

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行われている。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にかかわらず、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、さらには複数の館の情報を横断的に検索できることが望ましい[1]。

本稿では、このような情報検索における具体的なユースケースと、このときのオントロジの利用可能性について述べる。

2. 横断検索のためのフレームワーク

現状では、

- 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、たとえフォーマットの共通化はできたとしても、記述内容のレベルになると、共通化は困難である。

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レベルの統一がなければ、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。博物館に所属する専門家・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所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。

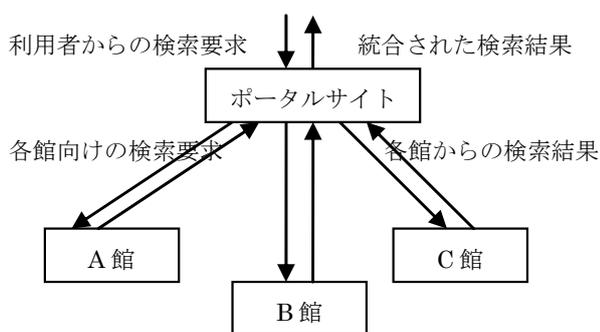


図1 横断検索の構成

4. ユースケースの検討

今回は、実在する博物館データを対象として、実際に想定される利用形態に対して、どのような問題があり、それに対処するためには何を考慮しなければならないかを検討した。例として、国立歴史民俗博物館で作成され、ウェブで公開されている「データベースれきはく」(<http://www.rekihaku.ac.jp/research/database.html>)から「館蔵紀州徳川家伝来楽器データベース」を取り上げ、ユースケースの検討を行った。以下では、単一のデータベースを対象としているが、これは横断検索時にポータルサイトが検索対象とする一つの館に相当する。

「館蔵紀州徳川家伝来楽器データベース」は歴博が所蔵する紀州徳川家伝来楽器コレクションをデータベース化したもので、登録件数は213件である。表1に楽器種毎の登録数を示す。

表1 楽器種別登録数

三管箱	1件	盛管用箱	1件
三管飾袋	2件	心葉	3件
笙箱	2件	笙	9件
袖笙	2件	箏	15件
龍笛	25件	能管	2件
高麗笛	8件	神楽笛	4件
笏拍子	6件	一節切	8件
洞簫	1件	明笛	4件
琵琶	23件	箏	5件
和琴(倭琴)	3件	七絃琴	4件
鼓瑟	1件	一絃琴	2件
琴台	1件	琴案	1件
太鼓	4件	釣太鼓	1件
太鼓撥	1件	鞆鼓	2件
鞆鼓胴	1件	鉦鼓	1件
錦鼓	1件	一鼓	1件
笛筒	2件	箏篋箱	2件
爪袋箱	2件	爪袋	5件
桴	4件	爪	1件
調子笛	1件	簧	1件
律囀	1件	調子笛	2件
調子笛・律管	4件	琵琶撥	1件
琴柱	1件	箏柱	9件
爪形	1件	倭琴柱	1件
楽譜	30件		

登録項目は、【資料番号】、【楽器種】、【資料名】、【資料名かな】、【時代 和暦】、【時代 西暦】、【作者】、【銘文】、【銘書著者】、【下絵等著者】、【数量】、【法量縦(cm)】、【法量横(cm)】、【法量高さ(cm)】、【法量径(cm)】、【品質・形状および加飾】、【伝来】、【修復歴】、【付属品 1～1 1】、【文書数】、【文書 1～1 7】、【備考】であり、対応する【資料画像】も格納されている。このうち、【資料番号】、【楽器種】、【資料名】、【資料名かな】、【時代和暦】、【作者】、【銘文】、【銘書著者】、【下絵等著者】、【品質・形状および加飾】、【伝来】の各項目の内容で検索可能となっている。検索語として入力された文字列が対象となる項目の値に含まれる場合に、検索結果として表示される文字列検索である。フリーワードを指定した場合は、上記検索対象項目のすべてが検索対象となる。

a) 様々な表記が存在するケース

「龍笛」という種類の楽器を検索したいとしよう。この場合、【楽器種】を対象にして、「龍笛」で検索をすると25件ヒットする。ところが、異表記の「竜笛」では0件になる。また、同じく【楽器種】に対して「りゅうてき」で検索しても0件になる。

【資料名かな】に対して「りゅうてき」で検索すると、29件ヒットするが、このうち4件は楽譜のインスタンスである。「龍笛」は「りょうてき」とも読むが、「りょうてき」では0件となる。広辞苑第五版によると「おうてき（横笛）」は「龍笛」の同義語とされるが、「おうてき」で検索をしてもヒットしない。c)に後述するように、フリーワード検索で「横笛」はヒットするが、これは「よこぶえ」のことであって、「龍笛」と同義ではない。

また、フリーワード検索で「龍笛」を検索すると、37件ヒットする。このうち、12件は「龍笛」のインスタンスではない。

<対応策>

これらは、表記の正規化の問題ととらえることができる。「館蔵紀州徳川家伝来楽器データベース」に関する限り、「竜笛」→「龍笛」、「りょうてき」→「りゅうてき」→「龍笛」といった正規化を施したうえで、【楽器種】に対して検索を実行すれば、目的のものが得られる。シソーラスを整備することで、異表記の存在は認識できる。

問題は、異表記中、当該データベースで採用されている表記(代表表記)がデータベース依存である点である。複数の館を対象にして、この問題に対応するためには、当該データベースで採用されている表記一覧(上の例では楽器種リスト)を取得して、シソーラス等を用いて、検索語を表記一覧中の代表表記に対応させるといった操作が必要になる。横断検索においては、これがポータルサイトのタスクになる。

