A Study on Language Design of Creating Compound Metadata Schama Method Based on Museum Information

Ryoji Akimoto
TOPPAN PRINTING CO., LTD.
Corporate R&D Division
1-3-3 Suido Bunkyo-ku Tokyo 112-8532
Japan
ryoji.akimoto@toppan.co.jp

Wataru Kameyama
Waseda University
Graduate School of Global Information and Telecommunication Studies
1011 Okuboyama Nishitomita Honjo-shi Saitama 367-0035
Japan
wataru@waseda.jp

Abstract: As information technology has progressed recently, the amount of contents that can be used increase. And, there is an enviroment in which lots of contents can be accessed at anytime and anywhere. In this situation, a technology is needed to extract useful information from various contents by using metadata in order to meet the requirements. In this paper, we summarize the current state of metadata technology and problems. Based on it, we propose the concept called Fuzzy Schema. Fuzzy Schema can flexibly construct metadata schema if necessary. And we also describe the design of Fuzzy Schema language taking an example of metadata schema of museum.

Key–Words: metadata, metadata schama, Fuzzy Schema, XML, content management, content distribution, museum information

1 Introduction

Recently, the information technology had progressed rapidly, and the amount of information that a user can handle increases rapidly, too. A lot of contents (Web pages, broadcasting, publication, music, image, and digital archives, etc.) are made by people not only for the personal use but companies or organizations. Especially, in the museum field, digitalization of cultural properties expressing of human behavior of ancient times is being progressed. Moreover, a ubiquitous information base is being maintained though which contents can be accessed by various devices at anytime and anywhere.

In this situation, a user demands to acquire contents from various resources and to use the acquired contents in compounding. In order to fulfil these, we need to utilize the data which describes the information about contents. This data is called "metadata". Until now, various types of metadata have been proposed considering the characteristic of contents. In particular, some international standards of metadata or industry standards of metadata have been proposed. However, these metadata schemas have some fundamental technical issues so that the these metadata schemas don't fulfil the practical requirements.

Therefore, in this paper, we clarify the technical issues of metadata, especially in the museum field, at first. Next, we propose the concept of "Fuzzy Schema" by which a compound metadata schema can be created for museum information in order to solve the issues. Furthermore, we discuss the language design for the Fuzzy Schema.

2 Present Situation and Issues of Metadata

2.1 Metadata and Metadata Schema

A content means some information expressed in digitalized data, e.g. text, picture, movie, music, and compound document that includes these in a complex way[1]. The followings are some of the concrete examples.

- Web contents in the Internet
- Movie contents, such as broadcast programs
- Image contents, such as paintings, landscapes in museum, etc.
- Publication contents, such as digital books, comics, etc.
- Music contents distributed through the Internet
- Digital archives of complex of contents

Metadata is defined as "(Structured) data about data"[2], i.e. it can be said that all description

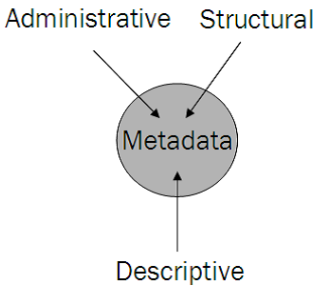

Figure 1: Categories of Metadata

about information resources is considered as metadata. Metadata can be classified into the following three categories from the point of information resources[3].

- Descriptive metadata
- Structural metadata
- Administrative metadata

The descriptive metadata describes metadata mainly for information retrieval, discovery, and profile of information resources. The structural metadata describes the form of information resources, or physical characteristic and its attribute. The administrative metadata describes information resources for managements, such as preservation of information resources particularly in museums, right management, and usage conditions.

Figure 1 shows this point of metadata.

In this paper, metadata means the data describes all the above three about a content.

Metadata schema defines metadata, which can be devided into following two types: the one defining an item name, scope, and data type, and the other defining the relation between items, and semantics.

Because use of metadata and metadata schema is increasing in recent years, international standards and industry standards have been being proposed.

An example is “General International Standard Archival Description (ISAD(G))[4]” which is a standard of archives description for managing and using various official documents. An another example is Dublin Core[5] which is a metadata schema for discovering information resouces on the Internet. Dublin Core defines 15 basic element types. Table 1 shows the Dublin Core 15 basic elements.

MPEG-7[6] is a metadata schema for multimedia which includes Dublin Core elements. MPEG-7 defines Description Definition Language(DDL) as an

Table 1: Dublin Core 15 Basic Element

Element	Definition
Title	A name given to a resource
Creator	An entity primarily responsible for making the content of a resource
Subject	A topic of the content of a resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of a resource
Date	A data of an event in the lifecycle of a resource
Type	The nature or genre of the content of a resource
Format	The physical or digital format of a resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to an another resource from which a resource is derived
Language	A language of the content of a resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of a resource
Rights	Information about rights held in and over a resource

XML subset for data description. MPEG-7 aims to be used by information retrieval of multimedia contents. Thus, MPEG-7 defines various information in a form of metadata in order to represent contents. As to be used in an information retrieval system, TV-anytime[7] releases a metadata schema to increase the efficiency of program retrieval used in Electronic Program Guide (EPG) or retrieval of contents distributed over the Internet.

