

Webからの画像の文脈情報の抽出と提示

是津 耕司^{†,††} 田中 克己[†]

† 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

†† 独立行政法人通信総合研究所 〒184-8795 東京都小金井市貫井北町4-2-1

E-mail: †{zettstu,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし Web上に存在する不特定多数の画像を検索する場合、従来の画像検索のように個々の画像の中身を正確に検索することは困難である。検索ユーザはまず画像に関連する話題を表すWebページ（あるいはWebページの一部）を検索し、それらを閲覧しながら適切な中身を持った画像を選択することを行っている。このギャップは、画像の検索可能性や検索結果の信頼性にとって無視できない問題になる。このギャップを埋めるため、本研究では画像がWeb上でどんなWebコンテンツ（テキストや他の画像）と関連付けられ一緒に使われているのかを表す“Web文脈”を提案する。画像のWeb文脈は、画像の周辺にあるWebコンテンツ集合によって表される。我々は、Webの構成要素（構造化文書とハイパーリンク）に基づき、画像とWebコンテンツの異なる関連性に応じて異なる種類のWeb文脈を定義する。Web文脈を検索することにより、検索ユーザは画像の使い方をWebから直接検索することが可能になる。また、Web文脈の中の一部として各画像を表示することにより、検索要求と画像の関連性を視覚的に捉えることができる。本稿では、画像のWeb文脈の抽出と可視化方法について説明する。

キーワード マルチメディアDB, 画像, Web文脈

Extraction and Visualization of Image Contexts from Web

Koji ZETTSTU^{†,††} and Katsumi TANAKA[†]

† Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan

†† Communications Research Laboratory

4-2-1 Nukui-Kitamachi, Koganei, Tokyo, 184-8795 Japan

E-mail: †{zettstu,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract The Web contains wide variety of images published by the general public. Thus, it is difficult for conventional image retrieval methods to retrieve images from the Web by specifying image contents precisely. Most of current search users first retrieve Web pages (or parts of Web pages) that contain related topics to the target image, then browse the Web pages for discovering the images with appropriate contents. This gap causes problems on search-ability and reliability of image retrieval. In order to fill the gap, we propose an approach for retrieving “Web contexts” of images, each of which represents with what kind of Web contents (e.g. texts and other images) an image is associated. A Web context is represented by a set of Web contents surrounding the image. We define three types of Web contexts for an image in accordance with different associations between an image and Web contents based on Web components (i.e. structured documents and hyperlinks). Our approach enables a search user to retrieve images by their usages directly from the Web. Besides, by showing each of result images as a part of the corresponding Web context, a search user can visually understand relevancy between query request and the result image. In this paper, we propose methods for extracting and visualizing Web contexts of images.

Key words Multimedia DB, image retrieval, Web context

1. はじめに

Webの普及に伴い、ビジネスや教育、報道など様々な分野

で、実用を目的とした様々な画像がWeb上に多数公開されている。近年、こうしたWeb上の画像を検索し、再利用することが試みられている。

従来の画像検索は、画像の物理的特徴量（色や形状）[1] [2] や画像の内容を記述したキーワードに基づいて、個々の画像を検索することを行ってきた。このアプローチでは、画像の中身を正確に特定することによって、検索精度の向上を図る。一方、Web 上には不特定多数の画像が存在し、個々の画像の中身を正確に特定することは非常に困難である。その為、今日 Web 上の画像を検索する際は、まず画像に関連する話題を含む Web ページや Web ページ内の部分領域を検索し、次にそれらの中に含まれる画像から、個々の画像の内容（被写体や構図など）を確認しながら、所望の画像を選択する。例えば、大自然を表す画像が欲しい場合、“自然”というキーワードで自然に関連する Web ページを検索し、更に検索結果の Web ページの中から大自然に関連する話題（森林や動物など）を含む段落を探す。そして、その段落の近辺に存在する画像の中から、適切な内容の画像を選択する（画像の中身は特定の動物の写真かもしれない）。このような検索を従来の画像検索で実現するには、個々の画像に対し“大自然”というキーワードを事前に記述しておかなければならない。しかし、画像の中身だけを頼りに、画像と関連する話題のキーワードを付与することは困難である。この問題は、Web 上の画像のように、不特定多数の画像を検索対象とする場合、内容記述と検索の間で検索対象にギャップが生じていることが原因である。即ち、内容記述では画像単体を検索対象に画像の中身を記述しているのに対し、検索では画像に関連する話題まで検索対象に含み、画像の周辺コンテンツ（画像を含む Web ページや段落のコンテンツ）を検索している。両者のギャップを埋めるには、画像が使われている話題を表す、画像の周辺コンテンツを検索できるようにする必要がある。その結果、今日、検索ユーザが手作業で行っている Web 上の画像の検索作業の負担を軽減することができると考えられる。本研究では、このような画像の周辺コンテンツのことを、Web 上における画像の使われ方を表す文脈という意味で、画像の“Web 文脈”と呼ぶ。

