

# 属性の組み合わせとその相性に基づく複合オブジェクト検索

佃 洗<sup>†,a</sup> 大島 裕明<sup>†,a</sup>  
山本 光穂<sup>‡,b</sup> 近藤 賢志<sup>‡,b</sup> 田中 克己<sup>†,a</sup>

† 京都大学大学院情報学研究科 ‡ 株式会社デンソーアイティラボラトリ

a) {tsukuda, ohshima, tanaka}@dl.kuis.kyoto-u.ac.jp b) {miyamamoto, skondo}@d-itlab.co.jp

**概要** 本稿では、複合オブジェクトの検索手法を提案する。本研究における複合オブジェクトとは、複数のオブジェクトの集合であり、例えば複数の観光地から成る観光コースや複数の料理から成る献立などがある。提案手法では、複合オブジェクトを構成する各オブジェクトの属性の組み合わせに着目し、ある観点における属性の“同一性”と“相補性”という考えを用いる。我々が対象とするクエリは“3つの観光地からなる観光コース”などである。本稿で提案する複合オブジェクト検索により、このようなクエリに対して“三十三間堂、大覚寺、大徳寺はいずれも本殿の様式という観点では入母屋造である”（同一性に基づく検索）や“壬生寺、南禅寺、大仙院はそれぞれ東洋の美術家という観点では長谷川等伯、狩野探幽、狩野元信と関連がある”（相補性に基づく検索）といった検索が可能となる。

**キーワード** 複合オブジェクト検索, 知識抽出, 相性

## 1 はじめに

Web 検索において、ユーザが検索する対象にはさまざまなものがある。“京都大学 HP”というクエリで検索するユーザが探しているのは単一の Web ページである。“脳梗塞”というクエリで検索するユーザが探しているのは脳梗塞に関する原因や症状、治療法など多岐に渡る場合もある。その場合、単一の Web ページで十分な情報が得られなければ、複数の Web ページを閲覧することになり、結果的にユーザは Web ページ集合を検索していると言える。“iPad Retina”というクエリで検索するユーザが探しているのは、“iPad Retina”という単一のオブジェクトに関する情報である。また、“京都観光コース”というクエリで検索するユーザが探しているのは、“金閣寺”や“清水寺”といったオブジェクトを要素とする複合オブジェクトである。複合オブジェクトを検索する際のユーザの検索意図は様々であるが、例えば以下のような検索意図があげられる。

1. あるユーザが京都に観光に行こうとしており、3箇所の観光地を巡る観光コースを知りたい。
2. あるユーザが休日に読む小説を購入しようとしており、2冊は購入するものが決まった。もう1冊読む時間がありそうなので、あと1冊追加するとしたらどの小説が良いか知りたい。
3. あるユーザがクラシック音楽に興味をもち、ある5枚のCDを購入しようとしている。しかし、すべてのCDを購入するのは費用が高すぎるので、

5枚の中から2枚削除するならどれが良いか知りたい。

上記以外にも多数の状況が考えられるが、一般的な検索エンジンを利用して上記のような検索を行うのは困難である。また、例えば観光コースを Web 上の情報から検索する場合、Web 上に記載されている観光コースしか検索の対象にならない。

これまで、単一オブジェクトの検索を目的とした研究は多数行われてきた [1] [2]。それらの研究では、オブジェクトとその属性を精度良く Web から抽出することを目的としているものが多い。複合オブジェクト検索の場合も、構成要素となるオブジェクトと、各オブジェクトの属性を求める必要がある。これに加えて、複合オブジェクト検索の場合、オブジェクトの組み合わせについて考慮しなければならない。例 (1) の場合、京都の観光地が全部で  $n$  個あったとすると、3つの観光地からなる観光コースの数は、観光地の順序を考慮しないとしても全部で  $nC_3$  個ある。例 (2) の場合も、追加の候補となるオブジェクトの数は膨大である。考えられる全ての組み合わせを検索結果としてユーザに提示するのはユーザの検索の負担が大きくなり適切でない。

