

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

[†] ASTEM RI/Kyoto

[‡] 国立歴史民俗博物館

[‡] National Museum of Japanese History

[§] パナソニックコミュニケーションズ

[§] Panasonic Communications Co., Ltd.

[¶] インフォコム

[¶] 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]。現状のほとんどの館では、それぞれが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[4]において共通索引の試みが行われている。国際的には、大掛かりになる可能性はあるが、CIDOC[5]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。

2. 横断検索のためのフレームワーク

定型的な共通フォーマットですべての情報を表現しようとする情報構造は、博物館情報の多様性を許容した

横断検索を困難にするため、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することはせず、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。これは、知識表現においてオントロジの統合と共有として認識される。

(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量に存在する。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. 情報記述内容レベルの相互変換

3.1 相互変換という考え方

先に述べた3階層のフレームワークでは、各館の独自性、多様性を許容しながら、様々なレベルでの情報の相互変換を行うことにより、複数の館が公開するそれぞれの収蔵品に関する情報(文書)を横断検索し、検索者が望む情報を得ることを可能にすることを目的としている。

情報記述構造レベルの相互変換に基づく共通化では、title, creatorなどの共通データ構造を用いた検索は可能であるが、検索対象となる各要素に書かれている内容については、対象としていない。そこで、たとえ同じ構造に変換できても、

(1) 同じ内容に対して、複数の館で異なる表記をしている場合

(2) 異なる内容に対して、同一の表記をしている場合に対応できない。記述内容の異同の判定には、比較的単純なものから意味内容に関する高度な判断を要するものまでが含まれるが、この問題に対処するためには、情報記述内容レベルの処理が必要となる[5]。

3.2 情報記述内容としての分類

情報記述内容の異同は、情報記述構造の様々なスロットにおいて問題となるが、ここでは特に横断検索において問題となる可能性の高い分類名に焦点を絞る。ユースケースとして、複数の館に対して、ある分類に属するものを横断検索するというタスクを設定した場合に、各館がそれぞれの収蔵品に対して、独自に分類し、それを記述しているという状況が想定される。

分類という行為には、行為者の思惟が反映される。物理的に同じものであっても、観点が異なれば分類の仕方は異なり得る。また、館の性質によっても分類の仕方は違ってくる可能性がある。さらに分類結果に対して付与される分類名の付け方も一様ではない可能性がある。これは、

- i. 分類されたクラスの一致・不一致
- ii. クラスにつけられた名称の一致・不一致

の二つの観点から整理することができる。iについて、ある館の分類の結果、生成されたクラスと完全に一致するクラスが別の館の分類結果に存在する保証はない。またiiについて、たとえ類似のクラスが存在しても、それらに同じ名称が付与されているとは限らない。

このような状況の下で、異なる館の情報を横断的に検索しようとするれば、分類体系間の対応関係をとらなければならない。

情報記述内容レベルのプロトタイピングを扱った文献[6]は、分類語彙表[7]による分類体系を共通構造とし、

それへの各館の独自の分類体系のマッピングに基づく横断検索を確認している。

もっと厳密なクラス定義、等価なクラスの発見、属性値に基づくインスタンスとクラスとの関連付けなどは、オントロジ技術を用いることによって可能になる。

4. 分類体系間の対応付け

4.1 分類体系に構造がない場合

一番単純な分類は、分類のための複数のクラスを準備し、各インスタンスがその中のどれに該当するかを記述するものである。この分類体系には構造がなく、フラットにクラスが並んでいるのみで、クラス間の関係は記述されない。

このような分類体系が複数あった場合に、それらの間の対応関係をとることを考えると、一方のクラスが他方のどのクラス(の集まり)に相当するかを記述することになる。

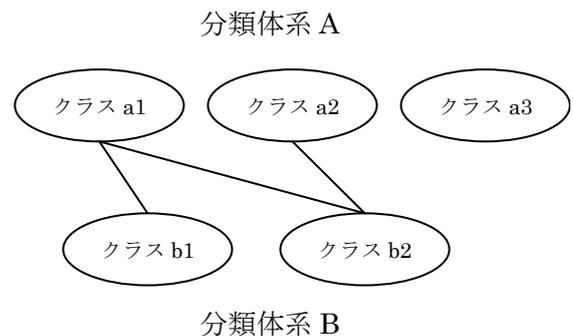


図 1 構造のない分類体系間の対応

このとき、一対一対応でない部分については、不完全性の問題が生じる。たとえば、図1で分類体系Bのクラスb1に対応するクラスとして、分類体系Aではクラスa1を得ることになるが、a1には分類体系Bのクラスb2に相当するインスタンスも含まれてしまっている。

