

博物館横断検索に向けた概念辞書の枠組みの検討

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[§] 河合 正樹[¶]

Atsushi YAMADA[†] Fumio ADACHI[‡] Yushi KOMACHI[§] and Masaki KAWAI[¶]

[†] 京都高度技術研究所

[‡] 国立歴史民俗博物館

[§] 大阪工業大学

[¶] インフォコム

[†] ASTEM RI/Kyoto

[‡] National Museum of Japanese History

[§] Osaka Institute of Technology

[¶] INFOCOM Corporation

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

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、このような横断検索において必要となる概念辞書の枠組みについて述べる。

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]。

単純な情報統合の方法として、収蔵品の関連情報を記述する共通フォーマットを定義し、各館がそれに従って情報提供できれば、横断検索の基盤は形成される。文化財情報については、文化財情報システムフォーラム[3]において共通索引の試みが行われている。国際的には、CIDOC[4]の取組みがある。

しかし実際に存在する博物館は多様であって、扱う対象も規模も異なる。対象を扱う態度は博物館によって異なり、提供する収蔵品関連情報も様々である。博物館による収蔵品に関する関連情報の違いは各館の独自性の現れであって、横断検索の名の下に共通化することは必ずしも適切ではなく、共通化は困難に近い。

そこに利用者側の要求が加わると、問題はさらに複雑になる。利用者の要求もまた多種多様であり、すべての要求に対する解を予め準備しておくことは現実的ではないからである。

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

収蔵品に関連する情報を記述する構造の共通化を図るために、各館に対して標準化された情報構造を採用することを求めたり、館固有のインハウス情報構造の利用を否定したりせず、各館の情報を共通構造に変換する。共通構造への変換ができれば、例えばデータベースのスキーマレベルで、A館のスロットXがB館のスロットYに対応するという情報共有が可能になる。

(2) 情報記述内容レベル

情報記述構造レベルでの情報共有が実現しても、そこに記述される内容の表記の統一がなけれ

ば、単純な文字列操作ではその同一性を判定できない。そこでこのレベルで、記述する語彙の相互変換と共有を図る[5]。この際に、概念レベルの変換辞書が必要となる[6][7]。

(3) 情報ナビゲーションレベル

オブジェクトとしての収蔵品間の関係情報を、オブジェクトへのリンクとして扱う。博物館に属する専門家・学芸員などがもつ知識には、このレベルの情報が多し。収蔵品を特定の視点から分類した情報、利用者に対して収蔵品の見方をガイドする情報などは、利用者ナビゲーションのシナリオ記述として、このレベルで記述する。これらは必ずしも博物館に属する専門家だけによって作成される必要はない。

3. 横断検索の構成

横断検索を実現するにあたっては、利用者からの検索要求を受け付ける窓口が必要になる。この機能をポータルサイトとして実現することを考える。

一方で、実際の検索対象データは各館にあるものとする。そうすると、利用者からの検索要求を受け付けたポータルサイトは、各館に対して検索要求を出し、各館から返ってきた検索結果をとりまとめて利用者に提示することになる。この関係を図1に示す。

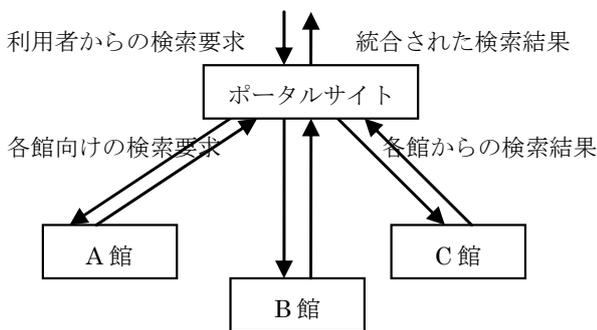


図1 横断検索の構成

ポータルサイトで行う検索として、ここでは

a) 名称を用いた検索

b) 分類を用いた検索

を考える。

このうち、収蔵品の名称を用いて検索する場合は、文字列の一致による検索を行うことになる。ポータルサイトは、利用者が入力した検索文字列を名称の部分として含むものを各館に問い合わせる。このとき、利用者は目的とするものの名称を知っているか（完全一致の場合）、目的とするものの名称に利用者が入力した文字列が含まれていなければならない（部分一致の場合）。また、部分一致の場合は、入力された文字列を部分文字列として含んでさえいけば、実際にそれが何であるかに関わりなく検索されることになる。このため、たとえば、「爪」の検索で「爪袋」も検索される[8]。

これに対して、収蔵品の分類を用いて検索を行う場合には、それが何であるかという情報を用いて検索を行うことができる可能性がある。このとき、分類の仕方やそこで用いられている語彙は各館毎に異なりうるため、ポータルサイトにおいて、これらと利用者要求との間の関連づけを行う必要がある[5]。また、先述したように、実際の検索対象データは各館にあるものとする。すなわち、ポータルサイトには、検索対象の個々の収蔵品に関する詳細な情報は存在しないという前提である。この場合のポータルサイトの動作は以下のようになる。

- (1) 各館が用いている分類語彙を問い合わせ、取得する
- (2) 利用者要求と、各館の分類語彙との対応を計算する
- (3) 利用者要求を各館毎の分類語彙に変換して、各館に問い合わせる

以下では、この関連づけのために概念辞書を用いることを検討する。

4. 概念辞書の検討

上述の分類を用いた横断検索を実現するために、以下のような条件の下で、必要となる概念辞書について検討を行った。

国立歴史民俗博物館の収蔵品データからいくつかのコレクションを選び、それぞれを異なる館と見なして、資料名称の第1階層ないし第2階層までを各館の分類語彙として用いた。こうして得られた各館の分類語彙を以下に示す。

