

博物館の展示物と見学者の興味を関連付ける情報の発見

楊 澤華[†] 山本 祐輔^{††} 山本 岳洋^{†††} 神門 典子^{††††} 大島 裕明[†]

[†] 兵庫県立大学 応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

^{††} 静岡大学 情報学部 〒432-8011 静岡県浜松市中区城北 3-5-1

^{†††} 京都大学 情報学研究科 〒606-8501 京都府京都市左京区吉田本町 36-1

^{††††} 国立情報学研究所 〒100-0003 東京都千代田区一ツ橋 2-1-2

E-mail: [†]{aa18x510,ohshima}@ai.u-hyogo.ac.jp, ^{††}yamamoto@inf.shizuoka.ac.jp,

^{†††}tyamamot@dl.kuis.kyoto-u.ac.jp, ^{††††}kando@nii.ac.jp

あらまし 本研究では、博物館や美術館の見学者が展示物に興味を持てるようになることを目的として、展示物の情報と、見学者が関心を持つことができる事項をつなぎ合わせる情報を Wikipedia のリンク構造から発見する手法を提案する。現在、いくつかの博物館や美術館では電子ガイドや電子ナビなどの情報端末が館内で利用できるようになっている。そのような端末では、展示物の詳細情報やおすすめの見学経路情報が提供される。紙のガイドや音声ガイドに比べて、情報端末ではより多くの情報が提供されるようになっているが、はたして、鑑賞体験をより記憶に残るものにすることに役立っているだろうか。本研究では、展示物への理解を深めるために、鑑賞者たる「私」と、展示物の間に何らかの関連性が感じられるようにし、見学者に鑑賞において能動性を持たせることを目指す。国立民族学博物館をフィールドとして、展示物のデータベースと Wikipedia などを用いて、すべての展示物と Wikipedia 項目などの世の中のあらゆる事柄に存在する様々な関連性を発見する手法を提案する。

キーワード 博物館鑑賞体験, Wikipedia 解析, パーソナライゼーション

1. はじめに

現在、博物館や美術館では、電子ガイドや電子ナビなどの情報端末が館内で利用できるようになっていることがある。そのような端末では、展示物の詳細情報をテキスト、映像、音声などで閲覧することができたり、おすすめの見学経路情報が提供されたりする。

たとえば、国立民族学博物館では、「みんぱく電子ガイド」^(注1)という端末を無料で借りることが可能である。

図1は、「みんぱく電子ガイド」である。ハードウェアはプレイステーションポータブルである。博物館内の様々な所に、



図1 国立民族学博物館のみんぱく電子ガイド

図2のようなビデオ番号パネルが設置されており、そこで示された3桁の番号を選択することによって、関連する2分程度の映像を閲覧することが可能である。

東京国立博物館では、「トーハクナビ」^(注2)というスマートフォン用アプリケーションが公開されており、見学コースの選択を行うことが可能である。選択された見学コースに従って、博物館内を見学すると、自動的にガイドが行われる。

このような情報端末では、紙のガイドや音声のみのガイドに比べて、より多くの情報が提供されるようになっている。しかし、はたして、博物館や美術館における鑑賞体験をより記憶に残るものにすることに貢献しているだろうか。博物館や美術館において展示物を鑑賞しているときには、より多くの情報に触れることになっていたとしても、しばらくするとほとんど忘れ



図2 国立民族学博物館のビデオ番号パネル

(注1)：国立民族学博物館 みんぱく電子ガイド：

<http://www.minpaku.ac.jp/nmuseum/exhibition/main/digitalguide>

(注2)：東京国立博物館 トーハクナビ：

https://www.tnm.jp/modules/r_free_page/index.php?id=1467

てしまうといったことにはならないだろうか。

本研究では、このような問題意識のもと、博物館や美術館の見学者が展示物に興味を持てるようになることを研究目的と設定する。そして、展示物の情報と、見学者が関心を持つことができる事項をつなぎ合わせる情報を発見する手法を提案する。

博物館や美術館には、展示物に特に強い興味を持っていないような見学者も多く訪れる。そのような人にとっては、能動性を持って鑑賞することは難しい。そこで、展示物への理解を深めるために、鑑賞者たる「私」と、展示物の間に何らかの関連性が感じられるようにし、見学者に鑑賞において能動性を持たせることを目指す。

本研究では、国立民族学博物館を主たるフィールドとして設定する。国立民族学博物館で展示されているあらゆる展示物を対象として、見学者に興味を持ってもらえるように仕向けることを目指す。見学者からは興味を持つことができる事項があらかじめ得られているものとして、国立民族学博物館の展示物とそれらの事項の間に存在する関連性を発見する手法を提案する。具体的には、国立民族学博物館の展示物のデータベースと Wikipedia を用いて、展示物と Wikipedia 項目のリンク構造を抽出し、その中から、提示すれば見学者が興味を持つことができるだろうと考えられるリンクを発見する。項目間のつながりを提示することによって、自分が興味を持っている事項と展示物の間に関連性を感じられることができるようになると考えている。

