Position Paper: A Study of Web Search Engine Bias and its Assessment

Ing-Xiang Chen

Dept. of Computer Sci. and Eng., Yuan Ze University
135 Yuan-Tung Road, Chungli
Taiwan, 320, ROC

sean@syslab.cse.yzu.edu.tw

Cheng-Zen Yang

Dept. of Computer Sci. and Eng., Yuan Ze University
135 Yuan-Tung Road, Chungli
Taiwan, 320, ROC

czyang@syslab.cse.yzu.edu.tw

ABSTRACT

Search engine bias has been seriously noticed in recent years. Several pioneering studies have reported that bias perceivably exists even with respect to the URLs in the search results. On the other hand, the potential bias with respect to the content of the search results has not been comprehensively studied. In this paper, we propose a two-dimensional approach to assess both the indexical bias and content bias existing in the search results. Statistical analyses have been further performed to present the significance of bias assessment. The results show that the content bias and indexical bias are both influential in the bias assessment, and they complement each other to provide a panoramic view with the two-dimensional representation.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *Performance Evaluation*

General Terms

Measurement

Keywords

search engine bias, indexical bias, content bias, information quality, automatic assessment.

1. INTRODUCTION

In recent years, an increasingly huge amount of information has been published and pervasively communicated over the World Wide Web (WWW). Web search engines have accordingly become the most important gateway to access the WWW and even an indispensable part of today's information society as well. According to [3][7], most users get used to few particular search interfaces, and thus mainly rely on these Web search engines to find the information. Unfortunately, due to some limitations of current search technology, different considerations of operating strategies, or even some political or cultural factors, Web search engines have their own preferences and prejudices to the Web information [10][11][12]. As a result, the information sources and content types indexed by different Web search engines are exhibited in an unbalanced condition. In the past studies [10][11][12], such unbalanced item selection in Web search engines is termed *search engine bias*.

In our observations, search engine bias can be incurred from three

aspects. The first source is from the diverse operating policies and the business strategies adopted in each search engine company. As mentioned in [1], such type of bias is more insidious than advertising. A recent hot piece of news demonstrates this type of bias from the event that Google in China distorts the reality of "Falun Gong" by removing the searched results. In this example, Google agrees to comply with showing in China to guard its business profits [4]. Second, the limitations of crawling, indexing, and ranking techniques may result in search engine bias. An interesting example shows that the phrase "Second Superpower" was once Googlewashed in only six weeks because webloggers spun the alternative meaning to produce sufficient PageRank to flood Google [9][13][17]. Third, the information provided by the search engines may be biased in some countries because of the opposed political standpoints, diverse cultural backgrounds, and different social custom. The blocking and filtering of Google in China [20][21] and the information filtering on Google in Saudi Arab, Germany, and France are the cases that politics biases the Web search engine [19][20].

As a search engine is an essential tool in the current cyber society, people are probably influenced by search engine bias without awareness when cognizing the information provided by the search engine. For example, some people may never get the information about certain popular brands when inquiring about the term "home refrigerators" via a search engine [11]. From the viewpoint of the entire information society, the marginalization of certain information limits the Web space and confines its functionality to a limited scope [6]. Consequently, many search engine users are unknowingly deprived of the right to fairly browse and access the WWW.

Recently, the issue of search engine bias has been noticed, and several studies have been proposed to investigate the measurement of search engine bias. In [10][11][12], an effective method is proposed to measure the search engine bias through comparing the URL of each indexed item retrieved by a search engine with that by a pool of search engines. The result of such search engine bias assessment is termed the *indexical bias*. Although the assessment of indexed URLs is an efficient and effective approach to predict search engine bias, assessing the indexical bias only provides a partial view of search engine bias. In our observations, two search engines with the same degree of indexical bias may return different page content and reveal the semantic differences. In such a case, the potential difference of overweighing specific content may result in significant content bias that cannot be presented by simply assessing the indexed URLs. In addition, if a search result contains redirection links to other URLs that are absent from the search result, these absent URLs still can be accessed via the redirection links. In this case, a search engine only reports the mediate URLs, and the search

engine may thus have a poor indexical bias performance but that is not true. However, analyzing the page content helps reveal a panoramic view of search engine bias.

