

Shigeo Sugimoto
Research Center for Knowledge Communities (RCKC), University of Tsukuba
(formerly University of Library and Information Science)¹
Tsukuba, Ibaraki, Japan
sugimoto@slis.tsukuba.ac.jp

1. Self-introduction and Background

I have been involved in digital library research at ULIS/U.Tsukuba and watching digital library R&D activities since early 90's. I have organized international symposia on DL in 1995, 1997 and 1999 at ULIS and International Conference on Dublin Core and Metadata Applications in 2001 in Tokyo, Japan. My primary research interests are metadata and technologies for metadata in the Web environment.

RCKC was established in October 2002. The mission of RCKC is research and development of information technologies to help communities in the networked information society. Digital library is an important component of knowledge and information infrastructure and it is one of the key research topics of RCKC. RCKC is encouraged to collaborate with communities in the real world to promote knowledge and technology transfer from academia to the communities and also to promote technology research in a real world environment.

Since mid-90's, the information environment in Japan has been very much changed as well as other countries. Recent significant changes from my viewpoint since 90's are:

- Mobile phones are widely used as email and WWW access media, e.g. i-mode. For example, some libraries provide OPAC access from mobile phones. Mobile phones with a digital camera have gained popularity; people take photographs and send them by email by a mobile phone.
- Internet connection has been improved very rapidly. Recent competitive market of ADSL connection has accelerated high-speed connection to the Internet from home.

The e-Japan program by the Japanese government is a key governmental activity to promote our nationwide IT environment. The report on strategies of e-Japan (II)² includes some keywords directly related to digital library, such as contents industries and digital archives, in addition to general terms. Ubiquitous information environment to be realized by small IC chips and high-performance network (IPv6) is another new keyword in new IT research. Other crucial topics include high-performance computing and robots, such as AIBO by SONY and ASIMO by HONDA gained popularity.

In the Japanese library community, DL has been and is recognized as a key issue. Digital collection including e-journals and digitized resources has been accepted as a core library service. Long-term preservation and resource access support services are recognized as important issues for libraries; for example, National Diet Library (NDL) is running an experimental Web archiving service and is discussing the legal deposit system of networked resources, and National Institute of Informatics (NII) and NDL have started portal services for Web resources. These services are also known challenging and expensive, though.

¹ University of Library and Information Science (ULIS) was merged with University of Tsukuba on October 1, 2002. The graduate and undergraduate programs of ULIS were continued as Graduate School of Library, Information and Media Studies and School of Library and Information Science, respectively. Research Center for Knowledge Communities was established at the time of the merger.

² As of June 1, 2003, a preliminary version of the second e-Japan strategy report (<http://www.kantei.go.jp/jp/singi/it2/pc/ejapan2.pdf>) is available for public comments. The report is written in Japanese.

Basic Perspectives

- In my basic understanding, digital library research should produce technologies and knowledge which are applicable to the real world of today and solve today's problem for (near) future.
- In the Digital Library Initiatives (DLI 1 & 2), Three Cs (Computing, Communication and Contents) shows fundamental viewpoints to the Digital Library Initiatives (DLI-1 & 2). I think addition of another C (Community) is important to make digital libraries research results really useful.
- There are many issues which were recognized important but not solved yet, e.g. preservation of digital/Web contents and interoperability among digital libraries. I think that we need more research on those unsolved issues.

Community oriented Information Technologies

WWW has created a global environment where any individual can publish and access information resources on the Internet. In one hand, search engines, e.g., Google, have been widely accepted as a basic service on WWW to search resources from the global WWW. These services connect every individual to the global WWW. On the other hand, the DL research community has developed technologies to assist an individual user to find and access resources, e.g., personalization of WWW and information extraction from huge amount of resources. These technologies help individuals access information resources on the Internet.

Human activities are primarily based on a community or on a few communities. A community is formed by people who share common interests. Traditionally, our communities have been formed based on physical restrictions, e.g., geographical distance and range of direct communication. In the networked information society, although the physical restrictions have less importance than before, it is unchanged that "community" is an important factor for our information and knowledge based activities. I think that information technologies to help us build a community-oriented information and knowledge environment are crucial.

The Internet and WWW provide flat environment, that is, Internet users can access to any resources open to the public using a uniform access method. However, this environment is very much different from organization of our communities. In a sense, users need only a small part of WWW but they are given too many resources and they have to evaluate resources by their own risk.

