Aggregating Subjective Measure of Web Search Quality with *PageRank*

Rashid Ali
Department of Computer Engineering
A. M. U., Aligarh-202002, India
rashidaliamu@rediffmail.com

M. M. Sufyan Beg
Department of Computer Science
University of California, Berkeley, CA 94720
mmsbeg@cs.berkeley.edu

Abstract - *Web Searching is one of the most popular activities on Internet. But as a number of search engines are available, there must be some procedure to evaluate them. In this paper, we present an effort in this regard. We intend to develop an evaluation system for web search results. We are taking into the consideration the "satisfaction" user gets when presented with search results. The feedback of the user is inferred from watching the actions of the user on the search results presented before him in response to his query. The implicit ranking given by the user is then compared with the original ranking given by the search engine and correlation coefficient is obtained. Then, parsing of URLs of web-search results is performed and PageRank is computed. Thus, we get a new ranking of the documents based on the PageRank. This ranking is also compared with the original ranking given by the search engine and correlation coefficient thus obtained is averaged with that obtained using user feedback. We repeat our procedure for a set of queries. We show our experimental results pertaining to seven public search engines and fifteen queries.*

Keywords: web search evaluation, user feedback, *pagerank*, rank correlation coefficient

1 Introduction

Internet has been very much popular for searching purpose. Everyday, millions of users search the web for some data and information using some query. A number of public search engines are available for this purpose. In an Internet search, the user writes some query and the search engine responds by returning a number of web pages that match the description. Since different search engines use different search algorithms and indexing techniques, they return different web pages in response to same query. Also, same web page is ranked differently by different search engines and returned at different positions in the list of search results. Since majority of users do not like to see behind some top ranked results, we can call a search engine better if that gives more relevant results at top few positions in response to a query. For this, search results need to be evaluated. The evaluation procedure may be subjective or objective. In the present work, we propose a web search evaluation system, which combines both the subjective as well as objective technique. For subjective evaluation, we take user's vote into account. For objective evaluation, we use *PageRank*. To evaluate the search engines subjectively, feedback from the users is to be taken. This feedback may be explicit or implicit. If the user is asked to fill up a feedback form regarding quality of search results, we can say the feedback is explicit in nature. But the problem is to obtain a correct feedback. This approach is too demanding from the user. A casual user might either fill the form carelessly or not fill it at all. We, therefore, devised a method to obtain the implicit feedback from the users. We watch the actions of the user on the search results presented before him in response to his query, and infer the feedback of the user there from. The feedback gives a ranking of the document by the user, which can then be compared with original ranking by search engine. The correlation coefficient thus obtained is a measure of search quality based on user feedback. We augment the subjective evaluation technique based on implicit user feedback as mentioned in the preceding paragraph with an objective evaluation based on *PageRank.*

1.1 Related Work

In the past, some efforts have already been made to evaluate search results from different search engines. In [1], [2] and [3], the size of indices, which indirectly estimates the performance of a search engine, is measured using uniform sample of web pages collected by carrying out random walks on the web. A search engine having larger index size has higher probability to give good search results. In [4] also, the relative size and overlap of search engines is found but by using random queries, which are generated from a lexicon of about 400,000 words, built from a broad crawl of roughly 300,000 documents in the *Yahoo* hierarchy. In [5] and [6], the search engines are compared using a standard query log like that of NEC research institute. In [7], a frozen 18.5 million page snapshots of part of the web is created for proper evaluation of web search systems. In [8], for two different sets of ad-hoc queries, the results from *AltaVista*, *Google* and *InfoSeek* are obtained. These results are automatically evaluated for relevance on the basis of vector space model. These results are found to agree with the manual evaluation of relevance based on precision. Precision evaluation of search engines is reported in [9]. But then, "precision" doesn't say anything about the ranking of the relevant documents in the search results. Also, there is no discussion on the satisfaction a user gets when presented with the search results.

An attempt has been made in [10], to augment the implicit user feedback based subjective evaluation with vector space model based objective evaluation. User feedback is implicit in the sense that it is inferred by watching the actions of the user on the search results presented before them in response to his query. In [11], the subjective evaluation is augmented with Boolean similarity measure based objective evaluation. The objective evaluations based techniques used in both [10] and [11] are content-based techniques. But, the connectivity based objective techniques such as *Google*'s *PageRank* are also of great importance. Therefore, in the present effort we try to combine the subjective technique with *PageRank* based objective technique of web search evaluation in order to get some new and effective picture of web search evaluation.

