

テキスト処理による画像の多義性解消と事典検索サイトへの応用

藤井 敦 石川 徹也

筑波大学大学院図書館情報メディア研究科

fujii@slis.tsukuba.ac.jp

1 はじめに

World Wide Web 上の情報を組織化して百科事典のように利用するために、筆者らは検索サイト「CYCLONE」を構築した [4, 8]¹。ユーザは、キーワード入力、質問文入力、可視化インタフェースなどによって、約 70 万語を収録したコンテンツから、言葉や事柄に関する説明テキストを効率良く検索することができる。

現在、CYCLONE のコンテンツを状況に応じた形式に変換するための諸研究を行っている。

まず、一つの見出し語に関する複数の説明を統合し、簡略化するための要約手法を提案した [3]。

本稿は、CYCLONE コンテンツのメディア変換に焦点を当てる。見出し語によっては、テキストは必ずしも最適な説明の手段ではない。例えば、動植物や装置は、写真やビデオを見た方が分かりやすい場合がある。

「エンカルタ²」のようなマルチメディア百科事典は、一つの見出し語について、テキスト、音声、画像、映像による多面的な説明コンテンツを提供する。しかし、人手によるコンテンツ構築は高価である。

本研究は、Web 上の画像情報を再構成して、CYCLONE コンテンツのマルチメディア化を目的とする。具体的には、同じ見出し語の同じ意味に関するテキストと画像を自動的に対応付ける。そこで「ハブ(装置/蛇)」のような多義語を対象とした「画像の多義性解消手法」を提案し、実験によって有効性を評価する。

2 CYCLONE の概要

図 1 に基づいて CYCLONE について説明する。事典コンテンツを構築するオフライン処理と、ユーザがコンテンツを検索するオンライン処理に分けて説明する。

オフライン処理では、「新語検出」によって見出し語の候補を Web から収集する。「検索」は見出し語を含むページを検索する。「抽出」は、HTML タグを用いて、見出し語に関する段落をページから抽出する。

「組織化」は、a) 分野への関連度、b) 説明らしい言語表現を含むかどうか、c) 説明らしいレイアウトかどうか、d) ページの信頼度という 4 つの尺度を統合して段落のスコアを計算する。そして、スコアに基づいて段落を分野に分類し、順位付けを行う。

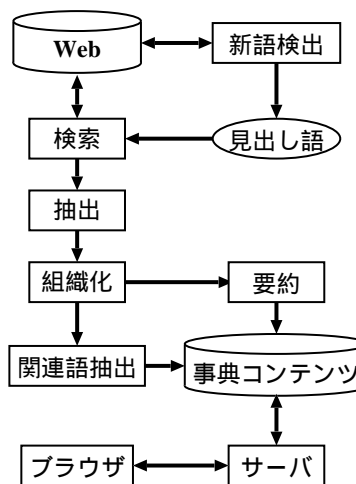


図 1: 検索サイト CYCLONE の概要

その結果「ハブ(装置/蛇)」のように分野によって意味が異なる多義語の説明を区別することができる。

分野の分類では、語の頻度情報に基づいて 22 の分野に関するモデルを作成し、段落に含まれる語を用いて、その段落に関連度が高い分野を選択する。クロスランゲージ社³の機械翻訳用辞書を用いて 20 専門分野のモデルを作成した。さらに、毎日新聞記事 CD-ROM のスポーツ面と芸能面から抽出した語を用いて「スポーツ」と「芸能」のモデルを作成した。

「関連語抽出」は、見出し語を特徴付ける関連語を取得する。関連語は、オンライン処理時に説明テキストを絞り込むために利用する。

「要約」は、一つの見出し語について、分野ごとに複数の説明段落を統合する。

オンライン処理では、ユーザが入力したキーワードが見出し語になかった場合に、文字列が部分一致する見出し語や同義語を提示することで、ユーザに必要な情報に誘導する。また、自然言語文による全文検索、質問応答、関連語グラフの可視化による検索機能がある。

本研究は、Web から収集した画像と事典コンテンツ中の説明テキストを対応付けることで、マルチメディア事典コンテンツを構築する。

¹<http://cyclone.slis.tsukuba.ac.jp/>

²<http://encarta.msn.com/>

³<http://www.crosslanguage.co.jp/>

3 画像の多義性解消

3.1 概要

本研究では、画像の内容解析を行わずに、テキスト解析によって画像の多義性を解消する。画像検索の研究においても、画像に付与されたキャプションだけを用いるテキスト駆動型の検索手法 [7] が提案されている。

