

博物館情報探索における到達容易性向上のための資料群分割の効果

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[§]

Atsushi YAMADA[†] Fumio ADACHI[‡] Yushi KOMACHI[§]

[†] 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

[§] 大阪工業大学

[§] Osaka Institute of Technology

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

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、利用者の検索要求から関連する収蔵品を見つけ出す際に、個々の収蔵品に対して付与されている資料名称をもとに計算される関連語を利用した探索を行う場合の到達容易性を向上させるために資料群を分割することの効果について述べる。

2. 横断検索のためのフレームワーク

収蔵品の横断検索のためには、個々の館が持つ情報を統合する仕組みが必要となる[2]。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

(2) 情報記述内容レベル

(3) 情報ナビゲーションレベル

このうち、(1)については、様々な取り組みがなされており、記述スキームの共有や標準化といった試みもある[3][4]。

これに対して、内容レベルの情報共有については、様々な収蔵品を扱うという事情から困難な問題として残っている。対象物が基本的に物であることから、図書のような全文検索といった手法も適用できない。オントロジ[6]や概念辞書[7]を用いた手法も検討されているが、収蔵品に関して様々なメタデータを付与することは博物館にとっても大きな負担となる。

このため、個々の博物館がなるべく簡単に用意できる情報をもとにして、内容レベルの情報共有を行い、横断検索を可能にすることが望まれる。

3. 資料群と資料名称

収蔵品に関して、博物館になるべく負担をかけずに収集可能なメタデータとして、資料群と資料名称を考える。

博物館では、様々な理由によって収蔵品を群として管理していることがある。このようにグループ化された資料の集まりを資料群と呼ぶことにする。資料群はさらに多階層に階層化されて管理されていることもある。[8]では資料群を分類と見なしているが、厳密な分類ではなくとも、グループ化された同じ資料群に属する収蔵品の間には何らかの関連があると考えられる。現状で、資料群の構成方法について、統一された基準は見あたらないが、大量の資料群を集めてくることによって、互いに関連する可能性の高い収蔵品を見いだす手がかりが得られるのではないかと考えている。

一方、個々の収蔵品には名称（資料名称）が付与されている。これと先の資料群を組み合わせると、資料群によってグループ化された資料名称の集まりができる。

[9]では資料群と資料名称との関係を、文書検索と同様の手法を用いて取り扱うことを提案している。また、[10]では、この考え方にに基づき、実データに対して、資料名称の形態素解析を行い、資料名称中の主要語を抽出して、関連語の計算を行っている。さらに[11]では、関連語の順位を用いた到達容易性の評価方法について示している。

本報告では、資料群を分割することで到達容易性がどのように変化するかについて実験を行い、その効果について述べる。

4. 資料群の分割による関連語の違い

はじめに、[11]と同じく、国立歴史民俗博物館の収蔵品データから、考古資料 21,935 点の資料名称を対象とし、主要語は人手によって付与した。また、一つの資料名称に対して、複数の主要語を割り当てることを許した。主要語の異なり語彙数は 751 であった。

次に、

1. コレクションを資料群とする
2. コレクションの階層構造によって資料群を分割する
3. コレクションの分割位置を人手で与える

の三つの方法で資料群を設定し、[10]と同様の方法で関連語の計算を行った。

一つめの、コレクションを資料群とする方法は、[11]と同様である。633 のコレクションをそのまま資料群として設定した。このとき、「土器」と関連があると計算された語彙は 353 個であった。このうち上位 20 位までのものを図 1 に示す。

二つめの、階層構造によるコレクションの分割は、多くのコレクションが

- A-1 落合計策考古資料コレクション
 - A-1-1 茂辺地遺跡出土品
 - A-1-1-1 注口縄文式土器

のように、階層構造を持っていることに注目し、階層構造で共通の親を持ち兄弟関係にある資料によって資料群を構築する方法である。たとえば、上の例では、A-1-1-1, A-1-1-2, ... が一つの資料群として設定される。このとき、兄弟関係にある資料が存在しない場合は資料群を構成しない。この結果、1001 個の資料群に分割された。このとき、「土器」と関連があると計算された語彙は 291 個となった。このうち上位 20 位までのものを図 2 に示す。

三つめの、人手によるコレクションの分割は、上記二つの方法の中間で、コレクションを分割するか否か、分割するとすればその位置を人手によって設定して資料群を構成する方法である。結果として 1032 個の資料群に分割されたが、この中には単独の資料からなり、兄弟関係にある資料が存在しないものも含まれる。このとき、「土器」と関連があると計算された語彙は 277 個となった。このう