そこで本研究では、集合としての“相性の良さ”を考慮した複合オブジェクト検索を目的とし、そのための手法を提案する。オブジェクト間の相性の良さを決める要因には様々なものがあるが、我々はオブジェクトのある観点における“同一性”と“相補性”に着目する。まず“同一性”について説明すると、例えば例 (1) において、“金閣寺”、“龍安寺”、“南禅寺”からなる観光コースは、“時代”という観点から見るといずれも“室町時代”に建設された建物である。“室町時代に関連のある観光コー

ス”としてこの観光コースをユーザに提示することは、ユーザが観光コースを決める上で有益な情報と言える。“相補性”については、例(1)において、“東寺”，“仁和寺”，“広隆寺”からなる観光コースは，“仏像”という観点から見るとそれぞれ“明王像”，“如来像”，“菩薩像”の種類の仏像がある。“仏像の様々な種類に関連のある観光コース”としてこの観光コースをユーザに提示することも、ユーザにとって有益な情報である。以上のように、ある観点において共通の属性をもつ複合オブジェクトは“同一性”によって相性が良いとし、ある観点において互いに異なる属性をもつ複合オブジェクトは“相補性”によって相性が良いとする。オブジェクトの追加および削除についても、追加・削除後の複合オブジェクトの相性の良さに基づいて追加・削除するオブジェクトを決定する。“同一性”および“相補性”については3章で詳しく述べる。

我々はドメインとして“京都市の重要文化財”を対象に実験を行った。実験では、例(1)のように複合オブジェクトの要素数を指定した検索と、例(2)および(3)のようにある複合オブジェクトに対するオブジェクトの追加および削除による検索を行い、検索結果の考察を行った。

本稿の構成は以下の通りである。2章では関連研究について述べる。3章では複合オブジェクトにおける“同一性”と“相補性”という観点から本研究で解くべき問題を定義する。4章では3章で定義した問題を解くための手法を説明する。5章で実験について述べ、6章ではまとめと今後の課題について述べる。

## 2 関連研究

### 2.1 相性に関する研究

心理学の分野では、対人関係における相性の良し悪しに関する研究が行われている [3] [4] [5] [6] [7]。これらの研究では、主に2つの観点から相性を規定している。1つ目は“類似性”であり、価値観や態度、性格などの相互の性質が類似しているほど、相互の対人魅力が高まるというものである。Byrne [3]をはじめとする対人魅力研究において、類似性の重要性が明らかにされている。2つ目は“相補性”であり、一方の性質が他方の足りない性質を補う場合に、相互の対人魅力が高まるというものである。相補性に関する研究として、中里ら [6] は内向型の人物は内向型の人物よりも外向型の人物がより魅力的であると感じるということを示している。また、Winch [7] は、配偶者選択の際には、類似性以上に相補性が重要であるとしている。これらの研究で述べられている“類似性”と“相補性”はそれぞれ、我々の研究における“同一性”と“相補性”に近いものといえる。

### 2.2 オブジェクト検索に関する研究

Nie らは、ユーザがデジタルカメラの商品名のように、オブジェクト名のクエリを入力した際に Web ページを検索結果として返すのでは検索精度が十分ではないため、オブジェクトレベルでの検索を実現する手法を提案した [2]。彼らは、Web ページの構造に着目し、大量の Web ページからまずオブジェクトの属性の候補を抽出し、確度の高いものをオブジェクトの属性としている。また Nie らは抽出した属性情報をもとにクエリに対してオブジェクトをランキングする手法も提案している [1]。彼らの手法では、オブジェクトが記述されている Web ページの重要度と、オブジェクトの属性間の参照関係に基づいてオブジェクトの重要度を求めている。彼らはオブジェクトとして論文を扱っているため、その属性である論文名や会議名、著者名の間には参照関係が存在する。しかし、一般のオブジェクトには必ずしも明確な参照関係は存在しないため、手法の適用範囲は限られる。

