Design of a Digital University Library: Mandala Library

Masakazu IMAI, Hisakazu HADA, Suguru YAMAGUCHI, Kosuke SATO,
Haruo TAKEMURA, Naokazu YOKOYA, Kunihiro CHIHARA and Tadao KASAMI

Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-01, Japan
imai@is.aist-nara.ac.jp
TEL: +81-7437-2-5371
FAX: +81-7437-2-5379

Abstract

The recent progress in computers and networks realizes the possibility of digital or electronic libraries. There are many researches on digital libraries. In this paper, we present the outline of the digital library project which is now in progress at Nara Institute of Science and Technology. The project, called Mandala Library, is planned to construct the first digital university library in Japan.

Nara Institute of Science and Technology (NAIST), established in 1992, is very young institute and does not have so many books and journals. NAIST consists of two graduate schools. According to the master plan of NAIST, the library of NAIST will open next year. Since NAIST consists of two graduate schools, the fields of researches conducted at NAIST are limited. These conditions make Mandala Library construct easily.

Mandala Library is a digital university library based on the technology of WWW. Using WWW, Mandala Library can provide easy access to not only primary sources but also other information on Internet. Users can have access to best database whose objective is suited for their purposes.

There are still significant problems, especially relating the copyright and Japanese language. In Mandala Library, the access history to all primary sources will be recorded. We expect to find a solution to the copyright problem in cooperation with publishers.

1. Introduction

The recent progress in computers and networks realizes the possibility of digital or electronic libraries. There are many researches on digital libraries [1, 2, 3]. In this paper, we present the outline of the digital library project which is now in progress at Nara Institute of Science and Technology. The project, called Mandala Library, is planned to construct the first digital library in Japan.

Nara Institute of Science and Technology (NAIST), established in 1992, is very young institute and does not have so many books and journals. NAIST consists of two graduate schools; Information Science and Biological Science. Each graduate school consists of 80 faculties and 350 graduate students. According to the master plan of NAIST, the library of NAIST will open next year. Since NAIST consists of two graduate schools, the fields of researches conducted at NAIST are information science and biological science.

We can assume the followings.

- Users are familiar to a workstation.
- The area which our library have to cover is restricted to information science and biological science.

Under these conditions, NAIST can construct a university library as a digital library more easily than other universities do.

We describe the Mandala Library project in this paper. We first present the role of university library in the following section. Then we show the concept of our project. After describing the infrastructure for a digital library at NAIST, we present the architecture of Mandala Library. Finally, we conclude.

2. Functions of University Library

University libraries have two major objectives. One is to collect scientific literature and library information for education and researches in the university using all available information sources. The other is to distribute academic information such as technical reports, master's theses and doctoral theses to outside of the university. The basic functions of university libraries are as follows:

- Ability to exchange scientific information internationally.
- Capability to use multi-media information. (conventionally, text on papers)
- Ability to collect information on sources of documents.
- Services at the nearest point from users.

3. Concept of Mandala Library

This section describes the outline of the NAIST university library, which is named "Mandala Library".

3.1 Characteristics

The Mandala Library is characterized as follows :

• Media Center —– When users have access to a material, they do not have to consider the difference of information sources.

• Desk-side Library —– Users retrieve primary sources from their workstations without going to a building of the library.

• 24-hours Library —– Users have access to the library any time.

3.2 Functions

Functions of Mandala Library are as follows :

1. Digitized primary sources —– All primary sources are digitized in Mandala Library. We will digitize books, journals, micro films and video sources. We will also make text data by using OCR (Optical Character Recognition) techniques if possible.

2. Retrieval and reading from all workstations in NAIST —– Users can retrieve and read not only primary sources but also video sources which are provided by retrieval server. The plural users can read same primary source at same time. Retrieval of specified documents is achieved by a full-text search of the text body of the documents which is generated by OCR.

3. Searching service with outside databases —– Users can search information which is not in Mandala Library using outside databases via Internet. Outside databases include commercial and semi-commercial databases and WWW (World Wide Web).

4. Notification to users —– Mandala Library system notifies a user that a new paper has arrived according to the keywords which are registered by the user.

5. Distributing information to the outside of NAIST —– Academic and Scientific information made in NAIST is distributed not only to the inside of the institute but also to other universities and research institutes.

3.3 User Interface

As described in next section, there are many types of workstations in our institute. This means that it is very difficult to make a new software for all workstations of NAIST. Therefore, we choose WWW client software as client software of Mandala Library. There are some other advantage of using WWW software. One advantage is that users can make access to other information provided by WWW via Internet. Users can have access to both Mandala Library and WWW on Internet in the same manner. This reduces loads of users. Recently, a graphical HTML editor has been released. Using such a HTML editor, librarians can reconstruct the WWW pages easily. This is another advantage of using WWW software.