In this paper, we examine the real bias events in the current Web environment and study the influences of search engine bias upon the information society. We assert that assessing the *content bias* through the content majorities and minorities existing in Web search engines as the other dimension can help evaluate search engine bias more thoroughly. Therefore, a two-dimensional assessment mechanism is proposed to assess search engine bias. In the experiments, the two-dimensional bias distribution and the statistical analyses sufficiently expound the bias performance of each search engine.

2. LITERATURE REVIEW

Recently, some pioneering studies have been conducted to discuss search engine bias by measuring the retrieved URLs of Web search engines. In 2002, Mowshowitz and Kawaguchi first proposed measuring the indexed URLs of a search engine to determine the search engine bias since they asserted that a Web search engine is a retrieval system containing a set of items that represent messages [10][11][12]. In their method, a vector-based statistical analysis is used to measure search engine bias by selecting a pool of Web search engines as an implicit norm, and comparing the occurring frequencies of the retrieved URLs by each search engine in the norm. Therefore, bias is assessed by calculating the deviation of URLs retrieved by a Web search engine from those of the norm.

In [11], a simple example is illustrated to assess indexical bias of three search engines with two queries and the top ten results of each query. Thus, a total of 60 URL entries were retrieved and analyzed, and 44 distinct URLs with occurring frequencies were transformed into the basis vector. The similarity between the two basis vectors was then calculated by using a cosine metric. The result of search engine bias is obtained by subtracting the cosine value from one and gains a result between 0 and 1 to represent the degree of bias.

Vaughan and Thelwall further used such a URL-based approach to investigate the causes of search engine coverage bias in different countries [18]. They asserted that the language of a site does not affect the search engine coverage bias but the visibility of the indexed sites. If a Web search engine has many high-visible sites, which means Web sites are linked by many other Web sites, the search engine has a high coverage ratio. Since they calculated the search engine coverage ratio based on the number of URLs retrieved by a search engine, the assessment still cannot clearly show how much information is covered. Furthermore, the experimental sites were retrieved only from three search engines with domain names from four countries with Chinese and English pages, and thus such few samples may not guarantee a universal truth in other countries.

In 2003, Chen and Yang used an adaptive vector model to explore the effects of content bias [2]. Since their study was targeted on the Web contents retrieved by each search engine, the content bias was normalized to present the bias degree. Although the assessment appropriately reveals content bias, the study ignores the normalization influences of contents among each retrieved item. Consequently, the content bias may be over-weighted with some rich-context items. Furthermore, the study cannot determine whether the results are statistically significant.

From the past literatures in search engine bias assessment, we argue that without considering the Web content, the bias assessment only tells users part of the reality. Besides, how to appropriately assess search engine bias from both views needs advanced study. In this paper, we propose an improved assessment method for content bias and in advance present a two-dimensional strategy for bias assessment.

3. THE BIAS ASSESSMENT METHOD

To assess the bias of a search engine, a norm should be first generated. In traditional content analysis studies, the norm is usually obtained with careful examinations of subject experts [5]. However, artificially examining Web page content to get the norm is impossible because the Web space is rapidly changing and the number of Web pages is extremely large. Therefore, an implicit norm is generally used in current studies [10][11][12]. The implicit norm is defined by a collection of search results of several representative search engines. To avoid unfairly favoring certain search engines, any search engine will not be considered if it uses other search engine's kernel without any refinement, or its indexing number is not comparably large enough.

Since assessing the retrieved URLs of search engines cannot represent the whole view of search engine bias, the assessment scheme needs to consider other expressions to satisfy the lack. In the current cyber-society, information is delivered to people through various Web pages. Although these Web pages are presented with photos, animations, and various multimedia technologies, the main content still consists of hypertextual information that is composed of different HTML tags [1]. Therefore, in our approach, the hypertextual content is assessed to reveal another bias aspect.

To appropriately present Web contents, we use a weighted vector approach to represent Web pages and compute the content bias. The following subsections elaborate the generation of an implicit bias norm, a two-dimensional assessment scheme, and a weighted vector approach for content bias assessment.

