# Developing a Meta-Information System for Hyperdocuments

Woojong Suh and Heeseok Lee
Graduate School of Management, KAIST

## Abstract

Recently, many organizations have attempted to construct hypermedia systems to expand their working areas to Internet-based virtual work places. It is important to manage hypermedia documents, hyperdocuments, effectively. Metadata play a critical role in document management. This paper proposes a meta-information system, HyDoMiS (Hyperdocument Meta-information System), employing metadata for managing hyperdocuments. HyDoMiS performs three functions: metadata management, search, and reporting. The metadata management function includes workflow analysis, document analysis, and database analysis. The system is designed to help system analysts or system administrators develop and maintain hypermedia application better.

## 1. Introduction

Today, hypermedia documents are growing explosively in many organizations because of many attempts to develop hypermedia systems employing intranets or extranets for electronic commerce. These hypermedia-based systems include hypermedia documents (hyperdocuments) for supporting organizational tasks. Unlike hyperdocuments often included in CD-ROM titles for games or educational materials, hyperdocuments employed in organizational tasks will be referred to as organizational hyperdocuments. Such documents typically play a critical role in business, in the form of, for example, invoices, checks, or orders. The maintenance of organizational hyperdocuments is becoming a burdensome task for many organizations today; managing organizational hyperdocuments needs is as important to the economic success of an organization as software maintenance is [3]. A hypermedia document, a special type of digital document, is based on the inter-linking of nodes such as multimedia components and other hypermedia documents [24]. That is, a hyperdocument is an application of hypertext technologies to multimedia documents [8]. In contrast to other digital documents, a hyperdocument has links to various nodes, so-called hyperlinks, which are used as path for the navigation, so that a hypermedia system may be called a navigation-based system [11].

Metadata have been conceived as a major weapon for managing hyperdocuments. Most of the previous studies on metadata for digital documents have investigated from a technical perspective such as information discovery [4, 6, 10, 14]. However, corporate digital documents are closely related to business tasks in an organization. In this context, organizational hyperdocuments typically have complex relationships with both information and processes for business. The organizational hyperdocuments of

the various business forms can impact on the speed of communications and the productivity of business processes. Accordingly, organizational hyperdocuments should be designed to support collaboration among workers in business processes. Furthermore, such documents should be considered from a technical aspect. The system resources used by organizational hyperdocuments are a considerable part of organizational assets. Therefore, the metadata for the documents need to be studied from an organizational perspective that considers business processes as well as human or system resources.

Two objectives of this paper are (i) to propose metadata classification and metadata schema for hyperdocuments and (ii) to implement a meta-information system on the basis of the schema. The system was designed to support the maintenance of organizational hyperdocuments.

## 2. Metadata and Meta-Information System

Metadata are generally known as data of data or information of information. Metadata for digital documents have been explored from various researches perspectives: mixed media [4], documents or services [10], multimedia representations [14], document objects [29], networked information [6], and web design [30]. Many past researches have concentrated on the use of metadata to support an access to media- and application-specific information of digital documents. These metadata are related to various system properties, such as video [12], images [1, 16], or speech and text document [9]. In contrast to these researches on media-specific metadata, there was a suggestion that a technique of using media-integrated metadata should be developed for the management of documents with heterogeneous properties integrally [9, 26]. It has thus attempted to develop systems with this purpose [21, 26]. Meanwhile, this paper focuses on controlling hyperdocuments, a kind of compound document. In this context, there has been a noticeable attempt to develop and enhance a metadata set, Dublin Core (Dublin Metadata Core Element Set), which deals with networked information resources, so-called DLO (Document-Like Objects) [7, 10, 31]. 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 [32, 33]. However, the major focus of Dublin Core elements is on information discovery, but they have some limitations in managing organizational hyperdocuments, as Murphy (1997) pointed out [22].