Because primary sources are digitized, users have to read primary source with CRT of a workstation. Since the resolution of CRT is low (typically less than 100 dpi) and characters generated by a workstation are not familiar to users, reading primary sources on the CRT is very hard work for users. In scientific papers, figures and tables play a very important role. We can not recognize the details of figures with the resolution of CRT. Additionally, users may be conservative and want to read primary sources printed on papers. Because of the reasons described above, we construct the system to show primary sources as images of the printings instead of text. In other words, the system is image based system. Of course, users may print the page image to read on papers with some copyright-related restrictions described later.

Mandala Library system provides a full-text search for the result of OCR. Using only secondary sources for searching keywords is conventional and implies a problem. The problem is that the keywords in secondary sources do not reflect actual content of literatures perfectly. It is very doubtful that all authors give appropriate keywords for their literatures. Important keywords appear several times in the body of literature. Even if OCR make a mistake at a location, an important keyword will be recognized correctly at other location. Thus, searching will succeed.

3.4 Copyright Policy

The problem relating copyright is one of the most significant problems in digital libraries. Our policy is that we never digitize printings without agreements of their publishers. Nor we never allow users to print the image data without agreements of publishers. Of course, people outside of NAIST can not have access to primary sources in Mandala Library. We are now negotiating with several publishers.

4 Infrastructure for a digital library at Nara Institute of Science and Technology

To realize a digital library, infrastructure of computer and network plays a very important role. NAIST, established in 1992, is very young institute. The infrastructure of our institute is named as **Mandala**. Mandala consists of computer systems, which are called as Mandala System, and networks which are called as Mandala Network.

4.1 Mandala System

Figure 1 shows the overview of Mandala System. Mandala System provides a workstation for every faculty and every graduate student. Users can use the workstations on their desks at any time and at any place in NAIST. Mandala System also has file servers to store home directories of users. The amount of user and file server is approximately 900 and 1 TB, respectively. There are other research specific sub-systems such as a small scale computing server, image processing system, knowledge system, etc. Users can use every system at any time and at any place via Mandala Network described below. Of course, the system runs 24 hours a day, 365 days a year.

4.2 Mandala Network

Figure 2 shows the outline of Mandala Network. The speed of the back-bone network is 1 Gbps. In the buildings of Graduate School of Information Science, each floor has 4 ethernets and 2 FDDIs. Mandala System is connected each other via Mandala Network by ethernet, FDDI or HIPPI. The workstations for users and the file servers are connected by ethernet and FDDI, respectively.

Mandala Network is connected to Internet via WIDE with the speed of 1.5 Mbps. WIDE is one of the largest experimental networks in Japan. Users in our institute can have access to information sources via Internet from every workstation attached to the Mandala Network.

5. Architecture of Mandala Library System

In this section, we describe the details of Mandala Library System. First, we describe the outline of the system. The details of each subsystem is then briefly presented.

5.1 Outline of Mandala Library System

Figure 3 shows the architecture of Mandala Library System which is composed of five subsystems. The subsytems are followings and the detail of each subsystem is described later.

- Input Subsystem
- Storage Subsystem
- Retrieval Subsystem
- Digital Video Subsystem
- Outer Database Access Subsystem

5.2 Input Subsystem

This subsystem provides the function of digitizing documents. In our project, Mandala Library System digitizes all page information as images. The objective of this subsystem is to digitize page information. Currently, primary sources are provided by printings, microfilms and CD-ROMs. Input subsystem consists of the following four kinds of devices. Some devices also have the capability of transforming scanned data into text by OCR.

• Image Scanners

These devices digitizes documents on papers automatically. Some documents are printed with color ink. The devices have to treat color information.

• Microfilm Input Devices

Backnumbers of many journals are provided by microfilms. Microfilm has very high durability and is best media to preserve. NAIST purchases backnumbers of journals as microfilms. The devices will digitize page information automatically on demand.

• CD-ROM Input Devices

Recently, some journals are provided by CD-ROMs. These devices treats documents already available in electronic form.

5.3 Storage Subsystem

The objective of the subsystem is to provide large-scale file servers to store the digitized primary sources together with structured information like authors and titles in a multi-media database. Because the capacity is very large, storage subsystem consists of three types of devices such as high access-speed device, medium access-speed device and low access-speed device. These devices totally realize a hierarchical storage system. The subsystem decides which data is in high access-speed device according to the access frequency of the data.