3.1 Bias Norm Generation

As the definition of bias in [10][11][12], an implicit norm used in our study is generated from the vector collection of a set of comparable search engines to approximate the ideal. The main reason of this approximation is because the changes in Web space are extremely frequent and divergent, and thus traditional methods of manually generating norms by subject experts are time-consuming and become impractical. On the other hand, search engines can be implicitly viewed as experts in reporting search results. The norms can be generated by selecting some representative search engines and synthesizing their search results. However, the selection of the representative search engines should be cautiously considered to avoid generating biased norms that will show favoritism on some specific search engines.

The selection of representative search engines is based on the following criteria:

1. The search engines are generally designed for different subject areas. Search engines for special domains are not considered. In addition, search engines, e.g. localized search engines, designed for specific users are also disregarded.
2. The search engines are comparable to each other and to the search engines to be assessed. Search engines are excluded if the number of the indexed pages is not large enough.
3. Search engines will not be considered if they use other search

engine's core without any refinement. For example, Lycos has started to use the crawling core provided by FAST in 1999. If both are selected to form the norms, their bias values are unfairly lower. However, if a search engine uses other's engine kernel but incorporates individual searching rules, it is still under consideration for it may provide different views.

4. Metasearch engines are under consideration if they have their own processing rules. We assume that these rules are not prejudiced in favor of certain search engines. In fact, if there exist prejudices, they will be revealed after the assessment, and the biased metasearch engine will be excluded.

3.2 The Two-dimensional Assessment Scheme

Since both indexical bias and content bias are important to represent the bias performance of a search engine, we assess search engine bias from both aspects and present search engine bias in a two-dimensional view. Figure 1 depicts the two-dimensional assessment process. For each query string, the corresponding query results are retrieved from Web search engines. Then the URL locator parses the search results and fetches the Web pages. The document parser extracts the feature words and computes the content vectors. Stop words are also filtered out in this stage. Finally, feature information is stored in the database for the following bias measurement.
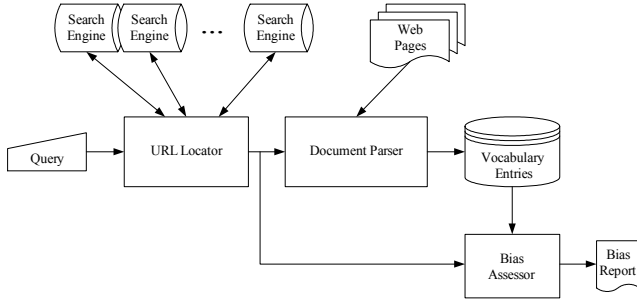

Figure 1: The assessment process of measuring search engine bias

The bias assessor collects two kinds of information: the URL indexes and the representative vocabulary vectors (*RVV*) for corresponding Web contents. The URL indexes are used to compute the indexical bias, and the *RVV* vectors are used to compute the content bias. After the assessment, the assessor generates bias reports.

3.3 The Weighted Vector Model

Web contents are mainly composed of different HTML tags that respectively represent their own specific meanings in Web pages. For example, a title tag represents the name of a Web page, which is shown in the browser window caption bar. Different headings represent differing importance in a Web page. In HTML there are six levels of headings. H1 is the most important; H2 is slightly less import, and so on down to H6, the least important [14]. In content bias assessment, how to represent a Web document plays an important role to reflect the reality of assessment.

Here we adopt a weighted vector approach to measure content bias [8]. It is based on a vector space model [15] but adapted to emphasize the feature information in Web pages. Because the features in <title>, <H1>, or <H2> tags usually indicate important information and are used more often in the Web documents, features in these tags are appropriately weighted to represent Web contents. Since the number of the total Web documents can only be estimated by sampling or assumption, this model is more

appropriate to represent and assess the contents of Web documents.

Since the search results are query-specific, query strings in different subjects are used to get corresponding representative vocabulary vectors *RVV* for search engines. Each *RVV* represents the search content of a search engine and is determined by examining the first *m* URL entry in the search result list. Every word in URL entries is parsed to filter out stop words and to extract feature words. The *RVV* consists of a series of vocabulary entries VE_i with eight fields: the *i*-th feature word, its overall frequency *f*, its document frequency *d*, the number of documents *n*, its title frequency *t*, its H1 frequency *H*, its H2 frequency *h*, and its score *S*. The score *S* is determined as follows:

$$S = (f + t \cdot w_t + H \cdot w_H + h \cdot w_h) \times \log(\frac{n}{d}) \qquad (1)$$

where w_t, w_H, and w_h are respective tag weights. The scores are used in similarity computations.

