Distributed Object-Based Applications for Document Management

Satoshi Wakayama
Yasuki Ito
Toshihiko Fukuda
Kanji Kato

OVERVIEW: We have developed an Enterprise Document Management System which supports document management in offices. This system is intended to construct business systems for the next generation. We designed this system with the aim of handling a document, in the mission critical business application system. The Enterprise Document Management System combines numeric information with documents. Our Enterprise Document Management System adopts a standardized specification called “DMA (Document Management Alliance) 1.0”. DMA specifies the document object model and the API (application programming interface). This standard has been proposed by DMA which is one of the task forces of AIIM (Association for Information and Image Management International). AIIM is the international standardization organization of document management. We expand the DMA specification and implement its facilities for managing structured documents such as SGML (standard generalized markup language) and XML (extensible markup language). This system is built on ORDB (object relational database), a database system in order to integrate document with numeric information. Database systems are added to the document management oriental functions, such as full-text retrieval. We are constructing a document oriented business application with the Enterprise Document Management System.

INTRODUCTION
WE write many documents, every day, but we do not understand how the documents are constructed. We use a lot of data, but we do not control these data.

For example, when we want to build a product catalog management system, we have to manage the document body and some meta-information such as price and brand name.

In modern on-line business systems, the meta-information is stored on RDB (relational database) and the document body is stored outside of RDB. And the document is bringing up to a more complicating one. We usually use many kinds of media in documents, such as “Text” and “Image.” And many authors write many components in a document. So, we have to manage the configuration of the meta-information and the document body.

We usually use document information in the mission critical business. Some Internet/intranet systems, such as WWW (World Wide Web) and Groupware, already use our business system.

We have decided to offer the “Enterprise Document Management System” in order to solve these problems. This system adopts the standardized specification of DMA (document management alliance) as a basic function in order to manage document information. It also expands DMA specification to support the function that is handling the structured document.

All document management functions, such as configuration management of a document, search, and partial extraction are built in DMA API and structured document interface. The “Enterprise Document Management System” is built on ORDB. This document describes the usability of our Enterprise Document Management System in business process.

BUSINESS ACTIVITIES BUILT ON DOCUMENTS
Based on surveys from the processing departments of various manufacturing companies, documents are necessary in regard to planning, production, and sales. They are one of the most important elements in each business process.

For example, people working in sales departments receive customer information for promoting and receiving purchase order activities. We should write down for share this information. And, this type of information should lead us.
A document is a result of an activity and this activity constructs the business process. We can understand “Why,” “When,” “Who,” “How” and “What” from a document. We use documents from previous activities, as information for current activities. Hence, business activities are controlled by documents.

With this point of view, we need to offer a system which enables associates to retrieve information from a document in order to help the organization process (Fig. 1).

**DATABASE CONSTRUCTION OF A DOCUMENT**

Subject of A Database Construction

A common framework is needed to build a database that includes general documents created by a word processor, etc., so that the documents can be shared and used several times. In our system, we have:

1. Allocated the sentence to record RDB
   This way ties the access method and the authorizing method to record.
2. Stored meta-information and documented the file system in RDB
   In this case, the document management system cannot manage the contents in a document. Therefore, they cannot be reused easily. People may use some kind of word processor to manage the contents in their documents. If we want to share a document, we have to prepare many conversion tools.

**Standardization of Access Methods and Content**

We adopt DMA as an access method and standard generalized markup language (SGML) as a data format with schema.

In RDB (relational database), SQL (structured query language) is the standardized access method consisting of robust search functions. The scheme which specifies the common data structure and data format independent of applications, plays a large role with regard to sharing and reusing numerical data. In document management systems, the access method and content format should also be standardized for sharing and reusing.

DMA was built in April 1995 as a task force of AIIM (Association for Information and Image Management International). The purpose of DMA is to define the specification of a document management system.
interface that is used by document handling application. At the end of 1997, “DMA1.0” became the standard specification of DMA.

SGML is a standardized information description language, which is ISO 8879:1986. It can describe a structure and a schema by itself. In W3C (World Wide Web Consortium), they built the information system with markup based on SGML, as XML.

Specification of DMA and Expansion by Hitachi Object Model and the Expansion of DMA

Fig. 2 shows the object model of DMA. It consists of an object group, a container, the relation between a container and a document, and a logical document. A logical document has version series, renditions and content elements.

Each object class has a fundamental role to realize the function mentioned above (Table 1), and operation corresponding to it is specified as the standardized API.

We expand the DMA specification in order to add structured document management functions to SGML and XML. Both SGML and XML document have an internal structure in a document (DTD: document type definition) which enables the internal data structure in the document to be in common and to be standardized (Fig. 3).

Our Enterprise Document Management System has a function for analyzing and managing SGML and XML documents that performs structure searches and extraction of elements.

We developed these functions of ORDB using “SGML TEXT plug in.”

System Configuration

Fig. 4 shows an example of the system configuration in the Web environment using the Enterprise Document Management System. In this distributed object-oriented business system, the Enterprise Document Management System is used as the middleware to manage and handle the document objects.

Table 1. The Role of DMA Objects

<table>
<thead>
<tr>
<th>Name of class</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>Including document and container</td>
</tr>
<tr>
<td>Configuration history</td>
<td>An object of controlling multi-version series</td>
</tr>
<tr>
<td>Version series</td>
<td>An object of controlling single version series</td>
</tr>
<tr>
<td>Document version</td>
<td>A document object</td>
</tr>
<tr>
<td>Rendition</td>
<td>Managing document format</td>
</tr>
<tr>
<td>Content element</td>
<td>Managing element of document</td>
</tr>
</tbody>
</table>

Fig. 2—The Object Model of DMA.

This figure shows the object model of DMA. The model of DMA is composed of five layers.

Fig. 3—SGML can Describe Relation of Element by Itself.

![Diagram of DMA Object Model](image-url)
APPLICATION EXAMPLE

We redesigned the business process for developing medicines based upon the Enterprise Document Management System.

In a pharmaceutical company, it takes about ten years to develop a new medicine. The business process for developing medicines consists of a plan, a study, a clinical test, and an application to the government. When the pharmaceutical company applies to the government, it has to arrange the materials made in these processes efficiently. The pharmaceutical companies in United States already use electronic media. Japanese pharmaceutical companies and the Japanese government are planning to use electronic media as well.

The management system used for developing medicines has been studied in order to develop a business system which supports the government.

We store and manage the results of the each developing process systematically and steadily for the long term. The results from the developing processes include the revised history of individual documents, which is a component of the application, and the relationship between the application, the report, and the experimental data.

CONCLUSIONS

The rapid progress of the standardization of structured documents, especially XML, is a drastic change in the document management system for the next generation because it enables the document itself to be utilized widely.

We have been observing this change closely, proposed the specification concerning the structured documents to DMA and contributed to the standardization.

Furthermore, we are going to accumulate the know-how and the parts of the system construction in common to each company based on the case studies from an actual system.

REFERENCES