1.2 Useful Definitions

Here, we list some definitions that are useful while evaluating search results.

Definition 1. Given a universe U and $S \subseteq U$, an ordered list (or simply, a list) l with respect to U is given as $l = [e_1, e_2, \ldots, e_{|s|}]$, with each $e_i \in S$, and $e_1 \succ e_2 \succ \ldots \succ e_{|s|}$, where "$\succ$" is some ordering relation on S. Also, for $j \in U \wedge j \in l$, let l(j) denote the position or rank of j, with a higher rank having a lower numbered position in the list. We may assign a unique identifier to each element in U and thus, without loss of generality we may get $U = \{1,2, \ldots, |U|\}$.

Definition 2. Full List: If a list contains all the elements in U, then it is said to be a full list.

Example 1. A full list l_f given as [e,a,d,c,b] has the ordering relation $e \succ a \succ d \succ c \succ b$. The Universe U may be taken as {1,2,3,4,5} with say a ≡ 1, b ≡ 2, c ≡ 3, d ≡ 4, e ≡ 5. With such an assumption, we have l_f = [5,1,4,3,2]. Here $l_f(5) \equiv l(e) = 1$, $l_f(1) \equiv l(a) = 2$, $l_f(4) \equiv l_f(d) = 3$, $l_f(3) \equiv l_f(c) = 4$, $l_f(2) \equiv l_f(b) = 5$.

Definition 3. Partial List: A list l_p containing elements, which are a strict subset of universe U, is called a partial list. We have a strict inequality $|l_p| < |U|$.

Definition 4. Spearman Rank Order Correlation coefficient [10]: Let the full lists $[u_1, u_2, \ldots u_n]$ and $[v_1, v_2, \ldots v_n]$ be the two rankings for some query Q. Spearman rank-order correlation coefficient (r_s) between these two rankings is defined as follows.

$$r_s = 1 - \frac{6\sum_{i=1}^{n}[l_f(u_i) - l_f(v_i)]^2}{n(n^2-1)} \quad (1)$$

The Spearman rank-order correlation coefficient (r_s) is a measure of closeness of two rankings. The coefficient r_s ranges between –1 and 1. When the two rankings are identical $r_s = 1$, and when one of the rankings is the inverse of the other then the $r_s = -1$.

Definition 5. Modified Spearman Rank Order Correlation coefficient: Without loss of generality, assume that full list be given as $[1,2,\ldots,n]$. Let the partial list be given as $[v_1, v_2, \ldots, v_m]$. Modified Spearman rank-order correlation coefficient (r_s') between these two rankings is defined as follows-

$$r_s' = 1 - \frac{\sum_{i=1}^{m}(i - v_i)^2}{m\left([\max_{j=1}^{m}\{v_j\}]^2 - 1\right)} \quad (2)$$

Example 2. For |U|=5, let the full list be l_f = {1,2,3,4,5} and the partial list l_p with |lp| = m =3 be l_p ={40,35,100}.

$$r_s' = 1 - \frac{(1-40)^2 + (2-35)^2 + (3-100)^2}{3\times\left([\max\{40,35,100\}]^2 - 1\right)} = 0.401$$

2 Web Search Evaluation Using User Feedback Vector Model

2.1 User Feedback Vector

The underlying principle of our approach [11] of subjective evaluation of search engines is to measure the "satisfaction" a user gets when presented with the search results. For this, we need to monitor the response of the user to the search results presented before him. We characterize the feedback of the user by a vector (*V,T,P,S,B,E,C*), which consists of the following.

(a) The sequence *V* in which the user visits the documents, $V=(v_1, v_2, \ldots v_N)$. If document *i* is the k^{th} document visited by the user, then we set $v_i = k$. If a document *i* is not visited by the user at all before the next query is submitted, the corresponding value of v_i is set to –1.