In addition, there are various metadata schemas used as industrial standards, such as XMDF[8] and BBeB[9] for digital books using XML, and MusicBrainz[10] for music distributions using RDF.

Furthermore, there is Open Archives Initiative Protocol Metadata Harvesting (OAI-PMH) [11] which defines communication protocols for metadata exchange. OAI-PMH harvests and saves the electronic intellectual products produced in university or research organization. In Japan, it is used in a system called “cultural heritage on-line”[12].

2.2 Issues of Metadata

Until now, lots of contents are created, released and sold. These contents are described by some standardized metadata schema depending on each application field. However, the following issues are not yet solved

though these standard metadata schema are proposed for solving them.

a Issue of metdata schema construction
Metadata is defined by a metadata schema. Thus, a content with metadata is managed by the metadata schema. The metadata schema defines item names, scope, data type, and the relation between items and semantics. Therefore, a specialist is needed who knows well the field in order to construct the metadata schema. This work takes very long. Thus, a significant delay is generated before completing the schema being ready to be use by users. In addition, it becomes difficult to change the metadata schema, if it starts to be used.

b Issue of metadata creation
In order to increase the usage of contents, a lot of metadata need to be distributed. In MPEG-7, some automatic metadata extraction from audio signal and video signal are specified. However, almost all metadata (description of attribute, data for maintenance management, etc.) are described and generated manually. Creating appropriate metadata requires time and effort. In addition, we also have to consider that increasing metadata contributes the increasing usage of contents.

c Issue of metadata mapping
In order to acquire information from several different databases or archives used in different applications, matching items of different metadata schemas is essentially required. Until now, matching items is performed by manual operations called ontology mapping [13] and metadata crosswalk[14]. And making wrappers for different metadata schemas has also been performed. However, the mismatching items by disagreement of specialists offen happens which makes the mapping frequently changed. Therefore, it is very difficult to keep the mapping consistency.

2.3 Metadata Situation in Museum

Museums deal with and keep various cultural properties broadly, e.g. history, arts, folk customs, industries, natural science, etc. Thus, museums have arranged and maintained collection information (i.e. museum metadata) for long time.

We have investigated about 20 collection management systems in museums in the world in order to figure out how museum metadata is treated, as shown in Table 2.

Every collection management system manages the following common two types metadata.

Table 2: Collection Management System

Item	Explanation
Collection info	Work name, author name, work year, size, etc.
Person info	Registrar, conservator, researcher, etc.
Media info	Tape, CD-ROM, DVD, etc.
Insurance info	Evaluation, insurance, etc.
Books info	Catalogue, publishing, bibliography, etc.
Ground info	Real estate, etc.
Accounting info	Accounting, accounting system, etc.
Loan info	Loan management for exhibition, etc.
Exhibition info	Exhibition load map, etc.
Event info	Related event, etc.

a Metadata about a collection itself

b Log metadata as generated with the operation of a collection

In addition to this, The thesaurus and dictionary which are linked to the system can be regarded as metadata.

After 1990s, an efficient operation of a museum has been highly requested, i.e. the information sharing between museums has been needed. As a result, some standard museum metadata has been proposed.

In the following, some of the metadata schemas used in museums are summarized.

I Structured Model for Museum Object Information

This is the model proposed by the Tokyo National Museum in Japan in November, 2005. The target of this model is tangible cultural properties, and to support the operation of the collection management in a museum. In addition, it provides information sharing between museums.

This model classifies objects to be mananged into four characteristics: “Identification” which identifies object uniquery, “Physical characteristic” which describes such as object form and/or size, “history” which describes such as object loan information and object restoration history, “relation / reference” which describes the literature and pertinent information about an object. This four characteristics consist of 34 attributes and 108 elements. In addition, This model defines the vocabulary for RDF Schema to describe museum metadata.

II ICOM/CIDOC

International Committee for Documentation (CIDOC) of International Council of Museums (ICOM) proposed "International Guidelines for Museum Object Information: The CIDOC Information Categories" in 1995. This is a guideline for describing museum metadata. The management process from entry to disposal is classified into 22 information groups and 77 information categories. In addition, in 1996, CIDOC Conceptural Reference Model (CRM) was developed, which is a framework for metadata exchange described by RDF/XML. Now, this model is available as ISO21127:2006.

