
Date 2nd version : 13/07/2006

# Using Tomorrow's Retrieval Technology to Explore the Heritage: *Bonding Past and Future* in the Case of Global Memory Net*

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

| | |
|---|---|
| **Meeting:** | **97 Information Technology with Audiovisual and Multimedia and National Libraries (part 2)** |
| **Simultaneous Interpretation:** | No |

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL

**20-24 August 2006, Seoul, Korea**

http://www.ifla.org/IV/ifla72/index.htm

## *Abstract*

*Since 2000, Global Memory Net (GMNet), supported by the US National Science Foundation/ International Digital Library Program, has developed to be an effective World culture and heritage Image Library and portal which provide instant access to the world's invaluable cultural, heritage and historical resources of libraries, museum, and archives.† It has its over 30 rich digital image collections in house as well as provides instant access to over 2400 world digital collections from over 80 countries at its reach.*

*Global Memory Net has been developed with a strong conceptual framework developed since 1993 and refined in the late 1990s and has utilized cutting-edge image retrieval technologies for providing innovative user-oriented information services. This paper focuses on a range of new and exciting technical developments to implement aggressive user-centered concepts and strategies for providing the kind of multi-format and multilingual access to multimedia resources not possible before.*

*Modified from a keynote speech entitled "New Digital Library Perspectives from International Library and Information Point of View," delivered at the *Annual Meeting of Library Directors of Universities and Colleges in Taiwan*, Tainan, May 19, 2006 [Chen, 2006a], and an invited speech entitled "Challenges for Developing a World Digital Library and Gateway: The Case of Global Memory Net," delivered *at Library in the Digital Age (LIDA 2006*), Dubrovnik, Croatia, May 31, 2006 [Chen, 2006b].

† See http: //www.memorynet.org/. Global Memory Net was launched for universal access in late June 2006.

## INTRODUCTION

Although since 1993, I have advocated the concept of a world digital library [Chen, 1993], yet my own serious R&D work related to the development of digital libraries did not start until the end of the 90s. As we entered an exciting digital era in the new millennium, we have begun to witness the exciting convergence of content, technology, and global collaboration. There are unprecedented potentials as well as challenges for developing digital libraries of all kinds. Fully realizing the importance of a triangular relationship between content, technology and user as shown in Figure 1, this paper will focus on the technical aspect of the challenges for developing a world digital image library and portal [Chen, 2006a, 2006b].

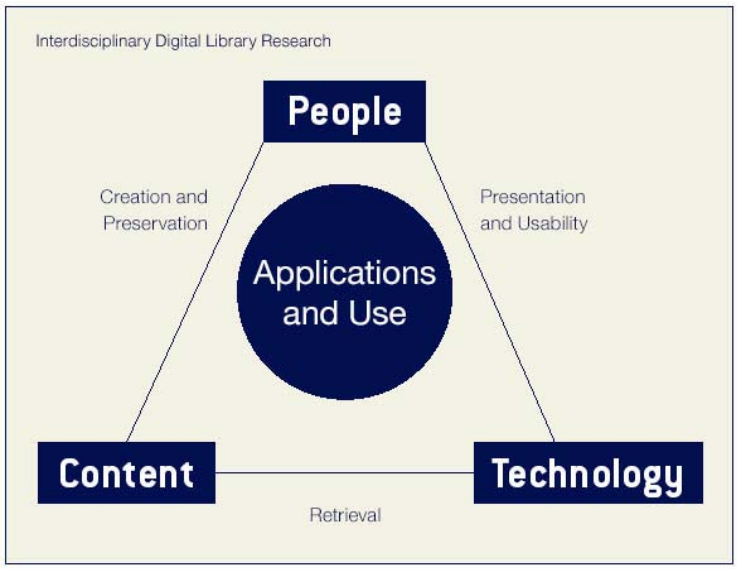

Figure 1. Conceptual Model of an Interdisciplinary Digital Library

## CHALLENGES FOR DEVELOPING A WORLD DIGITAL LIBRARY

In the case of Global Memory Net, since it is intended to be both a world digital image library and a gateway, we have conceptually set various requirements as our challenges for such digital library development. The following are some of the selected ones [Chen, 2006a, 2006b]:

- Instead of using web for publishing, we need to use web as a platform to enable more user participation,
- We need to provide seamlessly integrated multimedia information services in order to enrich user experience,
- Our system needs to provide innovative information services, and it is not to offer packaged software for building databases,
- User's need and behavior in information seeking should not be pre-determined, thus the search and retrieval capabilities have to be able to accommodate all kinds of needs,
- User needs to control the use of data, thus he/she can choose to search one single collection or multi-collections,
- What we offer should be a knowledge base and not database(s), therefore once simple information is found, user can and should be enriched with more related multimedia information, and mixable or re-mixable resources if desired,
- We need to provide freely accessible and searchable web resources, and not just links with them,
- To enhance user experience and ability to consume the obtained information, multilingual capabilities are essential for both presentation as well as retrieval,
- User should be able to lead from one tiny useful information to play big and link that to the world collections as well as world bibliographic and web resources,

- User should be able to address the content in any granularity,
- Geographical access to content should be provided,
- Intellectual property of the content provider should be protected,
- User should be able to use the discovered and retrieved data to develop his/her own project(s),
- User should be able to actively contribute to the library,
- Etc.

