22nd International Conference on Advanced Information Networking and Applications - Workshops

What's going on in search engine rankings?

Yasuaki Yoshida†, Takanori Ueda†, Takashi Tashiro†, Yu Hirate†,††and Hayato Yamana†,†††
†Graduate School of Fundamental Science and Engineering, Waseda University
3–4–1 Okubo, Shinjuku-ku, Tokyo 169–8555, Japan
††Media Network Center, Waseda University, 1–104 Totsuka-cho, Shinjuku-ku, Tokyo 169–8050, Japan
†††National Institute of Informatics, 2–1–2 Hitotsubashi, Chiyoda-ku, Tokyo 101–8430, Japan
{yoshida, ueda, ttashiro, hirate, yamana}@yama.info.waseda.ac.jp

Abstract

Many people use search engines every day to retrieve documents from the Web. Although the social influence of search engine rankings has become significant, ranking algorithms are not disclosed. In this paper, we have investigated three major search engine rankings by analyzing two kinds of data. One is the weekly ranking snapshots of top 250 Web pages we collected for almost one year by submitting 1,000 pre-selected queries; the other comprises back-linked Web pages gathered by our own Web crawling. As a result, we have confirmed that (1) several top 10 rankings are mutually similar, however, the following ranked Web pages are almost different, (2) ranking transitions have their own characteristics, and (3) each search engine's ranking has its own correlation with the number of back-linked Web pages.

1. Introduction

In recent years, since the number of Web pages has been increasing rapidly, it has become almost impossible to obtain information from the Web without using search engines. Since most people refer only to high-ranked Web pages, such as top 10 results, the information that we obtain from the Web might be biased according to search engine rankings. For example, if the Web pages that support one side's opinion about a topic were ranked high by search engines and the others were not, our information obtained from the Web might be biased to the former opinion. If this happens frequently, then it can be said that our knowledge is biased according to search engine rankings. Since search engine rankings have a great influence, it is important to understand how search engines rank Web pages and to check whether their rankings are biased.

To investigate the bias of search engines, we focused on the features of search engine rankings. In this study, we investigated the features of three major search engine rankings, Google [4], MSN [8], and Yahoo! JAPAN [12], by comparing ranking bias and correlation between rankings and back-links.

Ranking Bias: We collected weekly snapshots of top 250 ranked Web pages from Google, MSN, and Yahoo! JAPAN by submitting 1,000 pre-selected queries. The queries were chosen from 10 separate categories. In addition, to analyze the collected ranking snapshots, we propose a novel measure that indicates the bias of search engine rankings.

Correlation Between Rankings and Back-Links: We gathered back-linked Web pages, i.e., Web pages linked to top 250 ranked Web pages, using our own crawlers produced in the e-Society project [3]. Then, we investigated the correlation between their rankings and the number of back-links.

The remainder of this paper is organized as follows. In Section 2, we introduce related works. In Section 3, we propose a new method for evaluating the search engine rankings. In Section 4, collected ranking snapshots are described. Then, evaluation results are presented in Section 5, and conclusion is presented in Section 6.

2. Related Works

Bias, the measure of difference from an ideal ranking, which is approximated by the distribution of search results produced by a collection of search engines, was proposed in 2005 [18]. The following is the definition of *Bias*. Let $E = \{e_1, \cdots, e_n\}$ be a set of search engines, and $Q = \{q_1, \cdots, q_t\}$ be a set of queries. Then, $S_{i,j}$ is a set of URLs that are included in the ranking result obtained by submitting query q_j to search engine e_i. In addition, S is a union of all $S_{i,j}(1 \leq i \leq n, 1 \leq j \leq t)$. Let the elements of S be $u_1, \cdots, u_{|S|}$, where their order is not concsidered. Now,

