

## Database System for Data Mining Based on Meta Data

Hisashi Aomori, Hiroshi Ichijo, Tsuyoshi Otake and Mamoru Tanaka

Sophia University  
7-1 kioi-chiyoda, Tokyo, Japan  
Phone:+81-3-3238-3878, Fax:+81-3-3238-3321  
Email: aomori@mamoru.ee.sophia.ac.jp

**Abstract**— In this paper, we focus on database system for data mining. By using our proposed system, digitized data can be expressed by text like the MPEG-7, and search them by SQL query. The database system is implemented by PostgreSQL Server and Apache Web Server with Java Servlet, and we evaluated our approach.

### 1. Introduction

The progress of computer hardware and software allows today's computer system to make media easy to digitize, and huge amount of electronic media have been collected.

For data mining to be effective, it is important to include the humanity in the data exploration process and combine the flexibility, creativity, and general knowledge of the human with the computational power of computers. The basic idea of our work is to combine the flexible knowledge of human and digitized data[1].

Usually, electronic media is made of non text data, mining the valuable information hidden in them, however, the data mining is a difficult task. The technique of our work uses meta data of electronic media like the MPEG-7 can help to associate electronic media and meta data, and to combine the idea for the user. There are a large number of information which characterize electronic media, but essential information is little. We extract essential information of electronic media, and store them in the database as a meta data. The advantage of our system is that the user can search the electronic media with a simple query from the database. Moreover, we develop a simple web based application which is easy to use on our system.

The database system is implemented by PostgreSQL Server and Apache Web Server with Java Servlet, and we evaluated our approach. The web application searches the data by SQL query. We exemplify the advantage using a few examples, in short, our technique make most of the samples search easily.

### 2. Framework for associating electronic media and meta data

Meta data is a good method for describing digitized data, thereby helping access to them and to provide the essential information between the creator and the user[3]. For making our system flexible, we use Dublin Core's[2] scheme to associate electronic media and meta data.

Dublin Core Meta data Element Set (Dublin Core) is a meta data description scheme aimed to resource discovery in the Internet. Dublin Core defines 15 core elements to describe meta data of various information resources. It is gaining acceptance as a core meta data scheme for networked information resources. Dublin Core was proposed at the Meta data Workshop held at Dublin, Ohio in March 1995, and has been developed based on discussions at the series of workshops and on the mailing list. Because of Dublin Core flexibility, it is easy to associate to electronic media and meta data[4][5]. Table1 is Dublin Core's 15 core elements and their description.

### 3. Database system for Data Mining

Online analysis is an efficient way to access a database system. For using database effectively, our database system is used through the Internet. However, it might be difficult for online analysis to derive patterns like the decision tree from the stored data, thus our system is designed to support the user's decision. Fig.1 is overall process of our system. The process of the user's acquiring knowledge follows as below.

- The user mine electronic media from the database by the use of web application.
- The system returns the electronic media and relational resources.
- The user makes an evaluation for the returned data.
- If the user finds new relationship into the returned data, new knowledge is acquired.

Table 1: Dublin Core's 15 core elements

Element	Description
Creator	Person or organisation primarily responsible for creating the intellectual content of the resource.
Publisher	The entity responsible for making the resource available in its present form, such as a publishing house, a university department, or a corporate entity.
Contributor	Person or organisation not specified in a Creator element. e.g. editor, transcriber, illustrator.
Rights, Management	A rights management statement, an identifier that links to a rights management statement.
Title	The name given to the resource, usually by the creator or publisher.
Subject	The topic of the resource. Typically, will be expressed as keywords or phrases that describe the subject or content of the resource.
Date	A date associated with the creation or availability of the resource.
Identifier	A string or number used to uniquely identify the resource.
Description	A textual description of the content of the resource.
Source	The work, either print or electronic, from which this object is derived, if applicable.
Language	The language of the intellectual content of the resource.
Relation	Relationship to other resources.
Coverage	Spatial locations and temporal duration characteristic of the resource.
Type	The category of the resource.
Format	The data format of the resource, used to identify the software and possibly hardware that might be needed to display or operate the resource.

- If the returned data has less useful information, the user can refine the stored data.

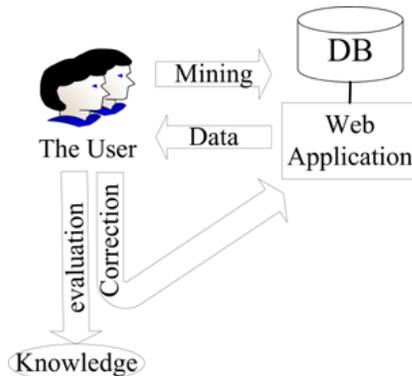


Fig. 1: overall process

### 3.1. Database system

We implemented our approach by the use of PostgreSQL Server and Apache Web Server with Java Servlet. Fig.2 is block diagram of our system. The user

can mine the data by using the web application. The web application searches the data by SQL query. Java application and web server are connected on the web server. Fig.3 is block diagram of web server.

