

Blog 文書中の地物画像を代替する言語的記述の抽出

服部 峻[†] 手塚 太郎[†] 田中 克己[†]

近年、誰でも、どんな計算機環境からでも、Web 文書などのデジタル情報へユニバーサルアクセスを可能にする技術の一つとして、画像メディアを出力できない非視覚的ブラウザのために、マルチメディアな Web 文書をテキスト版に変換する技術が必要とされている。本論文では、この小問題として、一般の Web 文書よりは HTML の構造が制約されている Blog 文書中の画像、特に建築物を含む写真に対して、その内容を説明する言語メディア表現に変換する手法について提案する。まず、対象の Blog 画像の特徴量や ALT 属性は利用せず、そのタイトルや周辺テキスト等のコンテキストに基づいて、その画像中のオブジェクト（建築物）名を特定する。次に、そのオブジェクトの外観に関する言語的な記述を Web マイニングにより抽出し、対象の Blog 画像の代替テキストとして呈示する。元々、画像メディアと言語メディアは表現能力などの特長が異なるため、完全に等価な表現メディア変換は不可能であるが、Web 文書から画像を単に取り除いてテキスト化するのではなく、その画像の内容とできる限り等価な言語メディア表現で代替して呈示できれば、有意義であると考えられる。

Extraction of Visual Description as the Linguistic Alternative to Image of Planimetric Feature in Blog Entry

SHUN HATTORI,[†] TARO TEZUKA[†] and KATSUMI TANAKA[†]

Recently, “ubiquitous society” demands the techniques to access information universally. One of the more important techniques allows us to access the Web via non-graphical browsers such as text or audio browser. In this paper, we propose a novel method to extract a visual description as the linguistic alternative to a photograph of planimetric feature in a blog entry by mining the Web/Blog, and to convert the original multimedia version of the blog entry into the text-only version by replacing each blog photograph with its visual description. Firstly, our method identifies the name of object (planimetric feature) which is taken in a blog photograph, based on not the content and ALT attribute of the blog photograph but its context such as its surrounding text and the title of the blog entry with it. Secondly, our method extracts a visual description of the identified object by mining the Web.

1. はじめに

近年、ユビキタス社会を実現するための大きな課題として、誰もが、どんな計算機環境からでも、デジタル情報にユニバーサルアクセス可能にする技術が求められている。その一つとして、画像メディアを出力できないテキストブラウザ¹⁾ や音声ブラウザ²⁾ といった非視覚的ブラウザからでも Web 文書にアクセスすることを可能にするために、画像メディアも混合した元々のマルチメディアな Web 文書を、言語的記述のみのテキスト版に変換する技術が必要である。

従来の非視覚的ブラウザでは、Web 画像の代わりに、その IMG 要素の ALT 属性の代替テキスト、SRC 属性の画像ファイル名、その画像がリンクアンカーで

あればリンク先の Web 文書のタイトルや URL などのテキスト情報で置換することで、マルチメディアな Web 文書をテキスト版に変換している。Web 文書中の画像が文字入りでメニューリンクとして使われている場合には、リンク先の Web 文書のタイトルや URL でその画像を置換すれば、元々のマルチメディアな Web 文書から視覚的に得られる情報とほぼ等価なテキスト版が得られる。また、単なる飾りとして使われている画像は、単純に除去するだけでも特に問題はない。しかしながら、Web 文書において画像自体が著者の最も伝えたい主要なコンテンツである場合には、従来の単純な置換方法では、元々のマルチメディアな Web 文書から視覚的に得られる情報とは程遠い価値のテキスト版しか得られない。一般の視覚的ブラウザからと非視覚的ブラウザからとでは、同一の Web 文書にアクセスしたとしても得られる情報に大きな差があり、依然としてデジタルデバイドの問題を払拭できていない。

[†] 京都大学大学院 情報学研究所 社会情報学専攻
Department of Social Informatics, Graduate School of Informatics, Kyoto University

この問題に対して、Web 画像の内容を説明する言語的な記述として、その画像中のオブジェクトの名称やオブジェクト間の位置関係といった高次の意味的な情報や、そのオブジェクトの色名、大きさ、形状名などの高次の視覚的な情報を対応付けることができれば、画像メディアと言語メディアは表現能力などの特長が元々異なるため完全に等価な表現メディア変換は不可能であるとしても、Web 文書の一層のマルチメディア化に伴うデジタルデバイドの問題をある程度緩和することができる我々は考える。画像の内容を表す言語的な記述を求める研究は、画像認識や画像理解の分野で古くから盛んに行われているが、光学的文字認識(OCR)や特定の対象領域に限定したオブジェクト認識などを除き、一般のオブジェクトを認識する実用的な手法は未だ考案されていない。一方で、画像検索の分野では、各 Web 画像に対して、その画像に関連するキーワードや説明文といったテキスト情報を索引付けする様々な手法が提案されているが、非視覚的ブラウザからのユニバーサルアクセスを実現するために、画像の代替テキストとして十分適当なテキスト情報を抽出できている研究例は見当たらない。

