

博物館情報を用いた複合的メタデータスキーマ作成に関する一考察

秋元 良仁 亀山 渉

早稲田大学大学院国際情報通信研究科 〒367-0035 埼玉県本庄市西富田大久保山1101

E-mail: ryoji@fuji.waseda.jp, wataru@waseda.jp

あらまし 近年の情報技術の進展に伴い、利用できるコンテンツの量は飛躍的に増大している。また、これら膨大なコンテンツにいつでもどこでもアクセスできるユビキタスな環境も整いつつある。このような状況では、メタデータを用いて多種多様なコンテンツから利用者の要求に応じて意味のある情報を横断的に紡ぎ出す技術が求められる。そこで、本稿ではまずメタデータの現状と問題点を整理する。それをもとに、必要に応じて柔軟にメタデータスキーマを構築できる「ファジー・スキーマ」という概念を提案する。更に、複数の博物館の資料情報管理用メタデータスキーマを例に、ファジー・スキーマの言語設計の指針を示す。

キーワード メタデータ、メタデータスキーマ、ファジー・スキーマ、コンテンツ管理/コンテンツ流通、博物館情報

A Study on Creating Compound Metadata Schema Based on Museum Information

Ryoji AKIMOTO and Wataru KAMEYAMA

Graduate School of Global Information and Telecommunication Studies, Waseda University 1011

Okuboyama, Nishitomida, Honjo-shi, Saitama 367-0035 Japan

E-mail: ryoji@fuji.waseda.jp, wataru@waseda.jp

Abstract Recently, information technology has progressed. Then, the amount of contents that can be used increases. And, there is an environment that can be accessed huge contents anytime and anywhere. In such a situation, it is necessary that the technology extract useful information according to requirements from various contents by using metadata. In this paper, we summarize the current state of metadata and problem. Based on it, we propose the concept that is called Fuzzy schema. Fuzzy schema can flexibly construct metadata schema if necessary. And then, we describe the design of Fuzzy schema language as an example of metadata schema of museum.

Key words Metadata, Metadata Schema, Fuzzy Schema, Contents Management/Contents Distribution, Museum Information

1. はじめに

近年、情報技術の活用が急速に進展し、利用者が取り扱うことのできる情報量は飛躍的に増大している。100億ページを超えると言われるWebコンテンツを始め、放送・出版・音楽・映像等の多種多様なコンテンツ、これらの複合体としてのデジタル・アーカイブ等、個人のみならず、企業や関係機関が編集を行い提供されるコンテンツが多数存在する。また、近年ではこれらのコンテンツに様々なデバイスを用いていつでもどこでもアクセスできる、ユビキタスな情報環境基盤が整いつつある。

このような状況ではメタデータを横断的・複合的に用いて多種多様なコンテンツから利用者の要求に応じて意味のある情報を紡ぎ出す技術が求められる。しかし、メタデータはコンテン

ツによってあらかじめ定められた体系(メタデータスキーマ)を持ち、異なる体系間で意味的なマッピングを行うことは非常に困難な状況にある。

そこで、本研究ではまずメタデータの現状と問題点を整理する。それをもとに、必要に応じて柔軟にメタデータスキーマを構築できる「ファジー・スキーマ」という概念を提案する。更に、複数の博物館の資料情報管理用メタデータスキーマを例に、ファジー・スキーマの言語設計の指針を示す。

以下、2節でメタデータの現状と問題点について概観する。特に後半では博物館分野におけるメタデータの現状と問題点を整理する。次に3節でファジー・スキーマの検討を行い、言語設計の指針を考察する。最後に4節でまとめと今後の課題を述べる。

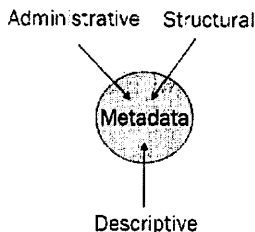


図1 メタデータ
Fig.1 metadata

2. メタデータの現状と問題点

2.1 メタデータとメタデータスキーマ

コンテンツとは、デジタルデータで表現された文章や画像、映像、音楽、これらを複合的に組み合わせた情報のことを言う[1]。コンテンツは多種多様に存在する。例えば、インターネット上に多数存在する Web コンテンツ、デジタル放送やワンセグに代表される映像コンテンツ、美術館・博物館の絵画や収蔵品、様々なキャラクターや風景等の静止画コンテンツ、電子書籍やコミック等の出版コンテンツ、Web を通じて配信される音楽コンテンツ、更にこれらのコンテンツの複合体としてのデジタル・アーカイブ等がある。

メタデータとは、「データに関する(構造化された)データ、(Structured) data about data」と定義される[2]。即ち、何らかの情報資源に関する記述は全てメタデータであると言うことができる。メタデータの記述は、情報資源をどのように表現するかによって、Descriptive メタデータ、Structural メタデータ、Administrative メタデータの3種に分類することができる[3](図1参照)。

Descriptive メタデータは主として情報検索等、情報資源の発見や特定するために記述されるメタデータである。Structural メタデータは、情報資源の形状や物理的な特性等、情報資源の属性に関する内容記述であり、Administrative メタデータは情報資源の保存、権利関係や利用条件等、情報資源管理に関する記述がなされる。

本稿では、これら3種概念を基本とし、特にコンテンツに関する情報を記述したデータのことをメタデータと言う。

メタデータは定義が広範であり、用途目的も様々であることから、それぞれのコンテンツの特性に合わせた形でメタデータの記述項目と記述形式が提案されてきた。本稿では記述項目と記述形式を併せてメタデータスキーマと呼ぶ。

