Multidimensional Models of Information Need

Yunjie (Calvin) Xu
National University of Singapore
xuyj@comp.nus.edu.sg

Joseph K.H. Tan
National University of Singapore
joseph tan@ nus.edu.sg

Abstract

User studies in information science have recognized relevance as a multidimensional construct. An implication of multidimensional relevance is that a user's information need should be modeled by multiple data structures to represent different relevance dimensions. While the extant literature has attempted to model multiple dimensions of a user's information need, the fundamental assumption that a multidimensional model is better than a uni-dimensional model has not been addressed. This study seeks to test this assumption. Our results indicate that a retrieval system that models both topicality and the novelty dimension of a users' information need outperforms a system with a uni-dimensional model.

Keywords: information retrieval, information behavior, relevance, topicality, novelty, redundancy.

This paper was accepted for publication at the Journal of Information & Knowledge Management in 2008 and appeared in Vol. 7, No. 4 (2008) 1–14. This is a working paper version.

Multidimensional Models of Information Need

1. Introduction

Knowledge is a primary resource for boosting organizational performance (Alavi & Leidner, 2001). Many knowledge management initiatives set as their objectives, ways to better manage the capture, storage, distribution and application of knowledge. Information retrieval (IR) is identified as one of the major research topics in knowledge management (Alavi & Leidner, 2001). However, traditional information retrieval techniques are troubled by the typically short user queries which often lead to inferior retrieval results. For example, Spink et al. (2000) observed that most queries contain only 2-3 words. A users' information need is not fully reflected in queries because users are not aware of the existence of a rich contextual backdrop against which they specify their queries, and because they are unable or unwilling to describe their information need. Even if they are to articulate their information need, the outcome is often incomplete and inaccurate (Belkin 1980).

In our study, an *information need* is defined as the awareness of the gap between an extrinsic demand for information and a seeker's intrinsic knowledge. This definition is closely aligned with the notion of the anomalous state of knowledge (Belkin et al. 1982). A user's query is a manifestation of a latent information need, but not the information need per se. Users seeks information to satisfy their information need.

How to better represent users' information need is an important research question in IR. The recognition of users' inability to specify adequate queries leads to a growing increase in literature on query expansion techniques (e.g., relevance feedback, collaborative query recommendation) with the objective to better capture users' tacit information need. However, the effectiveness of various techniques is still considered mixed (Ruthven, 2003). The majority of extant research has regarded information need as a uni-dimensional construct which can be represented by a single data structure such as a vector of terms (Ruthven, 2003). However, a multidimensional view of information need has been proposed recently (Yang, Zhang, Carbonell & Jin, 2002; Zhang, Callan, & Minka, 2002, Xie, 2000; Xu, 2007). While these studies have made an important assumption that a multidimensional model of users' information need is better than a uni-dimensional one, they have not empirically compared a multidimensional model of information need with one that is uni-dimensional. Therefore, a strong theoretical claim on the superiority of multidimensional relevance cannot be made. The purpose of our study is to test the assumption that a multidimensional model of users' information need is better than one that is uni-dimensional. This study also compares a few different multidimensional models of users' information need.

This paper is organized as follows: First, the notions of multidimensional information need and relevance judgment are reviewed (Grice, 1975; Xu & Chen, 2006). Then, hypotheses of this study are introduced. A set of four IR systems which correspond to the hypotheses is then introduced. After that, a user study is reported and the performances of the four IR systems are compared. To conclude, implications are discussed.

2. Conceptual Background

2.1 Information Need and Relevance

2.1.1 Information Need

The extant literature has conceptualized information need in various ways. Based on its content, an information need has been conceptualized as a monolithic need or a hierarchy of interrelated subneeds (Xie, 2000; Xu, 2007). Traditional methods like the vector space model or the language model (Ponte & Croft, 1998) treat a query as the only surrogate of an information need. In other words, a document is directly compared against the query. Recently, an information need has come to be regarded as a multi-level construct consisting of a long-term goal and many short-term goals (Xie, 2000). Xu (2007) defined them as the general information need and subneeds, and postulated that users adopt a divide-and-conquer strategy in information retrieval. Based on temporal stability, an information need can be assumed to be static or dynamic. A static information need does not change over time and is not related to the past information needs of a user. IR tests with a standard Text Retrieval Conference (TREC, trec.nist.gov) dataset often assume information needs are static. A dynamic information need assumes that a user's information needs are related over time. Therefore, a later query is often a revision of an earlier query. Some popular query revision tactics include making morphological changes, broadening a term, narrowing a term, and using a related term (Bates, 1979; Fidel, 1991). Finally, based on users' relevance judgment of documents, information needs can be conceptualized as uni-dimensional or multidimensional.