2. 関連研究

これまで、博物館における展示物の鑑賞インタフェースや鑑賞体験に焦点を当てた研究が提案されてきた。博物館の展示物を鑑賞する際の問題点として、見学者の知識が不十分であり、展示物に関する説明を読んだとしても十分に理解できないという点があげられる。この問題に対して Yamada らは、CEAX Voyager というシステムを提案している [1]。CEAX Voyager は、教育現場において、子どもが文化財について学ぶことを支援することを目的としたシステムである。このシステムは、見学者の知識がなくとも文化財の探索ができるように、画像の類似性で検索したり、見学者が自由に年代や場所と言った文化財を分類する軸を選んだり、多様な観点から自由に文化財を探索することができる。そのシステムでは文化財の解説文を平易な文で表現することで、専門的な知識がなくとも展示物の理解ができるような仕組みが提案されている [2]。

また、博物館での実際の鑑賞行動に関する研究として、ユーザ個人の興味や嗜好に合わせて鑑賞体験を個人化 (personalization) する手法もこれまで議論・提案されてきた [3] [4] [5] [6] [7]。たとえば、Hatala と Wakkary は、音声による解説文を聞きながら展示物を鑑賞するような鑑賞体験において、見学者がどのような音声ガイドを聞いたのかやどの展示物に移動したのかといった情報からユーザの興味を推定し、ユーザの興味を持ちそうな展示物を推薦するシステムを提案している [6]。Bohnert と Ingrid Zukerman [5] はメルボルン博物館において、見学者がどのように展示物から展示物へと移動していったのかの経「路

表 1 国立民族学博物館データベースの項目件数

データベース名	項目件数
標本資料目録データベース	285,135 件
標本資料詳細情報データベース	72,428 件

情報を収集し、収集したデータを基に既存の協調推薦アルゴリズムを用いることで、見学者が興味を持つ展示物を推薦する手法を提案している。

本研究も、上記であげたような博物館の鑑賞体験の個人化に関する研究と捉えることができる。一方、本研究とこうした研究の大きな違いは、これまでの研究の多くが見学者の博物館における行動から見学者の興味を推定し、見学者が興味を持ちそうな展示物を推薦するというアプローチであるのに対して、本研究は、博物館とは独立に見学者がすでに持っている興味（本研究では Wikipedia を利用）と展示物との関連を見つけることで、見学者の興味を喚起するというアプローチであるという点にある。

3. 利用するデータ

本研究では、国立民族学博物館の展示物の情報として、国立民族学博物館が公開しているデータベースを利用する。また、見学者の興味を表したり、見学者に自分の興味と展示物の関連を示すために Wikipedia を利用する。今回は、Wikipedia 日本語版を対象とする。本節では、それぞれのデータについて説明する。

3.1 国立民族学博物館のデータベース

国立民族学博物館のデータベースは標本資料目録データベースや標本資料詳細情報データベースなどから構成されている。これらのデータベースの項目件数について、表 1 に示す。

標本資料目録データベースは、国立民族学博物館の展示物についての概要データが格納されたデータベースである。国立民族学博物館の展示物について、最も網羅されているデータベースであると考えてよい。

標本資料詳細情報データベースは、網羅性は低いものの、展示物の「特徴」や「使用材料」といった、比較的詳細な属性についての情報が格納されているデータベースである。

本研究では、可能な限り多くの展示物を対象とするために、標本資料目録データベースを用いることとする。表 2 は、標本資料目録データベースに存在する属性とその属性値の例を示したものである。これらの属性以外にも、「民族」という属性があり、値を持つ展示物も存在する。また、代表写真やその他の写真があるもの、寸法情報や撮影・計測日などの情報が存在する場合もある。「関連情報」には他のデータベースへのリンク情報が記載されている。このように、必ずしもすべての属性に対して値があるわけではないが、「標本番号」「標本名」「受入年度」「記載日」についてはデータの欠損がないようである。

属性「OWC」と属性「OCM」は Human Relations Area Files ^(注3) による分類である。「OWC」は、Outline of World

(注3) : Human Relations Area Files の略称は HRAF であり「フラーフ」

表 2 標本資料目録データベースの属性

属性	属性値の例
標本番号	H0002381
標本名	カヌー用 櫓 カヌーヨウ カイ
地域	パプアニューギニア ラム川流域<使用>
OWC	OJ33
OCM	501
受入年度	1975
記載日	2018-11-05
関連情報	標本資料記事索引データベース (7 件)

表 3 Wikipedia 日本語版のデータの概要

ファイル説明	データ数
ページ	1,824,962 件
ページ間のリンク	88,276,200 件
カテゴリ数	213,700 件

Cultures という分類方法であり、4 桁のコードで世界の地域を表す。「OCM」は人間の様々な活動の分類方法であり、3 桁の数字で表される。「OWC」と「OCM」については、現在は利用していないが、有用な分類であるため今後利用することを検討している。

