

# Ranking Web Documents with Dynamic Evaluation by Expert Groups

Sea Woo Kim<sup>1</sup> and Chin-Wan Chung<sup>2</sup>

<sup>1</sup> Division of Information and Communication Engineering,  
373-1, Kusong-Dong, Yusong-Gu, Taejon 305-701, Korea  
[seawoo@islab.kaist.ac.kr](mailto:seawoo@islab.kaist.ac.kr)

<sup>2</sup> Division of Computer Science,  
373-1, Kusong-Dong, Yusong-Gu, Taejon 305-701, Korea  
[chungcw@islab.kaist.ac.kr](mailto:chungcw@islab.kaist.ac.kr)

**Abstract.** In spite of the wide use of the Internet, it is difficult to develop desirable web documents evaluation that reflects users' needs. Many automatic ranking systems have used this citation system to measure the relative importance of consumer products or documents. However, the automatic citation analysis has a limitation in that it does not truly reflect the importance of the varying viewpoints of human evaluation. Therefore, human evaluations of web documents are very helpful in finding relevant information in a specific domain. Currently, human evaluation is done by a single expert or general users without considering the degree of domain knowledge of evaluators. In this paper, we suggest that a dynamic group of experts for a certain web document be automatically created among users to evaluate domain specific web documents. The experts have dynamic authority weights depending on their performance of the ranking evaluation. In addition, we develop an evaluation effectiveness measure for ranking processes. This evaluation by a group of experts provides more accurate search results and can be a good measure of user preferences when the size of users' feedback is small. Also, dynamic change of authority weight provides the evaluation effectiveness of experts. Furthermore, dynamic change of authority weight provides the evaluation effectiveness of experts.

## 1 Introduction

Together with the development of the Internet and the popularity of WWW, Web document recommendation systems have drawn significant attention. Many Web document comparison search engines have been introduced and developed, but still have difficulty in providing completely relevant answers to the general subject of queries. The main reason is not due to the lack of data but rather an excess of data. This data has a variety of characteristics and information that is not specific to users. This is what led to the necessity of web document rankings. For the time being, the web documents are listed by using a ranking measure called connectivity analysis. This improves the quality of search results. An example of this method is the IBM HITS system. This method works on the assumption that a document that is cited

many times is popular and important. The hub scores of its parent pages are summed into its authority score and the authority scores of its children pages are summed into its hub score [4][5]. By iteration, it determines highly referenced pages.

Another ranking method is called Page Rank. This method works in two phases. First, it computes a ranking for every document based on the web connectivity graph [2] with a random walk traversal. Second, it considers the relative importance of the document by checking ranks of back link pages. When the document has back links to pages with high authority such as [www.yahoo.com](http://www.yahoo.com), it is ranked as highly important.

However, even these methods have a limitation. They do not truly reflect the importance of the varying viewpoints of human evaluation. There are many cases where simple citation counting does not reflect our common sense concept of importance [2]. Also documents, products and academic publication papers are significantly different in terms of citation analysis. Some search engines employ the method based on textual similarity [1][8]. Normally they count the frequency of terms in a domain to decide lexical affinities instead of using advanced natural language processing techniques. And yet even textual similarity analysis has its limitations.

While the above approaches consider topological links of the web, a combination of a broad search of the entire web with domain-specific textual and topological scoring of results is suggested [1]. Aridor has specialized knowledge agents for specific domains to extract the most relevant documents. Similarly we are interested in ranking Web documents in a specific domain. In our approach, the search engine goes through the process of a domain-specific web search and shows the list of popular documents for each specialized subject.

In this paper we use a method to evaluate web documents by a group of human agents[21], and we call it an expert group. We believe that the importance and level of authority of consumer products or documents should be determined by interactions between human and consumer products or documents. Furthermore, domain groups should be responsible for document ranking for each category. This domain group should have authorities to evaluate documents. This approach will overcome the disadvantages of the automatic ranking method through incomplete information processing based on citation authority or lexical affinities, which ignore the content of web documents. Another benefit is that when the size of evaluation feedback of general users is not big enough, this method can give the initial measurement of users' preferences.

