Exploiting Image Contents in Web Search

Zhi-Hua Zhou and **Hong-Bin Dai**
National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
{zhouzh, daihb}@lamda.nju.edu.cn

Abstract

Web search is a challenging task. Previous research mainly exploits texts on the Web pages or link information between the pages, while multimedia information is largely ignored. This paper proposes a new framework for Web search, which exploits image contents to help improve the search performance. In this framework, candidate images are retrieved at first by considering their associated text information. Then, images related to the query are identified by analyzing the density of the visual feature space. After that, an image-based rank of the Web pages is generated, which is combined with the traditional keyword-based search result to produce the final search result. Experiments demonstrate the promise of the proposed framework.

1 Introduction

Web has become the largest information repository over the world. Due to its huge volume, users easily feel lost and difficult to achieve what they need from the repository. In consequence, research on Web search has attracted more and more attention.

Earlier works on Web search mainly focused on exploiting *text* information since most Web pages contain abundant text contents. In general, each Web page is treated as a text document and Web is considered as a huge document collection. Thus, techniques from conventional text retrieval, such as Vector Space Model (VSM) [Raghavan and Wong, 1986] and TFIDF [Salton and Buckley, 1988], were employed to search for documents with text contents similar to the query. Later, considering that Web is a hyperlinked environment and the links between Web pages encode the importance of the pages to some extent, *link* information is utilized in many approaches, such as PAGERANK [Brin and Page, 1998] and HITS [Kleinberg, 1999], as a helpful evidence to make the pages which are relevant to the query stand out.

It is noteworthy that in addition to text and link, other modalities of information, such as image, video and audio, also convey rich and important information about contents of Web pages. In fact, Web is a multimedia environment. When possible, the author of a Web page would like to describe the topic using different information modalities with different advantages; a user would like to exploit the most convenient and prominent information modality to comprehend a Web page. For example, an author has put a picture of *tiger* to the Web page shown in Figure 1; when a user browses this page, even without reading the texts on the page, he or she would quickly understand that the page is about the animal *tiger*, since the first impression of the page comes from a first glimpse at the vivid image. However, although the importance of information modalities other than text and link has been well recognized, only a few works [Woodruff *et al.*, 2001; Xue *et al.*, 2006] have tried to use them in Web search. In particular, few works address the issue of exploiting visual contents of images to help improve Web search performance.

Figure 1: A Web page on *tiger*.

This paper is motivated by the fact that if a Web page is judged to be relevant to a query in multiple information modalities simultaneously, then the page has bigger chance to be really relevant than another page which is judged to be relevant in only the text modality. For example, when a user searches for the animal *tiger*, a Web page containing a tiger photo as well as the word 'tiger' would have a higher chance to satisfy the query than a page containing the word 'tiger' without a tiger picture. In other words, related images in Web pages are regarded as additional evidence in judging the relevance of those pages in the search process.

In the proposed WEBSEIC (Web Search Exploiting Image

Contents) framework, candidate images are retrieved from Web pages by considering the text information associated to them at first. Then, by analyzing the visual contents of these images, some images which have a high chance to be relevant to the query are identified. Based on these related images, an *image-based rank* of the Web pages is generated, which is then merged with a traditional *keyword-based rank* of the Web pages to produce the final search result.

The rest of this paper is organized as follows. Section 2 briefly introduces some related works. Section 3 presents the WEBSEIC framework. Section 4 reports on the experiments. Finally, Section 5 concludes.

2 Related Works

Traditional Web search is based on query term matching since queries are composed of textual keywords while Web pages usually contain abundant texts. Statistics such as term occurrence frequency are often calculated to examine the text contents of Web pages and many techniques from text retrieval such as VSM [Raghavan and Wong, 1986] and TFIDF [Salton and Buckley, 1988] have been utilized. Several new techniques have been developed recently. Considering that typical documents may contain multiple drifting topics and have varying lengths, Cai et al. [2004] took a block of a Web page as a single semantic unit instead of the whole page and proposed a block-based Web search strategy. Ntoulas et al. [2005] proposed a new Web search engine called *Infocious* which improves the Web search through linguistic analysis to resolve the content ambiguity of Web pages and to rate the quality of the retrieved Web page.

