

UTILIZATION OF DIGITAL LIBRARY AS KNOWLEDGE-SHARING DEVELOPMENT SUPPORT

Umi Laili Yuhana¹, Supriadi²

^{1,2}Informatics Department, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember

Email : ¹yuhana@its-sby.edu, ²xsupri05@yahoo.co.id

ABSTRACT

The library development in terms of data and documents that stored is starting from the traditional library, which only consists of a collection of books without catalog, and then appears semi-modern library using the catalog (index). Current development is the emergence of digital libraries, which has advantages in speed because the data is digital-oriented. The development of the digital library is in line with the motivation in knowledge management that getting lots of attention at this time. One important area in knowledge sharing is how to encourage people to share what they know. Knowledge is usually referred as the power source, and by not sharing it, then someone will have more value in an organization or company so that lessen the possibility to be replaced in the position. To overcome this, it is important to encourage knowledge sharing activities and provide understanding of knowledge sharing benefits to people who have knowledge.

Utilization of digital libraries as knowledge sharing development support is the author's ideas to overcome the lack of motivation in knowledge sharing, as mentioned above. Research and application in the field has been carried out as envisaged in the literature as literature review on this paper. Through this paper, the study will be conducted on the digital library as knowledge sharing development support.

Keywords: Digital Library, Knowledge Sharing

1. INTRODUCTION

Digital library is a library that stores data that included books (any posts), images, sound files in electronic form and distribute it using the protocol through a network of electronic computers. The term digital library itself has the same meaning as the electronic library and virtual library. While the

term that often used today is the digital library, this can be seen frequently, the emergence of that term in a workshop, conference or symposium with that name.

Digital library research began growing rapidly since 1990 accompanied by advances in technology that allows computer network initialization information from one place to another in a short time. Starting with Workshop on Digital Libraries in 1994, several other conferences such as Digital Libraries (DL) sponsored by ACM, and the Advances in Digital Libraries (ADL) sponsored by IEEE / NASA / NLM, continuously held. Many workshops, symposiums, conferences here prove that the number of researchers began to study this field.

One important area in knowledge sharing is how to encourage people to share what they know. Knowledge is usually referred to as the powers source, and by not sharing it, then someone will have more value in an organization or company so that the possibility to replace in the position. To overcome this, it is important to encourage knowledge sharing activities and provide understanding of knowledge sharing benefits to people who have knowledge. One suggestion given is to make clear relationship between knowledge sharing and business profit. Because of this, will motivate someone to take the time to share what they know.

Next, is the task of the organization or company to provide contribution value on each individual that share their knowledge. With that, more contributions will appear because it was clear that knowledge sharing does not mean that knowledge is lost. Sharing knowledge will provide new knowledge and increase the value of the organization and individual. Some of the other needs to support the knowledge sharing environment is the support, motivation, a forum to share information and make it accessible to all parties, and enough time to share their knowledge. All that requirements will be easily fulfilled by the library facilities, while the requirement for speed and ease of access can be fulfilled by the digital library.

Some of the competitive advantage of the digital library such as speed and ease of access is a requirement needed to support knowledge sharing. Because of that, implementation of digital library technology can be used

to support the development of knowledge sharing in Indonesia. As a first step, the implementation can be done in the universities environment in which intellectual ability and the infrastructure are ready.

1.1 Digital Library

Digital Library is collection management of digital objects. Digital object is created or collected based on the principles of collection development and is set such that allows the user to access and exploit the sources, such as library materials in general. Digital objects to be treated as the source remains stable within a very long and processes applied to it will ensure appropriate quality and survivability.

Digital library research project on its core areas is documents digitalization and development system for digital documents. On documents digitalization, is examined how to digitalize the type of document and digital documents storage such as full text and page image. While the field of system development in the DL, is examined on the design and implementation of the system to manipulate the data in the database, for example, the research system architecture for good the digital library.

The following are the main characteristics of the digital library:

- Variations of digital information sources
- Digital library reduce the physical space needs
- Users can access them wherever he is
- Users can create personal collections with the facilities provided by digital library
- Provides access to distributed information sources
- The same information source can be used by many users at the same time
- The paradigm changes in the ownership and use
- Development of the collection is based on the potential usefulness, filtering mechanism and the appropriate
- The ability to handle the collection of multilingual
- Minimize manual processes and utilize process automation

- Provides search facility and better retrieval
- Digital information can be used differently by different parties
- A digital library erases time, space and language barrier

1.2 Knowledge Sharing

Knowledge sharing is one of the types of knowledge development cycle, in cycles of knowledge sharing there are 8 as seen in the figure on 1, which are: Identify, collect, classify, organize / store, share / disseminate, access, use / exploit, and generate which describes the process of how we obtain, filter and re-create knowledge. Other literature says that knowledge sharing is the use of organizational resources to collect and summarize all the knowledge that already exists, which is focused on the terms of cooperation.