III CDWA

Categories for Description of Works of Art (CDWA) is metadata schema which is specialized in arts. CDWA was specified as a collaborative project of Art Information Task Force and College Art Association, in 1996. Metadata is defined by XML Schema.

IV MDA SPECTRUM

In UK, there is standard procedure on which about 75% of museum is based. It is called "SPECTRUM" which describes 20 procedures from entry to disposal. In Museum Documentation Association (MDA), not only the procedure but the vocabulary of metadata is proposed.

V Other Standard Museum Metadata

Table 3 shows some other standards of museum metadata.

Table 3: Metadata Schemas for Museum

kind	note
AFRICOM Handbook of Standards[15]	The standard manual of metadata description which about 120 museum in Africa created together for a theft, pillage, and smuggling prevention, in 1996.
AMICO Data Directory[16]	Data specification by Art Museum Image Consortium, developed in 2001.
CIMI profile	Standing for Computer Interchange of Museum Information Profile based on Z39.50, developed in 1995.
DC.Culture[17]	Developed by the European MINERVA project, based on Dublin Core, aiming for digitalization and discernment management of data.
Museum Handbook[18]	Published by the U.S. national park service about general handling data, such as management of museum data, preservation, recording, and access, developed in 1990.
Object ID[19]	The metadata schema for theft prevention in which information required for tracing is described, developed in 1997.
RLG Cultural Materials[20]	Jointly developed by about 40 American organizations with the total of 250,000 metadata described by Encoded Archival Description (EAD) and Metadata Object Description Schema (MODS) by that cross search can be performed between metadata.
VRA[21]	Virtual Resources Association Core Categories for Visual Resources, developed in 2000.
National standard cultural property of south korea[22]	Providing 16 mandatory items and 116 optional items used by about 90 South Korea organizations, developed in 1995.

2.4 Issues of Metadata in Museum

There are issues in the above-mentioned metadata schemas in museums. From the point that information sharing between museums is required and important, the issue of metadata mapping is important[23] [24].

3 Fuzzy Schema

3.1 Proposal of Fuzzy Schema

We have investigated the issues and the related trend of metadata and metadata schema. The issues can be summerized into the three points: a) issue of metadata schema construction, b) issue of metadata creation, and c) issue of metadata mapping. In order to solve these issues, we propose the "Fuzzy Schema" as follows.

When a person uses some contents, a matadata schema is created dynamically against the content in his/her brain, according to the situation, the environment, and the his/her experience. Moreover, the schema may be adjusted flexibly from time to time suitable for different metadata and contents if needed. In contrast, the existing contents management methods manage metadata based on the fixed metadata schemas specified before, and changing or reorganization of items are not usually possible. The schemas are not dynamic and not flexible compared to the way of a person using contents by generating dynamic and flexible schemas. Threfore, in this paper, in order to map different metadata flexibly like the way of a person, the concept "Fuzzy Schema" to create flexible metadata schema is proposed.

First, "Fuzzy Schema" enables item-mapping be-

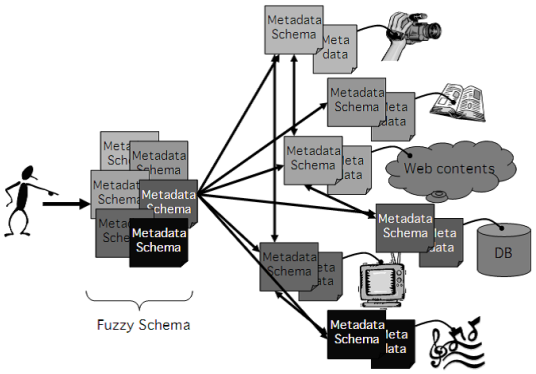

Figure 2: Concept of Fuzzy Schema

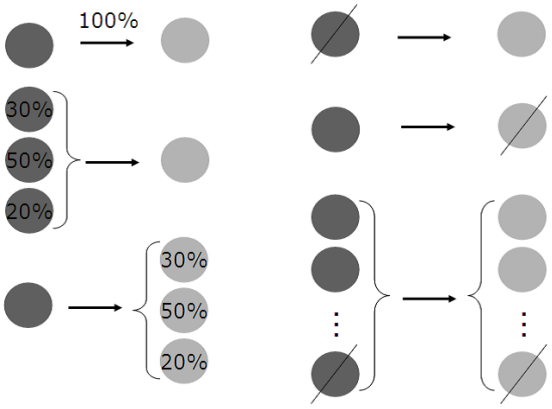

Figure 3: Mapping Pattern

tween different matadata and metadata schemas. The mapping can be categorized into “mapping patterns” that can be found in the comparison between different museum metadata and schemas. Second, it allows the ambiguous mapping between items by defining “the degree of ambiguity” which is a measure showing the similarity of different items of different schemas. The more accurate metadata mapping may be achieved by adjusting “the degree of ambiguity”.

