

博物館の展示解説に対する興味の具体化を支援する情報探索方法の検討

森本はるか 渡辺 靖彦 松本浩輝 西村 涼 岡田 至弘

龍谷大学 理工学部 情報メディア学科

t13m076@mail.ryukoku.ac.jp, watanabe@rins.ryukoku.ac.jp, t100440@mail.ryukoku.ac.jp,
r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

1 はじめに

高度情報化社会の到来にともない、IT 技術を積極的に利用して生涯にわたる多様で個性的な学習を支援する環境を整備することの重要性が高まっている。生涯にわたる学習の拠点として、さまざまな知的資源が蓄積されている博物館がある。博物館の展示はその知的資源を社会に還元する仕組みとして重要で、展示内容の理解を促進させるためにガイドシステムが近年積極的に導入され、盛んに利用されている。しかし、博物館で現在用いられているガイドシステムでは、展示の解説はガイドシステムからユーザに一方的に与えられ、ユーザの多様で個性的な興味や関心に対応した情報提供や学習支援は行なわれていない。このため、展示物とガイドシステムの説明に対してせっかく芽生えた興味や関心がモヤモヤした状態から具体化されず、そのまま失われてしまう場合も少なくないと考えられる。そこでわれわれは、言語化されていないモヤモヤした状態の興味や関心をユーザがその場で具体化することを支援して、創造的で個性的な発見に結びつけることができるガイドシステムを作成することをめざしている (図 1)。提案する研究は、生涯にわたる学習を支援する環境の形成に役立つと期待できる。

言語化されていないモヤモヤした状態の興味や関心をユーザがその場で具体化することを支援するためには、ユーザの自由な連想に対応して情報を示すことが重要である。しかし、現在のガイドシステムでは、情報の基本単位がそれぞれの展示についての解説である。これではユーザの自由な連想にもとづく情報探索には対応できないおそれがある。情報の基本単位が大きすぎて、ユーザの知りたい情報がその他の情報の中に埋没してしまい、その情報にたどりつく前にモヤモヤした状態の興味や関心が失われてしまうかもしれない。したがって、情報の基本単位をもっと小さなものにすることが重要である。さらに、それらの情報の基本単位にユーザが簡単にアクセスできるような工夫も重要である。なぜなら、博物館でユーザが携帯できるガイドシステムはユーザビリティが低く、ユーザの入力・操作の負担が大きくなるおそれがあるからである。したがって、ユーザビリティが低い携帯情報端末でもユーザの入力・操作の負担が小さくなるように、モヤモヤした状態の興味や関心を具体化するための情報を指でふれて選択できるなど、直感的な操作による情報探索の工夫が重要である。

本研究では、情報の基本単位を解説文の段落とし、自由な連想にもとづく情報探索が直感的な操作でできる方法について検討する。

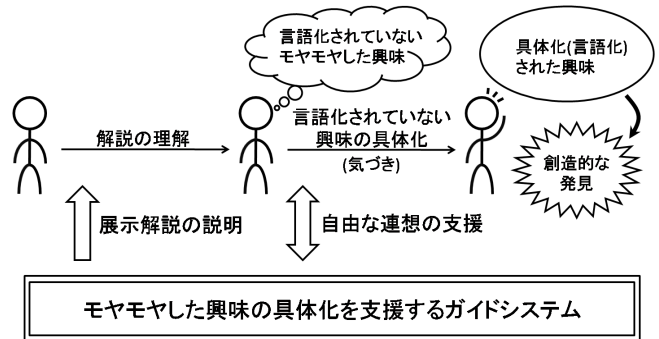


図 1 展示解説に対するモヤモヤした状態の興味の具体化を支援するガイドシステム

2 関連研究

博物館で現在用いられているガイドシステムやこれまでに提案されている知的ガイドシステム [1] [2] では、年齢・言語・履歴などによる情報の個人化は行われているが、展示の解説はガイドシステムからユーザに一方的に与えられ、ユーザの多様で個性的な興味や関心に対応した情報提供や学習支援は行なわれていない。一方、ユーザの興味や関心に対応した文書作成や情報検索などを支援するため、関連語の抽出や辞書の作成、クエリ拡張や検索結果の表示方法などについての研究がさかんに行われている。有田らは単語クラスタリングを行って階層型のシソーラスを構築し、それを利用して類義語ナビゲーション機能を実現した [3]。中田らは、単語から連想語、連想語から文例、文例から言語表現をそれぞれ導き出す言語表現支援システムを提案した [4]。徳田らは文書から抽出した共起データを格納するシソーラス管理システムを開発し、それがユーザの連想支援に有効であることを示した [5]。しかしこれらの研究では、ユーザを支援する場所についての検討が行われていない。また、支援する場所における情報端末のユーザビリティも考慮されていない。

