Image Retrieval and Disambiguation for Encyclopedic Web Search

Atsushi Fujii and Tetsuya Ishikawa
Graduate School of Library, Information and Media Studies
University of Tsukuba
1-2 Kasuga, Tsukuba, 305-8550, Japan

Abstract

To produce multimedia encyclopedic content, we propose a method to search the Web for images associated with a specific word sense. We use text in an HTML file which links to an image as a pseudo-caption for the image and perform text-based indexing and retrieval. We use term descriptions in a Web search site called "CYCLONE" as queries and correspond images and texts based on word senses.

1 Introduction

The World Wide Web, which contains an enormous volume of up-to-date information, is a promising source to obtain encyclopedic knowledge. It has become common to consult the Web for specific keywords, instead of using conventional dictionaries and encyclopedias. However, existing Web search engines often retrieve extraneous pages containing low-quality, unreliable, and misleading information.

Fujii and Ishikawa [2001] proposed an automatic method to extract term descriptions from the Web and classify multiple descriptions into domains and word senses. Using this method, Fujii and Ishikawa have built a Web search site called "CYCLONE"[1], in which a user can efficiently obtain an encyclopedic term description in a specific word sense. Over 700,000 Japanese terms are currently indexed as headwords.

However, to explain certain headwords, specifically those related to an entity, such as devices and animals, it is effective to present a picture depicting the entity, in addition to text descriptions.

In this paper, we propose a method to integrate images on the Web and text descriptions in CYCLONE. We resolve the ambiguity of the meaning of an image by text analysis, so that images for a polysemous word, such as "hub (network device and center of wheel)", are classified on the basis of word senses. Our research is a step toward the automatic compilation of multimedia encyclopedias, such as Encarta[2].

2 Description-based Image Disambiguation

Existing search engines, such as Google and Yahoo!, use texts in HTML files for retrieval purposes, instead of performing content-based image retrieval. Given a text query, an image file, such as GIF and JPEG files, linked from an HTML file including one or more query terms is selected as a candidate image. In other words, text content in an HTML file is used as a "pseudo-caption" for a target image.

However, the text-based image retrieval cannot distinguish images depicting different entities, if a query term is polysemous. For example, images of network devices and images of axles can potentially be retrieved together in response to the query "hub".

To resolve this problem, we use a text description for a specific word sense. Term descriptions in CYCLONE, which are organized on the basis of word senses, provide informative contexts for word sense disambiguation. For example, if an HTML file includes words "LAN" and "cable", an image linked from this HTML file is likely to depict a network device more than an axle.

We need to crawl the Web and cache a large number of images and HTML files linking to those images. We shall call these HTML files "hyper HTML files". However, this process is computationally expensive. We experimentally use Yahoo! Japan[3] to obtain pairs of an image and its hyper HTML file, by submitting a target term (e.g., "hub") as a query. We discard all HTML tags in the hyper HTML files and extract the text content, from which we produce a text index.

For indexing and retrieval purposes, any best-match text retrieval method, such as the vector space model and probabilistic model, can be used. We experimentally use Okapi BM25. We use content words, such as nouns, extracted from text as index terms, and perform word-based indexing. We use the ChaSen morphological analyzer[4] to extract content words, because Japanese sentences lack lexical segmentation. The same method is used to extract terms from queries.

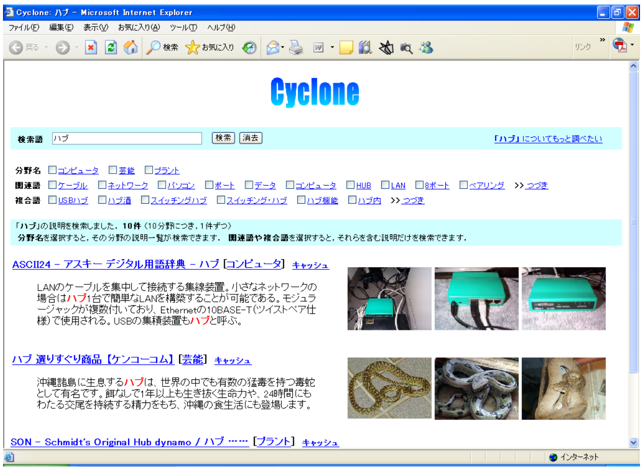
However, unlike a text-based image retrieval which uses well-organized captions [Smeaton and Quigley, 1996], not all words in an HTML file are related to an image. We use different term weights depending on the region in an HTML file. In principle, a decreasing function which assigns a value to each word depending on the proximity between the word and an anchor (i.e., <IMG> and <A> in HTML) to a target image can be used. In practice, we multiply the weight of a word M

[1] http://cyclone.slis.tsukuba.ac.jp/

[2] http://encarta.msn.com/

[3] http://www.yahoo.co.jp/

[4] http://chasen.aist-nara.ac.jp/


Cyclone: ハブ - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る
検索
お気に入り
リンク
インターネット

Cyclone

検索語 ハブ 検索 消去 「ハブ」についてもっと調べたい

分野名 □コンピュータ □芸能 □プラント

関連語 □ケーブル □ネットワーク □パソコン □ポート □データ □コンピュータ □HUB □LAN □8ポート □ペアリング >> つづき

複合語 □USBハブ □ハブ酒 □スイッチングハブ □スイッチング・ハブ □ハブ機能 □ハブ内 >> つづき

「**ハブ**」の説明を検索しました. **10件** (10分野につき, 1件ずつ)
分野名を選択すると, その分野の説明一覧が検索できます. **関連語や複合語**を選択すると, それらを含む説明だけを検索できます.