2.1.2 Relevance

The notion of relevance judgment is closely related to the notion of information need. If an information need is the gap between an extrinsic demand for information and a seeker's intrinsic stock of information, the relevance judgment of a document is the perceived degree of match between the document content and the information need, in other words,, how effectively a document can fill the gap (Saracevic, 1975; Xu, 2007; Xu & Chen, 2006). Because information needs are often unobservable, users' relevance judgments of documents manifest their information need. Besides being a subjective perception, relevance is considered as multidimensional and dynamic. The dimensions (i.e., criteria) employed in relevance judgments reflect the dimensions of users' information need. An extensive literature review of relevance can be found in (Borlund, 2003; Mizaro, 1997; Saracevic, 1975, 2007; Xu & Chen, 2006).

Of particular importance to this study is the multidimensionality of relevance. Numerous relevance criteria have been identified in extant literature such as document availability, novelty, currency, information quality, content topicality, presentation quality, and source characteristics (Bateman, 1998; Schamber, 1994). It is theoretically unnecessary and empirically infeasible to incorporate all these criteria as dimensions of relevance judgment in the design of IR systems. To reduce the dimensionality of relevance to a manageable size, Xu and Chen (2006) suggested that Grice's (1975) inferential communication theory could serve as a foundational theory to summarize the relevance judgment criteria. In summary, Grice (1975) suggested that communication is successful if the information delivered is reliable, on-topic, novel, understandable, and of appropriate scope. Empirical verifications of the five criteria in relevance judgment (Xu & Chen, 2006) suggest that (1) most of the relevance criteria proposed earlier can statistically be grouped into the five criteria, and (2) topicality and novelty are the two major criteria in predicting the relevance judgment of a document. Topicality is defined as the extent to which a document addresses the topic of the information need, and novelty is defined as the extent to which the content of a retrieved document is new to the user and different from what the user has known previously. Their finding is consistent with an earlier hypothesis made by Boyce (1982), which postulated that users follow a two-stage retrieval process. In the first stage, documents are filtered by topicality. In the second stage, documents are sorted by 'informativeness', a notion very close to novelty. Therefore, this study focuses on topicality and novelty as the two major dimensions of relevance judgments.

Topicality and novelty differ in degree of subjectivity and dynamics. Borlund (2003) defines topicality as the appropriateness of a document belonging to a subject area as indicated by a user's query. Different judges can often agree upon whether the topical content of a document is related to a query. Vakkari (2003) suggests that the topicality judgment of a document does not change within one search session. Therefore, for a given search task, topicality is relatively objective and stable over time. In contrast, novelty is more subjective and volatile due to users' different background knowledge (Barry, 1994; Bateman, 1998). While different users might collectively consider a document as on-topic, they might disagree on its novelty due to varied prior knowledge. A novel document can cause a significant change in users' cognition, which in turn affects their information need and their criteria for relevance judgment for subsequent documents (Harter, 1992). Therefore, in a list of documents, a document's novelty is even affected by the order it assumes (Zhang et al. 2002). Finally, learning speed differs for different individuals and tasks, depending on an individual's motivation and task urgency. The volatility and subjectivity of novelty implies that an IR system constantly has to monitor a user's behavior to capture and incorporate the novelty aspect of a user's information need. Empirical studies (Xu & Liu, 2007) have confirmed that novelty is more volatile than topicality in a search session.

2.1.3 Past Studies on the Dimensionality of Information Need

In order to better represent users' information need, most studies focus on the algorithms for query expansion, but a few directly target the understanding of users' information need. The extant research has found that (1) an information need, when it is not so simple as to be expressed in one query, often contains a hierarchy of subneeds and subtopics (Xie, 2000; Xu, 2007); and (2) because of users' multidimensional relevance perceptions of documents, an information need can be modeled as multidimensional too (Xu & Yin, 2008). Xu and Yin (2008) further suggest that a user's novelty judgment is directed rather than undirected in the sense that what was considered novel in the last query would still be considered novel in the immediate subsequent rounds. In contrast, undirected novelty refers psychologically to users regarding anything that has not been encountered before as novel; hence novelty contrasts with a document's redundancy in terms of documents inspected in the past (Yang, Zhang, Carbonell & Jin, 2002; Zhang, Callan, & Minka, 2002). Based on these hypotheses of user behavior, an IR system could use separate data structures to model topicality and novelty of users' information need (Xu & Yin, 2008). However, Xu and Yin (2008) did not empirically demonstrate that a multidimensional model of information need is better than one that is uni-dimensional.

2.2 Single Vector Presentation of Information Need

Why is a single data structure, based on the assumption of uni-dimensional information need, ineffective to represent users' information need? Traditional IR systems, such as those using the vector space model, have been criticized for failing to capture the multidimensionality, subjectivity and dynamics of users' relevance judgment (Borlund, 2003; Cosijn & Ingwersen, 2000; Saracevic, 1975). Such criticisms may be partially overstated, as IR systems do offer methods to capture subjectivity and dynamics -- the most popular method being the use of manual relevance feedback (Rocchio, 1971) whereby users read and indicate which documents are relevant for their queries and such evaluations are used to augment their queries in the next round. However, the criticism is correct in pointing out the inability of most relevance feedback systems to capture the multidimensionality of relevance judgment. Typically, a single vector is used to represent a user's information need in a relevance feedback system. The user's information need starts with a query vector. Over rounds of relevance feedback, the query vector is updated by document vectors to produce a new and hopefully better description of the user's information need. This vector is termed the *user profile* here. This user profile, over time, can capture the dynamics of user's information need. The key limitation of representing a user profile with a single vector is that it biases the retrieval result towards those on-topic documents without an adequate consideration for novelty. This point is illustrated with the vector space model in the following section.

