

Metadata and Meta-Information System for Hypermedia Documents

Woojong Suh and Heeseok Lee
Graduate School of Management, KAIST
wjsuh@chollian.net, hlee@msd.kaist.ac.kr.

Abstract

Recently, many organizations have attempted to construct hypermedia systems to expand their working areas to Internet-based virtual work places. For the effective management of the hypermedia applications, it is important to develop a technique for managing hypermedia documents, hyperdocuments. This paper employs metadata as it has been conceived as a key approach in document management. Hence, this paper proposes a meta-information system based on metadata, HyDoMIS, for the purpose of hyperdocument management. This system contains a repository for hyperdocuments, which is based on metadata schema and classification. HyDoMIS performs functions such as metadata management, searching and reporting.

1. Introduction

Today, hypermedia documents are growing explosively according to a lot of attempts to construct hypermedia systems such as intranets or extranets for electronic commerce by many organizations. Therefore, managing hyperdocuments is important. A hypermedia document, a special type of digital documents, is based on the interlinking of nodes such as multimedia components and other hypermedia documents. That is, it is an application of the hypertext to multimedia documents [Flukiger, 1995]. This paper refers to a hypermedia document as a hyperdocument. Different from general digital documents, a hyperdocument has a unique characteristic of links to other ones by anchors, so that a hypermedia system may be called a navigation-based system [Isakowitz & Bieber, 1995]. Moreover, most hyperdocuments in a business application typically have functions for business works as interfaces of their systems. Therefore, managing the hyperdocuments may be a major approach in maintaining the hypermedia systems.

For managing hyperdocuments, metadata can be conceived as a significant and major tool. Most of the previous researches on metadata for digital documents have focused on information discovery [Chen et al., 1994; Hakala et al., 1996; Kashyap & Sheth, 1996; Dempsey & Heery, 1997]. However, corporate digital documents are tightly related to business works in an organization, so that the metadata needs to be considered from an organizational perspective (processes, or human and system resources).

Therefore, this paper has the objectives of proposing metadata classification and metadata schema for hyperdocuments and implementing meta-informa-

tion system based on the schema. These objectives will be performed considering not only information discovery but also organizational memory of systems or processes.

2. Metadata and Meta-Information System for Digital Documents

2.1 Metadata

The concept of digital documents is being conceived as a more evolved perspective that a future document will become a conceptual unit record linked to a network or Web [Sprague, 1995]. Accordingly, this paper defines the term, digital document, as recorded digital information unit formed with a single or composite structure in networked environments based on computer systems. If a digital document contains various media such as text, table, image, video, animation, and sound, it may be called a multimedia document. Such a multimedia document may be often called a compound document from the structural aspect [Sprague, 1995; Kalakota & Whinston, 1995].

Metadata is generally known as data of data or information of information. Metadata for digital documents has been described in various researches by focusing on mixed media [Chen et al., 1994], documents or services [Hakala et al., 1996], multimedia representations [Kashyap & Sheth, 1996], document object [Shutton], networked information [Dempsey & Heery, 1997], and web design [Tim, 1997].

Many researches in the past have concentrated on the use of metadata to support access to media- and application-specific information of digital documents. These metadata are concerned with various properties such as video [Jain & Hamper, 1994], images [Anderson & Stonebraker, 1994; Kiyoki et al., 1994], speech and text document [Glavitsch et al., 1994]. Different from these studies, this paper concentrates on hyperdocuments focusing on business applications as a type of compound documents. It would appear that there are not many works yet from this perspective.

As an attempt to make metadata standard for networked information resources, DLO (Document-Like Objects) [Weibel et al., 1995], including hyperdocuments, Dublin Core (Dublin Metadata Core Element Set) is noticeable [Dempsey & Weibel, 1996; Hakala et al., 1996; Weibel et al., 1995]. In the fourth workshop in 1997, the original 13-element Dublin Core defined in the first workshop in 1995 is extended to a 15-element set of version 2 with slightly modified

element names [Weibel & Miller, 1997a, 1997b; Weibel & Iannella, 1997]. Dublin Core elements are also focused on information discovery. Therefore, the elements have limitations for managing organizational documents as Murphy (1997) pointed out.