(b) The time t_i that a user spends examining the document *i*. We denote the vector $(t_1, t_2, \ldots, t_N)$ by *T*. For a document that is not visited, the corresponding entry in the array *T* is 0.

(c) Whether or not the user prints the document *i*. This is denoted by the Boolean p_i. We shall denote the vector $(p_1, p_2, \ldots, p_N)$ by *P*.

(d) Whether or not the user saves the document i. This is denoted by the Boolean s_i. We shall denote the vector $(s_1, s_2, \ldots, s_N)$ by S.

(e) Whether or not the user book-marked the document i. This is denoted by the Boolean b_i. We shall denote the vector $(b_1, b_2, \ldots, b_N)$ by B.

(f) Whether or not the user e-mailed the document v to someone. This is denoted by the Boolean e_i. We shall denote the vector $(e_1, e_2, \ldots, e_N)$ by E.

(g) The number of words that the user copied and pasted elsewhere. We denote the vector $(c_1, c_2, \ldots, c_N)$ by C.

The motivation behind collecting this feedback is the belief that a well-educated user is likely to select the more appropriate documents early in the resource discovery process. Similarly, the time that a user spends examining a document, and whether or not he prints, saves, bookmarks, e-mails it to someone else or copies and pastes a portion of the document, indicate the level of importance that document holds for the specified query.

2.2 Search Quality Measure (SQM) using User Feedback Vector

We propose to compute the following weighted sum σ_j for each document j selected by the user.

$$\sigma_j = \left(w_V \frac{1}{2^{(i-1)}} + w_T \frac{t_j}{t_j^{max}} + w_P p_j + w_S s_j + w_B b_j + w_E e_j + w_C \frac{c_j}{c_j^{total}} \right) \qquad (3)$$

Where t_j^{max} represents the maximum time a user is expected to spend in examining the document j, and c_j^{total} is the total number of words in the document j. Here, w_V, w_T, w_P, w_S, w_B, w_E and w_C, all lying between 0 and 1, give the respective weightages we want to give to each of the seven components of the feedback vector. The sum σ_j represents the importance of document j. The intuition behind this formulation is as follows. The importance of the document should decrease monotonically with the postponement being afforded by the user in picking it up. More the time spent by the user in glancing through the document, more important that must be for him. If the user is printing the document, or saving it, or book-marking it, or e-mailing it to someone else, or copying and pasting a portion of the document, it must be having some importance in the eyes of the user. A combination of the above seven factors by simply taking their weighted sum gives the overall importance the document holds in the eyes of the user. As regards the maximum time a user is expected to spend in examining the document j, we clarify that this is taken to be directly proportional to the size of the document. We assume that an average user reads at a speed of about 10 bytes per second. It may be noted that depending on his preferences and practice, the user would set the importance of the different components of the feedback vector. For instance, if a user does not have a printer at his disposal, then there is no sense in setting up the importance weight (w_P) corresponding to the printing feedback component (P). It may, however, be noted that the component of the feedback vector corresponding to the sequence of clicking, always remains to be the prime one and so w_V must always be 1.

Now, sorting the documents on the descending values of σ_j will yield a sequence $\Re_{UF}$. Let the full list $\Re_{SE}$ be the sequence in which the documents were initially short-listed. Without loss of generality, it could be assumed that $\Re_{SE} = (1,2,3,\ldots,N)$, where N is the total number of documents listed in the result. We compare the sequences $\Re_{UF}$ and $\Re_{SE}$., and find *Modified Spearman Rank Order Correlation Coefficient* (r_s'). We repeat this procedure for a representative set of queries and take the average of r_s'. The resulting average value of r_s' is the required quantitative measure of the search quality (SQM). The above procedure is illustrated in Figure 1.