978-0-7695-3096-3/08 $25.00 © 2008 IEEE
DOI 10.1109/WAINA.2008.114

let $\boldsymbol{r}_{i,j} = [x_{i,j,1}, x_{i,j,2}, \cdots, x_{i,j,|S|}] (1 \leq i \leq n, 1 \leq j \leq t)$ be a one-dimensional vector that represents the ranking result by submitting query q_j to search engine e_i. Then, each element $x_{i,j,k}$ of $\boldsymbol{r}_{i,j}$ is represented by expression (1).

$$x_{i,j,k} = \begin{cases} 0 & (u_k \notin S_{i,j}) \\ 1 & (u_k \in S_{i,j}) \end{cases} \tag{1}$$

$\boldsymbol{R}_i$, the ranking result vector of search engine e_i, and $\boldsymbol{R}$, the ideal ranking result vector, are defined by expressions (2) and (3).

$$\boldsymbol{R}_i = \sum_{j=1}^{t} \boldsymbol{r}_{i,j} \tag{2}$$

$$\boldsymbol{R} = \sum_{i=1}^{n} \boldsymbol{R}_i \tag{3}$$

The cosine similarity expression (4) is adopted as a measure of similarity between vector $\boldsymbol{R}_i$ and $\boldsymbol{R}$.

$$s(\boldsymbol{v}, \boldsymbol{w}) = \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{|\boldsymbol{v}||\boldsymbol{w}|} \tag{4}$$

The *Bias* $b(e_i)$ of search engine e_i is defined as "1 minus the cosine similarity between the ideal ranking result and the ranking result of e_i," as shown in expression (5).

$$b(e_i) = 1 - s(\boldsymbol{R}_i, \boldsymbol{R}) \tag{5}$$

A large value of $b(e_i)$ signifies "high-biased" ranking, while a small value of $b(e_i)$ denotes "low-biased" ranking.

In [18], ranking snapshots were collected by submitting 50 queries to 16 search engines. The authors concluded that the ranking results of Ah-ha, MSN, Sprinks, Teoma, and TrueSearch have high *Bias*, while that of AOLSearch, FastSearch, Google, Netscape, and Yahoo! have low *Bias*. However, the ranks of results are not considered in [18]. Since most users of search engines refer only to the top ranked Web pages, the *Bias* measure fails to reflect users' behavior. In addition, the ranking transitions are not investigated, and therefore, the investigation of [18] is inadequate.

Another study on the transition of search engine rankings was initially conducted in 2006 [13]. In this study, the authors collected daily ranking snapshots of top 10 results by submitting three queries to Google, Yahoo!, and Teoma in the case of page search, and by submitting two queries to Google Image, Yahoo! Image, and Picsearch in the case of image search. The collected ranking snapshots were analyzed using $overlap$, Spearman's footrule [14, 15], Fagin's measure [16], and M measure [13]. However, the results of the investigation are unreliable because of the relatively small scale of the investigation, which comprised only five queries.

The differences between the Web User Interface (WUI) rankings and the Application Programming Interface (API) rankings were investigated in 2007 [17]. Ranking snapshots were collected by submitting 100 queries: 50 queries were selected from general terms, and 50 queries were selected from computer science terms. The authors collected daily snapshots of top 100 rankings with regard to the 100 queries by submitting to Google, MSN, and Yahoo!. The collected ranking snapshots were analyzed using $overlap$, Fagin's measure [16], and M measure [13]. In addition, the number of back-links to 100 URLs was compared. The differences in the indexes of the search engines between WUI and API were estimated by counting the number of indexed URLs in the 100 sites and computing the ratio of indexed URLs. However, the results of the investigation were unreliable because of the small scale of the investigation, which comprised only 100 queries. In addition, while the main purpose of [17] was to compare search engine rankings between WUI and API, our main purpose is to draw comparisons between search engine rankings more generally.

As described above, there have been problems with previous studies when investigating search engine rankings, such as the defining a measurement that is not a reflection of users' behavior [18], the relatively small scale of the investigations [13, 17], and difference in the purpose of investigations, e.g., API $vs.$ WUI [17]. On the other hand, our investigation adopted large-scale data such as weekly snapshots of the top 250 rankings for almost one year, and also adopts a novel measure that reflects users' behavior.