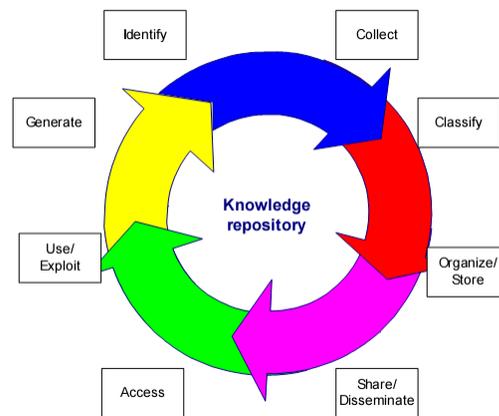


Figure 1 Knowledge development cycle

Knowledge-sharing strategies can be grouped into three categories, namely writing, speaking, and information technology. Some of the strategies included in the category of writing is the publication of research, scientific reports, books, newsletters, and other print media. While the speaking is speaking in conferences, lectures, presentations, workshops, sessions and meetings conversation. Meanwhile, the website, online forum and mailing list are included in the information technology category

Knowledge sharing activities should be able to improve access to information, facilitate communication and encourage active participation in learning and community. Implementation of knowledge sharing strategy may not be able to meet all these goals, but if knowledge sharing is intended to continue to be used in

the future, the need to set goals which must be met first. Sometimes, the choice must be determined between the most important goals and leave the others.

The process and strategy to create knowledge sharing will be different based on individuals, organizations and regions. This means that a priority goal can vary significantly in different location, time and user.

Some of the benefits that can be obtained with knowledge-sharing is as follows:

- Prevent "reinvention of the wheel"
- Minimize the possibility of loss of knowledge
- Allows the distribution of best practices
- Building social capital

1.3 Text Based Digital Library

The amount of document in digital library usually grows rapidly overtime. How to store, manage and search these documents within the digital library is a challenging problem. Documents in digital library are stored as semi-structured data, while in the traditional relational database it is store as structured data. Relational database management system cannot manage semi-structured data efficiently and cannot satisfy the requirement of content-based text retrieval.

A lot of research works have been done about semi-structured data, such as data modeling, query language for text retrieval, index methods and text retrieval algorithms and similarity search algorithms. These research results have been use a lot in digital library system. SSREADER Digital Library, the National Digital Library and Wanfang Database are popular digital libraries in China. All the digital libraries classify the document into several classes and support query inside a given class. Metadata search and full-text search through a single keyword or expression are both supported in these digital libraries.

Others examples of digital libraries are Greenstone Digital Library, UC Berkeley Digital Library, Tufts Digital Library, ACM digital library, NCSTRL etc. Similar functions are supported in these digital libraries, such as metadata searching, full-text searching, document classification and browsing.

1.4 Video Based Digital Library

Digital video is a critical challenge in the digital library. The challenge comes as the size and the lack of textual features, so that adds the difficulty in retrieval process.

Other important projects, including IBM's CueVideo, which has an variety of techniques Segmentation, indexing, and user interface developed Almaden and Watson labs. There is also the Digital Video Multimedia Group in Columbia who worked on the project relating to the making of the video Summaries automatically. The Multimedia Information Retrieval Group at Dublin City University is developing Fischlar Project, which provides a video broadcast to the university community. This group is also developing innovative user interface for the video repository. While the European Union's ECHO Project to develop research in the field of metadata scheme.

One of the example digital library system that is capable of handling digital collections-based multimedia (digital video) is the Open Video Digital Library or OVDL. OVDL using the MySQL database interface module and the Agile View PHP is used as middleware.

OVDL have a browsing interface that allows access based on genres (documentaries, educational, lecture, ephemerals, historical), duration (less than a minute, 1-2 minutes, 2-5 minutes, 5-10 minutes and more than 10 minutes), color (color or black and white), sound (with sound or silent) and the contribution organization (CMU, Internet Archive, etc). For each category, the number of documents in the category is given.