To meet these challenges, in the case of Global Memory Net (www.memorynet.org), we have developed the latest version of the ***i***nteractive **M**ultimedia **C**ontent Retrieval **S**ystem (tentatively called *i*-M-C-S) by incorporating many innovative concepts and functions. This paper will articulate how these challenges are being addressed. They include some of those listed in the following:

1) Using the Web as a platform to integrate seamlessly all types of multimedia resources.
2) Digital images can not only be retrieved in traditional ways by metadata fields, but also by cutting-edge content-based retrieval methods (CBIR).
3) Once the desired image(s) is (are) located, associated sound, digital video, textual annotation, as well as expanded bibliographical and Web resources can be obtained by the simple clicks of the mouse.
4) Available resources can be retrieved by searching a single collection, or by cross-collection (or multi-collection) search.
5) Multilingual presentation as well as retrieval should be possible.
6) Geographical retrieval should be provided.
7) While universal access is the goal, the users should be possible to contribute their resources as well.

## THE DEVELOPMENT OF GLOBAL MEMORY NET'S *i*-M-C-S SYSTEM

The ***i***nteractive **M**ultimedia **C**ontent retrieval **S**ystem (*i*-M-C-S) has been developed in house with its beginning as a Linux/MySQL/PHP-based system with a functionality scheme as shown in the following:

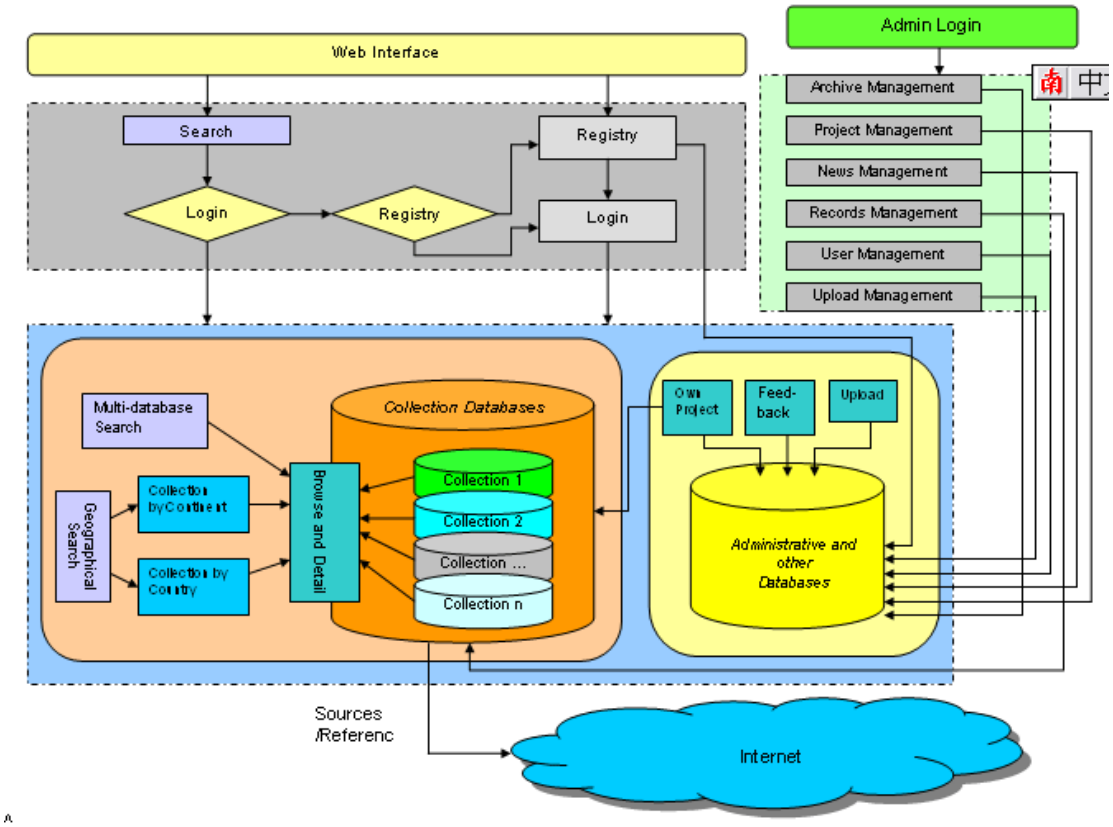


Figure 2. Functionality Scheme of Global Memory Net [Zhang and Chen, 2005].

As time goes on, more and more functionalities are added, and currently the latest version (Versions 4.5) can address literally all the challenges listed above. We shall elaborate on these features in the following section.