Table 1: An example based on the vector space model

Documents (relevance score)	**Mobile**	**Phone**	**Threat**
Q	1	1	1
D_1 (1)	2	2	3
D_2 (0.5)	2	3	2
D_3 (0.5)	2	3	2
User profile, *P*	5	6	6

For example, consider the following scenario: a user wants to find out if mobile phone usage will affect his son's health. Assume that he already knows radiation emitted from mobile phones may be a potential health threat, but he is not sure if this radiation poses severe health risks to his son. Therefore, the topic is mobile phone radiation and the novel information to search for concerns the threat of radiation (e.g., how harmful it is, what is the harm caused by radiation, how to protect oneself from such threats, etc.). He goes to an online search engine and submits a query Q, which consists of three terms: "mobile," "phone," and "threat." Assume documents D_1 and D_2 are returned as shown in Table 1. In his query Q, "mobile" and "phone" are necessary terms to ensure topicality (hence topicality terms) whereas "threat" represents the aspect of new knowledge to be retrieved (hence novelty term). Therefore, documents detailing "threat" should be considered more novel. Assume the term frequency for D_1 and D_2 are distributed as in Table 1, ignoring inverse document frequency used in the vector space model for simplicity, and that D_1 and D_2 produce the same cosine score with the query, although D_1 seems to address the "threat" aspect in greater detail. Therefore, a novel document is not necessarily rewarded in the vector space model.

Now, consider a retrieval system that supports relevance feedback. Assume the user assigns relevance scores to three documents D_1, D_2 and D_3 that are retrieved by the system as in Table 1. D_1 has a higher relevance score for its favorable term distribution. The updated user profile *P* is then formed by a simple weighted sum $P = Q+1\times D_1+0.5\times D_2+0.5\times D_3$. However, even with the weights, the updated profile is still unable to differentiate between the importance of "phone" and "threat" as both end up with the same score, although "threat" represents the actual aspect of interest. The inability of such a profile adjustment procedure to capture a user's information need seems contrived. However, this situation is common. First, users' queries tend to contain more general topicality terms than novelty terms. This is because they do not know how to pick the right term for the unknown information in query construction. Second, queries with general topicality terms tend to retrieve documents that match the general topic rather than the desired yet-unknown subtopic. Because there are more documents that match the general topic among the retrieved documents than those that match the specific subtopic, in the long run, the updating strategy outlined above makes the topicality terms stand out, but blunts the novelty terms because they appear only in a small set of documents. Ultimately, the user profile converges to the general topic rather than the desired aspect which is novel to the user.

This occurs because topicality is more stable in a search session, and is strengthened in successive rounds of retrieval. Novelty, however, is fluid and focused, and therefore is not always strengthened. Users' tendency to use general terms to specify queries is another reason that favors topicality more than novelty.

The above reasoning leads us to propose that relevance feedback systems are biased towards topicality. The hypothesis can be tested by checking whether a relevance feedback system and a topicality feedback system perform similarly. A topicality feedback system is an IR system that asks users to manually assign documents a score based on topicality judgment rather than relevance judgment. Hence, our first hypothesis is:

H1: A relevance feedback system and a topicality feedback system are not significantly different in performance.

2.3 Incorporating Novelty in IR

If relevance is multidimensional, it seems reasonable to represent a user's information need with two separate vectors: one representing topicality and the other representing novelty. Such vectors are called the *topicality profile* and the *novelty profile*. The next question is: How to operationalize a user's novelty profile with a vector and how to integrate topicality-based and novelty-based retrieval?

System-oriented studies propose that novelty should be defined as the amount of relevant information in the current document that is not covered by relevant documents previously read, that is, novelty is the opposite of *redundancy* (Brants et al., 2003; Kumaran & Allan, 2004; Yang et al., 2002, Zhang et al., 2002). The implicit assumption is that users can completely assimilate the information of a document they encounter, and do not have a planned direction for the next subtopic. Hence, novelty seeking is *undirected*. For example, if a user's initial document is "mobile phone threat", in the next round, a document containing "mobile phone threat driving accident" would be as novel as a document containing "mobile phone threat radiation tumor", because after removing the redundant words the two new documents have an equal amount of new information. However, the novel information in the latter matches the information need better than the former.

Following Xu and Yin (2008), this study contends that when looking for on-topic documents, users often want to read more on the current subtopic until sufficient information has been collected. Also, because users cannot fully assimilate the content of documents read previously, documents of similar content can enhance understanding or reaffirm previous understanding. Only when information for the current subtopic is sufficient would users switch to a new subtopic. Therefore, it is more reasonable to regard novelty seeking as *directed in the short term*. To operationalize novelty, this study seeks to identify the novelty terms in documents read in the last round. A novelty profile so built is termed a *directed novelty profile*.