Yumoto らはユーザの求める情報の全容を表す Web ページ集合を発見する全容検索を提案した [8] [9]。ページをオブジェクトとみなすと、彼らの研究は複合オブジェクトの検索といえる。彼らの目的が、クエリに関する情報を網羅するできるだけ少ないページ集合を求めることであるのに対して、我々は特定の観点に対するオブジェクトの属性間での同一性や相補性に基づいて複合オブジェクトを求めることを目的としている点異なる。

## 3 問題定義

本章では、まず複合オブジェクトの“同一性”と“相補性”について定義し、次にそれらに基づく複合オブジェクト検索問題を定義する。これらを定義するにあたり、あるドメイン  $D$  が与えられたとき、以下の項目については既知であるとする。

- ドメイン  $D$  に属するオブジェクトの集合  $O_D$ 。例えばドメインを“京都市の観光地”とすると“金閣寺”や“清水寺”などが  $O_D$  の要素になる。
- オブジェクト  $o_i \in O_D$  の属性値集合  $A_{o_i}$ 。例えば、オブジェクト“金閣寺”の属性値集合の要素としては“室町時代”や“足利義満”，“右京区”などがあげられる。
- ドメイン  $D$  における属性名集合  $V_D$ 。ドメイン“京都市の観光地”では“時代”や“戦国大名”，“仏教の宗派”などが  $V_D$  の要素になる。
- 属性名  $v_i \in V_D$  の属性値集合  $A_{v_i}$ 。例えば，“時代”という属性名の属性値集合の要素としては“平安時代”や“室町時代”，“明治時代”などがあげられる。

### 3.1 同一性に基づく複合オブジェクトの相性

ある複合オブジェクト  $S \subseteq O_D$  が同一性において相性が良いというのを、以下の2つの条件を満たす属性名  $v_i \in V_D$  が1つ以上存在することと定義する。

$$\forall o_k \in S, |A_{v_i} \cap A_{o_k}| = 1. \quad (1)$$

$$\left| \bigcup_{o_k \in S} A_{v_i} \cap A_{o_k} \right| = 1. \quad (2)$$

1つ目の条件は、 $S$  内のいずれのオブジェクトも属性名  $v_i$  の属性値を1つもっていることを表す。2つ目の条件は、その属性値が  $S$  内のすべてのオブジェクトで共通していることを表す。

### 3.2 相補性に基づく複合オブジェクトの相性

ある複合オブジェクト  $S \subseteq O_D$  が相補性において相性が良いというのを、以下の2つの条件を満たす属性名  $v_i \in V_D$  が1つ以上存在することと定義する。

$$\forall o_k \in S, |A_{v_i} \cap A_{o_k}| = 1. \quad (3)$$

$$\left| \bigcup_{o_k \in S} A_{v_i} \cap A_{o_k} \right| = |S|. \quad (4)$$

1つ目の条件は、 $S$  内のいずれのオブジェクトも属性名  $v_i$  の属性値を1つもっていることを表す。2つ目の条件は、その属性値が  $S$  内のすべてのオブジェクトで異なることを表す。

### 3.3 同一性と相補性に基づく複合オブジェクト検索

#### 3.3.1 複合オブジェクト検索

本研究で対象とする複合オブジェクト検索では、入力として複合オブジェクトのサイズ  $k$  を受け取る。観光コースの検索の場合、ユーザが観光コースに含めたい観光地の数に相当する。  $k$  が与えられたとき、同一性に基づく複合オブジェクト検索は次の条件を満たす複合オブジェクト  $S$  を発見する問題と定義される：式 (1) および (2) を満たす属性名  $v_i \in V_D$  が1つ以上存在し、かつ  $|S| = k$ 。また、相補性に基づく複合オブジェクト検索は次の条件を満たす複合オブジェクト  $S$  を発見する問題と定義される：式 (3) および (4) を満たす属性名  $v_i \in V_D$  が1つ以上存在し、かつ  $|S| = k$ 。