A number of studies have described metadata roles in various technical aspects: document type or system environments [6, 10, 14, 23, 26, 31]. The emphasis of most roles discussed is on efficiency in document access control or interoperability of heterogeneous documents. However, the requirements for document metadata are relatively extended to an organizational memory (OM) beyond document discovery, because organizational documents are a major part of organizational memory [20, 22, 28]. In this context, meta-information system can evolve toward managing OM by extending the metadata scope of organizational documents to capture their history in terms of business functions, communication mechanisms, or technical artifacts, beyond focusing on contents discovery. These memories may provide

knowledge to support various decision for controlling communication mechanisms in a business process, linking to the past responsible workers, or maintaining the hypermedia applications. Metadata roles can be summarized in three levels (operation, system, and organization) as follows.

<Table 1> Metadata Roles

| Level | Metadata Roles |
|---|---|
| **Operation** | ● Easy and fast access<br>● Increased accuracy |
| **System** | ● Interoperability under heterogeneous environment<br>● Document maintenance<br>● Document distribution |
| **Organization** | ● Increased reusability of information and knowledge resources<br>● Increased capacity of business management<br>● Increased organizational memory |

A meta-information system may be defined as a system implemented for realizing metadata roles on the basis of metadata databases. The meta-information systems can be characterized by information resources to be controlled or services supported by the system. The meta-information systems may be categorized into three domains: application-oriented, hybrid, or management-oriented, from the perspective of their applicable purposes.

The application-oriented meta-information systems focus on application functions, and metadata are used for supporting the functions. Therefore, metadata schemas are primarily determined on the basis of requirements of functions in a system. One example is a type of web search engine, a so-called meta-search engine. The main task of meta-search engines is information search, so that their metadata are determined so as to support their search mechanisms. The main users of this system domain may be application end-users.

In contrast, the management-oriented meta-information systems play a major role in supporting the reuse and maintenance of managerial considerations concerned with information, knowledge, or system resources of an organization. In a document-oriented aspect, a management-oriented meta-information system should serve managerial capabilities for the system- and business-related information or knowledge, through the management of the metadata on organizational documents. From this perspective, HyDoMiS is an example in this domain. The main users of a system in this domain may be system analysts, information managers, or system administrators.

The systems of hybrid domain pay attention to metadata for the managerial purposes, as well as specific application functions. Accordingly, the major users may include not only application but also end-users or application specialists. Examples of this domain include EDMS (Electronic Document Management System) [20, 27, 28]. The EDMS requires metadata as essential components for the document handling [29]. This system places emphasis on the productivity of processes through the efficient control of document life cycle that includes creation, retrieval, transmission, and expiration.

## 3. Metadata Classification and Elements for Hyperdocuments

A metadata classification, a fundamental framework for providing metadata elements, can reflect the scope of their roles. Bohm and Rakow (1994) [2] proposed a metadata classification for multimedia documents, and it has impacted on other studies. The classification focuses on the representation of media type, content, relationships among document components, history, and location. There was another series of research concerning a classification of metadata for information access and retrieval with its applicable systems [13, 14, 21, 25, 26]. The metadata classification provided by these studies is specified by whether or not metadata are dependent on content, and whether they are dependent directly or indirectly if they are dependent on content. These two kinds of classification impact on other studies, such as on that of model for medical documentation by the use of hypermedia [5] and on the quality of service management in distributed multimedia systems [15].

However, the metadata classifications of these studies place major emphasis on the use of metadata for information discovery, so that they have some limitations for the management of organizational hyperdocuments for productivity. This study proposes a metadata classification for organizational hyperdocuments, as follows:

- *Content-dependent Metadata*: These metadata are used to enable understanding the contents of documents. The metadata include information that depends on (i) the contents directly, and (ii) semantic meanings not directly based on the content of the document.
- *Workflow-dependent Metadata*: These metadata provide information about workflow related to an organizational hyperdocument.
- *Format-dependent Metadata*: These metadata describe information about formats related to organizational hyperdocuments, as well as hypermedia components, such as nodes and interface sources.
- *System-dependent Metadata*: These metadata provide information concerned with storage- and software-related information of system resources, such as hyperdocuments, interface sources, and databases.
- *Log-dependent Metadata*: These metadata describe information about the history and the status of organizational hyperdocuments.