## 2 Methodologies

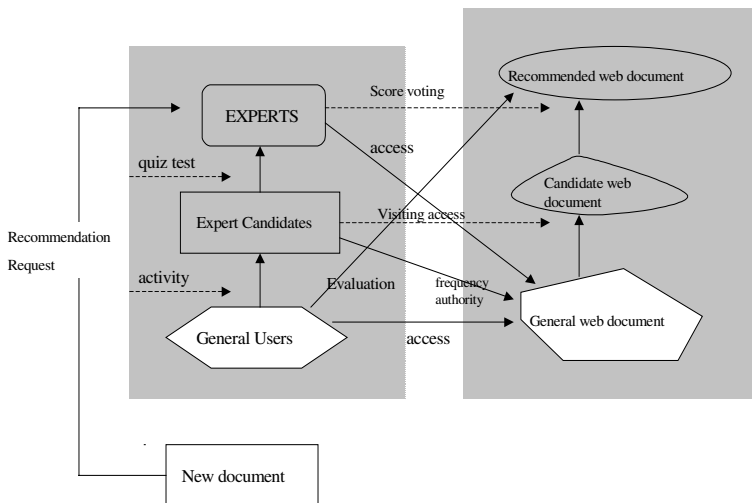
### 2.1 Model of Recommending Documents

We define a group of people with high authority as an expert group. This expert group is automatically promoted from the general users to evaluate web documents on a specific category. There are three groups in three levels for each category, the general user group, the expert candidate group, and the expert group as shown in Figure 1. The expert candidates are chosen among active users. The access count of web docu-

ments are recorded, and user participation is measured and classified. Access count to important products and the participation level of a user determine the activity of a user. This activity is the major factor to decide experts and their potential for future activity. An expert candidate will be given a test to measure his/her knowledge about a certain domain. Then the results of the test are calculated to decide if he/she can be an expert. If a candidate is qualified for the expert-pool, then he/she will be allowed to score and rank web documents that are to be evaluated.

Similar to the three user groups, we have three groups of web documents in three levels for each category, the general document group, the candidate document group, and the recommended document group as shown in figure 1. A pool of candidates for the expert groups are nominated and to be upgraded for the expert group for each category. A category is created considering subjects and it can be subsequently refined into many categories when the access to it exceeds a threshold. Each web document contains the information of how many search engines being utilized in our search engine are referring to the document, and keeps a record of how many times online users have accessed the web document using our search engine.

If a web document is referred to many times in many search engines under a given category, then it can be highly ranked. This kind of web document is considered as a candidate document that needs to be evaluated by an expert group. If a web document is accessed many times by users recently, then it can be highly ranked. This web documents one of the candidate documents to be evaluated by an expert group. This approach is likely to produce more fine-grained and reliable results to a new environment of web patterns. For every category there is a list of recommended documents evaluated by an expert group, which are sorted by score. Human experts decide which of the candidate document are to be promoted for the recommended documents.