The links between Web pages have also been utilized in Web search. Page et al. [1998] indicated that Web is a hyperlinked environment and the links between Web pages encode the importance of the pages to some extent. Based on this recognition, the PAGERANK [Brin and Page, 1998] algorithm has achieved great success. Kleinberg [1999] investigated the symmetric relationship between Web pages and indicated that 'good' pages should have a lot of links coming from 'good' pages. Based on this recognition, the HITS algorithm [Kleinberg, 1999] searches for 'good' pages recursively. Silva et al. [2000] have shown that Web search performance can be improved by using both the text and link information simultaneously.

Multimedia information retrieval has been studied for many years. In particular, many researchers have investigated image retrieval from Web [Frankel *et al.*, 1996; Smith and Chang, 1996; Sclaroff *et al.*, 1997; Mukherjea *et al.*, 1999]. However, only a few works have been done in exploiting the image modality in Web search. Woodruff et al. [2001] proposed to use an enhanced thumbnail to help the user quickly find the Web pages he or she expects. This enhanced thumbnail is an image, where the whole Web page is resized to fit into this image and important text is highlighted in order to be readable for the user. Xue et al. [2006] proposed to present *image snippets* along with text snippets to the user such that it is much easier and more accurate for the user to identify the Web pages he or she expects and to reformulate the initial query. An image snippet of a Web page is one of the images in the page that is both representative of the theme of the page and at the same time closely related to the query. Note that those works mainly use the image modality to help present the search results, while few works have addressed the issue of exploiting visual contents of images in the Web search process.

As images and texts in the same Web page are often related, there are usually some text information partially describing the image contents, such as image names or words around images, which can be regarded as natural descriptions for images in Web pages [Harmandas *et al.*, 1997]. Many works have tried to exploit the relationship between texts and images in the same page to identify related images. Coelho et al. [2004] extended a Bayesian network for this purpose. By combining distinct sources of textual descriptions associated with images, their method achieves better performance than using only a single source of textual descriptions. Wang et al. [2004] regarded web images as one type of objects and the surrounding texts as another type. By constructing the link structure and exploiting the mutual reinforcement between them, both visual content and text description are combined for identifying related Web images.

3 The WEBSEIC Framework

The WEBSEIC framework is shown in Figure 2. Given a query, WEBSEIC executes a typical keyword-based search [1] to obtain a traditional keyword-based rank of Web pages. Then, it retrieves candidate images from Web pages. Images having good chance to be relevant to the query are identified and used to generate an image-based rank of Web pages. Finally, the image-based and keyword-based ranks are combined to produce the final rank of the Web pages.

The basic process of WEBSEIC is similar to that of traditional Web search except that some additional operations are required, as highlighted in the grey boxes in Figure 2. The following subsections will describe how to realize these operations.

3.1 Retrieving Candidate Images

Several existing approaches can be used to retrieve the candidate related images. In this paper, the approach developed by Coelho et al. [2004] is employed. This approach attempts to retrieve images related to a query from Web pages by combining distinct sources of textual descriptions, which generalizes a Bayesian network-based text retrieval method presented by Ribeiro-Neto and Muntz [1996].

Let $\boldsymbol{q} = (w_{q1}, w_{q2}, \cdots, w_{qt})$ denote the query which is represented as a t-dimensional textual vector, and $\boldsymbol{I}_i$ $(i = 1, \cdots, N)$ denote the images for which the distinct pieces of evidence are combined through a disjunctive operator. Let $\boldsymbol{d}_i$ $(i = 1, \cdots, N)$ denote the text evidences extracted from description tags including filenames, ALT attribute of IMG tag and anchors, and $\boldsymbol{s}_i$ $(i = 1, \cdots, N)$ denote the

[1] Here we use *keyword-based search* to denote current search strategies where a user inputs some keywords to express his/her information requirements. Note that keyword-based search usually uses text as well as link information and is not an equivalence to the simple *query term matching*.