メタデータスキーマの活用は、近年その意識が高まり、幾つかの国際標準や業界標準が制定されている。

例えば、アーカイブズ記述の国際標準に General International Standard Archival Description (ISAD(G): 国際標準記録資料記述)[4]がある。これは様々な公式文書や資料を管理・活用するためのメタデータスキーマである。また、図書館の分野

表1 Dublin Core の基本 15 エレメントセット

エレメント名	定義
title(タイトル)	情報資源に与えられた名称
creator(作成者)	情報資源の内容の作成に主たる責任を持つ主体
subject(キーワード)	情報資源の内容のトピック
description(内容記述)	情報資源の内容の説明・記述
publisher(公開者)	情報資源を公開することに対して責任を持つ主体
contributor(寄与者)	情報資源の内容に寄与、貢献した主体
date(日付)	情報資源のライフサイクルにおける事象の日付
type(資源タイプ)	情報資源の内容の性質またはジャンル
format(記録形式)	情報資源の物理的形態ないしデジタル形態での表現形式
identifier(資源識別子)	情報資源を一意的に特定する識別子
source(出処)	情報資源が作り出される源の情報資源への参照
language(言語)	情報資源の内容を表す言語
relation(関係)	関連情報資源への参照
coverage(時空間範囲)	情報資源の内容が及ぶ範囲
rights(権利管理)	情報資源に含まれる(また関わる)権利情報

では、インターネット上の情報資源発見のためのメタデータスキーマに Dublin Core [5]がある。Dublin Core はインターネット上におけるリソースに関する情報を記述するための15の基本要素タイプを中心とするメタデータスキーマである(表1参照)。Dublin Core をベースとした映像音声等のマルチメディア記述用のメタデータスキーマに MPEG-7 [6]がある。MPEG-7 では、データ記述言語として XML のサブセットを Description Definition Language(DDL)として定義し、マルチメディアコンテンツの検索を目的に、コンテンツを表現する様々な情報をメタデータで定義している。また、TV-Anytime [7]では、Electronic Program Guide(EPG)を用いた放送番組検索のみならず、インターネットで配信されるコンテンツも含めて効率良く検索できることを目指してメタデータの標準化が行われてきた。

この他にも、例えば電子書籍で利用される XML 形式の XMDf [8]や BBeB [9]、音楽配信用に用いられる RDF 形式の MusicBrainz [10]等、業界で標準として利用される種々のメタデータスキーマが存在する。

また、これらのメタデータスキーマを相互に活用するためのメタデータ交換用の通信プロトコルとして、Open Archives Initiative Protocol Metadata Harvesting(OAI-PMH) [11]がある。OAI-PMH は大学や研究機関等で生産された電子知的生産物を収穫・保存し、原則無償でインターネット上に公開する機能を持つ。近年では文化遺産オンラインに応用される等、幅広い活用が見込まれている。

2.2 メタデータの問題点

これまで、コンテンツは Web やシステムを通じて公開、あるいは販売されることに主眼が置かれてきた。各コンテンツを表現するメタデータは各分野で標準化されたメタデータスキーマに基づいている。しかしながら、これらの標準化されたメタデータを利用したとしても、以下に述べるようなメタデータに関する根本的な技術的課題は解決できない。

a) メタデータスキーマ構築の問題

メタデータはあらかじめ定義されたメタデータスキーマを持ち、コンテンツはメタデータスキーマに基づいて運用される。メタデータスキーマはメタデータの基礎となるものであり、何をどのように管理すべきか、その項目名や適用範囲、あるいはデータ型等が厳密に定義される。そのため、通常メタデータスキーマの構築は、その分野に精通した専門家が行う。仮にメタデータスキーマがデジュール標準であった場合、策定プロセスや標準内容はオープンであり、かつ専門家による高度な知識が活用されているため、その内容は信頼性の高いものになる。しかしその反面、コンセンサスによる意思決定は相当な時間を要するため、標準の普及とソフトウェア等プロダクツの普及の間にタイムラグが生じやすい。また、一度メタデータスキーマが構築され、普及すると、新たな問題や追加事項に対して柔軟にメタデータスキーマを変更し対応することが困難となる。

b) メタデータ作成の問題

メタデータの記述を考えてみると、コンテンツの属性記述や維持・管理用の情報等、テキスト部分のメタデータ作成は、例えば MPEG-7 におけるビジュアル信号やオーディオ信号から自動的に取得できる特徴量と異なり、人手により記述を行っている現状がある。前述したように、現在既に多種多様かつ膨大な量のコンテンツが存在しており、これらのコンテンツに対して適切なメタデータを作成、付与する作業は人的・金銭的・時間的に見て非常に負荷の高い作業と言える。

c) メタデータ・マッピングの問題

多様なコンテンツから意味のある情報を紡ぎ出すためには、複数の異なるメタデータスキーマ間を横断する形でコンテンツが活用されなければならない。しかし、メタデータスキーマは分野や応用ごとに固定的に形作られているため、相互のメタデータスキーマにおいて知識のマッピングを行う必要がある。これは議論を要する問題であり、異論のない整合の取れたマッピングを実現することは非常に困難である。