Web上の画像ファイルは、通常、HTMLファイルから、 や <A> などのタグによってリンクされている。そこで、リンク元のHTMLファイルに含まれるテキストを「擬似キャプション」として利用する。

既存の検索エンジン (Google や Yahoo! など) も擬似キャプションによるテキスト駆動型の画像検索を行う。しかし「ハブ」のような多義語を入力すると、装置や蛇の画像を区別することができず、混在して検索されてしまう。入力された「ハブ」と擬似キャプション中の「ハブ」が、どの語義で使われているかを自動的に特定することができないためである。

この問題を解決するためには、単語の多義性解消が有効である。多義性解消には、辞書の定義文を利用する手法 [5] がある。ある語義に関する定義文は、その語義を特徴付ける手掛かり語を含んでいる。

例えば「ハブ (装置)」の定義文に「LAN」や「ケーブル」などの手掛かり語が含まれていたとする。擬似キャプションに「ハブ」「LAN」「ケーブル」が現れれば、当キャプション中の「ハブ」は装置の語義で使われている可能性が高い。さらに、リンク先の画像は、装置に関する画像である可能性が高い。すなわち、テキスト情報を用いて、画像の多義性を間接的に解消することができる。

CYCLONEの説明テキストは語義 (分野) によって分類されている。説明テキストを辞書の定義文として利用すれば、上記手法によって画像の多義性解消が可能である。具体的には、ある説明テキストに含まれる手掛かり語を含むようなHTMLファイルを検索し、そこからリンクされている画像を説明テキストと対応付ける。

既存の辞書では収録語数が限られるのに対して、CYCLONEは新語や専門用語を多数収録しているため、網羅性の点において有利である。

3.2 手法

本研究で提案する手法は、以下の手順で画像の多義性を解消し、CYCLONEの説明テキストと対応付ける。

1. 対象の見出し語を含むHTMLファイルとリンク先の画像ファイルを検索する。
2. その見出し語について、ある語義 (分野) に対応する説明テキストを入力し、類似する候補を上記のHTMLファイル群から検索する。
3. 検索された候補からリンクされている画像を取得して、説明テキストと対応付ける。

対象の見出し語に関する画像情報をテキストによって検索するためには、GIF形式やJPEG形式の画像ファイルとリンク元HTMLファイルの対をWebから網羅的に収集し、事前に索引付けを行う必要がある。

しかし、この作業は高価である。そこで、限定された見出し語に対して、Yahoo! Japan⁴を利用して関連する画像ファイルとリンク元HTMLファイルを検索し、実験を行っている。

次に、収集されたHTMLファイル群からテキスト情報 (擬似キャプション) を抽出し、索引付けを行う。その結果、テキスト検索によって、特定のHTMLファイルとリンク先の画像を取得することが可能になる。

入力となる検索質問は、キーワードではなく、テキストである。そこで、完全一致 (exact match) 手法ではなく、最良一致 (best match) 手法に基づく類似文書検索を行う。ベクトル空間モデルや確率型モデルなどの最良一致手法は全て利用可能である。現在は実験的に確率型モデル [6] を利用して、説明テキストに関連するHTMLファイルを検索し、スコアに基づいて順位付けを行う。

索引付けでは、「茶釜⁵」を用いて擬似キャプションを形態素解析し、名詞や動詞などの内容語を抽出して索引語とする。検索質問となる説明テキストからも同様の手法で索引語を抽出し、検索に利用する。

しかし、擬似キャプション中の全テキストがリンク先の画像に関連しているとは限らない。例えば、複数の話題に関するページでは、ページ内の位置によっては、リンク先画像と無関係なテキストもある。

一般に、アンカー (や <A> などによる画像へのリンク) 周辺のテキストはリンク先の画像への関連が高く、アンカーからの距離が遠くなるほど関連は低くなる。そこで、索引付けの段階で、アンカーからの距離に基づいて索引語の重みを計算する。

原理的には、アンカーからの距離が遠くなるほど小さい値を与えるような減少関数を用いて索引語の重みを計算すればよい。実際は、アンカーから N 文字以内の範囲にある索引語の重みを一律 M 倍する ($M > 1$)。 M と N はパラメタである。