3. Rank Weighted Bias

In this section, we propose a new bias measure to reflect users' behavior, called "*Rank Weighted Bias*." The measurement is based on *Bias* [18].

The ranking results are simply regarded as a set of URLs in the definition of *Bias* [18]. This means that the ranking positions of URLs are not considered. As we described above, most users refer only to high-ranked results, i.e., high-ranked Web pages. When the high-ranked results of two rankings resemble each other, the two rankings seem to provide similar results for users even if the low-ranked results are different. On the other hand, even if the low-ranked results are similar, the two rankings are totally different for users when the high-ranked results are different. This means that the bias measure should consider a concept of ranking in order to reflect users' behavior.

Rank Weighted Bias is extended from the *Bias* measure [18] with regard to each rank of the resultant URLs. The following is its definition. In *Rank Weighted Bias*, the value of the element in the ranking result vector $\boldsymbol{r}_{i,j}$, i.e., $x_{i,j,k}$, is replaced with the weighted values $x'_{i,j,k}$. The weighted value $x'_{i,j,k}$ is defined in expression (6).

$$x'_{i,j,k} = \begin{cases} 0 & (u_k \notin S_{i,j}) \\ m - p_{i,j,k} + 1 & (u_k \in S_{i,j}) \end{cases} \tag{6}$$

where

m : the maximum number of collected URLs per query
$p_{i,j,k}$ *: the rank of* u_k *(query:* q_j*, search engine:* e_i*)*

For example, when we collect the top 250 ranked URLs, the top ranked URL has a value of 250, the second ranked URL has a value of 249, etc. In this way, $\boldsymbol{r}_{i,j}$ is also replaced with $\boldsymbol{r'}_{i,j}$.

The remainder of the definition is the same as that of *Bias*. The *weighted* ranking result vector $\boldsymbol{R_w}_i$ of search engine e_i and the *weighted* ideal ranking result vector $\boldsymbol{R_w}$ are defined by expressions (7) and (8).

$$\boldsymbol{R_w}_i = \sum_{j=1}^{t} \boldsymbol{r'}_{i,j} \tag{7}$$

$$\boldsymbol{R_w} = \sum_{i=1}^{n} \boldsymbol{R_w}_i \tag{8}$$

The *Rank Weighted Bias* $b_w(e_i)$ of search engine e_i is defined as "1 minus the cosine similarity between the *weighted* ideal ranking result and the *weighted* ranking result of search engine e_i," as shown in expression (9).

$$b_w(e_i) = 1 - s(\boldsymbol{R_w}_i, \boldsymbol{R_w}) \tag{9}$$

In the same way as for *Bias* $b(e_i)$, a large value of $b_w(e_i)$ denotes "high-biased" ranking, while a small value of $b_w(e_i)$ signifies "low-biased" ranking.

Extension to Measure Ranking Transition

We also propose a quantitative measure for ranking transition. Here, we extend $\boldsymbol{R}_i$ and $\boldsymbol{R_w}_i$ to indicate the time period when the ranking result is collected. Let $\boldsymbol{R}_{i,t}$ be a vector that consists of the ranking result retrieved by search engine e_i at the time period $t (t \in N)$. In the same way, let $\boldsymbol{R_w}_{i,t}$ be a vector that consists of weighted ranking results retrieved by search engine e_i at the time period $t (t \in N)$. The similarity between $\boldsymbol{R}_{i,t}$ and $\boldsymbol{R}_{i,t+1}$ is defined by expression (10) using cosine similarity.

$$Sim(e_i, t) = s(\boldsymbol{R}_{i,t}, \boldsymbol{R}_{i,t+1}) \tag{10}$$

Moreover, the rank weighted similarity between $\boldsymbol{R_w}_{i,t}$ and $\boldsymbol{R_w}_{i,t+1}$ is defined by expression (11).

$$Sim_w(e_i, t) = s(\boldsymbol{R_w}_{i,t}, \boldsymbol{R_w}_{i,t+1}) \tag{11}$$

Both Sim and Sim_w enable us to quantitatively evaluate the ranking transition.