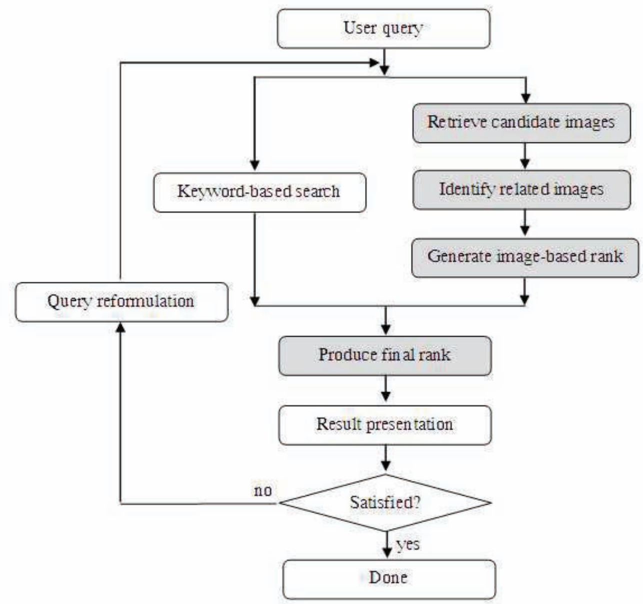

Figure 2: The WEBSEIC framework.

Figure 3: The top 10 Web images retrieved for the query *tiger animal* after executing the process described in Section 3.1.

text evidences extracted using surrounding passages. $\boldsymbol{d}_i$ and $\boldsymbol{s}_i$ are all t-dimensional textual vectors, i.e. $\boldsymbol{d}_i = (w_{d_i,1}, w_{d_i,2}, \cdots, w_{d_i,t})$ and $\boldsymbol{s}_i = (w_{s_i,1}, w_{s_i,2}, \cdots, w_{s_i,t})$.

Given $\boldsymbol{q}$, the conditional probability of each source of evidence being observed can be computed according to Eq. 1.

$$P(\boldsymbol{e}_j|\boldsymbol{q}) = \frac{\sum_{k=1}^{t} w_{jk} w_{qk}}{\sqrt{\sum_{k=1}^{t} w_{jk}^2}\sqrt{\sum_{k=1}^{t} w_{qk}^2}} \tag{1}$$

where $\boldsymbol{e}_j$ could be $\boldsymbol{d}_j$ or $\boldsymbol{s}_j$, each representing the corresponding source of text evidences associated with image $\boldsymbol{I}_j$. For convenience, denote these conditional probabilities by Pd_{jq} and Ps_{jq}, respectively. Then the conditional probability of image $\boldsymbol{I}_j$ being observed, given $\boldsymbol{q}$, can be calculated according to Eq. 2, where η is a normalization factor.

$$P(\boldsymbol{I}_j|\boldsymbol{q}) = \eta \times [1 - (1 - Pd_{jq})(1 - Ps_{jq})] \tag{2}$$

In this paper, description tags associated with images and 40 surrounding words are used as sources of text evidences, as recommended by Coelho et al. [2004]. For simplicity and efficiency, only the keywords in the query are considered. Moreover, considering that tiny images are usually advertisements or tips, images whose height or width is less than 60 pixels are filtered out.

Note that in the original form of Coelho et al.'s approach, text evidences extracted using metatags of Web pages are also used. However, Coelho et al. [2004] found in experiments that combining description tags and surrounding passages only would gain higher precision than including metatags at the same time in Web image retrieval. So, here we do not use metatags.

It is evident that Eq. 2 can be used to estimate the relevances of Web images to the query. However, such an estimation is not very reliable since it only utilizes the text information while the contents of Web pages are rather complex and the text descriptions are often ambiguous. For example, an image with the file name 'tiger' could be a photo of an animal, a mask like tiger face, a bottle of tiger-brand beer, or even a picture of a team named 'tiger'. Figure 3 shows the top 10 images retrieved according to the above process for the query *tiger animal*. It can be seen that there are lots of irrelevant images. So, in the WEBSEIC framework, the above process is only used to retrieve candidates of related images.