3.2 Wikipedia 日本語版のデータ

本研究では、見学者の興味を表すために Wikipedia を利用する。具体的には、あらかじめ、見学者が関心を持つ Wikipedia のカテゴリが与えられていることを前提とする。また、標本資料目録データベースから対象の展示物の情報を取得したのち、それを Wikipedia ページに対応づけるということを行う。そして、それらの間にある「見学者が関心を持つことができる」リンクを発見するというを行う。

このように、主な課題は、Wikipedia におけるリンク解析である。我々が今回利用するのは Wikipedia 日本語版である。公開されているダンプデータから、Wikipedia のグラフを取得する。利用した Wikipedia データの概要を表 3 にまとめる。

4. Wikipedia リンク解析による展示物の関連情報取得

本節では、まず、本研究の問題定義を行う。次に、前処理について説明し、実質的にどのような問題を解くのかということについて述べる。そして、その後、提案手法のアプローチについて説明する。

4.1 問題定義

本研究で取り組む問題の inputs は、以下の 2 つである。

- ある一つの展示物
- 見学者の興味

出力は、入力された展示物と、見学者の興味の間にある関係性を示す情報である。これは、博物館の見学者が、ある一つの展示物を見ているときに、自分の興味と関連づけて展示物を見る

ことができると、より関心を持って鑑賞を行うことができるだろうという考えに基づいたものである。本研究では、この問題を Wikipedia 上のリンクのパス発見の問題としてとらえる。すなわち、

- ある一つの展示物の標本名を表す Wikipedia ページ z
- ある一つの展示物の地域を表す Wikipedia ページ r
- 見学者の興味を表す Wikipedia ページ集合 M

を入力とし、それらの間にある

- Wikipedia 上のリンクパス p

を出力とする問題とする。ただし、出力される Wikipedia 上のリンクパス p は、 z 、または r から、いずれかの $m_j \in M$ への単純道である。そして、 p を見ることによって、見学者が入力された展示物に、より興味を持つことができるようなものでなくてはならないとする。このような Wikipedia 上のリンクのパス発見の問題と考えると、実際の展示物に対して、その展示物を表す Wikipedia ページ集合を取得することと、ある見学者に対して、その見学者の興味を表す Wikipedia ページ集合を取得することが必要となる。次節で、この問題の理想的な出力例について述べた後、必要となる Wikipedia 集合の取得手法について説明する。

4.2 提案アプローチにおける具体例

たとえば、国立民族学博物館には、図 4.2 のような展示物が存在する。これは、パプアニューギニアで使われていた、船を漕ぐための「櫓」である。この展示物は、標本番号「H0002381」であり、標本資料目録データベースにおける情報は、先ほど示した表 2 のとおりである。フリガナを除く標本名は「カヌー用 櫓」であり、地域は「パプアニューギニア ラム川流域<使用>」である。ここでは、この「櫓」を見学者が見ているということ为例として用いて、提案アプローチでどのような情報を発見しようとしているのかを説明する。

また、ここでは見学者は「釣り」に興味を持っているということを表明していることを想定する。

本研究における問題の、理想的な出力の例を図 4.2 を用いて説明する。

対象の展示物のデータベース情報から、「パプアニューギニア」と「櫓」を抽出する。そして、これらのそれぞれの Wikipedia ページが取得できたものとする。見学者は、「釣り」という Wikipedia カテゴリを指定しているとし、それらに関連する



図3 パプアニューギニアの櫓

ページ集合が得られるものとする。たとえば、「ナマズ」や「ウナギ」などの Wikipedia ページが含まれるものとする。

それらの間をつなぐ Wikipedia のリンクで構成されるパスには、たとえば、「パプアニューギニア」-「ヒクイドリ」-「カタツムリ」-「ミミズ」-「ナマズ」というものが存在している。このようなパスからは、以下のような情報を得ることが可能である。

「パプアニューギニア」に生息する「ヒクイドリ」は「カタツムリ」を食べる。ある「カタツムリ」は餌として「ミミズ」を食べる。また、「ミミズ」は「ナマズ」を釣るための餌として使われる。

このような情報が得られたとすると、見学者が目の前に見ている「櫓」を単に興味の持てないものではなく、自分が興味を持っている「釣り」と関連するものとして見るができるのではないかと考える。さらには、自分の興味である「釣り」に関連する「ナマズ」と「パプアニューギニア」に生息する「ヒクイドリ」に関係があると、興味を広げることも可能であるのではなだろうか。

