

## 異分野を横断し流通する学術コンテンツのメタデータ設計

吉永 早織<sup>†‡</sup> 古川 康一<sup>‡</sup> 嶋津 恵子<sup>†</sup>慶應義塾大学デジタルメディア・コンテンツ統合研究機構<sup>†</sup> 慶應義塾大学政策・メディア研究科<sup>‡</sup>

## 1. はじめに

我々は利用者が想定した内容を裏づける情報や、意外性のある情報を入手できる情報システムの実現を目指し、検索されたコンテンツを利用者の考えるコンテキストに沿って並べ替える機能を構築中である。これにより、直接的な回答となるコンテンツが探し出せない/存在しない場合でも必要な情報の入手が可能となる。このとき、専門外のコンテンツも取り扱えることで、より強力な支援を実現できる。専門外のコンテンツを有効に活用するには、そのコンテンツに何が書かれているかということ、利用者が理解できることが必要である。専門家でない利用者が専門知識を理解し、情報を有効利用する方法として KM(Knowledge Management) では 5W1H に従って情報共有が必要とされ[1]、さらに[2]はこれをコンテンツ流通に応用させるべく特に人(who)、地名(where)、時(when)に相当するメタデータの利用を提案している。一方、慶應義塾福沢研究センターが所有するある写真コンテンツは、「福沢諭吉筆『西航記』(後略)」「遣欧使節団に随行(後略)」「1861~1862」というメタデータを持つ。しかし一般の利用者がこのコンテンツを探したい場合「福沢諭吉 ヨーロッパ 19世紀」などが用いられるであろう。そこで我々は、メタデータの値として使用される用語の専門性を下げ、一般の利用者が理解できるレベルに変換することが必要であると考えた。今回、コンテンツの持つメタデータを対象に、「地名」、「時」として分類できるメタデータの値の粒度を変えて別のメタデータとして持たせた。具体的には、「地名」に分類できるものの値に対し、それが存在する大陸名を値とする別のメタデータを、また「時」に分類できるメタデータの値に対し、それを西暦表記による年月日を値とする別のメタデータと世紀表記を値とするものの2つのメタデータを追加した。作成したメタデータを用いて、コンテンツを閲覧しても一般の利用者がその内容や作成背景を把握できないと考えられる専門性の高いコンテンツを利用し、検索実験を行った結果、我々の提案するメタデータの構成の有効性を確認した。

## 2. メタデータの例

国立国会図書館のデジタルアーカイブポータル<sup>1</sup>中の30件のコンテンツを対象に、メタデータの値が「地名」、もしくは「時」を表すものが存在するか、それらは複数個存在するか、そして複数個存在する場合、同一の内容を表現(粒度)を変えたものが値になっているかを調べた(Table1)。全コンテンツ中、同一の内容で粒度を変えた値をメタデータとして持つものは10分の1であった。

Table1 メタデータの例

	メタデータを持つコンテンツ	左記中で「地名」、「時」である複数のメタデータを持つコンテンツ	左記中で同じ意味で粒度を変えた値をもつコンテンツ
地名	10/30	6/10	3/6
時	30/30	8/30	0/8

## 3. 拡張メタデータ構成案

次の方法で拡張版メタデータを作成した。

「地名」メタデータの拡張: 「地名」であるメタデータすべてに対しそれが存在する「大陸名」を値とするものを手作業で作成(Fig.1)する。

「時」メタデータの拡張: 「時」であるメタデータすべてを対象に、「世紀」表現を値とするものを手作業で作成する(Fig.2)。この方法で作成される拡張版メタデータの構成イメージは Fig.3 の通り。

## 4. 比較実験と考察

さらに、拡張版メタデータを利用したコンテンツ検索の実験を行った。「地名」と「時」のメタデータを持つ160件のコンテンツを用意した(C1)。さらに同じ160件のコンテンツを別に用意し、前述した拡張版メタデータを付与した(C2)。この2セットのコンテンツを対象にそれぞれ同じ20例の検索を行い、検索結果の精度と再現率を比較した(Table2, Table3)。そして、C1とC2のコンテンツセット別に精度と再現率の平均を算出し両者を比較した(Table4)。C1を対象にした検索の平均精度は24%、同再現率は33%。C2を対象にした検索の平均精度は19%、同再現率は平均78%であった。どちらも20のテスト事例に

Metadata Design for Academic Digital Contents  
<sup>†‡</sup>Saori Yoshinaga, <sup>‡</sup>Koichi Furukawa, <sup>†</sup>Keiko Shimazu (Keio Univ.)