When a directed novelty profile is built, documents in a corpus can be evaluated based on their similarity to the directed novelty profile, and a directed novelty score can be calculated to rank documents. How should the directed novelty-based document rank and topicality-based document rank be integrated to form a final ranking? Two options exist. In the first option, a topicality score and a directed novelty score can be calculated for each document and a weighted sum could be obtained as an overall relevance score (Zhai et al., 2003). In the second option, the corpus can be ranked first by topicality. Then, only the top few documents are regarded as on-topic and are retained. The retained documents can be ranked by directed novelty (Zhang et al., 2002). Essentially, the first option regards topicality and directed novelty as compensatory, while the second option regards them as non-compensatory and are steps in the decision-making process. This study regards the second approach as more reasonable based on the following reasons. First, in past behavioral research on relevance judgment, topicality has been considered the first criterion before other factors. As Froehlich (1994, p.129) highlighted, "all relevance judgments *start* with topically relevant materials (which is an appropriate first step of a system), but then diverse criteria come into play." In psychology, it has been observed that people first screen documents based on topicality before they start reading (Pirolli & Card, 1999). The empirical study of Xu and Yin (2008) posits that a stepwise decision process is more consistent with users' actual behavior than a compensatory one.

Finally, this study considers directed novelty and redundancy to be not mutually exclusive. It is still desirable to avoid redundancy, especially in the longer term. When the search for a subtopic or even the general topic is satisfied, retrieving more documents on the same topic is meaningless. Therefore, it still helps to filter out or reduce the weight of documents that are similar to documents read in a distant past. Redundancy is considered as an undirected sub-dimension of novelty in this study.

An IR system that maintains separate topicality and directed novelty profiles in a feedback process and integrates them in a stepwise way is termed a directed novelty-augmented system. If a redundancy profile is used to further enhance the system, it is termed a redundancy-augmented system. Because a multidimensional model of users' information need allows for differences in subjectivity and dynamics of topicality and novelty, it is hypothesized to be better than a uni-dimensional model. The following hypotheses are proposed:

H2. The performance of a directed novelty-augmented system is better than that of a topicality feedback system.

H3: The performance of a redundancy-augmented system is better than that of a directed novelty-augmented system.

3. Systems

To test the above hypotheses, four different systems were designed: the relevance feedback system, the topicality feedback system, the directed novelty augmented system, and the redundancy augmented system. All were built on the vector space model. The four systems differed in their assumptions of user behavior and how user profiles were constructed and maintained. Table 2 summarizes the differences among the four systems. These systems are largely based on the algorithms reported in Xu and Yin (2008). To make this study self-contained, major features of those algorithms are repeated here. A user study was then conducted to compare the perceived performance of the four systems in terms of the relevance, topicality and novelty of documents retrieved as judged by system users. In the following sections, the design of each system is explained. It is assumed that these systems are used in an interactive information retrieval context and documents come in rounds. In other words, a system retrieves a batch of documents (e.g., 10) in each round, a user evaluates them, and then the system retrieves the next batch.

3.1 Relevance Feedback System

The first system is the relevance feedback system (RFS). It is based on the traditional Rocchio feedback system (Rocchio, 1971). In the relevance feedback system, a user profile is represented with a term vector. The user's profile starts with the initial query and is subsequently adjusted according to the user's relevance evaluation of documents encountered. This profile is called the *relevance profile* P^{REL}.

The relevance profile updating strategy is also based on Rocchio's relevance feedback. In a retrieval session consisting of a number of rounds, if the user evaluates and assigns relevance scores to a set of documents returned in each round, the document vectors and relevance scores are used to update the initial profile. Only positive feedback is used in our system. Therefore:

$$P_t^{REL} = P_{t-1}^{REL} + \frac{1}{|R_t|}\sum_{d_i \in R_t} d_i REL_i$$

where t denotes the feedback round and P_t^{REL} is the relevance profile at round t. R_t is the set of documents examined at round t, $|R|$ is the number of documents in the set, d_i is a document vector based on TFIDF term weighting and REL_i is the subjective relevance score assigned by the user to document i. All documents, except the initial query, contribute equally to the updated profile if the size of R is fixed. The differential contribution of a document to the profile is determined by the product of its term weights and the user's relevance perception $RELi$.

Relevance evaluation of an unseen document j by the system is calculated using the similarity between the relevance profile and the document vector. It can be expressed as:

$$Rel(d_i) = sim(d_i, P_t^{REL})$$

Cosine similarity function is used in all our systems as it is one of the most effective measures.

3.2 Topicality Feedback System

The second system is the topicality feedback system (TFS). It is similar to the relevance feedback system except that in updating the user topicality profile (P^T), instead of using the relevance score assigned by users, the topicality score is used. The topicality profile also starts with the initial query and is subsequently updated by the documents encountered. The comparison of RFS and TFS allows us to test the first hypothesis.