本研究の位置付けを図 1 に示す。本研究の目的は、画像の周辺コンテンツを使って画像の中身をより正確に特定し、従来の画像単位の検索精度を向上させることではない。中身を正確に特定することが困難な Web 上の画像に対し、個々の画像を検索するのではなく、画像の周辺コンテンツを検索対象とすることで、ある話題に関連した画像を検索できるようにすることである。例えば、Web 上の画像をクリップアートとして利用することを考えた場合、通常我々は、クリップアートを挿入したい箇所の内容に相応しい画像を検索する。従来のクリップ画像検索では、画像の様々な使われ方を想定して、各画像に関連する話題を予めキーワードで記述しておく必要があった。これに対し、画像の Web 文脈を検索すれば、挿入したい箇所の話題に関連した画像を、その周辺コンテンツと共に取得することができる。Web 文脈は画像がどのようなコンテンツと一緒に使われているかを表しているのので、挿入箇所に対する画像の相応しさを周辺コンテンツから直接評価することができる。

本研究の成果の概要は以下のとおりである：

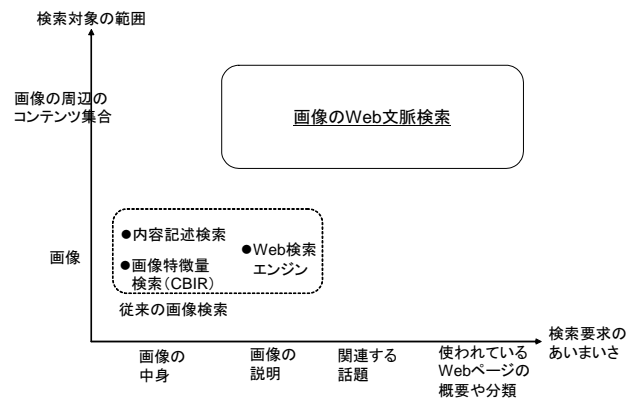


図 1 本アプローチの位置付け

Fig.1 Focused area of our approach.

- 画像が Web 上でどんなコンテンツと一緒に使われているかという文脈情報に基づいて、Web 上の画像を検索するアプローチを提案した。
- Web の構成要素（構造化文書とハイパーリンク）に基づき、画像の文脈情報を表す 3 種類の Web コンテンツ集合を、画像の“Web 文脈”として定義した。
- 画像の Web 文脈を抽出し可視化する方法を提案した。また、抽出された Web 文脈を検索する方法を提案した。
- 実際に抽出された Web 文脈の内容を検証し、提案したアプローチの評価を行った。

本稿は以下のように構成される。まず第 2 章で、関連研究と本研究との違いを説明する。第 3 章で、画像の Web 文脈の定義とその抽出方法について説明する。第 4 章では、抽出された Web 文脈を可視化する方法について説明する。第 5 章では、Web 上の画像の検索シナリオに沿って、画像の Web 文脈を検索する方法について述べる。第 6 章では、実証実験により抽出された Web 文脈の内容とその有用性に関する評価について述べる。最後に第 7 章でまとめと今後の課題について述べる。

2. 関連研究

Web 検索に“文脈”を用いるアイデアは、これまでも幾つかの研究で提案されてきた [3] [4]。それらの多くは、検索履歴や閲覧中の Web ページの内容に基づいて検索ユーザの興味に関連するキーワードを発見し、検索ユーザが発行する検索要求にそれらのキーワードを追加することによって検索対象（Web ページ）を絞り込むことを行っている。これらの手法における“文脈”は、検索要求に付加されたキーワード集合である。また、これらの手法は、検索対象に十分な内容記述が与えられていることを前提にしている。これに対し、我々のアプローチは、検索対象である画像に十分な内容記述が付けられない状況下で、画像の周辺コンテンツにまで検索対象を広げて検索を行う。我々のアプローチにおける“文脈”は、画像の周辺に存在する Web コンテンツ集合である。従って、従来の文脈を用いた Web 検索とは、文脈の内容も検索目的も異なる。