#### 3.3.2 要素の追加・削除に基づく複合オブジェクト検索

本研究で対象とする追加複合オブジェクト検索では、入力として複合オブジェクト  $T$  と、追加するオブジェクトの数  $k$  を受け取る。観光コースの検索の場合、ユーザが既に観光に行くことを決めている観光地集合が  $T$  に、追加したい観光地の数が  $k$  に相当する。この入力に対して、 $T$  に  $k$  個の要素を追加してもなお、ある属性名の

元で同一性または相補性を満たす複合オブジェクトを求める。

削除複合オブジェクト検索では、入力として複合オブジェクト  $T$  と、削除するオブジェクトの数  $k$  を受け取る。この入力に対して、 $T$  から  $k$  個の要素を削除してもなお、ある属性名の元で同一性または相補性を満たす複合オブジェクトを求める。

## 4 提案手法

3章で定義した各問題を解決するためには、3章で既知としていた  $O_D, A_{o_i}, V_D, A_{v_i}$  を求める必要がある。次節以降で、それぞれを求める方法について述べる。

### 4.1 ドメインに属する複合オブジェクトの取得

本研究では、あるドメインに属する複合オブジェクトを取得するために、ALAGIN フォーラムから提供されている上位語階層データ<sup>1</sup>を用いる。このデータは、Wikipedia<sup>2</sup>で記事の見出し語やカテゴリ名となっている名詞句をその上位語、下位語関係に基づいて階層化したものであり、223,772 個の上位語と 2,751,046 個の下位語から成る。これを用いることで、ある語の下位語集合を求めることが可能となる。例えば、“京都市の重要文化財”という語（ドメイン）に対して、“清水寺”や“金閣寺”など合計 168 個の下位語が得られる。

### 4.2 オブジェクトの属性集合の取得

本研究では  $o_i$  が見出し語となっている Wikipedia の記事中でリンクが貼られている全ての語をそのオブジェクトの属性語集合として収集する。これは、Wikipedia の記事には見出し語に関連のある情報のみが記述されており、ノイズとなる語が比較的含まれにくいためである。

### 4.3 ドメイン内の観点集合の取得

本研究では、大きく2つの段階に分けてドメイン内の属性名集合を求める。まず、 $O_D$  内の全てのオブジェクトの全属性値集合、すなわち  $\bigcup_{o_i \in O_D} A_{o_i}$  に対してクラスタリングを行う。次に、各クラスタの名称、つまり属性名を求める。次項以降で、それぞれの手法について述べる。

#### 4.3.1 属性値集合のクラスタリング

$O_D$  内の全てのオブジェクトの全属性値集合  $\bigcup_{o_i \in O_D} A_{o_i}$  を  $A_D$  とすると、 $A_D$  内の属性値のクラスタリングを行うためには、まず任意の2つの属性値  $t_i \in A_D$  と  $t_j \in A_D$  の距離を定義する必要がある。本研究では、属性値間の距離を次式により定義する。

$$d(t_i, t_j) = \frac{1}{|H_{t_i} \cap H_{t_j}|}. \quad (5)$$

<sup>1</sup><http://nlpwww.nict.go.jp/corpus/>

<sup>2</sup><http://ja.wikipedia.org/>

表 1 “京都市の重要文化財” ドメインにおけるクラスタのラベルとクラスタに属する属性の例.