3 モヤモヤした状態の興味や関心を具体化するための情報の抽出

モヤモヤした状態の興味や関心をユーザが具体化するためには、自由な連想にもとづく情報探索ができることが重要である。

現在のガイドシステムでは、情報の基本単位がそれぞれ

坂本龍馬

坂本 龍馬（さかもと りょうま、天保6年11月15日（1836年1月3日）－慶応3年11月15日（1867年12月10日）は、江戸時代末期志士、土佐藩郷士。

諱は直陰（なおかげ）のちに直柔（なおなり）。通称は龍馬。他に才谷 梅太郎（さいたに うめたろう、さいだに うめたろう）などの変名がある（手紙と変名の項を参照）。

土佐郷士に生まれ、脱藩した後は志士として活動し、貿易会社と政治組織を兼ねた亀山社中（後の海援隊）を結成した。薩長同盟の斡旋、大政奉還の成立に尽力するなど倒幕および明治維新に影響を与えた。大政奉還成立の1ヶ月後に近江屋事件で暗殺された。1891年（明治24年）4月8日、正四位を追贈される。

（以下略）

図2 wikipediaの項目の例

の展示についての解説である。これではユーザの自由な連想にもとづく情報探索には対応できないおそれがある。例えば、ユーザの知りたいことが、その展示解説のたった1つの段落でしか説明されていないことがある。この場合、知りたいことが説明されている段落にユーザがたどりつくまでに長い時間がかかったり、たどりつく前にユーザが読むのをあきらめてしまうおそれがある。そこで本研究では、ガイドシステムでの情報の基本単位を展示解説の段落とする。

興味や関心がまだモヤモヤした状態のユーザには、知りたいことに関連するキーワードを具体的に思いつくことはむずかしい。しかし、具体的なキーワードが示されれば、それらの中から選択することはむずかしくない。さらに、それらのキーワードが「だれが」「いつ」「どこで」などの問いと結びつくように分類されていれば、ユーザの負担はさらに小さくなる。そこで本研究では、解説文の段落から、そこで説明に用いられている用語を取り出して分類する方法について検討する。なお、博物館の展示解説を大量に入手することはむずかしい。そこで本研究では、展示解説のかわりに wikipedia の項目を用いる。これは、展示解説に類似する内容の文書を大量にかつ自由に利用できるからである。図2に wikipedia の項目の例を示す。図2に示すように、wikipedia の項目は、見出し語と項目説明文から構成されている。項目説明文の先頭には、その項目の概要が述べられている。

3.1 解説文の段落からの関連する用語の抽出

wikipedia の項目の説明文はハイパーテキストである。ここでは、関連する項目へのリンクを、ウィキリンクとよばれるマークアップを用いて表現している。そこで、このウィキリンクを用いて、項目の説明文の各段落から説明に用いられている用語を取り出す。ただし、用語の文字列の一部が省略されることがあるという問題がある。特に人名を表す用語の場合、姓または名前が省略されることがある。例えば、wikipedia の「坂本龍馬」の項目説明文では、

「近藤長次郎」と「近藤勇」の2人の「近藤」について説明がある。このとき、名前が省略された「近藤」が、「近藤長次郎」と「近藤勇」のどちらを意味するのかを推定して用語の抽出を行わなければ、ユーザの求める情報が説明されている段落にたどりつけない。そこで、wikipedia の項目の説明文から関連する項目の見出し語を以下の方法で取り出す。

(step 1) ウィキリンクでマークアップされた項目の見出し語を対象項目の説明文から取り出す。

(step 2) (step 1) で取り出した見出し語について、ウィキリンクでマークアップされていないものを対象項目の説明文から取り出す。

(step 3) 項目の説明文の最初の段落を調べ、(step 1) で取り出した見出し語を意味的に区切る。例えば、図2の「坂本龍馬」の説明文の最初の段落では、太字で表現された「坂本」と「龍馬」の間に半角空白が用いられていて、ここに意味的な区切りがあることが示されている。この情報を用いて、見出し語を意味的に区切り、その部分文字列を取り出す。

(step 4) (step 3) で区切られた見出し語の部分文字列を対象項目の説明文から取り出す。そして、その部分文字列が表すのは、(step 3) でその部分文字列を取り出した見出し語のうち、その部分文字列よりも前で最後に用いられた見出し語であると推定する。例えば、「坂本龍馬」の説明文では、名前が省略された「近藤」が1回あらわれる。そして、「近藤長次郎」と「近藤勇」のうち、その「近藤」の前で最後に用いられていたのは「近藤長次郎」であった。そこでこの「近藤」は「近藤長次郎」を表していると推定する。なお、部分文字列よりも前で、その部分文字列を取り出すことができる見出し語が1つも用いられていなければ、その部分文字列は取り出さない。