After all *RVV* vectors are computed, necessary empty entries are inserted to make the entries in *RVV* exactly corresponding to the entries in the norm for similarity computation. Then the cosine function is used to compute the similarity between RVV_i of *i*-th search engine and the norm *N*:

$$Sim(RVV_i, N) = \cos(RVV_i, N) =$$

$$\frac{\sum_j S_{RVV_{i,j}} \cdot S_{N,j}}{\sqrt{\sum_j S_{RVV_{i,j}}^2} \sqrt{\sum_j S_{N,j}^2}} \qquad (2)$$

where $S_{RVV_{i,j}}$ is the *j*-th entry score of RVV_i, and $S_{N,j}$ is the *j*-th entry score of the norm. Finally, the content bias value $CB(RVV_i, N)$ is defined as

$$CB(RVV_i, N) = 1 - Sim(RVV_i, N) \qquad (3)$$

4. EXPERIMENTS AND DISCUSSIONS

We have conducted experiments to study bias in currently famous search engines with the proposed two-dimensional assessment scheme. Ten search engines are included in the assessment studies: About, AltaVista, Excite, Google, Inktomi, Lycos, MSN, Overture, Teoma, and Yahoo. To compute *RVV* vectors, the top *m*=10 URLs from search results are processed because it is shown that the first result screen is requested for 85% of the queries [16], and it usually shows the top ten results. To generate the norm, we used a weighted term-frequency-inversedocument-frequency (TF-IDF) strategy to select the feature information from the ten search engines. The size of *N* is thus adaptive to different queries to appropriately represent the norm.

We have conducted experiments to measure the biases of ten general search engines. The indexical bias is assessed according to the approach proposed by Mowshowitz and Kawaguchi [10][11][12]. The content bias is assessed according to the proposed weighted vector model. In the experiments, queries from different subjects were tested. Two of the experimental results are reported and discussed here. The first is a summarization of ten hot queries. This study shows the average bias performance of Web search engines according to their content bias and indexical bias values. The second is a case study on overwhelming redefinition power of search engines reported in [13]. In this experiment, the two-dimensional assessment shows that most

search engines report similar indexical and content bias ranking except Overture.

4.1 The Assessment Results of Hot Queries

In this experiment, we randomly chose ten hot queries from Lycos 50 [22]. For each of them, we collected 100 Web pages from ten search engines. The queries are "Final Fantasy", "Harry Potter", "Iraq", "Jennifer Lopez", "Las Vegas", "Lord of the Rings", "NASCAR", "SARS", "Tattoos", and "The Bible". The assessment results of their indexical bias and content bias values are shown in Table 1 and Table 2.
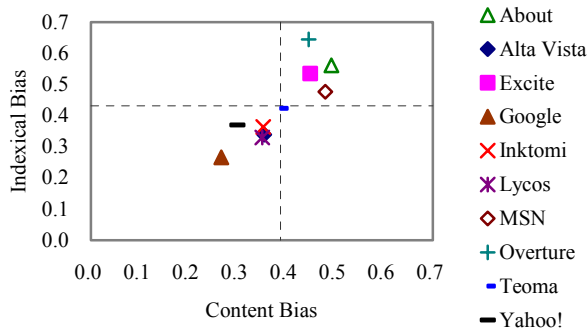

Figure 2: The two-dimensional analysis of the ten hot queries from Lycos 50

In Figure 2, the average bias performance is further displayed in a two-dimensional diagram. In the figure, two additional dotted lines are used to represent the respective statistic mean values of bias. The results show that Google has the lowest indexical and content bias value, which means that Google outperforms others in bias performance. The best bias performance in Google represents that both the sites and the contents it retrieved are the majority on the Web and may satisfy the most user needs. From the average results, we found that most of the search engines show similar bias rankings in both indexical bias and content bias.

However, when we review the bias performance of Yahoo!, we can see that it has quite good content bias performance, which is ranked as the second best, but only has a medium indexical bias ranking. Such insistent bias performance shows that Yahoo! can discover other similar major contents from different Web sites. However, such differences cannot be revealed when users only consider the indexical bias as the panorama of search engine bias. In our experiments, a one-way analysis of variance (ANOVA) was conducted to analyze the statistical significance on bias performance among each search engine. The ANOVA analyses in Table 5 and Table6 indicate that the content bias of Yahoo! is more statistically significant than the indexical bias.