3.2 Identifying Related Images

Given a query, a rank of Web images can be generated by executing the process described in Section 3.1. Among the top-ranked images, such as the top 400 images, there should be some relevant images though as indicated above, they may not rank high due to the ambiguity of text descriptions and their complex relationship with images. At the same time, there are also many irrelevant images whose associated text descriptions looks relevant to the textual query. Since the rank is only generated by considering the text information, it is anticipated that further analysis on the visual contents of these images can help make the relevant images stand out. Techniques from the area of content-based image retrieval (CBIR) [Smeulders *et al.*, 2000] are helpful for this purpose.

In CBIR, each image is represented by a feature vector which can be regarded as a point in an image feature space. The query image can also be represented by a point in the space, which expresses the target concept required by the user. Then, similarities between images and the query image can be measured, which yields a rank where the higher the rank, the more relevant the image to the target concept. However, in Web search the query is not an image and therefore the point corresponding to the target concept is not known.

Fortunately, considering that all the relevant images should

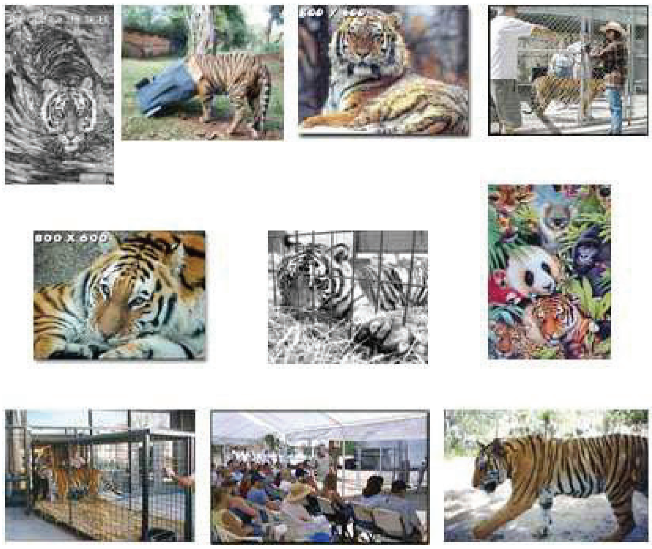

Figure 4: The top 10 Web images retrieved for the query *tiger animal* after executing the process described in Section 3.2.

be related to the target concept while the irrelevant images may be related to different themes though some associated texts look similar, the distribution of the top-ranked images obtained by executing the process described in Section 3.1 can be exploited. Generally, since all the relevant images are related to the target concept, the feature vectors corresponding to them should be relatively close in the feature space; while the feature vectors corresponding to the irrelevant images should be relatively scattered since they are related to different themes. Thus, looking for high density region in the image feature space could be helpful to identify related images.

Let $\mathcal{C} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{|\mathcal{C}|}\}$ denote the set of top-ranked candidate images each is represented by a d-dimensional feature vector, and $\boldsymbol{z}$ denotes a point in the feature space. This paper employs a simple method to estimate the density [Fukunaga and Hostetler, 1975], as shown in Eq. 3.

$$Density(\boldsymbol{z}) = \sum_{i=1}^{|\mathcal{C}|} e^{-\sum_{j=1}^{d} |\boldsymbol{z}_j - \boldsymbol{x}_{ij}|^2} \tag{3}$$

In WEBSEIC, the density of every top-ranked candidate image is estimated according to Eq. 3. Then, the images are sorted according to their densities and half of the candidate images with lower densities are filtered out. This process is repeated until a specified number of images remain, and these images are regarded as relevant images.

Figure 4 shows the top 10 images retrieved for the query *tiger animal* after executing the process described in this section. Here $\mathcal{C}$ contains the top-400 images generated through the process described in Section 3.1. RGB color histogram is used as visual features and each top-ranked candidate image is represented as a feature vector with 3×64 dimensions where 64 bins are used for each color. Comparing Figures 3 and 4 it can be found that after executing the process described in this section, the quality of the retrieved images have been apparently improved. Greater improvements can be anticipated by using stronger features.