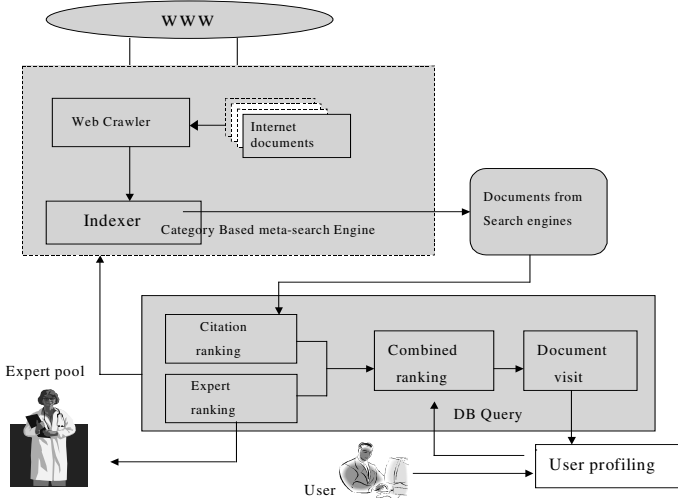


**Fig. 1.** Diagram of user group and documents group

## 2.2 Dynamic Evaluation by Expert Group

A meta-search engine that is run to collect addresses of cited web document from conventional search engines as shown in Figure 2. List of web documents are collected from the each search engine directory. Web crawler automatically performs this process. At the time of the query, the ranking by the number of citations from shopping malls and expert ranking are combined. The combined rank documents are shown to the users. The visiting log of users are recorded and monitored. The method of employing an expert group is based on the idea that for a given decision task requiring expert knowledge, many experts may be better than one if their individual judgments are properly combined.

In our system, experts decide whether a document should be classified into a top ranked list for a given category. A simple way is the majority voting, where each expert has a binary vote for a document and then the document obtaining equal to or greater than half of the votes are classified into a top ranked list [7]. The result of the decision for documents is stored along with their addresses.



**Fig. 2.** System Architecture

Another possible method we can consider is a weighted linear combination. A weighted linear sum of expert voting results yields the final decision documents. In this paper, we take the adaptive weighted linear combination method, where the individual contributions of members in the expert group are weighted by their judgment performance. The evaluations of all the experts are summed with weighted linear combinations. The expert-voted results will dynamically change depending on each expert's performance and effectiveness. Our approach of expert group decision is similar to a classifier committee concept in automatic text categorization [9]. Their methods use classifiers based on various statistical or learning techniques instead of

human interaction and decision. The explanation of choosing group members and determining authority i.e. effectiveness for each expert is as follows.:

We have a visiting access matrix  $C=[C_{ij}]$  between users and Web documents, which is given by

$$C_{ij} = \begin{cases} 1 & \text{If a user } u_i \text{ visits a search engine } d_j \\ 0 & \text{otherwise} \end{cases}$$

$C_{ij}$  is binary to prevent spam effects. Since it is binary one user's frequent vote does not influence the visiting access matrix. From this information, we define an activity measure  $I_a(u_i)$  for a user  $u_i$  and a visiting access measure  $I_v(\beta dj)$  for a web document  $dj$  as follows :

$$I_a(u_i) = \sum_{j=1}^{N_d} C_{ij} \quad , \quad I_v(dj) = \sum_{i=1}^{N_u} C_{ij}$$

where  $N_d$  is the number of documents and  $N_u$  is the number of users registered in the meta-search engine.

Our meta-search engine extracts web documents from  $N_s$ -existing search engines, which are denoted as  $s_i$ ,  $i=1, \dots, N_s$ . Then we define a frequency authority measure  $I_f(d_j)$  based on the search engines over a web document  $d_j$  as follows :

$$I_f(d_i) = \sum_{j=1}^{N_s} \delta_j m_{ij}$$

where  $\delta_i$  is a weight for each search engine, initially set to 1.

The frequency authority measure represents the frequency of a document in a meta-search engine.

If this measure  $I_f(dj)$  is larger than a threshold, then we can assume that the document  $p_j$  is of good quality and importance. Using the visiting access matrix  $C=[c_{ij}]$  and the consumer product matrix  $M=[m_{ij}]$ , we can calculate the popularity and performance of each search engine used in the meta-search engine. Search engine frequency matrix  $Y=[y_{ij}]$  is defined as  $Y = C \cdot M^T$  then weight  $\delta_k$  for each search engine can be updated as follows:

$$\delta_k = \frac{\sum_{i=1}^{N_u} y_{ik}}{\sum_{j=1}^{N_s} \sum_{i=1}^{N_u} y_{ij}}$$