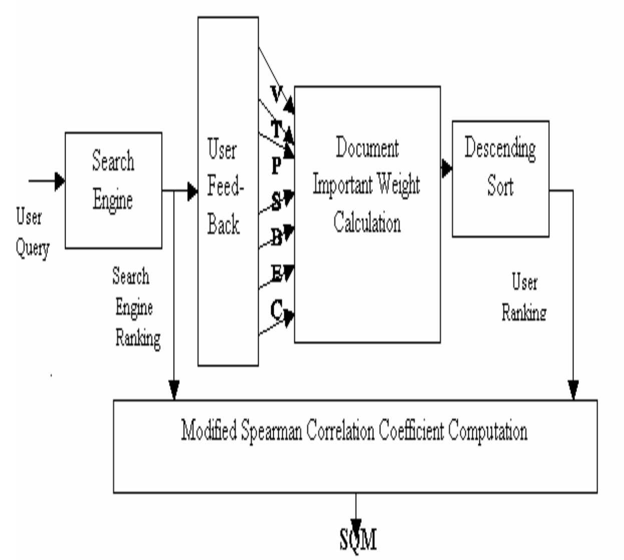

Figure 1: Subjective Search Quality Evaluation

We must note here that it is a very common practice that the user views only those documents whose snippets displayed before him by the search engine he finds to be worth viewing. *Modified Spearman Rank Order Correlation Coefficient* is a better choice than *Spearman Rank Order Correlation Coefficient* to measure the closeness of the two rankings. Since, *it* is capable of working on a full list and a partial list and the sequence $\Re_{SE}$ of documents viewed by a user is almost always a partial list. Use of *Modified Spearman Rank Order Correlation Coefficient* (r_s') saves computational efforts both in conversion of partial list to full list and also in the computation of r_s' for truncated lists.

3 Web Search Evaluation Using *PageRank*

For using *PageRank* for the evaluation of search engines, we need first to get URLs of documents, which is to be parsed out from web search results.

3.1 Parsing URLs from Web Search Results

As per parsing URLs out of the web search results is concerned , we emphasize that the process vary from search engine to search engine. As different search engines design their web site as well the result page differently, the method to parse out URLs from their result page vary. For example, the way we parse URLs out of result page of *Yahoo,* is different from that of *Google*. Hence, we have to have a look at the structure of result page of each search engine and devise the method accordingly. But, one thing common in all is that the URLs of documents in response to a query are present in the result page and we have to parse them out for our purpose.

3.2 PageRank

PageRank is a connectivity based technique to describe the importance of a web page and hence the document. *PageRank* says nothing about the content or size of a page. *PageRank* exploits the link structure of the web to rank the importance of web pages, and in turn allot some numeric values to represent their importance. The whole web is assumed to be a graph with web pages denoting the nodes of the graph and the links as edges in between. Now, if page A has more back links (other pages pointing to page A) than page B, page A is considered more important than page B. In fact, when one page points to another page, it is casting a vote for the other page. Now, more the votes cast in favor of a page, more important the page should be. Along with the number of back links a page has, *PageRank* also depends on the rankings of the pages that contain back links. So a higher ranked page will naturally make jump high the ranking of pages that it links to than a page with a lower ranking. For example, if page C and page D each have one back link from page A and page B, respectively, and page A has a higher rank than page B, then page C will rank higher than page D. Hence, "The importance of the page that is casting the vote determines how important the vote itself is"[12].

PageRanks form a probability distribution over web pages, so the sum of all web pages' *PageRanks* will be one. *PageRank* can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web. *PageRank* is one of the methods *Google* uses to determine a page's relevance or importance. *PageRank* of a page can be seen with the *Google* toolbar. But the toolbar *PageRank* only goes from 0 to 10 and seems to be something like on a logarithmic scale.

Once we have parsed out the URLs of the documents in the result page of search, we can get their *PageRanks* value using *PageRank* checker tool[13]. Now, sorting the documents in the decreasing order of their *PageRank* values, we obtain a new ranking $\Re_{PR}$ of the documents. We compare the ranking $\Re_{PR}$ with the original ranking $\Re_{SE}$ given by search engine and find *Modified Spearman Rank Order Correlation Coefficient* (r_s'). We repeat this procedure for a representative set of queries and take the average of r_s'. The resulting average value of r_s' is the required objective measure of the search quality (OQM). The above-mentioned procedure is shown in fig. 2.