このように、展示物が見学者の興味から遠いものであったとしても、何らかの関連を強制的に見せることが、見学者が展示物に感心を持つことにつながると考えている。

4.3 展示物を表す Wikipedia ページ集合の取得

提案手法では、まず、ある展示物が与えられると、標本資料目録データベースから、その展示物の

- 標本名
- 地域

の2つの属性の値を取得する。これら2つの属性から、それぞれ Wikipedia ページを1つ取得する。

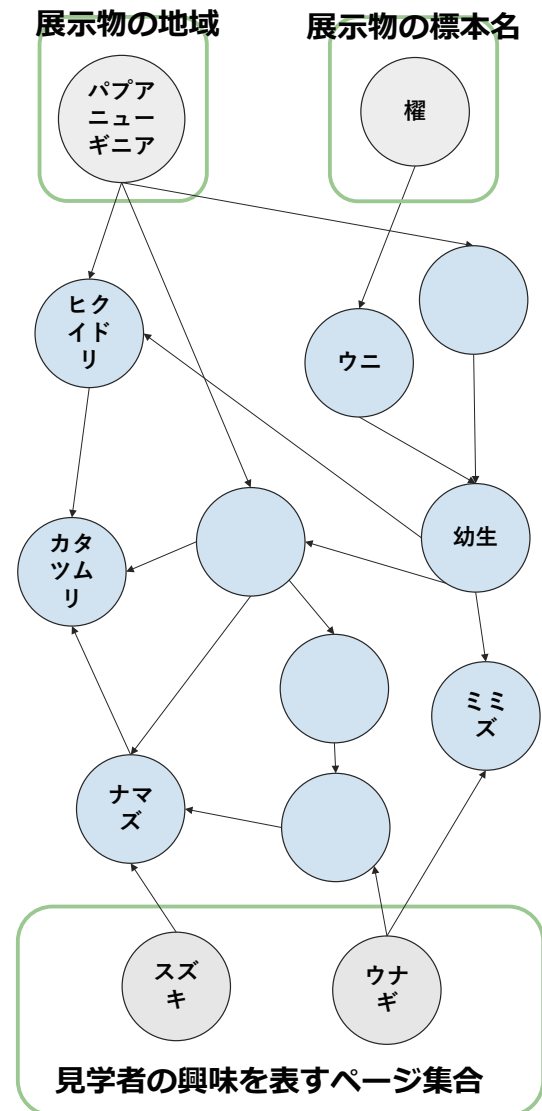


図4 「パプアニューギニア」「櫓」「釣り」に関する Wikipedia パス

まず、標本名から Wikipedia ページを取得する手法を説明する。標本名としては、たとえば、「櫓 \n カイ」や「カヌー用 櫓 \n カヌーヨウ カイ」のような文字列が得られる。まず、得られた文字列から改行文字「\n」の部分を削除し、フリガナの部分を取り除く。改行文字よりも前の部分には、半角スペースで区切られた標本名の漢字表記が記載されている。そのうち、半角スペースで区切られた最後の部分には、その展示物の大まかな種類を表す語が現れていることがほとんどである。そこで、半角スペースで区切られた最後の部分のみを取得する。たとえば、前述の2つの標本名の例の場合、ともに、「櫓」という語が取得される。このようにして得られた語に対応する Wikipedia ページを検索して取得することで、標本名から Wikipedia ページが得られることになる。

次に、地域から Wikipedia ページを取得する手法を説明する。地域としては、たとえば、「パプアニューギニア ラム川流域 <使用>」や「日本国 石川県 <使用>」のような文字列が得られる。こちらでは、国名など、より広い地域を表す地域名から始まり、半角スペースで区切られながら、より狭い地域を表す

地域名が表記されている．そこで，半角スペースで区切られた部分の最初の部分のみを取得する．たとえば，前述の2つの地域の場合，「バプアニューギニア」と「日本国」という語が取得される．このようにして得られた語に対応する Wikipedia ページを検索して取得することで，地域から Wikipedia ページが得られることになる．

以上の手法により，ある展示物が与えられると，展示物の標本名と地域と2つの Wikipedia ページが取得されることになる．標本名の Wikipedia ページは z と地域の Wikipedia ページは r と表すこととする．

4.4 見学者の興味を表す Wikipedia ページ集合の取得

見学者は，何らかの方法で自らの興味を表すこととする．本研究では，見学者があらかじめ興味を持つ Wikipedia カテゴリをいくつか指定するということとする．そして，与えられた Wikipedia カテゴリから Wikipedia ページ集合を取得する．

たとえば，見学者が「釣り」に興味を持っているとする．Wikipedia には，「釣り」というカテゴリが存在する^(注4)．このカテゴリには，下位カテゴリが7件存在し，また，このカテゴリに直接属する Wikipedia ページが46件存在する．

カテゴリによっては，直接属する Wikipedia ページが非常に少ない場合がある．そこで，

- 与えられたカテゴリに直接属する Wikipedia ページ
- 与えられたカテゴリの下位カテゴリに直接属する Wikipedia ページ

をすべて取得し，それらを見学者の興味を表す Wikipedia ページ集合 M とする．たとえば，「釣り」カテゴリの場合，上記のように得られた Wikipedia ページ数，すなわち， $|M|$ は366となる．