4. Web Ranking Snapshots

In this section, we demonstrate the method of collecting the ranking snapshots that we have analyzed.

We selected Google, MSN, and Yahoo! JAPAN as the search engines for the investigation because these three search engines have large market shares in Japan. We used their APIs to collect search results [5, 1, 11]. We submitted 1,000 pre-selected queries to each search engine once a week[1] and collected the top 250 ranking results.

Query Selection Process

To evaluate the ranking results precisely, we selected 1,000 queries in total from 10 categories (100 queries/category) in August 2006.

First, we selected 10 categories: "Animals," "Company Names," "Plants," "Universities," "Food Names," "Trademarks," "Names of Japanese People," "Product Names," "Names of English People," and "Place Names." Then, in each category, we gathered words from conventional dictionaries and Web sites. We used a Japanese trademark database [6] for the "Trademarks" category. We used Yahoo! JAPAN [12], Livedoor [7], and Happy Market [9] for the "Product Names" category, and for the other categories, we used Wikipedia [10].

To select 100 queries from each category unintentionally, we adopted various popularities as the criterion. Here, we simulated "popularities" to the number of ranking results. The detailed selection process is as follows:

1. Submit all the gathered words to Yahoo! JAPAN as queries, and collect the number of result pages to simulate their popularities. Here, the most popular word returns m Web pages.

2. In each category, divide a set of words into $[\log_{10} m]$ groups in the following manner: The n-th group consists of a set of words whose number of search results is from 10^{n-1} to $10^n - 1$, where $1 \leq n \leq [\log_{10} m]$.

3. In each category, select $\frac{100}{[\log_{10} m]}$ queries randomly from every group.

Table 1 shows the number of words in each category and the number of selected words for the queries in each category. Fig. 1 shows the ratio of Japanese queries and English queries that were selected.

5. Experimental Results

In this section, we present the experimental results. All ranking snapshots were collected from October 2006 to September 2007. Each period of collection is shown in Table 2. As shown in Table 2, each period is labeled with a unique symbol from T1 to T22.

[1] In fact, the intervals of ranking snapshots are not weekly because of the fluctuation of time when collecting ranking results.

Table 1. The number of words in each category and the number of selected words for queries

	Num. of words in categories	Num. of selected words
Animals	10123	107
Company Names	2864	100
Plants	433	101
Universities	963	100
Food Names	505	100
Trademarks	745	106
Names of Japanese People	14097	100
Product Names	194	100
Names of English People	10273	100
Place Names	1976	100

Table 2. Observation periods and their symbols

period	symbol	period	symbol
10/12/06-10/16/06	T1	03/23/07-03/30/07	T12
10/22/06-10/26/06	T2	04/06/07-04/11/07	T13
10/31/06-11/05/06	T3	04/14/07-04/18/07	T14
11/06/06-11/11/06	T4	06/12/07-06/16/07	T15
11/15/06-11/25/06	T5	06/23/07-06/28/07	T16
11/30/06-12/11/06	T6	06/30/07-07/09/07	T17
01/01/07-01/06/07	T7	07/11/07-07/16/07	T18
01/13/07-01/20/07	T8	07/21/07-07/27/07	T19
01/24/07-01/28/07	T9	08/01/07-08/08/07	T20
02/01/07-02/07/07	T10	08/21/07-08/29/07	T21
02/15/07-02/20/07	T11	09/02/07-09/10/07	T22

5.1. Rankings Bias

Both the *Bias* and the *Rank Weighted Bias* of all search engines were compared. We computed the average value of the *Bias* and *Rank Weighted Bias* of each search engine in every category throughout the collection periods. Fig. 2 shows the average value of the *Bias* and *Rank Weighted Bias* of each search engine. Fig. 2 shows the following three points. First, all search engines have small values of *Rank Weighted Bias* in comparison with those of *Bias* in every category. This means that the ranking results of each search engine tend to be similar to each other in higher rankings. Second, Google shows the biggest difference between *Bias* and *Rank Weighted Bias* compared to the other search engines. This means that Google returns original low-ranked results, while its high-ranked results are similar to those of

