

博物館情報の知的横断検索のためのフレームワーク

画像電子学会 VMA 研究会 博物館・美術館 DTD-SG

山田 篤 (京都高度技術研究所) 安達 文夫 (国立歴史民俗博物館) 海田 茂 (ネクストソリューション)

今門 政記、河合 正樹 (インフォコム) 小町 祐史 (松下電送システム)

1. はじめに

画像電子学会 VMA 研究会「博物館・美術館文書の文書型定義 SG (DTD-SG)」は、博物館、美術館の文書情報の電子化が推進されている状況の下で、今後それらの情報が広く公開される場合に必要になる共通 DTD の開発を目標として、博物館、美術館の文書情報の構造(体裁上の構造と意味上の構造)を調査研究するために設置された。本稿では、複数の館が公開する文書情報を横断検索するためのフレームワークに関する本研究グループの検討結果について報告する。

2. 横断検索に向けた分析

近年、博物館・美術館情報の電子化が進み、一部の館ではインターネットを通じて館蔵品に関する情報の提供や検索サービスを行っている[†]。現状ではこれらのサービスはそれぞれの館が独自に行っているため、利用者側から見た場合、

- 検索方法が館によってまちまちである
- ある館の中だけで検索することは容易だが、別の館のもつ情報と関連付けて見ることが難しい

といった問題点がある。利用者にとっては、各館の差異にとらわれることなく、シームレスに検索ができることが望ましい。これを横断検索と呼ぶことにすると、これを可能ならしむためには、現状のような各館の自己努力だけでは不十分であり、全体を統合する何らかの仕組みが必要となるが、ここで、その統合の方法が問題となる。

もっとも単純には、すべての館の館蔵品に関する情報を記述するための共通のフォーマットを定義し、各館がそれに従い情報を提供することができれば、それを集中管理するか分散処理するかは純粋に技術的な問題であり、横断検索の基盤は形成されることになる。実際に文化財情報については文化財情報システムフォーラム^[1]において共通索引という試みが行われている。また、国際的には、非常に大掛かりになる可能性はあるが、CIDOC^[2]と呼ばれる取り組みがある。

ところで、世の中に実際に存在する博物館は多種多様である。その扱っている対象や規模も様々であれば、それを

扱う態度も様々である。そもそも命名や分類という行為は私が世界をどのように分節化して考えるかという認知を伴う思惟の表明であり、ア priori に客観的ではありえない。そういう意味で、館による違いは各館の独自性を表しているとも考えられる。

たとえば、ある館では美術品を主な対象としているため、館蔵品のことを作品と呼んでいるのに対し、別の館では史料と呼んでいる。そもそも美術品と史料を同じく扱おうとすると問題があるのであって、特定の分野、たとえば美術品ならば美術品だけに限定していけばよいというアプローチもあろうが、たとえば古書を対象にしても、実際には、その言語的な内容に興味を持つ者もいれば、文字そのものや紙面への割り付け情報に興味をもつもの、紙質などの材質、綴じ方や体裁に興味を持つもの等々である。このように、同一の対象に対してもその興味専門によって違った見方をすることが普通である。この場合、heterogeneous であることが重要なのであって、所蔵者の観点と他者の観点が常に一致するとは限らない。さらに、ある観点で認識した結果をどのように表現するかというレベルでも多様性は生ずる。たとえば、年代を時代名で表すのか、世紀で表記するのか、それとも西暦を用いるのかといった点にも表現者の判断が含まれている。

これは館の多様性のみならず、利用者層の多様性にもつながる問題である。ある検索の仕組みを作るときに、一般にはターゲティングが重要であると考えられているが、では博物館において、専門家向け、一般向け、さらには大人向け、子供向けといったように、それぞれ情報をカスタマイズしていかねばならないのだろうか。

このように考えてみると、これは「多様な情報をいかに共有するか」という一般的な問題であることがわかる。我々はインターネット上で現在実際に公開されている情報を分析し、シンタックスレベルでの共通化は可能だが、セマンティクスレベルでは今後相当の努力が必要となると考えている。

3. 情報共有のためのフレームワーク

様々な多様性に対応した横断検索を考えた場合に定型的なフォーマットですべての情報を表現しようとする monolithic な構造では対応が困難である。このため、我々は共通のシンタックスを用いながら、様々なレベルの多様性を許すような枠組みとして、以下の3階層からなるフレ

[†]デジタルミュージアムといった場合には館蔵品そのもののデジタル化という方向も見逃せないが、本稿では館蔵品そのものではなく、これに関する情報の取り扱いに限定する。また、館側の立場に立った館蔵品の管理情報の取り扱いについても本稿の範囲外とする (Cf. CIMI^[3])。