4.5 提案手法のアプローチ

展示物の地域と標本名の Wikipedia ページ z ， r と見学者の興味を表すページ集合 M を入力として，リンクパス p を出力する手法には様々なものが考えられる．本節では提案手法のアプローチを説明する．

提案手法ではまず，重要な経由点と考えられる Wikipedia ページ t を取得する．そして， z または r から t への単純道を取得し，さらに， t からいずれかの $m_j \in M$ への単純道を取得し，それらを接続することによって出力である p を構成する．

重要な経由点 t は，以下にあげる2つの性質を持つべきであると考えられる．

- z または r と強い関連を持つ．
- M の多くと強い関連を持つ．

1点目は，展示物の地域，または展示物の標本名とある程度の関連性があることを示す．2点目は，見学者が興味を持つことができることを示す．

このような性質を持つ t を取得するため，本研究ではある Wikipedia ページ集合が与えられたときに，それらと関連が強い Wikipedia 集合を取得する手法として，Personalized PageRank アルゴリズム（以下，PPR と表す）を用いる．PPR は，

Topic-sensitive PageRank [8] や TrustRank [9] で利用されているアルゴリズムであり，PageRank アルゴリズムにおいてランダムジャンプで向かう先がすべてのノードではなく，指定されたいくつかのページに限定されているものと解釈できるものである．

次節で，PPR を用いて，シードとなる Wikipedia ページ集合が与えられた時に，それらと関連が強い Wikipedia ページが高い値を持つようなスコアを計算する手法について説明する．

4.6 Personalized PageRank による関連ページ取得

本節では，ある Wikipedia ページ集合がシードページ集合として与えられた時に，シードページ集合と関連が深いページ集合を取得する手法を説明する．

Wikipedia リンググラフにおいて，ある Wikipedia ページ v の PageRank 値は，パワーメソッドによって以下の式で計算される．

$$PR_{i+1}(v) = \alpha \sum_{u \in B_v} \frac{PR_i(u)}{|F_u|} + (1 - \alpha) \frac{1}{N}. \quad (1)$$

ただし， $PR_i(v)$ は i 番目のイテレーションにおける v の PageRank 値であり， F_u はページ u からリンクされている Wikipedia ページ集合を表す．また， B_v はページ v にリンクしている Wikipedia ページ集合を表す． α は，いわゆるランダムウォーク成分の割合を決めるものである． N はすべての Wikipedia ページの数を表す．ダンピングファクター α は，通常 0.85 が用いられるが，本研究でもこの値を用いている．式の第二項はダンピングファクターと呼ばれ，ランダムジャンプ成分を表すものである．PageRank においては，すべてのページに等しくランダムジャンプするということが表された式となっている．

PageRank と PPR の違いは，このランダムジャンプ成分の部分である． S をシードページ集合としたとき，新たなダンピングファクター $b_v(S)$ を導入し，あるページ v の PPR の値は，以下のように計算される．

$$PPR_{i+1}(v, S) = \alpha \sum_{u \in B_v} \frac{PPR_i(u)}{|F_u|} + (1 - \alpha) b_v(S), \quad (2)$$

$$b_v(S) = \begin{cases} 0 & (v \notin S) \\ \frac{1}{|S|} & (v \in S) \end{cases} \quad (3)$$

PPR が高いページ集合を取得することで，与えられたシードページと強く関連する Wikipedia ページ集合を発見することが可能となる．すなわち，展示物に関連するページや，見学者の興味と関連するページと関連が強いページを取得することが可能となる．しかし，実際に，PPR を適用してみると，いくつかの種類のノイズが混ざってしまうこととなる．

一種類目のノイズは，たとえ，ランダムジャンプの先をシードページ集合に限定したとしても，もともと高い PageRank 値を持つページは，高い PPR 値を取得しがちであるということである．二種類目のノイズは，シードページ集合に含まれるページが，ダンピングファクターによる値を得ることによって，最終的にある程度高い PPR 値をえることが多いということである．

(注4)：「釣り」カテゴリ：<https://ja.wikipedia.org/wiki/Category:釣り>

一種類目のノイズについては、二種類の方法を用いて対応を行う。一つ目は、PPR 値から、元の PageRank 値を引くことである。これによって、元の PageRank 値よりも、どの程度 PPR によって値が増えたかということが得られたスコアの意味となる。以下に、そのようにして得られるスコア Δ の計算式を示す。

$$\Delta(v, S) = \text{PPR}(v, S) - \text{PR}(v) \quad (4)$$

二つ目は、もともと高い PageRank 値を持っているページを、関連ページとして取得される結果のページから排除する方法である。具体的には、PageRank 値が上位 10,000 件以内の Wikipedia ページは、結果から排除する。これは、そのような Wikipedia ページが、良い中継点 t となることがあまりないと考えたためである。たとえば、そのようなページには、「日本」「英語」「アメリカ合衆国」「2006 年」などが含まれる。