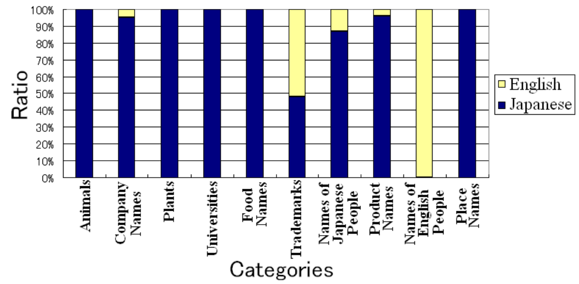


Figure 1. Ratio of queries written in Japanese and English

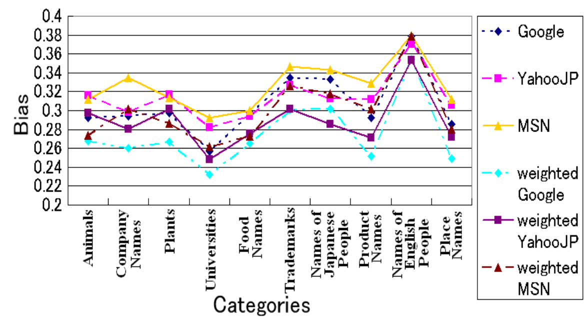


Figure 2. Average *Bias* and *Rank Weighted Bias* in all categories

the other search engines. Third, each search engine has different tendencies in both *Bias* and *Rank Weighted Bias* depending on the category; MSN has the highest *Bias* in most categories. Yahoo! JAPAN has the highest *Bias* in the "Animals" and "Plants" categories.

In the "Trademarks," "Names of Japanese People," and "Names of English People" categories, all three search engines have higher *Bias* than in the other categories, since these three categories contain more English queries than the others, as shown in Fig. 1. In fact, this is shown in Fig. 3. Fig. 3 shows the values of *Bias* and *Rank Weighted Bias* computed using only Japanese queries and only English queries in the "Trademarks" category that contains Japanese and English queries in almost the same ratio. All search engines show lower *Bias* for Japanese queries than for English queries.

5.2. Ranking Transition

We evaluated the ranking transition of each search engine with Sim and Sim_w (defined in Section 3). Fig. 4 shows the average similarity value in all the categories. All three search engines show larger values of Sim_w compared to Sim. This means that the ranking results for higher rank-

