



## OTHER ARTICLE

# An opinion-based decision model for recommender systems

Sea Woo Kim

*Division of Information and Communication Engineering,  
Korea Advanced Institute of Science and Technology, Seoul, South Korea*

Chin-Wan Chung

*Department of Computer Science,  
Korea Advanced Institute of Science and Technology, Seoul, South Korea, and*

DaeEun Kim

*School of Electrical and Electronic Engineering, Yonsei University,  
Seoul, South Korea*

### Abstract

**Purpose** – A good recommender system helps users find items of interest on the web and can provide recommendations based on user preferences. In contrast to automatic technology-generated recommender systems, this paper aims to use dynamic expert groups that are automatically formed to recommend domain-specific documents for general users. In addition, it aims to test several effectiveness measures of rank order to determine if the top-ranked lists recommended by the experts were reliable.

**Design/methodology/approach** – In the approach, expert groups evaluate web documents to provide a recommender system for general users. The authority and make-up of the expert group are adjusted through user feedback. The system also uses various measures to gauge the difference between the opinions of experts and those of general users to improve the evaluation effectiveness.

**Findings** – The proposed system is efficient when there is major support from experts and general users. The recommender system is especially effective where there is a limited amount of evaluation data from general users.

**Originality/value** – This is an original study of how to effectively recommend web documents to users based on the opinions of human experts. Simulation results were provided to show the effectiveness of the dynamic expert group for recommender systems.

**Keywords** Information retrieval, Skills, Worldwide web

**Paper type** Research paper



### Introduction

The development of recommender systems as a means of information retrieval has emerged as an important issue of the internet, and has drawn attention both from academics and the commercial sector.

This research was supported by the Ministry of Knowledge Economy, Korea under the Information Technology Research Center support programme supervised by the Institute of Information Technology Advancement (grant number IITA-2008-C1090-0801-0031).

---

An example of the use of such systems is to recommend new products or items of interest to online customers, using customer preferences. When customers have no personal experience of an item or class of items, they are often interested in retrieving information about the items or products that many others have ordered or used. However, many recommender systems have focused on retrieving information considering the preferences of just one or a few customers, and sometimes there may be no information about user preferences for a recommender system to draw on. Recommender systems can also be applied to retrieve relevant web documents. Web documents need many users' evaluations in order to be recommendable. A simple citation search for a relevant document may not reflect well the crucial point of a document.

We need a different approach to recommender systems. In our research we have explored a method in which human agents collect useful information and provide it to general users.

A group of human users or experts can cooperate to determine whether a web document includes useful information and rank the web documents in order. Such information can be made available to general users as recommendations. The feedback of the users can reorganise the group and refine the knowledge level that a group of human experts provides. This kind of adaptive organisation and feedback loop will give users access to expert knowledge in a particular domain, and will have a filtering effect on biased opinions from just a few people.

Information retrieval systems often give numeric scores to documents and then rank them based on the scores in order to make recommendations to users. There have been several approaches to the information retrieval system, the most common of which are the vector space model, the probabilistic model and the inference network model. In the vector space model, a document is represented by a vector of terms, that is, words and phrases (Salton *et al.*, 1975). The model calculates the similarity between a query and a document. The angle between the query vector and the document vector can be measured using the cosine property (the dot product of two vectors is involved with the cosine angle). More similar vectors will have a numeric cosine value close to 1. The probabilistic model estimates the probability of the relevance of a document to a query (Robertson, 1977) and documents can be ranked based on the relevance probability. In the inference network model, an inference process is applied to model the information retrieval (Turtle and Croft, 1991), where a document instantiates a term with a certain strength, and it accumulates the credit from multiple terms to assign a numeric score to the document. Then the strength of instantiation is taken as the weight of the term in the document.

The above scoring methods can assist in the automatic evaluation of documents. But while this kind of numeric assignment can give a rough evaluation or coarse information retrieval, in many cases it cannot provide accurate information about given documents.

Many common web searches retrieve a very small number of relevant documents. Topic distillation is a special kind of topical relevance search where the user wishes to find a few key websites rather than every relevant webpage. Because these types of searches are so common, web search evaluations have come to focus on tasks where there are a very few relevant documents. Evaluations with just a few relevant documents pose special challenges for current metrics (Soboroff, 2006). The development of intelligent information retrieval techniques has large impact potential in many domains (Fern *et al.*, 2007).

---

Recommender systems, as one information retrieval technique, can be broadly categorised into content-based and collaborative filtering systems (Hill *et al.*, 1995; Resnick *et al.*, 1994; Shardanand and Maes, 1995; Soboroff *et al.*, 1999). Content-based filtering methods use textual descriptions of documents or items to be recommended. A user's profile is associated with the content of the documents that the user has already rated. The features of documents are extracted by information retrieval, pattern recognition or machine learning techniques. Then the content-based system recommends documents that match the user's profile (Delgado *et al.*, 1998; Soboroff *et al.*, 1999).

In contrast, collaborative filtering systems are based on user ratings rather than the features of the documents (Breese *et al.*, 1998; Soboroff *et al.*, 1999; Shardanand and Maes, 1995). These systems predict the ratings of a user for given documents or items, depending on the ratings of other users with similar preferences to the user. Collaborative filtering systems, such as GroupLens (Resnick *et al.*, 1994; Konstan *et al.*, 1997), can be part of recommender systems for online shopping sites. They recommend items to users, using the history of products that similar users have ordered or have viewed.

Most recommender systems use analysis of the user's preferences. Such systems require the user to judge many items in order to obtain the user's preferences. In general, many online customers or users are interested in other users' opinions or ratings about items that belong to a certain category. For instance, many e-commerce customers like to see the top-ranked lists of rating scores of many users for retail items in order to help them make a purchase decision. However, recommender systems still have difficulty providing relevant rating information before they receive a large number of user evaluations or feedbacks.

In this paper, we provide a new method for evaluating web documents using a representative board of human agents (an "expert group"). This is different from automatic recommender systems with software agents or feature extractions. We suggest that dynamic expert groups should be created from among users to evaluate domain-specific documents for webpage ranking, and that the group members should have dynamic authority weights depending on the performance of their ranking evaluations. This method will be quite effective in recommending web documents or items that many users have not already evaluated – in such cases it is difficult for automatic recommender services to provide effective recommendations. Because in our approach users with expertise in a domain category evaluate the documents, it is not feasible to replace human agents with intelligent software agents.