Metadata elements for digital documents have been typically determined differently according to the characteristics of documents and purposes of their systems. This paper focuses on hyperdocument maintenance in consideration of processes and system artifacts. From this perspective, detailed metadata elements will be provided under the classification suggested above. Most of the organizational hyperdocuments in business applications typically perform complex functions that are often connected with a corporate database for business tasks in a workflow. For the maintenance of such documents, metadata elements may be specified as follows.

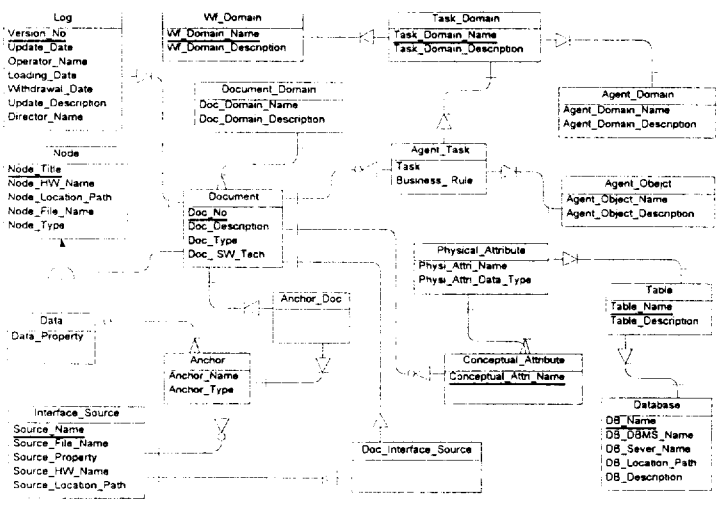
<Table 2> Metadata Elements of Organizational Hyperdocuments

| Classifications | Elements |
|---|---|
| **Content-dependent** | [Document] Title, Description, Document Domain Name, Conceptual Attribute Name [Anchor] Name [Data Node] Title [Interface-Source] Name |
| **Workflow-dependent** | Task Domain Name, Task, Agent Domain Name, Agent Object Name, Business Rule |
| **Format-dependent** | [Document] Type [Anchor] Type [Node] Type, [Data Node] Property [Interface-Source] Property [DB] Physical Attribute Type |
| **System-dependent** | [Document] File Name, H/W Name, Location Path, S/W Technology [Data Node] File Name, H/W Name, Location Path [Interface-Source] File Name, Storage, Location Path [Database] Name, H/W Name, Location Path, Table Name, Table Type, Physical Attribute Name, DBMS Name |
| **Log-dependent** | Document Number, Version Number, Loading Date, Withdrawal Date, Update Date, Update Description, Director, Operator |

## 4. Implementing HyDoMiS

This paper introduces a meta-information system for organizational hyperdocuments, called Hyperdocument Meta-information System (HyDoMiS), constructed so as to manage organizational hyperdocuments effectively through their metadata. This system may affect economic success in maintaining organizational hypermedia applications based on intranet or extranet. HyDoMiS was constructed as a Web server based on Internet Information Server (IIS) 4.0 for multi-users such as developers or system administrators. These users can access the HyDoMiS through Web browsers in the client side. The Visual Basic script based on ASP technology was used primarily for implementing functions for dynamic navigation, metadata controls (creating, editing, and deleting), search and reporting. The metadata DB was developed with Microsoft SQL Server 6.5.

The metadata schema is produced based on metadata elements proposed in Table 2 (in section 3). The schema was designed by E-R diagram for implementing a metadata database for HyDoMiS as shown in Figure 1. The schema was centered on document entity.

