

PageSense: Style-wise Web Page Advertising

Lusong Li
State Key Lab of Software
Development Environment
Beihang University
Beijing 100191, P. R. China
lilusong@gmail.com

Tao Mei
Microsoft Research Asia
Beijing 100190
P. R. China
tmei@microsoft.com

Xiang Niu
School of Computer Science
and Engineering
Beihang University
Beijing 100191, P. R. China
niuxiang@nlsde.buaa.edu.cn

Chong-Wah Ngo
Department of Computer
Science
City University of Hong Kong
Kowloon, Hong Kong
cwngo@cs.cityu.edu.hk

ABSTRACT

This paper presents an innovative style-wise advertising platform for web page. Web page “style” mainly refers to visual effects, such as color and layout. Unlike the most popular ad-network such as Google AdSense which needs publishers to change the original structure of their pages and define the position and style of the embedded ads manually, style-wise page advertising aims to automatically deliver style-consistent ads at proper positions within the web page, without breaking the layout of the original page. Our system is motivated from the fact that almost 90% web pages contain blank regions without any content. Given a web page with some blank regions, style-wise page advertising is able to detect the suitable blank areas for advertising, rank the ads according to the semantic relevance and web page style, and embed relevant ads into these nonintrusive areas. Style-wise page advertising represents one of the first attempts towards contextual advertising which enables the publishers to save effort when applying online advertising services.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Selection process*; H.3.5 [Information Storage and Retrieval]: Online Information Services—*Web-based services*

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Web page advertising, web page style, multimodal relevance

1. INTRODUCTION

Web advertising uses the World Wide Web for the expressed purpose of delivering marketing messages to attract

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

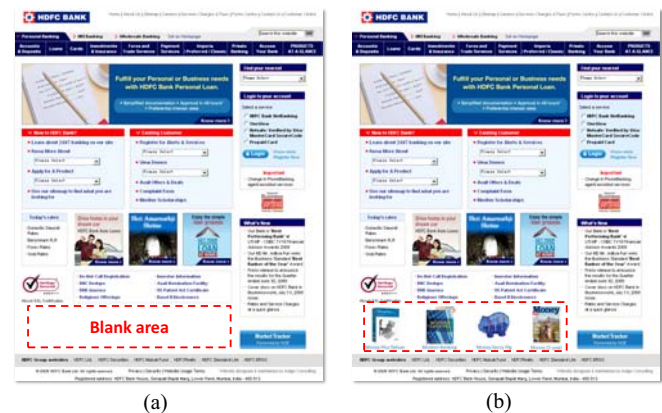


Figure 1: (a) A blank area within the web page. The highlighted rectangle area indicates the blank area. (b) Some associated ads are inserted in the blank area within the web page. The highlighted rectangle area indicates the inserted ads which are relevant and style-consistent with the ad landing page.

customers. We have witnessed a fast growing and great potential market of online advertising in recent years. Many existing ad-networks such as Google AdSense [1] have provided contextual advertising services which associates advertisements with an web page. With AdSense, the owners of websites can display contextually-relevant, unobtrusive Google ads on their website’s pages to earn money. However, AdSense ads are located in the fixed corner of the Web page. Moreover, it need the publishers to define the style of the delivered ads to match their own pages. For image ads, AdSense does not consider whether the ad images match the style of the host page. We argue that inserting ads at preserved and fixed positions destroys the visually appealing appearance and structure of the original web page. We also believe that the ads which are relevant to the page style will reduce the intrusiveness of the ads.