テキスト駆動型の画像検索 [7] は人間が注意深く記述したキャプションを利用する。それに対して、本手法はHTMLファイル中の不要なテキストに対して頑健である。

図2は、見出し語「ハブ」に対する多義性解消と対応付けの結果である。ここでは、装置と蛇に関する画像が区別され、適切な説明テキストに対応付けられている。

本手法において、入力テキストと検索対象テキストを交換することができる。すなわち、「ハブ」を含む擬似キャプションを検索質問として、類似する説明テキストを検索し、対応付けを行うことも可能である。しかし、本研究の目的は、説明テキストに対応する画像を最低1件見つけることであり、「ハブ」を含む擬似キャプション全てを説明テキストに対応付けることではない。目的に応じて、入力と検索対象を決定することが重要である。

⁴<http://www.yahoo.co.jp/>

⁵<http://chasen.aist-nara.ac.jp/>



図 2: 「ハブ」に対する説明テキストと画像の対応付け結果

3.3 応用

本研究の主たる目的は、説明テキストと画像を対応付けて、ユーザに多面的な説明を提供することである。しかし、説明テキストと画像を対応付けることで、いくつかの応用が考えられ、副次的な効果を得ることができる。

まず、画像を「絞り込み情報」として利用することが可能である。CYCLONEでは関連語を絞り込み情報として提供している。すなわち、対象の見出し語に関する説明テキストのうち、ユーザが選択した関連語を含むテキストだけを提示する。

画像による絞り込みも目的は同じである。ただし、技術的な差異がある。ユーザは、説明に対応付けられた複数の画像から、情報要求に関連する画像を選択する。すると、その画像に対応する擬似キャプションが検索質問となり、CYCLONEの説明テキストが検索され、関連する説明テキストが提示される。

その結果、説明テキストと画像の対応付けに誤りがあっても、ユーザの指定によって適切な説明テキストに絞り込むことができる。ユーザが画像を見て、欲しい情報が否かを容易に判断できる場合に有効な手段である。

もう一つの応用は、「マルチメディア質問応答」である。CYCLONEは、自然言語による質問文が入力されると、回答になる見出し語や関連する見出し語を検索する。

そこで、検索された見出し語に対応付けられた画像を提示すれば、テキストによる質問に対して画像で回答することが可能である。

4 評価実験

提案手法の有効性を評価するために、テストコレクションを作成して実験を行った。テスト用の入力として、CYCLONEに見出し語として収録されている多義語で、かつ複数の語義に対して画像で提示することに意味がある言葉を選択した。以下に示す見出し語 22 件をテスト入力として使用した。

アパッチ、アポロ、カブチーノ、サーバー、サイクロン、サラトガ、サルサ、シーザー、タイタン、ドライバー、ハブ、パール、パイソン、ファントム、ヘラクレス、ポセイドン、マウス、マウスピース、マグナム、ミール、ミキサー、モンブラン

各見出し語に対して、Yahoo! Japan を用いて上位 100 件の画像とリンク元の HTML ファイルを収集した。各画像に対して、リンク元 HTML ファイルの内容を適宜参照しながら、人手で語義を付与した。見出し語と関係のない画像には「ノイズ」という語義を付与した。

各見出し語を用いて CYCLONE を検索し、各分野から上位 1 件の説明テキストを抽出し、検索質問として利用した。ただし、画像に付与された語義と無関係な説明テキストは削除した。その結果、「マウス」と「ポセイドン」に対して、「ノイズ」以外の語義はそれぞれ 1 つとなった。

検索質問数は 198 となった。見出し語あたりの平均語義数は「ノイズ」も語義に含めると 3.6 だった。

本実験の評価尺度は、検索質問 198 件のうち、どれだけの質問に対して適切な画像を対応付けることができたかを示す「精度」である。以下の 3 手法を実装し、精度を比較した。

- Yahoo! Japan で検索された順番に画像を整列する。画像の多義性を解消することができないため、同じ見出し語に対応する複数の語義には、同じ画像リストが出力される（ベースライン）。
- 本手法を利用して、画像の多義性を解消する。ただし、アンカーからの位置に基づく索引語の重み付けは適用しない（本手法 1）。
- 本手法で、かつアンカーからの位置に基づく索引語の重み付けを適用する（本手法 2）。予備実験を通して、 $M = 7$ 、 $N = 400$ とした。