Our recommender system with dynamic expert groups may be extended to challenge search engine designs and image retrieval problems. Many search engines find relevant information and its importance by applying automatic citation analysis to the general subject of a query. The hypertext connectivity of web documents has been a good measure for automatic web citation analysis. This method works on the assumption that a webpage that is cited many times is popular and important. Many automatic page-ranking systems have used this citation metric to decide the relative importance of web documents. The IBM HITS system maintains a hub and an authority score for every document (Kleinberg, 1998). A method called PageRank computes a ranking for every web document based on a web connectivity graph (Brin and Page, 1998) with the random walk traversal. It also considers the relative

---

importance by checking the rank of documents – a document is ranked as highly important when the document has backlinks from documents with high authority, such as the Yahoo homepage.

However, automatic citation analysis is limited in that it does not reflect well the importance of a document from a human perspective. There are many cases where simple citation counting does not reflect our commonsense concept of importance (Brin and Page, 1998). This research addresses this problem by exploring a method of ranking based on human interactions, where a pool of expert human agents are used to evaluate web documents and their authority is dynamically determined through user feedback on their performance.

Rocchio (1971) proposed relevance feedback for query modification, where users judge the relevance of a document for a query and leave feedback. The system then updates the query based on the feedback. This has been shown to be quite effective in query modification. Following this idea, we apply the relevance feedback of users to the ranked documents provided by the expert group. The feedback information will modify the authority weight of the expert group members. As a result, the decisions of the expert group will reflect the feedback of users as time passes.

In this paper, we suggest a novel recommender system based on human interactions. All the key decisions follow human opinions from a specialised or “expert” group, so more reasonable recommendations can be made available in situations that are vague because few users have evaluated an item. Automatic selection or ejection of expert members based on their performance can be used to maintain the expertise of the group. The relevant documents provided by the expert group are sorted in rank order. To check the effectiveness of the system, we have developed several effectiveness measures based on rank order. In this paper we validate our approach with simulations of user feedback and expert group reorganisation, and evaluate the results using the new effectiveness measures. Our preliminary work was published in conference proceedings (Kim and Kim, 2001; Kim and Chung, 2001).

## Proposed method

### *Dynamic authority weights of experts*

We define a group of people with high authority and much expertise in a special field as an “expert group”.

Figure 1 shows a framework for a search engine with our recommender system. A meta-search engine is used to collect good web documents from the conventional search engines (e.g. Yahoo, AltaVista, Excite and InfoSeek). The addresses of the documents cited in the search engines are stored in the document database. Also recorded for each web document are details of how many search engines in the meta-search engine referred to the document, and how many times online users had accessed the web document using the search engine.

For every category there is a list of top-ranked documents rated by an expert group, which are sorted by score. Authoritative webpages are determined by human expert group members. The experts examine the content of candidate webpages that are highly referenced among web documents or have been accessed by many users. 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

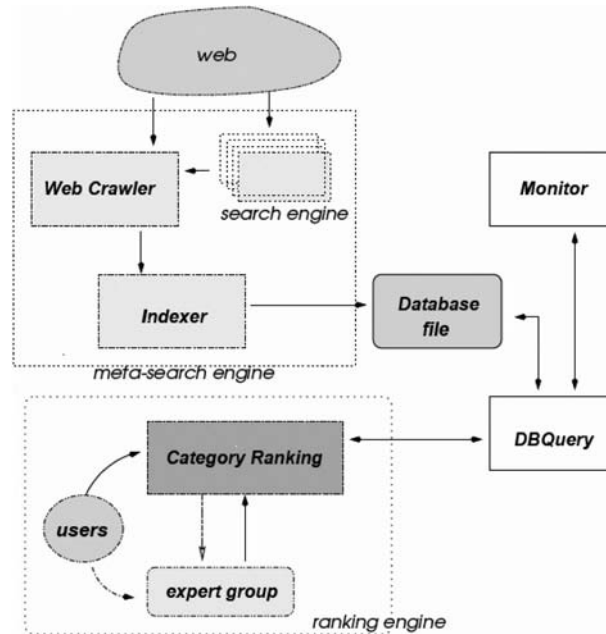


Figure 1.  
Search engine architecture

judgments are properly combined. In our system, experts decide whether a web document should be classified as a recommended document for a given category. A simple way to combine the experts' individual judgements is majority voting (Liere and Tadepalli, 1997; Li and Jain, 1998), where each expert has a binary vote for each web document and the documents obtaining equal to or greater than half of the votes are classified into a top-ranked list.

An alternative method is a weighted linear combination, where a weighted linear sum of expert voting yields the collaborative net-effect ratings of documents. In this paper, we take the adaptive weighted linear combination method, where the individual contributions of members of the expert groups are weighted by their evaluation performance. All the experts' evaluations are summed with weighted linear combinations. The expert rating results will dynamically change depending on each expert's performance. Our approach to expert group decision-making is similar to the classifier committee concept of Li and Jain (1998) and Sebastiani (1999), except that their methods use classifiers based on various statistical or learning techniques instead of human interactions and decisions. This weighted measure is useful even when the number of experts is not fixed.

How to choose experts and decide authority weights is an issue. Initially, experts will be selected from among the users who have most frequently rated products or documents. A positive authority weight will be assigned to each expert member. The voting results of experts will determine a score over a given document. The score ranking will reflect the importance or popularity of the document. As time goes on, the authority weight will be changed depending on users' feedback. An expert will receive a higher authority weight if his or her opinion agrees with those of general users, and

otherwise, a lower authority weight. If the authority weight becomes negative, the corresponding expert will be dropped from the representative board and a new member will be chosen from among users who have highest participation in evaluating expert opinions. If there is more than one user who has provided the most frequent feedback, one user will be randomly chosen from among them. In this way, the constitution of the expert group is dynamically changed.

We define a rating score matrix  $X = [X_{ij}]$ , when the  $i$ -th expert rates a web document  $d_j$  with a score  $X = [X_{ij}]$ . For each web document  $d_j$ , the voting score of an expert committee is given as follows:

$$V(d_j) = \sum_{i=1}^{N_e} r_i \chi_{ij} = \sum_{i=1}^{N_e} \frac{w_i}{\sum_{k=1}^{N_e} w_k} \chi_{ij}$$

where  $N_e$  is the number of experts for a given category and  $r_i$  is the relative authority for the  $i$ -th expert member of the expert pool, and  $w_i$  is the authority weight for the  $i$ -th expert member. We suppose  $w_i$  should always be positive. The weight  $w_i$  is a dynamic factor, and it represents each expert's authority to evaluate documents. A higher authority weight indicates that the expert has more influence in a voting decision.