The candidate users that are to be promoted to experts for a category can be determined by checking the measure  $I_a$  calculated during a given period. Every week or so, this activity measure is updated. Candidates are required to pass a test to become an expert. The web documents of good quality or importance are determined by checking the frequency authority measure  $I_f$  and visiting access measure  $I_v$  over web documents. Thus, we can decide candidate document among general documents using the following measure:

$$I_c = \alpha I_f + \beta I_v + \gamma$$

where  $\alpha, \beta, \gamma$  are scaling factors. If  $I_c$  is larger than a threshold  $\tau$ , the corresponding document will be accepted as a candidate document. The selected documents wait for an evaluation by experts for a given category. For each candidate document, experts are required to evaluate the document and give them scores. An evaluation score matrix is defined as  $X = [\chi_{ij}]$  when the  $i$ -th expert evaluates a consumer product  $p_j$  with

a voting score  $\chi_{ij}$ . We have a weighted importance or authority over experts for each category. The evaluation score matrix shows a relation between experts and candidate documents. The weight is determined by their activity measure  $Ia$ , test score, and other factors. Experts are assigned the same weight at the initial stage as experts are not differentiated at this moment.

The weight is dynamically changed by their activity and feedback from online users about recommended documents through voting results.

This weighted measure is useful even when the number of expert members is not fixed. Thus, for each document  $d_j$  listed as a candidate document the weighted authority voting score is defined as follows:

$$V(d_j) = \sum_{k=1}^{N_e} \frac{w_k}{\sum_{i=1}^{N_e} w_i} \chi_{kj} \quad w_k = E I_a(u_k) + H$$

where  $N_e$  is the number of experts for a given category, and  $r_k$  is the relative authority for the  $k$ -th expert in the expert pool, and  $w_k$  is the weight calculated using activity measure, test scores and other career factors for the  $k$ -th expert member. And  $E, H$  are scaling factors.

The weight  $w_k$  is a dynamic factor, and it discriminates bad experts from good experts in terms of their activity and users' voting results. When some experts show little participation in voting or evaluate incorrectly, their authority weight  $w_k$  becomes smaller. For example, when an expert voting score is larger than the weighted average voting score and the average score is smaller than the desirable rank score, the expert status is rewarded, otherwise penalized. Therefore, some experts receive rewards and others receive penalties depending on the weighted average voting score of experts. All experts receive penalties or rewards in our application. It is possible that some experts with too many penalties are excluded from the expert group and new experts are added to the group.

We define an error measure  $E$  as a squared sum of differences between desired voting scores and actual voting scores as follows:

$$E = \frac{1}{2} \sum_{j=1}^n (V(d_j) - V'(d_j))^2 = \frac{1}{2} \sum_{j=1}^n \left( \sum_{k=1}^{N_e} \frac{w_k}{\sum_{i=1}^{N_e} w_i} \chi_{kj} - V'(d_j) \right)^2$$

where  $n$  is the number of documents evaluated by users,  $V'(d_j)$  is the desired voting score for an expert-voting document  $d_j$ . We assume  $V'(d_j)$  is the average score evaluated by all users, but in reality it is rarely possible to receive the feedback from all users.

We chose the coefficient  $1/2$  to make its gradient formula simpler, which will be shown later. We assumed this value could be determined by the feedback from general on-line users. Voting scores of the experts should reflect common ideas of users about ranking and satisfy desires of many users to find proper relative appropriate information, because an expert is a representative of general users and has a extensive expert knowledge in a specific domain.