While the search for the facility, supported by 2 types of search process. Quick search that only needs a keyword and detailed search that involved metadata.

After the user determines the desired category by using the search or browse interface, the information is displayed and the detail preview of some segment is displayed. This is important when handling video files with a large size. Because it will take some time to download or view them, and will be a bit useless if the video is not viewed or downloaded as you wish. To reduce the possibility of that event, display detail information about video and preview the storyboard can be a solution.

2. ANALYSYS, DESIGN AND IMPLEMENTATION

System will be divided into 2 parts based on major functions of this application. Here is the part :

1. Web Access (Front-End)

This part is the interface that will be used by visitors to access the website collection. Collection of the website is divided into 2 types, namely a collection of text documents with the PDF file format and a collection of multimedia documents with FLV file format.

2. Web Management (Back-End)

This section is an interface for administrator to do the website system management. With this interface, admin can do Member management, document management, category management and news management.

Users will be divided into 3 categories, namely: admin, registered user and website visitor who does not register (guests).

Here some features of the system:

- Document uploading for registered user.
- Document reviewing for registered user.
- Simple Searching, for text documents that supported by search engine capabilities
- Secure document text, so it was not easy to do copy-paste
- Accessing Abstract Text Document and Multimedia Storyboard to visitors who do not register (guest)
- Downloading Documents Text and Multimedia, to the user who register (registered)
- Browsing By Title
- Browsing By Category
- Searching With Metadata
- Accessing News
- Member Management
- Category Management
- Document Management
- News Management

2.1 Requirement Specification

In general, the requirements that would be fulfilled by the existence of the Digital Library system are:

1. Providing system management.
2. Providing document management.
3. Access to text documents.
4. Access to multimedia documents.
5. Search documents quickly.
6. Document protection.
7. Document sharing.
8. Document reviewing.

Besides that, the authors also map the non-functional requirement that is expected to be fulfilled by this application. The requirements are as follows:

1. Security

Security is the ability to be able to protect the application system. The requirement is this application can protect the documents to not be easy to do copy-paste, and can only be accessed by users who have been registered on the system.

2. Accessibility

Applications can be accessed by anyone that has been registered into the system. The process of access can be done easily because it requires only a web browser.

3. Manageability

Collection that used as the main data on this system may be arranged in such a way. The arrangement process can be done with the interface that is easy to operate.

4. Performance

Measuring the performance of this system is more emphasis on the response time needed during the search process. With the implementation of the concept of indexing and search engine features, it is expected that the response time on document searching process is 1 than 2 seconds with the accuracy or relevance between keywords and the search results.

2.2 System Architecture

In the implementation, the Digital Library system has a device and network topology as described by figure2

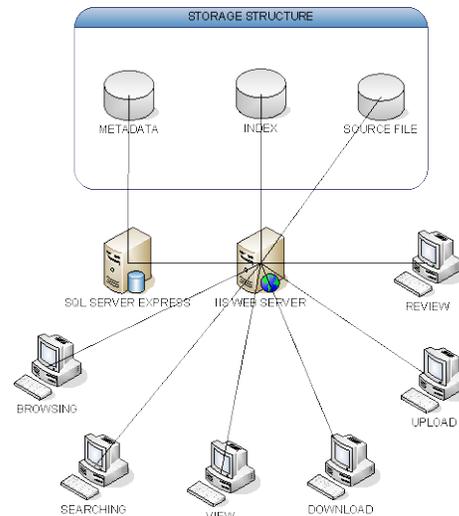


Figure 2 System Topology

As described by figure 2, digital library system can provide most requirements needed by knowledge sharing process. As explained before on knowledge development cycles, there are several processes on knowledge development. The system can support knowledge repository by providing storage structure that involved database and file system. Other knowledge management process such as generate, identify, collect, classify, organize and store can be supported by the system that has document management system. The share/disseminate, access and use/exploit process will be supported by system feature that includes browsing, searching, view, download, upload and review.