5.4 Retrieval Subsystem

Retrieval Subsystem provides a full-text search of the total text body of the documents which is obtained by OCR in the Input subsystem. Retrieval Subsystem also provides access to all data in Storage Subsystem for clients.

5.5 Digital Video Subsystem

Mandala Library collects video information such as video teaching materials, visual result of researches, lectures and video data for research. This video subsystem provides function of creating, editing and presenting video data in Mandala Library. The subsystem includes an MPEG-2 encoder, decoders and a video server.

5.6 Outer Database Access Subsystem

It is very difficult for Mandala Library to hold complete bibliography. If a user want to search bibliography using available on-line databases of third vender, Mandala Library System has to support them. The subsystem also collect informations on histories of usage of third vender's database.

6. Conclusion

In this paper, we have described the outline of the digital library project at Nara Institute of Science and Technology. Our digital library, Mandala Library, is now in the planning stage and will be in operation from April 1996. Though the building of the library is built in 1996, Mandala Library System will be constructed for three years, from 1996 to 1998. Mandala Library is the first digital university library in Japan.

Mandala Library is a digital library based on the technology of WWW. Using WWW, Mandala Library can provide easy access to not only primary sources but also other information on Internet. Users can have access to best database whose objective is suited for their purposes. Mandala Library uses various advanced technologies, some of which are now under development. We hope the development goes smoothly. We continue to employ new technologies if they are suitable for Mandala Library.

There are significant problems, especially relating the copyright and Japanese language. In Mandala Library, the access history to all primary sources will be recorded. We expect to find a solution to the copyright problem in cooperation with publishers.

References

[1] Proceedings of Digital Libraries '94, 1994.

[2] G. A. Story, et. al. : "The RightPages Image-Based Electronic Library for Alerting and Browsing", IEEE Computer, Vol. 25, No. 9, pp. 17-26, 1992.

[3] R. Bayer : "Design of a Multi-Server Active Digital Library System", Proc. of Int. Sympo. on Advanced Database Technologies and Their Integration, pp. 284-291, Nara, Oct., 1994