4.2 分類体系に構造がある場合

一つの分類体系の中で、クラスどうしの関係を記述することにより、分類体系に構造を持たせ、この構造を用いて、異なる体系間の対応付けを行うことが考えられる。

分類体系内でのクラス間の関係として典型的なものとして、クラス-下位クラスの関係に基づくクラス階層の構成(is-a階層)がある。このとき、下位クラスに属するインスタンスは同時にその上位クラスのインスタンスでもある。

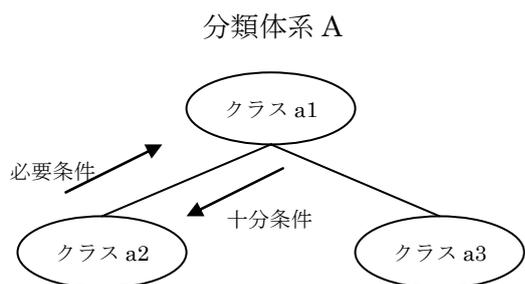


図 2 分類体系の階層化

たとえば、図 2 において、クラス a2 がクラス a1 の下位クラスであった場合に、クラス a1 のインスタンスであることはクラス a2 のインスタンスであることの必要条件となり、逆は十分条件となる。

今、検索結果集合を M, 必要とされる集合を N としたとき、

$$\text{再現率} = (N \cap M) / N$$

$$\text{適合率} = (N \cap M) / M$$

である。このとき、必要条件だけで検索すれば、再現率は最大になるが、適合率は低くなる。逆に、十分条件だけで検索すれば、適合率は最大になるが、再現率は低くなる。このことから、分類階層を検索にどのように利用するかは、目的に応じて使い分ける必要があることがわかる。

4.3 特性の利用

分類階層を利用したとしても、対応関係はクラス単位でしかとれず、他の分類体系において独立したクラスとして表現されていないクラスは、完全な形では表現することができない。

そこで、個々のインスタンスが持つ特性を利用することにより、その値の制限により匿名のクラスを構成して、対応をとることが考えられる。

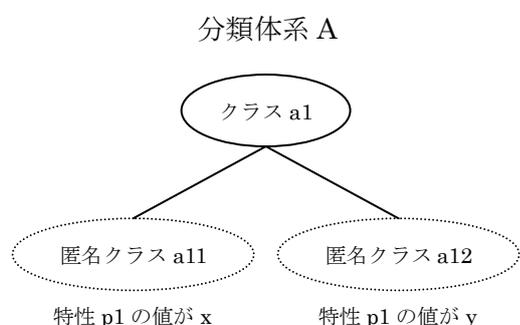


図 3 特性の利用によるクラスの構成

たとえば、図 3 において、分類体系 A ではクラス a1

はそれ以上細分化されていなかったが、a1 に属するインスタンスのもつ特性 p1 の値に注目して、分類体系 A には元々存在していなかった匿名クラス a11 と a12 を構成している。

5. オントロジの利用

5.1 オントロジ記述言語による表現

上に述べた分類体系およびそれらの間の対応付けをウェブオントロジ言語 OWL[8]を用いて表現することを考える。

分類体系に構造がない場合は、単に名前付きクラスの記述が並ぶだけである。二つの分類体系間の対応関係は、ある体系のクラスないしクラスの和集合または積集合と別の体系のクラスを等価クラス公理で結ぶことにより表現する。

分類体系に構造がある場合は、分類階層を下位クラス公理を使って表現する。異なる分類体系間のクラスの対応関係を等価クラス公理で結ぶ点では同じである。

さらに特性を用いる場合は、特性を定義し、特性の制限によるクラス記述を用いて、匿名クラスを記述することにより、元々の分類体系には存在しなかったクラスを構成して、それを等価クラス公理で結びつけることができる。