It is witnessed that web pages usually have some blank regions without any informative content, as shown in Figure 1(a). Table 1 shows the ad formats of Google AdSense. The

Table 1: Google AdSense Ad Formats

Ad type	Format	Size
Half Banner	234x60	14,040
Button	125x125	15,625
Small Rectangle	180x150	27,000
Banner	468 x 60	28,080
Vertical Banner	120 x 240	28,800
Small Square	200 x 200	40,000
Square	250 x 250	62,500
Leaderboard	728 x 90	65,520
Skyscraper	120x600	72,000
Medium Rectangle	300 x 250	75,000
Large Rectangle	336 x 280	94,080
Wide Skyscraper	160x600	96,000

size of ad format ranges from about 10,000 to 100,000 pixels. Based on the sizes that Google ads use, we conduct an experiment to investigate if web pages contain blank areas which are suitable for delivering ads. According to Alexa.com's traffic statistics, we analyze the top 1,000 web sites' home-pages and find that large percentage of web pages have blank areas which are large enough for inserting ads. Figure 2 shows the results. We believe that there are two benefits of delivering ads on the blank regions of the web page: 1) The blank regions sometimes make the web page unbeautiful. The embedded ads which are consistent with the style of ad landing page will probably complement the blank area and improve the user's viewing experience; 2) No predefined ad blocks are to be preserved which enables the publishers to save effort when applying contextual advertising service.

Motivated by the above observations, we demonstrate our novel advertising system in this work. The proposed system, named PageSense, supports contextual web page advertising by associating the most relevant ads to web page content as well as web page style. In order to minimize intrusiveness to the users, the ads should be delivered on the blank areas within the web page so that the ads don't affect users' viewing experience. The ad insertion positions are detected based on image analysis. Figure 1(b) shows a snapshot of the example subscribing our style-wise advertising service. It is observed that the relevant ads have been inserted into the blank region within the web page.

2. RELATED WORK

Many efforts have been devoted to online advertising which can be categorized into two dimensions, ad relevance matching and ad position detection.

Research on ad relevance matching can be further classified into three directions from the perspective of what the ads are matched against: keyword-targeted, content-targeted, and user-targeted advertising. Typical keyword-targeted advertising analyzes a web page or query to find prominent keywords, and then match these keywords against the words for which advertisers bid [2][3][4]. In content-targeted advertising (also called "contextual advertising"), the ads are associated with the web page content rather than the keywords [5][6][7][8]. In user-targeted advertising, the ads are driven based on user profile and demography [9], or behavior [10][11].

In addition to the relevance, it is important to insert the

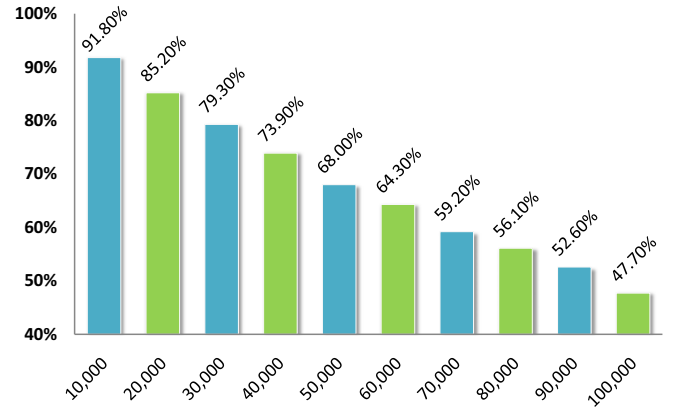


Figure 2: Based on the sizes of Google ad format which range from about 10,000 to 100,000 pixels, we investigate the percentage of web pages contain blank areas big enough for delivering ads.

ad on a non-intrusive position. Mei et al. proposed how to detect appropriate insertion positions in the video stream [12] and within the image [13] to reduce the intrusiveness.

3. SYSTEM OVERVIEW

Figure 3 illustrates the framework of demonstration system. The system consists of three major components: (1) ad matching engine, (2) ad position detection engine, and (3) ad delivery engine.

The ad matching engine matches the web page with ads based on category relevance, text relevance and style consistency. Ad insertion point detection engine detects the blank area for ad insertion, so that the informative content of the host page will not be occluded and users may not feel intrusive to the inserted advertisement. The ad delivery engine delivers and renders the relevant ads at blank areas within web page.