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

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

where  $n$  is the number of documents evaluated by users,  $V'(d_j)$  is the users' voting score for an expert-voted document  $d_j$ . We assume  $V'(d_j)$  is the average over all user scores, but in reality it is rarely possible to receive feedback from all users. The authority weight for each expert is changed every session, which is a given period of time, and at the same time  $V'(d_j)$  can be approximated by the central limit theorem with a set of  $\tilde{V}'(d_j)$ , which is the average user rating during the given session.

We use a gradient-descent method over the error measure  $E$  with respect to a weight  $w_i$  and the gradient is given by:

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

where  $S = \sum_{k=1}^{N_e} w_k$  is the sum of weights, and  $\Delta_j = [V(d_j) - \tilde{V}'(d_j)]$  is the difference between the predicted voting score and the users' rating score during a session for a document  $d_j$ :

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

We apply the similar scheme shown in error back-propagation of multiplayer perceptrons (Haykin, 1999) to our approach. If we update the weights of experts with the feedback of users about a web document  $d_j$ , the weight is changed each session by the following dynamic equation:

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

where  $\eta$  is a learning rate proportional to the number of user ratings per session and  $\alpha$  is the momentum constant.

The above equation says how to reward or penalise the authority weights of experts for their share of responsibility for any error. According to the equation, the weight change involves the correlation between a voting score difference among experts and the error difference. For example, when both an expert-voted score and the desirable-rank score are larger than the weighted average voting score, or both of them are smaller than the average score, the expert is rewarded; if otherwise, the expert is penalised. In this case some experts have rewards and others receive penalties depending on the weighted average voting score of the expert group.

### *Evaluation of effectiveness*

When dynamic authority weights are assigned to experts for a category, the expert group ratings can form a ranking list in order. We need to determine if the given ranking list is reliable. Reliable ranking means that good experts have been selected for an expert group and they recommend relevant documents or items to general users. We evaluate the prediction performance of expert groups in terms of effectiveness – that is, a measure of the agreement between expert groups and users – in ranking a test set of web documents. We assume there are many users to evaluate the top-ranked lists in contrast to a small number of experts in a category group.

We suggest several effectiveness measures that are related to the agreement in rank order between expert ratings and user ratings. They are rank order window measure, rank function measure, and  $F_\beta$  measure with rank order partition. We compared these with Spearman's correlation measure, which is a common measure in the information retrieval field.

*Rank order window measure.* Given a sample query or category, we can represent the effectiveness as the percentage of top-ranked lists that user ratings rank in the same or very close position as an expert group does. Given top-ranked web documents  $D = \{d_1, d_2, \dots, d_n\}$ , we can define effectiveness  $\lambda_\delta$  with rank order window  $\delta(d_k)$  as:

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

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

where  $d_k$  is the  $k$ -th web document from the test set for a given category, and  $\delta(d_k)$  is the width of the window centred in the rank  $\mu(d_k)$  assigned by the ratings of experts for  $d_k$ .  $Q(d_k)$  is the rank position of the average rating score of users for a document  $d_k$ .  $S(d_k)$  calculates the rate of the rank order difference in the window  $[\mu(d_k) - \delta(d_k), \mu(d_k) + \delta(d_k)]$ .

For this measure, we directly compare the rank of documents that the expert group provides with the rank given by users. For each of the top-ranked documents the experts recommend, we calculate how much the rank position is changed by user feedback. However, we check the position change within a window size.

*Rank function measure.* Given web resources  $D = \{d_1, d_2, \dots, d_n\}$ , and a set of all rank functions  $\Phi$  over the set  $D$ , we suppose that  $d_1, d_2, \dots, d_n$  is decreasingly ordered by their weighted rating values according to experts' evaluations. We define a measure  $\rho$  to evaluate a ranking function over given ranked web documents  $D$  as follows:

$$\rho(\phi, d_k) = \text{Card}(\{d_i \in D | (1 \leq i < k) \wedge \phi(d_i) < \phi(d_k)\})$$

where  $\text{Card}$  is a cardinality function to count the number of elements in a given set, and  $\phi$  is a rank function over web resources  $D$ , which gives a sorting order.

We define a user satisfaction function  $\Psi$  over expert-voted ranked sites  $D$  as follows:

$$\psi(\bar{\phi}) = \frac{\sum_{k=1}^n \rho(\bar{\phi}, d_k)}{(n-1)(n-2)/2}$$

where  $\bar{\phi}$  is the rank function obtained from the result of all user ratings for  $n$  documents, and  $0 \leq \Psi \leq 1$ .

Similar to the rank order window measure, we compare the user feedback rank and the expert group rank for a document. We calculate the distance of rank difference for each document and sum all the distances for ranked documents that the expert group recommend.

*F<sub>β</sub> measure with rank order partition.* The evaluation of search effectiveness in a document is one of the essential components in information retrieval. In the objective evaluation of retrieval techniques, two properties – precision and recall – have been accepted as general-purpose evaluation criteria. Precision is the conditional probability that when a document is predicted to be in a positive class, it truly belongs in this class. Recall is the conditional probability that a document belonging to a positive class is truly classified in this class (Raghavan *et al.*, 1989; Sebastiani, 1999). A good information retrieval system will have high precision and recall. Researchers have sometimes applied a variation of precision and recall, since the two properties have a trade-off depending on their application. The two properties can be assembled with the  $F_\beta$ -measure, a combination of precision and recall (Raghavan *et al.*, 1989; Sebastiani, 1999).

We suggest a variation of precision and recall for the rank order system. We first partition recommended documents by their rank order and make classes. We define a positive class  $i$  as the top  $[10(i-1)+1 \sim 10i]$  ranked documents by expert voting and a negative class as the others. For example, class 2 documents are the 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 ratings and user ratings, and those probabilities in our application can be calculated with transition instances between classes. A transition instance  $p_{ij}$  is defined as the number of instances that are predicted to be in class  $i$  by expert ratings, but that belong to class  $j$  by user ratings. Here we give the distance penalty among the classes, since we consider rank order relations. If the actual rating class  $j$  is closer to the predicted class  $i$ , then we give higher precision probability:



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

$$\bar{P} = \frac{\sum_{i=1}^m P_i}{m}, \quad \bar{R} = \frac{\sum_{i=1}^m R_i}{m}$$

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 the effectiveness measure can be computed using the value of  $0 \leq \beta \leq \infty$  (van Rijsbergen, 1979; Cohen and Singer, 1999; Yang, 2000):

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

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 results can be seen to have many false responses from the feedback of general users or many new documents positioned in the top ranks. If  $F_\beta$  is close to one, then top-ranked sites have good feedback from general users and little change occurs in the top-ranked lists.