そこで、本論文では、この小問題として、Blog 文書中の画像、特に建築物を含む写真を、その内容を説明する言語メディア表現に変換する手法を提案する。

まず、対象の Blog 画像の特徴量や ALT 属性は利用せず、そのタイトルや周辺テキスト等のコンテキストに基づいて、その画像中のオブジェクト(建築物)名を特定する。HTML4.0 以降、IMG 要素の ALT 属性が必須となり、一般企業の Web 文書などでは適切な ALT 属性が付与されている場合も増えて来ている。しかし、Blog 文書中の画像では皆無に等しく、記述されている場合であっても、撮影デバイスが自動的に付与したファイル名などの英数字の羅列であり、画像の内容を推測するための情報を何ら与えない場合が多いため、画像の ALT 属性の活用を本論文では想定しない。また、一般の Web 文書ではなく Blog 文書に限定した理由は、非常に多種多様な文書構造やメディアで混成されている一般の Web 文書と比べて、Blog 文書は使用できる HTML のタグが制限されており、文書構造も多くの場合に直線的であること、その文書が投稿された時間を簡単に取得できること等である。

次に、そのオブジェクトの外観に関する言語的な記述を Web マイニングにより抽出し、その Blog 画像の代替テキストとして呈示する。我々はこれまで、道案内補助を目的に、地物の外観情報の抽出に関する研究を行って来ており³⁾、本稿では抽出手法の改善を図る。

本論文の以下の構成を示す。まず、2 章では、本論文に関連する研究をいくつか紹介し、提案手法との比較も行う。3 章では、対象の地物の全体的な外観情報を抽出する手法について、4 章では、対象の地物を構成する要素名、及び、その要素毎の外観情報を抽出する手法について提案する。5 章では、提案手法に関する評価実験とその考察を行う。最後に、6 章で本論文をまとめ、今後の課題についても述べる。

2. 関連研究

2.1 画像検索のための索引語・説明文の抽出

画像検索は、画像に対して予め索引付けされたテキスト情報に基づいて意味的なキーワードで検索する手法(TBIR: Text-Based Image Retrieval)と、画像自身の内容(視覚的な特徴量)に基づいて互いに類似した画像を検索する手法(CBIR: Content-Based Image Retrieval)^{4),5)}とに大別することができる。本論文は、地物画像に対して、写っている地物やその構成要素の名称といった高次の意味的な情報や、その色名、大きさ、形状名などの高次の視覚的な表現記述を抽出しており、前者の TBIR と非常に関連が強い。後者の CBIR では、色、テクスチャ、形状などの低次の視覚的な特徴量に基づいて、類似度計算が行われる。

近年、デジタルカメラやカメラ付き携帯電話の普及により、Web 上には大量の画像が存在しているが、これらをキーワードで画像検索するためには、各画像に対して予めテキスト情報で索引付けしておく必要がある。画像認識や画像理解といったコンピュータ・ビジョンの分野では、一般的な画像に対して、写っているオブジェクトの名称、オブジェクト間の位置関係など、その画像の内容に関する高次の意味的な情報を抽出する一般物体認識の研究が盛んに行われているが、Web 上の多種多様な画像に対して適用することは現時点では容易ではない^{6),7)}。また、画像の低次の視覚的な特徴量と印象語とをマッピングすることにより、ユーザの印象語に基づく画像検索の研究も行われている^{8),9)}。

一方、テキストマイニングや Web マイニングの分野では、各 Web 画像に対して、それ自身(IMG 要素)の ALT 属性に記述された代替テキストや SRC 属性に記述されたファイル名、その画像を含む Web ページのタイトル、その周辺テキストや HTML タグの構造などを解析することにより、索引語や説明文といったテキスト情報を対応付ける手法が提案されている¹⁰⁾。以下、いくつかの関連研究について詳細に述べていく。

是津らは、各 Web 画像に対して、その前後にある

テキストや画像、ハイパーリンクによって関連付けられた周辺コンテンツから抜き出したリンクアンカーや Web ページのタイトルといった画像の「Web 文脈」を抽出し、画像検索の結果と共に呈示することにより、ユーザの検索要求と画像との関連性を視覚的に捉えること可能なシステムを提案している¹¹⁾。

相良らは、Web 画像を含むページ中で強く関連するテキスト領域をまず抽出し、そのテキスト領域内の重要文に基づく評価尺度と、画像からの距離やテキスト領域中での出現頻度に基づく従来の評価尺度とを組み合わせることで、人の感覚により近い、より高精度の索引付けが実現できることを示している¹²⁾。