これまでにも、オントロジー・マッピング [12] やメタデータ・クロスワーク [13] といった、メタデータの各項目を対応させて異なるメタデータスキーマ間のメタデータの相互利用を行おうとする研究は多く存在する。これらの多くでは、専門家同士がその対応関係を議論し、1項目ずつデータをマッピングしていく手法が採られている。しかし、専門家同士の議論が平行線に進む場合や、専門家であってもどこにマッピングすれば良いのか不明瞭な場合、議論に第三者の視点に加わることでマッピングの対応を変化せざるを得ない場合等、1対1のメタデータ・マッピングが困難な状況がよく発生する。このような場合に強引なマッピングを行うことも想定され、また実際に行われているが、そのような行為は双方のメタデータスキーマから必要な情報の欠落や無駄な情報の増加を引き起こす可能性がある。

2.3 博物館分野におけるメタデータの現状

博物館は歴史・芸術・民族・産業・自然科学等、人類の営みを表現した有形・無形の文化財を幅広く取り扱う。そのため、古くから台帳あるいは目録という形で文化財に関する資料情報(以下、博物館メタデータ)が整理されてきた。

表 2 一般的な収蔵品管理システムが取り扱う情報

Table 2 Information of collection management system

情報名称	内容
収蔵品情報	作品名、作者名、制作年、サイズ等
人物情報	登録者、修復者、研究者等関係者情報
メディア情報	テープや CD-ROM、DVD 等の媒体管理
保険情報	評価・保険情報
図書情報	カタログ、出版物、参考文献管理情報
用地情報	不動産文化財情報
会計情報	会計システムと連携した会計情報
貸出・借入情報	展示会等による収蔵品の貸借管理情報
展示会情報	展示会行程情報
イベント情報	関連展示会情報

筆者らは、博物館メタデータの利用がどのように行われているのか、博物館で利用されている国内外約 20 の収蔵品管理用のシステムについて調査を行った [14]。その結果、表 2 に示す情報が一般的な収蔵品管理システムが管理する情報であることがわかった。

一般的な収蔵品管理システムが取り扱う博物館メタデータは、収蔵品そのものに関するメタデータと収蔵品を操作する際に生じる履歴メタデータに大別することができる。

具体的には、前者は作品/作者名、物理的なサイズ、材質や数量、テープや CD-ROM 等の記録媒体情報、収蔵品評価情報といった現物に関するメタデータがこれに当たる。また、後者は博物館が業務を遂行する上で基盤となり、かつ時系列に並べることができるメタデータ(例えば、ある収蔵品の受け入れから破棄までの履歴情報や、展示会情報、展示会等による収蔵品の貸借情報等)と言える。

また、これらの博物館メタデータとリンクする形である特定の分野(美術・民族・工芸・建築・考古学等)に特化したソーラスや辞書を有するシステムもあり、これらも含めて博物館メタデータと言うことができる。

近年、博物館メタデータの記述方式を整理・モデル化(メタデータスキーマ化)し、システム化する動きがある。これまで、各博物館はそれぞれ独自の記述項目・記述形式で博物館メタデータを蓄積し、システム化してきた。しかし、より効率的な博物館の業務遂行や博物館間での情報共有の必要性から、標準的な博物館メタデータの利用が求められている。

以下、具体的に検討されている博物館分野におけるメタデータスキーマの動向を整理する。

i) ミュージアム資料情報構造化モデル

国内における標準化の動向として、2005 年 11 月、東京国立博物館を中心とした博物館情報処理に関する調査研究プロジェクトチームより「ミュージアム資料情報構造化モデル」[15]が公表された。このモデルは人文科学系博物館の、移動可能な有形文化財を対象としている。対象の各種メタデータを整理・モデル化し、メタデータスキーマを構築することで効率的な博物館収蔵品に関する業務支援と博物館同士の情報共有の基盤とすることを目指している。

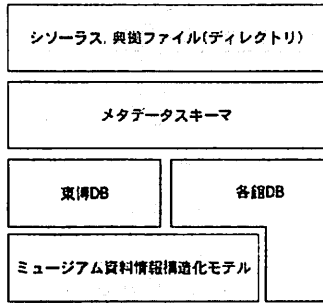


図2 ミュージアム資料情報構造化モデルにおける情報共有
Fig.2 Information Sharing of Structured Model for Museum Object Information

このモデルは、対象を一意的に識別する「特定・特定」、対象そのものを記述する「物理的特性」、対象の貸借情報や修復情報を記述する「履歴」、対象に関する様々な文献・関連情報を記述する「関連・参照」の4特性に分け、更にそれを34属性、108の詳細要素に細分化し構成される(表3参照)。

このモデルは後述する世界各国で提案されているメタデータスキーマを参考に、日本の博物館分野における実状を考慮した抽象度の高い概念的なモデルとなっている。情報共有を行うためには、このモデルに基づいた各博物館で利用されるデータベースがあり、その上位に RDF Schema [16] によって記述されるメタデータスキーマが構築される。更にその上位にはシソーラス・典拠ファイルといったレイヤがあり、RDF Schema とこれらシソーラスを活用することで博物館間の情報共有がなされることが想定されている(図2)。なお、現在博物館情報処理に関する調査研究プロジェクトチームでは、RDF Schema によるメタデータスキーマの策定中である。