3.3 Directed Novelty-Augmented System

In a directed novelty-augmented system (DNAS), a user's profile consists of two sub-profiles: the topicality profile and the directed novelty profile. Different from the topicality profile, a novelty profile does not start with the user's initial query because what the user is pursuing as a subtopic is at that point not yet evident. Rather, the novelty profile is built on the user's feedback of the first round of retrieved documents. When the documents in the first round have been assigned novelty scores, simplistically speaking, the document vectors can be weight-averaged to produce a vector in a way similar to the case of relevance feedback. However, the main thrust of directed novelty perception is that users are interested in a particular subtopic of a more general search task. Therefore, there is a need to differentiate subtopics embedded in the retrieved document set. The documents with high novelty scores should be identified, and the terms that are representative of the novelty aspect, rather than the more general topicality terms, should be identified and included in the directed novelty profile.

The probabilistic F4 measure proposed by Robertson and Spark-Jones (1976) was adopted. The F4 measure of a term t_j is the ratio of the odds that a relevant document contains the term and the odds that an irrelevant document contains it. Because F4 assigns a weight to a term based on its relative probability in relevant and irrelevant documents, the F4 measure can be regarded as a local classification measure. The idea of an F4 measure could be adapted to differentiate novel and non-novel documents. The weight of t_j, which indicates how effectively the term can differentiate novel and non-novel documents, is:

$$w_{t_j} = \log\frac{P(t_j \mid N)}{1-P(t_j \mid N)} - \log\frac{P(t_j \mid \overline{N})}{1-P(t_j \mid \overline{N})} = \log\frac{r_j/(R-r_j)}{(n_j-r_j)/(S-n_j-R+r_j)}$$

where $P(t_j \mid N)$ is the probability of novel documents containing t_j, $P(t_j \mid \overline{N})$ is the probability of non-novel documents containing t_j, r_j is the count of novel documents containing term t_j, R is the count of all novel documents, n_j is the count of all documents containing t_j, and S is the total number of documents in the set.

The original F4 measure assumes binary document classification. An adaption is needed to cater to partially novel documents. The adaption is to allow a partially novel document to contribute to both the novel and the non-novel sets according to its novelty score. For example, if the maximum possible novelty score is 7, a document with a novelty score of 6 would contribute 6/7 of a document to the novel document set and 1/7 of a document to the non-novel document set. This adaption makes r_j the sum of novelty 'fractions' of documents containing term t_j, R the sum of novelty scores for all documents regardless of the terms contained, n_j the number of documents containing term t_j, and S the total number of documents in the set.

One disadvantage of using the F4 measure is that it identifies only novel terms regardless of whether they are on- or off-topic. In an IR system, novelty is meaningful only if topicality is ensured first, or is at least partially present (Xu & Chen, 2006). Therefore, the F4 term weight is multiplied with the corresponding weight of the term in the topicality:

$$w_{t_j} = w_{t_j}^{F4} \times w_{t_j}^{P^T}$$

Because off-topic terms are now discounted due to their low weights in a topicality profile, the new term weight can be regarded as a topicality-conditioned directed novelty weight.

Based on the novelty terms identified and weighed at round t, a set of terms with directed novelty weights forms the *local directed novelty profile* P^N_{Rt} based on a document set (R_t). This local profile is then used to build and update a *global directed novelty profile* P_t^N using the following formula:

$$P_t^N = (1-\beta)P_{t-1}^N + \beta P_{R_t}^N$$

where *ß* is an updating parameter. *ß* is arbitrarily set to 0.8, which means that the novelty profile is largely based on the last round of feedback. While arbitrary, as reported by Xu and Yin (2008), the system performance is not sensitive to the parameter in the range of approximately 0.8. The global directed novelty profile is the operationalization of the *directed novelty profile*. With a global directed novelty profile, the novelty of unseen documents can be calculated with the simple cosine score between the document vector and the profile.

Finally, topicality and novelty evaluations of a document in a corpus need to be integrated. As we argued previously, a stepwise integration makes more sense from the user behavior perspective. Thus, the relevance evaluation of a document is defined as:

$$Rel(d_i) = \begin{cases} 0, & \text{if } sim(d_i, P^T) < s* \\ sim(d_i, P^N), & \text{if } sim(d_i, P^T) > s* \end{cases}$$

where $s*$ is a topicality cutoff value. $s*$ is set as the topicality level of the 20th document, that is, the top 20 documents are considered as on-topic. Again, 20 is an arbitrary number. Spink et al. (2000) observed that most users only browse the first 20 returned documents. Documents are then re-ranked by novelty.