一種類目のノイズについては、シードページ集合に含まれるページが PPR においてダンピングファクターとして得るスコアを引くことで対応する。すなわち、実際のスコア計算は以下の式となる。

$$\Delta'(v, S) = \text{PPR}(v, S) - \text{PR}(v) - b_v(S) \quad (5)$$

この手法によって、与えられたシードページ集合と強い関連を持つ Wikipedia ページ集合を取得することが可能となる。以下では、実際に、展示物や見学者の興味と関連するページ集合を取得する手法を説明する。

4.7 展示物と関連するページ集合の取得

先述した通り、展示物に関連する Wikipedia ページとして、標本名の Wikipedia ページ z と、地域の Wikipedia ページ r が得られる。今、取得したいのは、これらのページと関連が強いページである。可能な手法の一つとしては、 z や r をシードページとして $\Delta'(v, S)$ を計算するということが考えられる。しかし、実際にはシードページが一つだけであると、PPR によって関連ページをうまく取得することは難しい。そこで、 z や r を基にして、より大きなシードページ集合を取得するということを行う。PPR によって、 z や r が大きな値を持つためには、 z や r にリンクしているページ集合を取得し、それらをシードページ集合とするという手法が考えられる。

すなわち、 z にリンクしているページ集合 Z' を取得し、それをシードページ集合とし、 $\Delta'(v, Z')$ を計算することで、展示物の標本名と関連が強いページ集合を取得する。そして、 r にリンクしているページ集合 R' を取得し、それをシードページ集合とし、 $\Delta'(v, R')$ を計算することで、展示物の地域と関連が強いページ集合を取得する。

たとえば、 z が「権」の時、「権」にリンクしている Wikipedia ページは 118 ページ存在する。すなわち $|Z'| = 118$ である。実際に、 $|Z'|$ に含まれるいくつかのページについて、表 4 に示す。

そして、 r が「パプアニューギニア」の時、「パプアニューギニア」にリンクしている Wikipedia ページは 1142 ページ存在する。すなわち $|R'| = 1442$ である。実際に、 $|R'|$ に含まれる

表 4 「権」にリンクしているページの例

ページ
オール
艦
高知県を舞台とした作品一覧
船
健磐龍命

表 5 「パプアニューギニア」にリンクしているページの例

ページ
珊瑚海海戦
北マリアナ諸島
ヤシ
オーストラリア
ヤイロチョウ科

表 6 「権」と関連が強い Wikipedia ページ

ページ v	$\Delta'(v, Z')$
権	$4.038e^{-03}$
カヤック	$5.581e^{-04}$
舵	$3.850e^{-04}$
帆	$3.480e^{-04}$
マスト	$3.382e^{-04}$
和船	$3.367e^{-04}$
SOPHIA	$2.942e^{-04}$
ボート	$2.927e^{-04}$
宮本武蔵	$2.759e^{-04}$
OAR	$2.694e^{-04}$
琉球古武術	$2.686e^{-04}$
琉球古武道	$2.555e^{-04}$
小早	$2.527e^{-04}$
竜骨	$2.526e^{-04}$
カッター	$2.482e^{-04}$

いくつかのページについて、表 5 に示す。

このようにして得られた $|Z'|$ をシードページとして、 $\Delta'(v, Z')$ を計算し、その値が大きいページを取得する。そのようにして取得した「権」に対して関連が強い Wikipedia ページを表 6 に示す。

上位のページには、海洋関係のページが現れていることが見て取れる。また、一方で、「健磐龍命」や「高知県を舞台とした作品一覧」といったページがシードページ集合に含まれている影響から、海洋関係以外のページも現れている。「宮本武蔵」など「権」にピンポイントに関係しているページも取得されることが分かる。

次に、 $|R'|$ をシードページとして、 $\Delta'(v, R')$ を計算し、その値が大きいページを取得する。「パプアニューギニア」に対して関連が強い Wikipedia ページを表 7 に示す。

「パプアニューギニア」と同じく太平洋に位置する「ピトケアン諸島」のような地名が上位に出ているだけでなく、「スズメ目」や「ヤム」など「パプアニューギニア」に生息する生物などに関連する語も取得されている。

表 7 「パプアニューギニア」と関連が強い Wikipedia ページ

ページ v	$\Delta'(v, R')$
パプア州	$1.291e^{-03}$
ピトケアン諸島	$1.263e^{-03}$
タロイモ	$1.230e^{-03}$
委任統治	$1.023e^{-03}$
スズメ目	$9.540e^{-05}$
コブラ	$9.333e^{-05}$
ポートモレスビー作戦	$9.159e^{-05}$
環礁	$7.811e^{-05}$
ヤム	$7.338e^{-05}$
珊瑚海	$7.252e^{-05}$
ヤマイモ	$7.249e^{-05}$
スラウェシ島	$7.209e^{-05}$
トランス・ニューギニア語族	$7.138e^{-05}$
オーストラリア国防軍	$7.105e^{-05}$
コーラル・シー諸島	$6.852e^{-05}$