また、検索も単に部分文字列で行ってしまうと、たとえば「爪」に対して「爪袋」もヒットしてしまうため、【楽器種】のような制御されたフィールドの場合は、完全一致のほうが望ましい。

場合によっては、単一データベース内でも、表記の揺れが存在するかもしれないが、これはデータベース内の整備の問題である。

さらに、「横笛」(「おうてき」と「よこぶえ」)のように複数の意味的な可能性がある場合には、そもそも意味的な曖昧性を解決しなければ代表表記への言い換えが

できない。

b) 記述形式に違いがあるケース

江戸時代の楽器を検索したいとしよう。【時代 和暦】を対象として「江戸時代」で検索すると39件ヒットする。一方で、【時代 和暦】項目に「天保13年」といった記述がなされているデータがあり、これは先の検索ではヒットしない。

また、実際には【時代 西暦】では検索できないが、何年から何年までといった区間指定や、「19世紀」といった表現での検索も難しい。これは【時代 和暦】でも同様で、「江戸時代中期以降」といった指定では検索できない。

<対応策>

当該データベースで、どのような記述形式が採用されているかはまちまちである。また、上例のように、単一データベース内でも時代表記と年号表記といった粒度の違いが観測されることも少なくない。手順としては、ほぼa)と同様であるが、言語的な言い換え(代表表記への置換)ではなく、ある種の計算が必要になる点が異なる。たとえば、「天保年間」の収蔵品を検索したい場合、「天保13年」はこれに含まれるが、「江戸時代」は逆に「天保年間」を含む。また、「江戸時代前期」は「天保年間」を含まないといった関係があることを判定する必要がある。

c) 概念階層が必要となるケース

専門家でない一般のユーザが、「笛」を検索したいとしよう。

考えられる一つの方法として、【楽器種】に対して「笛」で検索すると49件ヒットする。内訳は「龍笛」、「高麗笛」、「神楽笛」、「明笛」、「笛筒」、「調子笛」、「調子笛・律管」のインスタンスで、要するに部分文字列として「笛」を含んでいるものであり、このうち、「笛筒」は明らかに「笛」ではない。

これに対して、フリーワード検索で「笛」を検索すると、90件ヒットする。この場合、「箏篋」、「一節切」などの名称に「笛」という文字を含まない「笛」も入るが、「楽譜」、「琵琶」などのインスタンスもヒットしている。

次に、もう少し絞り込んで、「横笛(よこぶえ)」を検索したいとしよう。【楽器種】に対して「横笛」で検索すると0件だが、フリーワード検索で「横笛」を検索する

と46件ヒットする。これは【品質・形状および加飾】の記述による。実際に「龍笛」、「高麗笛」、「神楽笛」といった「横笛」に関する記述には「横笛」という文字列が含まれていたため、これらはヒットしているが、一方で「楽譜」のインスタンスも検索結果に含まれてしまっている。

また、「吹きもの」や「管楽器」といった検索語では、フリーワード検索でも何もヒットしない。

<対応策>

言葉が表している意味にかかわる問題である。安直な対応としては、「笛」や「横笛」といった想定される検索語に対して、予め【楽器種】との対応表を作成しておくというものがある。検索要求が予め容易に想定されるか、提示したメニューの選択肢の中から選ばせるような場合には、この方法も有効である。

一般的には、その言葉がどのような概念を表しているか、さらに概念間のis-a階層に関する知識が必要になる。

横断検索を考えた場合には、ある館で採用している分類体系が他の館と異なることにより、横断的に検索できないという問題になる。これに対して[6]では、各館の分類オントロジのマッピングという考え方で、オントロジを用いた各館の分類体系の差異の吸収方法について論じている。この概略を以下に示す。

まず、各館の分類体系のオントロジ記述を作成する。これをポータルサイトで、ひとつに統合する際に対応関係を記述して、その差異を吸収する。各館のオントロジは、ウェブオントロジの取り込み機構を利用してポータルサイトで統合する。

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、それが各館のデータベース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[7][8]。

d) 関連による修飾が必要となるケース

何かと「同納」されているものを検索したいとしよう。フリーワード検索で「同納」を検索すると8件ヒットする。これは【備考】の記述によるものである。実際に、どれとどれが同納されているか（関連しているか）は一つ一つのインスタンスにあたる必要がある。また、これは細かな点であるが、【資料番号】H-46-55の【備考】にはH-46-40と同納との記載があるが、H-46-40の【備考】

欄にはそのような記載がなく、【付属品1】（外箱に関する記述）にH-46-55と同納との記載があるため、フリーワード検索でもヒットしなかった。

<対応策>

これもまた、オントロジの利用形態の一つと考えられる。内容的には第3層の情報ナビゲーションレベルで想定している収藏品間の関係の記述にあたる。このとき、関係を記述したオントロジは館外の第三者によって外在化することも考えられる。

外箱等の付属品との関係で調べる場合も基本的に同様であるが、これは登録単位の問題とも関連する。付属品は単独の収藏品（エントリ）としては、データ化されていないため、指定方法に工夫を要する。

5. おわりに

本稿では、博物館、美術館の収藏品の横断検索におけるユースケースの検討を行った。

本稿で述べた処理の切り分けとして、それぞれの館が独自に自館について詳細なオントロジを構築できるとは考えにくい。むしろ、各館は収藏品に関する素性データと、タクソノミ的な大まかなラベルづけを提供し、第三者的なポータルサイトがそれらを元に、横断的なオントロジを構築する方法が現実的ではないかと考えられる。

オントロジなどの知識表現を適切に用いたポータルサイトを構築することにより、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の構築の可能性はある。

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] 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.
- [8] 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.