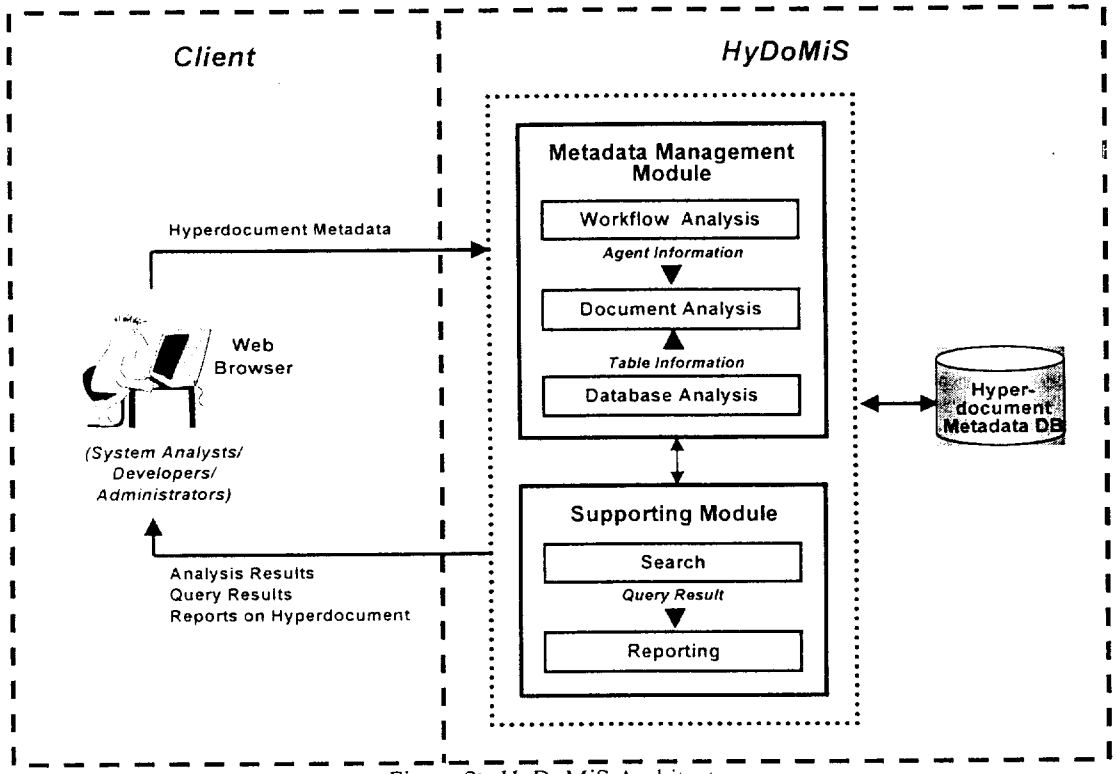



<Figure 1> Metadata DB Schema of HyDoMiS

This schema represents complex relationships among the components employed in hyper-document operations, and contains managerial factors for the control of task or system. Such schema contents can be captured effectively through the processes of the hypermedia development metho-dology, WHDM (Workflow-Based Hypermedia Development Metho-dology) [19] rather than other

methodologies such as VHDM (View-based Hypermedia Design Methodology) [17] or SOHDM (Scenario-based Object-Oriented Hypermedia Design Methodology) [18]. WHDM employs a document-based workflow model to capture the requirements for organizational hyperdocuments to be implemented.

HyDoMiS consists of two main modules: metadata management and a supporting module. These modules have their sub-modules, as shown in Figure 2. The metadata management module takes responsibilities for metadata handling such as creating, editing, or deleting with some analysis mechanisms. The supporting module serves two types of functions for searching a hyperdocument and reporting its meta-information. These functions are based on a hyperdocument metadata database.