2.2 Meta-Information System

Many researches have described metadata roles in various aspects according to the focused document types or system environments [Dempsey & Heery, 1997; Hakala et al., 1996; Kashyap & Sheth, 1996; Neuss & Kent, 1995; Shklar et al. 1995; Weibel et al., 1995; Weibel & Miller, 1997a, 1997b]. Most roles discussed in these researches emphasized efficiency in document controls. However, the requirements for documents metadata are being enlarged to an organizational memory beyond document discovery [Meier & Sprague, 1996; Murphy, 1998]. From these studies, metadata roles can be divided into three levels: operation, system, and organization level. They can be summarized as in Table 1.

<Table 1> Summary of Metadata Roles

Level	Metadata Roles
Operation Level	<ul style="list-style-type: none"> ● Easy and fast access ● Increase of accuracy
System Level	<ul style="list-style-type: none"> ● Interoperability of heterogeneous environment ● Document maintenance ● Document distribution
Organization Level	<ul style="list-style-type: none"> ● Increase of reusability for information and knowledge resources ● Increase of capacity of business management ● Increase of organizational memory

A meta-information system is a system that realizes these metadata roles. Although the systems have various purposes and metadata schemas, they can be categorized into three domains: application-oriented,

hybrid, repository-oriented, depending upon whether the focus is on application functions or metadata schema as shown in Table 2.

The application-oriented meta-information systems focus on application functions, and metadata is used for supporting those functions. Therefore, metadata schema is determined based on requirements of the functions of the systems.

In contrast, the repository-oriented meta-information systems have major concerns on the contents of metadata schemas, and may often use supplementary functions for management of metadata. Although the concept of a repository has been defined differently [Ashrafi and Kuilboer, 1995; Moriarty, 1993; Tannenbaum, 1994], these definitions tend to be abstract, and have different perspective levels. Therefore, we redefine a repository as a physical storage area holding metadata based on information and knowledge resources such as data, models, or documents for organizational memory. In this sense, digital documents repository-oriented systems should attempt to manage organizational knowledge as well as information related to the documents in the long term. HyDoMIS proposed in this paper belongs to this domain.

The hybrid domain systems place emphasis on both a metadata schema and applicable functions based on it. As a standard for the metadata, USMARC (United States Machine-Readable Recognition Characteristics [MARC, 1998] is often regarded as critical for most of digital library systems. Another example may be an EDMS (Electronic Document Management System) [Sprague, 1995; Meier & Sprague, 1996]. These domain systems are mainly used for supporting major business functions. Accordingly, in this domain metadata scope is set to support document life-cycle control as well as information search.

The three types of meta-information systems for digital documents described until now, may be summarized as in Table 2.

<Table 2> Summary of Meta-Information System Domains for Digital Documents

Criteria \ Domain	Application-oriented	Hybrid	Repository-oriented
Main Focus	Application Functions	Metadata and Functions	Metadata Schemas
Major Task of System	Information Search and Exchange	Support of Business Functions	Management of Business and System Infrastructure
Major Users	Application End-Users	Application End-Users, Application Specialist	System Analyst, Information Manager, System Administrator
Example	Web Search Engine, E-mail System	Digital Library System, EDMS	HyDoMIS
Major Task of Metadata	Supporting Application Functions	Supporting Document Life-Cycle Control	Supporting Organization Memory

3. Metadata Classification for Hyperdocuments

A metadata classification plays a major role in determining the scope of its roles. Therefore, the metadata classification becomes a fundamental framework to provide useful metadata elements.

The past researches on metadata classification for digital documents [Bohm & Rakow, 1994; Shklar et al., 1995; Kashyap and Sheth, 1996] are inadequate to capture hyperdocument features from the technical and structural aspects as well as processes for organizational works. Therefore, we propose a metadata classification that can overcome these inadequacies in hyperdocuments as follows.