2.3 Implementation

To fulfill the requirement as explained earlier, there are several technology that used. The technology used by this system are as follows:

a. Extraction of PDF files to text form

Because the system uses text-based search engine, then the PDF file must first be extracted into text form. To do this, the library for PDF manipulation is used. That library is PDFBox.

b. The process of text based document indexing

The process of indexing and searching in this system are handled by the Lucene. With Lucene, we can design our own index which will be made in accordance with our needs.

c. Text based encryption

The aim of encryption on this system is to make the pdf file content not easy to copy-and-paste. To do this another PDF manipulation library is used. That is iText.

d. Content-based searching on text-based document

As mentioned earlier the process of searching using Lucene library. In addition to rapid searching capability, Lucene can provide high relevancy between the query and search results.

e. Extraction process of video frame as storyboard of video based document

The frame extraction process use a library namely ffmpeg. With ffmpeg we can access video frame based on the time. For example, we can take video frame on minutes 1, seconds 200 and so on. Later we use this technique to generate storyboard from video based document.

f. Security with access rights.

Users who do not registered on the system have only limited access that is only has access to get abstract of text based document and storyboard of multimedia document. While the internal user can access document fully, and do the download.

3. RESULT

Testing is done on a system built using ASP.NET technology and SQL Server 2005 Express Edition. Testing includes functional and non-functional testing.

3.1 Functional Testing

From the results of the testing, all the features of the system can run well in accordance with their needs. System is capable of handling text-based and multimedia document well. Document management functions are also running well and can be used to store and manage documents in the system. Access documents also running well. Functions of browsing, searching, uploading, downloading, and access to documents all runs as expected.

3.2 Non-Functional Testing

Non-functionality testing is done to see whether the performance of the application in accordance with the expectation. Non-functional testing includes some non-functional requirements of the security and performance.

Three main security features are:

- Access Rights
- Protection of the copy-paste
- Absolute URL protection

From the results of the testing, all features work well in the protection of the document. Users can only access the document and the features in accordance with their access rights. Documents stored in the system also not easy to be copy-and-paste. While the enhancement of the protection of absolute URL, give extra protection from the download process of the user who does not registered.

System performance is measured from three main factors, namely:

- Time needed by the indexing process.
- Comparison between searching with indexing and searching without indexing.
- The relevance between keywords and the search results.

From the results of the testing, the time required for the indexing process is less than 0.5 seconds for each document. While the comparison between searching with and without implementation of indexing concept give satisfactory results that is almost 1:2 or better. Testing results for relevance between keywords and the search results, the results also satisfy the relevance between keyword and the search results that is quite good.

4. CONCLUSIONS

Based on the testing we conclude that:

1. Web-based digital library system can provide system management, document management, access to text documents, access to multimedia document, searching documents quickly, document protection, document sharing and document reviewing that can be used to support the development of knowledge sharing.
2. Secure text document can be implemented using the access rights and encryption.
3. Text documents searching using the concept of indexing and search engine can accelerate the search process.
4. Text documents searching using content-based search engine can increase the relevance between documents and keywords.
5. Digital library that implemented with web-based system can handle text and multimedia document well.

REFERENCE

- [1] Gospodnetic, O and Hatcher, E. 2004. Lucene in Action. Manning.
- [2] Lowagie, Bruno. 2007. iText in Action. Manning.
- [3] Marchionini, G., dan Geisler, G. 2002. "The Open Video Digital Library". D-Lib Magazine Vol 8, Desember, 12.
- [4] Mazurek, C, Stroinski, M, Szuber, S. Digital Library for Multimedia Content Management. Pozna_Supercomputing and Networking Centre, ul. Noskowskiego 10, 61-704 Pozna_, POLAND.
- [5] Setiarso, Bambang. 2007. Knowledge Sharing and Digital Libraries in Indonesia : Facing a lot of Constraints and Challenges to Disseminate The S&T Information For the Society. IlmuKomputer.Com.
- [6] Singh, Sukhdev. 2003. Digital Library: Definition To Implementation. Lecture Delivered at Ranganathan Research Circle. Delhi.
- [7] Tsui, T, Chapman, S, Schnirer, L, Stewart, S. 2006. A Handbook on Knowledge Sharing: Strategies and Recommendations for Researchers, Policymakers, and Service Providers. Community-University Partnership.
- [8] Wahono, Romi Satria. 2006. *Teknologi Informasi untuk Perpustakaan : Perpustakaan Digital dan Sistem Otomasi Perpustakaan*. IlmuKomputer.Com.
- [9] Wahono, Romi Satria. 2003. *Menengok Proyek Digital Library*. IlmuKomputer.Com
- [10] Yan, Y., Jianzhong, Li, dan Zhongliang, Kan. 2005. "The Design and Impelementation of Parallel Digital Library Managament System". Jurnal of Computer Science 1, 2:157-163.