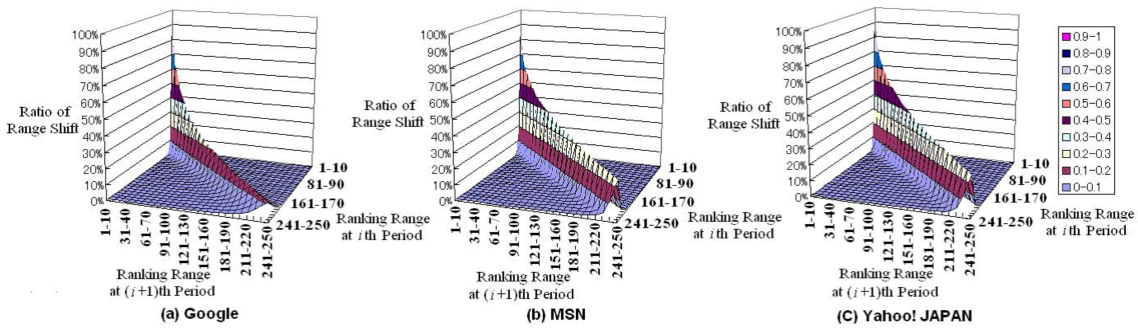


Figure 5. Distribution of ranking shift

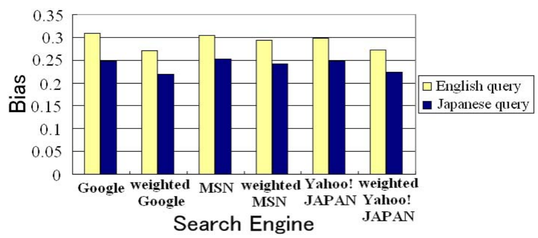


Figure 3. $Bias$ **in Japanese queries and English queries**

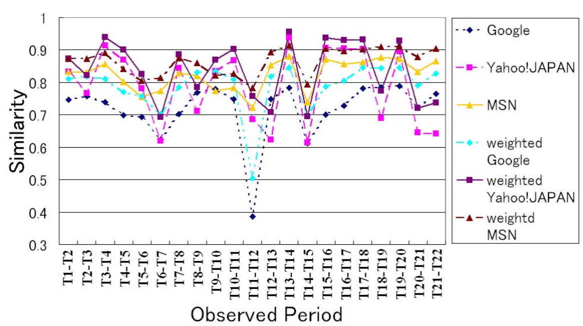


Figure 4. Ranking transition with Sim **and** Sim_w

ings tend to resemble each other more than those for lower rankings. In addition, Google always exhibits a low value of similarity. This shows that the ranking transition of Google is the most intense. The troughs of the graph, such as T6-T7 or T14-T15 of Yahoo! JAPAN, means that the ranking results changed significantly. [2]

To analyze the ranking transition in detail, we compared the ranking results of 1,000 queries with the same search

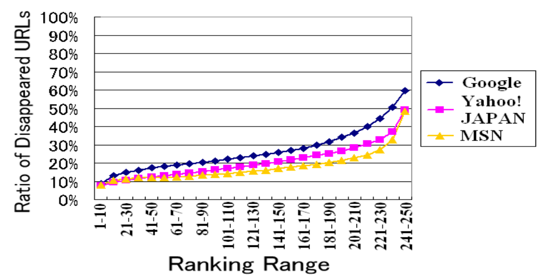


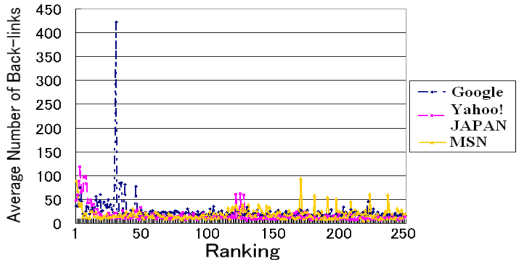
Figure 6. Ratio of the missing URLs from top 250 in the next period

engine in the i-th and $(i+1)$-th periods, where $1 \leq i \leq 21$. For effective summarization, the top 250 ranked results were divided into 25 ranges, such as the top 1 to 10, the top 11 to 20, $\cdots$, up to the top 241 to 250. Then, for all i-th period results, we found out their changed ranking positions in the $(i+1)$-th period in order to check their ranking shifts. The results are shown in Fig. 5. Moreover, Fig. 6 shows the ratio of the missing URLs from the top 250 ranked results in the next period. Figs. 5 and 6 indicate the following two points. First, high-ranked URLs tend to stay in their ranking position, while low-ranked URLs change their ranking position dramatically. Second, while all search engines display a peak on their respective diagonal lines in Fig. 5, Google has the lowest peak among the three search engines. This shows that Google changes its ranking results the most.

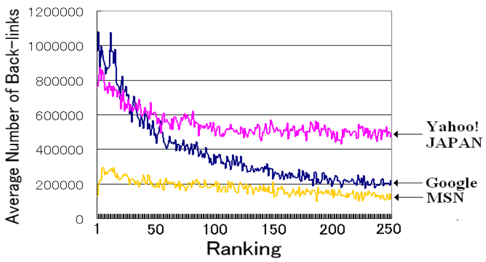
5.3. Correlation Between Rankings and Back-Links

To analyze the feature of ranking, we investigated the correlation between rankings and the number of back-links. This is because major ranking algorithms, such as PageRank [19], utilize back-links to rank their search results. To compute the number of back-links, we used our crawled

[2] There is also a trough at T11-T12 for Google. We changed the search options of Google during this term, which might have resulted in the low similarity.