ち上位 20 位までのものを図 3 に示す。

関連語	関連度	順位
鉢	8.03913828624174	1
深鉢	7.86422961961796	2
石鉢	7.83571978935358	3
縄文式土器	7.47299868019516	4
敲石	7.38722016253306	5
土器	7.2092556540883	6
磨石	6.91231106632793	7
石錘	6.60617903218393	8
石斧	6.53347584629854	9
壺	6.38438041261366	10
尖頭器	6.26528183115589	11
浅鉢	6.23886951547771	12
石器	6.00940483766365	13
礫	5.97652008832945	14
石匕	5.88919695337085	15
石錐	5.73358667002854	16
剥片	5.66514528536452	17
石皿	5.57998733397973	18
斧	5.56782684783571	19
石槍	5.36499175059727	20

図 1 コレクションで資料群を設定した場合の「土器」の関連語

関連語	関連度	順位
深鉢	10.399308068734	1
土器	10.0849511069683	2
鉢	9.44369574700631	3
縄文式土器	9.41568418968622	4
壺	7.26668895273861	5
浅鉢	6.54921114305041	6
磨石	5.8666578219108	7
尖頭器	4.96492710345645	8
凹石	4.8950389760587	9
石匕	4.76935571314622	10
石錘	4.72496044662438	11
敲石	4.68730544067559	12
香炉	4.58133608242022	13
礫	4.50594840818759	14
石篋	4.48650164496014	15
石斧	4.28139996064648	16
石皿	3.88162237652794	17
壺	3.79629638680769	18
斧	3.73769583349627	19
土偶	3.73768152061492	20

図 2 コレクションの階層構造によって資料群を設定した場合の「土器」の関連語

関連語	関連度	順位
土器	10.589237616961	1
深鉢	10.4654286968898	2
鉢	10.0537858340324	3
縄文式土器	9.19309312173201	4
浅鉢	7.44322338613634	5
壺	6.8611030441309	6
磨石	6.52230676828483	7
敲石	5.71627771615203	8
凹石	5.5997226451075	9
石斧	4.95443605474348	10
石錘	4.86062963933407	11
石匕	4.81294780147157	12
香炉	4.78051939845698	13
石皿	4.62999857406622	14
石鏃	4.57495559946432	15
磔	4.54405360462472	16
尖頭器	4.33311119929153	17
斧	4.30891275386338	18
石器	4.24883234402577	19
皿	4.2331746688721	20

図 3 人手によってコレクションを分割し資料群を設定した場合の「土器」の関連語

これらを比べてみると、資料群を分割することによって、計算される関連語に違いが生じていることがわかる。このとき、多数の資料を含む資料群の場合に、階層構造によって、

- ・より関連性の高い資料が兄弟関係になっている
- ・同じコレクション内でも、関連性の薄い資料は異なる枝に分かれる

という性質をもつとすると、適切にコレクションを分割し、資料群を設定することによって、適切な関連語が抽出されて、間接的にも関連する語への到達容易性が向上することが期待される。

5. 到達容易性への影響

そこで、[11]と同様に、ある語を入力した場合に、その関連語を関連度の高いものから順に提示することを考え、先頭の語からはじめて、一つ下位の語をみる確率を p とし、 n 番目の語に到達する確率を p^{n-1} 、ある語の関連語を見に行く確率を q (ただし、 $p+q<1$) として、それぞれの資料群設定における到達容易性を評価してみる。 $p=0.89$, $q=p^{20}$ として、関連語は第 60 位まで、別の関連語を辿る回数は 4 回までに制限して、到達可能な全経路の計算を行った。三つの資料群の設定のそれぞれについて、例として、「土器」から「石鏃」に至る経路のうち、上位のものを以下に示す。また、このとき、全経路の確率の和はそれぞれ

1. コレクションをそのまま資料群とした場合：0.168589930619463
2. 階層構造で資料群を分割した場合：0.047174548107095
3. 人手で分割位置を設定した場合：0.0744339146642573

であった。次に、「石器」から「石鏃」に至る全経路の確率の和はそれぞれ

1. コレクションをそのまま資料群とした場合：0.163546525303156
2. 階層構造で資料群を分割した場合：0.132500091056194
3. 人手で分割位置を設定した場合：0.164118432685646

