

# Competitive Analysis from Click-Through Log

Gang Wang<sup>1</sup>, Jian Hu<sup>1</sup>, Yunzhang Zhu<sup>2</sup>, Hua Li<sup>1</sup> and Zheng Chen<sup>1</sup>

<sup>1</sup>Microsoft Research Asia  
No. 49 Zhichun Road,

Beijing 100190, P.R. China

{gawa, jianh, huli, zhengc}@microsoft.com

<sup>2</sup> Department of Fundamental Science,  
Tsinghua University,

Beijing 100084, P.R. China

zhuyz05@mails.tsinghua.edu.cn

## ABSTRACT

Existing keyword suggestion tools from various search engine companies could automatically suggest keywords related to the advertisers' products or services, counting in simple statistics of the keywords, such as search volume, cost per click (CPC), etc. However, the nature of the generalized Second Price Auction suggests that better understanding the competitors' keyword selection and bidding strategies better helps to win the auction, other than only relying on general search statistics. In this paper, we propose a novel keyword suggestion strategy, called *Competitive Analysis*, to explore the keyword based competition relationships among advertisers and eventually help advertisers to build campaigns with better performance. The experimental results demonstrate that the proposed Competitive Analysis can both help advertisers to promote their product selling and generate more revenue to the search engine companies.

## Categories and Subject Descriptors

H.3.5 [Information System]: Information Storage and Retrieval—Online Information Services; I.5 [Pattern Classification];

## General Terms

Algorithm, Design, Performance, Experimentation

## Keywords

Keyword advertising, keyword suggestion, competitive analysis.

## 1. INTRODUCTION

Search Engine Advertising is on the basis of the advertisers' bidding on keywords and driven by the generalized Second Price Auction mechanism [1]. The advertiser creates an advertisement through providing search engine the displayed ads title, description text and URL together with a set of keywords to bid for the advertisement, which are usually related to the product or service it wishes to advertise. However, it has been difficult for advertisers to precisely identify their requirements on keywords which can introduce high user interests and traffics, only relying on the general traffic statistics on some related keywords. Moreover, a top advertiser would be more interested in its competitor's concerns and attempts so as trying to gain the advantages over the competitors and boost their own advertisements. As the main motivation of this paper, the proposed Competitor Analysis is to provide advertiser the competitive analysis tool to automatically analyze and understand the competition relationship among a group of advertisers and the related keyword competitions as well.

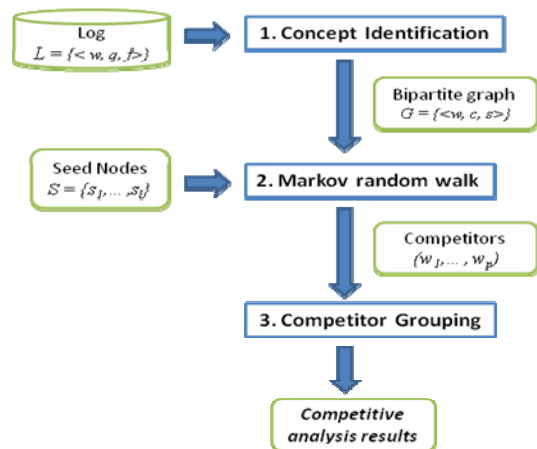
Copyright is held by the author/owner(s).  
WWW 2009, April 20–24, 2009, Madrid, Spain.  
ACM 978-1-60558-487-4/09/04.

## 2. COMPETITIVE ANALYSIS METHOD

In general search engine, if a query  $q$  leads users to two different advertiser's websites  $w_a$  and  $w_b$ , then  $w_a$  and  $w_b$  are competitors with each other on the query  $q$ . Thus, the search click-through data, which is the log data generated by Web search engine, is a good resource to discover the competitive relations. Our approach consists of three major steps, which is illustrated in Figure 1:

1. In order to filter out the noise from log data, the first step is to extract a number of concept keywords from queries. The concept keywords have compact and precise meanings and can be considered as the website abstraction. Hence, each website is represented by a vector of concept keywords.
2. In the second step, we build a bipartite graph based on the results from the first step. One partition of the graph consists of the extracted concept keywords and another partition consists of advertisers' websites. The Markov random walk algorithm [2] is then used to measure the degree of competition according to website proximities in the bipartite graph.
3. In the last step, we choose a set of important keyword groups and arrange competitors into them according to the results from the Markov random walk algorithm.

**Figure 1:** The algorithm of competitive analysis consists of three major steps.



## 3. EXPERIMENTS

The search click-through log we use for the experiments is collected from a commercial search engine Live Search for one month. Table 1 illustrates the competitive results for two websites "expedia.com" and "shoes.com". The first task of the competitive analysis is to provide a list of the ranked websites competing against the seed node. Since the degree of competition for each competitor is recognized, we choose the top 20 competitors displayed.

Moreover, the competitive analysis produces a list of ranked keywords on which the seed node competes with other websites, and the top 10 keywords are presented in the paper. It is crucial for an advertiser to identify which websites are most competitive on a certain keyword so that the advertiser can make the accurate strategy to target on the right audience and drag the traffic from its competitors. Therefore, we arrange the competitors into different keyword categories. For each keyword category, the websites are also ranked according to the degree of competition on this keyword.