Now we use a gradient-descent method over error measure  $E$  with respect to a weight  $w_k$  and the gradient is given by

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \left( \frac{1}{2} \sum_{j=1}^n [V(d_j) - V'(d_j)]^2 \right)$$

$$\begin{aligned}
&= \frac{1}{2} \bullet 2 \sum_{j=1}^n [V(d_j) - V(\gamma(d_j))] \frac{\partial V(d_j)}{\partial \omega_i} \\
&= \sum_{j=1}^n \Delta_j \frac{\partial V(d_j)}{\partial \omega_i} = \sum_{j=1}^n [\chi_{ij} - V(d_j)] \frac{\Delta_j}{S}
\end{aligned}$$

where  $s = \sum_{k=1}^{N_e} w_k$  is the sum of weights, and  $\Delta_j = [v(d_j) - V(\gamma(d_j))]$  is the difference between the predicted voting score and the users evaluation score for a document  $d_j$ .

If we update weights of experts for ranking by feedback from users about a document  $d_j$ , the weight is changed by the following dynamic equation

$$\omega_i(t+1) = \omega_i(t) - \eta [\chi_{ij} - V(d_j)] \frac{\Delta_j}{S}$$

### 2.3 Evaluation Effectiveness

User effectiveness is measured by calculating how close the evaluation of the general users and experts. In this paper we use four measures for calculating user effectiveness.

**1. Rank Order Window Measure.** Given a sample query or category, we can represent the effectiveness as the percentage of top-ranked list of user ratings which rank in the same or very close position as an expert group does. A window is defined as a cluster of a certain group. Given top-ranked products  $P = \{p_1, p_2, \dots, p_n\}$ , we can define effectiveness  $\Lambda_\delta$  of rank-order window  $\delta(p_k)$  where  $p_k$  is the  $k$ -th web document from the test set for a given category, and  $\delta(d_k)$  is the width of the window centered in the rank  $\delta(p_k)$  assigned by the ratings of experts for  $d_k$ .

$$S(p_k) = 1 - \frac{1}{\delta(d_k)} \min(\delta(d_k), \left| \sum_{i=\mu(p_k)-\delta(d_k)}^{\mu(p_k)+\delta(d_k)} \frac{\mu(d_i) - Q(d_i)}{2\delta(d_k) + 1} \right|)$$

$$\Lambda_\delta = \frac{\sum_{k=1}^n S(d_k)}{n}$$

where  $Q(d_i)$  is the rank position of the average rating score of users for a document  $d_i$ .  $S(d_k)$  calculates the rank order difference in the window  $[\mu(d_k) - \delta(d_k), \mu(d_k) + \delta(d_k)]$ . In the window the evaluation difference between the expert and general user can be bigger than the size of window. In that case  $\delta(d_k)$  is the minimum value. In the equation, we added 1 to  $2\delta(d_k)$  in denominator. We added 1 because we need non-zero denominator at any time.

**2. F<sub>γ</sub> measure with Rank Order Partition.** Evaluation effectiveness also can be described in terms of precision and recall widely used in information retrieval. Precision is the conditional probability that when a document is predicted to be in a positive class, it truly belongs to this class. Recall is the conditional probability that a document belonging to positive class is truly classified into this class [10,11]. We partition recommended documents by their rank order and make classes. We define a

positive class  $i$  as top  $[(i-1)*10+1 \sim i*10]$  ranked documents by expert voting and a negative class as the others. For Example, class 2 documents are top  $[11 \sim 20]$  ranked documents.

The precision probability  $P_i$  and recall probability  $R_i$  for ranking site class  $i$  may be estimated using the contingency relations between expert rating and user rating, and those probabilities in our application can be calculated with transition instances between classes. A transition instance  $p_{ij}$  is define as the number of instances that are predicted to be in class  $i$  by user ratings.

$$P_i = \frac{p_{ij}}{\sum_{j=1}^m p_{ij} \cdot |i - j + 1|}, \quad R_i = \frac{p_{ij}}{\sum_{j=1}^m p_{ij} \cdot |i - j + 1|}$$

where  $m$  is the number of classes, and  $\bar{P}, \bar{R}$  are the average precision and recall probabilities, respectively. The distance between classes is considered to calculate  $P_i, R_i$ . Then effectiveness can be computed using the value of  $F_\beta$  for  $0 \leq \beta \leq \infty$  [3,14,20].