竹内らは、各 Web 画像の ALT 属性を利用し、その代替テキスト中の語と同一 Web ページ中の語との共起関係を求めることで、その Web 画像により深く関連する複数の文を同一 Web ページ中の文章の中から抽出し、各 Web 画像に対して、抽出された全ての文をその説明文として対応付けする手法を提案している¹³⁾。ALT 属性が記述されていない場合には適用できず、索引付けできる Web 画像が少ないといった問題があったが、「類似した画像には、類似したキーワードが与えられている」という経験則にも着目し、類似画像検索による画像的な特徴と、語の共起関係に基づくキーワードの拡張による言語的な特徴との両方を併用することで、Web 画像の説明文を、同一の Web ページのテキスト部分から抽出する手法を提案している¹⁴⁾。しかしながら、Web 画像の説明文として、同一ページ中のテキスト部分だけから抽出しているため、同一ページ中に画像の説明として適当な文が元々含まれていない場合、上手く抽出できない。成功例として挙げられている画像の説明文には、その画像の見た目を説明するような記述は含まれておらず、画像中のオブジェクト（地物など）に関する様々な説明文であり、そのオブジェクトの外観以外の情報も混合してしまっている。また、非視覚的ブラウザによるアクセスにおいて、同一ページ中のテキスト部分から抽出した文章で画像を代替しても冗長なだけである。一方、我々は、画像中の地物の名称をその周辺テキストを基に特定した後、その画像を含む Blog エントリーに限らず Web 全体から、その地物の外観に関する言語的な記述をマイニングしている。同一ページ内にその地物の外観に関する言語的な記述が含まれていない場合でも抽出することが可能であり、非視覚的ブラウザのための画像の代替テキストの自動生成を実現できる。

岡田らは、画像データに対する感性検索を行うために、名詞だけでなく形容詞もキーワード候補として抽

出している¹⁵⁾。抽出された形容詞の中には、オブジェクトの外観情報として、高次の視覚的な表現記述や印象語を含んでいる場合もあるが、あくまで同一ページ中のテキストだけから抽出しているため、画像の内容に関する視覚的な説明が同一ページ中に元々含まれていなければ抽出することができないという問題が残る。

2.2 地物の特徴抽出

地物の特徴として、外観情報の他にランドマーク性や体験を抽出する研究がある。ランドマークとは、空間を移動する際に目的地を見付けるための目印となる地物のことである。手塚らは、人々が頻繁に目的地にして広く知られている地物であるオーソリティ型のランドマークを文書頻度に基づいて、空間的な位置関係を把握するために重要となる地理オブジェクトであり経路説明に頻繁に用いられるハブ型のランドマークを周辺の地名との共起度に基づいて、Web から抽出する手法を提案している¹⁹⁾。さらに、店舗などの開店している時間帯を抽出し、与えられた時間に活動している場所を可視化するシステム「ChronoSearch²⁰⁾」を実装している。一方、倉島らは、ある場所ですぐに行った体験に関する文書を多く含み、且つ、日時と伴に記録されるという Blog の特性に着目し、特定の場所について記述された Blog 文書のテキストから、その場所の訪問者による体験を、時間、空間、動作、対象属性の間の相関ルールマイニングによって抽出する手法を提案している²¹⁾。ある場所の上で行われた典型的な人々の活動を把握することが可能である。

地物に限らず、何らかの一般の対象物が与えられた時に、その評判情報を Web や Blog などの文書から抽出する研究も盛んに行われている。小林らは、評価対象表現、属性表現、評価表現の共起パターンを利用することで、これら領域依存の表現セットを効率的に収集する手法を提案している²²⁾。藤村らは、Web 文書から評判情報を抽出し、肯定・否定・非評価を判定する分類器を構築している²³⁾。また、赤木らは、ある対象に対して、どのような観点から評価されているかを表す評価属性の名称を抽出している²⁴⁾。本論文では、「八坂神社」の「西楼門」は「朱塗り」というように、ある地物に対して、その地物の構成要素、及び、その外観的な評価を表す語の組を抽出しており、一般オブジェクトに対する評価情報抽出の研究と非常に関連しているが、実世界の地物の外観は、いつ、何処から、どんな状況下で見るとかに依って大きく変化する場合がある点特徴的である。但し、本論文では、時空間や状況に依存した外観情報の抽出までは行っていない。

3. 地物の外観情報の抽出

ある地物が典型的にどのように見られているかを表す外観情報を、Web マイニングにより抽出する手法について述べる。地物の外観情報としては、建物全体としての色、大きさ、高さ（階数）、形状、建物の構成要素である外壁や屋根などの色、材質、大きさ、形状などが考えられる。本論文では特に、地物の色、材質、階数を抽出する手法について述べる。Blog 文書中の地物画像の周辺テキスト中には、その画像に写っている地物の名称に関しては記述されている場合が多く、その地物名を特定した後、その地物の外観情報を本章の手法により抽出し、その地物画像を代替する言語的記述として置換する。本論文では、検索エンジンがメタデータとして持つ検索質問の検索件数とスニペット（要約文）を用いて、地物の外観情報を Web マイニングする。本手法は次の二つの基本ステップから成る。