Google (<http://www.google.com>) や Altavista (<http://www.altavista.com>), FAST (<http://www.alltheweb.com>) などの Web 検索エンジンでは, Web ページ内の画像の周辺テキスト (surrounding text) に対してキーワード検索を行うことで, キーワードによる画像検索を実現している. また [5] では, 画像を含む Web ページのリンク元ページから画像のキーワードを抽出することを行っている. 内容記述が難しい Web 上の画像に対し, その周辺情報を使って検索するアプローチは, 本研究と基本的に同じである. しかし, これらの手法では, キーワードと画像がどのような関係にあるかが検索ユーザに示されず, 検索ユーザは検索結果に対し十分な確信を持つことができない. その為, 検索ユーザは, 画像が実際に使われている Web ページを訪れ, キーワードと画像がどのような関係にあるのかを逐一確認しなければならない. 我々のアプローチでは, 画像と周辺コンテンツの関係に基づいて Web 文脈を分類し, Web 文脈の内容を可視化することによって, キーワードと画像の関係を直感的に把握できるようにする.

3. 画像の Web 文脈の抽出

3.1 Web 文脈の種類

画像の Web 文脈は, その画像の実際の使用箇所の周辺に存在し, その画像と関連付けられた Web コンテンツ (テキストや他の画像など) の集合によって表される. 我々のアプローチでは, 画像と特定の周辺コンテンツとの関係を事前に定義した辞書に相当するものを用いず, Web の構成要素, 即ち HTML や XML などの文書構造とハイパーリンクによって画像と関連付けられた周辺コンテンツを使って Web 文脈を表現する. これらの構成要素に基づいて特定された Web 文脈は, Web ページの製作者によって付けられた, 最も汎用的な画像と周辺コンテンツとの関連性を表現していると考えられる.

我々は, Web の文書構造とリンク構造に基づいて, 以下の 3 種類の Web 文脈を定義した. ここで, 対象となる画像を “キー画像” と呼ぶ.

Surrounding context: Web ページ内で画像の前後に位置する Web コンテンツ集合によって, 画像に関する説明や一緒に扱われている内容を表す.

Structural context: Web ページの文書構造に沿って, キー画像の上位レベル (先祖) のタグ領域, 下位レベル (子孫) のタグ領域, および同じレベル (兄弟) のタグ領域に含まれるコンテンツ集合によって, 画像と関連付けられた話題を表す. 例えば, キー画像の上位レベルのタグ領域は, キー画像を引用している話題の概要を表すと考えられ, 一方, 下位レベルのタグ領域はその詳細を表すと考えられる. また同じレベルのタグ領域は, 画像を引用している話題と話題と並行して扱われている話題を表していると考えられる.

Referential context: ハイパーリンクに沿って, キー画像を参照するリンク系路上の Web コンテンツ集合を使って, キー画像が使われている Web ページに対する, 周囲の Web ページ

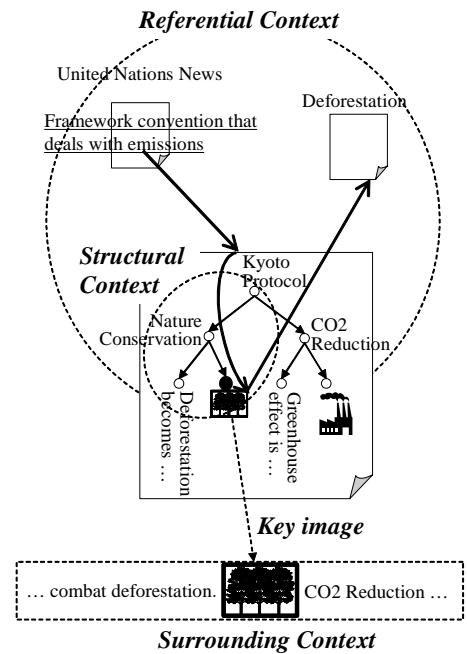


図 2 3 種類の画像の Web 文脈

Fig. 2 Three types of Web contexts for an image.

からの見られ方 (主題や分類, 評判など) を表す. リンク系路上の Web コンテンツとは, リンクアンカーと Web ページタイトルである. また, キー画像自身がリンクアンカーとして使われている場合, キー画像から参照されるリンク系路上の Web コンテンツ集合も抽出し, キー画像を参照するリンク経路と合わせた一連のリンク経路の中でキー画像の位置付けを表す.