*Spearman's correlation measure.* Spearman's rank order correlation measure, which is a popular measure for information retrieval systems, checks whether rank-ordered data is correlated. Let  $x_i$  be the rank of a document  $d_i$  in  $D = \{d_1, d_2, \dots, d_n\}$  by expert ratings and  $y_i$  the rank of  $d_i$  by user ratings. The non-parametric correlation is defined to be the linear correlation coefficient of the ranks:

$$r_s = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

where  $\bar{x}$ ,  $\bar{y}$  are the average of  $x_i$ ,  $y_i$ , respectively. A value of  $r_s = 1$  indicates a completely positive correlation, which is desirable in our application, and  $r_s = 0$  indicates no correlation of the data.

*User confidence level.* We maintain a rating record for each user. Many users evaluate web documents, but it is difficult to extract user preferences due to the lack of rating information. Before reflecting each user evaluation, we need to check if each user has rated a sufficient number of documents and the user evaluations are reliable. Thus, if we assume a rating score level from 1 to  $m$ , the confidence level  $C$  for a user  $u$  is defined as follows:

$$C(u) = -\sum_{i=1}^m p(i) \log p(i)$$

where  $p(i)$  is the probability that the user  $u$  rates documents with score  $i$ , and it can be calculated by counting the number of documents with score  $i$  among all the documents that the user  $u$  has rated.

The confidence level of a user is an entropy measurement to check the distribution of score ratings. If it is more equally distributed, it is more likely that the user has given a sufficient number of ratings and also that the user has unbiased criteria for evaluations. For example, if a user consistently puts only low scores or high scores on

---

web documents, the user has a low confidence level. The rating information of users who keep low confidence levels for many sessions will not be considered for the database of rating scores as well as the analysis of user preferences.

## Experiments

### *Metrics experiment*

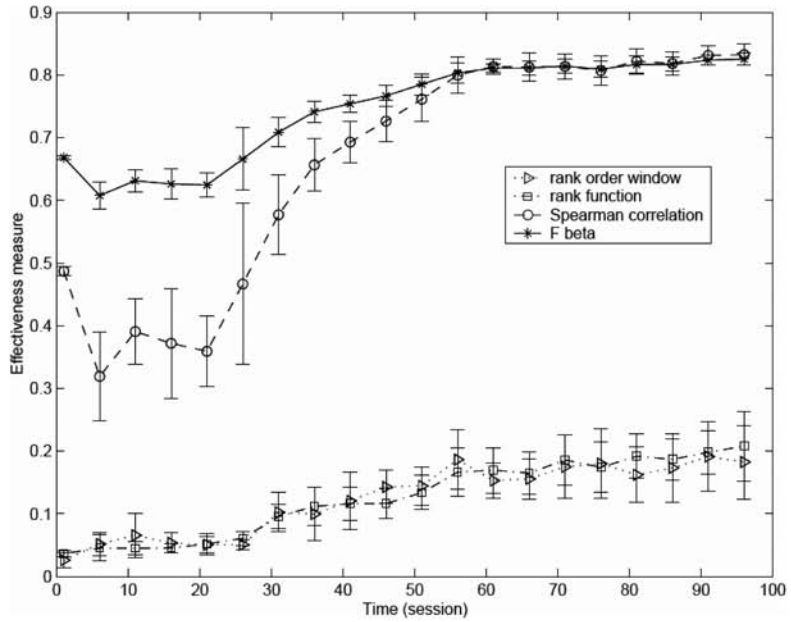
We simulated the dynamic process of web document ranking and the changing constitution of expert groups depending on their performance. (The prediction performance of expert groups in reality will remain for future works.) The purpose of the simulation test was to confirm that the decisions of dynamic expert groups reflect general users' opinions or ratings, and that this approach has the potential to recommend documents that have not been rated yet. The results of the test also show how effective the system is.

In the simulation, we assumed: ten categories to need expert groups, a maximum of ten experts for each group, 10,000 web documents ( $d_k, k = 1, \dots, 10,000$ ) in the movie search engine, and 500 random users logged into our search engine. We modelled random log-in patterns of online users as a Poisson process (Ross, 2000; Taylor and Karlin, 1998). Each user had an arrival rate, in other words, an access rate and a transaction processing time, thus we defined the arrival rate  $\lambda_i$  for a user  $u_i$  for  $I = 1, \dots, 500$ . For each user  $u_i$ , the probability that the user accessed the search engine document within time  $\Delta t$  was  $P_i = 1 - e^{-\lambda_i \Delta t}$  where  $\Delta t$  is the basic time unit. This Poisson process closely resembles the pattern of incoming random patterns. Thus, our simulation work is valid for the real system.

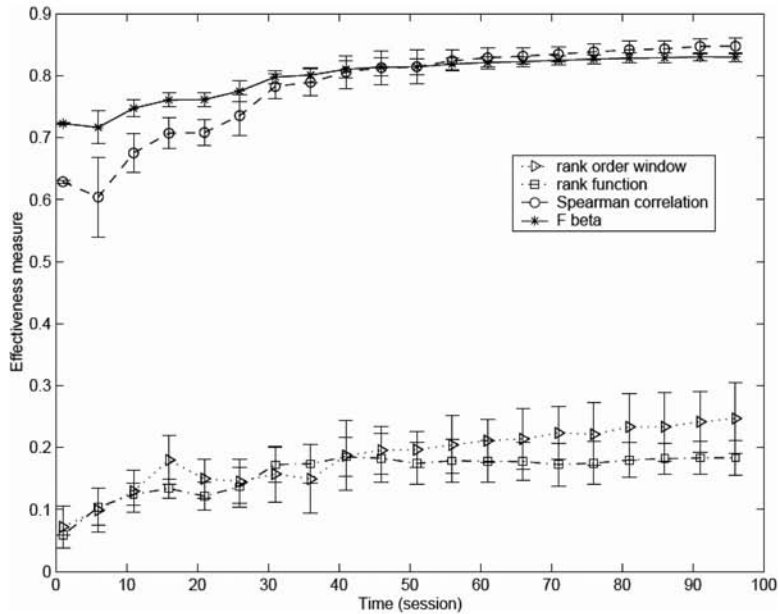
For every session we selected the top-100 ranked documents recommended by an expert group for each category and applied our effectiveness measures to the top-ranked lists. For the rank order window measure, we used window size  $\delta(d_k) = 4$ . For the  $F_\beta$  measure, we grouped the top-100 ranked documents into ten classes, each of which contained ten documents.