In Table 3 and Table4, the ANOVA results of the averaged indexical bias and content bias are presented to display the statistical significance between the experimental search engines. Both of the ANOVA results reveal statistical significance of the ten search engines over the hot query terms ($p \leq 0.05$). The p-values in the table measure the credibility of the null hypothesis. The null hypothesis here means that there is no significant difference between each search engine. If the p-value is less than or equal to the widely accepted value 0.05, the null hypothesis is rejected.

Since there is significant difference among the search engines, we further analyze the variance across different hot query terms. Table 5 and Table 6 show the ANOVA results of indexical bias and content bias between each search engine over the ten hot query terms. Table 5 further indicates that About, AltaVista, Google, Lycos, and Overture are significant, and Table 6 presents that About, Google, MSN, and Yahoo! are significant. From the ANOVA analyses, the original indexical bias of MSN and Yahoo! is less significant, but the content bias assessment can reveal the complementary information. The two-dimensional assessment scheme tells users a panoramic view of search engine bias.

Table 1: The indexical bias of ten hot queries randomly chosen from Lycos 50.

Queries	About	AltaVista	Excite	Google	Inktomi	Lycos	MSN	Overture	Teoma	Yahoo!
Final Fantasy	0.5895	0.1876	0.5194	0.1876	0.3488	0.2403	0.4339	0.7054	0.4573	0.2713
Harry Potter	0.5669	0.3098	0.5837	0.2253	0.3098	0.3275	0.4299	0.7758	0.3755	0.4181
Iraq	0.7231	0.2560	0.5328	0.3252	0.2733	0.3771	0.4809	0.3771	0.4463	0.4290
Jennifer Lopez	0.5878	0.3681	0.5835	0.2606	0.3864	0.2448	0.5123	0.3078	0.3550	0.2134
Las Vegas	0.6985	0.3439	0.5921	0.1488	0.2375	0.3793	0.5744	0.8049	0.3261	0.2552
Lord of the Rings	0.5493	0.2558	0.5659	0.2074	0.2924	0.2093	0.4418	0.7829	0.3953	0.2093
NASCAR	0.3745	0.3897	0.4318	0.2982	0.3816	0.4150	0.4652	0.7493	0.4819	0.2829
SARS	0.4206	0.4902	0.3309	0.2874	0.4743	0.4902	0.3526	0.6655	0.5691	0.5018
Tattoos	0.5017	0.3355	0.6543	0.3995	0.5633	0.2903	0.4177	0.5847	0.4177	0.4905
The Bible	0.6059	0.4518	0.5546	0.3148	0.3662	0.3245	0.6511	0.6917	0.3995	0.6247
Average:	0.5618	0.3388	0.5349	0.2655	0.3634	0.3298	0.4760	0.6445	0.4224	0.3696

Table 2: The content bias of ten hot queries randomly chosen from Lycos 50.

Queries	About	AltaVista	Excite	Google	Inktomi	Lycos	MSN	Overture	Teoma	Yahoo!
Final Fantasy	0.5629	0.4535	0.3315	0.3507	0.5545	0.2724	0.4396	0.2961	0.5030	0.3481
Harry Potter	0.5315	0.3028	0.4498	0.3181	0.4985	0.3555	0.4461	0.4346	0.3332	0.5443
Iraq	0.4301	0.1651	0.5557	0.2250	0.1605	0.2213	0.5390	0.4403	0.2461	0.1711
Jennifer Lopez	0.4723	0.4193	0.4524	0.3150	0.5921	0.3450	0.3959	0.2441	0.3914	0.3138
Las Vegas	0.4656	0.4252	0.3303	0.1831	0.1971	0.2080	0.5267	0.5286	0.2201	0.2036
Lord of the Rings	0.5853	0.2030	0.2622	0.1516	0.1801	0.1966	0.5129	0.4509	0.2440	0.1573
NASCAR	0.3318	0.2210	0.4724	0.1743	0.1995	0.2195	0.5005	0.6139	0.2515	0.1950
SARS	0.4373	0.6965	0.5769	0.3784	0.6521	0.7361	0.4259	0.5443	0.6819	0.3854
Tattoos	0.5270	0.4733	0.4989	0.3351	0.3145	0.3425	0.3472	0.3732	0.3907	0.4654
The Bible	0.5829	0.1874	0.5639	0.2394	0.1815	0.6096	0.6647	0.5358	0.6202	0.2126
Average:	0.4927	0.3547	0.4494	0.2671	0.3530	0.3507	0.4798	0.4462	0.3882	0.2997