In addition, realizing that more content collaborators would want to develop their own digital image collections, we have developed a generic version of *i*-M-C-S to enable our content collaborators to develop their own digital collections easily by simply plug in what is prompted after the basic image database has been created. This was tested in Croatia at a workshop offered at University of Zadar after LIDA 2006 conference [Chen and Badurina, 2006], and it was proven to be very successful. Furthermore, this generic version of *i*-M-C-S is a 3-tier system, which is developed to meet the needs of interested organizations depending on the technical sophistication and capabilities of each of them. The 3 tiers are:

- **Tier One**: This system has almost all the systems features of the current Global Memory Net. It required a server, with staff possessing knowledge and expertise in Linux, PHP, and MySQL etc.
- **Tier Two**: This system will use PC as a server, and the functionality is reduced from those of Tier One due to PC's functionalities.

- **Tier Three**: This is the lowest level which will start an institution with much less technical capabilities to start developing its digital collection database(s), knowing whether that all digital library applications require functional digital databases.

## FEATURES OF GLOBAL MEMORY NET [Chen, 2006a, 2006b]

To elaborate on what was discussed above, while it is impossible to cover all features, we shall select a few in the following with appropriate illustrations:

- *Web is used as a platform*, and not as a publishing medium.

  Figure 3 shows how Global Memory Net is available to universal access by using the Web as a platform.


Figure 3. The Home Page of the Global Memory Net.

- *Instant access to rich image collections* – One can access to over 30 image collections in GMNet with over 20,000 images as well as over 2400 digital collections from over 80 countries. Although GMNet's own current collections have focused in culture, history, and heritage, the World Digital Collection in GMNet has included over 2400 digital collections cover all subject areas. This includes over 100 collections from UNESCO's Memory of the World, over 290 collections from the US Library of Congress, as well as those from other major national libraries, archives, museums, academic institutions, etc. In other words, the world's rich resources are instantly accessible at a simple click of the mouse. Figure 3 shows how these collections can be accessed by selecting the collection from the left blue panel of "Collections". This panel is enlarged in Figure 4 in order to be able to

read the details. Although the panel shows only 9 major collections, but many more can be accessed by clicking "more".


Figure 4. The Navigational Panel Showing How GMNet Collections Can Be Accessed →

- *Easy and flexible traditional search* – Select a collection of interest, and use traditional search by any or all of the metadata fields when one knows the precise information to search in this collection. In this case, the Google syntax is used.
- *User can explore the unknown collection and learn its coverage* – For an unknown collection of curiosity to the user, it is impossible for one to search either by author, title, subject or keyword as normally requested. In this case, we provide the capability for the user to browse the collection. But more significantly, the "random" feature permits one to explore and learn the coverage of this collection. In seconds, one will know considerably more about the collection though the variety of images displayed as well as the words showing up for the titles. These words can then be used as possible keywords for further retrieval. For example, Figure 5 shows the great diversity of images covered in the UNESCO's Memory of the World collections. A quick glance of this screen, one can have a good idea on the subjects involved. Should there be a topic of interest, then one can move to the next step. Otherwise, further random access will yield more topics of coverage.
- *Instant retrieval of similar images of interest by using cutting-edge retrieval techniques* – Once a picture of interest is spotted, one is given three choices – 1) to find pictures of the same color or shape using the cutting-edge content-based image retrieval (CBIR) or the same topics, or 2) to zoom the image for larger sizes and more details, or 3) to have some descriptive information, as shown in Figure 6 on the right. Our CBIR uses SIMPLIcity of Prof. James Z. Wang of Penn State University. For example, if the image shown on Figure 5 related to "Curing Diseases" is of interest, one will be able to ask for more information as shown in the enlarged illustration shown in Figure 6. The user can choose to click on any of the three choices for much more information.