表 8 「釣り」と関連が強い Wikipedia ページ

ページ v	$\Delta'(v, M')$
コイ	$1.308e^{-03}$
ルアー	$1.166e^{-03}$
ミミズ	$1.062e^{-03}$
イカナゴ	$8.717e^{-05}$
アユ	$8.392e^{-05}$
ハエ	$6.185e^{-05}$
FishBase	$5.847e^{-05}$
汽水域	$5.668e^{-05}$
ブラックバス	$5.602e^{-05}$
脊索動物門	$5.363e^{-05}$
つるの剛士	$5.319e^{-05}$
淡水	$5.300e^{-05}$
カイコ	$5.166e^{-05}$
カツオ	$5.107e^{-05}$
キンギョ	$5.074e^{-05}$

4.8 見学者の興味と関連するページ集合の取得

見学者の興味と関連する Wikipedia ページとしては、指定されたカテゴリやその下位カテゴリに属する Wikipedia ページ集合 M が与えられる。先ほどの展示物の場合と同様に、任意の $m_j \in M$ にリンクしているページ集合を取得し、それを M' と表すこととする。

Wikipedia カテゴリの「釣り」が見学者の興味として与えられたとき、 $|M| = 366$ であったが、 $|M'| = 16,818$ となる。 M' をシードページ集合として、 $\Delta'(v, M')$ を計算することで、見学者の興味と関連が強いページ集合を取得する。

「釣り」に対して関連が強い Wikipedia ページを表 8 に示す。

釣りに関連する道具である「ルアー」や、餌として使われる「ミミズ」が得られていることが分かる。また、「コイ」や「アユ」などの具体的な魚も上位に現れていることが分かる。

4.9 重要な経由点

このようにして得られた展示物と関連するページ集合と、見学者の興味と関連するページ集合から、重要な経由点 t を発見

する。先述した t が持つべき性質は、以下のように説明することが可能である。

- Z' をシードページ集合としたときの $\Delta'(v, Z')$ の値が大きい。
- R' をシードページ集合としたときの $\Delta'(v, R')$ の値が大きい。
- M' をシードページ集合としたときの $\Delta'(v, M')$ の値が大きい。

グラフ $G = (V, E)$ において、部分ノード集合 $S \subset V$ を PPR におけるランダムジャンプ先としたとき、関数 $\Delta'(v_i, S)$ は、 V の任意のノード v_i ($1 \leq i \leq |V|$) に対して、ある実数 r_i^S を返す関数である。ここで、 $|V|$ は集合 V のノード数を表す。これを以下のように、

$$\{r_i^S\}_{i \leq |V|} := \Delta'(v_i, S), \quad (6)$$

と表記する。このある実数列 $\{r_i^S\}_{i \leq |V|}$ により定まる V の全順序を V の S によるページのランキング $Rank(V, S)$ と定義する。これらの記号を利用して、かつ、Wikipedia のリンク構造を G と表すとすると、提案手法のアプローチでは、まず、シードページ集合 R' , Z' , M' を利用して、

- (1) $Rank(V, R')$ を算出
- (2) $Rank(V, Z')$ を算出
- (3) $Rank(V, M')$ を算出

することを行う。

我々は、上記の $Rank(V, R')$, $Rank(V, Z')$, $Rank(V, M')$ を求めるのに必要な $\{r_i^{R'}\}_{i \leq |V|}$, $\{r_i^{Z'}\}_{i \leq |V|}$, $\{r_i^{M'}\}_{i \leq |V|}$ をそれぞれの各ノード v_i に関して掛け合わせた

$$\{r_i^{R'Z'M'}\}_{i \leq |V|} = r_i^{R'} \cdot r_i^{Z'} \cdot r_i^{M'}, \quad (7)$$

をノード v_i に関する関連度 $r_i^{R'Z'M'}$ と定義する。この $\{r_i^{R'Z'M'}\}_{i \leq |V|}$ を降順で並べかえた全順序集合 T を得る。我々は、 T の上位のノードを重要な経由点 t と定義する。

実際に、 z を「樺」、 r を「パプアニューギニア」、 M のものになる Wikipedia カテゴリを「釣り」としたとき、取得される T に含まれる上位のページを表 9 に示す。

結果を見ると、「貝類」などの種類名と「ミミズ」のような具体的なものの名称が混在していることが分かった。

4.10 パスの取得

重要と考えられる経由点 t が得られると、それを基にしてパス p を取得する。この t から展示物の標本名 z 、または、地域 r へのパスを取得し、また、 t から見学者の興味を表すページ集合の任意のページ $m_j \in M$ までのパスを取得し、それらを連結することによって得られる Wikipedia のグラフ G の単純道を出力のパス p とする。

