

部分的分類知識の統合による博物館情報の横断検索の提案

画像電子学会 画像ミュージアム研究会 博物館・美術館 DTD-SG

山田 篤[†] 安達 文夫[‡] 小町 祐史[§]

Atsushi YAMADA[†] Fumio ADACHI[‡] Yushi KOMACHI[§]

[†] 京都高度技術研究所

[†] ASTEM RI/Kyoto

[‡] 国立歴史民俗博物館

[‡] National Museum of Japanese History

[§] 大阪工業大学

[§] Osaka Institute of Technology

E-mail: [†] yamada@astem.or.jp, [‡] adachi@rekihaku.ac.jp, [§] komachi@y-adagio.com

1. はじめに

博物館・美術館情報の電子化が進み、ネットワークを通じて個々の館の収蔵品に関する情報の提供サービス、検索サービスなどが開始されている。博物館情報の利用者にとっては、どの館にアクセスするかを意識せず、各館の差異を意識せずにシームレスに検索ができること、つまり横断検索できることが望ましい[1]。

本稿では、各館から提供される分類知識を統合することにより、このような横断検索を実現する方法について述べる。

2. 横断検索のためのフレームワーク

現状では、ほとんどの館がそれぞれ独自のサービスを提供しているため、

- 検索方法が館によって異なる
- ある館の情報を他の館の情報と関連付けて見ることが難しい

などの問題点があり、横断検索を困難にしている。そこで横断検索のためには、各館の情報を統合する仕組みが必要となる[2]。

画一的な共通フォーマットを用いずに、多様性を許容する情報構造として、次の3レベルに階層化される情報共有のフレームワークが提案されている[1]。

- (1) 情報記述構造レベル
- (2) 情報記述内容レベル
- (3) 情報ナビゲーションレベル

このうち、(1)については、様々な取り組みがなされており、記述スキームの共有や標準化といった試みもある[3][4]。

これに対して、内容レベルの情報共有については、様々な収蔵品を扱うという事情から困難な問題として残っている。対象物が基本的に物であることから、図書のような全文検索といった手法も適用できない。オントロジ[6]や概念辞書[9]を用いた手法も検討されているが、収蔵品に関して様々なメタデータを付与することは博物館にとっても大きな負担となる。

このため、個々の博物館がなるべく簡単に用意できる情報をもとにして、内容レベルの情報共有を行い、横断検索を可能にすることが望まれる。

3. 分類知識

収蔵品に関して、博物館になるべく負担をかけずに収集可能なメタデータとして、収蔵品の分類に関する情報を考える。

ある館が所蔵する収蔵品に関して、主にそれらを管理する目的で分類がなされ、分類名(ラベル)が付与されている状況を考える。このようなラベルを横断検索において利用することを検討する。

このような分類は博物館毎、あるいは博物館の学芸員毎、コレクション毎になされるため、他との関係において、そのまま使えるとは限らない。たとえば、A館の分類名XとB館の同じ分類名Xが同じ内容を表していることは保証できない。各

館は自らのカバーする範囲においてラベルをつけるため、これらの分類知識は部分的なものにならざるをえない。

よって、これらの部分的な知識を統合し利用するための仕組みとして、ポータルサイトを用いた統合を考える。ポータルサイトにおいて、各館から得られる分類知識を統合し、それらの間の差異を吸収して、横断検索に用いる。

4. 分類知識の統合実験

個々の博物館から提供される分類知識の統合方法について検討するために、簡単な予備実験を行った。

国立歴史民俗博物館の収蔵品データから衣類に関する 15 のコレクションを選び、それぞれのコレクションに含まれる資料名称などをもとに、コレクション毎に分類語彙を設定した。例を図 1 に示す。

```
<collection id="H-35" name="野村正治郎衣裳コレクション">
  <set name="衣裳">
    <set name="小袖" />
    <set name="振袖" />
    <set name="帷子" />
    <set name="被衣" />
    <set name="襦袢" />
    <set name="腰巻" />
  </set>
</collection>
<collection id="H-36" name="徳川（有栖川宮）実枝子所用衣裳">
  <set name="衣裳">
    <set name="袷" />
    <set name="帯" />
    <set name="綴織帯" />
    <set name="振袖" />
  </set>
</collection>
```