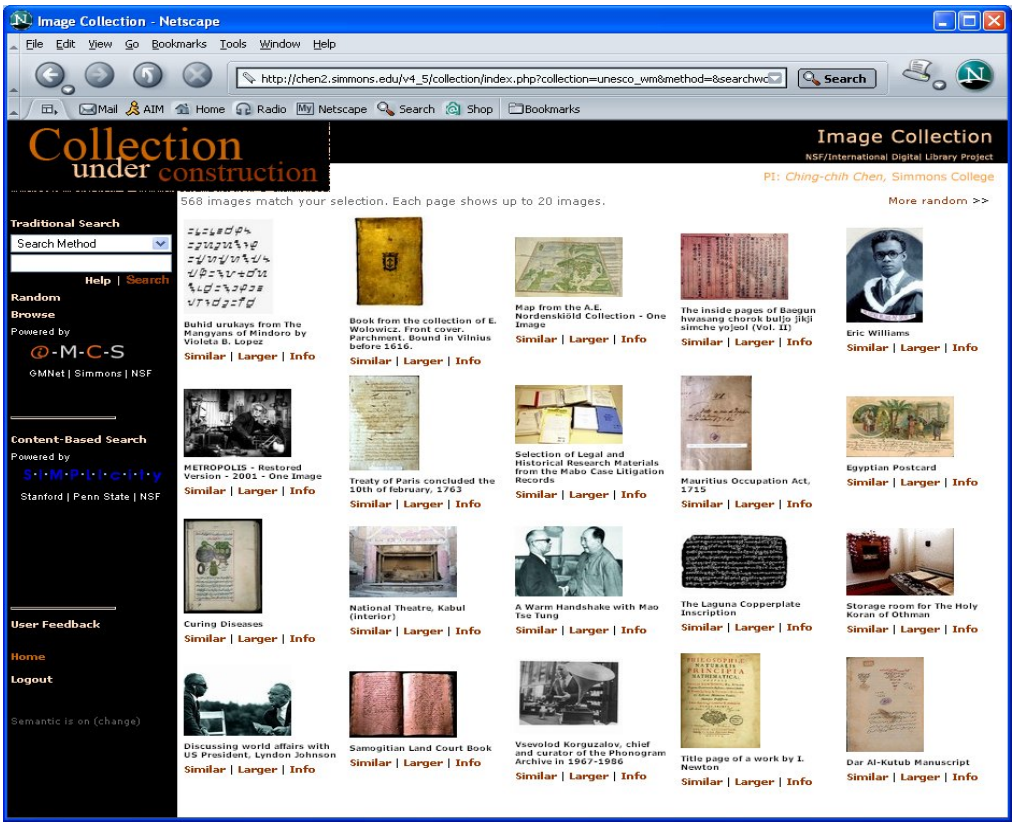


Figure 5. For UNESCO's World Memory, "Random" feature provides an instant display of a great diversity of topics covered

- "Similar" – A simple click on "Similar" will yield all the images of similar color and shape through the use a content-based image retrieval technique, SIMPLIcity. Or, we can also retrieve images of the similar titles or keywords if SIMPLIcity is not used. This is an important feature which is able to provide instantly a large number of similar images to the user without having to type anything on the keyboard. For someone who knows nothing about the subject or the inclusion of the collection, undoubtedly this opens up a new horizon for learning to the user.


Figure 6. The Chosen Image on "Curing Diseases"

- "Larger" – By clicking this, instant magnification of a chosen image in defined areas is shown. Depending on the resolution of the image, some images can be zoomed over a dozen of time. Note the small icon of the rare Japanese Inoh Map of the Library of Congress' Asian Division (Figure 8). All image icons are so small that nothing can be seen. Yet, any of these tiny image icons can be magnified to 12 times with a great deal of invaluably significant information to the scholars.