Step 1: 抽出対象の地物名と抽出したい外観属性の表現テンプレートを基に構成した検索質問を検索エンジンに投げ、その検索結果ページの要約部を解析し、抽出したい外観属性の値となる外観語の候補を絞る。本章の次節以降で詳細に述べる。

Step 2: 外観属性値の各候補語 (v) に対して、検索エンジンの検索件数を用いて、対象の地物 (p) への相応度を以下で定義する式により評価する。

$$S_p(v) = \frac{df(p \wedge v)}{df(v)}$$

但し、 $DF(q)$ は、検索エンジンのコーパス中、検索質問 (q) を満たす文書の件数を表す。

3.1 地物の色情報の抽出

全く同じ地物を見たとしても、人に依って感じ方は異なり、その色を表現する仕方も千差万別である。逆に、ある地物の色に関する記述として呈示された色名から連想する色のイメージも人に依って異なるため、地物の色という外観情報を Blog 画像の代替テキストとしてユーザに呈示しても等価な表現メディア変換になるとは限らない。しかしながら、そもそも同じ画像から得られる情報は人に依って異なることと比べれば、言語的な記述の方が個々人に依る受け取り方の変動は小さいため、Blog 画像の代替テキストとして地物の色情報は、ある程度は有用であると考えられる。色に関する認知データをユーザ毎に事前に取得しておくことで、個人適応化することも考えられ、今後の課題としたい。

まず、一般的に良く使用され、ユーザの多くが正しく理解できる色名の辞書を用意する。本論文では、色

名辞書として、日本工業規格 (JIS) で規定している「物体色の色名 (JIS Z 8102:2001)」に現在登録されている 269 種の慣用色名を参考にした。主に工業製品に対して用いられる色の名称ではあるが、工業製品の利用者にとって分かり易い色名が選択されており、普段、利用者自身が色を表現するのに用いる名称を網羅していると考えた。登録されている慣用色名は、「黒」「白」「赤」「青」「ピンク」のように語尾に「色」が付いていない色名と、「黄色」「茶色」「肌色」「朱色」「金色」のように語尾に「色」が付いている色名とに二分できる。語尾に「色」が付いていない色名も、語尾に「色」を付けて用いることもあるが、語尾に「色」を付けない方が一般的である。一方、語尾に「色」が付いている色名も、語尾に「色」を付けないで用いることもあるが、「茶」や「肌」のように別の意味に変わってしまい、色名だけを表さなくなってしまう色名もある。

ある地物の外観を記述する表現としては、次のような二種類のタイプが考えられる。但し、構成要素とは、地物の外観を構成する外壁や屋根、門などを指す。

Type 1:

- 「(色名) の (地物名)」
- 「(色名) の (構成要素) の (地物名)」
- 「(地物名) の (色名) の (構成要素)」など

Type 2:

- 「(地物名) の (色名)」
- 「(地物名) の (構成要素) の (色名)」
- 「(地物名) の色は (色名) である」など

色名の語尾に「色」が付く場合「色の」という格助詞付きのテンプレートを Type 1 から得ることができ、最も単純には、「色の (地物名)」という検索質問を検索エンジンに投げた検索結果ページから「色の」の前に直結する語を切り出す方法が考えられるが、検索結果が 0 件である場合も多い。そこで、検索条件を少し緩め、「色の」という表現テンプレートを使い、Web 検索エンジンに「(地物名) AND “色の”」という検索質問を投げた結果の上位数件のスニペット中から「色の」の前に直結する語を切り出し、色名辞書に登録されている色名だけを候補語とする。しかし、上述の方法で地物の色名の候補を絞ると、一般に語尾に「色」を付けない色名を漏らしてしまうという問題が残る。例えば、ある地物の外観を表す色名として最も相応しい色名が「赤」であり、その地物について文書を記述した者も「赤色」という表現を使用していないと、「赤」という色名を候補から漏らしてしまい、最適な色名を解として求める機会を失ってしまう。

この問題への対策としては、一般に語尾に「色」を付けないで使用する色名は、表現テンプレートを用いた絞り込みの結果に依らず候補語に含めるという方法が考えられる。検索エンジンに投げなくてはならない検索質問の数は増大してしまうが、最適な色名を候補から漏らさないことと計算時間を抑制することとはトレードオフの関係である。語尾に「色」が付いている色名の一つが最適な色名である場合においても、上位何件までを表現テンプレートで切り抜くかに依って最適な色名を漏らしてしまう可能性は残るため、精度を重視する場合には、色名辞書に登録されている色名の全てを Step 2 の候補語とすべきである。