<sup>1</sup> <http://www.dap.ndl.go.jp/home/>

よる数値のばらつきは大きくなく同程度であった。つまり拡張版メタデータを付与したコンテンツを対象にした検索では、妥当性の高いものが漏れなく出力される傾向にあるといえる。

Fig.1 “地名”の拡張メタデータ作成手順

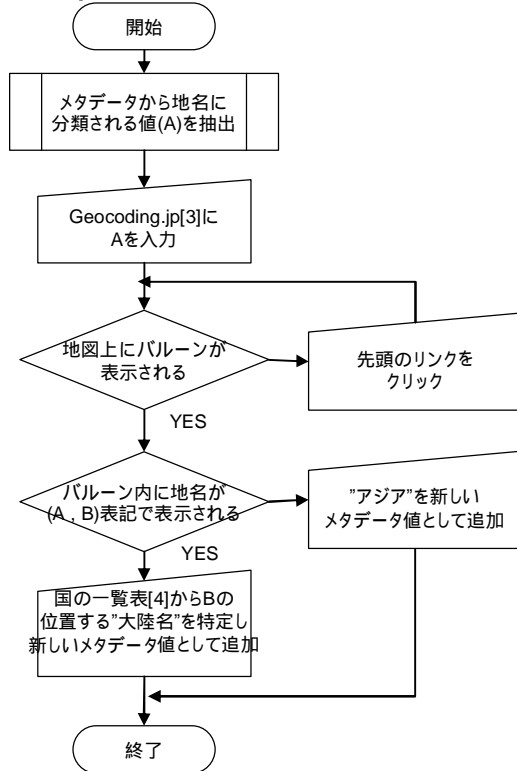


Fig.2 “時”の拡張メタデータ作成手順

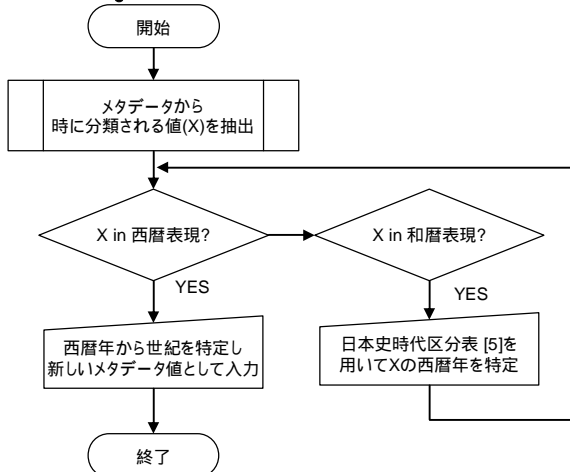


Fig.3 拡張版メタデータの構成イメージ

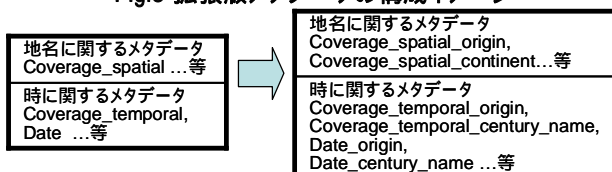


Table 2 実験のテストケース(部分)

実験 No	検索目的	検索キーワード	TargetContents 数 <sup>1</sup>
1	中世ヨーロッパの邸宅の見取り図を見たい	ヨーロッパ 貴族 邸宅 図	3
2	日本と西洋の人体解剖図の描き方の違いを見たい	日本 ヨーロッパ 人体 解剖	8
3	18世紀の貴族の結婚生活の様式を知りたい	18世紀 貴族 結婚生活	4

Table 3 実験結果(部分)

実験 No	検索に使ったメタデータ	精度 <sup>2</sup>	再現率 <sup>3</sup>
1	C1	16	33
	C2	30	100
2	C1	40	25
	C2	16	50
3	C1	20	50
	C2	17	75

Table 4 実験結果の平均

検索に使ったメタデータ	精度	再現率
C1	24	33
C2	19	78

## 5. まとめ

“地名”, “時” に注目した拡張版メタデータの構成を提案した。この拡張版メタデータをコンテンツに付与しておくこと、妥当性の高いものが漏れなく出力される傾向にあることがわかった。

## 謝辞

本稿は、文部科学省科学技術振興調整費の支援による研究の一部である。

## 参考文献

- [1] Hansen, M.T., Nohria, N., Tierney, T.: What's Your Strategy for Managing Knowledge?. Harvard Business Review, March April, pp106-117(1999).
- [2] Keiko Shimazu, Tatsuya Arisawa : Interdisciplinary Contents Management using 5W1H Interface for Metadata, International Conference on Web Intelligence (2006).
- [3] <http://www.geocoding.jp/>
- [4] [http://ja.wikipedia.org/wiki/国の一覧\\_\(大陸別\)](http://ja.wikipedia.org/wiki/国の一覧_(大陸別))
- [5] <http://ja.wikipedia.org/wiki/日本史時代区分表>

<sup>1</sup> TargetContents 数: 検索目的に合致しているコンテンツの数

<sup>2</sup> 精度: (実際に検索された TargetContents 数) ÷ (出力全コンテンツ数) × 100

<sup>3</sup> 再現率: (実際に検索された TargetContents 数 ÷ TargetContents 数) × 100