Table 3: ANOVA result of the indexical bias between Web search engines

	Sum of Squares	Degree of Freedom	Mean Square	F-ration	p-value
Between Groups	1.301	9	0.145	12.687	0.000
Within Groups	1.025	90	0.011		
Total	2.326	99			

Table 4: ANOVA result of the content bias between Web search engines

	Sum of Squares	Degree of Freedom	Mean Square	F-ration	p-value
Between Groups	0.527	9	0.059	3.036	0.003
Within Groups	1.736	90	0.019		
Total	2.263	99			

Table 5: ANOVA result of the indexical bias across hot terms

Engine	About	AltaVista	Excite	Google	Inktomi	Lycos	MSN	Overture	Teoma	Yahoo!
p-value	0.002	0.023	0.089	0.000	0.072	0.014	0.163	0.000	0.429	0.092

Table 6: ANOVA result of the content bias across hot terms

Engine	About	AltaVista	Excite	Google	Inktomi	Lycos	MSN	Overture	Teoma	Yahoo!
p-value	0.010	0.232	0.089	0.003	0.221	0.206	0.021	0.101	0.499	0.025

4.2 The Case of "Second Superpower"

To further assess the bias event happening on the Web, we used a real Googlewashed event happening on the Web to assess the bias performance of Web search engines. In this experiment, we once retrieved the search results and the Web pages from these ten search engines about one month later after the event happened. As reported in [13], Tyler's original concept of "Second Superpower" was flooded by Google with Moore's alternative definition in seven weeks. As a matter of fact, the idea of "second superpower" first appeared in the New York Times written by Tyler to describe the global anti-war protests [17]. After a while, Moore's essay used the term to describe another totally different meaning, the influence of the Internet and other interactive media [9].

In Figure 3, the two-dimensional assessment result shows that the Googlewashed effect indeed lowers the bias performance of Google. The two-dimensional analysis also reflects that the Googlewashed effect was perceptible to Google and Yahoo! since Yahoo! once cooperated with Google at that time (Actually, Yahoo is the same to Google in this query).
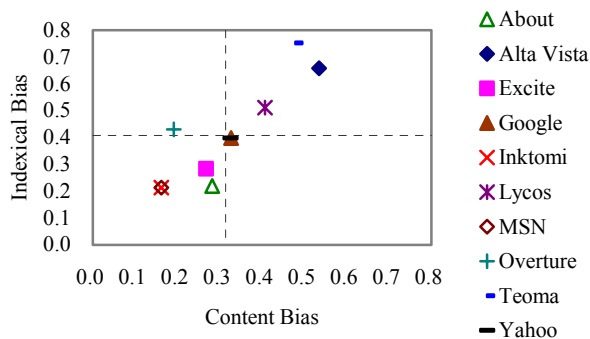
Figure 3: The bias result of "Second Superpower"

Interestingly, Figure 3 shows that the indexical bias ranking of Overture is relatively higher than its content bias. After manually reviewing the total of 100 Web pages for this query, we discovered that there are actually several definitions about "Second Superpower," not just Tyler's and Moore's. Although most contents retrieved by Overture point to the major viewpoints appearing in the norm, they are retrieved from diverse URLs but not mirror sites, and thus the search results incur a high indexical bias value. In this study, it shows that the indexical bias cannot tell us the whole story, but a two-dimensional scheme reflects a more comprehensive view of search engine bias.

5. CONCLUSION

Since Web search engines have become an essential gateway to the Internet, their favor or bias of Web contents has deeply affected users' browsing behavior and may influence their sight of viewing the Web. Recently, some studies of search engine bias have been proposed to measure the deviation of sites retrieved by a Web search engine from the norm for each specific query. These studies have presented an efficient way to assess search engine bias. However, such assessment method ignores the content information in Web pages and thus cannot present the search engine bias thoroughly.