図 1 分類語彙の例

たとえば、野村正治郎衣裳コレクションでは、衣裳、小袖、振袖、帷子、被衣、襦袢、腰巻といっ

た分類語が用いられ、このうち、衣裳とそれ以外の語の間に階層関係があったことを表している。

15 コレクションについて、使用された分類語彙はのべ 291 語、このうち、ユニークなものは 169 語あった。さらにこのうち、126 語は出現回数が 1 回のみであり、1 つのコレクションでしか用いられていなかった。よって、他のコレクションと共通して現れた語は 43 語であった。

次にそれぞれのユニークな語について、どのコレクションで用いられていたかを抽出した。例を図 2 に示す。

```
<set name="衣裳">
  <collection id="H-35" name="野村正治郎衣裳コレクション" />
  <collection id="H-36" name="徳川（有栖川宮）実枝子所用衣裳" />
  <collection id="H-1694" name="近・現代女性洋装衣裳" />
</set>
<set name="小袖">
  <collection id="H-35" name="野村正治郎衣裳コレクション" />
  <collection id="H-1359" name="幕末・明治初期服飾資料" />
</set>
<set name="振袖">
  <collection id="H-35" name="野村正治郎衣裳コレクション" />
  <collection id="H-36" name="徳川（有栖川宮）実枝子所用衣裳" />
  <collection id="H-1359" name="幕末・明治初期服飾資料" />
</set>
```

図 2 共通分類語彙の例

たとえば、衣裳という分類語彙は、H-35（野村正治郎衣裳コレクション）、H-36（徳川（有栖川宮）実枝子所用衣裳）、H-1694（近・現代女性洋装衣裳）で用いられていることを表している。

最後に、共通語彙を手がかりにして、コレクション間の類似度を計算した。コレクション A と B の間の類似度を次のように定義した。

$$\frac{A \text{ と } B \text{ の共通語彙数} \times 2}{A \text{ の語彙数} + B \text{ の語彙数}}$$

すべての語彙が共通であれば、100%となり、共通した語彙が全くなければ0%になる。このようにして計算した類似度の例を図3に示す。

```
<collection id="H-35" name="野村正治郎衣裳コレクション">
  <collection id="H-36" name="徳川(有栖川宮)実枝子所用衣裳" similarity="33%" />
  <collection id="H-1182" name="海軍礼服" similarity="0%" />
  <collection id="H-1314" name="柳沢家伝来服飾資料" similarity="0%" />
  <collection id="H-1359" name="幕末・明治初期服飾資料" similarity="18%" />
  <collection id="H-1395" name="大正末～昭和初期着物類" similarity="0%" />
  <collection id="H-1694" name="近・現代女性洋装衣裳" similarity="9%" />
  <collection id="F-7" name="青森県の衣生活用具" similarity="8%" />
  <collection id="F-8" name="青森県南地方の仕事着コレクション" similarity="0%" />
  <collection id="F-148" name="婚礼衣裳・婚礼用具及び生活用具" similarity="10%" />
```

図 3 類似度の計算

たとえば、この例では、H-35(野村正治郎衣裳コレクション)との類似度が、H-36(徳川(有栖川宮)実枝子所用衣裳)は33%、H-1359(幕末・明治初期服飾資料)は18%、H-1694(近・現代女性洋装衣裳)は9%、F-7(青森県の衣生活用具)は8%、F-148(婚礼衣裳・婚礼用具及び生活用具)は10%、F-211(秋田の衣服)は9%、F-338(加藤家衣生活資料)は5%で、残りのH-1182(海軍礼服)、H-1314(柳沢家伝来服飾資料)、H-1395(大正末～昭和初期着物類)、F-8(青森県南地方の仕事着コレクション)、F-150(秋田・青森・岩手の仕事着及びこぎん・ひしざし模様見本)、F-220(津軽・南部の衣服)、F-337(東京近郊農

家の衣生活資料)はすべて0%であったことを表している。

今回の実験では、振袖などの和服を集めたコレクションと、生活衣料を集めたコレクションが混在していたが、完全には分離されないものの、それぞれを多く含むコレクションどうしが高い類似度で結びつく傾向があった。

一方、分類語として複数の意味で使われる語があると、それを介して異なるものが結びつく可能性もあった。ただし、それが全体に占める割合が小さければ、類似度への影響も少なくなる。