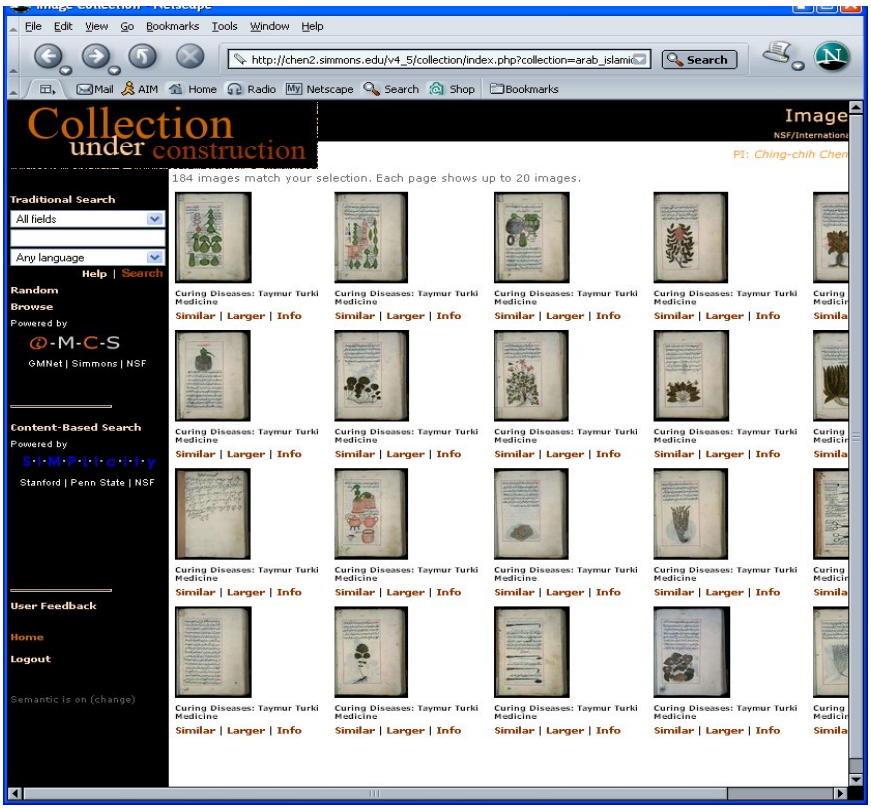


Figure 7. Instant display of all similar images of the same color and shape

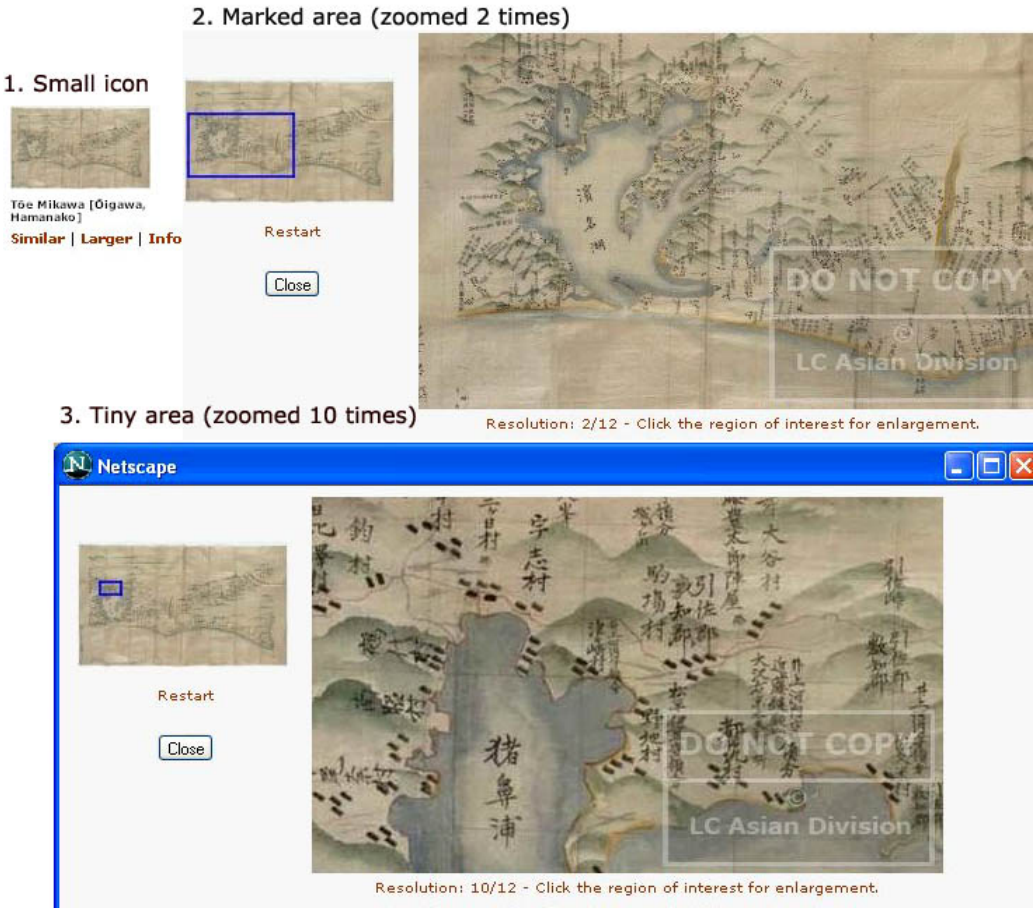


Figure 8. This figure shows the multiple levels of magnification.
Note also the dynamically generated digital watermark.

- "Info" – "Info" will yield instant descriptive information about the image, as shown in Figure 9 for the image on "curing diseases", and Figure 10 on a Japanese Waka poem. This information can be in multilingual forms. In addition, a simple click on the URL of the source, the user will go immediately to the Web page so he/she can browse and search information there.


Figure 9. Descriptive information (metadata) of the Islamic "curing diseases" image available in both English and French


Figure 10. Descriptive information (metadata) of a Japanese Waka poem image from the Tsurumi Collection available in both English and Japanese.

- *Dynamically generated digital watermark for copyright protection of the content provider* – See Figure 8 above. This provision has encouraged more content providers to feel comfortable in providing their treasures in digital form on the Web.

- *Seamless integration of multimedia information* – If relevant resource information on a chosen image is available in formats other than textual annotation, the user can then retrieve the relevant audio, video, etc. again by a single click of the mouse, as shown in Figure 11.

Figure 11. The "Info" screen shows the availability of 3 digital videos and 2 PDF document files for instant retrieval. Some also include sound and other files

- *User can choose to conduct either a single collection or multi-collection search* – For example, if one wants to find information on Naxi, one of the 56 minority ethnic groups in China, one can choose to search Library of Congress' Naxi Manuscripts Collection, or one can do multi-collection search. The later will search all the 30 some collections in GMNet as well as the 2400 digital collections in the world. The search results will instantly be shown (see Figure 12). We can see that, in addition to 1027 images found in LC' Naxi Manuscripts, one can also locate additional resources and images in other collections, such as:
  - 2 Web sites in the World Digital Collection,
  - 3 images in UNESCO's World memory,
  - 9 images in World Musical Instruments,
  - 10 images in Chinese 56 Ethnic Groups.