ii) ICOM/CIDOC

海外においても、博物館のためのメタデータスキーマはいくつか開発・提案されている。

International Council of Museums(ICOM:国際博物館会議)の International Committee for Documentation(CIDOC:国際ドキュメンテーション委員会)では、1995年に「博物館資料情報のための国際ガイドライン CIDOC 情報カテゴリ」[17]を発表している。このガイドラインでは、メタデータを記述するためのガイドラインとして、博物館における資料の受入から管理、廃棄に至るまでの管理プロセスを22の情報グループと74の情報カテゴリに分類して示している。CIDOCではガイドライン発表の翌年以降、ガイドラインを盛り込んだ形のモデルとして、「CIDOC Conceptual Reference Model(CRM:概念参照モデル)」[18]の開発・提案を行っている。このモデルは各博物館が持つメタデータの交換・統合を行うためのフレームワークであり、RDF/XML形式で記述ができる。なお、現在このモデルはISO化(ISO21127:2006)されている。

iii) MDA SPECTRUM

英国では、Museum Documentation Association(MDA:博

表3 ミュージアム資料情報構造化モデル

Table 3 Structured Model for Museum Object Information

性格	属性名	役割
識別・特定	識別子	記述単位を一意的に識別する記号、番号。
	資料番号	組織によって資料に付された記号、番号。
	名称	資料の名称、呼称、タイトル。
	分類	資料の分野、種別。
	用途	貸借・考古資料などで資料が本来持っていた機能。
	様式	資料が作られているスタイル、流儀。
物理的特性	品型形状	「材質」「技法」「形状」をまとめて記述する。この3つをそれぞれ記述する場合は省略。
	材質	資料を構成する材料、材質。
	技法	制作に用いられている技法。
	形状	資料の形状の型型。
	員数	資料の数量、点数。
	計測値	数値で表現できる計測値。寸法や重量。
	部分	資料の部分。下位の記述単位への参照。
	保存状態	資料の保存状態。
	付属品	資料に付属する物品。付属文書や箱。
	印章・銘記	資料に直接書き込まれた文字や印。
	履歴	制作
出土・発見		資料の出土、発見に関する情報。
来歴		資料の伝来、所有、使用の歴史。
取得		購入、寄贈などにより資料が管理下に置かれることになった際の記録。
毀損・処分		修復、売却、破損、盗難などにより資料が管理下に置かれなくなった際の記録。
受入		寄託、借入などにより資料を受け入れた際の記録。
調査		資料の調査履歴。
修復		資料の修復履歴。
展示		資料を公開した際の記録。
所在		資料が保管されている場所、収蔵庫、貸出先などを含む。
価格評価		資料に対する価格評価の履歴
受賞・相定	資料が受けた賞の履歴や文化財指定の履歴。	
関連・参照	権利	所有権、著作権、複製権など権利についての記述。
	関連資料	他の資料への参照。関連する記述単位への参照も含む。
	文献	関連する文書、刊行された図書、論文等への参照。
	画像	写真などの視覚的二次資料。
	記述ノート	その他の情報についての文章による記述。
記述作成	記述の作成者、変更履歴など。	

物館ドキュメンテーション協会)が国内の博物館メタデータの標準化を推進している。MDAは1994年、資料の受入から廃棄に至るまでの具体的な資料管理手順20項目を標準化し、SPECTRUM [19]として発行(2006年現在の最新版ではVersion3.0)している。MDAでは、手続きのみならずメタデータの語彙やシソーラス等も含めたMDAデータ標準を提案している。

英国ではSPECTRUMを用いて収蔵品登録を行わなければ博物館登録が行えず、国や自治体から補助金等の交付を受けることができなくなる。この制度とSPECTRUMによって、英国内75%の博物館がメタデータ作成を進めている[3]。

iv) CDWA

特に美術分野におけるメタデータスキーマとして、Art Information Task Force (Getty Research Instituteの美術史情

表 4 博物館メタデータスキーマの種類
Table 4 Metadata Schema for Museum

種類	注記
手続上の標準	
Museum Handbook	米国立公園管理局は 1990 年、博物館資料の管理、保存、記録化、アクセス等全般的な取り扱いに関するハンドブックを発行している。
AFRICOM Handbook of Standards	盗難、略奪、密輸防止のため、アフリカの約 120 の博物館が共同で作成したメタデータ記述の標準マニュアル、ICOM/CIDOC が指導協力、1996 年発行。
データ構造上の標準	
RLG Cultural Materials [23]	米国の約 40 の参加機関 (大学や博物館等) の統計 25 万点に上るメタデータを Encoded Archival Description (EAD) や Metadata Object Description Schema (MODS) の XML 形式で記述することで検索検索が可能なシステムを開発している。
DC:Culture [22]	資料の電子化・識別管理のため、欧州の MINERVA プロジェクトで開発、Dublin Core をベースとしている。
博物館遺物管理電算化のための遺物分類標準 [24]	Dublin Core をベースとした基本 16 項目、オプション 116 項目を要したメタデータスキーマであり、韓国内の博物館や大学等約 90 の参加機関で共通に利用されている。1995 年発行。
Object ID	博物館資料の追跡調査 (盗難防止のため) を行えるよう、メタデータとして資料 (作品、盗難品) を特定する最低限の情報を記述する。1997 年発行。
CIMI profile	Computer Interchange of Museum Information Profile は Z39.50 プロトコルをベースに博物館メタデータを定義している。1995 年発行。
芸術/視覚情報記述の標準	
VRA	Virtual Resources Association Core Categories for Visual Resources. 美術作品記述のためのカテゴリ、2000 年発行
AMICO Data Directory	Art Museum Image Consortium による美術館画像のためのデータ仕様、2001 年発行。