$$F_\beta = \frac{(\beta^2 + 1) \cdot \bar{P} \cdot \bar{R}}{\beta^2 \cdot \bar{P} + \bar{R}}$$

To balance precision and recall, a value  $\beta=1$  is used in our experiments. If  $F_\beta$  is close to zero, then the current documents ranked in a class through expert voting result can be seen to have many false responses from feedback of general online users or many new documents positioned to the top ranks. If  $F_\beta$  is close to one, then top-ranked sites have good feedback from general users and little change occurs on the top ranked lists.

## 2.4 Expert Group Evaluation Algorithms

The general users do evaluations of consumer product ranking and it can be calculated by two methods. First, the user can use the same scale as a user does to evaluate the web documents. In this paper, we define the method which use the same scale of evaluation for expert group and general users as “absolute method”, and define the method which use the same scale of evaluation for expert group and general users as “relative method”. Table 1 shows the absolute method.

The evaluation of web documents and its ranking process for a search engine is shown in Table 2. It shows the algorithm for the growth of top ranked documents and the dynamic change of authority weights of experts. Table 2 shows the relative method.

## 3 Experiments

Actual experiment is performed using two different methods: user’s absolute feedback and relative feedback. Then four effectiveness measures are used such as rank order window, rank function, spearman’s correlations and  $F_\beta$  measure. There are 200 web documents. We used movie category with 100 users and 5 experts.



**Table 1.** Algorithm for an absolute feedback ranking

<p><b>Repeat</b></p> <ul style="list-style-type: none"> <li>- choose candidate members among users with the measure <math>I_a</math></li> <li>- choose expert members among candidates using <math>I_a</math> and quiz test</li> <li>- decide authority weight for each expert,  <math>w_k = \alpha I_a + \gamma</math></li> <li>- choose candidate documents among general documents with the measure <math>\alpha I_v + \beta I_i + \gamma &gt; \tau</math></li> <li>- list the ranked documents.</li> <li>- when a user evaluates each ranked document the authority weight for each expert is updated by  <math>w_k(t+1) = w_k(t) + \eta H_e d(V(d_i), V'(d_i)) \chi_{ki}</math></li> <li>- create classes for ranking documents.</li> <li>- calculate precision <math>P_i</math> and recall <math>R_i</math> for every class <math>I</math>.</li> <li>- calculate the evaluation effectiveness with the measure <math>f_\beta</math>.</li> <li>- If <math>f_\beta</math> is close to 1, then it reaches desirable status</li> </ul> <p><b>Until</b> performance is satisfiable</p>
---

**Table 2.** Algorithm for a relative feedback ranking

<p><b>Repeat</b></p> <ul style="list-style-type: none"> <li>- The same as absolute method</li> <li>- when a user evaluates each ranked document the authority weight for each expert is updated using the equation <math display="block">H_e(a,b) = \begin{cases} 1 &amp; \text{if } a-b \geq e \\ -1 &amp; \text{if } a-b \leq -e \\ 0 &amp; \text{otherwise} \end{cases}</math> </li> <li>- The same as absolute method</li> </ul>
--

Figure 3 shows the experiment using the absolute feedback algorithm. The result is similar to that of the simulation. Figure 3 also shows the experiment using user effectiveness evaluation methods using the absolute feedback method. The general users just express their preference by scale of 10 to 30 regarding a certain web document's quality. This feedback influences the weight of each expert for next session. As time passes, all the measurements converge to a certain value. Therefore after we found no significant change the evaluation of expert is stabilized and can be trusted by the general users.