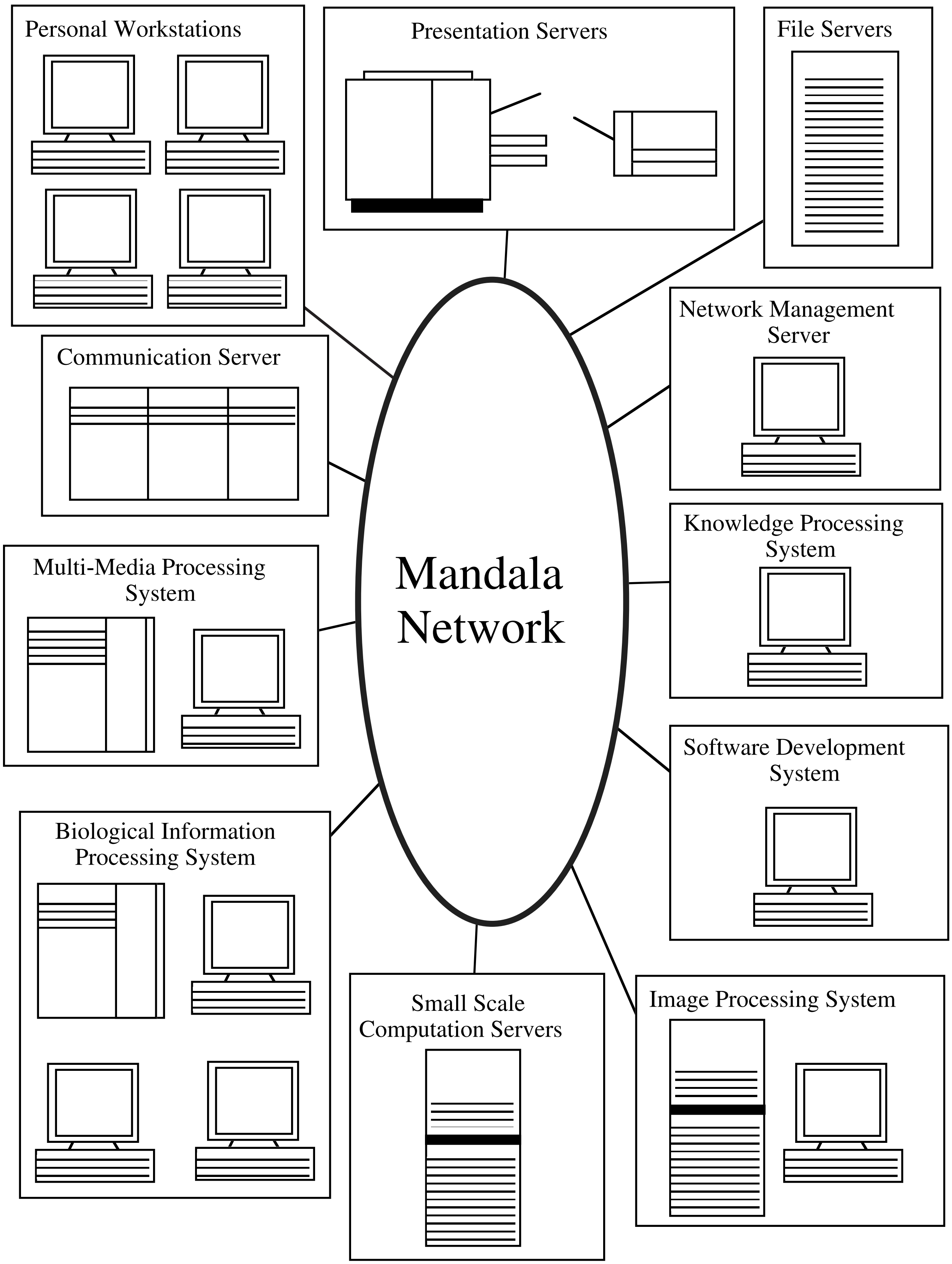

Figure 1 An Overview of Mandala System

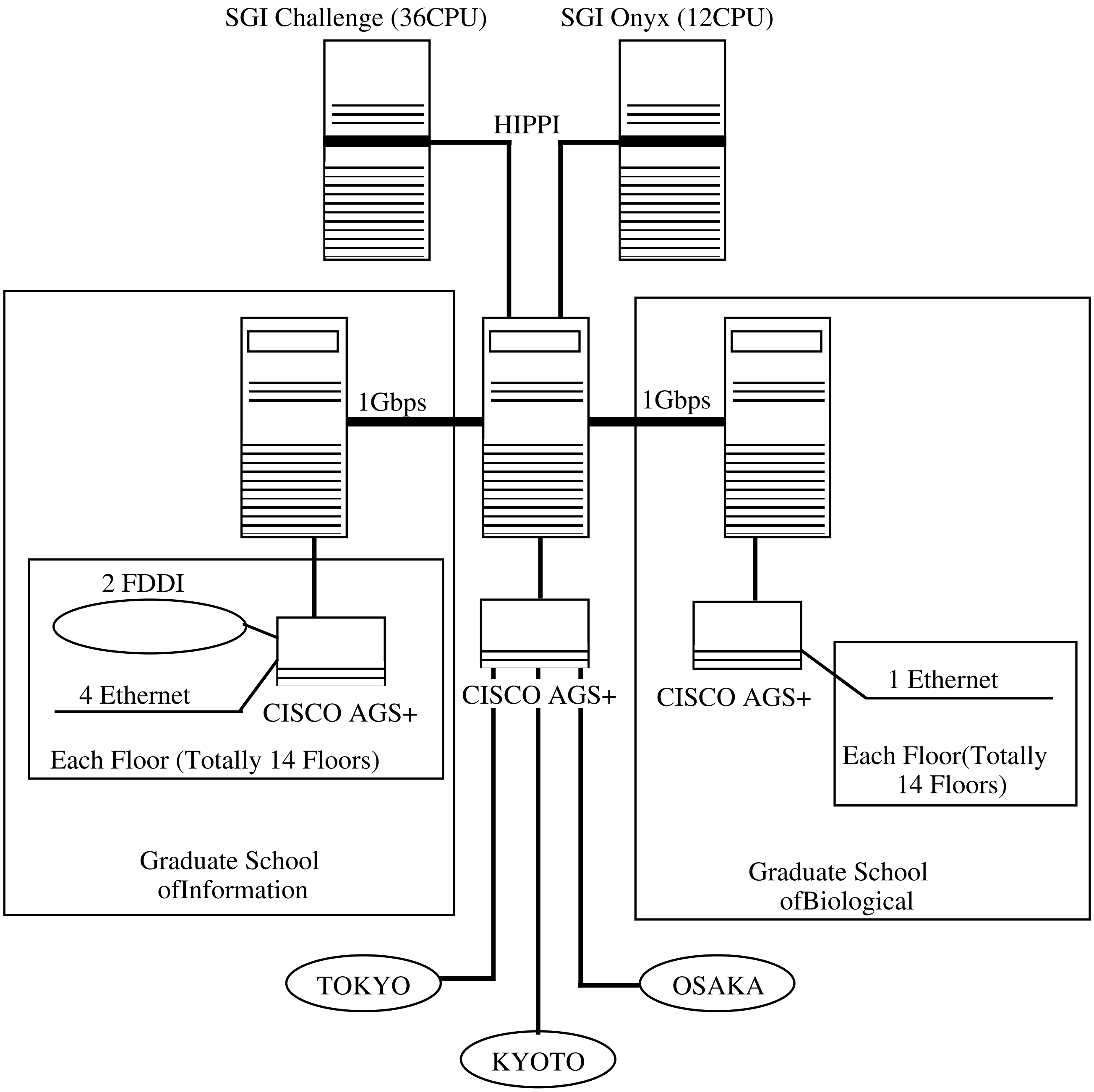

