

博物館情報探索における資料群分割の到達容易性による評価

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と、収蔵品をグループ化した資料群をもとに計算される関連語を用いて探索を行う場合に、到達容易性を比較することによる資料群の分割の評価について述べる。

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グループ分け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、[12]では、資料群の分割の仕方を変えることにより、到達容易性が変化することを示している。

本報告では、専門家の知見をもとに到達容易性を比較することによって、資料群の分割を評価する方法について述べる。

4. 関連語の計算とゴールの設定

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。そして、[12]と同様、

分割 1：コレクションを資料群とする

分割 2：コレクションの階層構造によって資料群を分割する

分割 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。分割の仕方を変えることにより、導かれる関連語も変化する。

次に、望ましい検索結果を設定するために、「石器」、「石斧」、「土器」、「土偶」のそれぞれの語の関連語群を専門家に提示し、それぞれが関連語としてふさわしいかどうかを○×で判定してもらった。「土器」の関連語の判定結果の一部を図 1 に示す。×が付与された語は、専門家にとっては、「土器」との関連がないと判定された語であるため、「土器」から出発した場合に、到達容易性が低くなることが望ましいものであることが想定される。

関連語	順位	○×
鉢	1	
深鉢	2	
石鏝	3	×
縄文式土器	4	
敲石	5	×
土器	6	
磨石	7	×
石錘	8	×
石斧	9	×
壺	10	
尖頭器	11	×
浅鉢	12	
石器	13	×
礫	14	×
石ヒ	15	×
石錐	16	×
剥片	17	×
石皿	18	×
斧	19	×
石槍	20	×

図 1 専門家による「土器」の関連語の判定結果

5. 到達容易性による評価

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の計算を行った。 $p=0.89$, $q=p^{20}$ 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、「土器」から「石鏝」に至る到達可能な全経路の計算を行い、全経路の確率の和を求めたところ、

分割 1 (コレクションをそのまま資料群とした場合) : 0.168589930619463

分割 2 (階層構造で資料群を分割した場合) : 0.047174548107095

分割 3 (人手で分割位置を設定した場合) : 0.0744339146642573

となった。

専門家によって、「石鏝」は「土器」の関連語としてはふさわしくないと判定されているため、「土器」から「石鏝」へ至る到達可能性は低い方が望ましいとすると、望ましい分割は順に、分割 2 > 分割 3 > 分割 1 となる。

次に、「土器」の関連語としては比較的下位で、専門家が関連語としてふさわしいと判定しているものの中から「高杯」を選び、「土器」から「高杯」に至る到達可能な全経路の計算を行い、全経路の確率の和を求めたところ、

分割 1 (コレクションをそのまま資料群とした場合) : 0.00049785187787727

分割 2 (階層構造で資料群を分割した場合) : 0.00142262644849791

分割 3 (人手で分割位置を設定した場合) : 0.00203064928687728

となった。こちらは逆に、到達容易性が高いほど望ましいとすると、望ましい分割は順に、分割 3 > 分割 2 > 分割 1 となる。ただし、異なる分割間で、到達容易性の絶対値で比較することには問題もある。

そこで、様々な語から出発して「高杯」に至る到達容易性を 3 種の分割のそれぞれについて求め、同じ分割の中でそれらの比較を試みる。出発語として「石器」、「土器」、「土偶」、「埴輪」、「銅鐸」、「銅剣」、「銅鏡」を設定した場合の計算結果を図 2 に示す。

	分割 1	分割 2	分割 3
石器	0.000194056	0.000369437	0.000549275
土器	0.000497852	0.001422626	0.002030649
土偶	0.000345532	0.000186265	0.000621021
埴輪	9.90E-03	0.015456021	0.028347691
銅鐸	0.001193007	0.000708322	0.001994376
銅剣	0.006404267	0.000203992	0.008341554
銅鏡	0.002636213	0.003337144	0.009224251

図 2 様々な語から「高杯」への到達容易性の分割による違い

それぞれの分割において、到達容易性の高い順に並べてみると、

分割 1 : 埴輪 > 銅剣 > 銅鏡 > 銅鐸 > 土器 > 土偶 > 石器

分割 2 : 埴輪 > 銅鏡 > 土器 > 銅鐸 > 石器 > 銅剣 > 土偶

分割 3 : 埴輪 > 銅鏡 > 銅剣 > 土器 > 銅鐸 > 土偶 > 石器

となる。

これらを比べてみると、「土器」の相対的順位が最も高いのは分割 2 であるが、同時に「石器」の順位も高くなっている。専門家により「高杯」は「石器」の関連度としてはふさわしくないと判定されているため、「石器」から出発した場合の到達容易性は低い方が望ましいことを勘案すると、分割 3 のほうがよりよい分割になる。

同様に、「銅戈」に至る到達容易性を 3 種の分割のそれぞれについて求め、「石器」、「土器」、「土偶」、「埴輪」、「銅鐸」、「銅剣」、「銅鏡」から出発した場合の計算結果を図 3 に示す。

	分割 1	分割 2	分割 3
石器	3.24E-07	3.01E-07	7.13E-07
土器	1.86E-07	2.29E-07	4.12E-07
土偶	4.44E-07	4.55E-08	5.89E-07
埴輪	4.31E-06	3.00E-05	4.88E-05
銅鐸	0.102808997	0.102965398	0.104108761
銅剣	0.087795157	0.091310143	0.091240786
銅鏡	0.000912182	0.00119148	0.001086877

図 3 様々な語から「銅戈」への到達容易性の分割による違い

それぞれの分割において、到達容易性の高い順に並べてみると、

分割 1 : 銅鐸 > 銅剣 > 銅鏡 >> 埴輪 > 土偶 > 石器 > 土器
 分割 2 : 銅鐸 > 銅剣 > 銅鏡 >> 埴輪 > 石器 > 土器 > 土偶
 分割 3 : 銅鐸 > 銅剣 > 銅鏡 >> 埴輪 > 石器 > 土偶 > 土器

となり、3種の分割の間で相対順位の違いはほとんど見られない。実際には、いずれの分割においても「埴輪」以下からはほとんど到達する可能性はないと見るべきである。もともと「銅戈」は出現数が少なく、関連語の数も少ないため、このような語に対して分割を変えることの効果は見られなかった。

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、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、専門家の知見をもとに到達容易性を比較することによって資料群の分割を評価する方法について述べた。今回は予め設定した分割の評価を行ったが、この評価方法によって、よりよい分割を導くにはどうすればよいか、専門家の知見でふさわしくないと考えた関連語を除くにはどうすればよいか、といった点について検討することにより、専門家の知見を反映した検索が実現されると考えている。

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.
- [12] 山田篤, 他:博物館情報検索における到達容易性向上のための資料群分割の効果, 画像ミュージアム研究会, 2011-02.