4. SYSTEM IMPLEMENTATION

4.1 Ad Ranking

Conventional contextual advertising primarily matches ads to web pages based on categories or prominent keywords which are regarded as semantic meaning. Besides the semantic relevance between the ad and ad landing page, the ad should be consistent with the style of web page. Therefore, given a web page P , we introduce the following three relevance items for each ad x .

- **Category relevance** $R_C(P, x)$: It measures if the ad x and ad landing page P belong to the same or compatible categories. The web categorization is based on the SVM classifier on text and web link analysis[14].
- **Keyword relevance** $R_T(P, x)$: It measures the relevance between the keywords or named entities extracted from the web page and the text information associated with the ad.
- **Style relevance** $R_V(P, x)$: It measures the style consistency between the ad and the page. The ad is assumed to have similar appearance with the page, so

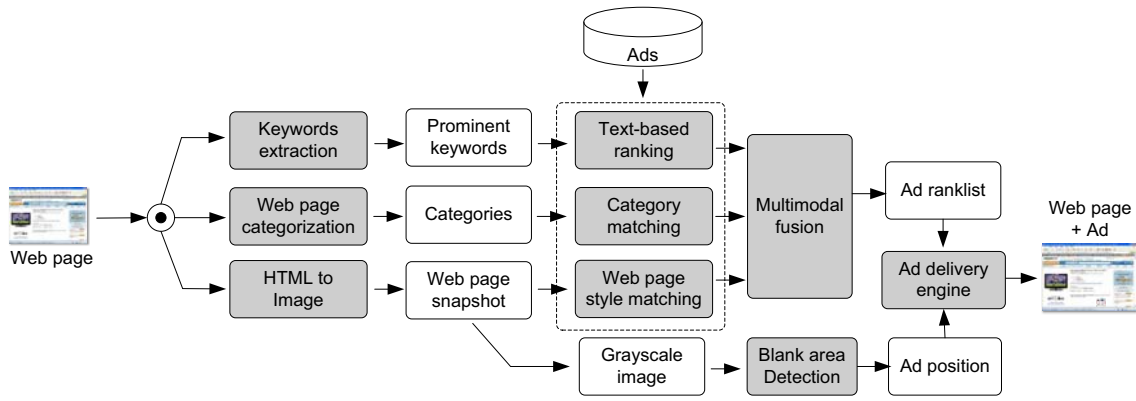


Figure 3: System framework of PageSense.

that the users may perceive the ad as a natural part of original page. To measure the visual similarity between x and P , we compute the L1 distance defined in a HSV color space between x and P [15].

We fuse the relevance scores from category, keyword and style to give a more reasonable judgment, i.e. multimodal relevance. It is determined by the following formulation:

$$R(P, x) = R_C(P, x) + R_T(P, x) + R_V(P, x) \quad (1)$$

Since our system supports multiple ads on the same web page, The N ads $\mathcal{X} = \{x_i\}_{i=1}^N$ with the highest scores of equation (1) are selected as relevant ads which will be delivered to the web page. N is decided by the publisher and the default value is 10.

4.2 Ad Position Detection

The process of ad position detection is shown in Figure 4. First, the web page is crawled and converted into snapshot image. Secondly, the snapshot image is processed and converted into grey-scale image. Third, we detect the blank area within the grey-scale image based on image analysis. The blank area is the candidate position for delivering ads. Finally, we calculate the ad position relative to the original page.

4.3 Ad Delivery

Publishers only need to insert a piece of HTML codes at the bottom of their own web pages, then relevant ads will emerge on the blank regions of the web page.

The inserted HTML code is able to download a javascript file which handles the ad delivery. First, the current page's URL is sent back to a web service in the back-end. Second, the web service will crawl the web page and detect the blank areas. Third, the HTML layers embedded with ads will be delivered to the browser and overlay on the blank areas.

According to the size of the insertion position, PageSense can choose one ad or multiple ads arranged in a rectangle. It also supports a variety of rendering effects such as displaying multiple ads which appear in turn like slideshow, fade-in/fade-out, dissolve-in/dissolve-out, and so on.