- *Content-oriented Metadata:* These metadata are used to understand the information of documents. The metadata include (i) information that depends on the contents directly and (ii) semantic meanings in a manner that may not be directly based on the content of the document as well.
- *Management-oriented Metadata:* These metadata are concerned with management of hypermedia applications.
 - *Workflow-dependent Metadata:* These metadata provide information about workflow related to a hyperdocument.
 - *Format-dependent Metadata:* These metadata describe information about formats related to hyperdocuments, as well as components such as nodes and interface sources.
 - *Storage-dependent Metadata:* These metadata provide information about storage concerned with hyperdocuments, and components, databases, and programs related to them
 - *Software-dependent Metadata:* These metadata provide information about software for document functions.
 - *Function-dependent Metadata:* These metadata describe information about functions related to hyperdocuments or their components.
 - *Log-dependent Metadata:* These metadata describe information about the history and the status of hyperdocuments.

This classification is divided into two categories of content-oriented and management-oriented metadata. The content-oriented metadata have been used for information discovery in most meta-information systems. On the other hand, as Dempsey and Heery (1996) addressed technical and business aspects of metadata, the proposed classification emphasizes the managerial roles of metadata. Therefore, management-oriented metadata is classified more specifically to reflect structural and technical features as well as business processes.

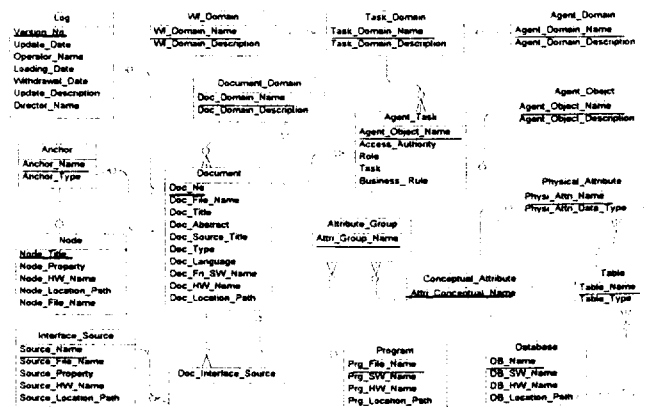
4. Metadata Elements for Hyperdocuments

Generally, metadata elements for digital documents are determined differently in their systems according to the characteristics of documents and purposes of the systems. In hyperdocuments, the metadata elements may be specified as in Table 3.

<Table 3> Metadata Specifications of Hyperdocument

Classifications		Elements
Content-oriented		[Document] Title, Abstract, Source Title, Attribute Group Name, Document Domain, Conceptual Attribute Name [Anchor] Name [Node] Title [Interface-Source] Name
Management-oriented	Workflow-dependent	Workflow Domain, Task Domain, Task, Agent Domain, Agent Object, Role, Business Rule
	Format-dependent	[Document] Language, Type [Anchor] Type [Node] Property [Interface-Source] Property [DB] Table Type, Physical Attribute Type
	Storage-dependent	[Document] File Name, H/W Name, Location Path [Program] File Name, H/W Name, Location Path [Node] File Name, H/W Name, Location Path [Interface-Source] File Name, H/W Name, Location Path [Database] Name, H/W Name, Location Path, Table Name, Physical Attribute Name
	Software-dependent	[Document] Functional S/W Name [Program] S/W Name [DB] S/W Name
	Function-dependent	[Document] Access Authority [Interface-Source] Function Description
	Log-dependent	Document Number, Version Number, Loading Date, Withdrawal Date, Update Date, Update Description, Director, Operator

These metadata elements are designed by E-R diagram for implementing a repository of HyDoMIS as shown in Figure 1. The repository was designed centered on the entity, "document."