この方法を wikipedia の「坂本龍馬」の項目説明文に適用すると、744個の見出し語を抽出できた。このうち、2回以上用いられていた見出し語は366個である。またこれらの見出し語から取り出した部分文字列、すなわち文字列の一部が省略されている見出し語71個が188回検出された。このうち、170回について、部分文字列が表す見出し語を正しく推定することができた。

3.2 解説文の見出し語の分類

wikipedia の項目説明文から取り出した見出し語が、以下に示す意味を表すかどうかの判定を機械学習で行う。

- 人物
- 時間
- 場所
- 組織
- 作品
- できごと

この意味分類は、ユーザが展示の解説を情報探索する時、「だれが」「いつ」「どこで」などの問いと結びつけて用語を

表1 見出し語の意味についての人手による判定結果

分類	人手による判定結果		
	Yes	No	合計
人物	2118	3751	5869
時間	285	5584	5869
場所	660	5209	5869
組織	463	5406	5869
作品	562	5307	5869
できごと	427	5442	5869

表2 人名項目の判定に用いる素性

番号	素性の種類
s1	項目説明文の概要の形態素 1-gram
s2	項目説明文の概要の形態素 2-gram
s3	項目説明文の概要の形態素の数
s4	項目説明文の概要の最初の文の形態素 1-gram
s5	項目説明文の概要の最初の文の形態素 2-gram
s6	項目説明文の概要の最初の文の形態素の数
s7	項目説明文の概要の最初の文の最後の形態素
s8	項目説明文の概要の文字 1-gram
s9	項目説明文の概要の文字 2-gram
s10	項目説明文の概要の文字 3-gram
s11	項目説明文の概要の長さ
s12	項目説明文の概要の最初の文の文字 1-gram
s13	項目説明文の概要の最初の文の文字 2-gram
s14	項目説明文の概要の最初の文の文字 3-gram
s15	項目説明文の概要の最初の文の長さ
s16	項目見出し語の文字 1-gram
s17	項目見出し語の文字 2-gram
s18	項目見出し語の文字 3-gram
s19	項目見出し語の長さ

選択できるように設定した。図2に示すように、wikipediaの項目は、見出し語と項目説明文から構成されている。項目説明文の先頭には、その項目の概要が述べられている。本研究では、この概要と見出し語を用いて機械学習を行う。

項目の見出し語の分類に用いる実験データを以下の手順で作成した。まず、wikipediaの項目の見出し語を5869件取り出し、(1)人物、(2)時間、(3)場所、(4)組織、(5)作品、(6)できごと、であるかどうかを人手で判定した。人手による判定では、見出し語だけではなく、項目説明文を必ず見て判定した。人手による判定結果の内訳を表1に示す。次に、この5869件の見出し語とその説明文から表2に示す素性を取り出した。s1~s7は項目説明文の概要に対する形態素解析結果を素性として利用している。s8~s10、s12~s14、およびs16~s18は項目説明文の概要および見出し語の文字 n-gram を素性として利用している。文字 3-gram を使用しているのは、小高らが文字 n-gram のうち日本語には 3-gram が効果的であると報告しているからである [6]。s4~s7 および s12~s15 は項目説明文の概

表3 見出し語の意味分類の実験結果

分類	機械学習による結果		
	正解率	適合率	再現率
人物	99.1%	99.0%	98.7%
時間	99.4%	99.6%	87.4%
場所	98.3%	95.5%	89.5%
組織	96.8%	89.9%	67.0%
作品	98.5%	95.4%	88.8%
できごと	98.9%	97.9%	87.4%

要の最初の文についての情報を素性に利用している。

この実験データに対して、SVMによる学習と10分割のクロスバリデーションによる評価を行った。SVMにはTinySVM^{*1}を用いた。また、形態素解析にはJUMAN [7]を用いた。この分類実験の結果を表3に示す。素性選択の実験も行ったが、表2の素性で有効でないものはなかった。

4 モヤモヤした状態の興味の具体化を支援する情報探索システム

博物館の展示に対するモヤモヤした状態の興味や関心を具体化するのを支援するため、以下の点を考慮した情報探索システムをPHPとRDBを用いて作成した。

- 情報探索の基本単位を解説文の段落とする
- 関連する用語を「だれが」「いつ」「どこで」などの問いと結びつくように分類して示し、自由な連想をやすくする
- ユーザビリティが低い情報端末でも、情報へのアクセスの負担を小さくする

