

博物館情報横断検索に向けた概念辞書構造の検討

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構造について述べる。

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。

さらに、たとえ共通化ができたとしても、それと利用者側の要求との間のすりあわせも問題になる。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の間の関係情報を、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い。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。

このとき、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。

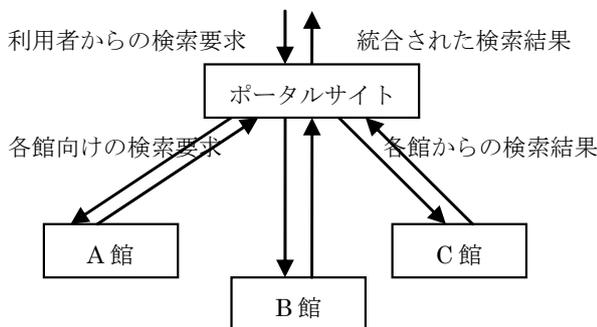


図1 横断検索の構成

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ため、それが何であるかという観点からの検索は難しい[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、分類の仕方やそこで用いられている語彙が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ことにより、意味内容による横断検索の可能性がある[5]。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のもとで、ポータルサイトの動作は以下ようになる。

- (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

[9]では、この関連づけのために概念辞書を用いることが検討されているが、利用者要求と資料名称(分類語)との対応づけについては述べられていない。以下では、利用者要求と資料分類を関連づけるための概念辞書の構造について考察する。

4. 概念辞書の構造

国立歴史民俗博物館の収蔵品データの中から、[9]において検討課題として残されていた「青森県の衣生活用具」コレクション(資料番号 F-7, 数量 1518)について、概念辞書として WordNet [10]を用いて、利用者要求との対応をとることを試みた。

WordNet はプリンストン大学の認知科学研究所で運営されている英語の概念辞書である。synset と呼ばれる同義語を単位として、上位語(hypernym)、下位語(hyponym)の情報をもっているため、[9]で述べられている概念の上位・下位

関係を用いた検索に用いることができる。

実際に行った作業は以下のとおりである。

- (1) WordNet から clothing をルートとする概念階層を抽出し、一意に id をふる。
- (2) 「青森県の衣生活用具」コレクション中の資料名称毎に WordNet の synset id を人手で対応づける。
- (3) 利用者要求として想定される一般的な検索語に対して WordNet の synset id を人手で対応づける。
- (4) 利用者要求を表す検索語から synset id を介して、「青森県の衣生活用具」コレクション中の資料を検索する。

なお、2 と 3 は完全に独立に作業を行った。

WordNet の階層的な id 付けの例を図 2 に示す。

[1] clothing, article of clothing, vesture, wear, wearable, habiliment
...
[1.12] garment
[1.12.1] breechcloth, breechlout, loincloth
...
[1.12.20] overgarment, outer garment
[1.12.20.3] coat
[1.12.20.3.8] jacket
[1.12.20.3.8.1] banyan, banian
[1.12.20.3.8.2] bed jacket
[1.12.20.3.8.3] blazer, sport jacket, sport coat, sports jacket, sports coat
...

図 2 WordNet の例

次に「青森県の衣生活用具」コレクションの概要を示す。

番号	画像	資料名称	コレクション名称
71	無	編笠	青森県の衣生活用具
72	無	編笠	青森県の衣生活用具
73	無	短着	青森県の衣生活用具
74	無	短着	青森県の衣生活用具
75	有	長着	青森県の衣生活用具
76	有	長着	青森県の衣生活用具
77	無	短着	青森県の衣生活用具
78	有	短着	青森県の衣生活用具
79	無	短着	青森県の衣生活用具
80	有	短着	青森県の衣生活用具

図 3 「青森県の衣生活用具」コレクション資料名称として同じものをもつものがあるた

め、資料名称の異なり数は 147 であった。これらの異なり資料名称に対して WordNet の id を対応づける。対応付けの一部を図 4 に示す。

資料名称	WordNet id
風呂敷	対応なし
耳かけ	1.15.9
編笠	1.15.3.5
笠	1.15.3.5
藁ぼっち	対応なし
短着	1.12.24.4
長着	1.12.24.4
着物	1.12.24.4
羽織	1.12.24.4
袖無し	1.12.24.4
短着・股引き	1.12.43
山シャツ	1.4.14
負紐	1.1.1
れんじゃく	1.1
籠用紐	1.1.1

図 4 資料名称と WordNet の対応例

一方、利用者要求で用いられることが想定される一般的な検索語についても同様に WordNet の id を対応づける。この一部を図 5 に示す。

検索語	WordNet との対応
衣類	1
上着	1.12.20.3
シャツ	1.12.31
肌着	1.12.44.11
ズボン	1.12.42
外套	1.12.20.3.7
帽子	1.15.1
襟巻	1.12.28
手袋	1.14.1
靴下	1.11.1.1
足袋	1.11.1.1.5

図 5 検索語と WordNet の対応例

以上の準備のもとで、検索語から WordNet を経由した検索のシミュレーションを行った。

例として、「衣類」という検索語が入力されたとすると、図 5 より対応する synset id として 1 を取得する。このとき、上位概念はすべての下位概念を含むため、1.* という id をもつ synset が検索対象となる。図 4 よりこの形の id に対応する資料名称を取得し、図 3 に示したコレクションから該当する資料名称をもつものを取り出す。この

手法を用いて、「上着」という検索語から、「合羽」「外套」「蓑」「コート」という資料名称が、「手袋」という検索語から、「手袋」「藁手袋」「毛手袋」という資料名称が、「靴下」という検索語から、「足袋」「ぼこ足袋」という資料名称がそれぞれ検索された。

5. 考察

作業中、最大の問題となったのは、日本語の検索語ないし資料名称をどのように WordNet に対応させるかということであった。実際に 147 の異なり資料名称のうち、76 は対応をつけることができなかった。

この中には、用法に関する問題もあった。たとえば、「風呂敷」という資料名称は clothing の体型の中には対応づけることができなかったが、具体的に資料の備考欄を見てみると「ほっかむり」とあり、headscarf の一種と見なせることがわかった。資料名称だけからは、対応が判断できない例である。

また、WordNet の階層が必ずしも日本語の語感とあわないという問題もあった。たとえば、「足袋」が「靴下」の一種であることに違和感を覚えるという意見もあった。

同様に、「かんじき」や「長沓」「下駄」といった資料名称に対しては適切なものがないため footwear という一般的な概念を対応させたが、そうすると、これらの間の区別ができないばかりか、「足袋」も footwear の一種（下位概念）であることから、上位は下位を含むというロジックを適用すると、これらがすべて同時に検索されてしまうという問題が起こった。

今回は英語版の WordNet を改変せずにそのまま用いたため、生じた問題もあるが、概念辞書において、上位・下位の概念間の関係をどのように設計し利用するかは今後も検討が必要である。

また、予め完璧な概念辞書を用意しておくことは困難であるから、運用時には、実際に利用者が検索を行った結果をもとにフィードバックをして、概念辞書との対応づけを漸次更新していくよ

うな仕組みを採用することが望ましいと考えられる。

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、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構造に関する検討を行った。

概念辞書を用いることで、検索語と検索対象の分類語彙との間の対応付けを円滑に行えることがわかったが、上位・下位の関係を注意深く設計しないと予期せぬ検索結果が得られることもわかった。

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を行うために、ポータルサイトにおいて利用可能な概念辞書をどのように整備していくかは、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.
- [10] WordNet (<http://wordnet.princeton.edu/>).