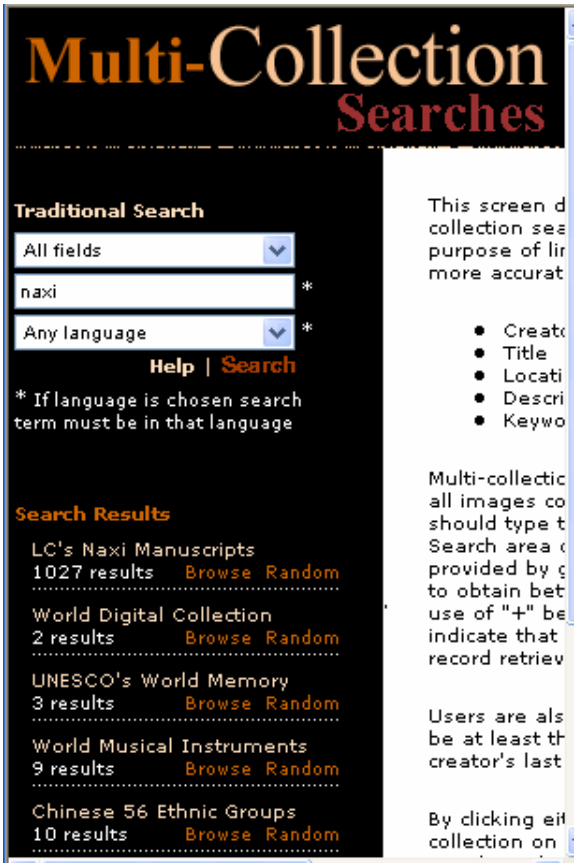


Figure 12. Display of multi-collection search results

- *User can have much more in-depth learning of the subject by using expanded bibliographical and web resources* – For example, as shown in Figure 13, if one is interested in finding more information on an image related to "Timbuktu" such as books articles, etc., one can instantly find relevant books from the half-billion bibliographical records of OCLC/World Cat, or web resources like Google Scholar, Wikipedia, Internet Archive, Million Books etc… When the desired books are found in the OCLC/WorldCat, one will be able to locate them in libraries nearby the user. They can then be obtained via interlibrary loan, etc. Google Scholar and Wikipedia will be able to provide full-text information instantly.

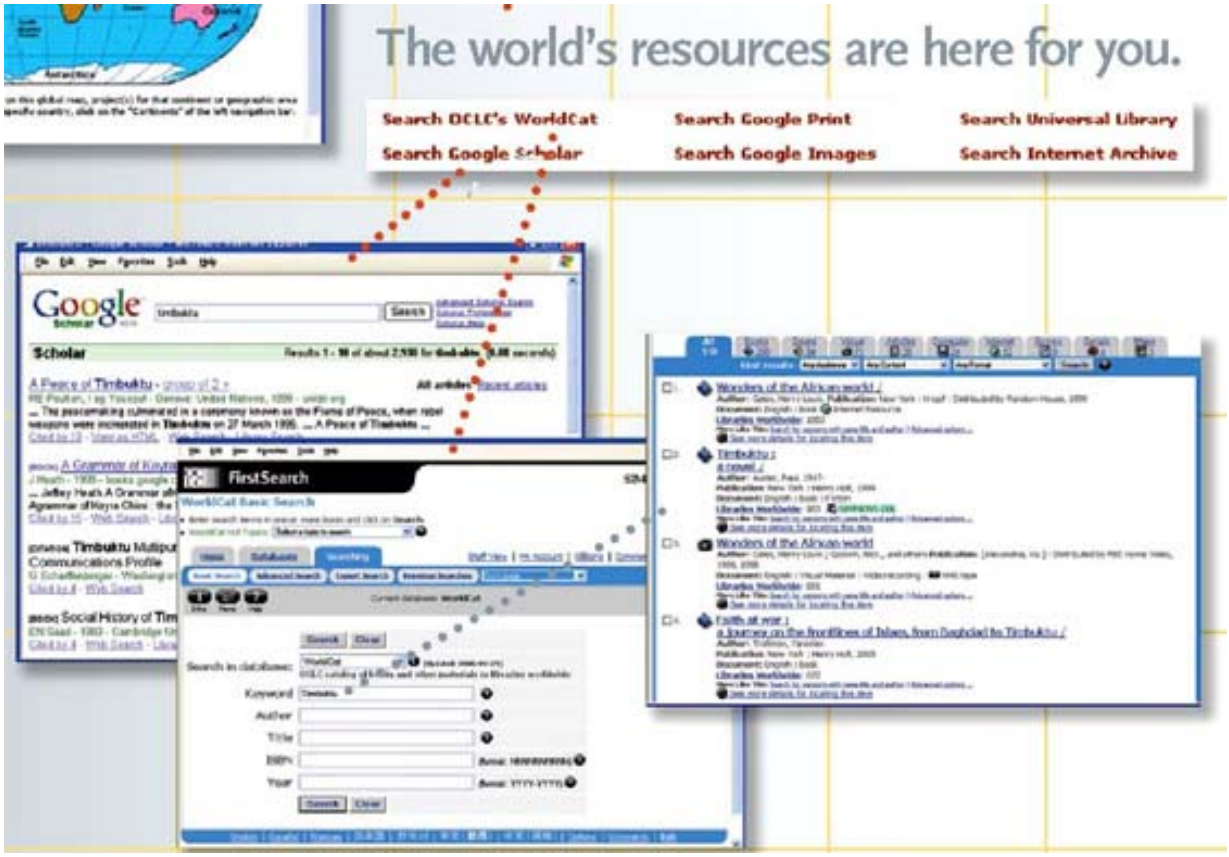


Figure 13. Searching the World's Bibliographical and Other Resources

- *Multilingual display of descriptive information as well as multilingual retrieval* – Currently GMNet already has contents in English, Chinese, Croatian, French, Italian, Japanese, Spanish, Thai, Vietnamese, etc., as shown in Figure 14. This facilitates greatly the use of GMNet resources because the choses language is familiar to him/her.