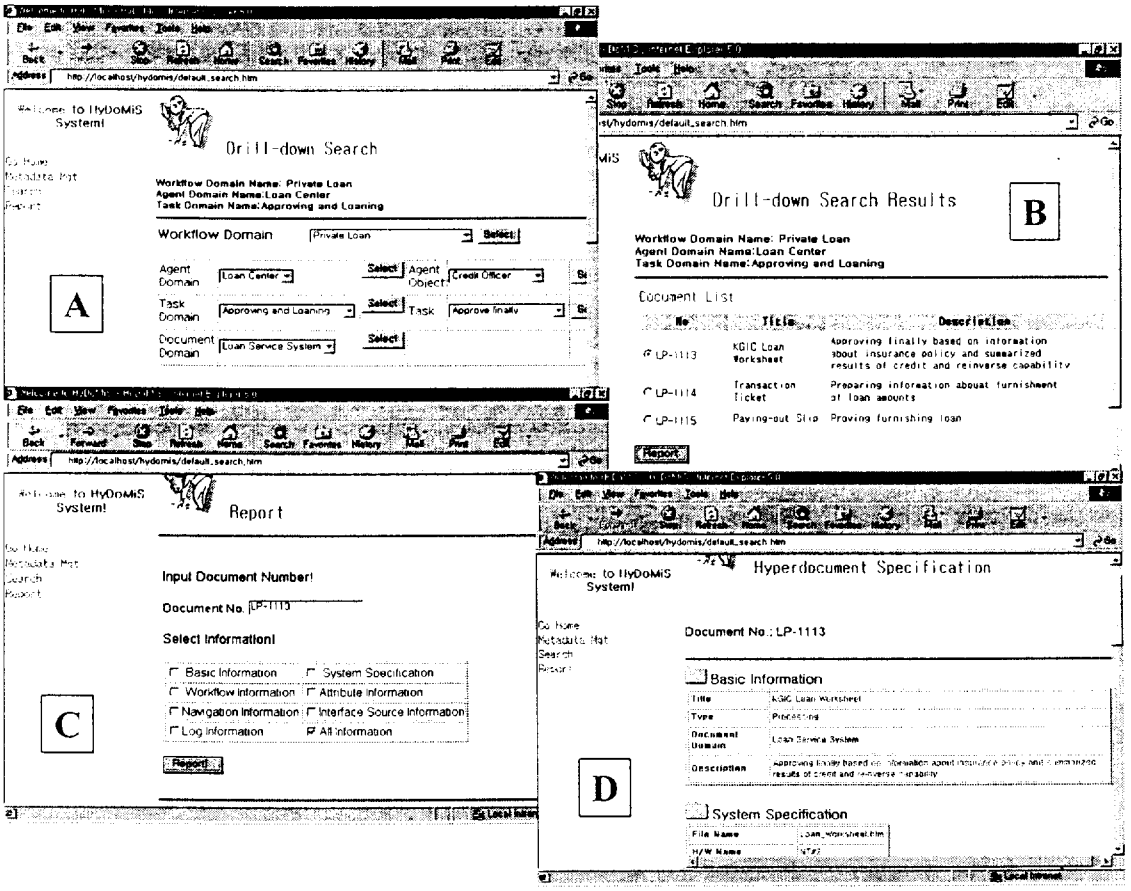



<Figure 2> HyDoMiS Architecture

Workflow information enables us to understand the hyperdocuments' roles in their processes. This information is concerned with workflow domains, agents, tasks, and business rules of an organizational hyperdocument. The document management module is composed of five sub-modules related to a hypermedia system, data attributes, navigation, interface sources, and log information. The system-related information focuses on the hardware, and the software technologies. The attribute information is concerned with data provided by a corporate database. This sub-module, therefore, can provide complete information, so long as database-related information is provided from the database information sub-module. The navigation information sub-module provides meta-information in terms of two kinds of nodes (destination node and source node). That is, for a certain hyperdocument, we can get not only information of source nodes which can go the hyperdocument, but also information about destination nodes which are differentiated by two types, document node and data node. The interface source

information is useful to increase the reusability of multimedia components represented on hyperdocuments. The database information module deals with information about databases connected to hyperdocuments. The information produced in this module can be used for the control of the connection between hyperdocuments and a database in the document module.