In this paper, we assert that both indexical bias and content bias are important to present search bias. Therefore, we study the content bias existing in current popular Web search engines and propose a two-dimensional assessment scheme to complement the lack of indexical bias. The experimental results have shown that such a two-dimensional scheme can notice the blind spot of one-dimensional bias assessment approach and provide users with a more thorough view to search engine bias. Statistical analyses further present that such a two-dimensional scheme can fulfill the task of bias assessment and reveal more advanced information about search engine bias.

6. REFERENCES

[1] Brin, S., and Page, L., The Anatomy of Large-Scale Hypertextual Web Search Engine. In *Proceedings of the 7th International World Wide Web Conference* (Brisbane, Australia, 1998), ACM Press, New York, 107-117.

[2] Chen, I.-X. and Yang, C.-Z., Evaluating Content Bias and Indexical Bias in Web Search Engines. In *Proceedings of International Conference on Informatics, Cybernetics and Systems* (ICICS 2003) (Kaohsiung, Taiwan, ROC, 2003), 1597-1605.

[3] Gikandi D., Maximizing Search Engine Positioning (April 2, 1999); www.webdevelopersjournal.com/articles/search_engines.html

[4] Google Censors Itself for China. *BBC News* (Jan. 26, 2006); news.bbc.co.uk/1/hi/technology/4645596.stm.

[5] Holsti, O.R., *Content Analysis for the Social Science and Humanities*. 1st ed. Addison-Wesley Publishing Co., 1969.

[6] Introna, L. and Nissenbaum, H., Shaping the Web: Why the Politics of Search Engines Matters, *The Information Society*, 16, 3 (2000), 1-17.

[7] iProspect Search Engine User Attitudes (April-May, 2004); www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf.

[8] Jenkins, C., and Inman, D., Adaptive Automatic Classification on the Web. In *Proceedings of the 11th International Workshop on Database and Expert Systems Applications* (Greenwich, London, U.K., 2000), 504-511.

[9] Moore, J.F., The Second Superpower Rears its Beautiful Head (March 31, 2003); cyber.law.harvard.edu/people/jmoore/secondsuperpower.html.

[10] Mowshowitz, A., and Kawaguchi, A., Assessing Bias in Search Engines. *Information Processing & Management*, 38, 1 (Jan. 2002), 141-156.

[11] Mowshowitz, A., and Kawaguchi, A., Bias on the Web. *Commun. ACM*, 45, 9 (Sep. 2002), 56-60.

[12] Mowshowitz, A., and Kawaguchi, A., Measuring Search Engine Bias. *Information Processing & Management*, 41, 5 (Sep. 2005), 1193-1205.

[13] Orlowski, A., Anti-war Slogan Coined, Repurposed and Googlewashed . . . in 42 Days. *The Register* (April 3, 2003); www.theregister.co.uk/content/6/30087.html.

[14] Raggett, D., Getting Started with HTML, *W3C Consortium* (May 24, 2005); www.w3.org/MarkUp/Guide/.

[15] Salton, G., Wong, A., and Yang, C. S., A Vector Space Model for Automatic Indexing. *Commun. ACM*, 18, 11 (Nov. 1975), 613-620.

[16] Silverstein, C., Henzinger, M., Marais, H., and Moricz, M., Analysis of a Very Large AltaVista Query Log, *ACM SIGIR Forum*, 33, 1 (Fall 1999), 6-12.

[17] Tyler, P.E., A New Power in the Streets. *New York Times* (Feb. 17, 2003); foi.missouri.edu/voicesdissent/newpower.html.

[18] Vaughan, L. and Thelwall, M., Search Engine Coverage Bias: Evidence and Possible Causes, *Information Processing & Management*, 40, 4, (July 2004), 693-707.

[19] Zittrain, J. and Edelman, B., Documentation of Internet Filtering in Saudi Arabia, (Sep. 12, 2002); cyber.law.harvard.edu/filtering/saudiarabia/.

[20] Zittrain, J. and Edelman, B., Localized Google search result exclusions, (Oct. 26, 2002); cyber.law.harvard.edu/filtering/google/.

[21] Zittrain, J. and Edelman, B., Internet Filtering in China. *IEEE Internet Computing*, 7, 2 (March/April, 2003), 70-77.

[22] 50.lycos.com.