ラベル	クラスタサイズ	属性の例
日本の大学	55	同志社大学, 上智大学, 関西大学, 早稲田大学, 帝京大学, 近畿大学
東洋の美術家	19	俵屋宗達, 雪舟, 尾形光琳, 与謝蕪村, 狩野探幽, 長谷川等伯, 堂本印象
日本の地方公共団体	18	久留米市, 伊賀市, 奈良市, 桑名市, 長野市, 沖縄市, 松山市, 津市, 鳥取市
仏教系の主な伝統宗教	10	臨済宗, 天台宗, 浄土宗, 真言宗, 浄土真宗, 曹洞宗, 日蓮宗, 律宗
日本の神	6	雷神, 風神, スサノオ, ヤマトノオロチ, ツクヨミ, 八咫鳥
本殿の様式	6	入母屋造, 流造, 祇園造, 権現造, 春日造, 寝殿造
戦国甲斐武田氏の城	2	箕輪城, 高遠城

ここで,  $H_{t_i}$  は 4.1 節で述べた上位語階層データにおける  $t_i$  の上位語集合である. つまり, 2つの属性値がより多くの上位語を共有しているほど, その2つの属性値は類似しており, 距離は近いと考える. 2つの属性値  $t_i$  と  $t_j$  が上位語を共有していない場合, つまり  $|H_{t_i} \cap H_{t_j}| = 0$  の場合, 十分大きな値  $h$  を用いて  $d(t_i, t_j) = h$  とする.

クラスタリングには最長距離法を用いる. 本研究ではクラスタ間の距離の最小値が閾値  $r$  より大きくなった時点でクラスタリングをやめるようにする. クラスタリング後, クラスタのサイズが2以上のものを集め, これをドメイン  $D$  におけるクラスタ集合  $C_D$  とする.

#### 4.3.2 属性値集合のラベリング

クラスタ  $C_i \in C_D$  のラベルは,  $C_i$  に含まれる全ての属性値に共通の上位語から選択する. そのような上位語が複数あるときに, 最も適切なクラスタのラベルの性質について考える. 例えば,  $C_i$  に含まれる語が“俵屋宗達”, “狩野探幽”, “長谷川等伯”の3語であり, これら全てに共通する上位語が“男性”と“江戸時代の画家”であったとする. 本研究の目的が, 複合オブジェクトとしての相性の良さを観点とともにユーザに提示することであることを考えると, 例えばあるコースが“男性という観点において相補性が満たされている”と提示するよりもなぜなら, “江戸時代の画家という観点において相補性が満たされている”と提示する方がユーザにとって情報量が多く, 有益であると考えられる.

以上より, 本研究では次式によりクラスタ  $C_i$  のラベル  $l_i$  を求める.

$$l_i = \operatorname{argmin}_{t_j \in H_{C_i}} |L_{t_j}|. \quad (6)$$

ここで,  $H_{C_i} = \bigcap_{t_j \in C_i} H_{t_j}$ , つまり  $C_i$  内の全ての語に共通する上位語集合であり,  $L_{t_j}$  は 4.1 節の上位語階層データにより得られる  $t_j$  の下位語集合である.

#### 4.4 属性名に対する属性値集合の取得

属性名  $v_i \in V_D$  に対する属性値集合は, 4.3 節で得られるクラスタリングの結果をもとに,  $v_i$  のラベルが付与されたクラスタ内の属性値集合とする.

## 5 実験

提案手法の有効性を確かめるため, “京都の観光コースを決める”というユーザの意図を想定して実験を行った. 本実験では京都の観光地を多く含む“京都市の重要文化財”という語をドメイン名として用いた. 5.1 節では, 4章で提案した各手法の結果を示す. 5.2 節では, 3章で定義した問題に対する結果を示す.

### 5.1 ドメインにおけるデータ収集

まず, 4.1 節で述べた方法により, ドメイン“京都市の重要文化財”に属する全てのオブジェクトを取得する. オブジェクトの総数は168であった. 属性値の数はオブジェクトによって大きく異なり, オブジェクトの一般的な知名度が高いほど Wikipedia の記述量が多く, 属性値の数も多いという傾向が見られた. 本稿では各オブジェクトの属性値をその多寡に関わらず全て用いたが, 属性値の数が少ないオブジェクトについては Wikipedia 以外の情報源から属性値を取得するといったことが今後の課題としてあげられる.