図 2 に, これら 3 種類の文脈の例を示す. 図 2 において, surrounding context では, キー画像の前後に “... combat deforestation” や “CO2 Reduction ...” が存在する. この surrounding context からは, キー画像が森林破壊や CO2 削減に関するコンテンツと一緒に使われていることが分かる. 次に structural context では, キー画像の兄弟ノードに “Deforestation becomes ...” があり, 親ノードには “Nature Conservation”, さらに先祖ノードには “Kyoto Protocol” が存在する. この structural context からは, キー画像が京都議定書における自然保護の話題の中で, 森林破壊の話題の近くで使われていることが分かる. 最後に referential context では, キー画像が “Kyoto Protocol” というタイトルの Web ページで使われて (インライン画像として参照されて) おり, さらにそのページは “United Nations News” というタイトルの Web ページから “Framework convention that deals with emissions” というアンカーによって参照されている. さらにこの画像をアンカーとして “Deforestation” というタイトルの Web ページを参照している. この referential context からは, キー画像が国連のニュースで取り上げられる (重要な) 排ガス規制に関する会議の Web ページで使われており, 森林破壊に関係していることが分かる.

3.2 周辺コンテンツの抽出

Web 文脈の種類ごとに、キー画像の周辺から以下の Web コンテンツを抽出する：

Surrounding context： Web ページ内のコンテンツの書き下し順に沿って、キー画像の前後に位置するキーワードや画像をキー画像に近いものから順に抽出する。

Structural context： Web 文書構造の木構造表現に基づいて、キー画像の属するノードから、その親子・兄弟ノード、さらにそれらの親子・兄弟ノード、... という順にノードを辿り、各ノードに含まれるキーワードおよび画像を抽出する。

Referential context： キー画像からハイパーリンクを逆に辿って、キー画像を含む Web 文書のタイトル（インライン画像リンク）、その Web 文書をリンクしているリンクアンカー、そのリンクアンカーを含む Web 文書のタイトル、... という順に周辺コンテンツを抽出する。また、キー画像がリンクアンカーとして使われている場合、キー画像からリンクされている Web 文書のタイトルも取得する。多重リンクが存在する場合は、分岐してそれぞれのリンク経路から個別に周辺コンテンツを抽出する。

抽出された個々の Web コンテンツ（個々のキーワードや画像）は、“周辺オブジェクト”と呼ばれる。内容が同じでも出現位置が異なれば、別々の周辺オブジェクトとして扱われる。

ある画像の文脈情報は、その画像の近くにあり、かつ、その画像の周辺にのみ集中して現れる周辺オブジェクトによって特徴付けられると考えられる。従って、その様な周辺オブジェクトを一定以上含む極小範囲を、この画像の Web 文脈の抽出範囲とする。ある画像 I の文脈 $C(I)$ は、以下の条件を満たす画像 I の周辺範囲 R に含まれる周辺オブジェクトの集合として定義される：

- (1) 文脈 $C(I)$ に含まれるすべての周辺オブジェクトは、画像 I を含む周辺範囲 R の中に出現する。
- (2) 周辺範囲 R の中に出現する周辺オブジェクトの“文脈貢献度”の合計が既定の閾値を超えている。
- (3) 周辺範囲 R のどの部分範囲 R' も上記 1 および 2 を満たさない。

周辺オブジェクト o の文脈 $C(I)$ に対する文脈貢献度は、オブジェクト o が文脈 $C(I)$ をどの程度特徴付けているかを表し、オブジェクト o の表わすキーワードや画像が文脈 $C(I)$ のみに集中して出現する程高くなり、またオブジェクト o がキー画像 I の近くに出現する程高くなる。周辺オブジェクト o の文脈 $C(I)$ に対する文脈貢献度は、以下の式で求められる：

$$ccontrib(o, C(I)) = adensity(o, C(I)) \frac{b}{distance(o, I)}$$

ここで、 $adensity(o, C(I))$ は周辺オブジェクト o の文脈 $C(I)$ に対する“出現密度”であり、 $distance(o, I)$ は周辺オブジェク