なお、今回の予備実験では、コレクション毎の類似度の計算を行ったが、階層構造を持つ分類知識の場合は、共通語彙間の類似度の計算を行うことができる。たとえば、A館とB館で共通して衣裳という分類語彙が用いられており、それぞれの語の下に階層的にさらに分類語彙が用意されているような場合には、それぞれの語の子孫で共通の分類語彙をもとに、上と同様の手法で類似度を計算することができる。

5. 横断検索への利用

前節では、機械的な処理により、互いに独立した複数の館・コレクションの分類語彙を、類似度という尺度を導入して結びつけた。本節では、これを横断検索にいかに関与するかという観点から検討を加える。

最も典型的な利用例は、類似したものを探すとことである。たとえば、今、振袖を検索していて、これに類似するもの(類概念)で検索を行いたいとしよう。ところが、ある館では振袖の上位概念は衣裳であったり、別の館では着物であったりする。また、中には振袖を含むが、全く異なる分類になっているものもあり得る。そこで、振袖と共通の上位概念をもつもので類似度の高いものから順に表示するようなインターフェースがあれば、可能性の高いものから順に見ていくことができる。

次に、あるコレクションを見ているときに、類似度の高い別の館のコレクションを推薦すると

いった利用法が考えられる。これも、分類語を直接指定するのではなしに、今見ているものから探索範囲を漸次広げていくといった検索が可能になる。

今回、考慮していない問題として、利用者の検索語彙と、博物館の分類語彙をどのようにすりあわせるかという問題がある。上記の例でも、何か特定のものにたどり着いてから、それに類似するものを検索していくというストーリーを想定している。たとえば、博物館側はより専門的な語彙で記述し、利用者はより一般的な語彙で検索する場合に、利用者が入力する語彙と、博物館側で想定している分類語彙との間にギャップが生じ、うまく検索できない可能性がある。本手法では、博物館から提供される分類語彙が増えれば増えるほど、そのカバーする範囲も増えていくという想定にたっているが、利用者の語彙体系も類似度計算の対象にするといった工夫が必要になるかもしれない。

本稿では、類似度の計算を完全に一致する語彙のみをもとにして行っただが、分類語の中から主要語（ヘッド）を切り出したり、類義語や上位・下位概念を記述した辞書やシソーラスが利用可能であれば、それらを用いたりすることにより、さらに精緻な類似度判定を行うこともできる。

また、機械的に結びつけたものどうしが本当に関連するものか否かを、利用者からのフィードバックによって修正し漸次更新していく枠組みも必要であると考えている。

6. おわりに

本稿では、博物館、美術館の収蔵品の横断検索において、博物館から提供可能な分類知識を用いて、記述内容に基づく検索を行う方法について検討を行った。

個々の館から提供される分類知識は部分的であるが、共通語彙を手がかりにして、これを他の館の分類知識と関連づけることにより、全体を統合することで、他の館の収蔵品と関連づけた横断検索に利用することができる

博物館から提供される専門的な語彙と、利用者が用いる一般的な検索語彙をいかに関連づけるかや、利用者からのフィードバックによる館をまたがる関連性の修正や更新は、今後の課題である。

文 献

- [1] 山田篤, 他: 博物館情報の知的横断検索のためのフレームワーク, 画電年次大会, 2002-06.
- [2] 山本泰則, 中川隆: 博物館資料情報共有の試み, 画電年次大会, 2004-06.
- [3] 文化財情報システムフォーラム (<http://www.tnm.go.jp/bnca/>).
- [4] The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC) (<http://www.cidoc.icom.org/>).
- [5] 山田篤, 他: 博物館情報の分類マッピングを用いた横断検索, 画電年次大会, 2004-06.
- [6] 山田篤, 他: 博物館情報の横断検索におけるオントロジ利用の試み, 画像ミュージアム研究会, 2005-03.
- [7] 山田篤, 他: 博物館情報横断検索における分散オントロジの検討, 画像ミュージアム研究会, 2006-03.
- [8] 山田篤, 他: 博物館情報検索のためのオントロジ・ユースケースの検討, 画電年次大会, 2006-06.
- [9] 山田篤, 他: 博物館横断検索に向けた概念辞書の枠組みの検討, 画像ミュージアム研究会, 2007-03.