報プログラムと College Art Association の共同プロジェクト) が 1996 年に発行した Categories for Description of Works of Art (CDWA) [20] がある。Task Force では現在、CDWA を XML Schema で表現するための CDWA Lite [21] が開発されている。

v) その他

博物館用のメタデータスキーマは 1990 年代以降、様々な提案がなされている。以下、表 4 に博物館メタデータスキーマの種類を示す。

2.4 博物館分野におけるメタデータの問題点と関連研究

博物館分野においても、メタデータスキーマ構築の問題、メタデータ作成の問題、メタデータ・マッピングの問題が存在する。特に、博物館間における情報共有の必要性の高まりから、メタデータ・マッピングに関する話題は議論を呼んでいる。

国内におけるメタデータ・マッピングに関する先行研究として、Dublin Core と Z39.50 プロトコルに基づく人文科学系データベースの統合検索に関する報告 [25] がある。この報告は人文

科学系研究機関が保有する、異なる設計のデータベース間の接続・検索を目的としている。各機関のデータベースのレコード・フィールドを Dublin Core メタデータにマッピングし、情報検索用プロトコルの国際標準である Z39.50 を使い、サーバを介して各データベース間の情報検索を行う。当初はデータベース接続が中心の課題であったが、接続確認後はメタデータ・マッピングが課題の中心となっている。この研究では事例研究として、専門知識を有する被験者に対して、国文学研究資料館の国文学論文目録データを Dublin Core にマッピングしてもらう実験を行っている。この実験の結果として、被験者のリソースの捉え方によりマッピングの結果に大きな差異が生じることが明らかになっており、中間的なメタデータを介することが提案されている。

上記先行研究に関連して、国立歴史民俗博物館では、中間的なメタデータとしてミュージアム資料情報構造化モデルを用いたマッピングの実験を行っている [26]。実験では、国立歴史民俗博物館が有する 6 つのデータベースのレコード・フィールドにミュージアム資料情報構造化モデルの属性をマッピングしている。結果として、物理的特性を表現する項目に対しては不確実性が低かったものの、制作や出土・発見といった一部属性は拡張が必要であるとしている。

あらかじめシステム内部に標準的なメタデータスキーマを用意し、各博物館のメタデータスキーマとのマッピングを現場の学芸員に行ってもらうことで、結果として各博物館間のメタデータスキーマの差異を吸収しようとする試みもある。凸版印刷では、ミュージアム資料情報構造化モデルをベースとした ASP タイプの収蔵品管理システムを提案している [27]。システムはフレームワークとして標準的なメタデータスキーマを提供する。利用者 (各博物館学芸員) はこのフレームワークを参考に項目名の変更等のラベル付けが行えるため、自館の運用を維持したまま、ASP システムを介することで他博物館とのデータ共有が可能となる。

メタデータ・マッピングの話題として、専門家による 1 対 1 の項目マッピング以外に、シソーラスやオントロジを用いた知識処理によりメタデータ・マッピングを行おうとする試みもある。

国立情報学研究所では、米国の美術分野の標準的シソーラスである the Getty's Art & Architecture Thesaurus (AAT) を軸に、博物館メタデータを記述した Dublin Core, VRA, CDWA, MPEG-7 のデータを分類する試みが行われている [28]。

また、博物館情報横断検索に分散オントロジを適用を検討する報告もある [29]。この報告では、標準的なメタデータスキーマを直接利用せず、多様性を許容する情報構造として、3 層の情報共有フレームワークを用いている。まず第 1 層のフレームワークでは、各博物館のメタデータスキーマの構造にある共通構造に変換を行う。ここでは記述内容の同一性は考えない。次に、第 2 層のフレームワークで記述されている語彙の相互変換と共有が行われる。最後に第 3 層のフレームワークで相互の関係情報をオブジェクトのリンクとして捉え、利用者ナビゲーションを行う。この報告では、現在第 2 層におけるオントロジ

の適用可能性が議論されている。

これらオントロジやシソーラスを用いたメタデータ・マッピングの議論は始まったばかりであり、今後博物館における情報共有を促進する上で有効な手段となり得るか、動向を見ていく必要がある。

メタデータ・マッピングの問題以外にも、各博物館におけるメタデータ作成が進んでいないとの報告もあり、メタデータ作成に関する問題も深刻である。

この他にも、実際の博物館における収藏品管理システム運用に関する問題点として、各館のシステムは館内 LAN のみに接続されているか、あるいはスタンドアロンで運用されている場合が多く、他館や他機関とのシステム連携はあまり見られないという点も挙げられる。更に、各館のシステムは担当者の有無に関わらずサーバシステムの管理者が必要であり、そのことが高い負荷になっているとの報告もある。

これらは直接博物館メタデータに関わる問題ではないが、今後、博物館メタデータの共有や相互利用を行う際に考慮しなければならない問題の1つである。

3. ファジー・スキーマ

3.1 柔軟な知識管理体系に基づくコンテンツ管理技術の提案

これまで、メタデータとメタデータスキーマ、特に博物館に特化した分野における問題点と関連動向を概観してきた。

そこでは、メタデータスキーマを構築する難しさ、メタデータを作成する手間、既に作られているメタデータスキーマを相互に活用する難しさ、が問題となっている。

今後、これらの問題に加えて、コンテンツの制作者・提供者からコンテンツの消費者までシームレスなコンテンツ流通を促進するために、コンテンツの権利を明確にするとともに、メタデータの信頼性を保障する技術の確立も必要となる。