3.2 色以外の外観情報の抽出

地物の外観的な材質としては、「木」「石」「レンガ」「ブロック」「タイル」等がある。地物の色情報の抽出で用いた「色の」というテンプレートに対して、「造りの」や「造の」といったテンプレートを用いることで、地物の外観的な材質を表す語の候補を絞ることができる。しかし、地物の内部構造を表す「鉄骨造」「鉄筋コンクリート(RC)造」などや、地物の建築様式を表す「寝殿造」「武家造」「合掌造」なども候補として取得され得る。前者に関しては、地物の外観情報としては全く相応しくない。後者に関しては、地物の材質という属性名に対する属性値としては不適切ではあるが、これらの情報から地物の外観を連想することはできるため、全く無意味であるとは言えない。地物の材質と様式とを分別する手法については今後の課題とする。次に、地物の階数は数値であり、「階建て」「階建」といった表現テンプレートを用いることで、地物の階数を表す語の候補を絞ることができる。

4. 地物の構成要素とその外観情報の抽出

Blog 文書中の地物画像の周辺テキスト中には、その画像に写っている地物の名称に関しては記述されている場合が多い。しかしながら、その地物の特にどの構成要素について撮られた写真であるかまでは記述されていない場合もある。このような画像と言語的記述のギャップは、著者にとって、地物名を読書に伝えることは重要であっても、その地物の特定の構成要素名を伝えることは重要でない場合に起こり得るが、一方で、その地物の画像を掲載しているということは、その地物の外観情報は伝えたいと考えてのはずである。ある地物が多数の構成要素を持ち、その構成要素毎に外観は異なる場合も多い。対象の地物画像に対して、

その地物名しか特定せず、その地物の全体的な外観情報を抽出する手法に比べ、その地物の構成要素名までを特定し、その地物の特定の構成要素の外観情報を抽出する手法の方が、その地物画像を代替する言語的記述として、より正確かつ詳細な情報である。もちろん、地物名さえ記述されていない場合には、その画像自体の内容を解析しない限り、その画像に写っている地物の名称および外観情報を抽出することは不可能である。本論文では、Blog 文書中の地物画像の周辺テキスト中にその地物名は記述されており特定できるという前提で、その周辺テキストだけではなく Web 全体から、その地物を構成する要素の名称、そして、その地物の特定の構成要素の外観情報を抽出することを想定する。

4.1 地物の構成要素の抽出

ある地物を構成する要素名のリストを Web マイニングする手法として、[“(地物名)の”]という検索質問を検索エンジンに投げた結果の上位数件のスニペットから“(地物名)の”以降に続く名詞を形態素解析により切り出す手法が考えられる。しかしながら、日本語の文法における「の」は様々な目的で使用されるため、単純に、Google 等の Web 検索エンジンに投げた検索結果を解析しても、地物の構成要素とは何ら関係のないキーワードも多数抽出されてしまい上手くいかない。一方で、我々は“(地物名)の”という記述が画像の周辺に在る場合“(地物名)の”に続く名詞がその地物の構成要素を表していることが多いことを発見した。そこで、[“(地物名)の”]という検索質問を Web 検索エンジンではなく画像検索エンジンに投げ、その上位数件のスニペットから“(地物名)の”に続く名詞を形態素解析により切り出し、その地物の構成要素の候補語とする。次に、各候補語に対して、[“(地物名)の(構成要素名)”]という検索質問を画像検索エンジンに投げて件数を得る。そして、画像検索エンジンに[“(地物名)の”]という検索質問を投げた検索件数に比べて十分に大きな占有割合を持つ候補語を、その地物の構成要素の名称として採択する。

4.2 地物の構成要素の外観情報の抽出

地物名とその構成要素名が既知の場合に、その地物の構成要素の外観情報を抽出する手法について述べる。基本的には“(地物名)の(外観記述)(構成要素名)”という形式(表現テンプレート)の記述を抽出する。

まず、[“(地物名)の*(構成要素名)”]という検索質問を Web 検索エンジンに投げた上位数件のスニペットから“(地物名)の”と“(構成要素名)”の間の

テキストを切り出し、外観候補とする。但し、「*」は Google 等の「ワイルドカード」検索を意味する。

次に、外観情報の各候補に対して、Web 検索エンジンに [“(地物名)の(外観候補)(構成要素名)”] という検索質問を投げて検索件数を取得する。そして、[“(地物名)の*(構成要素名)”] という検索質問を Web 検索エンジンに投げた検索件数に比べて、十分に大きな占有割合を持つ外観候補だけを、対象の地物の構成要素の外観情報として採択する。

5. 評価実験と考察

本論文では、Blog 文書中の地物画像を代替する言語的な記述を得るための基盤技術として、地物の外観情報の抽出、地物の構成要素名の抽出、地物の構成要素の外観情報の抽出について提案した。本章では、各々の抽出手法について評価実験を行い、順に考察する。

5.1 地物の外観情報の抽出例

地物の外観情報の抽出例として、以下のような外観的な特徴を持つ京都祇園の「八坂神社」を取り上げる。

- 西から近付くと、大きな朱色の西楼門が見える。
- 南から近付くと、まず、大きな石鳥居が見え、さらに近付くと、その奥に朱色の南楼門が見えて来る。
- 本殿や鳥居などの全体が緑色の森に包まれている。従って、色名として「朱色」「緑」「赤」を正解とする。