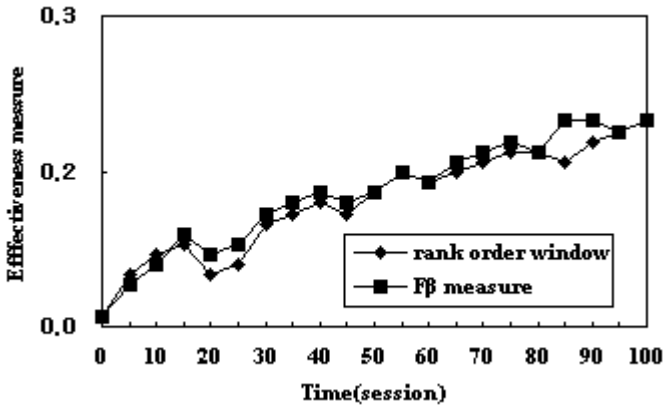


Fig. 3. User's Absolute Feedback

After the feedback of general users, an expert with low score is excluded from the expert group. In many cases, the users are not in the same scale as the expert group. Therefore, it is not easy to compare the two groups scores to figure out the correlation. This leads to the introduction of relative feedback by the general user.

Figure 4 shows the user's relative feedback method. The general users just express their preference by three categories; good, neutral, or bad regarding a certain product's quality. This feedback influences the weight of each expert for next session. After the feedback of general consumers, an expert with a low score is excluded from the expert group. For both absolute feedback and relative feedback methods the weight for each expert converges as session increases. Also we found that the product evaluation results of experts and general users are getting close to each other. This increases the user effectiveness as this paper defined in section 2.

## 4 Conclusion and Future Works

In this paper, we have shown a dynamic evaluation for consumer product ranking through human interaction. Expert group is automatically formed among users. Each expert has his own authority to evaluate web documents. This authority is dynamically changed using feedback of users. Expert group evaluation is more attractive compared to other search engines because it can receive more human feedback.

Our search engine is very domain-specific to improve the performance of search results and show the top ranked products. We have clustered top rank documents by their rank order, and thus we applied expert evaluation effectiveness to those clusters to test if our ranking list is reliable. Thus, we achieve the early web document recommendation system when responses of general users are small. Evaluation effectiveness measures show that our system satisfies the general users.

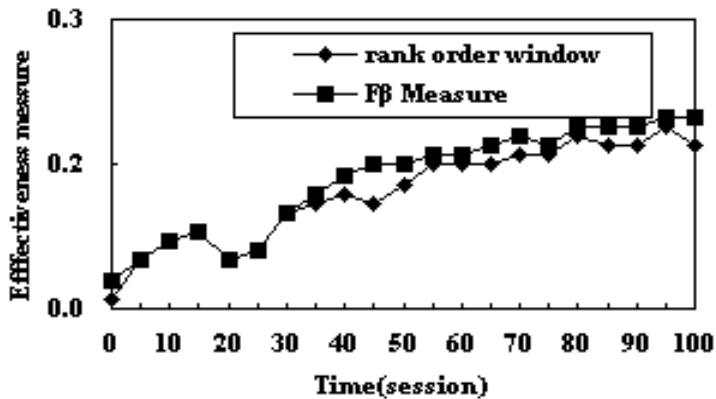


Fig. 4. User's Relative feedback

Future works include applying this idea to design a meta-search engine for specific domain: movies, music, and shopping mall documents. In addition, all documents should be classified in advance before the proposed system is applied. Currently, our categorization of all documents into specific domains is user-defined but not automatic. This is one of the text categorization problems [20]. We plan to develop a text categorization process to filter out documents irrelevant to a domain for a given query.

Also, comparison with PageRank or other methods should be included. Using the user effectiveness measures as we suggested, a search engine of web documents or web products will be prototyped [22][23].

Furthermore, users will evaluate using the mobile Short Message Service(SMS) that will improve higher response rate. Currently, web documents are evaluated on the web and we found the users feels more comfortable to evaluate them through the SMS service. Web service ranking is another area that can be expanded from our ranking model.