If there is no blank area within the page, the ads will float on the bottom right corner of the browser and only appear for certain duration, so that the users can still see the whole web page.

5. EXPERIMENTS

We have collected 10,127 real selling products from Amazon.com as ads and top 1,000 websites' homepages based on Alexa.com's traffic statistics, among which 100 randomly selected homepages are used for PageSense evaluations. We also collect 10 pages which subscribe AdSense service for comparison. The evaluation involves 30 human subjects, including 15 males and 15 females. These subjects are university students, teachers, medical people, IT people and finance people. The students and teachers are major in a wide variety of backgrounds, such as aeronautics, mathematics, computer science, business, physics, art, language, law, history, biology and so on. All the subjects' ages range from 19 to 40.

In our experiments, we compared each page of PageSense to each page of Google AdSense. As a result, there are 1,000 page pairs. Each pair is judged by three different subjects. The human subjects are asked to give scores (1-5) to show their satisfactions with respective to the following perspectives:

1) Position: How do you feel the position of the advertisement? (5: very good; 4: good to; 3: may be; 2: not good; 1: bad)

2) Impression: Does the ad impress you? (5: very impressive; 4: impressive; 3: may be; 2: not too impressive; 1: unimpressive)

3) Acceptance: If you own a web page, are you willing to put it on your site? (5: very willing to; 4: willing to; 3: may be; 2: not too willing to; 1: not willing to)

4) Overall effectiveness: Is it an effective way to deliver the ad? How do you think of the overall effectiveness of advertising? (5: very good; 4: good to; 3: may be; 2: not good; 1: bad)

Figure 5 shows the evaluation results. The higher score indicates the higher satisfaction. It is observed that AdSense outperformed PageSense with slight margin in terms of position, while PageSense achieved better satisfactions than AdSense with respect to other perspectives, i.e. impression, acceptance and overall effectiveness. Because AdSense's ad positions are defined manually, while PageSense's are detected automatically, it is reasonable to see that AdSense achieved the higher score than PageSense in terms of position. The score of PageSense is indeed closely approaching that of Google AdSense, though the former does not involve

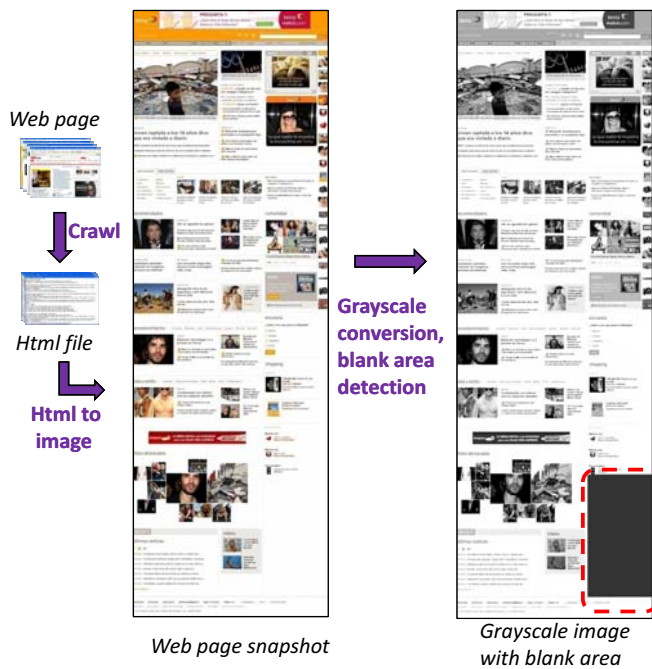


Figure 4: Ad insertion position detection. The highlighted rectangle area indicates the candidate ad position.

manual insertion. This has proved that the algorithm of ad position detection for PageSense is satisfactory.

The PageSense delivers the image-based ads which support a variety of rendering effects. Moreover, PageSense ads are consistent with the ad landing page's color style. Most of the evaluators think PageSense ads are more impressive than AdSense ads.