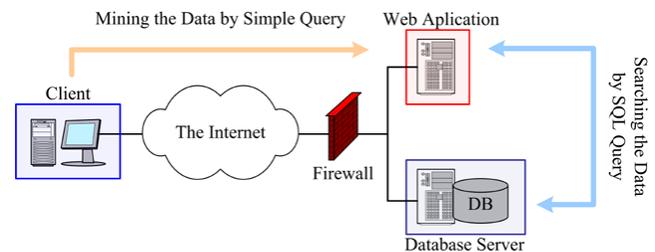


Fig. 2: block diagram of our system

### 3.2. Process of storing data

Storing data in database is very important task for our system. The overall process of this task follows as below.

- Extract information of electronic media

- Associate electronic media and meta data in accordance with Dublin Core
- Extract essential meta data in accordance with data type
- Put electronic media and relational resources on the database server's specific directory
- Store meta data and relational resources by using web application's form

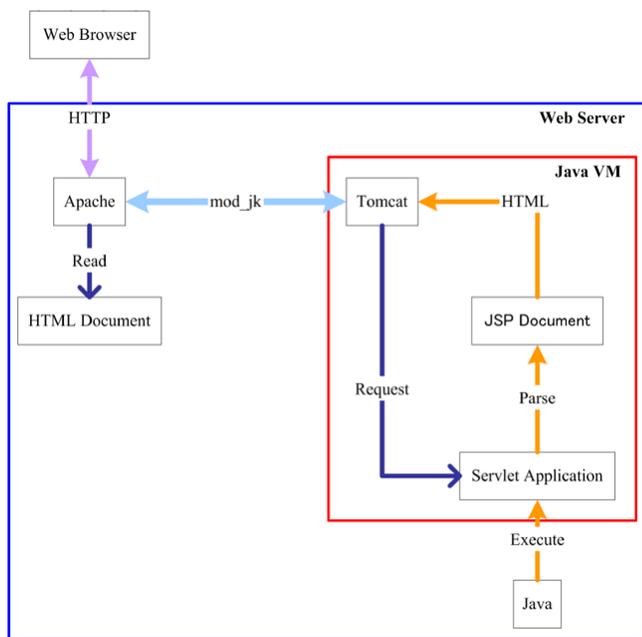


Fig. 3: block diagram in web server

## 4. Implementation

In our work, we focus on mining landscape system.

### 4.1. Meta data and database scheme

For mining landscape, we extract some information and associate meta data. To identify landscape, some information as follows are required, and we define database scheme.

- Title : title of landscape
- Date : date of creation
- Description : description of the place or object
- Relation : relational resources (sound data and text data )
- Format : format of image

## 4.2. Results

We store meta data and relational resources by using web application's form. Fig.4 is web application's form.

The screenshot shows a web form for registering new media. It includes fields for Title, Comment, Date (Year and Month), Region Information, Supplemental Information, File Location, File Type, and Audio Information. There are also buttons for '編集実行' (Execute Edit) and 'リセット' (Reset).

Fig. 4: Web application's form

To search landscape, the user selects some information from the web application's form. The result is returned in html format. The user selects the title of landscape, then, the landscape is identified. Fig.5 is form of searching for landscape.

The screenshot displays search results for 'メディアを検索' (Search Media). It shows a list of 200 items with columns for '地域名' (Region Name), '年号' (Year), and 'タイトル' (Title). The results include various hiking routes and scenic spots. There are also search filters for '地域名' and '年号'.

Fig. 5: Form of searching for landscape

If the user clicks the title, the user can view all of landscape's data. Fig.6 is all data of landscape.



Fig. 6: All data of landscape

The user can edit the stored data. Fig.7 is edit form for landscape's data.



Fig. 7: Edit form of the data of landscape

media, and we obtain an effective result. Our system might be able to apply for web museum, review of sickness, file mining system.

## References

- [1] Hiroshi ICHIJO, Norio FURUYA, Kazuo SHIMADA, Kenichi OGAWA, Masahiko IIZUKA, "A Multimedia System for Medical Information," Medical Imaging Technology Vol.13 No.5, pp.762-769, 1995.9
- [2] The web site of Dublin Core, <http://dublincore.org/>
- [3] The web site of The World Wide Web Consortium, <http://www.w3.org/Metadata/>
- [4] Schable P. and Smeaton, "An International Research Agenda for Digital Libraries", <http://www.iei.pi.cnr.it/DELOS/NSF/Brussrep.htm>, 1998.10
- [5] Lagoze C., "A Container Architecture for Diverse Sets of Metadata", D-Lib Magazine, <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>, 1996.7/8

## 5. Conclusions

In this paper, we propose the database system based on meta data for data mining. This system helps the user's decision. In order to estimate our system, we apply this system to mining landscape system. We show the advantage using a meta data for mining electronic