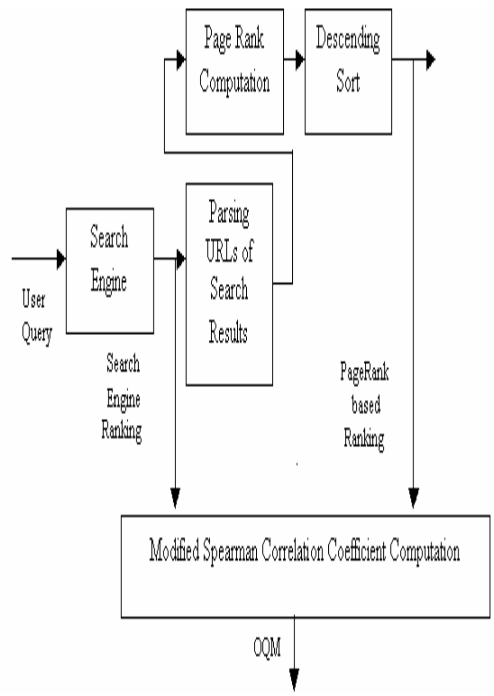

Figure 2: Objective Search Quality Evaluation

As we will be computing the *PageRanks* for URLs of the documents picked up by the user only, the sequence $\Re_{PR}$ will always be a partial list. So, *Modified Spearman Rank Order Correlation Coefficient* (r_s') would be a better choice than *Spearman Rank Order Correlation Coefficient* (r_s') as explained earlier.

We aggregate the subjective measure (SQM) with the objective measure (OQM) by averaging the values of correlation coefficients.

4 Experiments And Results

We experimented with a few queries on seven popular search engines, namely, *AltaVista, DirectHit, Excite, Google, HotBot, Lycos* and *Yahoo*. For the sake of simplicity, we have obtained all our results with the weights in equation (3) being w_V=1, w_T=1, w_P=1, w_S=1, w_B=1, w_E=1 and w_C=1. For example, the observation corresponding to the query *similarity measure for resource discovery* is given in Table 1.

Table 1: User Feedback Model Results for the Query: *similarity measure for resource discovery*

Search Engine	User Feedback V,T,P,S,B,E,C	Document Weight (σ_j)	Rank Co-efficient (r_s')
AltaVista	1,0.2,0,1,1,0,0.0	3.200	0.666667
	2,0.2,0,1,1,1,0.0	3.700	
DirectHit	1.0.2,0,1,1,0,0.0	3.200	0.645833
	5,0.2,0,1,1,1,0.0	3.700	
Excite	6,0.2,0,1,1,0,0.0	3.200	0.745833
	4,0.2,0,1,1,1,0.0	3.700	
	9,0.2,0,1,1,0,0.0	2.490	
Google	1,0.4,0,0,1,0,0.0	2.400	0.930556
	3,0.3,0,0,1,0,0.0	1.800	
	5,0.3,1,0,0,0,0.0	1.550	
HotBot	2,0.2,0,1,1,0,0.0	3.200	0.666667
Lycos	2,0.2,0,1,1,0,0.0	3.200	0.875000
	3,0.5,1,0,1,0,0.0	3.000	
	7,0.5,1,0,0,0,0.0	1.750	
Yahoo	2,0.2,0,1,1,0,0.0	3.200	0.829167
	4,0.2,0,0,1,0,0.0	1.700	
	9.0.3,0,0,0,0,0.0	0.550	

Table 1 shows that from the results of *Google*, the first document was picked up first by the user, the document was read by the user for 40% of the time required to read it completely. It was neither printed, nor saved but bookmarked. It was neither e-mailed to any friend, nor was any of its portions copied and pasted elsewhere. The user then picked the third document listed by *Google* and spent on it 30% of the time required actually to read it completely. It was neither printed nor it was saved but book marked. It was not e-mailed and none of its portion was copied and pasted too. The user then picked the fifth document listed by *Google* and spent on it 30% of the time required actually to read it completely. It was printed but not saved. Neither it was bookmarked nor it was emailed. None of its portion was copied and pasted too. This gives an importance weight (σ_j) of 2.400, 1.800 and 1.550 to the first, third and fifth documents rspectively. So the implicit ranking given by user is document 1 $\succ$ document 3 $\succ$ document 5, where "$\succ$" indicates the "more relevant than". This is compared with the original search engine ranking (1,2,3) to give the *Modified Spearman Rank Order Correlation Coefficient* (r_s'=0.930556) for *Google* in the subjective model. This way the value of r_s' is found for rest of the search engines as shown in Table 1 for the query *similarity measure for resource discovery*.