Figure 2 An Overview of Mandala Network

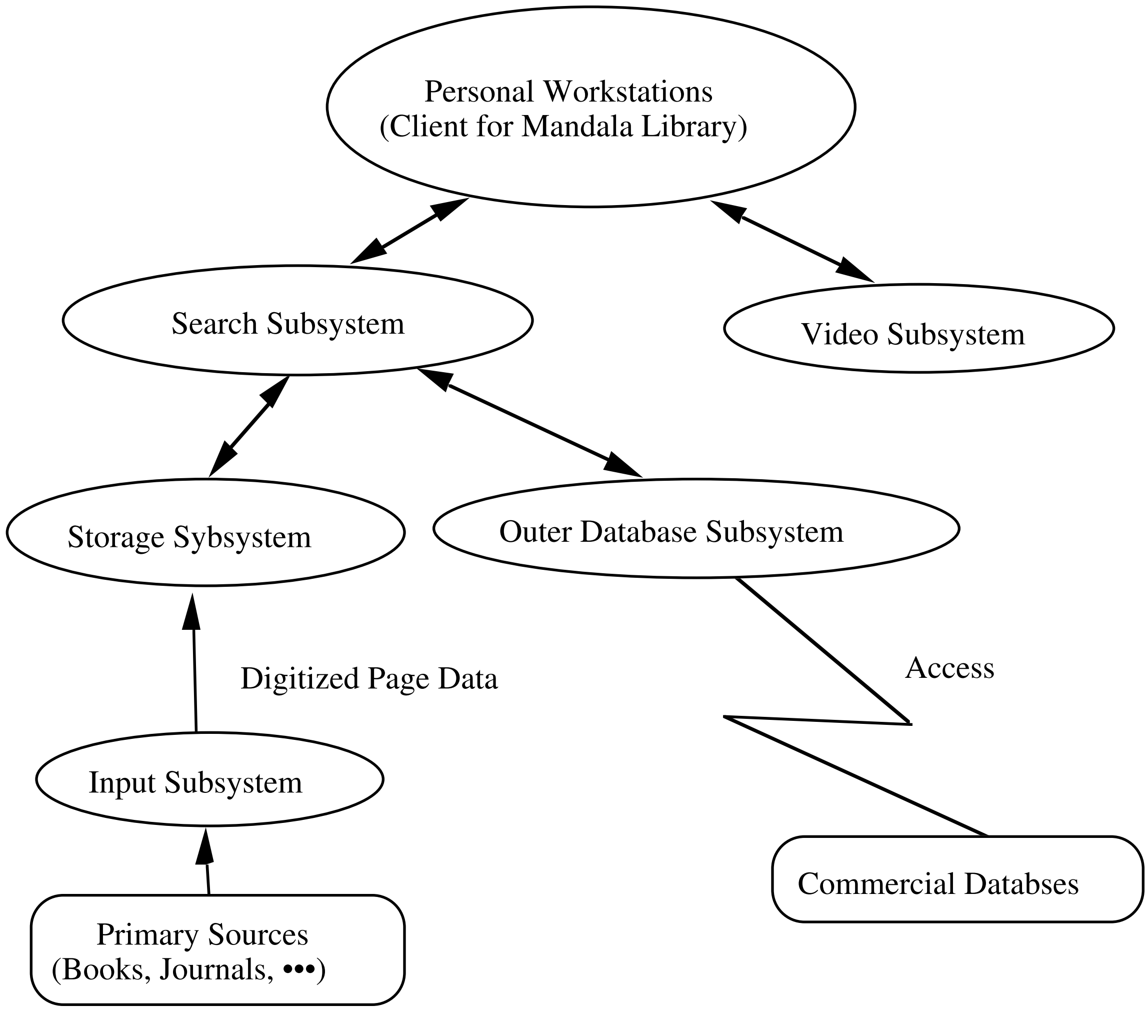

Figure 3 Architecture of Mandala Library