「八坂神社」の色情報を抽出するためにまず、色名の候補を絞る。[“八坂神社” AND “色の”] という検索質問を Google の Blog 検索エンジンに投げ、その検索結果の上位 50 件のスニペットから「色の」の直前の名詞を切り出す。色名辞書に含まれている語だけを候補語とすると、「ピンク」「朱」「オレンジ」「赤」「グレー」「セピア」「緑」「若草」「珊瑚」などの色名が取得され、「仏教」「二」「地域」などは棄却される。

次に、色名の各候補について「八坂神社」という地物名に対する相応度を算出すると表 1 のようになる。各色に対する「色」付きと「色」無しの名称とを別々に扱うと、相応度は「朱色」「朱」「緑」「青」の順となり、「青」が「赤」や「赤色」よりも大きな値となっているが、Web において「八坂神社」と「青蓮院」との共起性が非常に強く、[“八坂神社” AND “青”] という検索質問の件数に、「青」単独ではなく「青蓮院」を含む Web 文書も含まれているためである。「色」付きと「色」無しの名称の相応度の中で小さい方を各色の概念全体としての相応度とすると、「朱色」「緑」「赤」「青」の順となり、ロバスト性を高めることができる。

表 1 地物名 ($p = \text{“八坂神社”}$) に対する色名 (v) の相応度
Table 1 Suitability of Color-Name (v) for Planimetric Feature's Name ($p = \text{“Yasaka-Jinja”}$)

v	$df(p \wedge v) / df(v)$	$S_p(v) \cdot 10^3$	Rank	
赤	150 / 1310664	0.1144	9	3
赤色	7 / 49357	0.1418	8	
朱	40 / 64119	0.6238	2	1
朱色	40 / 11939	3.3504	1	
ピンク	66 / 779861	0.0846	11	5
ピンク色	15 / 70785	0.2120	6	
オレンジ	34 / 471717	0.0721	12	6
オレンジ色	11 / 63656	0.1728	7	
緑	163 / 450345	0.3619	3	2
緑色	21 / 96170	0.2183	5	
青	182 / 604921	0.3009	4	4
青色	5 / 52556	0.0951	10	

5.2 地物の構成要素の抽出

いくつかの著名な地物名に対して、その構成要素の抽出を試みる。Google の画像検索エンジンに [“(地物名)の”] という検索質問を投げ、その検索結果の上位 100 件のスニペットから「(地物名)の」に続く名詞句を抽出する。そして、抽出された地物の構成要素名とその後ろに、[“(地物名)の(構成要素名)”] という検索質問による件数を付すと表 2 のようになる。但し、「写真」「イメージ」「画像」などは予め除去している。これは、Google の画像検索エンジンが、例えば [“(地物名)の写真”] という検索質問を投げたとしても、[“(地物名)の”] として扱うためである。

Web 画像検索エンジンに対して同様に抽出した場合に比べれば、地物を構成する要素の名称以外のノイズは少なくなっているが、依然として問題が残っている。しかしながら、「近く」「そば」「前」「後」「奥」「中」といった位置関係を表す語、「東」「西」「南」「北」といった方角を表す語、「朝」「昼」「夜」といった時間帯を表す語、「春」「夏」「秋」「冬」といった季節を表す語を含む名詞句をストップワードとして除去しさえすれば、ほとんどのノイズを除去することが可能である。例えば、「八坂神社」に対する「近く」「奥」や「平安神宮」に対する「中」などは、検索件数だけを見ると大きな値であり、その地物の構成要素として採択される危険性がより高いが、ストップワード対象として除去することができる。

逆に、「八坂神社」に対する「南楼門」や「金閣寺」に対する「屋根」など、地物の構成要素名として本来は相応しいが、その検索件数が少ないために棄却されてしまうものも見られる。もちろん、その地物に対して、その構成要素が写真を撮る対象になり難いことを表しているのかもしれない。しかしながら、地物の構