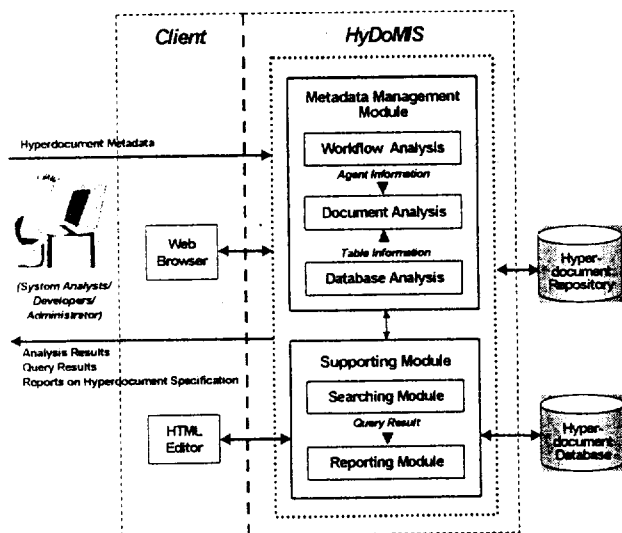


<Figure 1> Repository Schema for HyDoMIS

5. A Meta-Information System

This section introduces the prototype of a meta-information system for hyperdocuments, called Hyperdocument Meta-Information System (HyDoMIS). This system was constructed as a Web server based on Internet Information Server (IIS) 4.0 for multi-users such as developers or system administrators. Visual Basic Script was used to implement functional modules for menu or data control of a repository. The repository was developed by the use of Microsoft SQL Server 6.5.

HyDoMIS consists of two main modules: metadata management and supporting module. These modules have their sub-modules, as shown in Figure 2. The metadata management module takes responsibilities on metadata handling such as creating, editing, or deleting with some analysis mechanisms. The supporting module serves two types of search methods (by attributes and keywords) and reports on full document specification of a query result.



<Figure 2> HyDoMIS Architecture

6. Conclusions

Recently, the hyperdocuments are increasing rapidly as many organizations expand their business based on Internet. These documents should be conceived important for information resources and knowledge resources in an organization. Thus, it is important to manage hyperdocuments effectively.

In this paper, the concept of a meta-information system was defined. Then, a metadata classification for hyperdocuments from an organizational perspective was proposed to serve as a framework for specifying elements. Based on these researches, we are implementing the prototype of a meta-information system, HyDoMIS (Hyperdocument Meta-Information System). This system, a repository-oriented type, is expected to effectively maintain hypermedia applications. It can become a tool for organizational memory in the long term.