属性値集合のクラスタリングは, 4.3.1 項で述べた閾値  $r$  の値によって結果が異なる.  $r$  の値を  $\frac{1}{3}$ ,  $\frac{1}{4}$ ,  $\frac{1}{5}$ ,  $\frac{1}{6}$  としたときに得られたクラスタリングの結果を著者が確認したところ,  $r = \frac{1}{4}$  のときに最も適切なクラスタに分割されていると判断したため, これ以降では  $r = \frac{1}{4}$  におけるクラスタリングの結果を用いる. このときの総クラスタ数は436個, クラスタの最小サイズは2, 最大サイズは55, 平均サイズは4.1であった. また,  $r = \frac{1}{4}$  のときに得られる各クラスタのラベルとクラスタに属する属性値の例を表1に示す. 最もサイズの大きいクラスタは, 日本の大学に関するクラスタであった. これは, 本実験で対象としたドメインの中に“龍谷大学”や“同志社大学”というオブジェクトが含まれており, それらの属性値集合に多数の大学が含まれていたためである. しかし, 観光において大学という観点はあまり重要でないと言える. 今後は, オブジェクトとしてより適切なもののみを用いることでノイズとなるような観点を除くことができると考えている.

表 2 複合オブジェクトのサイズを指定した検索において得られた結果数.

サイズ	オブジェクト下の全組合せ数	相性のタイプ	相性の良い組合せ数	総結果数
2	14,028	同一性	10,993	20,686
		相補性	7,173	11,308
3	776,216	同一性	197,649	232,761
		相補性	24,608	25,810

## 5.2 複合オブジェクトの検索

### 5.2.1 要素数に基づく複合オブジェクト検索

まず、要素数を指定した複合オブジェクト検索について述べる。手法上は、要素数に制約はないが、要素数が増えるにつれて考慮すべきオブジェクトの組み合わせの数は指数的に増加する。本実験では比較的短時間で結果が得られるように要素数を2および3とした。要素数が増加した場合に計算量を低減する方法が今後の課題としてあげられる。

要素数を2および3としたときの検索結果数を表2に示す。要素数が2の場合、オブジェクトの全組み合わせ数が14,028個であるのに対して、同一性の観点をもつ複合オブジェクトは10,993個存在している。1つの複合オブジェクトが複数の属性名において同一性を満たす場合もあるため、表中の“総結果数”は同一性を満たす全複合オブジェクトの全属性名の数などを表している。この結果を見ると、同一性や相補性を満たす全ての複合オブジェクトを検索結果として提示するには依然として検索結果数が多すぎると言える。この問題を解決するには、同一性や相補性を満たす複合オブジェクトを何らかの尺度に基づいてランキングして提示する必要がある。

検索結果として実際に得られた複合オブジェクトの例を表3に示す。提案手法により、アニメや歌人、本殿の様式など、多様な観点から同一性や相補性に基づく観光地の集合が検索できていることがわかる。

検索結果として適切でないものには大きく分けて2つの傾向が見られた。1つ目は、属性は適切であるが観点が適切でない場合である。例えば、表3において要素数が3で同一性を満たす複合オブジェクトのひとつに“紅葉”を属性値としてもつものがある。“紅葉”は観光において観光地の有用な属性値であるが、この属性値に対する属性名は“キャラクター”となっている。上位語階層データの中には、“紅葉”の上位語として“季節現象”のようにより適したものが存在し、こちらを提示する方が好ましい。この問題を解決するためには、ドメインと各属性名の関連度をあらかじめ求め、関連度を考慮して適切な上位語を求めるといった方法がある。2つ目は、観点として“地名”、属性として“京都市”のように、ドメ

イン内の多くのオブジェクトがもつ属性が同一性の観点として提示される場合である。この問題の解決策としては、各属性値の出現確率を求め、確率の高い属性値はオブジェクト間の相性を考慮するうえで重視しないといったことが考えられる。