Figure 2 shows the plots of effectiveness with four different measures, as the dynamic process of ranking evaluations continued. The expert group members and their knowledge levels were fixed for each category, and a random sequence of user ratings given. The results show the agreement level between the expert groups and the users in ranking documents according to the queries or categories. The simulation was run ten times for each category, and only two category results are displayed among ten categories. The figures show the average performance results with 95 per cent confidence intervals.

The results of the rank order window measure were similar to those of the rank function measure while the results of the Spearman's correlation measure were similar to those of the  $F_\beta$  measure. The rank order window and rank function measure can be seen as micro-view evaluations of rank order difference, and the others as the macro-view. In Figure 2, the rank function or rank order window measure has a relatively low value, compared with the Spearman's correlation or  $F_\beta$  measure. It is because the micro-view evaluations would need a more perfect rank match with desired ranks to reach 1. Even if there is a small number of elements with considerable rank difference, the measures will be severely influenced. However, it is notable that the four measures show a similar trend of curves, as the decision performance increased.



(a) Results of the effectiveness measures under different categories – category A



(b) Results of the effectiveness measures under different categories – category B

**Figure 2.**  
Results of the effectiveness measures under different categories – category A and category B

---

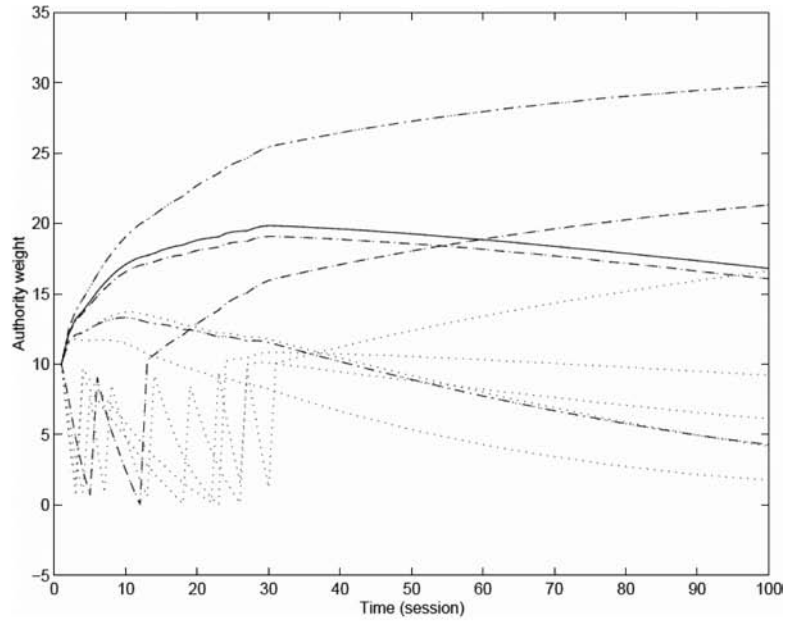
Figure 3 shows the transitions of experts' authority weights according to their rating performance. Each expert has an initial weight of 10 and a maximum weight of 30 is allowed to prevent too much authority across only a few experts. We assume that when a weight becomes negative, the corresponding expert is dropped from the group. In the simulation experiments some experts kept high authority weights for a while and then yielded their authority levels to other experts. Many experts with poor performance were dropped from the expert groups and then new members added – many oscillating curves of authority weights are seen between 0 and 10 in Figure 4. As a result, the above process plays a role in filtering out poor experts and keeping good experts as time passes. As the iteration of weight change continues for a long session, the authority weight may become stabilised, as shown in Figure 4 – there is no newcomer in the expert group.

Figure 5(b) shows an example of the agreement of rank order between the expert ratings and the user ratings in evaluating documents, while there is no regular pattern of agreement in the initial state as shown in Figure 5(a). After applying the adaptive change of authority weights, the rank order prediction of experts became close to the rank order of user ratings.

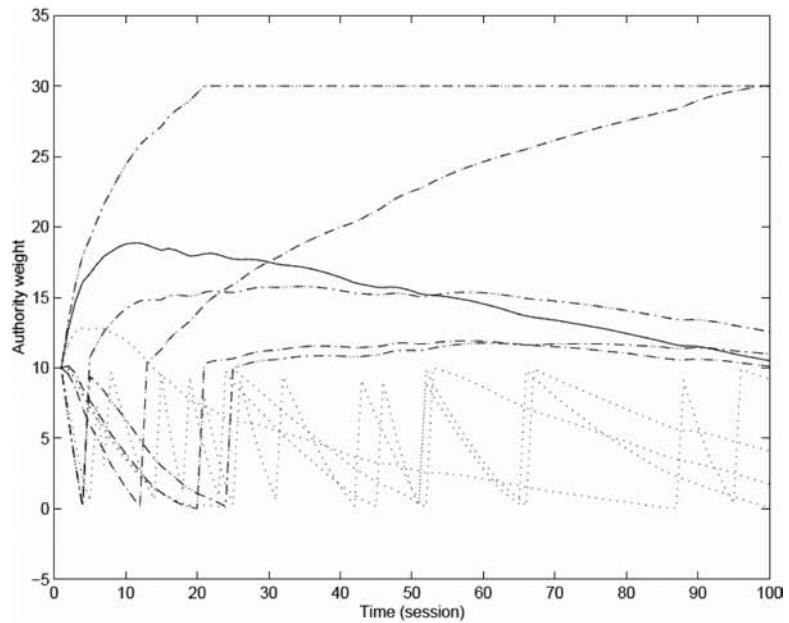
### Discussion

The most popular and largest information provider is the world wide web. The common approach to finding information on the web is to use search engines, which explore HTML pages for specific category information and its relevant documents, normally through keyword-based queries. These days weblogs are becoming more and more popular among web users. Weblogs are websites for comments, information, events or news on a particular subject, which are often organised and maintained by individuals. Interestingly, weblogs can be maintained in an interactive fashion with general users, that is, blog readers often leave comments or feedback about a specific weblog opinion. In this respect, our opinion-based decision approach has a strong potential for application. However, there has been little research on users' online attitudes about and organisation of opinions. We argue that many opinions from weblog readers can be evaluated using the expert group model and effectiveness metrics that we suggest. For the experiments, we designed a new type of weblog opinion representation, called "Blog Cosmos" (which is not shown in this paper), and we were able to find expert groups or opinion leaders on a particular subject weblog with dynamic authority weights for those members.

