS-0219271.

REFERENCES

1. Chen, Ching-chih, "Digital library development in the US and global reach: Roles of federal government agencies and PITAC," *Proceedings of the Digital Library - IT Opportunities and Challenges in the New Millennium: An International Conference*, Beijing, July 9-11, 2002. Beijing, China: National Library of China, 2002. pp. 17-27 (Chinese version). pp. 28-41 (English version).

2. Chen, Ching-chih. *The First Emperor of China*. Santa Monica, CA: The Voyager Company, 1991. This is a videodisc product and is the popular version of this title [A

Cindy Award videodisc product]. Chen, Ching-chih. *The First Emperor of China: A multimedia digital CD*. [Mac and Windows dual version]. New York: The Voyager Co., Septebmer 1994. *MacUser* (October 1994) selected this as one of the "50 Best CDs of 1994".

3. Chen, Ching-chih, "The promise of international digital library collaboration for innovative use of invaluable resources," in *Human Information Behaviour & Competences for Digital Libraries,* Keynote in *Proceedings of the Libraries in the Digital Age*, May 25-29, 2004, Dubrovnik and Mljet, Croatia. pp. 7-15.

4. Chen, Ching-chih, "*Global Memory Net* Offers Users the World Instantly," *Library Times International*, 21 (1): 1-4 (July 2004).

5. 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.

6. 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, India, February 29, 2004.

7. Soo, V.W., C.Y. Lee, C.C. Yeh, and Ching-chih Chen, "Using sharable ontology to retrieve historical images," *ACM/IEEE JCDL Proceedings, Portland*, OR, July 15-18, *2002*. pp. 197-198.

8. Soo, V.W., C.Y. Lee, C.C. Lin, S.L. Chen, and Ching-chih 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 pages.

9. Chen, Ching-chih and James 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: International Conference*, Beijing, July 9-11, 2002. pp. 454-462.

10. Wang, James Z., Jia Li and Ching-chih Chen, "Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritage," *Proceedings of the. ACM Multimedia, Workshop on Multimedia Information Retrieval***,** Juan Les Pins, France, December 2002. 6 pp.

11. Wactlar, Howard D. and Ching-chih 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.

12. Wang, James Z., Kurt Grieb, Ya Zhang, Ching-chih Chen, Yixin Chen, and Jia Li, "Machine annotation and retrieval for digital imagery of historical materials." Submitted to *International Journal on Digital Libraries*' *Special Issue on Multimedia* Centents, 2004.

13. *A Collection of Hu Boxiang's Painting: Memory of Hu Boxiang's Centennary Birthday Celebration.* Shanghai: Shanghai Publishing House of Calligraphy and Painting. January 1997.

14. Baglioni, Piero, Rodorico Giorgi, and Ching-chih Chen, "Nanoparticle technology saves cultural relics: Potential for a multimedia digital library," presented at the *DELOS/NSF Workshop on Multimedia in Digital Libraries,* Chania, Crete, Greece, June 2-3, 2003. http://memorynet.org/pdf/baglioni_crete.pdf/

15. 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

Dr. Ching-chih Chen is Professor of the Graduate School of Library and Information Science, Simmons College, Boston, USA. A sought-after international consultant and speaker in over 40 countries, she is an author and editor of over 35 books and more than 180 scholarly journal articles. She produced the award winning interactive videodisc and multimedia CD entitled, *The First Emperor of China*. 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. The Proceedings of *NIT 2001*, held at Tsinghua University, Beijing, was published as *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration* by Tsinghua University Press in 2001.

Since 1993, she has been advocating the global digital library concept by linking libraries, museums and archives all over the world together, and this Global Digital Library Initiative has helped the development of digital libraries in numerous countries. Since 2000, she has led a NSF/International Digital Library Project, *Chinese Memory Net (CMNet)*. She is also the co-PI with Prof. Raj Reddy of the China-*US Million Book Digital Library Project.* She is a member of the Advisory Committee of DELOS (European Digital Library Network) and co-Chaired the *DELOS/NSF Working Group on Digital Imagery for Significant Historical, Cultural and Heritage Materials*. She has been advocating the need for international consortium in making cultural and heritage digital contents accessible to users. To this end, *Chinese Memory Net*, serving as a model for archiving, content building in specifically image and video areas, as well as international collaboration, has grown now to be *Global Memory Net*, with collaborators from different part of the world.

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. During 1997-2002, she served as a member of the US President's Information Technology Advisory Committee.

A sought after international speaker, in the last six month alone, she was a keynote speaker at the *International Conference on Digital Libraries in Delhi, India*; the Invited Annual Lecturer of the *Annual Lecture in Informatics in Bangalore*, India; the invited speaker at the University of Mysore, Mysore, India; a keynote speaker at the *Libraries in the Digital Age (LIDA 2004)*: International Conference, Dubrovnik and Mljet, Croatia; and an invited speaker at the ACRL Asian and African Division of American Library Association, and the Chinese American Librarians' Association Meeting in Orlando, Florida.