ト o とキー画像 I の距離である。 b は定数である。周辺オブジェクト o の文脈 $C(I)$ に対する出現密度は、オブジェクト o が表わすキーワードや画像が文脈 $C(I)$ にどの程度集中的に出現しているかを示す尺度であり、以下の式によって求められる：

$$adensity(o, C(I)) = \frac{N_{E(o, C(I))}}{N_{E(o, U)}}$$

ここで、 $N_{E(o, C(I))}$ は文脈 $C(I)$ の中で周辺オブジェクト o と同一のキーワードや画像を表す周辺オブジェクトの出現数を示し、 $N_{E(o, U)}$ はその様な周辺オブジェクトの全てのキー画像の文脈集合 U における出現数を示す。例えば、もし 2 つの周辺オブジェクトが同じ語幹を持つキーワードやある閾値以上の類似度を持つ画像を表していれば、それらは同一であると見なされる。なお、事前に全てのキー画像の文脈を特定することは困難であるため、実際には U を、文脈の抽出対象となる Web コンテンツ全体（Web 文書集合）に含まれるキーワードおよび画像の集合によって近似する。一方、周辺オブジェクト o とキー画像 I との距離 $distance(o, I)$ は以下のように定義される：

Surrounding context における距離 キー画像 I と周辺オブジェクト o の間に存在するオブジェクト数（キーワードおよび画像の数）。

Structural context における距離： キー画像 I から周辺オブジェクト o を含むノードに至るまでの間に辿る親子・兄弟関係の数。

Referential context における距離： キー画像 I から周辺オブジェクト o を含む文書タイトルもしくはリンクアンカーに至るまでの間に辿る Web 文書およびリンクアンカーの数。

4. Web 文脈の可視化

抽出された Web 文脈は、キー画像とその周辺オブジェクト集合（キーワードおよび画像）から構成される。Web 文脈の可視化の目的は、周辺オブジェクトによって表される画像の説明や、話題、Web ページの分類などがどの程度キー画像と密接に関連しているかを視覚的に示すことである。その為には、以下の点が重要であると考えられる：

- (1) キー画像を中心に、各周辺オブジェクトとキー画像との距離の近さを示すことで、キー画像がどの周辺オブジェクトからどの程度強く関連しているかを直感的に把握できるようにする。
- (2) キー画像と周辺オブジェクト、および周辺オブジェクトとどうしの繋がりを連続的に示すことで、それらを集約した内容、例えば話題の流れや階層的な分類の中でのキー画像の位置付け（役割など）を把握できるようにする。
- (3) マルチメディア情報を含む Web 文脈では、内容や出現順序が部分的に異なる周辺オブジェクト集合を使って同じ内容を表す場合が多い。その為、Web 文脈を個別に表示するのではなく、複数の Web 文脈を同時に比較可能な形で表

示し、それらの類似性や差異に基づいて個々の文脈の特徴を正確に把握できるようにする。

テキストの文脈表示方法としてよく知られている KWIC (keyword in context) [6] は、あるキーワードの前後に出現する周辺テキストを様々な文章から抽出して一覧表示することで、キーワードの多様な文脈を一目で把握できるようにしている。KWIC の表示方法は、テキストを対象に、上記の要件 (3) を実現している。一方、要件 (1)(2) については、テキストの特性上、前後の周辺テキストを表示することで自然に実現されている。そこで、Web 文脈の表示では、KWIC の表示方法を以下の点で拡張した表示方法を提案する。

- 画像を含むマルチメディア情報の表示
- 文書構造やリンク構造によって構造化された周辺オブジェクトを比較可能な形式で表示

Web 文脈の表示では、文脈の種類ごとに、KWIC と同様に複数の Web 文脈を一覧表示する。一覧表示の各行は各々の Web 文脈を表し、キー画像を中心に周辺オブジェクトを前後に並べて表示する。これにより、Web 文脈の表示に関する要件 (3) を満たし、様々な文脈の比較が容易になる。一方、要件 (1) を満たすため、文脈の種類ごとに決まるキー画像と周辺オブジェクトとの距離 (第 3.2 節参照) に基づき、キー画像と距離の短い周辺オブジェクトほどキー画像の近くに表示されるように並べる。更に、Web 文脈の表示に関する要件 (2) を満たすため、文脈の種類ごとに、キー画像の前後に周辺オブジェクトを以下のように配置する：

Surrounding context の表示： Web ページ内のコンテンツの書き下し順に沿って、キー画像の前に位置する周辺オブジェクトをキー画像の前に表示し、キー画像の後ろに位置する周辺オブジェクトをキー画像の後ろに表示する。

Structural context の表示： Web ページの文書構造に沿って、キー画像の親 (先祖) ノードに位置する周辺オブジェクトをキー画像の前に表示し、キー画像の子 (子孫) ノードに位置する周辺オブジェクトをキー画像の後ろに表示する。