表 2 地物の構成要素の抽出例

Table 2 Component Extraction for Planimetric Feature

地物名	抽出された地物の構成要素の上位とその検索件数（降順ソート）
八坂神社（592）	境内（36）、祇園祭（26）、鳥居（24）、近く（21）、枝垂れ桜（12）、本殿（11）、奥（11）、西楼門（10）、おけら参り（8）、正門（7）、紅葉（7）、節分祭（7）、入り口（6）、森（6）、仁王（6）、大晦日（6）、南楼門（3）、末社（4）、絵馬（2）、拝殿（1）
平安神宮（576）	桜（153）、鳥居（58）、大鳥居（55）、神苑（37）、庭園（24）、菖蒲（18）、四季（16）、枝垂れ桜（9）、門（8）、中（8）、応天門（7）、お庭（6）、睡蓮（6）紅葉（6）、初詣（3）、祭礼（3）、雪（3）、おすすめ（2）、紅枝垂桜（2）、朱（2）、蓮（1）、祭り（1）、鯉（1）
金閣寺（598）	庭園（77）、紅葉（77）、前（31）、近く（15）、製作（12）、雪景色（10）、池（9）、茶店（9）、杜若（7）、見学（7）、中（7）、庭（6）、鳳凰（6）、茶室（5）、一枚天井（5）、巻（5）、ホテル（4）、見どころ（4）、もみじ（4）、不動明王（3）、撮影（3）、屋根（2）、お守り（2）、そば（2）、様子（2）、こけら葺き屋根（1）、雨樋（1）、晩秋（1）、盛夏（1）
銀閣寺（673）	紅葉（79）、風景（41）、近く（40）、参道（28）、前（16）、苔（15）、庭園（13）、後（11）、入口（8）、ススキ（6）、雪（5）、そば（5）、中（5）、門（4）、新緑（4）、北（4）、石庭（3）、茶室（3）、味（3）、駐車場（3）、うめ（2）、夜間公開（2）、手前（2）、帰り（2）

表 3 地物「八坂神社」の構成要素の外観情報の抽出例

Table 3 Visual Description Extraction for Components of “Yasaka-Jinja”

構成要素名	外観情報の抽出例
境内（145）	祇園祭（11）、さして広くない（7）、小高い（1）、拝殿前（1）
祇園祭（443）	祭礼である（67）、祭礼（36）、氏子は（22）、有名な（9）、祭り（8）、お祭り（3）、守谷（3）、流れを汲む（2）、お祭り（1）、恒例（1）、主催する（1）
鳥居（116）	南の（98）、東の（88）、石（64）、赤い（62）、南側の（43）、南門（37）、赤（12）、正面（7）、北東の（7）、南側（5）、肥前（5）、入口（2）、入り口の（2）、祇園（1）、石の（1）
本殿（49）	鳥居と（18）、拝殿と（16）、正門から入って（8）、西楼門から（4）、南楼門から（2）
西楼門（44）	正門である（19）、西側にある（6）、朱塗りの（4）、主門である（4）、顔となった（1）、正面玄関とも言える（1）

成要素の抽出手法として今後、適合率だけでなく再現率も上げていくためには、これらの写真を撮る対象になり難いような構成要素名も抽出できる必要がある。

5.3 地物の構成要素の外観情報の抽出

「八坂神社」という地物名の上位にランクされている構成要素のいくつかに対して、その外観情報を抽出すると表 3 のようになる。「祇園祭」という構成要素名に対して抽出された候補には、適当な外観情報は一つも含まれていない。「祇園祭」はそもそも、「八坂神社」の構成要素ではなく、「八坂神社」で行われるイベントの一つであり、他の構成要素の例とは一線を隔す。

「八坂神社の」と「(構成要素名)」との間のテキストをそのまま候補としているが、係り受け解析を用いて「(構成要素名)」を修飾しているテキストだけに限定したり、位置関係や方角などを表す語を含む候補を除去したりすれば、より精度を上げることができる。しかし、Blog 文書中の地物画像を代替するための地物の外観情報として、まだ十分とは言えない。従って、今後の課題として、抽出のための表現テンプレートを改良し、地物の外観情報をより効率的に収集することができる検索質問を探索していく必要がある。

6. おわりに

本論文では、Blog 文書中の画像、特に著名な建築物を含む写真を、その内容を説明する言語的な記述に変換する手法について提案した。まず、対象の Blog 画像の特徴量や ALT 属性は利用せず、そのタイトルや周辺テキスト等に基づいて、その画像中のオブジェクト（建築物）名を特定する。次に、各オブジェクト名の外観に関する言語的な記述を Web マイニングにより抽出し、その Blog 画像の代替テキストとして呈示する。デジタル情報へのユニバーサルアクセス技術の一つとして、画像メディアを出力できない非視覚的ブラウザのために、マルチメディアな Web 文書をテキスト版に変換する技術が必要であるが、本提案手法の適用により、Web 文書から画像を単に取り除いてテキスト化するのではなく、その画像とできる限り等価な言語メディア表現で代替することができ、デジタルデバイドの改善に貢献できると考える。

今後の課題としては、より大規模な評価実験を行い、地物の外観語としての相応度を量る尺度の改良や、地物の外観情報をより効率的に収集するための検索質問の構成方法などについて検討を続ける。

謝辞 本研究の一部は、文部科学省 21 世紀 COE 拠点形成プログラム「知識社会基盤構築のための情報学拠点形成」(リーダー: 田中克己, 平成 14~18 年度), 及び、文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」, 異メディア・アーカイブの横断的検索・統合ソフトウェア開発(研究代表者: 田中克己), 及び、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」, 計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(研究代表者: 田中克己, A01-00-02, 課題番号: 18049041) により, ここに記して謝意を表す。