ームワークを提案する。

- I) 情報構造レベル
- II) 情報記述レベル
- III) 情報ナビゲーションレベル

レベルIでは情報を記述するための構造(テンプレート)の共有と相互変換を行う。レベルIIでは情報の記述内容のボキャブラリの共有と相互変換を行う。そしてレベルIIIではナビゲーションのための外在化された関係情報(リンク)の共有と相互変換を行う。以下、順に説明する。

3.1. 構造の共有

このレベルでは、館藏品(オブジェクト)に関する情報をどのように記述するかについて、記述の構造レベルでの共通化をはかる。いわゆる DTD レベルの共有である。ただし、これは各館毎(インハウス)の独自の DTD を否定するものではない。すなわち、標準的な唯一の構造を強制するのではなく、各館の保有する情報がマッピング可能な構造を探ることになる。このマッピングは XSLT^[4]等による構造変換により実現可能である。また、2館毎に変換規則を作成するのではなく、共通構造を中心としたスター型の構成をとることにより、効率化をはかる。

このレベルでの共有を実現することで、たとえばデータベースのスキーマレベルで、A館のxというスロットはB館のyというスロットに対応するといった対応をとることができるようになる。

3.2. ボキャブラリの共有

構造レベルでの共有が実現したとしても、そこに書かれている値について、同じ内容が異なる表現がなされていれば、単純な文字列操作ではその同一性は判定できない。そこで次のレベルとして、記述するボキャブラリの共有及び相互変換をはかる必要がある。これは知識表現の分野では、一般にオントロジーの共有と統合の問題として認識されているものである。ここでも全体で唯一のオントロジーを用いるのではなく、最低限互いに情報を交換するための相互変換を実現するための仕組みを導入することが望ましい。オントロジーについては理論的な検討から、表現方法についてもフレームベースの DAML+OIL^[5]等、様々な研究がなされており、ここではその必要性を指摘するにとどめたい。

なお、日本語に固有の問題として表記とその読みの対応という問題があるが、それもこのレベルに含める[‡]。

3.3. 関係情報の共有

このレベルではオブジェクト間の関係情報を記述する。これは形態としては他のオブジェクトへのリンクという形をとる。たとえば自ら執筆した解説文からオブジェクトへのリンク等である。ここでオブジェクト自身に関する情

[‡]固有名詞とその読み等は構造レベルに組み込んでよいだろうが、子供による検索を考えると、どんな易しい単語でもひらがなでもひけることが必要である。

報と関係情報(リレーション)を区別している。

博物館がもつ情報のうち、専門家・学芸員等が持つ知識はこのレベルに属する。オブジェクトを特定の観点から分類したり、このように見て欲しいといった情報は、利用者をナビゲートするためのシナリオ記述としてここで記述する。これは必ずしも専門家のみが作成する必要はなく、小学生向けの教材としてのナビゲーションや個人的な覚書としての記述等、多種多様なものが存在するだろう。技術的には、リンクそのものについては行外リンクとして URI ベースの XLink^[6]を用いることができるが、シナリオ記述については別途検討が必要となる。

4. おわりに

本稿では様々な多様性に対応するための博物館情報の横断検索のためのフレームワークに関する検討結果について述べた。

素材としてのレベルIができれば、キーワードによるグループ化、クラスタ化による総覧の作成は容易にできる。そこから先はセマンティクスにかかわる問題である。

レベルIIIのナビゲーションを作成する際にも、まず検索を行うところから始まるだろう。

最後にレベルIIIの情報は現在のインターネットと同じく非常に多様なものになることが予想される。場合によっては専門家の間でも意見の相違が見られることがあるだろう。このとき、複数の知識が矛盾なく結びつく(結合)ことは可能かという問題が起こる。両立可能な情報ばかりとは限らない場合に、互いに矛盾するかもしれない、分散して存在する知識を前に、それらを単純に結合するのではなく、アラインメントをとる、さらには必要に応じて変換、変形して自らに必要な情報を適切に発見し利用するための枠組みを考えることは、人間同士のインタラクションや、発見、関係構築へとつながり、非常に興味深い課題である。

そのためには、まずレベルIから順にプロトタイプによる実証を行うとともに、表現されていない情報は検索できないという点で、情報を表現することのメリットを訴えながら、本フレームワークの実現に向けた検討を進めていきたい。

参考文献

- [1] 文化財情報システムフォーラム
(<http://www.tnm.go.jp/bnca/>)
- [2] CIDOC (<http://www.cidoc.icom.org/>)
- [3] CIMI (<http://www.cimi.org>)
- [4] XSLT (<http://www.w3.org/TR/xslt>)
- [5] DAML+OIL
(<http://www.w3.org/TR/daml+oil-reference>)
- [6] XLink (<http://www.w3.org/TR/xlink/>)