【A館】野村正治郎衣裳コレクション

【B館】中世後期武家縫箔打掛衣裳 復元

【C館】婚礼衣裳・婚礼用具及び生活用具

└婚礼衣裳

└婚礼用具及び生活用具

【D館】秋田・青森・岩手の仕事着及びこぎん・ひしぎし模様見本

└秋田・青森・岩手の仕事着

└津軽こぎん・南部ひしぎし

└津軽こぎん実測図・津軽つづれこぎん模様見本

└南部つづれひしぎし模様見本

【E館】青森県の衣生活用具

各館で用いられている分類語彙として、ポータルサイトではこれらの語彙が取得されたものとする。

次に、利用者要求として「衣装」が入力されたとする。すると、ポータルサイトでは各館の分類語彙との照合を行い、各館向けの検索に用いる語彙を決定する。

これを順に見ていくと、A館の分類語彙「野村正治郎衣裳コレクション」に対しては、形態素解析及び係り受け解析を行い、「衣裳」を抽出することができたとする。すると、「衣装」と「衣裳」の関係が判断できればよいことになる。ここで概念辞書に「衣装」の類義語として「衣裳」が書かれていれば、利用者要求の「衣装」について、A館に対しては「野村正治郎衣裳コレクション」に含まれる収蔵品を要求すればよいことになる。B館についても同様に処理できる。類義語という観点からは、「衣服」「装束」といった語が用いられた場合も同様である。

次にC館について見てみると、第1層の「婚礼

衣裳・婚礼用具及び生活用具」では「衣裳」と「用具」が並置されており、「衣装」とは一致しない。もしもこの分類語彙しか存在しない場合は、E館における処理とほぼ同様となるが、C館については第2層の分類語彙が存在するため、これを見つめる。「婚礼衣裳」は先と同様「衣裳」を取り出し、「衣裳」を「衣装」の類義語と判断することで、この分類語彙をもつ収蔵品を要求対象に含めることができる。一方、「婚礼用具及び生活用具」は「用具」を取り出し、「用具」と「衣装」の関係を判定することになる。概念辞書の構成としては、これらが isa 階層（上位、下位の関係）において異なる枝に属することがわかれば、「衣装」の検索において「用具」を検索対象に含める必要がないことが判定できる。ただし、一般には「衣装」は「用具」の下位概念である可能性もある。その場合は、上位概念に拡張した検索を行うか否かという問題になる。

D館についても、第1層の「秋田・青森・岩手の仕事着及びこぎん・ひしぎし模様見本」ではなく、第2層語について見ていく。一つめの「秋田・青森・岩手の仕事着」からは主要語として「仕事着」を取り出す。次に「仕事着」と「衣装」の関係を判断する。今度は概念辞書を用いて「仕事着」が「衣装」の下位概念（直接の子でなくても、isa階層をたどって行って到達できる子孫であればよい）であることが判定できれば、「衣装」の検索において「仕事着」を検索対象に含めてよいことがわかる。次の「津軽こぎん・南部ひしぎし」からは「こぎん」「ひしぎし」を抽出する。これらはどちらかというと専門的な語彙であるため、一般的な概念辞書の記述対象であるかは疑問であるが、「刺し子」の一種である。「刺し子」と「衣装」の関係として、isa階層の上位・下位の関係が得られなかったとすると、「衣装」の検索において、これらは対象外と判断できる。しかし、実際には、ここには isa ではないが、「衣装」で用いられる「布」の装飾に「刺し子」が用いられるという関係が存在している。そのため、「津軽こぎん・南部ひしぎし」に分類されている収蔵品を見

てみると、実は「衣装」であったという可能性も存在する。しかしながら、これは各収蔵品を見てみないとわからないことである。残りの「津軽こぎん実測図・津軽つづれこぎん模様見本」「南部つづれひしぎし模様見本」については、「実測図」「模様見本」を取り出し、これらが「衣装」と上位・下位の関係にないことから検索対象に含めないという判断をする。

最後の E 館については、「青森県の衣生活用具」という語彙しか得られていない。この判定はきわめて困難であるが、仮に「衣装」が「衣生活用具」の下位概念であるとわかったとすると、「衣生活用具」のうち、一部は「衣装」である可能性がある。しかし、個々の収蔵品を見ずに、この判定をすることは不可能である。

今回の検討では、実際には分類語ではなく、資料名称を分類語彙と見なして検討を行ったため、必ずしも分類語彙としては適切ではないものが出現している可能性は否めない。ただし、各館における分類方法として、オントロジとして厳密な上位・下位の概念に基づく階層化がなされるとは限らず、実際にはそういった分類との対応づけも必要となる可能性がある。

このため、予め完璧な概念辞書を用意しておくというよりも、運用を通じて、実際に検索を行った結果をもとにしたフィードバックを行い、概念辞書との対応づけについて、各館毎に適応していくといった方法も考慮する必要がある。

5. 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、記述内容に踏み込んだ検索時に必要になると考えられる概念辞書の枠組みに関する検討を行った。

概念辞書に要求される表現能力としては

- ・ 類義語の判定に用いることができる
- ・ 上位・下位の概念間の判定に用いることができる

ことがあげられる。

ポータルサイトにおいて、このような概念辞書

を整備していくことにより、各館毎の差異や利用者要求の多様性を考慮した横断検索が行える仕組みを構築できる可能性がある。

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.