(a) Back-Links per Page



(b) Back-Links per Site

Figure 7. Correlation between the number of back-links and the ranking

Web pages gathered by the e-Society project [2] from September 2006 to February 2007. In this investigation, we selected about 160 million Web pages randomly from the 1.4 billion gathered Web pages in order to simplify the calculation of the number of back-links.

Figs. 7 show the correlation between rankings and the number of back-links per page (a), and per site (b), respectively. In Fig. 7 (a), ranking results show no correlation with the number of back-links per page.[3] There are two possible reasons for this: (1) The page set gathered by e-Society does not accord with the indexes of search engines and (2) rankings are not computed according to back-links per page. However, we can see a correlation in Fig. 7 (b). Sites whose pages are in the higher rankings tend to have more back-links per site. This implies that the rankings are computed according to back-links per site. In addition, Google shows the strongest correlation, while MSN displays the weakest correlation.

6. Conclusion

In this paper, we have investigated the features of the rankings of three major search engines, Google, MSN, and Yahoo! JAPAN, using the concept of *Rank Weighted Bias* proposed in this paper. Our experimental results reveal the following three points. (1) Results ranked in high positions by each search engine tend to be similar to each other. (2) Results ranked in high positions, such as the top 10, have a high probability of retaining their positions. (3) All search engines display their correlation between rankings and the number of back-links; Google exhibits a strong correlation, while MSN shows a weak correlation.

[3] Google shows a peak value around the 30th ranking. This anomaly may be a result of the low coverage of our Web crawling dataset.

Acknowledgement

This work was initially supported by METI as a part of prior research of Information Grand Voyage Project. After that, this work has been supported by MEXT both as a part e-Society project and as a part of info-plosion project.

References

[1] Developer center. http://search.msn.com/developer/default.aspx.
[2] E-society project: About our web crawler. http://www.yama.info.waseda.ac.jp/~yamana/es/index_eng.htm.
[3] E-society project: the development of fundamental software. http://cif.iis.u-tokyo.ac.jp/e-society/.
[4] Google. http://www.google.co.jp/.
[5] Google soap search api. http://code.google.com/apis/soapsearch/.
[6] Japanese trademark database. http://www1.ipdl.ncipi.go.jp/chomei/search_j.cgi.
[7] Livedoor department store: product category list. http://depart.livedoor.com/all_item.
[8] Msn japan. http://jp.msn.com/.
[9] Product category list. http://www.hapima.com/prdct/.
[10] Wikipedia. http://ja.wikipedia.org/wiki/.
[11] Yahoo! developer network. http://developer.yahoo.co.jp/.
[12] Yahoo! japan. http://www.yahoo.co.jp/.
[13] J. Bar-Ilan, M. Mat-Hassan, and M. Levene. Methods for comparing rankings of search engine results. *Computer Networks*, Vol.50:pp.1448–1463, 2006.
[14] P. Diaconis and R. L. Graham. Spearman's footrule as a measure of disarray. *J. of the Royal Statistical Society, Series B (Methodological)*, Vol.39:pp.262–268, 1977.
[15] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proc. of the 10th WWW Conf.*, pages pp.613–622, May 2001.
[16] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. *SIAM J. on Discrete Mathematics*, Vol.17:pp.134–160, 2003.
[17] F. McCown and M. L. Nelson. Agreeing to disagree: Search engines and their public interfaces. In *ACM IEEE Joint Conf. on Digital Libraries (JCDL 2007)*, 2007.
[18] A. Mowshowitz and A. Kawaguchi. Measuring search engine bias. *Information Processing and Management*, Vol.41:pp.1193–1204, Sep. 2005.
[19] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford University, 1998.