Referential context の表示： ハイパーリンクに沿って、キー画像に向かって参照するリンク系路上に存在する周辺オブジェクトをキー画像の前に、キー画像から参照されるリンク系路上に存在する周辺オブジェクトをキー画像の後ろに表示する。

図 3 に、図 2 の例を文脈の種類ごとに上記の方法で表示した例を示す。この様に周辺オブジェクトをキー画像の前後に配置することによって、前から後ろに向かって各文脈を見れば、説明の流れや話題の展開、階層的な分類の中でのキー画像の位置付けを容易に把握することができる。

文脈一覧の中から所望の Web 文脈を見つけ易くするために、文脈一覧はキー画像の類似度 (画像特徴量の類似度)、もしくは周辺オブジェクトの辞書順に基づいてソートされる。周辺オ

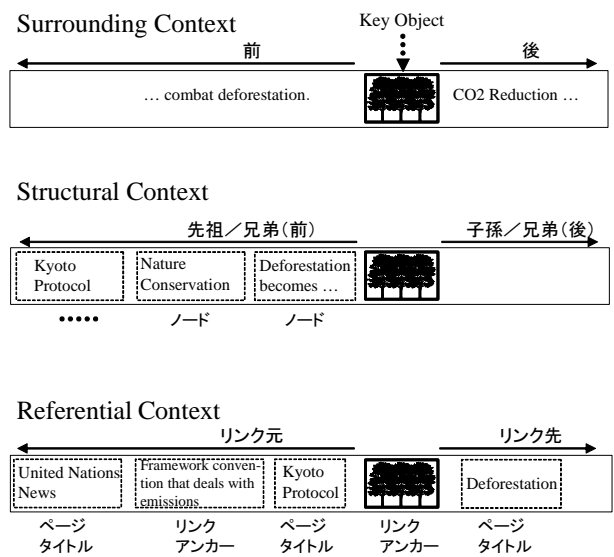


図 3 各種類の Web 文脈の表示
Fig. 3 Visualizing each type of Web context.

ブジェクトの辞書順に基づくソートでは、キー画像に近い周辺オブジェクトを順に第 1 ソートキー、第 2 ソートキー、... とし、周辺オブジェクトがキーワードの場合は単語の辞書順に、画像の場合は全てのキーワードの後に続けて表示する。

5. Web 文脈の検索

Web 文脈を用いた画像検索は、以下のシナリオに沿って行われる：

- (1) 画像の説明や関連する話題、使用されている Web ページの概要など、所望の画像に関連するキーワードを指定し、そのキーワード含む Web 文脈を検索する。
- (2) Web 文脈の内容とそこに含まれるキー画像の中身を確認しながら、所望の Web 文脈とキー画像を選択する。所望の Web 文脈が見つからなければ (1) を繰り返す。Web 文脈は見つかったが画像の中身が所望のものとは異なる場合、(3) に進む。
- (3) 同じような文脈で使われている、中身の異なる代替のキー画像を探すため、選択された Web 文脈と類似する他の Web 文脈を検索する。その後 (2) を繰り返す。

上記の検索シナリオのうち、ステップ (2) は、前章で述べた Web 文脈の一覧表示で可能である。本章では、ステップ (1) とステップ (3) について述べる。

5.1 Web 文脈のキーワード検索

検索シナリオのステップ (1) では、検索要求に指定されたキーワードにヒットする周辺オブジェクトを含む Web 文脈を検索する。検索結果に含まれる各 Web 文脈に対し、キーワードにヒットする周辺オブジェクトの文脈貢献度の合計を計算し、その値に基づいて検索結果のランキングを行う。その結果、検索要求に指定されたキーワードによって最も良く特徴付けられる Web 文脈が、高位にランキングされることになる。

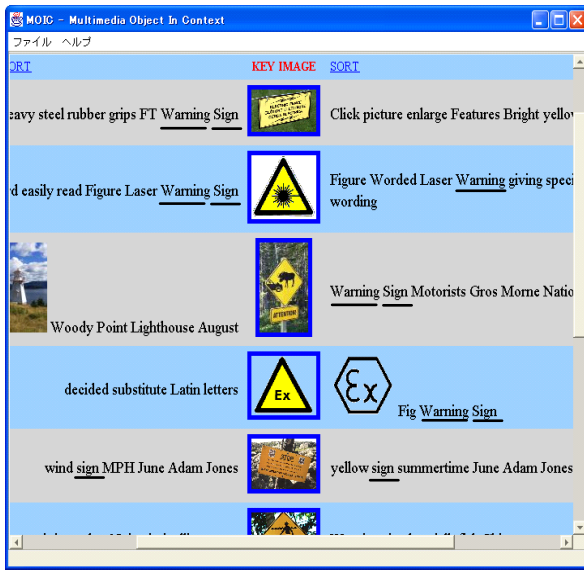


図 4 Web 文脈の類似検索例

Fig. 4 Search by Web context similarity.