3.3 Generating Image-Based Rank

After obtaining $\mathcal{D}$, a set of images with the highest densities, the center $\boldsymbol{y}$ can be derived by averaging the corresponding feature vectors, which is then regarded as the point corresponding to the target concept.

On this basis, the visual contents of images in Web pages can be exploited as information evidences for Web search. For a page containing candidate image, the image-based relevance of the page to the query can be computed according to Eq. 4, where $\boldsymbol{x}$ denotes the d-dimensional feature vector corresponding to the image, and σ_j denotes the standard deviation of the j-th dimension on $\mathcal{D}$.

$$ImageRel(\boldsymbol{x}) = \sqrt{\sum_{j=1}^{d} \left[\frac{1}{\sigma_j} |\boldsymbol{x}_j - \boldsymbol{y}_j| \right]^2} \tag{4}$$

If a Web page contains multiple candidate images, then its image-based relevance is the biggest $ImageRel$ value of those images; if a Web page does not contain any image, then its image-based relevance is set to a default value. Then, by ranking the Web pages according to their image-based relevances, an image-based rank of the Web pages is obtained.

3.4 Producing Final Rank

Now we have two ranks for the Web pages, i.e. a keyword-based rank and an image-based rank. The smaller the rank, the higher the relevance. It is evident that since these ranks are generated using different information, the combination of them would be more accurate and reliable.

Let $R_k(g)$ and $R_i(g)$ denote the keyword-based rank and image-based rank of a Web page g, respectively. Then, the final rank of g can be computed according to Eq. 5, where $AdaR_k$ and $AdaR_i$ are calculated according to Eqs. 6 and 7, respectively.

$$R(g) = \frac{AdaR_k\,(g) + AdaR_k\,(g)}{2} \tag{5}$$

$$AdaR_k(g) = \left(\sum\nolimits_g \frac{1}{R_k(g) + \alpha_k} \right)^{-1} \left(\frac{1}{R_k(g) + \alpha_k} \right) \tag{6}$$

$$AdaR_i(g) = \left(\sum\nolimits_g \frac{1}{R_i(g) + \alpha_i} \right)^{-1} \left(\frac{1}{R_i(g) + \alpha_i} \right) \tag{7}$$

The first terms at the right end of Eqs. 6 and 7 are normalization terms. Note that instead of directly combining the keyword-based and image-based ranks, α_k and α_i are used to tradeoff their relative contributions. The values of α_k and α_i can be determined empirically.

Table 1: Precisions of the compared techniques (B: BASELINE, W: WEBSEIC).

Retrieved pages	*jaguar aircraft*			*jaguar animal*			*peafowl bird*			*penguin shoe*			*rose flower*		
	B	W	improv.	B	W	improv.	B	W	improv.	B	W	improv.	B	W	improv.
Top 10	.700	.800	14.3%	.600	.900	50.0%	.700	.900	28.6%	.400	.500	25.0%	.500	.700	40.0%
Top 20	.400	.800	100.0%	.600	.900	50.0%	.800	.800	0.0%	.400	.550	37.5%	.550	.650	18.2%
Top 30	.400	.733	83.3%	.633	.833	31.6%	.667	.800	20.0%	.300	.433	44.4%	.567	.667	17.6%
Top 40	.400	.725	81.3%	.600	.750	25.0%	.625	.725	16.0%	.325	.400	23.1%	.600	.650	8.3%
Top 50	.460	.740	60.9%	.600	.780	30.0%	.600	.720	20.0%	.280	.400	42.9%	.560	.600	7.1%

4 Experiments

In the experiments, the *Google* search engine [Brin and Page, 1998] is used as the baseline for comparison. Moreover, it is used as the keyword-based search module in the WEBSEIC framework shown in Figure 2. In other words, we want to see whether the powerful Google search engine can be improved by exploiting image contents.