こうした問題を解決するために本研究では、目的に応じて柔軟に形を変化させる知識管理体系を構築する技術を提案する。

人間が思考をする際、想起される情報群は置かれている状況や環境、経験に応じて変化する。人間は必要とする情報(コンテンツ)に関する知識管理体系(メタデータスキーマ)を状況や環境、経験に応じてその都度創り出し、その中で必要に応じて柔軟な取捨選択を行っていると考えられる。これに対して、現存するコンテンツ管理の手法を概観してみると、コンテンツにはあらかじめ決められたメタデータスキーマに基づくメタデータが付与されており、その内容項目の変更や改編等は通常行われぬ。つまり、そこには人間が思考する際に創り出すような、ダイナミックで柔軟性に富んだ知識管理体系は存在せず、固定で柔軟性に乏しい知識管理体系が存在するのみである。そしてこのことが、メタデータの利用における大きな障害となっている。

あらかじめ決められた知識管理体系を「静的なメタデータスキーマ」と呼ぶならば、人間が想起する柔軟な知識管理体系はコンテキストに応じた「動的なメタデータスキーマ」と言うことができよう。本研究では、双方のメタデータスキーマの定性

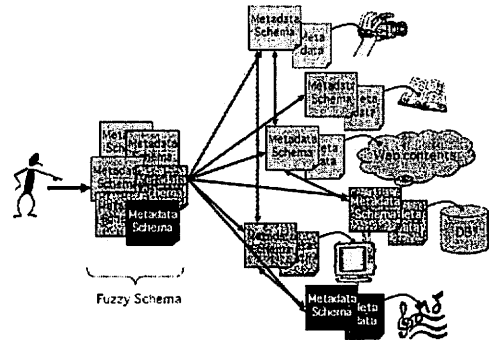


図3 柔軟な知識管理体系に基づくコンテンツ管理
Fig.3 Contents Management Technology Based on Flexible Knowledge Management System

的性質を捉え、どこまでが静的に管理されるべき体系であり、どこからがコンテキストに応じて動的に創り出される体系であるのか、その最適値の算出を定量的に行うことを考える。そして、その最適値算出方法をベースとし、必要に応じて静的なメタデータスキーマと動的なメタデータスキーマを緩やかにチューニングできる「ファジー・スキーマ」という概念を提案する。

ファジー・スキーマを用いることで、目的に応じて形を変化させることのできる柔軟な知識管理体系の構築が可能となる。ファジー・スキーマでは、ある程度自動的にメタデータスキーマを作成することが可能となるため、これまで専門家による高度な知識と膨大な時間を必要としてきたメタデータスキーマ構築を簡易化することが期待できる。

また、動的なメタデータスキーマの構築により、これまで人手によっていたメタデータ作成の(半)自動化も見込まれる。従って、メタデータ作成の負荷軽減が期待できる。

更に、メタデータスキーマを複数個連携させる手法を考える。それぞれのコンテンツは、そのコンテンツが何を表現しているのかを示すメタデータを持つ。これまで、ユーザはコンテンツごとのメタデータを用いてコンテンツにアクセスしてきた。先に述べたように、ファジー・スキーマでは、管理目的に応じて緩やかにメタデータスキーマを構築し直すことができる。そこで、あるメタデータスキーマと周辺のメタデータスキーマの双方を複合的に再編成する。1対1のメタデータ・マッピングとは異なるため、再編成による必要情報の欠落や無駄な情報の増加は存在しない。また、再編成に伴い、ユーザは各コンテンツのメタデータスキーマの差異を意識することなく、かつ複合的にコンテンツへのアクセスが可能となることを期待できる(図3参照)。

これらを実現するために、本研究では具体的に使用されている幾つかのメタデータスキーマを例に、その複合利用(ファジー・スキーマ化)を検討する。具体的には、複数の博物館の資料情報管理用メタデータスキーマを用い、横断利用が可能なメタデータスキーマの作成の指針とその問題点について考察する。

3.2 アプローチ

ファジー・スキーマを用いて、様々なメタデータスキーマを意識することなく横断的にコンテンツを利用するためには、いくつかのアプローチが考えられる。

3.2.1 オントロジを用いたアプローチ

メタデータスキーマ間でのメタデータ・マッピングをオントロジやシソーラスを用いて行う方法がある。

シソーラスは単語の上位/下位、全体/部分、同義/類義等の関係を分類し体系付けた辞書であり、古くから図書館では検索等に用いられてきた。博物館の分野には、AATやThesaurus of Geographic Names(TGN)、Thesaurus of Graphic Material(TGM)といった強力なシソーラスが数多く存在する。各メタデータスキーマにこれらのシソーラスを共通のシソーラスとして用いることで、例えば検索時に、メタデータスキーマを経由してある程度検索キーと近い知識を引き出すことが可能であろう。

また、オントロジは哲学の世界では「本質または実在に関する理論」と位置付けられ、それはしばしば計算機の世界では「概念化のための明示的な仕様」[30]とされている。Semantic Webはコンテンツの意味も含めて定義されたWebであり、現在のWebの拡張として設計されている。Semantic Webでは、コンテンツに含まれる概念の形式化記述を行うために、オントロジに基づくアプローチがなされており、実際W3Cではオントロジ記述言語OWLの開発も行われた。

Semantic Webの取り組みはかつての人工知能における知識表現と異なり、多様な不整合性を認める立場を取っている。つまり、Webというフィールドの中でコンテンツ記述言語と形式的な推論が可能なOWLを用い、知識表現を行おうとしている。