By using these two mechanisms, a dynamic schema mapping of different schemas would be possible. Consequently, it is expected that reducing the burden of creating metadata schema and converting metadata from a schema to another. Figure 2 shows this concept of Fuzzy Schema.

3.2 Language Design of Fuzzy Schema

In order to design Fuzzy Schema, we are examining a possible form of language using XML by which the mapping pattern and the degree of ambiguity can be described. We are currently under the process to complete the final design of the Fuzzy Schema language.

The mapping pattern can be regarded as a mathematical projection. In our case, it is the pattern of item relationship in metadata defined by metadata schema. As mentioned above, the mapping has been performed by some techniques and/or specialists in paractical applications. However, only the surjective mapping and the injective mapping have been mostly done. In addition, the disagreement of mapping between items is often raised by the manual mapping. Fuzzy Schema can provide a solution to this situation. Figure 3 shows some mapping patterns using Fuzzy Schema representation.

4 Conclusion

In this paper, the issues of metadata in museums are clarified, that are a) the issue of metadata schema construction, b) the issue of metadata creation, and c) the issue of metadata mapping. In order to solve these issues, we propose the concept of ”Fuzzy Schema” by which a flexible metadata schema can be created. In addition, the idea of mapping patterns and the degree of ambiguity is introduced.

For the future study, the concrete description of the mapping patterns and an algorithm of automatically generating “the degree of ambiguity” will be done as well as the completion of the Fuzzy Schema language design.

References:

[1] Ministry of Economy, Trade and Industry of Japan: *“Digital Content White Paper 2005”*, Digital Content Association of Japan(2005)(in Japanese).

[2] Dempsy *et al.: “Metadata: A Current View of Practice and Issues”*, J. of Documentation, Vol. 54, No. 2, pp. 145-172(Mar. 1998).

[3] Japan Society of Library and Information Science: “Library catalogue and Metadata”, Bensey Publishing(2004-10)(in Japanese).

[4] ISAD(G):General International Standard Archival Description(cited 2006-06-28) URL:http://www.ica.org/biblio/isad_g_2e.pdf

[5] Dublin Core Metadata Initiative(cited 2006-09-28) URL:http://dublincore.org/

[6] MPEG-7(cited 2006-09-28)(in Japanese)
URL:http://www.itscj.ipsj.or.jp/mpeg7/

[7] TV-Anytime(cited 2006-09-28)
URL:http://tv-anytime.org/

[8] Kitamura *et al. : "Trend in Electronic Publishing and XMDF Technology"*, SHARP Technical Journal, Vol. 84(2002-12)(in Japanese).

[9] Broad Band e-Book(cited 2006-09-28)(in Japanese)
URL:http://ps.canon-sol.jp/bc/kikaku.htm/

[10] MusicBrainz(cited 2006-09-28)
URL:http://musicbrainz.org/

[11] The Open Archives Initiative Protocol for Metadata Harvesting(cited 2006-09-28)(in Japanese)
URL:http://www.openarchives.org/

[12] Cultural Heritage Online(cited 2006-09-28)(in Japanese)
URL:http://bunka.nii.ac.jp/

[13] Cheng Zhu *et al.: "Ontology Mapping For Interaction in Agent Society"*, Proc. of IEEE Int'l Conf. on SCC'04, pp. 619-622(15-18. Sept. 2004).

[14] Diane I. Hillmann *et al.: "Metadata in Practice"*, American Library Association(2004)

[15] International Council of African Museums(cited 2006-09-28)
URL:http://www.africom.museum/

[16] Art Museum Image Consortium(cited 2006-09-28)
URL:http://www.amico.org/

[17] MINERVA DC.Culture(cited 2006-09-28)
URL:http://www.minervaeurope.org/DC.Culture.htm

[18] Museum Handbook(cited 2006-09-28)
URL:http://www.cr.nps.gov/museum/publications/handbook.html

[19] Object ID(cited 2006-09-28)
URL:http://www.object-id.com/

[20] Research Libraries Group(cited 2006-09-28)
URL:http://www.rlg.org/

[21] Visual Resources Association(cited 2006-09-28)
URL:http://www.vraweb.org/index.html

[22] Takubo: *"On Digitization and Networking of Cultural Property Information in South Korea"*, The Bulletin of JADS. No. 13, pp. 10-17(2006-03)(in Japanese).

[23] Hara *et al.: "Case Study on the Data-Mapping for the Resource Sharing System"*, J. IPS. Japan. Vol. 2005, No. 76, pp. 31-38(2005-07)(in Japanese).

[24] Yamada *et al.: "A Study on distributed ontology in museum information crossing search"*, J. IIEE. Japan.(2006-03)(in Japanese).