5.2 Web 文脈の類似検索

検索シナリオのステップ (3) では、検索要求に指定された Web 文脈を特徴付ける周辺オブジェクトと同様の周辺オブジェクトで特徴付けられている他の Web 文脈を検索する。その為に、各 Web 文脈と検索要求に指定された Web 文脈との間で、周辺オブジェクトの文脈貢献度の類似性を評価する。

画像 I_i の Web 文脈 $C(I_i)$ と画像 I_j の Web 文脈 $C(I_j)$ の類似度は以下のようにして求められる。まず、 $E = \{e_1, e_2, \dots, e_n\}$ を、文脈 $C(I_i)$ と文脈 $C(I_j)$ に含まれるキーワードおよび画像の集合として表す。次に、文脈 $C(I_i)$ を、 E の各要素を成分とする特徴ベクトル $f(C(I_i)) = (v_1, v_2, \dots, v_n)$ で表わす。 v_k は、文脈 $C(I_i)$ の中で e_k と同一のキーワードもしくは画像を表わす周辺オブジェクトの文脈貢献度の合計である。同様にして、文脈 $C(I_j)$ も特徴ベクトル $f(C(I_j))$ で表わす。最終的に、文脈 $C(I_i)$ と文脈 $C(I_j)$ の類似度は、対応する特徴ベクトル $f(C(I_i))$ と $f(C(I_j))$ のコサイン相関値 [7] として算出される。

図 4 に、Web 文脈の類似検索を実際に行った検索結果の例を示す。図 4 は、文脈一覧の先頭に示された文脈を検索要求に指定し、これに類似する文脈を検索した結果を示している。文脈一覧は、検索要求（先頭の文脈）に対する各文脈の類似度に従ってソートされている。図 4 では、検索要求の文脈に含まれる単語（周辺コンテンツ）と同一の単語が下線で示されている。例えば、2 番目の文脈では検索要求の文脈と同じ単語が 3 回出現するが、3 番目の文脈では 2 回しか出現しない。そのため、3 番目の文脈は、2 番目の文脈より検索要求の文脈に対する類似度が低くなっている。一方、4 番目の文脈では、検索要求の文脈と同じ単語が 3 番目の文脈と同じ回数出現するが、出現位置がキー画像より遠くなっている。そのため、4 番目の文脈は、3 番目の文脈より検索要求の文脈に対する類似度が低くなっている。

6. 評価と考察

6.1 3 種類の Web 文脈の内容の評価

実際に Web から 3 種類の Web 文脈を抽出した結果を図 5 に示す。図 5 では、同じキー画像集合に対する 3 種類の Web 文脈が示されている。Surrounding context では、キー画像に関する説明を表すようなコンテンツが、比較的多く含まれている（“sign”, “warning sign”, など）。Referential context では、キー画像が使われている背景を表すようなコンテンツが多く含まれている（“Las Vegas”, “Newfoundland”, “Shark Museum” など）。Structural context は、他の 2 つの文脈の特徴を併せ持ったような内容を表している。明確な特徴が現れなかったのは、HTML が文書の論理構造をそれ程正確に反映していないことが原因と考えられる。今後増加すると予想される XML では、より厳密に文書の論理構造が反映されるため、状況は改善されていくと考えられる。

6.2 Web 文脈の有用性の評価

Web 上の画像を探索するユーザにとって画像の Web 文脈が実際にどの様に役立つのかを、実験によって検証した。まず、検索された Web 文脈の内容が画像と関連しているかどうかを、被験者に 5 段階（1:関連していない～ 5:関連している）で評価してもらった。15 名の被験者の評価値の平均は 3.73 で中立（評価値 3）より高く、提案した Web 文脈が画像と関連性する内容を示していることが分かった。被験者のコメントから、多くの被験者が、検索結果の文脈一覧において、キー画像と周辺オブジェクトに含まれる他の画像との組み合わせに基づいて、キー画像の分類や区別を視覚的に行っていることがわかった。