The search module provides the numbers and titles of searched hyperdocuments as a result. The search module employs two approaches: drill-down search and keyword search. The drill-down search uses a mechanism to narrow the domain to find a hyperdocument. This search module is implemented as following Figure 3.




<Figure 3> Screens for Drill-down Search and Reports

This search can be performed by the use of various domains in terms of workflow, database, navigation, interface source, and log information, as shown in Screen A. The result produced (Screen B) in a search module can be transferred to a reporting module automatically, in order to generate a report on searched document. The report (Screen D) can provide meta-information whose nature will depend on which reporting domains are selected (Screen C). The keyword search uses a typed keyword complying with the selected items, such as workflow name, document title, abstract, or anchor name.

## 5. Conclusions

Recently, many organizations have expanded their business by the use of Internet technologies. Organizational hyperdocuments are critical resources for such organizations. Managing the documents may impact on the success of business.

In this paper, we propose a meta-information system, HyDoMiS (Hyperdocument Meta-information System), on the basis of metadata database for the effective management of hypermedia resources. In order to generate a metadata schema for HyDoMiS, a metadata classification for hyperdocuments is studied, and metadata elements are thus specified. The metadata schema is developed from an organizational perspective in terms of processes and technical artifacts. HyDoMiS is constructed as a Web server for a variety of users, such as system analysts, information managers, or system administrators. HyDoMiS is expected to become a repository for organizational memory, in the long term.

Another contribution of this paper is to redefine the complex relationships among the components employed in hyperdocument operations, and to capture the definitions into a metadata schema for the HyDoMiS metadata database.

On the basis of the current research, we are in the process of incorporating SGML (Standard Generalized Markup Language) into HyDoMiS functions. Another research challenge is to apply HyDoMiS to a futuristic knowledge repository.

## References


[1] J.T. Anderson and M. Stonebraker, "Sequoia 2000 metadata schema for satellite images," *ACM SIGMOD Record*, 23(4), 1994, pp. 42-48.

[2] K. Bohm and T.C. Rakow, "Metadata for multimedia documents," *ACM SIGMOD Record*, 23(4), 1994, pp. 21-26.

[3] P. Brereton, D. Budgen, and G. Hamilton, "Hypertext: The next maintenance mountain," *IEEE Computer*, Dec. 1998, pp. 49-55.

[4] F. Chen, M. Hearst, J. Kupiec, J. Pederson and L. Wilcox, "Metadata for mixed-media access," *ACM SIGMOD Record*, 23(4), 1994, pp. 64-71.

[5] F. Consorti, P. Merialdo, and G. Sindoni, "Metadata reference model for medical documentation: A hypermedia proposal," *Proceedings of the 1st IEEE Metadata Conference*, (http://www.computer.muni.cz/conferen/meta96/sindoni/ieee.html), Apr. 16-18, 1996.

[6] L. Dempsey and R. Heery, "A review of metadata: A survey of current resource description formats," (http://www.ukoln.ac.uk/metadata/DESIRE/overview/), Mar. 1997.

[7] L. Dempsey and S.L. Weibel, "The Warwik metadata workshop: A framework for the deployment of resource description," *D-Lib Magazine* (http://www.dlib.org/dlib/july96/ 07webel.html), Jul./Aug. 1996.

[8] F. Fluckiger, *Understanding Networked Multimedia: Applications and Technology*, Prentice Hall, 1995.

[9] U. Glavitsch, P. Schauble and M. Wechsler, "Metadata for integrating speech documents in a text retrieval system," *ACMI SIGMOD Record*, 23(4), 1994, pp. 57-63.

[10] J. Hakala, O. Husby, and T. Koch, "Warwick framework and Dublin core set provide a comprehensive infrastructure for network resource description," (http://www.bibsys.no/warwick.html), 1996.