- *User can conduct geographical searches by continent and/or country* – Information on all digital collections can be accessed by continents as shown on the top part of Figure 15. Furthermore, over 230 countries can be searched and accessed by country as shown in the lower part of Figure 15. Unfortunately at the moment, only 80 of the 230 plus countries have digital collections. Many national libraries in the world do not even have Web Page available. This is why GMNet has a seperate collection on the world's National Libraries as shown in Figure 4.


Figure 14. Multilingual capabilities

In addition to the features mentioned above, it is worthy to stress that GMNet positions its user to [Chen, 2006b]:

- Participate actively during his/her own information seeking process,
- Decide how he/she would like to use the discovered or retrieved information,
- Develop his/her own project(s) by choosing, mixing and remixing the retrieved images, and
- Be an active content contributor to GMNet.

For example, upload capabilities are provided should one has valuable image(s) to contribute. What is also important is the provision for the user to utilize the resources of GMNet to make projects of their own. This is particularly useful for teachers, for example, they can retrieve the needed images of given subjects, go the classes for the scheduled times, and simply sign on to GMNet and deliver the lectures.

Figure 15. Geographical Access to GMNet

## CONCLUSION

In this paper, we have shown how new technologies have been uses to bond the past and the present together. Since its beginning in 2000, this National Science Foundation's International Digital Library Project has expanded to be a world image digital library as well as a global gateway to the world culture, history and heritage. Although the conceptual model of a global digital library was conceived as early as 1993, and it was only further refined in the late 1990s. The development of GMNet began at a time when there was no mention of Web 2.0.

The concept of "Web 2.0" began with a conference brainstorming session between O'Reilly and MediaLive International in 2004, and "in the year and a half since, the term "Web 2.0" has clearly taken hold, with more than 9.5 million citations in Google." [O'Reilly, 2005]. Although there are still a lot of disagreement on what Web 2.0 is all about and for, for GMNet, we have actually found considerable agreement between the core concepts of GMNet and the Web 2.0 core as shown in Figure 16, as well as some of the ideas that radiate out from that core.

It is clear to us that in developing the next-generation web-based applications, including digital Libraries, one should keep in mind the keywords such as "user," "services," "participation," "knowledge base," "open access," etc., and NOT "publishing," "packaged software," and "database(s)," etc.

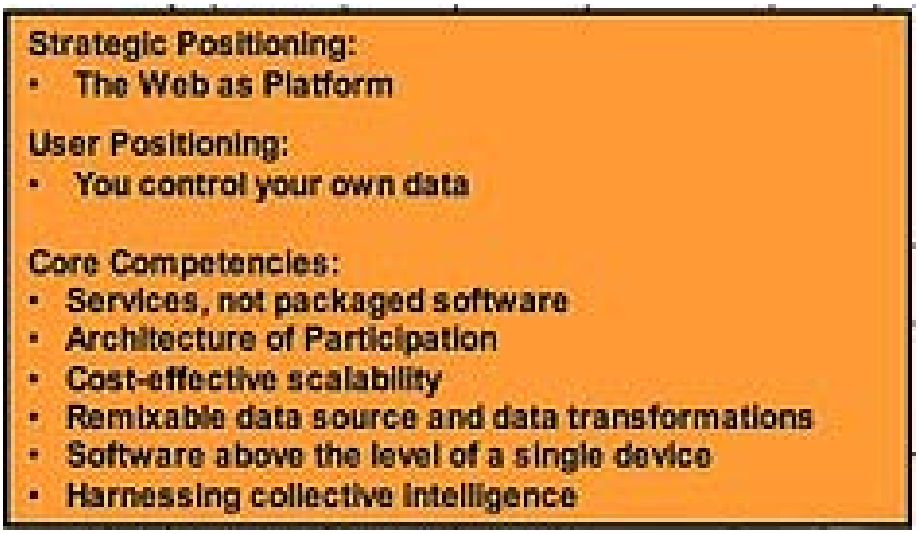


Figure 16. Part of the O'Reilly "meme" map showing the Web 2.0 Core.

## ACKNOWLEDGMENT

Global Memory Net is supported by the National Science Foundation's International Digital Library Program under the Grant Nos. NSF/CISE/IIS-9905833 and NSF/CISE/IIS-0333036.