3.4 Redundancy Augmented System

It is possible to further enhance a directed novelty augmented system with redundancy avoidance, leading to the redundancy augmented system (RAS). First, a redundancy profile P^{RD} needs to be established. The idea of the maximum marginal relevance (MMR) model as proposed by Carbonell and Goldstein (1998) is adapted for this study. In this model, the redundancy profile is defined as a collection of previously read documents. Newly evaluated documents are added to the profile at each round. A "novel" document is defined as one that is dissimilar to those in the redundancy profile. Mathematically, let R be the set of documents seen before, d_i be a new document to be evaluated, the redundancy score of d_i, $Rd(d_i|R)$ is defined as:

$$Rd(d_i|R) = max_{dj \in R}\, sim(d_i, d_j).$$

Similar to directed novelty, the redundancy calculation formula suggests that redundancy is not applicable to the first round of retrieval when no documents have been read. It is only after the first round of retrieval that we can place the read documents into a redundancy profile to evaluate the subsequent round. However, if previously read documents are used to evaluate the redundancy of the round that immediately follows, as many previous studies did (Carbonell & Goldstein, 1998; Zhai et al., 2003), this practice conflicts with the directed novelty assumption. To give sufficient time for a subtopic to be satisfied, at a specific round of retrieval, document redundancy is evaluated based on the redundancy profile built k (k>1) rounds ago. Because there is no prior study suggesting how large k should be, k is arbitrarily set at 2 in this study. In other words, before round 3, no redundancy checking occurs. However, in round 3, documents read in round 1 will be used to evaluate the redundancy of new documents to be retrieved. For new documents in round 4, documents read in round 1 and 2 will be used for redundancy checking.

Following that, redundancy evaluation needs to be integrated with topicality and directed novelty evaluations. Similar to the directed novelty augmented system, documents are first ranked by topicality. The top $s*$ (i.e., 20) documents are then ranked by a combined score of directed novelty and redundancy rankings. The basic idea is to penalize those with a high redundancy score. The redundancy score is subtracted from the directed novelty score, so that those documents that are similar to well-studied subtopics receive a low evaluation. However, as directed novelty and redundancy are calculated with different formulas; direct subtraction of scores could be misleading. Therefore, it is necessary to normalize the redundancy score and novelty score. Particularly, for a document j in one round,

$$Rd_j^* = \frac{Rd_j - Rd_{\min}}{Rd_{\max} - Rd_{\min}} \qquad N_j^* = \frac{N_j - N_{\min}}{N_{\max} - N_{\min}}$$

where Rd^*_j is the normalized redundancy score, Rd_j is the raw redundancy score, Rd_{max} and Rd_{min} are maximum and minimum redundancy scores in the $s*$ documents. Similarly, N^*_j is the normalized novelty score, N_j is the raw novelty score, and N_{max} and N_{min} are maximum and minimum directed novelty scores in the $s*$ documents. The final relevance score of a document is:

$$Rel(d_i) = \begin{cases} 0, & \text{if } sim(d_i, P^T) < s* \\ N_j^* - Rd_j^*, & \text{if } sim(d_i, P^T) > s* \end{cases}$$

4. User Study

4.1 Search Task and Search Procedure

A user study was conducted with the above systems. Testing with a standard TREC dataset was not adequate for our purposes because our systems involved subjective topicality, novelty, and relevance judgment by users. The same study design of Xu and Yin (2008), including search task and corpus, was followed. Again, for the sake of self-containment, the study design is repeated here.

University students were recruited as participants to test our systems. The search topic for participants was "mobile phone radiation and health" which was relevant to the local population with a 97.8% mobile phone service subscription rate. The search task was described as follows:

> Assume you are taking a health education class. The final examination, which accounts for 50% of the total grade, is to **search** for and **study** online documents on "**mobile phone radiation and health**." The relevance of a document depends on how much it addresses the following issues:
>
> 1. Does the use of a mobile phone pose radiation threats to the user's health?
> 2. Why is there such/or no such radiation threat to health?
> 3. What is the proper way to use a mobile phone to protect your health from radiation?
>
> You need to search for documents with the search engine provided. After a list of documents (**60 documents in six pages**) is returned by the search engine, please read each document in order, and evaluate each document in terms of whether it is **on-topic, novel (provides new knowledge)** to you, and of an **overall usefulness**. **You also will be asked to take a short online examination on the topic of "mobile phone radiation and health" after the search.**

When participants came to our research laboratory, they were first introduced to the search task. Participants were then asked to evaluate their knowledge of the search topic on a printed survey form with three subjective questions on a seven-point scale (e.g., "I consider myself an expert in this topic area," 1=strongly disagree, 7=strongly agree). Next, they were given the definitions of topicality, novelty and relevance, which were also stated in the instructions:

> A document is **on-topic** if it talks about something related to your information need. However, an on-topic document can have as little or as much content related to your information need. A document is **novel** if it provides knowledge that is **new** to you. A document is **overall useful** if it makes a major contribution to your information need, you expect it to contribute substantially to your quiz grade, and you try to memorize its content.