このアプローチは先行研究にも見られるように、Webのように限定された範囲内においては各メタデータスキーマを意味的に関連付けることができると考えられる。

しかし、これらのアプローチには問題もある。オントロジやシソーラスの維持管理と記述範囲の問題である。

辞書、あるいはオントロジは、新しい発見があるたびに、変更・追加・削除等のアップデートを繰り返して整備されなければならない。ある発見により自動的に辞書やオントロジがアップデートされるような仕組みがない限り、日々更新される内容の整合性が保証されない。

また、その記述範囲も問題となる。特に博物館分野において、その記述対象となるのは歴史・芸術・民族・産業・自然科学等、人間の営みを表現したほぼ全ての事象であり、Webに限定した範囲での記述で賄いきれるとは言いにくい。仮に記述できたとしても、OWL Fullのように計算の完全性や決定可能性が保証されていない状況ではそれが有効なものであるか、検証が難しい。

3.2.2 アノテーションを用いたアプローチ

Semantic Webにおけるオントロジに基づくアプローチは、概念を普遍的なものとして取り扱うため、トップダウン的なアプローチとすることができる。これに対して、ある特定のコンテンツに対する情報であるアノテーションやロコミ情報を用い

たアプローチはボトムアップ的なアプローチと言うことができる。

アノテーションはコンテンツに対するメタコンテンツとして、古くからWebページに言語的な注釈を付けるという方向で議論されてきた。WebページにそのWebページが何であるのかを記すアノテーションを付与し、要約や翻訳といった自然言語処理のためのリソースとして用いる方法である。W3CではAnnotea[31]と呼ばれるプロジェクトが進行している。これは、WebコンテンツにRDF形式の注釈を付けることで、注釈の付与および注釈の共有を行おうとするプロジェクトであり、メタデータを共有することで共同作業における利便性や効率の向上を目指している。

また、近年のWeblogにおけるコメントやトラックバック等の書き込み情報(いわゆるロコミ情報)もアノテーションの一種と見ることができる。つまり、Weblogにポストされたあるトピックに対して、それを評価する情報を人間が加えていくことで、そのトピックに対する周辺情報が蓄積され、最終的にそれがそのトピックの信頼性へとつながる。

アノテーションやロコミ情報等のコンテンツに対する言語的捕捉情報を解析することで、例えばあるアノテーションを持つWebページとあるロコミ情報を持つWebページの意味的な近さを得る等、コンテンツの内容理解を促進するものと考えられる。

これは博物館における複数のメタデータスキーマ間でのメタデータ・マッピングへの応用が可能である。つまり、あるメタデータを持つ情報に対して多数のアノテーションが付与されれば、メタデータが示す以上の情報を得ることができる。この情報を元に、自然言語処理等の言語的解析によりメタデータ間の近さを図り、マッピングすることが可能であると考えられる。

しかし、アノテーションによるアプローチにも問題がある。そもそも誰がアノテートするのかという問題とアノテーション自体の信頼性の保証である。

コンテンツに対する捕捉の情報であるアノテーションはコンテンツに熟知した人間、例えばコンテンツ制作者でなければ付与できない。しかし、図書館や博物館の世界で目録情報を作成する専門家がいることを考えればわかる通り、コンテンツ制作者はあまりコンテンツに対する捕捉情報の付与を行わない。更に、Weblogの書き込み情報は一般的に匿名性が高く、一概に信用できる情報は少ない。

3.2.3 統計処理手法を用いたアプローチ

これらに対して、本研究では統計処理手法によるアプローチを取る。

本研究が提案するファジー・スキーマでは、まずあるメタデータスキーマを、静的な部分と動的な部分を切り分ける最適値を定量的に算出することを行う。その過程において動的に生成できる部分が他のメタデータスキーマのそれとどの程度オーバーラップしているのか、「あいまい度」のようなものを計算することで異なるメタデータスキーマ間で確率的なメタデータ・マッピングが可能なのではないかと考えた。

これを実現するためには、あるメタデータスキーマ内で定義

される要素が他の要素とどの程度の距離にあるのか、その類似度を調べる必要がある。具体的にはパターン認識や統計処理で用いられているサポートベクターマシンや主成分分析、重回帰分析などが有効であると考えられる。

更に、前項のアプローチはいずれも一長一短があるものの、単なる1対1の項目マッピングに比べれば有効性があると思われる。そこで、本研究ではこれらのアプローチも統計処理的な手法の比較対象として用いることを検討している。

3.3 博物館情報を用いた複合的メタデータスキーマの作成
統計処理手法を用いた確率的メタデータ・マッピング実現のため、ファジー・スキーマの基本的な言語設計を行っている。

具体的には、実際に博物館で利用されているメタデータスキーマを用い、手作業によるメタデータスキーマの分類を行うことで自動化や統計処理の手がかりとする。

現在検討しているメタデータスキーマを以下に示す。

- ミュージアム資料情報構造化モデル
- CIDOC/CRM
- CIDOC/IC
- Dublin Core
- MDA SPECTRUM
- CDWA
- 博物館遺物管理電算化のための遺物分類標準

これらの中で特に、様々なメタデータスキーマの手本となっている Dublin Core と、日本において今後中心的なメタデータスキーマになるであろうミュージアム資料情報構造化モデルをベースに、特徴的と思われる部分を重点的に分類を進めて行く予定である。