### 5.2.2 要素の追加・削除に基づく複合オブジェクト検索

複合オブジェクトに対する追加要素の検索結果の例を表4に示す。紙面の都合上、追加要素の検索結果のみ掲載する。ここでも、計算量の問題のため、元の複合オブジェクトのサイズは追加の場合2または3とし、追加するオブジェクト数は1または2としている。

検索結果を見ると、観光コースを決めるうえで有効な観点や属性を発見できているものが多かったが、以下のような課題もある。1つ目は、5.2.1の考察で述べた課題と同様、ドメイン内の多くのオブジェクトがもつ属性値が提示されたり、属性名が適切でないという問題である。2つ目は、属性値として“摂政”や“関白”のように、抽象的な語が提示されるという問題である。ユーザの検索の支援という観点からは、具体的な属性値の方が望ましい。したがって、語の具体度を考慮した属性値や属性名の提示といった方法が解決策のひとつとして考えられる。

## 6 まとめ

本稿では複合オブジェクトにおけるオブジェクトの属性間の同一性と相補性という観点に着目し、サイズを指定した複合オブジェクトの検索、複合オブジェクトに対するオブジェクトの追加および削除の問題を定義し、検索のための手法を提案した。

また“京都市の重要文化財”を対象ドメインとして実験を行い、ユーザが複合オブジェクトを選択するうえで有効と考えられる観点および属性を求められることを確認した。

今後の課題として複合オブジェクトのランキングがあげられる。5.2.1項でも述べたように、検索結果に含める複合オブジェクトを同一性または相補性を満たすものに限ったとしても、その全てをユーザに提示するには多すぎると考えられる。複合オブジェクトのランキングを行うために、3つの観点から各複合オブジェクトのスコアを求めることを考えている。

1つ目は同一性または相補性としての“良さ”である。例えば、“京都市の重要文化財”というドメインにおいて多くのオブジェクトは“京都市”という属性値をもっているため、“京都市”という属性値で同一性を満たす確率は高い。一方、“狩野探幽”という属性値をもつオブジェクトは多くないため、“狩野探幽”という属性値で同一性を満たす確率は低い。このように、同一性や相

表3 複合オブジェクトのサイズを指定した検索において得られた結果の例.

サイズ	相性のタイプ	観点	オブジェクト (属性)
2	同一性	劇場版アニメ	仁和寺 (名探偵コナン_迷宮の十字路), 東福寺 (名探偵コナン_迷宮の十字路)
		小倉百人一首の歌人	退耕庵 (小野小町), 随心院 (小野小町)
	相補性	新撰組隊士	壬生寺 (近藤勇), 阿弥陀寺 (斎藤一)
		平安時代の日記	清水寺 (更級日記), 毘沙門堂 (明月記)
3	同一性	キャラクター	善峯寺 (紅葉), 常寂光寺 (紅葉), 高山寺 (紅葉)
		本殿の様式	三十三間堂 (入母屋造), 大覚寺 (入母屋造), 大徳寺 (入母屋造)
	相補性	真言系仏教宗派	仁和寺 (真言宗御室派), 東寺 (東寺真言宗), 大覚寺 (真言宗大覚寺派)
		東洋の美術家	壬生寺 (長谷川等伯), 南禅寺 (狩野探幽), 大仙院 (狩野元信)

表4 複合オブジェクトに対する追加複合オブジェクトの検索結果の例. 太字は追加されたオブジェクト.