参 考 文 献

- 1) Lynx: <http://lynx.browser.org/> (2006).
- 2) Home Page Reader: <http://www-06.ibm.com/jp/accessibility/soft/hpr.html> (2006).
- 3) 服部峻, 手塚太郎, 田中克己: サーチエンジンのメタデータを用いた道案内補助のための外観情報の抽出, 情報処理学会 研究報告「データベースシステム」, Vol.2006, No.78, pp.169-176 (2006).
- 4) V. Gudivada, and V. Raghavan: Content-Based Image Retrieval-Systems, IEEE Computer, Vol.28, No.9, pp.18-22 (1995).
- 5) 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司: 色や形状等の表層的特徴量にもとづく画像内容検索技術, 情報処理学会 論文誌「データベース」, Vol.40, No.SIG03 (TOD1), pp.171-184 (1999).
- 6) 彌富仁, 萩原将文: ファジー推論ニューラルネットワークを用いた風景画像からの知識抽出と認識, 電子情報通信学会 論文誌, Vol.82-D2, No.4, pp.685-693 (1999).
- 7) 柳井啓司: 一般画像自動分類の実現へ向けた World Wide Web からの画像知識の獲得, 人工知能学会誌, Vol.19, No.5, pp.429-439 (2004).
- 8) 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ: 印象語による絵画データベースの検索, 情報処理学会 論文誌, Vol.33, No.11, pp.1373-1383 (1992).
- 9) 八村広三郎, 英保茂: 色彩分布と印象語に基づく絵画データの検索, 情処学会 研報「人文科学とコンピュータ」, Vol.1995, No.91, pp.37-44 (1995).
- 10) 出原博, 藤本典幸, 竹野浩, 萩原兼一: WWW 画像検索における画像周辺の HTML 構文構造を考慮した画像説明文の抽出手法, 電子情報通信学会 技術研究報告, DE2005-136, pp.19-24 (2005).
- 11) 是津耕司, 田中克己: Web からの画像の文脈情報の抽出と提示, 第 14 回データ工学ワークショップ (DEWS'03) 論文集, 6-P-05 (2003).
- 12) 相良直樹, 砂山渡, 谷内田正彦: HTML テキストの重要文を用いた画像ラベリング手法, 信学会 論文誌, Vol.J87-D1, No.2, pp.145-153 (2004).
- 13) 竹内謹治, 黄瀬浩一, 松本啓之亮: 語の共起の統計情報を用いた画像と説明文の対応付け, 信学会総合大会講演論文集 2005, D-5-1 (2005).
- 14) 竹内謹治, 黄瀬浩一: 類似画像とキーワードを利用した Web 画像の説明文抽出, 情処学会 研報「自然言語処理」, Vol.2006, No.1, pp.7-12 (2006).
- 15) 岡田真, 浜田浩史, 宝珍輝尚: マルチメディアデータの効率的検索のためのキーワード自動抽出手法, 情報処理学会 研究報告「自然言語処理」, Vol.2005, No.94, pp.73-78 (2005).
- 16) C. Framkel, M. Swain, and V. Athitsos: Web-Seer: An Image Search Engine for the World Wide Web, Technical Report TR-96-14, University of Chicago (1996).
- 17) N. Rowe, and B. Frew: Automatic Caption Localization for Photographs on World-Wide Web Pages, Information Processing and Management, Vol.34, No.1, pp.95-107 (1998).
- 18) K. Koiso, T. Mori, H. Kawagishi, K. Tanaka, and T. Matsumoto: InfoLOD and LandMark: Spatial Presentation of Attribute Information and Computing Representative Objects for Spatial Data, International Journal of Cooperative Information Systems (IJCIS), Vol.9, No.1-2, pp.53-76 (2000).
- 19) T. Tezuka, and K. Tanaka: Landmark Extraction: a Web Mining Approach, Spatial Information Theory (COSIT'05), LNCS Vol.3693, pp.379-396 (2005).
- 20) T. Tezuka, and K. Tanaka: Temporal and Spatial Attribute Extraction from Web Documents and Time-Specific Regional Web Search System, Web and Wireless Geographical Information Systems (W2GIS'04), LNCS Vol.3428, pp.14-25 (2005).
- 21) T. Kurashima, T. Tezuka, and K. Tanaka: Blog Map of Experience: Extracting and Geographically Mapping Visitor Experiences from Urban Blogs, In: Proceedings of the 6th Web Information Systems Engineering (WISE'05), LNCS Vol.3806, pp.496-503 (2005).
- 22) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: テキストマイニングによる評価表現の収集, 情報処理学会 研究報告「自然言語処理」, Vol.2003, No.23, pp.77-84 (2003).
- 23) 藤村滋, 豊田正史, 喜連川優: 文の構造を考慮した評判抽出手法, 第 16 回データ工学ワークショップ (DEWS'05) 論文集, 6C-i8 (2005).
- 24) 赤木法生, 大島裕明, 小山聡, 田島敬史, 田中克己: レビューページ例からの属性抽出に基づくレビューページ検索, 第 17 回データ工学ワークショップ (DEWS'06) 論文集, 2C-i10 (2006).