Our opinion-based decision model has relevance for the personal knowledge management (PKM) field. PKM involves personal inquiry with tools of communication and networked collaboration skills. General users who are interested in a particular subject need to collect appropriate information to improve their knowledge of the subject. With our approach, ranked information about the category, provided by an expert group, is suggested to each knowledge worker (general user), and each worker evaluates the knowledge. The collected feedback of knowledge workers changes the members of the group of opinion leaders and ultimately reorganises the knowledge for the subject. Thus, knowledge workers are participating in social activities of deriving new knowledge and in individual activities through their preferences, and social interactions can produce new knowledge. In this approach, expert users or opinion



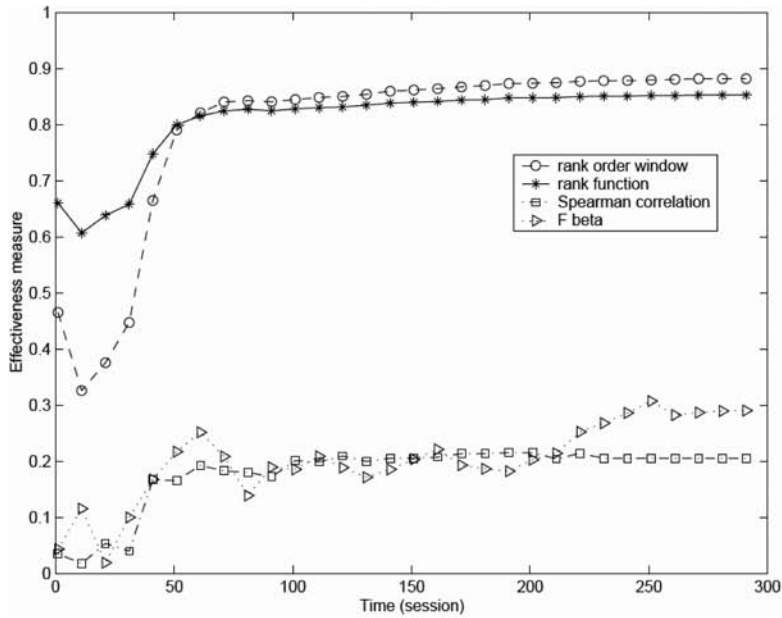
(a) An example of weight change for experts – category A



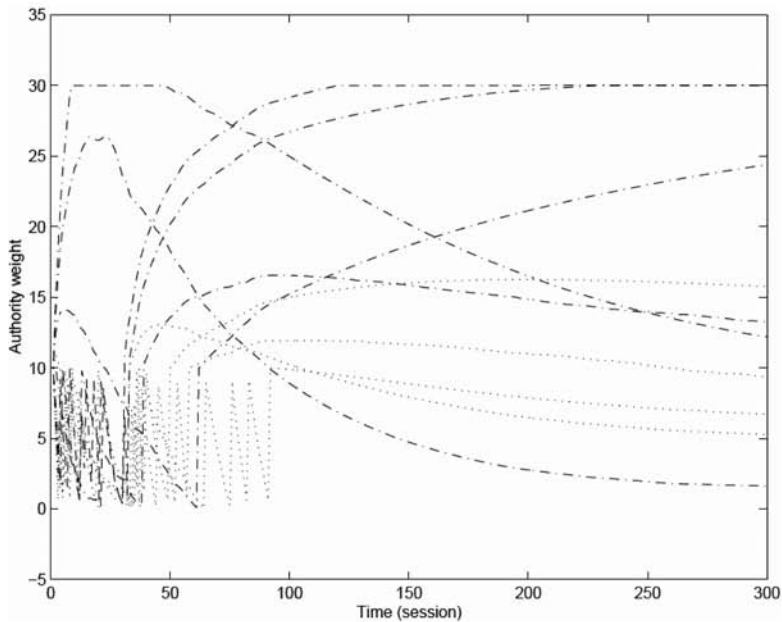
(b) An example of weight change for experts – category B

**Figure 3.**  
An example of weight  
change for experts –  
category A and category B

---

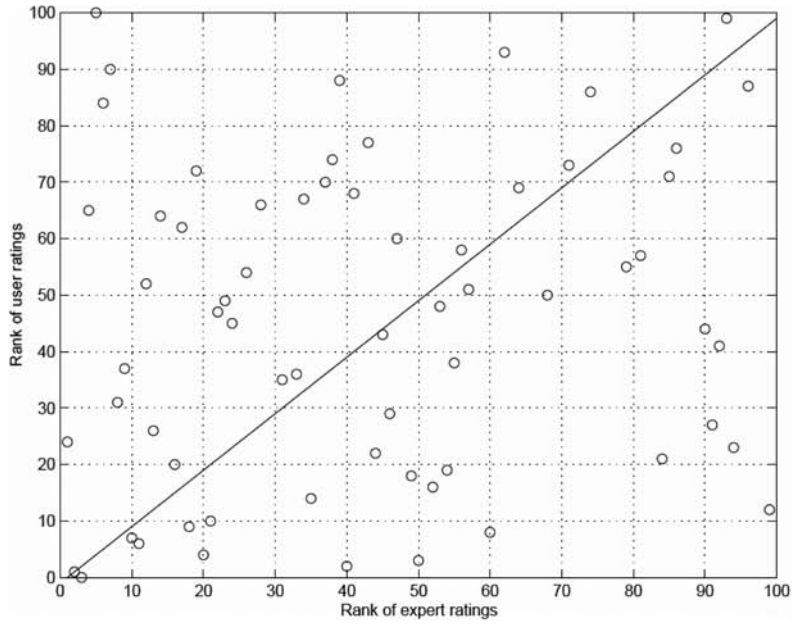


(a) An example of dynamic weights for expert group members and their performance – effectiveness performance