For the *PageRanks*, after parsing URLs of the documents, we compute the *PageRank* values for the documents picked by the user using *PageRank* checker tool[13].The observation corresponding to the query *similarity measure for resource discovery* is given in Table 2.

Table 2: *PageRank* Based Model Results for the Query: *similarity measure for resource discovery*

Search Engine	Picked Document	Page Rank	$\Re_{PR}$	Rank Coefficient (r_s')
AltaVista	1	0	1	1.000000
	2	0	2	
DirectHit	1	3	5	0.645833
	5	6	1	
Excite	6	2	4	0.745833
	4	4	6	
	9	0	9	
Google	1	3	3	0.763889
	3	4	5	
	5	4	1	
Hotbot	2	0	2	0.666667
Lycos	2	0	7	0.750000
	3	0	2	
	7	6	3	
Yahoo	2	0	2	0.829167
	4	0	4	
	9	0	9	

From the results of *Google*, first, third and fifth documents were picked first, second and third time respectively. The *PageRank* values for these documents are found to be 3, 4 and 4 respectively. So the ranking $\Re_{PR}$ based on *PageRank* is document 3 $\succ$ document 5 $\succ$ document 1, where "$\succ$" indicates the "more relevant than". This is compared with the original search engine ranking (1,2,3) to give the modified Spearman Rank Order Correlation Coefficient (r_s'=0.763889) for *Google* in the objective model. This way the value of r_s' is found for rest of the search engines as shown in Table 2 for the query *similarity measure for resource discovery*.

Table 3: List of Test Queries

1	Measuring Search Quality
2	Mining Access Patterns From Web Logs
3	Pattern Discovery From Web Transactions
4	Distributed Associations Rule Mining
5	Document Categorization Query Generation
6	Term Vector Database
7	Client –Directory-Server-Model
8	Similarity Measure For Resource Discovery
9	Hypertextual Web Search
10	IP Routing In Satellite Networks
11	Focussed Web Crawling
12	Concept Based Relevance Feedback For Information Retrieval
13	Parallel Sorting Neural Network
14	Spearman Rank Order Correlation Coefficient
15	Web Search Query Benchmark

Table 4:Userfeedback Model Modified Spearman Rank Correlation Coefficient (r_s') for the Queries Given in Table 3

Query	*AltaVista*	*DirectHit*	*Excite*	*Google*	*Hotbot*	*Lycos*	*Yahoo*
1	0.695000	0.889583	0.819865	0.964583	0.755556	0.766667	0.782313
2	0.700000	0.875000	0.832500	0.733333	0.773737	0.736111	0.916667
3	0.625000	1.000000	0.862434	0.917500	0.785354	0.666667	0.968750
4	0.666667	0.757576	0.706349	0.833333	0.937500	0.400000	0.628571
5	0.583333	0.222222	0.750000	0.687500	0.805556	0.666667	0.666667
6	1.000000	1.000000	0.781746	1.000000	0.876190	0.666667	0.795833
7	0.875000	1.000000	0.766667	0.983333	0.674603	0.250000	0.760417
8	0.666667	0.645833	0.745833	0.930556	0.666667	0.875000	0.829167
9	0.819048	0.820106	0.911458	0.977778	0.854167	0.914286	0.872500
10	0.550505	1.000000	0.200000	1.000000	0.181818	0.671717	0.790476
11	0.808081	0.880000	0.733333	0.875000	0.857143	0.977778	0.848485
12	0.859375	0.861111	0.947917	0.916667	0.285714	1.000000	0.930556
13	1.000000	0.515873	0.222222	1.000000	0.181818	0.181818	0.666667
14	1.000000	0.914286	0.916667	0.977778	0.687500	1.000000	0.845833
15	0.181818	0.181818	0.181818	0.500000	0.285714	0.181818	0.181818

Table 5: *PageRank* Based Model Modified Spearman Rank Correlation Coefficient (r_s') for the Queries Given in Table 3