表 9 で得られた経由点から、Wikipedia グラフにおいて、「樺」、「パプアニューギニア」、「釣り」関連のページで最短パスを取得した。たとえば、「パプアニューギニア」から「釣り」へのパスの一例は以下になる。

- パプアニューギニア
- ヒクイドリ

表 9 重要な経由点となる Wikipedia ページ

ページ	
貝類	$2.114e^{-13}$
淡水	$7.929e^{-14}$
ミミズ	$2.348e^{-14}$
鳥	$2.056e^{-14}$
棘皮動物	$1.895e^{-14}$
四肢動物	$1.880e^{-14}$
砂浜	$1.812e^{-14}$
ジョルジュ・キュヴィエ	$1.722e^{-14}$
始新世	$1.567e^{-14}$
羽毛	$1.498e^{-14}$
汽水域	$1.352e^{-14}$
硬骨魚綱	$1.246e^{-14}$
単子葉植物綱	$1.191e^{-14}$
分類群	$1.138e^{-14}$
ジョゼフ・ピトン・トゥルヌフォール	$1.100e^{-14}$

- カタツムリ
- ミミズ
- ナマズ
- 釣り（カテゴリ）

このようなパスを取得することが可能であることが明らかとなった。このようなパスの情報を基にして、Wikipedia や Web を使用することで、見学者を展示物への興味を引き付けさせることができるコンテンツを生成することが可能となると考えられる。

5. まとめと今後の課題

本研究では、国立民族学博物館を対象として、Wikipedia グラフを用いて、博物館の見学者の興味と展示物を結び付ける方法を提案した。いくつかの事例を通して、提案手法が展示物と関連が強いページ集合、見学者の興味と関連が強いページ集合を取得することができることを示した。そして、それらを組み合わせることによって、展示物と見学者の興味と関連する Wikipedia ページを取得し、それを基にしてパスを取得することができることが分かった。

しかし、まだ改善しなければならない点がいくつかある。たとえば、取得された経由点には、あまり具体的でない概念を表す語が現れる。より具体的で、興味を持つのに役立つ語を経由点として取得することがより適切であると考えられる。また、「英語」のような非常に一般的である語が取り除いたが、国名などのような一般的な語がまだ現れることも問題である。今後は、パラメータの調整やノイズと考えられるページを取り除く方法を検討する必要がある。また、パスについては現在候補を取得するにとどまっており、今後、良いパスを選択する手法が必要となる。また、提案手法を定量的に評価することも必要である。さらには、見学者の興味と展示物との関連パスから、真に見学者に興味を持ってもらえるかどうか、システムとして実装した後にユーザ実験を行う必要があると考えている。

謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP16H02906, JP16H01756, JP18H03494, JP18H03243, JP17K17832, JP18KT0097, JP18H03338, JSPS 先導的人文社会科学研究推進事業、ならびに、2018 年度国立情報学研究所公募型共同研究による助成を受けたものです。また、本研究の遂行のため国立民族学博物館よりご提供いただいたデータベースを利用しました。ここに記して謝意を表します。

文 献

- [1] Y. Taizo, A. Kenro, K. Noriko, F. Satoko, U. Yusuke, B. Takayuki, N. Shigemi, T. Takashi, H. Yuko, and J. Adachi, “Ceax’s learning support system to explore cultural heritage objects without keyword search,” Current Developments in Technology-Assisted Education, pp.318–322, 2006.
- [2] 山田太造, 相原健郎, 藤沢仁子, 神門典子, 上原祐介, 馬場孝之, 長田茂美, 安達淳, “学校教育における文化財コンテンツ利活用のための教育支援システム,” 日本教育工学会研究報告集, pp.2006(3):23–30, 2006.
- [3] K. Tsvi, K. Judy, and B. Kummerfeld, “Lifelong personalized museum experiences,” Pervasive User Modeling and Personalization (PUMP’10), pp.9–16, 2010.
- [4] S. Oliviero, Z. Massimo, B. Paolo, C. Charles, K. Antonio, K. Michael, K. Tsvi, N. Elena, and C. Rocchi, “Adaptive, intelligent presentation of information for the museum visitor in peach,” User Modeling and User-Adapted Interaction, pp.17(3):257–304, 2007.
- [5] B. Fabian and Z. Ingrid, “Non-intrusive personalisation of the museum experience,” In International Conference on User Modeling, Adaptation, and Personalization, pp.197–209, Springer, 2009.
- [6] H. Marek and W. Ron, “Ontology-based user modeling in an augmented audio reality system for museums,” User Modeling and User-Adapted Interaction, pp.15(3–4):339–380, 2005.
- [7] B. Fabian, “Personalising the museum experience,” Pervasive User Modeling and Personalization (PUMP 2010), p.33, 2010.
- [8] T.H. Haveliwala, “Topic-sensitive pagerank,” Proceedings of the 11th International Conference on World Wide Web, pp.517–526, 2002.
- [9] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, “Combating web spam with trustrank,” Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, pp.576–587, 2004.