4. まとめと今後の課題

本稿では、博物館を中心とするメタデータの問題を整理した。ここでは、メタデータスキーマを構築する難しさ、メタデータを作成する手間、既に作られているメタデータスキーマを相互に利用する難しさが問題となっていることがわかった。そこで、これらの課題を解決するために必要に応じて柔軟にメタデータスキーマを構築できる「ファジー・スキーマ」という概念の提案を行った。更に、ファジー・スキーマ実現のためのいくつかのアプローチを提示し、その言語設計の指針を示した。今後は、具体的な言語設計を行い、ファジー・スキーマの構築に取り組んで行く予定である。

文 献

- [1] 経済産業省商務情報政策局監修: "デジタルコンテンツ白書 2005", デジタルコンテンツ協会 (2005)
- [2] Dempsey et al.: "Metadata: A Current View of Practice and Issues", J. of Documentation, Vol. 54, No. 2, pp. 145-172(Mar. 1998).
- [3] 日本図書館情報学会研究委員会編: "図書館目録とメタデータ", 勉誠出版 (2004)
- [4] ISAD(G): General International Standard Archival Description (参照 2006-06-28)
< URL: http://www.ica.org/biblio/isad_g_2e.pdf >
- [5] Dublin Core Metadata Initiative (参照 2006-06-28)
< URL: <http://dublincore.org/> >
- [6] MPEG-7 (参照 2006-06-28)

- < URL: <http://www.itacj.ipsj.or.jp/mpeg7/> >
- [7] TV-Anytime (参照 2006-06-28)
< URL: <http://tv-anytime.org/> >
- [8] 北村義弘, 岩崎圭介, 田中秀明: "電子出版と X MDF 技術", シャープ技術報, 第 84 号 (2002-12)
- [9] Broad Band e-Book (参照 2006-06-28)
< URL: <http://ps.canon-sol.jp/bc/kikaku.htm> >
- [10] MusicBrainz (参照 2006-06-28)
< URL: <http://musicbrainz.org> >
- [11] The Open Archives Initiative Protocol for Metadata Harvesting (参照 2006-06-28)
< URL: <http://www.openarchives.org/> >
- [12] 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).
- [13] Diane I. Hillmann, Elaine L. Westbrooks: "Metadata in Practice", American Library Association (2004)
- [14] 秋元良仁, 鈴木理洋: "博物館情報の相互利用を目的とした文化財情報システムの提案", 情報処理学会人文科学とコンピュータシンポジウム「じんもんこん 2004」(デジタルアーカイブ- デジタル学術情報資源の共有と活用 -), (2004-12)
- [15] 村田良二: "ミュージアム資料情報構造化モデルの開発", 情報処理学会人文科学とコンピュータシンポジウム「じんもんこん 2005」(デジタルアーカイブ - その理論の深化と技術の応用 -), (2005-12)
- [16] RDF Schema (参照 2006-06-28)
< URL: <http://www.w3.org/TR/rdf-schema/> >
- [17] ICOM-CIDOC IC (参照 2006-06-28)
< URL: <http://www.willpowerinfo.myby.co.uk/cidoc/> >
- [18] The CIDOC Conceptual Reference Model (参照 2006-06-28)
< URL: <http://cidoc.ics.forth.gr/> >
- [19] MDA SPECTRUM (参照 2006-06-28)
< URL: <http://www.mda.org.uk/spectrum.htm> >
- [20] Categories for the Description of Works of Art (CDWA) (参照 2006-06-28)
< URL: http://www.getty.edu/research/conducting_research/standards/cdwa/ >
- [21] CDWA Lite: XML Schema Content for Contributing Records via the OAI Harvesting Protocol (Version 1.0) (参照 2006-06-28)
< URL: http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/index.html >
- [22] MINERVA DC.Culture (参照 2006-06-28)
< URL: <http://www.minervaeurope.org/DC.Culture.htm> >
- [23] Research Libraries Group (参照 2006-06-28)
< URL: <http://www.rlg.org/> >
- [24] 田窪直規: "韓国における文化財情報のデジタル化とネットワーク化について", アート・ドキュメンテーション研究, 第 13 号 (2006-03)
- [25] 原正一郎, 相田満, 入口敦志, et al.: "データベースの共有におけるデータマッピングの事例的研究", 情報処理学会研究報告, Vol. 2005, No. 76, pp. 31-38 (2005-07)
- [26] 安達文夫: "博物館資料情報の共有化のための共通メタデータへのマッピング", 画像電子学会, 第 4 回画像ミュージアム研究会 (2006-03)
- [27] 三嶋章浩, 斎藤伸雄, 秋元良仁: "収蔵品管理 ASP システムの試作と実行結果", アート・ドキュメンテーション学会, 第 17 回年次大会, pp. 45-47 (2006-06)
- [28] 藤沢仁子, Frederic Andres: "シソーラスによる複数のメタデータの統合", 情報処理学会研究報告, Vol. 2005, No. 105, pp. 33-40 (2005-10)
- [29] 山田篤, 安達文夫, 小町祐史, et al.: "博物館情報横断検索における分散オントロジーの検討", 画像電子学会, 第 4 回画像ミュージアム研究会 (2006-03)
- [30] AIDOS 著: "オントロジ技術入門 - ウェブオントロジと OWL -", 東京電機大学出版局 (2005-09)
- [31] Annotea Project (参照 2006-06-28)
< URL: <http://www.w3.org/2001/Annotea/> >