**Acknowledgement.** This work was supported in part by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program.

## References

- [1] Aridor, Y., Carmel, D., Lempel, R., Soffer, A. & Maarek, Y.(2000), Knowledge Agents on the Web, in Fourth International Workshop on cooperative Information Agents.
- [2] Brin, S. & Page, L.(1998), The anatomy of a large-scale hypertextual web search engine, in proceedings of the Seventh International World Wide Web Conference.
- [3] Cohen, W.& Singer, Y(1999), "Context-sensitive learning methods for text categorization", ACM transactions on Information Systems 17(2), 141-173.

- [4] Gibson, D., Kleinberg, J. & Raghavan, R.(1998), Inferring Web communities from link topology, in Proceedings of the ACM Conference on Hypertext and Hypermedia.
- [5] Kleinberg, J.(1998), Authoritative sources un a hyperlinked environment, In Proceedings of the ACM-SIAM Symposium on Discrete Algorithms.
- [6] Larkey, L. & Croft, W.(1996). Combining classifiers in text categorization, in Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.289–297
- [7] Li, Y.& Jain, A. (1998), "Classification of text documents", The Computer Journal 41(8), 537–546
- [8] Liere, R & Tadepalli, P.(1997), Active learning with committees for text categorization, in Proceedings of the AAAI-97, 14<sup>th</sup> Conference of the American Association for Artificial Intelligence, PP.591–596.
- [9] Maarek, Y&Smdja, F.(1989), Full text indexing based on lexical relations, an application: Software libraries, in Proceedings of SIGIR 89, ACM Press, pp.198–206
- [10] Raghavan, V., Bollmann, P. and Jung, G.(1989)."A critical investigation of recall and precision as measures of retrieval system performance", ACM transactions on Information Systems, 7(3):205–229
- [11] Sebastiani, F.(1999), Machine learning in automated text categorization: a survey, Technical Report, IEI-B4-31-1999, Istituto di Elaborazione dell'Informazione, C.N.R, Pisa, IT.
- [12] Van Rijsbergen, c.(1979), Information Retrieval, Butterworths, London.
- [13] Yang, Y.(1994), Expert network: Effective and efficient learning from human decisions in text categorization and retrieval, in Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 13–22.,1994
- [14] Yang, Y.(2000),"An evaluation of statistical approaches to text categorization", Information Retrieval 1(1-2),69–90., 2000
- [15] W. Hill, L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. In Proceedings of ACM CHI'95, pages 194–201,1995.
- [16] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: an open architecture for collaborative filtering of netnews. In Proceedings of Computer Supported Cooperative Work Conference, pages 175–186, 1994.
- [17] S.M. Ross. Introduction to Probability Models. Academic Press, London, 7<sup>th</sup> edition, 2000. U.Shardanand and P.Maes. Social information filtering: algorithms for automating word of mouth'. In proceedings of ACM CHI95,1995.
- [18] I. Soboroff, C. Nicholas, and editors M. Pazzani Web Aocument Selection. In Proceedings of the SIGIR-99 Workshop on Recommender Systems, Berkeley, California, 1999.
- [19] H.M. Taylor and S. Karlin. An Introduction to Stochastic Modeling. Academic Press, London, 3<sup>rd</sup> edition, 1998.
- [20] C.J van Rijsbergen. Information Retrieval. Butterworths, London, 1979.
- [21] S.W. Kim and C.W.Chung, "Web Document Ranking by Differentiated Expert Group Evaluation," HCI International 2001, pp.415–419, Aug. 2001
- [22] Anil L. Norri, E-business Architectures and Standards, Tutorial, VLDB'02, HongKong, China ,2002.
- [23] B. Benatallah and F. Casati, editors. Special Issue on Web Services. Distributed and Parallel Databases, An International journal 12(2-3), Sep. 2002.