The term "usefulness" was used instead of "relevance" because participants were unlikely to be aware of the academic definition of relevance and the notion of relevance is often known as "usefulness" by users in a search task (Fitzgerald & Galloway, 2001). Terms like "usefulness," "contribution to quiz" and "try to memorize its content" are also consistent with the situational and action-oriented nature of relevance (Xu & Chen, 2006). Regardless of the system they were assigned to, participants were asked to evaluate topicality, novelty and usefulness based on an eight-point scale (0=totally off topic, totally non-novel, or useless, 7= very on-topic, very novel, or very useful) as suggested in past literature (Xu & Chen, 2006).

Participants were randomly assigned to a system and all systems had identical interfaces. Then, after some hands-on experience with the assigned system based on an irrelevant search task, the search process started. However, for the search process, the search query was pre-specified in the query input box and the participants were told that they did not need to revise the search query (the input box was not editable). The search query was "mobile phone health." The initial query was kept the same in order to reduce system performance variance arising from different initial queries, so that the impact of search algorithms on system performance was not confounded by query differences. The participants were told that their task was only to evaluate the documents returned in terms of topicality, novelty, and usefulness. They were asked to evaluate 60 documents in 6 rounds.

When the initial query was submitted, the top 10 of the most relevant documents were returned and listed in one page showing only the titles and three evaluation boxes for the three factors (i.e., topicality, novelty, and relevance). Participants were asked to read and evaluate all 10 documents in terms of the three factors. The evaluations were recorded by the server. Using document evaluations, the four systems updated participants' user profiles based on different profile-updating strategies which are consistent with the assumptions of users' relevance judgment behavior. Finally, the systems returned another 10 documents with no repetition of previously used documents.

After the search, participants took a quiz of 10 multiple choice questions on the search task. They were also asked to evaluate their knowledge on the subject again with the same set of questions used in the pre-search survey. The search process lasted for 1.5-2 hours. Participants were paid for their participation. Moreover, the one with the best score in the after-search quiz won a prize of S$50. The use of the quiz and incentives was to encourage participants to become more involved in the search so that their judgments of topicality, novelty, and relevance were more reliable. Since the quiz was of no theoretical significance to the study, it will not be further discussed.

4.2 Testing Corpus

The corpus for the search task was collected from Google.com, using queries like "mobile phone health", "mobile phone radiation", "mobile phone safety," and "mobile phone safety precautions." Web pages linked from the first 20 results pages were downloaded and examined; duplicates and navigational pages were removed. In the end, 295 documents were kept for our experiment. Furthermore, these documents were assumed to be sufficient in satisfying users' information needs on the topic. These documents were then pre-processed: All HTML tags and irrelevant information such as headers, menus, footers and advertisements were removed. Only the main text of the document was left intact.

5. Results

There were 85 participants in our experiment -- 35 females and 50 males. Their average age was 20 and they were experienced users of search engines with an average of 5.43 years of experience. The number of participants for RFS, TFS, DNAS and RAS was 19, 24, 22 and 20 respectively. The slight imbalance in group sizes was due to the randomization when assigning a participant to a system. Participants also were experienced mobile phone users, with an average usage of 3.9 years. Learning about the search topic occurred during the experiment along the seven-point scale; participants began with an average score of 3.34 for their self-evaluated knowledge level and this score rose to 5.1 after the experiment, which was statistically significant ($p<.001$). The improvement in knowledge assured us that novelty played a role in the information retrieval process.

To compare the performance of systems, each participant's perception of topicality, novelty, and relevance of documents in each round was averaged over documents. Then, each evaluation was averaged over all participants for a system.

<<Insert Figure 1 about here>>

Figure 1. Performance of the four systems

Figure 1 illustrates the average relevance, topicality and novelty perceptions of the four systems over six rounds. First, there was a minor difference in the round 1 evaluation. It is to be noted that because the initial query was the same for all systems, and feedback came in only after the first round, the documents returned in the first round should not differ. Theoretically, the evaluation of the first round of documents should be very close across all the systems. The minor difference across the systems during the first round was the result of randomness. However, if the difference was random in nature, it should not be significant. To confirm that, an ANOVA test was conducted with each participant as the unit of analysis. Topicality, novelty, and relevance were the dependent variables; and systems, the independent variable (three dummy variables were used to represent the four systems). The results indicate the difference was insignificant for relevance ($F_{3,81}$=0.48, p=0.69), topicality ($F_{3,81}$=0.19, p=0.90) and novelty ($F_{3,81}$=0.95, p=0.42). Therefore, the initial difference was insignificant.

Each hypothesis compared the performance of the two systems on topicality, novelty and overall relevance perceptions. For the two systems to be compared, a multivariate analysis of covariance (MANCOVA) was conducted for each performance measure. MANCOVA was used because it is capable of handling correlated multiple dependent variables. In our experiment, for example, users' evaluations of systems in rounds 2-6 were compared. Evaluations of these rounds were correlated. If the performances of all four rounds were averaged, the detailed performance difference in each round would become lost. Another advantage of MANCOVA is that it allows the round 1 evaluation to be used as a covariate. It is to be noted that because there was no system difference in round 1, the round 1 evaluations reflected a user's idiosyncratic properties, such as background knowledge and leniency in giving a score. Treating round 1 evaluations as a covariate could statistically adjust their impacts on the later rounds (Kirk, 1995).