6. CONCLUSIONS

We have demonstrated a novel web page advertising platform named PageSense that supports not only "contextually relevant" but also "less-intrusive" online advertising. This platform automatically matches ads to the content and style of web page, delivers relevant ads on the blank area within the page, without breaking the original structure of the web page by only inserting a piece of code.

The future work includes further defining the style consistency, and making the ads user-targeted in terms of user behavior and social context.

7. ACKNOWLEDGMENTS

The research was supported by National 973 Program of China (Grant No.2005CB321901).

8. REFERENCES

- [1] AdSense. <http://www.google.com/adsense/>.
- [2] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5):22, 2007.
- [3] D. Shen, J.-T. Sun, Q. Yang, and Z. Chen. Building bridges for web query classification. In *Proceedings of SIGIR '06*, pages 131–138, 2006.
- [4] W.-t. Yih, J. Goodman, and V. R. Carvalho. Finding advertising keywords on web pages. In *Proceedings of WWW '06*, pages 213–222, 2006.
- [5] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. In *Proceedings of SIGIR '07*, pages 559–566, 2007.
- [6] A. Lacerda, M. Cristo, M. A. Gonçalves, W. Fan, N. Ziviani, and B. Ribeiro-Neto. Learning to advertise. In *Proceedings of SIGIR '06*, pages 549–556, 2006.
- [7] V. Murdock, M. Ciaramita, and V. Plachouras. A noisy-channel approach to contextual advertising. In *Proceedings of ADKDD '07*, pages 21–27, 2007.
- [8] B. Ribeiro-Neto, M. Cristo, P. B. Golgher, and E. Silva de Moura. Impedance coupling in content-targeted advertising. In *Proceedings of SIGIR '05*, pages 496–503, 2005.
- [9] J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen. Demographic prediction based on user's browsing behavior. In *Proceedings of WWW '07*, pages 151–160, 2007.
- [10] H. K. Dai, L. Zhao, Z. Nie, J.-R. Wen, L. Wang, and Y. Li. Detecting online commercial intention (oci). In *Proceedings of WWW '06*, pages 829–837, 2006.
- [11] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of WWW '07*, pages 521–530, 2007.
- [12] T. Mei, X.-S. Hua, L. Yang, and S. Li. VideoSense: towards effective online video advertising. In *Proceedings of ACM MULTIMEDIA '07*, pages 1075–1084, 2007.
- [13] T. Mei, X.-S. Hua, and S. Li. Contextual in-image advertising. In *Proceeding of ACM MULTIMEDIA '08*, pages 439–448, 2008.
- [14] G. Attardi, A. Gulli, and F. Sebastiani. Automatic web page categorization by link and context analysis. In *Proceedings of THAI-99*, pages 105–119, 1999.
- [15] T. Mei, X.-S. Hua, W. Lai, L. Yang, and et al. MSRA-USTC-SJTU at TRECVID 2007: High-level feature extraction and search. In *TREC Video Retrieval Evaluation Online Proceedings*, 2007.

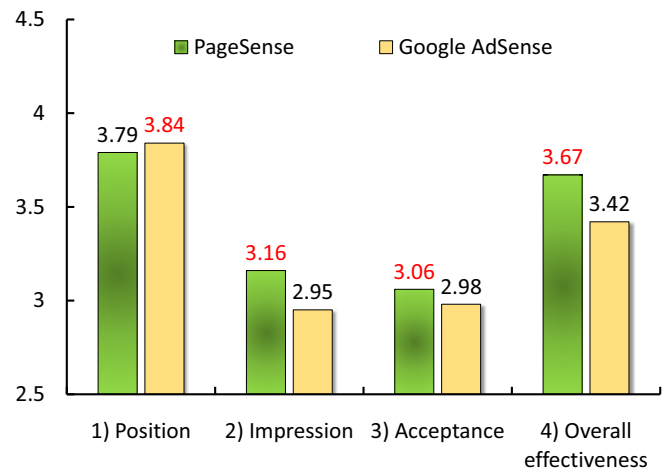


Figure 5: The evaluations on ad satisfactions.