Query	*AltaVista*	*DirectHit*	*Excite*	*Google*	*Hotbot*	*Lycos*	*Yahoo*
1	0.715000	0.977083	0.954545	0.889583	0.977778	1.000000	0.859410
2	0.791667	0.930556	0.852500	0.766667	0.822222	0.888889	0.916667
3	0.625000	0.916667	0.910053	0.952500	0.719697	0.866667	0.781250
4	0.791667	0.784512	0.706349	0.833333	0.687500	0.400000	0.771429
5	0.645833	0.222222	0.750000	0.937500	0.805556	0.666667	0.666667
6	1.000000	0.966667	0.884921	0.900000	0.838095	1.000000	0.845833
7	0.930556	1.000000	0.900000	0.833333	0.825397	0.250000	0.989583
8	1.000000	0.645833	0.745833	0.763889	0.666667	0.750000	0.829167
9	0.819048	0.788360	0.921875	0.933333	0.787500	0.982857	0.877500
10	0.550505	1.000000	0.200000	1.000000	0.181818	0.671717	0.771429
11	0.883838	0.775000	0.733333	0.875000	0.914286	0.933333b	0.905051
12	0.828125	0.819444	0.937500	0.750000	0.285714	1.000000	0.930556
13	1.000000	0.515873	0.222222	1.000000	0.181818	0.181818	0.666667
14	1.000000	0.742857	0.805556	0.933333	0.687500	0.916667	0.762500
15	0.181818	0.181818	0.181818	0.500000	0.285714	0.181818	0.181818

Table 6: Aggregated Modified Spearman Rank Correlation Coefficient (r_s') for the Queries Given In Table 3

Query	*AltaVista*	*DirectHit*	*Excite*	*Google*	*Hotbot*	*Lycos*	*Yahoo*
1	0.705000	0.933333	0.887205	0.927083	0.866667	0.883334	0.820861
2	0.745834	0.902778	0.842500	0.750000	0.797979	0.812500	0.916667
3	0.625000	0.958333	0.886244	0.935000	0.752526	0.766667	0.875000
4	0.729167	0.771044	0.706349	0.833333	0.812500	0.400000	0.700000
5	0.614583	0.222222	0.750000	0.812500	0.805556	0.666667	0.666667
6	1.000000	0.983333	0.833333	0.950000	0.857143	0.833333	0.820833
7	0.902778	1.000000	0.833333	0.908333	0.750000	0.250000	0.875000
8	0.833333	0.645833	0.745833	0.847223	0.666667	0.812500	0.829167
9	0.819048	0.804233	0.916667	0.955555	0.820834	0.948572	0.875000
10	0.550505	1.000000	0.200000	1.000000	0.181818	0.671717	0.780953
11	0.845959	0.827500	0.733333	0.875000	0.885715	0.955555	0.876768
12	0.843750	0.840277	0.942708	0.833333	0.285714	1.000000	0.930556
13	1.000000	0.515873	0.222222	1.000000	0.181818	0.181818	0.666667
14	1.000000	0.828571	0.861112	0.955555	0.687500	0.958333	0.804166
15	0.181818	0.181818	0.181818	0.500000	0.285714	0.181818	0.181818
Average	*0.759785*	*0.761010*	*0.702844*	*0.872194*	*0.642543*	*0.688188*	*0.774675*

We experimented with 15 queries in all. These queries are listed in Table 3 and their results for user feedback and *PageRank* based models are summarized in Table 4 and Table 5 respectively. The spearman correlation coefficients for the user feedback and the *PageRank* based models, that are listed in Tables 4 and 5 respectively, are averaged for each of search engines and queries and the aggregated rank correlation coefficient thus obtained is given in Table 6.

The results of Table 6 are pictorially represented in Figure 3. From Table 6 and Figure 3, we observe that *Google* gives the best performance, followed by *Yahoo*, *DirectHit*, *AltaVista*, *Excite*, *Lycos* and *Hotbot*, in that order.