元の複合オブジェクト	追加数	タイプ	観点	オブジェクト (属性)
{ 伏見稲荷大社, 東福寺 }	1	相補性	門	伏見稲荷大社 (鳥居), 東福寺 (三門), <b>八坂神社 (桜門)</b>
{ 鹿苑寺, 三十三間堂 }	2	相補性	東洋の美術家	鹿苑寺 (伊藤若冲), 三十三間堂 (俵屋宗達), <b>三千院 (竹内栖鳳), 南禅寺 (狩野探幽)</b>
{ 鹿苑寺, 天龍寺 }	1	同一性	宗教・文化人	鹿苑寺 (夢窓疎石), 天龍寺 (夢窓疎石), <b>法観寺 (夢窓疎石)</b>
{ 清水寺, 三十三間堂 }	2	同一性	織豊政権の大名	清水寺 (豊臣秀吉), 三十三間堂 (豊臣秀吉), <b>佛光寺 (豊臣秀吉), 春光院 (豊臣秀吉)</b>
{ 八坂神社, 東福寺, 等持院 }	1	相補性	延喜式による上国	八坂神社 (山城国), 東福寺 (駿河国), 等持院 (下野国), <b>西芳寺 (丹波国)</b>
{ 三千院, 東寺, 清浄華院 }	2	相補性	仏教系の主な伝統宗教	三千院 (天台宗), 東寺 (真言宗), 清浄華院 (浄土宗), <b>本圀寺 (日蓮宗), 平等寺 (曹洞宗)</b>
{ 東本願寺, 知恩院, 西本願寺 }	1	同一性	思想家	東本願寺 (空也), 知恩院 (空也), 西本願寺 (空也), <b>六波羅蜜寺 (空也)</b>
{ 天龍寺, 東寺, 聖護院 }	2	同一性	記念物	天龍寺 (史跡), 東寺 (史跡), 聖護院 (史跡), <b>大徳寺 (史跡), 二条城 (史跡)</b>

補性を満たす確率が低いほど“良い”と考え、スコア付けを行う。

2つ目はオブジェクトと属性の関連度である。例えば“京都市の重要文化財”というドメインにおいて“三千院”と“豊国神社”を含む複合オブジェクトに対して1つオブジェクトを追加する際、“戦国武将”という属性名において“豊臣秀吉”という属性値で同一性を満たすオブジェクトが“高台寺”と“方広寺”の2つあったとする。このような場合、豊臣秀吉と関連度のより高い方を追加候補として上位に提示する方が適切である。

3つ目はドメインと属性名の関連度である。“京都の観光地”というドメインにおいて“仏教”と“大学”という属性名があったとき、“京都の観光地”の関連の強い“仏教”に対する同一性や相補性を満たす複合オブジェクトを上位に提示する方が適切であると考えられる。

また、提案手法の有用性を測るために、ユーザを用いた評価実験も行う予定である。

## 謝辞

本研究の一部は、文部科学省科学研究費補助金 (課題番号 24240013, 24680008, 243993) によるものです。ここに記して謝意を表します。

## 参考文献

- [1] Z. Nie, Y. Zhang, J.-R. Wen and W.-Y. Ma: “Object-level ranking: bringing order to web objects”, Proceedings of the 14th international conference on World Wide Web, WWW '05, pp. 567-574 (2005).
- [2] Z. Nie, Y. Ma, S. Shi, J.-R. Wen and W.-Y. Ma: “Web object retrieval”, Proceedings of the 16th international conference on World Wide Web, WWW '07, pp. 81-90 (2007).
- [3] D. E. Byrne: “The attraction paradigm”, Academic Press (1971).
- [4] K. Endou, I. Yamane and H. Hori: “大学生の結婚に対する意識 (1): 性格特性の相性観について”, Tsukuba psychological research, **12**, pp. 85-91 (1990).
- [5] 岸本康孝: “Jung のタイプ論からみる大学生カップルの相性についての一考察”, Journal of clinical and educational psychology, **31**, 1, p. 109 (2005).
- [6] 田中国夫: “人格類似性と対人魅力-向性と欲求の次元-”, 心理学研究, **46**, pp. 109-117 (1975).
- [7] R. F. Winch: “Mate-selection; a study of complementary needs”, Harper (1958).
- [8] T. Yumoto and K. Tanaka: “Finding pertinent page-pairs from web search results”, ICADL, pp. 301-310 (2005).
- [9] T. Yumoto and K. Tanaka: “Page sets as web search answers”, ICADL, pp. 244-253 (2006).