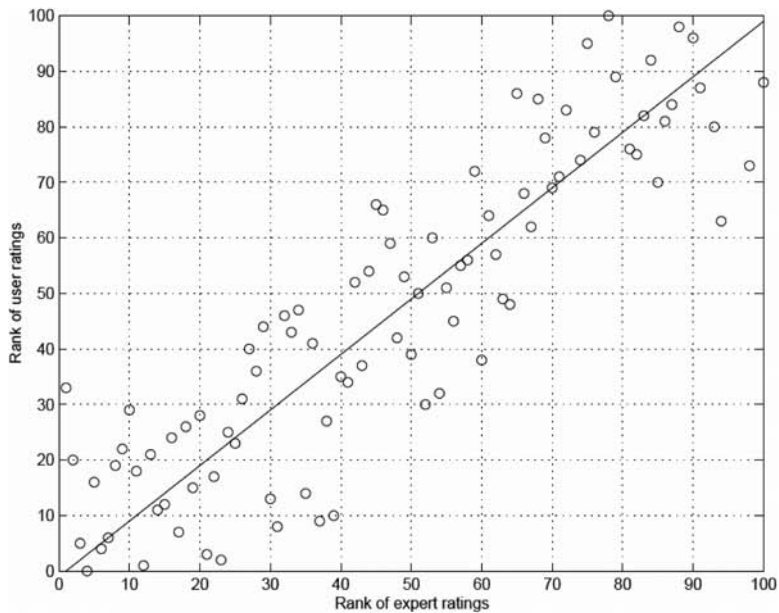


(b) An example of dynamic weights for expert group members and their performance – weight change

**Figure 4.**  
An example of dynamic weights for expert group members and their performance – (a) effectiveness performance; (b) weight change



(a) An example of distribution of rank orders by expert group ratings and user ratings – before weight change



(b) An example of distribution of rank orders by expert group ratings and user ratings – after weight change

**Figure 5.**  
An example of distribution of rank orders by expert group ratings and user ratings – (a) before weight change; (b) after weight change

---

---

leaders themselves are applying their personal knowledge in a collaborative form and the collected knowledge plays the role of social bookmark as a guide for general users.

In this paper, we have used a simulation to test the effectiveness of expert groups for webpage ranking. However, it is quite different from the real world in many respects. Some experts with good knowledge may have different opinions and views from general users. The access rate of each user is very irregular in reality, and the behaviour or access pattern of users will vary from category to category. Also, for some categories, it may be difficult to assemble good experts.

The recommendation feedback process of users enables us to maintain a group of desirable experts for a given category. When users evaluate recommended documents, the system needs more evaluations for better performance, but users may be reluctant to provide feedback for a large number of documents. Experts have their assignments to evaluate many documents. In addition, the scoring criteria of general users may be different from those of experts, which may make it difficult to compare numeric scores between expert groups and general users. Instead of numeric scores between expert groups and users, the rank order difference between two evaluators can be applied to the weight change equation. We leave this work for future research.

Our recommender system with expert groups adapted through user feedback can be applied to the ensemble model of classifiers instead of human agents. This would be useful in the field of text categorisation or collaborative filtering. Also our effectiveness measures based on rank order could be used as evaluation criteria for collaborative filtering with the analysis of user preferences. In our approach, the current score ratings of users range from 1 to 10. Instead of score ratings of documents, the rank information can be extracted from users' relative ratings, such as positive or negative, depending on their satisfaction with experts' recommendations. If we use the relative rating method, it would be easy to decide users' satisfaction levels. The use of relative ratings instead of score ratings can lead to a probabilistic method with prior probabilities where the expert-recommended documents keep their ranks.

In our model, we assume that the expert group is a representative board to provide opinions to general users in a democratic way. According to this concept, a good expert group should have the same or similar opinions to general users. For example, for a shopping website, the expert group should recommend a good shopping list under a specific category, and here, good experts will give recommendations that many users are likely to follow. Also, if the expert group is not established at the initial stage, we can use the suggested method with dynamic authority weights to choose the expert members in the representative board. In most cases, we can hardly obtain all users' opinions in time and so a small number of experts will provide the opinions on behalf of a larger number of users. This will reduce the time and cost of recommending documents or lists. This is a practical advantage of our approach.

As an alternative application, this system could be applied to the journal reviewing system. If the experts are well known, the expert groups will not need to be updated. However, if the expert group is not established well at the initial stage, then we can receive feedback from authors on whether reviewers' comments were helpful or not. Then we can evaluate the reviewers' expert knowledge and reorganise the review committee.



### Conclusion and future works

In this paper we suggest a new type of recommender system with an opinion-based decision model. Dynamic expert groups, automatically formed from among users, rank web documents. Each expert has his or her own authority to evaluate webpages. This authority is dynamically changed by the feedback of users.

We have clustered the top-ranked documents by their rank order, and applied various effectiveness measures to those clusters to evaluate whether the ranking list is reliable. As the user feedback and rating process continues, the dynamic authority weights increase for good experts and decrease for poor experts. This automatically chooses good experts for a given category and thus it improves the effectiveness measures that we suggest – rank order window measure, rank function measure, and  $F_\beta$  measure with rank order partition. Our effectiveness measures have the potential to select a pool of good experts. The system accumulates more user-feedback ratings of expert-recommended documents as time passes, and it can improve the selection of experts as well as use collaborative filtering methods with the analysis of user preferences. We intend to apply this approach to design a meta-search engine or weblog for movie, music and shopping mall sites. In many such areas, customers or users are interested in seeing the top-ranked documents or products.

The recommender system with dynamic expert groups will be a feasible solution to recommend items or documents for unspecified user fields that automatic recommender systems cannot cover. Future work should explore natural language processing of a document for categorisation. Furthermore, users could use the mobile Short Message Service (SMS) to evaluate web documents, which could increase the feedback rates of users.