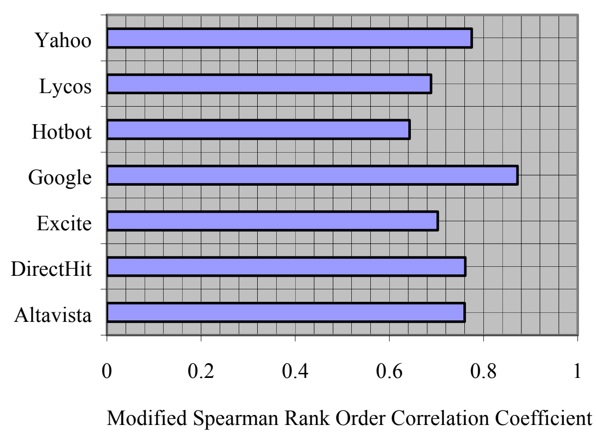

Figure 3: Performance of Search Engines based on Aggregated Model

5 Conclusion

We have tried to combine the user feedback based subjective evaluation with *PageRank* based objective evaluation for the public web search engines. For the subjective measure, we have used a method that observes the actions of the users on the search results presented before him and then infer his preferences there from. For the objective measure, we used the *PageRank* value based rankings of documents returned by the search engines. Our results for 15 queries and 7 public web search engines show that *Google* gives the best performance, followed by *Yahoo*, *DirectHit*, *AltaVista*, *Excite*, *Lycos* and *Hotbot*, in that order. We know that *Google* is primarily based on *PageRank*. But this study shows that even the combination of user feedback technique with *PageRank* renders *Google* at the top of the list. This clearly shows the superiority of *Google* over other public web search engine.

6 References

[1] M. R. Henzinger, A. Heydon, M. Mitzenmacher and M. Najork, "Measuring Index Quality Using Random Walks on the Web," *Computer Networks*, 31, 1999, pp. 1291-1303.

[2] M. R. Henzinger, A. Heydon, M. Mitzenmacher and M. Najork, "On Near Uniform URL Sampling," Proc. 9th International World Wide Web Conference (WWW9), pp. 295-308, May 2000.

[3] Z. Bar-Yossef, A. Berg, S. Chien, J. Fakcharoenphol and D. Weitz, "Approximating Aggregate Queries about Web Pages via Random Walks," Proc. 26th Very Large Data Bases (VLDB) Conference, Cairo, Egypt, pp. 535-544, September 2000.

[4] K. Bharat and A. Broder, "A Technique for Measuring the Relative Size and Overlap of Public Web Search Engines," Proc. 7th International World Wide Web Conference (WWW7), pp. 379-388, April 1998.

[5] S. Lawrence and C. L. Giles, "Searching the World Wide Web," *Science*, 5360(280), pp. 98-100, 1998.

[6] S. Lawrence and C. L. Giles., "Accessibility of Information on the Web," *Nature*, 400, pp. 107-109, 1999.

[7] D. Hawking, N. Craswell, P. Thistlewaite and D. Harman, "Results and Challenges in Web Search Evaluation," Proc. 8th International World Wide Web Conference (WWW8), Toronto, Canada, pp. 1321-1330, May 1999.

[8] L. Li and Y. Shang, "A New Method for Automatic Performance Comparison of Search Engines," *World Wide Web: Internet and Web Information Systems*, 3, pp. 241-247, 2000.

[9] L. Li and Y. Shang, "Precision Evaluation of Search Engines," *World Wide Web: Internet and Web Information Systems*, 5, pp. 159-173, 2002.

[10] M. M. S Beg and R. Ali, " A Subjective cum Objective Measure of Web Search Quality", Proc. 2005 International Conference on Intelligent Sensing and Information Processing (ICISIP-05), Chennai, India, pp 365-370, January 2005.

[11] R. Ali and M. M. S Beg, " Aggregating Subjective measure of Web Search quality with Boolean Similarity measures", Proc. 2nd World Congress on Lateral computing (WCLC-05), Bangalore, India, Dec. 2005.

[12] E. W. Weisstein "Spearman Rank Correlation Coefficient" From MathWorld - A Wolfram Web Resource, © 1999-2004 Wolfram Research, Inc. http://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html.

[13] M. M. S. Beg, *On Measurement and Enhancement of Web Search Quality,* Ph.D. thesis submitted to the Department of Electrical Engineering, I. I. T. Delhi, 2002.

[14] P.Craven, "Google's PageRank explained and how to make the most of it", http://www.webworkshop.net/pagerank.html.

[15] http://www.searchenginegenie.com/search-engine-page-rank-checking.php.