Fifteen queries are conducted, including *albacore fish*, *apple computer*, *apple fruit*, *bongo shoe*, *dove chocolate*, *eagle bird*, *geneva watch*, *shark fish*, *sunset landscape*, *tiger animal*, *jaguar aircraft*, *jaguar animal*, *peafowl bird*, *penguin shoe*, and *rose flower*. The former ten queries are used as training set on which the parameters α_k and α_i in Eq. 5 are determined, while the last five queries are used as test set.

For each test query the *precision* of the compared techniques are evaluated, which is the fraction of the number of retrieved related page to that of all the retrieved pages. Whether a retrieved page is really related to the query or not is examined by a human volunteer. Note that the conventional precision-recall graph is not used here since few users would be patient enough to browse more than the top 50 pages. The precisions on the top 10, 20, 30, 40, and 50 retrieved Web pages are tabulated in Table 1. Here 'improv.' denotes the improvement brought by exploiting image contents, which is computed by dividing the difference between the precisions of WEBSEIC and BASELINE by that of BASELINE.

It can be seen from Table 1 that exploiting image contents in the way of WEBSEIC can apparently improve the Web search performance. In about 75% cases the improvement is bigger than 25%. Considering that the comparison baseline is the powerful Google search engine, such a result is quite impressive.

Figure 5 depicts the average test performance of the compared techniques. It can be found that for the top two retrieved Web pages there is no difference between the precisions of WEBSEIC and BASELINE. This is easy to understand since the earliest retrieved Web pages are the most relevant pages which can usually be easily identified by pure keyword-based search. It can be found from Figure 5 that for later retrieved Web pages, the precision of WEBSEIC is apparently higher than that of BASELINE, which verifies the helpfulness of the exploitation of image contents.

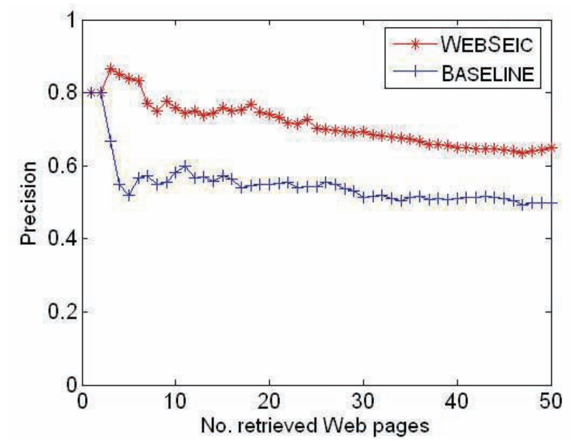

Figure 5: The average test performance of WEBSEIC and BASELINE.

5 Conclusion

In previous research on Web search, the information exploited were mainly the texts in the Web pages and the link structure between the pages. Considering that Web is a multimedia environment, this paper advocates to exploit multimedia information in the Web search process. In particular, this paper proposes a new Web search framework, in which the visual contents of images are exploited to help improve the search performance. Experiments show that this direction is promising for designing powerful Web search engines.

Since preparing the experimental data requires substantive human endeavors in determining whether the retrieved pages are really related to the query or not, in this paper the experiments are conducted on a limited number of queries. A large empirical study involving more human volunteers and more search queries will be executed in the future. Comparing the proposed framework with other Web search frameworks is also left for future work.

Note that all the operations in the WEBSEIC framework, such as retrieving candidate images, identifying related images, generating image-based rank and producing the final rank, can be realized in many forms other than those described in this paper. In fact, the main purpose of the paper is to show that exploiting image contents can bring gains to Web search, and therefore the current operations have not been finely polished. For example, the current process for analyzing the images is not fast enough. Another example is the lack of a mechanism to adaptively tradeoff the contributions of keyword-based and image-based ranks. So, it is evident that investigating effective and efficient realizations of WEBSEIC is an interesting issue for future work.

Moreover, the current technique is less helpful on abstract concepts such as *tiger economy* than on concrete entities such as *tiger* because the former can hardly be presented as an image. Improving the search performance on those abstract concepts using image information remains an open problem. Furthermore, extending the idea of WEBSEIC to design Web search frameworks which could exploit other information modalities, such as video and audio, is another interesting future issue.