REFERENCES

- [Anderson & Stonebraker, 1994] J.T. Anderson and M. Stonebraker, "Sequoia 2000 Metadata Schema for Satellite Images," *ACM SIGMOD Record*, Vol.23, No4, pp42-48, Dec. 1994.
- [Ashrafi & Kuilboer, 1995] N. Ashrafi and J-P. Kuilboer: The Information Repository: A Tool for Metadata Management, *Journal of Database Management*, pp3-11, Spr. 1995.
- [Bohm & Rakow, 1994] K.Bohm, and T.C.Rakow. "Metadata for Multimedia Documents," *ACM SIGMOD Record*, Vol.23, No4, Dec. 1994.
- [Chen et al., 1994] F. Chen, M. Hearst, J. Kupiec, J. Pederson and L. Wilcox, "Metadata for Mixed-Media Access," *ACM SIGMOD Record*, Vol.23, No.4, pp.64-71, Dec. 1994.
- [Dempsey & Weibel, 1996] L. Dempsey and S.L. Weibel. "The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description," *D-Lib Magazine*, Jul/Aug. <http://www.dlib.org/dlib/july96/07weibel.html>, 1996.
- [Dempsey & Heery, 1997] L. Dempsey and R. Heery. "A Review of Metadata: A Survey of Current Resource Description Formats." Mar. <http://www.ukoln.ac.uk/metadata/DESIRE/overview/>, 1997.
- [Fluckiger, 1995] F. Fluckiger, "Understanding Networked Multimedia: Application and Technology," Prentice Hall, 1995.
- [Glavitsch et al., 1994] U. Glavitsch, P. Schauble and M. Wechsler. "Metadata for Integrating Speech Documents in a Text Retrieval System," *ACM SIGMOD Record*, Vol.23, No.4, Dec. 1994.
- [Hakala et al., 1996] J. Hakala, O. Husby, and T. Kochi. "Warwick Framework and Dublin Core Set Provide a Comprehensive Infrastructure for Network Resource Description." <http://www.bibsys.no/warwick.html>, 1996.
- [Isakowitz & Bieber, 1995] T. Isakowitz and M. Bierber, "Special Section: Navigation in information-intensive environments," *Journal of Management Information Systems*, Vol.11, No.4, pp5-8, Spr. 1995.
- [Jain & Hampapur, 1994] R. Jain and A. Hampapur, "Metadata in Video Databases," *ACM SIGMOD Record*, Vol.23, No.4, pp.27-33, Dec. 1994.
- [Kalakota & Winston, 1995] Kalakota and Winston, "Multimedia Document Database: Representation," *Query Processing and Navigation, Journal of Database (in Korea)*, Vol.1, No.1, 1995.
- [Kashyap & Sheth, 1996] V. Kashyap, and A. Sheth, "Semantic Heterogeneity in Global Information Systems: The Role of Metadata. Context and Ontologies," *Cooperative Information Systems: Current Trends and Directions*, M. Papazoglou and G. Schlageter (Eds.), <http://lstdis.cs.uga.edu/~kashap/mikep-chapter.ps>, 1996.
- [MARC, 1998] Library Congress. "MARC Standards," <http://lcweb.loc.gov/marc/>, 1998.
- [Meier & Sprague, 1996] J. Meier and R. Sprague, "Towards a Better Understanding of Electronic Document Management," *Proceedings of the 29th Hawaii International Conference on System Science*, Vol.V, pp.53-61, 1996.
- [Moriarty, 1993] T. Moriarty, "Tools for Managing Metadata. *Database Programming and Design*," Vol.6, No.10, 1993.
- [Murphy, 1998] L.D. Murphy, "Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions," *Proceedings of the 31st Hawaii International Conference on System Science*, Vol.11, 1998.
- [Neuss and Kent, 1995] C. Neuss and R.E. Kent, "Conceptual Analysis of Resource Meta-information," *Proceedings of WWW'95*, <http://www.idg.fhg.ed/www/WWW95/papers/94/www3.htm>, 1995.
- [Shklar et al., 1995] L. Shklar, A. Sheth, V. Kashyap, and K. Shah, "InfoHarness: Use of Automatically Generated Metadata for Search and Retrieval of Heterogeneous Information," *Proceedings of CAISE 95*, Jun. 1995.
- [Sprague, 1995] R.H. Sprague, Jr., "Electronic document management: Challenges and Opportunities for Information Systems Managers," *MIS Quarterly*, pp.29-49, Mar. 1995.
- [Sutton, 1996] M.J.D. Sutton, "Document Management for the Enterprise - Principles, Techniques, and Applications," Wiley, 1996.
- [Tannenbaum, 1994] A. Tannenbaum, "Implementing a Corporate Repository: The Models Meet Reality," *John Wiley & Sons, Inc.*, 1994.
- [Tim, 1997] B.L.Tim, "Metadata Architecture," <http://www.w3.org/DesignIssues/Metadata/1997>.
- [Weibel, 1995] S. Weibel, Metadata. "The Foundations of Resource Description," *D-Lib Magazine*, <http://www.cnri.reston.va.us/home/dlib/July95/07weibel.html>, Jul. 1995.
- [Weibel et al, 1995] S. Weibel, J. Godby and E. Miller, "OCLC/NCSA Metadata Workshop Report," http://www.oclc.org:5040/oclc/research/metadata/dublin_core_report.html, 1995.
- [Weibel & Iannella, 1997] S. Weibel and R. Iannella, "The 4th Dublin Core Metadata Workshop Report," *D-Lib Magazine*, <http://www.dlib.org/dlib/june97/metadata/06weibel.html>, Jun. 1997.