[11] T. Isakowitz, and M. Bieber, "Special section: navigation in information-intensive environments," *Journal of Management Information Systems*, 11(4), 1995, pp. 5-8.

[12] R. Jain and A. Hampapur, "Metadata in video databases," *ACM SIGMOD Record*, 23(4), 1994, pp. 27-33.

[13] V. Kashyap, K. Shah, and A. Sheth, "Metadata for building the multimedia patch quilt," in *Multimedia Database System: Issue and Research Directions*, S. Jajodia and V. Subrahamaniam (Eds.), Springer-Verlag, 1995.

[14] V. Kashyap, and A. Sheth, "Semantic heterogeneity in global information systems: The role of metadata, context and ontologies," in *Cooperative Information Systems: Current Trends and Directions*, M. Papazoglou and G. Schlageter (Eds.), (http://lsdis.cs.uga.edu/~kashap/ mikep-chapter.ps), 1996.

[15] B. Kerherv, A. Pons, G.V. Bochmann and A. Hafid, "Maetadata modeling for quality of service management in distributed multimedia systems," *Proceedings of the 1st IEEE Metadata Conference*, (http://www.computer.muni.cz/conferen/meta96/sindoni/ieee. html), Apr. 1996.

[16] Y. Kiyoki, T. Kitagawa and T. Hayama, "A metadatabase system for semantic image search by a mathematical model of meaning," *ACM SIGMOD Record*, 23(4), 1994, pp. 34-41.

[17] H. Lee, J. Kim, Y. Kim, and S. Cho, "A view based hypermedia design methodology," *Journal of Database Management*, 10(2), 1999, pp. 3-13.

[18] H. Lee, C. Lee, and C. Yoo, "A scenario-based object-oriented methodology for developing hypermedia information systems," *Information and Management (In press)*, 1999.

[19] H. Lee and W. Suh, "A workflow-based methodology for developing hypermedia information systems," *Proceedings of the 5th International Conference of the Decision Sciences Institute (In press)*, Athens, Greece, Jul. 1999.

[20] J. Meier and R. Sprague, "Towards a better understanding of electronic document management," *Proceedings of the 29th Hawaii International Conference on System Science*, 5, 1996, pp. 53-61.

[21] E. Mena, V. Kashap, A. Sheth and A. Illarramendi, "OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies," *Proceedings of the First IFCIS International Conference on Cooperative Information Systems (CoopIS '96)*, Brussels, Belgium, Jun. 1996.

[22] L.D. Murphy, "Digital document metadata in Organizations: Roles, analytical, approaches, and future research directions," *Proceedings of the 31th Hawaii International Conference on System Science*, 2, 1998.

[23] 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.

[24] J. Nielsen, *Hypertext and hypermedia*, Academic Press Professional, 1993.

[25] A. Sheth and V. Kashyap, "Media-independent correlation of information: What? how?," *Proceedings of the 1st IEEE Metadata Conference*, (http://www.computer.muni.cz/conferen/ meta96/sindoni/ieee.html), Apr. 16-18, 1996.

[26] 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.

[27] R. H. Sprague, Jr. and C. B.C. McNurlin, *Information Systems Management in Praktice*, Prentice-Hall International Editions, 1993.

[28] R.H. Sprague, Jr., "Electronic document management: Challenges and opportunities for information systems managers," *MIS Quarterly*, Mar. 1995, pp.29-49.

[29] M.J.D. Sutton, *Document Management for the Enterprise - Principles, Techniques, and Applications*, Wiley, 1996.

[30] B.L. Tim, "Metadata architecture," (http://www.w3.org/DesignIssues/Madtadata.html), 1997.

[31] 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.

[32] S. Weibel and R. Iannella, "The 4th Dublin core metadata workshop report," *D-Lib Magazine*, (http://www.dlib.org/jun97/metadata/06weibel.html), Jun. 1995.

[33] S. Webel and E. Miller, "Description of Dublin core elements," (http://purl.oclc.org/ metadata/dublin_core/syntex.html), 1997.