### References

- Breese, J.S., Heckerman, D. and Kadie, C. (1998), "Empirical analysis of predictive algorithms for collaborative filtering", *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43-52.
- Brin, S. and Page, L. (1998), "The anatomy of a large-scale hypertextual web search engine", *WWW7: Proceedings of the Seventh International Conference on the World Wide Web*, Elsevier, Amsterdam, pp. 107-17.
- Cohen, W. and Singer, Y. (1999), "Context-sensitive learning methods for text categorization", *ACM Transactions on Information Systems*, Vol. 17 No. 2, pp. 141-73.
- Delgado, J., Ishii, N. and Ura, T. (1998), "Content-based collaborative information filtering: actively learning to classify and recommend documents", *Proceedings of Cooperative Information Agents*, Springer, Heidelberg, pp. 206-15.
- Fern, A., Natarajan, S., Judah, K. and Tadepalli, P. (2007), "A decision-theoretic model of assistance", *Proceedings of the International Joint Conference on Artificial Intelligence 2007*, Vol. 2007, IJCAI, Hyderabad, pp. 2-8.
- Haykin, S.S. (1999), *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.
- Hill, W., Stead, L., Rosenstein, M. and Furnas, G. (1995), "Recommending and evaluating choices in a virtual community of use", *Proceedings of the SIGCHI conference on Human Factors in Computing System*, ACM, New York, NY, pp. 194-201.

- 
- Kim, D. and Kim, S.W. (2001), "Dynamic models of expert groups to recommend web documents", *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*, Springer, Heidelberg, pp. 275-86.
- Kim, S.W. and Chung, C.-W. (2001), "Web document ranking by differentiated expert group evaluation", *Proceedings of the 9th International Conference on Human Computer Interaction, New Orleans, LA*, pp. 415-9.
- Kleinberg, J. (1998), "Authoritative sources in a hyperlinked environment", *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, NY, pp. 604-32.
- Konstan, J., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J. (1997), "GroupLens: applying filtering to usenet news", *Communications of the ACM*, Vol. 40 No. 3, pp. 77-87.
- Li, Y.H. and Jain, A.K. (1998), "Classification of text documents", *The Computer Journal*, Vol. 41 No. 8, pp. 537-46.
- Liere, R. and Tadepalli, P. (1997), "Active learning with committees for text categorization", *Proceedings of 14th Conference of the American Association for Artificial Intelligence, Rhode Island*, pp. 591-6.
- 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*, Vol. 7 No. 3, pp. 205-29.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994), "GroupLens: an open architecture for collaborative filtering of netnews", *Proceedings of Computer Supported Cooperative Work Conference*, pp. 175-86.
- Robertson, S.E. (1977), "The probabilistic ranking principle in IR", *Journal of Documentation*, Vol. 33, pp. 294-304.
- Rocchio, J.J. (1971), "Relevance feedback in information retrieval", in Salton, G. (Ed.), *The SMART Retrieval System - Experiments in Automatic Document Processing*, Prentice-Hall, Englewood Cliffs, NJ, pp. 313-23.
- Ross, S.M. (2000), *Introduction to Probability Models*, London.
- Salton, G., Wong, A. and Yang, C.S. (1975), "A vector space model for information retrieval", *Communications of the ACM*, Vol. 18 No. 11, pp. 613-20.
- Sebastiani, F. (1999), "Machine learning in automated text categorisation: a survey", Technical Report, IEL-B4-31-1999, Istituto di Elaborazione dell'Informazione, CNR, Pisa.
- Shardanand, U. and Maes, P. (1995), "Social information filtering: algorithms for automating 'word of mouth'", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, pp. 210-7.
- Soboroff, I. (2006), "Dynamic test collections: measuring search effectiveness on the live web", *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 6-11 August, Seattle, WA*.
- Soboroff, I., Nicholas, C. and Pazzani, M. (Eds) (1999), *Proceedings of the SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluations, Berkeley, CA*.
- Taylor, H.M. and Karlin, S. (1998), *An Introduction to Stochastic Modeling*, 3rd ed., Academic Press, London.
- Turtle, H.R. and Croft, W.B. (1991), "Evaluation of an inference network-based retrieval model", *ACM Transactions on Information Systems*, Vol. 9 No. 3, pp. 187-222.
- van Rijsbergen, C.J. (1979), *Information Retrieval*, Butterworths, London.
- Yang, Y. (2000), "An evaluation of statistical approaches to text categorization", *Information Retrieval*, Vol. 1 Nos 1-2, pp. 69-90.

---

**Further reading**

- Baize-Yates, R., Castillo, C., Martin, M. and Rodriguez, A. (2005), "Crawling a country: better strategies than breadth first for the web page ordering", *Proceedings of the 14th International Conference on the World Wide Web*, ACM, New York, NY, pp. 2-7.
- Birnbaum, M. and Stegner, S. (1979), "Source credibility in social judgment: bias, expertise and the judge's point of view", *Journal of Personality and Social Psychology*, Vol. 37 No. 1, pp. 48-74.
- Hawkins, D., Best, R. and Coney, K. (2004), *Consumer Behavior: Building Marketing Strategy*, McGraw-Hill, Boston, MA.
- Nakajima, S., Tatemura, J., Hino, Y., Hara, Y. and Tanaka, K. (2005), "Discovering important bloggers based on analyzing blog threads", *Proceedings of WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 10 May, Chiba, Japan.
- Rocha, C., Schwabe, D. and Poggi de Aragao, M. (2004), "A hybrid approach for searching in the semantic web", *Proceedings of the 13th International Conference on World Wide Web*, ACM, New York, NY, pp. 374-83.
- Setiono, R. and Liu, H. (1997), "Neurolinear: a system for extracting oblique decision rules from neural networks", *Proceedings of European Conference on Machine Learning, Lecture Notes in Computer Science*, Vol. 1224, pp. 221-33.

**About the authors**

Sea Woo Kim received his PhD from the Korea Advanced Institute of Science and Technology (KAIST). He is Assistant Professor at Soong Eui Women's College and has worked for Korea Telecom International. He is interested in human-computer interaction, the internet, artificial intelligence and information retrieval.

Chin-Wan Chung received his Masters and PhD degrees from the University of Michigan. He is Professor at the Korea Advanced Institute of Science and Technology (KAIST) and has worked at General Motors Research Laboratories. His areas of interest include the semantic web, sensor network databases and multimedia information.

DaeEun Kim received his BE and MSc in engineering from the Department of Computer Science and Engineering of Seoul National University and the University of Michigan at Ann Arbor, respectively. From 2002 to 2006 he was a research scientist at the Max Planck Institute for Human Cognitive and Brain Sciences. Currently he is Assistant Professor at Yonsei University. DaeEun Kim is the corresponding author and can be contacted at: daeeun@yonsei.ac.kr