次に、特定の中身を持つ画像を選択する目的に Web 文脈が有効であるかどうかを、先ほどと同じ方法で評価した。その結果、評価値の平均は 2.80 で中立より低く、直接の効果が認められなかった。被験者のコメントから、多くの被験者にとって、特定の中身を持つ画像を選択する作業では、被験者自身が持つ画像のイメージを画像の中身と直接比較することが中心となり、画像の文脈情報が使われる余地が少ないことが分かった。しかし、そのような場合でも、逐一画像の中身を確認する前処理として、ふさわしくない使われ方をしている画像を排除するために Web 文脈が積極的に用いられていることも分かった。

7. まとめと今後の課題

本稿では、Web 上の画像に対し、画像の周辺コンテンツを使って画像の使用状況を表す Web 文脈のコンセプトを提案し、Web の構成要素に基づく 3 種類の Web 文脈（surrounding context, structural context, referential context）を定義した。また、Web 文脈の抽出方法と可視化方法、および検索方法について説明した。また実証実験によって、Web 文脈の内容とその有用性の評価を行った。

Web を対象とする場合、スケーラビリティも重要な課題である。大規模な実験を通して、提案手法のスケーラビリティを向上させることが、今後の課題として挙げられる。また、文脈検索の基本的な考え方は、Web のみに限定されるものではない。

| Literal Context | | Media Object In | Structural Context | Multimedia Object | Referential Context |
|------------------------------|--|--------------------------------|-----------------------------|---------------------------------|---|
| Key image SORT | | Key image | | Key image | |
| Mail friend July | | SIGNS forbid trespassing | July | SIGNS Site Test trespassing | Las Vegas SUN Photo Warning |
| Body Point Lighthouse August | | Warning Sign Motorists Cross | August extends | August Warning Motorists | Photos Newfoundland |
| Area Coober Pedy | | warning sign scrawled | Coober | scrawled warning | Coober Pedy |
| Adults XLXXL | | SWIM01 WARNING SIGN | height hats size | SWIM01 Shirt Sizes WARNING | Fox fuseum Caps TShirts Rodney Fox Shark Museum |
| rubber grips FT Warning Sign | | Click picture enlarge Features | Toll Free | | Electric Fencing Accessories |
| EST | | Watch head | | | Kevin Bentley Home page |
| June Adam Jones | | yellow sign summertime | December Monday | Quote day Watch megaton warning | WASHINGTON Adam Jones photos Warning |
| s vandalised detail | | | Warning photos Adam Warning | sign yellow Adam | WASHINGTON Adam Jones photos Warning |
| | | | | | Multiple Links |

図 5 抽出された Web 文脈の例

Fig. 5 Sample result of Web context extraction.

今後、一般の文書編集アプリケーションにおけるクリップアート検索などにも、同様の技術を展開していく。

謝 辞

本研究は、一部平成 14 年度科研費特定領域研究 (2) 「Web の意味構造に基づく新しい Web 検索サービス方式に関する研究」(課題番号: 14019048, 代表: 田中克己) による。ここに記し謝意を表します。また、本研究は、一部 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」による。ここに記し謝意を表します。また、本研究は、一部独立行政法人通信総合研究所の研究施設を利用して行われている。ここに記し謝意を表します。

文 献

- [1] Flickner, M., Sawhney, H. S., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P.: Query by Image and Video Content: The QBIC System, *IEEE Computer*, Vol. 28, No. 9, pp. 23–32 (1995).
- [2] Smith, J. R. and Chang, S. F.: Querying by Color Regions Using the VisualSEEK Content-based Visual Query System, *Intelligent Multimedia Information Retrieval*, AAAI Press (1997).
- [3] Lawrence, S.: Context in Web Search, *IEEE Data Engineering Bulletin*, Vol. 23, No. 3, pp. 25–32 (2000).
- [4] Finklstien, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, S., Wolfman, G. and Ruppim, E.: Placing Search in Context: The Concept Revisited, *ACM Transactions on Information Systems*, Vol. 20, No. 1, pp. 116–131 (2002).
- [5] Harmandas, V., Sanderson, M. and Dunlop, M. D.: Image retrieval by hypertext links, *Proceedings of the ACM SIGIR*

'97 Conference on Research and Development in Information Retrieval, pp. 296–303 (1997).

- [6] Luhn, H. P.: Keyword in Context Index for Technical Literature (KWIC Index), *American Documentation*, No. 11, pp. 288–295 (1960).
- [7] Salton, G. and McGill, M.: Introduction to Modern Information Retrieval, *McGraw Hill* (1983).