Here are some examples of communities:

(1) Scholarly Community

A scholarly community shares scholarly knowledge in a certain domain. Community members find new things and evaluate them based on their knowledge. Scholarly journals and subject gateways are built based on the community/domain knowledge. Both services are important but expensive. Information technologies to apply the knowledge to find high quality resources and organize resources for efficient access are crucial.

(2) School Community

A K-12 school is a community. School members – students, teachers, parents and others – would create resources and put them on the Internet. The published resources would be mostly used by the community members or those who are living next to the community. It is not easy for users who are not close to the community to find the resources only by using conventional resource discovery services even if the resources fit to their interests, e.g., members of other K-12 schools.

(3) Metadata Community - Dublin Core

Dublin Core, which is a metadata schema well-known and widely used, is a good example of a diverse community. Dublin Core Metadata Element Set (DCES) is simple because it is designed for usage in the global network. On the other hand, each community adopting DCES can extend DCES in accordance with their requirements; for example, the library community in DC has their application profile, a Japanese community uses elements for Japanese specific information. In this

environment, technologies to share metadata and its schemas across communities are crucial. The global DC community contains local or domain specific communities and it has two conflicting issues - interoperability and localizability.

A community shares knowledge. In other words, a community maintains knowledge and information shared among and used by the community members. A community has own vocabularies, criteria to evaluate resources, and interpretation frameworks. Thus, a community has its semantic framework but current WWW and DL technologies do not manage the framework well yet. In this sense, the Semantic Web activity has a large potential to build basis for community-based semantic framework, although its top layers are hard to realize by short/mid-term research.

Digital Library as (Social) Middleware – high-level semantic interoperability

Digital libraries need not be directly visible for human users as a library service for library users. From this viewpoint, digital library could be realized as middleware. It is a challenging issue, though. I think that digital library interoperability technology researches, such as Stanford DLI project and OAI, are oriented to this direction. Semantic interoperability is a key issue. We need semantic interoperability in higher level to build digital libraries as middleware; for example, cross-domain or cross-community ontology, multilingual information access, accessibility issues for users with disabilities, and so forth.

Digital Library is a key component in our information society. Not only information technologies but also social and human aspects are crucial to realize really useful infrastructure. Intellectual property issues are obviously a social issue. Reference service requires high quality human intelligence. High quality services which can be provided only by human services should also be taken into account to realize digital libraries are middleware in our social system.

Summary – five questions

As I mentioned in the first section of this paper, Internet has been accepted as an information infrastructure. In Japan, non-business people including high-school students are using mobile phones not only for talk but also for email and WWW access. In this aspect, mobile phones are a gateway to the Internet for them. These users don't always need to find resources from the global network but from local communities. ("Local" means both physical/geographical locality and logical/intellectual locality.)

I think usability in a real word environment is crucial for DL research. Many traditional communities are using the Internet, and there are many new communities which exist only in the Internet. Each community has semantic framework to evaluate information resources. The semantic framework is implicitly given to community members. I think we need technologies to build different types of information and knowledge infrastructure, i.e., personal, community-wide, cross-community, and global.

From this viewpoint, I think that interesting issues for post-DL research would include the issues, community-oriented information and knowledge infrastructure, cross-community semantic interoperability, and digital library as middleware.

Five questions

- What are we trying to do? What is the problem we're trying to solve?
Develop technologies to build community-oriented information and knowledge infrastructure on the global network.
Develop technologies to solve cross-community semantic interoperability.

The Internet provides flat structure. On the other hand, our society has hierarchy and network of communities. Challenge is to build information and knowledge infrastructure based on the (chaotic) community structure on the flat Internet.

- How is it done today, and what are the limitations of current practice?

Each community has knowledge to interpret and evaluate resources. Collection building and subject gateway development are typical examples. In the current environment, these tasks are done manually, in general.

- What is new in our approach / technology, and why do we think it will be successful? What gives evidence that it will work?

We would need community-based and cross-community ontology. Controlled vocabulary, which is nearly equal to ontology in this context, has long history in the library world. It is expensive to build and maintain a large controlled vocabulary. Mark-up language technologies would be useful to build ontology, but light-weight tools to build ontology and flexible and powerful tools to manage multiple ontologies would be required.

- Assuming we are successful, what difference does it make?

Flat global Internet is not always useful. We can build information and knowledge infrastructure that fits to our community structure.

- How long will it take, how much will it cost, and what are (measurable) milestones, mid-term and final exams?

Development of Software tools would be a short term issue. Development of community-based and cross-community meta-information (or metadata) schemas would be a mid-term issue. Community oriented ontology building would be a mid-term to long-term development.