ASCII24 - アスキー デジタル用語辞典 - ハブ [コンピュータ] キャッシュ

LANのケーブルを集中して接続する集線装置。小さなネットワークの場合はハブ1台で簡単なLANを構築することが可能である。モジュラージャックが複数付いており、Ethernetの10BASE-T(ツイストペア仕様)で使用される。USBの集積装置もハブと呼ぶ。

ハブ 選りすぐり商品【ケンコーコム】[芸能] キャッシュ

沖縄諸島に生息するハブは、世界の中でも有数の猛毒を持つ毒蛇として有名です。餌なしで1年以上も生き抜く生命力や、24時間にもわたる交尾を持続する精力をもち、沖縄の食生活にも登場します。

SON - Schmidt's Original Hub dynamo / ハブ …… [プラント] キャッシュ

Figure 1: Example text descriptions and images for "*habu*".

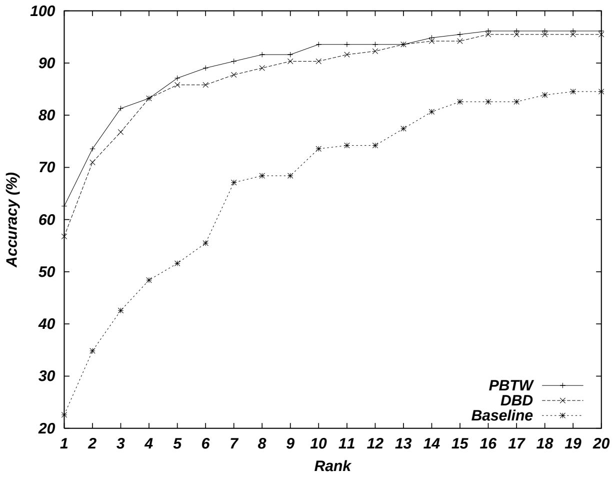


Figure 2: Rank-accuracy graphs for different methods.

times, if the number of characters between the word and the anchor to the image is less than N. We shall call this method "proximity-based term weighting (PBTW)". The values of M and N are determined empirically in Section 3.

Figure 1 depicts a successful example result for the Japanese term "*habu*", which is associated with multiple word senses, such as a network device, snake in *Okinawa*, airport, and center. In Figure 1, the first two paragraphs describe device and snake, respectively. The top three image candidates are associated with each paragraph.

3 Experimentation

To evaluate the accuracy of our method, we produced a test collection. First, we collected polysemous words used as test terms. For each test term, at least two word senses must be able to be depicted by image, because our purpose is to disambiguate the meaning of images. We collected 22 test terms.

Second, for each test term, we used Yahoo! Japan to search the Web for the top one hundred images and their hyper HTML files. Third, each image was manually annotated with a word sense, disregarding as to whether or not the sense is included in CYCLONE. The annotator was able to read the content of a hyper HTML file for the decision, if necessary. The images not associated with any word sense are annotated with "irrelevant". Fourth, for each test term, a query is produced from each description in CYCLONE. We used the top descriptions classified into each domain. The descriptions not associated with any word sense annotated to the images were discarded. The total number of queries was 155.

We compared the following three methods:

- a baseline method, which sorts candidate images for each query according to the ranking in Yahoo! Japan ("Baseline"),
- our method (the description-based disambiguation method), which uses descriptions in CYCLONE as queries ("DBD"),
- our method, which uses descriptions in CYCLONE as queries and performs the proximity-based term weighting method ("PBTW").

For the baseline method, which did not perform image disambiguation, the same list of candidate images was associated with the different senses for each test term. Through a preliminary experiment, we set $M = 7$ and $N = 400$ for PBTW.

Figure 2 shows the accuracy for each method in different ranks. The accuracy for rank r is the number of queries for which a correct image was found in the top r candidates and the total number of queries. DBD and PBTW significantly improved on the accuracy of the baseline method, irrespective of the rank. PBTW improved on the accuracy of DBD. As predicted, the words in close proximity to an anchor for a target image were important to disambiguate the meaning of the image. PBTW retrieved at least one correct image in the top ten candidates for 93.6% of the queries.

4 Conclusion

To compile multimedia encyclopedic content on the Web, we proposed a method to associate text descriptions in CYCLONE and image files based on word senses, and showed its effectiveness by means of experiments.

References

[Fujii and Ishikawa, 2001] Atsushi Fujii and Tetsuya Ishikawa. Organizing encyclopedic knowledge based on the Web and its application to question answering. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 196–203, 2001.

[Smeaton and Quigley, 1996] 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*, pages 174–180, 1996.