このシステムは、PCまたはタブレット端末での使用を想定している。開始画面の入力フォームに展示の名称や番号を入力すると、図3に示すように、以下の3つのモジュールから構成されるインターフェイスが示される。

- 気になるキーワードの表示部
- 検索結果の表示部
- 関連する用語の表示部

ユーザはまず、気になるキーワードをどの用語の分類から選ぶのかを決める。用語の分類は、「だれが」「いつ」「どこで」などの問いと結びつくように「人物」「時間」「場所」などを設定した。図3(a)は、ユーザが関連する用語の分類に「時間」を選んだところである。このとき、関連する用語の表示部には、「坂本龍馬」の項目説明文で用いられ、「時間」に分類された関連用語が示されている。次にユーザは、この関連用語の中から気になるキーワードを選ぶ。すると、そのキーワードは気になるキーワードの表示部で示され、そのキーワードを含む段落が検索結果の表示部に

^{*1} <http://chasen.org/taku/software/TinySVM/>

示される。図 3(b) では、ユーザが選んだ「安政」が気になるキーワードとして表示され、検索結果には「安政」を含む段落が表示されている。従来の全文検索システムでは、ユーザは検索結果を読み、そして追加するキーワードを考えなければならない。そして、関連がないキーワードを追加してしまい、情報探索に失敗するおそれもある。一方、このシステムでは、ユーザは関連する用語の中から気になるキーワードを探せばよい。関連する用語は、気になるキーワードが更新されるたびに、検索結果として出力された段落から取り出されたものに更新される。そして、関連がないキーワードを追加して情報探索に失敗するおそれは、このシステムにはない。図 3(c) は、「場所」に分類されている「江戸」が気になるキーワードに追加されたところである。「時間」に分類されている「安政」というキーワードを見て、「坂本龍馬は安政のころどうしていたのだろう」という興味から情報探索をスタートしたユーザは、その問いに関連する用語として「江戸」を示されて、「そのころ、坂本龍馬は江戸にいたのか」とか「江戸で坂本龍馬はなにをしていたのだろう」と関心を具体化・詳細化していくことができる。

モヤモヤした状態の興味や関心を具体化する情報に、自由な連想にもとづいて、また、ユーザビリティの低い携帯情報端末からでも簡単にアクセスすることができる情報探索システムを作成した。このシステムは従来のガイドシステムとは異なり、展示解説や関連する項目説明文をそれらのテキストの先頭から順に読んだり聞いたりしなくても、モヤモヤした状態の興味や関心を具体化する情報を探することができる。また、ガイドシステムから一方的に情報を与えられていたときに比べ、多様で個性的な興味や関心にもとづいて主体的に博物館の展示やその解説を利用することができると思われる。

参考文献

- [1] 安田, 井上, 岸佳, 長尾: 自動走行可能な移動体によるミュージアムでの鑑賞体験の個人化, 情報処理学会第 72 回全国大会講演論文集, 3ZA-4, (2010).
- [2] 高橋, 益岡, 深谷, 伊藤, 片桐: ubiNEXT: 自由選択学習を支援する展示ガイドシステム, 人工知能学会第 19 回全国大会論文集, 2A3-01, (2005).
- [3] 有田, 菊池, 白井: 検索語の共起情報を利用した単語クラスタリングと Web 検索への応用, 情報処理学会研究報告, 2007-NL-180, pp.115-120, (2007).
- [4] 中田, 葛, 吉村: キーワードおよびその連想語による文例の検索に関する考察, 情報処理学会研究報告, 2004-CH-61, pp.5-12, (2004).
- [5] 徳田, 間瀬, 辻: デジタルドキュメントにおける共起データを用いた検索ターム連想支援について, 情報処理学会研究報告, 97-DD-10, pp.15-22, (1997).
- [6] 小高, 村田, 高, 諏訪, 白井, 高橋, 黒岩, 小倉: n-gram を用いた学生レポート評価手法の提案, 電子情報通信学会論文集, Vol.J86-D-I No.9, pp.702-705, (2003).
- [7] 黒橋, 河原: 日本語形態素解析システム JUMAN version 5.1 使用説明書, 京都大学, (2005).



(a) 「坂本龍馬」の「時間」に関連する用語を表示させる



(b) 気になるキーワードに「安政」を選ぶ



(c) 気になるキーワードに「江戸」を追加する

図 3 モヤモヤした状態の興味の具体化を支援する情報探索システム