## REFERENCES

[Chen, 1993]. Chen, Ching-chih, "Technological Potentials for the Global Digital Library," In: *Proceedings of the International Conference on National Libraries - Towards the 21st Century, Taipei, Taiwan, April 20-24, 1993*. Taipei, Taiwan: National Central Library. pp. 859-868.

[Chen et al, 2005]. Chen, Ching-chih, Howard Wactlar, James Z. Wang and Kevin Kiernan, "Digital Imagery for Significant Cultural and Historical Materials - An Emerging Research Field Bridging People, Culture, and Technologies," *International Journal on Digital Libraries*, 5: 275-285 (2005).

[Chen, 2006a]. Chen, Ching-chih, "New Digital Library Perspectives from International Library and Information Point of View," *Proceedings of the Annual Meeting of Library Directors of Universities and Colleges in Taiwan*, Tainan, May 19, 2006.

[Chen, 2006b]. Chen, Ching-chih, "Challenges for Developing a World Digital Library and Gateway: The Case of Global Memory Net," *Proceedings of Library in the Digital Age (LIDA 2006*), Dubrovnik, Croatia, May 29-June 1, 2006 [Chen, 2006b].

[Chen and Badurina, 2006]. Ching-chih Chen and Boris Badurina, "Developing Your Own Digital Image Library: Practical Approach," A Post-*LIDA (Library in the Digital Age) 2006 Workshop*, Zadar, Croatia, June 6-7, 2006.

[O'Reilly, 2005]. O'Reilly, Tom, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, " September 30, 2005.
http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[Zhang and Chen, 2004]. Zhang, Shengqiang and Ching-chih Chen, "Global Memory Net and development of digital image information management system: Experience and practice," *Journal of Zhejiang University SCIENCE*, 6A (11): 1216-1220. November 2005.

Photo credit: Justin Knight

**Ching-Chih Chen** is Professor of the Graduate School of Library and Information Science, Simmons College, Boston, and is a consultant and speaker to over 40 countries. She is the author/editor of more than 35 books and over 200 journal articles in areas of new information technologies, such as global digital libraries, multimedia technology, digital imaging, interactive videodisc technology, global information infrastructure, information management, and information resources, etc. She produced the award winning interactive videodisc and multimedia CD entitled *The First Emperor of China*, supported by the US National Endowment for Humanities (NEH). Currently she is leading two major NSF/International Digital Library Projects (IDLP): (1) Global Memory Net, a gateway to the world cultural, historical, and heritage multimedia resources, with collaborators from different part of the world, and (2) International Collaboration to Advance User-oriented Technologies for Managing and Distributing Images in Digital Libraries. She is also co-PI, with Prof. Raj Reddy of Carnegie Mellon University, of the China-US Million Book Digital Library Project.

A Fellow of the American Association for the Advancement of Science, she was appointed by President Clinton in February 1997 to serve as a member of the U.S. President's Information Technology Advisory Committee (PITAC). PITAC was established by a new Presidential Executive Order. Under both Presidents Clinton and Bush during 1997 to December 2002, she co-chaired the PITAC Subcommittee on International Issues, and was a member of the PITAC Subcommittees on Next Generation Internet (NGI) and IT*2 Initiative Review; and Panels on Digital Divide, Digital Library, Learning of the Future, and Individual Security. She also chaired the PITAC's activity on Digital Divide for Smaller Institutions. During 1987 to 2001, Dr. Chen was Chief Organizer of a series of 12 International Conferences on New Information Technology (NIT) in many continents of the world. The outcome of NIT '99 (Taipei) and NIT'2001 (Beijing) are the two-volume books related to the development of Global Digital Libraries – *IT and Global Digital Library Development* (1999) and *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration* (2001). She is a recipient of many major awards, and was also elected in 1985 as Fellow of the American Association for the Advancement of Science. She served as an Honorary Professor of Tsinghua University in Beijing from August 1999 to 2002 and University of Hainan, China since 2004. Active in the digital library area she was the co-Chair of the 4th ACM/IEEE Joint Conference on Digital Libraries (JCDL) of 2004 held in Tucson, Arizona in June 2004. She was on the Advisory Board of DELOS (the European Digital Network for Excellence), serving as the US Co-Chair of the NSF/DELOS Working Group in Digital Imagery for Significant Cultural, Historical and Heritage Materials, and served as the co-editor for the *Journal of Digital Library*'s Special Issue on Multimedia Contents in Digital Libraries (February 2006). A sought-after international speaker, in the last two years alone, she delivered keynote speeches and made presentations at many international conferences including those in Delhi and Bangalore, India; Dubrovnik, Croatia; Buenos Aires, Argentina; Beijing, Shanghai and Hangzhou, China; Kawasaki, Japan; Tainan, Taiwan, etc. She is on the advisory board of the major China Digital Library Project of the National Library of China; in October 2005, she was appointed as a consultant to OCLC for its Global Digital Initiative (http://www.oclc.org/news/releases/200520.htm); and receivedthe coveted LITA/OCLC Kilgour Award from the Library Information Technology Asso-ciation in June 2006 (http://www.ala.org/ala/pressreleases2006/april2006/2006KilgourAward.htm).