Hypermedia Document Management: A Metadata and Meta-Information System

WOOJONG SUH, Korea Advanced Institute of Science and Technology, Korea
HEESEOK LEE, Korea Advanced Institute of Science and Technology, Korea

Recently, many organizations have attempted to build hypermedia systems to expand their working areas into Internet-based virtual workplaces. Thus, it is important to manage corporate hypermedia documents effectively. Metadata plays a critical role for managing these documents. This paper identifies metadata roles and components to build a metadata schema. Furthermore, a meta-information system, HyDoMiS (Hyperdocument Meta-information System) is proposed by the use of this metadata schema. HyDoMiS performs three functions: metadata management, search, and reporting. The metadata management function is concerned with workflow, documents, and databases. The system is more likely to help implement and maintain hypermedia information systems effectively.

INTRODUCTION

Today, hypermedia documents are growing explosively in many organizations because of the large number of attempts to develop systems employing intranets or extranets for electronic commerce (EC). These systems include hypermedia documents (hyperdocuments) for supporting organizational tasks. Hyperdocuments employed for such tasks are referred to as organizational hyperdocuments (OHDs). They typically play a critical role in business, in the form of, for example, invoices, checks, or orders. The maintenance of OHD is becoming a burdensome task; managing their needs is as important to economic success as is software maintenance (Brereton et al., 1998). A hypermedia document—a special type of digital document—is based on the interlinking of nodes, such as multimedia components, etc. (Nielsen, 1993); i.e., it is an application of hypertext technologies employing multimedia components (Fluckiger, 1995). In contrast to other digital documents, a hyperdocument has links to various nodes, “hyperlinks,” which are used as a path for the navigation.

Most of the previous studies on metadata for digital documents have investigated the topic from a technical perspective, such as information discovery. However, corporate digital documents are closely related to business tasks in an organization. In this context, OHDs typically have complex relationships with both information and business processes. The OHDs can impact the speed of communications and the productivity of business processes. Accordingly, OHDs should be designed to support collaboration among workers in business processes. This aspect needs to be considered in defining metadata of the OHDs for their effective management. Furthermore, such documents should also be considered from a technical aspect. The system resources used by OHDs are a considerable part of the organizational assets.

The two objectives of this paper are (i) to propose metadata classification and metadata schema for OHDs, and (ii) to implement a meta-information system on the basis of the schema. The system was designed to support the maintenance of OHDs. Our research is rooted in previous studies on the various types of multimedia documents so as to capture the more generic perspective of metadata.

METADATA AND META-INFORMATION SYSTEM

Metadata is generally known as data about data (or information about information). Metadata for digital documents has been explored from various research perspectives: mixed media (Chen et al., 1994), multimedia representations (Kashyap & Seth, 1996), document objects (Sutton, 1996), and networked information resources (Dempsey & Weibel, 1996). Much past research has concentrated on the use of metadata to support access to media- and application-specific documents. This metadata describes various system properties, such as video (Jain & Hampapur, 1994), images (Anderson & Stonebraker, 1994; Kiyoki et al., 1994), or speech and text document (Glavitsch et al., 1994). In contrast to these, it has been suggested that media-integrated metadata should be developed for the management of documents with heterogeneous properties. There have been attempts to do this (Mena et al., 1996; Shklar et al., 1995).
These studies have described metadata roles in various technical aspects from the perspective of document types or system environments. Efficiency in document access control or interoperability of heterogeneous documents has been discussed as the prime problems of these systems. A set of hyperdocument metadata, the Dublin Core (Dublin Metadata Core Element Set) (Dempsey & Weibel, 1996; Weibel et al., 1995; Weibel & Iannella, 1997), has also focused on the information discovery; these are some of the difficulties in managing OHDs (Murphy, 1998).

Metadata of OHDs should be considered beyond the technical aspects by including an organizational aspect toward organizational memory (OM) because they are a major source of organizational memory (Meier & Sprague, 1996; Murphy, 1998; Sprague, 1995). The concept of OM has many facets, but most authors agree that OM must support decisions by using OM techniques for managing an organization’s information or knowledge of the past (Shum, 1997; Stein & Zwass, 1995; Wijnhoven, 1998). In this context, a meta-information system for OHDs can evolve toward managing OM by extending their metadata scope 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 decisions for controlling communication mechanisms in a business process, linking to the previous responsible workers, or maintaining the hypermedia applications. Metadata roles can be summarized in three levels (operation, system, and organization) as shown in Table 1.

A meta-information system can be characterized by information resources (to be controlled) or supported services; they service three types of domains: application-oriented, hybrid, or management-oriented.

Application-oriented meta-information systems use metadata to support the application functions. Therefore, metadata schemas are primarily determined on the basis of system requirements. One example is a type of Web search engine. Its main task is to search for information. The main users of this system domain may be application end-users.

In contrast, management-oriented meta-information systems play a major role in supporting the reuse and maintenance of managerial resources. In a document-oriented environment, these systems should serve managerial capabilities for the system- and business-related information or knowledge, through the management of the metadata on organizational documents; major users may be system analysts, information managers, or system administrators.

Hybrid domain systems pay attention to metadata for the managerial purposes, as well as specific application functions. Accordingly, the major users may not only include application managers but also end-users or application specialists. Examples of this domain include EDMS (Electronic Document Management System), which requires metadata as an essential component for document handling (Sutton, 1996).

**METADATA CLASSIFICATION AND ELEMENTS FOR HYPERDOCUMENTS**

**Metadata Classification**

Metadata classification can be perceived as a fundamental framework for providing metadata elements. The roles of the elements are determined by the classification coverage. Bohm and Rakow (1994) proposed a metadata classification for multimedia documents. It focuses on the representation of media type, content, relationships among document components, history, and location. Another metadata classification is specified by the dependency on the content (Kashyap & Sheth, 1996; Mena et al., 1996; Shklar et al., 1995). On the basis of these two kinds of classification, metadata for managing medical documentation using hypermedia (Consorti et al., 1996) and quality of service in distributed multimedia systems (Kerherv et al., 1996) have also been developed.

This study proposes the following metadata classification for organizational hyperdocuments:

- **Content-Dependent Metadata**: This metadata is used to enable understanding of the content of documents. The metadata includes information that depends on (i) the content directly, and (ii) semantic meanings based on the content of the document indirectly.
- **Workflow-Dependent Metadata**: This metadata provides information about workflow related to an organizational hyperdocument.
- **Format-Dependent Metadata**: This metadata describes information on formats related to organizational hyperdocuments, as well as hypermedia components, such as nodes and interface sources.
- **System-Dependent Metadata**: This metadata provides information concerned with storage- and software-related information on system resources, such as hyperdocuments, interface sources, and databases.