Acknowledgement

The authors wish to thank the anonymous reviewers for their constructive comments. This work was supported by the National Science Foundation of China (60505013), the National Science Fund for Distinguished Young Scholars of China (60325207), the Jiangsu Science Foundation (BK2005412), and the Foundation for the Author of National Excellent Doctoral Dissertation of China (200343).

References

[Brin and Page, 1998] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

[Cai *et al.*, 2004] D. Cai, S.-P. Yu, J.-R. Wen, and W.-Y. Ma. Block based web search. In *Proceeding of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 456–463, Sheffield, UK, 2004.

[Coelho *et al.*, 2004] T. A. S. Coelho, P. P. Calado, L. V. Souza, B. Ribeiro-Neto, and R. Muntz. Image retrieval using multiple evidence ranking. *IEEE Transactions on Knowledge and Data Engineering*, 16(4):408–417, 2004.

[Frankel *et al.*, 1996] C. Frankel, M. J. Swain, and V. Athitsos. WebSeer: An image search engine for the World Wide Web. Technical Report 96-14, Department of Computer Science, University of Chicago, Chicago, IL, August 1996.

[Frankel *et al.*, 1998] C. Frankel, M. J. Swain, and V. Athitsos. The PageRank citation ranking: Bringing order to the Web. Technical Report 1999-0120, Stanford Digital Library Technologies Project, Standord University, Stanford, CA, January 1998.

[Fukunaga and Hostetler, 1975] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975.

[Harmandas *et al.*, 1997] V. Harmandas, M. Sanderson, and M. D. Dunlop. Image retrieval by hypertext links. In *Proceeding of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 296–303, Philadelphia, PA, 1997.

[Kleinberg, 1999] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[Mukherjea *et al.*, 1999] S. Mukherjea, K. Hirata, and Y. Hara. AMORE: A World Wide Web image retrieval engine. *World Wide Web*, 2(3):115–132, 1999.

[Ntoulas *et al.*, 2005] A. Ntoulas, G. Chao, and J. Cho. The infocious web search engine: Improving web searching through linguistic analysis. In *Proceeding of the 14th International Conference on World Wide Web*, pages 840–849, Chiba, Japan, 2005.

[Raghavan and Wong, 1986] V. V. Raghavan and S. K. M. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Science*, 37(5):279–287, 1986.

[Ribeiro-Neto and Muntz, 1996] B. A. Ribeiro-Neto and R. Muntz. A belief network model for ir. In *Proceeding of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 253–260, Zurich, Switzerland, 1996.

[Salton and Buckley, 1988] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513–523, 1988.

[Sclaroff *et al.*, 1997] S. Sclaroff, L. Taycher, and M. L. Cascia. ImageRover: A content-based image browser for the World Wide Web. In *Proceeding of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 2–9, San Juan, Puerto Rico, 1997.

[Silva *et al.*, 2000] I. Silva, B. A. Ribeiro-Neto, P. Calado, E. S. de Moura, and N. Ziviani. Link-based and content-based evidential information in a belief network model. In *Proceeding of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 96–103, Athens, Greece, 2000.

[Smeulders *et al.*, 2000] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[Smith and Chang, 1996] J. R. Smith and S.-F. Chang. Searching for images and videos on the World Wide Web. Technical Report 459-96-25, Center for Telecommunications Research, Columbia University, New York, NY, August 1996.

[Wang *et al.*, 2004] X.-J. Wang, W.-Y. Ma, G.-R. Xue, and X. Li. Multi-model similarity propagation and its application for Web image retrieval. In *Proceeding of the 12th ACM International Conference on Multimedia*, pages 944–951, New York, NY, 2004.

[Woodruff *et al.*, 2001] A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrison, and P. Pirolli. Using thumbnails to search the Web. In *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems*, pages 198–205, Seatle, WA, 2001.

[Xue *et al.*, 2006] X.-B. Xue, Z.-H. Zhou, and Z. Zhang. Improve Web search using image snippets. In *Proceeding of the 21st National Conference on Artificial Intelligence*, pages 1431–1436, Boston, WA, 2006.