であった。

確率	経路 (順位)
0.0770158558936133	石鏃(3)
0.00666454163935484	鉢(1) 石鏃(4)
0.00418148277690937	深鉢(2) 石鏃(7)
0.00418148277690937	敲石(5) 石鏃(4)
0.00418148277690937	土器(6) 石鏃(3)
0.00331215250758991	縄文式土器(4) 石鏃(7)
0.00294781573175502	石斧(9) 石鏃(3)
0.00233496484112315	磨石(7) 石鏃(7)
0.00233496484112315	石錘(8) 石鏃(6)
0.00184952565065365	尖頭器(11) 石鏃(5)
0.00184952565065365	石器(13) 石鏃(3)
0.00164607782908175	石匕(15) 石鏃(2)
0.00130385824841565	石錐(16) 石鏃(3)
0.00103278611857004	壺(10) 石鏃(11)
0.00103278611857004	礫(14) 石鏃(7)

図 4 コレクションで資料群を設定した場合の「土器」から「石鏃」へ至る経路

確率	経路 (順位)
0.00945366624540205	石鏃(21)
0.0020781187085996	尖頭器(8) 石鏃(7)
0.00184952565065365	石匕(10) 石鏃(6)
0.00130385824841565	磨石(7) 石鏃(12)
0.00116043384108993	深鉢(1) 石鏃(19)
0.000919179645527333	石錘(11) 石鏃(11)
0.000818069884519326	土器(2) 石鏃(21)
0.000728082197222201	鉢(3) 石鏃(21)
0.000728082197222201	縄文式土器(4) 石鏃(20)
0.000647993155527758	礫(14) 石鏃(11)
0.000647993155527758	石篋(15) 石鏃(10)
0.000647993155527758	石錐(23) 石鏃(2)
0.000576713908419705	石斧(16) 石鏃(10)
0.000513275378493538	凹石(9) 石鏃(18)
0.000513275378493538	石器(24) 石鏃(3)

図 5 コレクションの階層構造によって資料群を設定した場合の「土器」から「石鏃」へ至る経路

確率	経路 (順位)
0.0190221773280876	石鏃(15)
0.00233496484112315	石匕(12) 石鏃(2)
0.00184952565065365	土器(1) 石鏃(15)
0.00184952565065365	縄文式土器(4) 石鏃(12)
0.00184952565065365	磨石(7) 石鏃(9)
0.00184952565065365	敲石(8) 石鏃(8)
0.00184952565065365	石斧(10) 石鏃(6)
0.00164607782908175	深鉢(2) 石鏃(15)
0.00164607782908175	石錘(11) 石鏃(6)
0.00116043384108993	鉢(3) 石鏃(17)
0.00103278611857004	石器(19) 石鏃(2)
0.000919179645527333	尖頭器(17) 石鏃(5)
0.000728082197222201	凹石(9) 石鏃(15)
0.000728082197222201	礫(16) 石鏃(8)
0.000728082197222201	斧(18) 石鏃(6)

図 6 人手によってコレクションを分割し資料群を設定した場合の「土器」から「石鏃」へ至る経路

この結果を見ると、資料群を分割することによって、「土器」から「石鏃」への到達容易性は減少する一方で、「石器」から「石鏃」への到達容易性は分割による影響をほとんど受けていない。分割により、意味的により近いものへの到達容易性が相対的に上がったと考えられる。階層構造による一律の分割と人手による分割を比べてみると、階層構造による分割では、「石器」から「石鏃」への到達容易性も若干減少している。これは過分割の影響ではないかと考えられる。

2. 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、資料名称と資料群に基づき計算された関連語を用いてある語から別の語へ辿っていく場合に、到達容易性を向上させるために資料群を分割することの効果について述べた。資料群の分割を適切に設定することで、より関連性の高い語への到達容易性が向上することが確かめられた。どのようにすれば、最適な資料群の分割が得られるかは今後の検討課題である。

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.
- [8] 山田篤，他：部分的分類知識の統合による博物館情報の横断検索の提案，画像ミュージアム研究会，2008-02.
- [9] 山田篤，他：博物館資料群中の語の共起関係を用いた関連語抽出，画像ミュージアム研究会，2009-03.
- [10] 山田篤，他：博物館資料群中の語の共起関係を用いた関連語抽出における主要語選定の効果，画電年次大会，2009-05.
- [11] 山田篤，他：博物館情報探索における関連語の順位を考慮した到達容易性の評価，画電年次大会，2010-06.