- The competitors of “expedia.com” are plentiful and broad in concept as well including travel agencies, airline companies, hotels, travel insurance companies, etc. If “expedia.com” would like to launch an ad campaign to promote the hotel service (no.3 in keyword category), the top competitors are “hotels.com”, “disney.go.com” and “tripadvisor.com”. Thus, “expedia.com” should collect their ad information such as ad coverage and bidding so as to gain the advantages over these competitors. If “expedia.com” would like to attract the traffic from those persons seeking for the last minute flight (no.7 in keyword category), it should pay more attentions to the ads from “lastminutedeals.com”, “lastminutetravel.com” and “smartertravel.com”.
- “Shoes.com” mainly sells shoes and related products online. The competitor composition of “shoes.com” is quite interesting with “payless.com” in a dominating position. The

keyword category varies from general shoes categories such as “women shoes”, “men shoes” and “kid shoes” to brand based categories like “converse shoes”, “simpson shoes”, “adidas shoes” and etc. If “shoes.com” wants to build an ad campaign targeting to female customers, they really need to look into the top competitors at “women shoes” category, which are totally different from the category of “men shoes”.

#### 4. RELATED WORKS

Keyword substitution is the most direct technique for keyword suggestion. The methods transforming keyword completely different through the click-through log analysis[3], latent semantic indexing (LSI) [4] and pseudo-relevance feedback [5]. Keyword generation is another important technique which suggests keywords through exploiting the content from web pages. Turney *et al.* [6] used keyword candidates to query the search engine, and then use the number of hit documents as an additional feature for keyword generation. Kelleher [7] used the link information of a web page, thus design a “semantic ratio” feature with to anchor text to extract keyword. Yih *et al.* [8] described a supervised approach that uses a number of features such as TF and IDF to learn how to extract keywords from web pages.

**Table 1:** The competitive analysis results on two input websites. Top 20 competitors are shown on the left side hand for each input website. The number in bracket indicates the degree of competition. Top 10 competing keywords are shown in the middle. The websites that competing on a keyword are arranged into this keyword, and top 3 competitors are selected.

expedia.com					
Top 20 Competitors:		Keyword Category:	#1	#2	#3
1. travelers.com (8.8)	11. travelocity.com (3.0)	1. travel	travelers.com	travel.state.gov	travel.com
2. cheapflights.com (7.0)	12. disney.go.com (2.8)	2. flight	cheapflights.com	travel.com	flights.com
3. travel.state.gov (6.2)	13. flytcomm.com (2.8)	3. hotel	hotels.com	disney.go.com	tripadvisor.com
4. travel.com (5.5)	14. cheaptickets.com (2.6)	4. airfare	expedia.com	orbitz.com	travelocity.com
5. flights.com (4.8)	15. delta.com (2.6)	5. vacation	aaa.com	disney.go.com	tripadvisor.com
6. aaa.com (4.7)	16. sidestep.com (2.4)	6. air ticket	cheapair.com	cheaptickets.com	cheapair.com
7. hotels.com (4.6)	17. priceline.com (2.3)	7. last minute	lastminutedeals.com	lastminutetravel.com	smartertravel.com
8. orbitz.com (4.3)	18. kayak.com (2.0)	8. airline flight	usairways.com	flytcomm.com	delta.com
9. usairways.com (3.9)	19. tripadvisor.com (2.0)	9. air fare	cheapflights.com	priceline.com	airfarewatchdog.com
10. cheapair.com (3.6)	20. maps.live.com (1.9)	10. cheap	cheapflights.com	cheapair.com	cheaptickets.com
shoes.com					
Top 20 Competitors:		Keyword Category:	#1	#2	#3
1. payless.com (25.6)	11. shoesunlimited.com (1.7)	1. shoes	payless.com	shoebuy.com	zappos.com
2. shoebuy.com (9.1)	12. planetshoes.com (1.7)	2. women shoes	shoebuy.com	famousfootwear.com	bakersshoes.com
3. zappos.com (8.1)	13. crocs.com (1.5)	3. mudd shoes	shoesunlimited.com	shoes.about.com	shoes.style365.com
4. dsw.com (5.6)	14. shoecarnival.com (1.5)	4. simpson shoes	onlineshoes.com	jessicasimpsoncollection.com	simpson-heat.com
5. onlineshoes.com (5.2)	15. shop.nordstrom.com (1.4)	5. converse shoes	converse.com	discountshoesonsale.com	myairshoes.com
6. nike.com (3.3)	16. qualitybridalshoes.com (1.4)	6. kids shoes	payless.com	shoesunlimited.com	shoes.about.com
7. jessicasimpsoncollection.com (2.8)	17. shoes.about.com (1.3)	7. men shoes	intrends.com	ajkings.com	styledrops.com
8. famousfootwear.com (2.1)	18. edhardyshop.com (1.3)	8. adidas shoes	finishline.com	discountshoesonsale.com	myairshoes.com
9. finishline.com (1.9)	19. shoes.style365.com (1.2)	9. bass shoes	ghbass.com	shoestation.com	awwus.com
10. bakersshoes.com (1.7)	20. converse.com (1.2)	10. discount shoes	zappos.com	elite-shoes-discount.com	discountshoesonsale.com

#### REFERENCES

[1] V. Krishna. *Auction Theory*. Academic Press, 2002.

[2] J. Norris. *Markov Chains*. Cambridge University Press, 1998.

[3] R. Baeza-Yates and A. Tiberi. Extracting semantic relations from query logs. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD-07)*, 2007.

[4] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391-407, 1990.

[5] M. Mitra, A. Singhal, and C. Buckley. Improving automatic query expansion. In *Proceedings of the 21th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR-98)*, 1998.

[6] Turney, P.D., Coherent Keyphrase extraction via web mining. In *Proc. of the 16th International Joint Conference on Artificial Intelligence (IJCAI-03)*, 2003.

[7] Kelleher, D. and Luz, S., Automatic hypertext keyphrase detection. In *Proc. of the 18th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 2005.

[8] Yih, W., Goodman, J. and Carvalho, V.R., Finding advertising keywords on web pages. In *Proc. of the 15th World Wide Web Conference (WWW-06)*, 2006.