ただし、検索質問ごとに複数の画像が順位付きでリスト化されるため、実際は、出力と見なす画像の件数（順位）と精度のグラフを描いた。

図 3 は、各手法の順位-精度グラフである。図 3 より、「本手法 1」と「本手法 2」は、順位によらず、ベースライン手法の精度を向上させた。画像の多義性解消手法は、画像と説明テキストの対応付けに効果的であった。

「本手法 2」は「本手法 1」の精度を向上させた。アンカーからの位置による重み付けも、画像と説明テキストの対応付けに効果的であった。上位 10 件の画像を出力した場合、「本手法 2」は 93.6% の検索質問（説明テキスト）に対して適切な画像を対応付けることができた。

本研究の提案手法は、「マウス」や「ポセイドン」のように多義性がない見出し語についても、不要な画像を削除するフィルタリングとして機能することが分かった。

5 おわりに

本研究は、事典検索サイト CYCLONE のコンテンツをマルチメディア化することを目的とし、テキスト解析に基づく画像の多義性解消手法を提案した。

画像の内容解析とテキストを併用した画像検索 [1] や、画像の内容解析を行い、その画像に対するキャプションに含まれる単語の多義性を解消する手法 [2] が提案されている。これらの手法を応用して、画像の多義性解消精度を向上させることが今後の研究課題である。

擬似キャプションを利用した検索手法は、Web 上の音声ファイルやビデオファイルの検索にも応用が可能である。今後は、実験によって有効性を評価する必要がある。

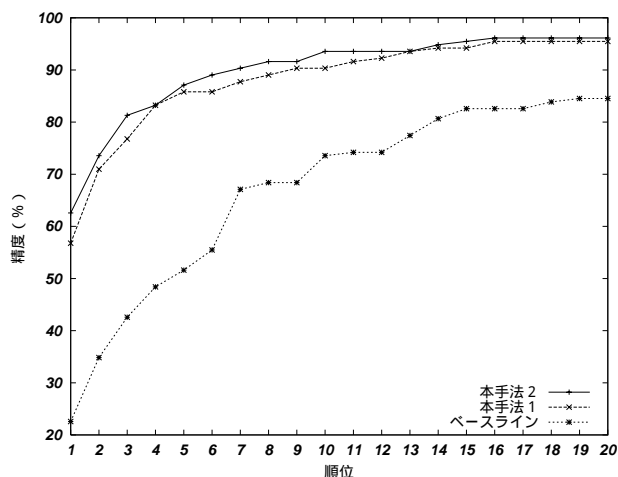


図 3: 順位-精度グラフ

謝辞

本研究の一部は、文部科学省科学研究費補助金特定領域研究 (2) (課題番号 16016206) によって実施された。

参考文献

- [1] Carmen Alvarez, Ahmed Id Oumohmed, Max Mignotte, and Jian-Yun Nie. Toward cross-language and cross-media image retrieval. In *Working Notes for the CLEF 2004 Workshop*, 2004.
- [2] Kobus Barnard, Matthew Johnson, and David Forsyth. Word sense disambiguation with pictures. In *Proceedings of the HTL-NAACL 2003 Workshop on Learning Word Meaning from Non-Linguistic Data*, 2003.
- [3] Atsushi Fujii and Tetsuya Ishikawa. Summarizing encyclopedic term descriptions on the Web. In *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 645–651, 2004.
- [4] Atsushi Fujii and Tetsuya Ishikawa. Cyclone: An encyclopedic Web search site. In *Proceedings of the 14th International World Wide Web Conference*, 2005. (To appear).
- [5] Joe A. Guthrie, Louise Guthrie, Yorick Wilks, and Homa Aidinejad. Subject-dependent co-occurrence and word sense disambiguation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 146–152, 1991.
- [6] S.E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 232–241, 1994.
- [7] Alan F. Smeaton and Ian Quigley. Experiments on using semantic distances between words in image caption retrieval. In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 174–180, 1996.
- [8] 藤井敦, 伊藤克亘, 石川徹也. Web マイニングによる事典的コンテンツの構築と多様なアクセス手法. 電子情報通信学会技術研究報告, DE2004-6, pp. 31–36, 2004.