5.2 OWL を用いた記述例

OWL を用いた記述を作成するために、[6]と同じく食器を題材にした仮想的な三つの博物館を定義し、それぞれの収蔵品に関するメタデータを記述した。各館の分類体系の OWL 記述を図 4 に示す。また、これで表現されているクラス階層を可視化したものを図 5 に示す。

分類自体に注目すると、扱っている対象が共通であっても、それぞれの分類基準や命名法は異なっていることが見て取れる。たとえば、A 館のクラス「皿」は、B 館のクラス「平板食器」に対応すると思われるが、その下位クラスの分け方は異なっている。C 館においては、「皿」に相当する上位のクラスは存在せず、「和食器」「洋食器」「中華食器」のそれぞれの下位クラスの中に散在している。

ここで、A 館のクラス「大皿」に対応するクラスを B 館、C 館から見いだすことを考えてみる。B 館では「大型平板食器」がこれにあたりと考えられる。C 館では「和皿」「洋皿」「中華皿」はこれに相当すると思われるが、「中華皿」については判断ができない。そこで大きさに関わる特性の値を利用して、「中華皿」の下に匿名のクラスを構成し、これを A 館のクラス「大皿」に対応させるという方法をとる。B 館の「大型平板食器」や、C 館の「和

大皿」「洋大皿」についても、必要ならば特性による制限を追加してもよい。

5.3 実現に向けた考察

最後に、以上述べてきた分類体系間の対応付けをどのように実現して、横断検索において利用するかについて述べる。

まず、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表示することになる。この関係を図 6 に示す。

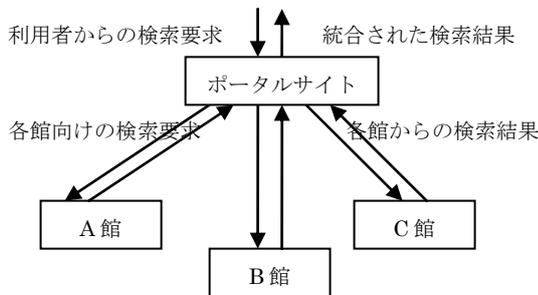


図 6 横断検索の構成

このとき、オントロジ間のマッピングは、理論的にはポータルサイトと各館の間のどこで行われてもよい。ここで仮に、各館に対しては各館毎の語彙を使って検索要求を出すことにすると、対応付けと変換作業はポータルサイト内で実行されることになる。このとき、OWL のウェブオントロジとしての分散オントロジ管理機構が意味をもつことになる。

先に示したように、各館で採用している分類体系の OWL 記述は各館毎に作成される。これをポータルサイトでひとつに統合することになるが、このときにウェブオントロジの取り込み機構を利用する。具体的には、各館毎に記述された OWL 記述を取り込んで、名前空間で区別しながら、ポータルサイトで作成するオントロジの中で対応関係を記述していく。これを図 7 に示す。

ポータルサイトでは、関連づけのためのオントロジを用いて検索要求に対応する各館のクラス記述を求めるとともに、それが各館の DB でどのように格納されているかを情報記述構造レベルで解決し、各館向けの検索要求を作成する。

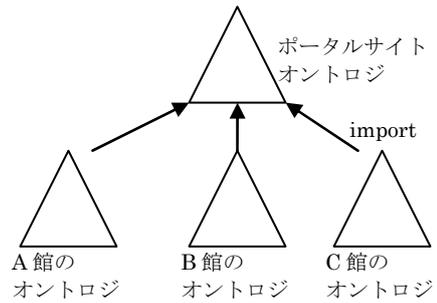


図 7 オントロジの取り込み

このとき、今回は議論の対象外としたが、利用者がポータルサイトで用いる検索用語彙と各館のオントロジの間の相互変換も必要となる。

6. おわりに

本稿では、博物館、美術館が公開する文書情報の横断検索を目指して、情報記述内容レベルでの相互変換に、オントロジを用いる方法について述べた。これにより、各館毎に異なる分類体系を採用していても、それらの間の横断検索をより柔軟に行える仕組みを構築できる可能性がある。今後、より具体的なオントロジの利用方法について検討していく予定である。

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：博物館情報横断検索のための記述内容レベル相互変換，画電年次大会，2003-06.
- [6] 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.
- [7] 分類語彙表，国立国語研究所，1964.
- [8] OWL Web Ontology Language Reference, W3C Rec., 2004-02, (<http://www.w3.org/TR/owl-ref/>).