<<Insert Table 3 about here>>

Table 3a records the comparison between relevance feedback system (RFS) and the topicality feedback system (TFS). Table 3b records the marginal means of each round for the two systems. The results indicate that round 1 evaluation had a significant effect on the evaluations of the later rounds, which was expected. However, for all three performance measures, there was no significant difference between the two systems as indicated by the p-values in the "TFS or RFS" column. Although Figure 1 indicates that the topicality feedback system performs marginally better than the relevance feedback system, after adjusting for the round 1 evaluation, the difference was insignificant, in support of Hypothesis 1.

Also following the MANCOVA procedure, the directed novelty augmented system (DNAS) and the topicality feedback system (TFS) were compared. Table 4a records the results. It indicates that the directed novelty augmented system had a significant effect on both relevance ($F_{5,34}$=2.54, p=0.05) and novelty ($F_{5,34}$=7.88, p<0.01), but not on topicality.

<<Insert Table 4 about here>>

Table 4b illustrates the marginal mean performance in each round. It shows that the improvement in relevance and novelty mainly came from rounds 3 and 4. For the other rounds, although DNAS generally performed better, the difference was not significant. The generally insignificant improvement in topicality and the significant improvement in novelty suggest that the improved relevance perception was more likely to be due to the improved novelty when a directed novelty profile was used to augment the topicality profile. Therefore, Hypothesis 2 was supported.

While they were not hypothesized, it is also meaningful to conduct a post hoc comparison of DNAS and RFS. The result was similar to the comparison between DNAS and TFS: DNAS significantly outperformed RFS on relevance ($F_{5,34}$=2.54, p=0.05) and novelty ($F_{5,34}$=7.88, p<0.01), but not on topicality. Together with the comparison between TFS and RFS, the result indicates that a multidimensional model of users' information need is better than one that is uni-dimensional based on either general relevance or topicality evaluations.

Table 5 records the results for Hypothesis 3. To compare the directed novelty augmented system (DNAS) and redundancy augmented system (RAS), the round 2 evaluation also was added as a covariate, because the documents returned in the first two rounds were the same.

<<Insert Table 5 about here>>

The result indicates that there was no significant difference in the performance of the two systems regarding all three performance measures. Therefore, Hypothesis 3 was not supported; further enhancing that a directed novelty augmented system with redundancy seemed to produce no improvement. However, it should be noted that documents read two rounds previously were regarded as redundant and placed in the redundancy profile. This arbitrary aging speed could be too tight in the experimental setting, and may have caused the redundancy profile to show no improvement. Further testing is needed to see if the redundancy profile helps in a longer retrieval session or in multi-session retrievals. Finally, round 1 evaluations turned out to be insignificant while round 2 evaluations were significant. That was because round 2 was a more recent behavior to predict a user's future behavior.

6. Discussion and Conclusion

With an aim to design an IR system that can better capture users' information need, this study proposes that a multidimensional model of users' information need is better than a uni-dimensional model based on relevance or topicality feedback. Furthermore, a users' information need can be modeled with a topicality profile, a directed novelty profile, and a redundancy profile. The user study indicates that there is no significant difference between a traditional relevance feedback system and a topicality feedback system, suggesting that relevance feedback in IR is indeed biased towards topicality. In support of Hypothesis 2, the user study indicates that adding a directed novelty profile makes a significant difference; therefore a multidimensional model of users' information need is better than one that is uni-dimensional.. This is because the directed novelty profile captures what is interesting to the user at the right moment, even though the user might not be able to fully specify it in queries. Finally, further adding a redundancy profile does not improve the performance of a directed novelty augmented system, thus lending no support to Hypothesis 3.

However, the results of this study should be interpreted within its limitations. First, the sample size in the experiment is not large. It is desirable to further test these systems with a larger sample. Second, the use of a small corpus and a single search task limits the generalizability of findings. Third, because the purpose of this study is to empirically test the modeling of users' information need rather than proposing an efficient algorithm, (although the speed of the four systems were similar in a small-scale testing environment), there is a need to further improve the directed novelty augmented system in future research in order to deliver a more effective system for actual use. Fourth, the use of a short retrieval session (i.e., six rounds of document evaluation) might lead to the ineffectiveness of redundancy in further boosting system performance. Fifth, this study considered only topicality and novelty of relevance judgment. Other dimensions of relevance, namely reliability, understandability and scope (Xu & Chen, 2006), are not tested. Future research should also find ways to capture these aspects. Finally, while a few systems are proposed, none is designed to handle multiple search tasks. In actual situations, users often conduct multiple tasks and multi-topic searches.. How to capture and identify the nuances among multiple information needs is an interesting question for future research.

Despite these limitations, this study offers important implications towards our understanding of information needs. It lays a system-based empirical foundation for studies on multidimensionality of relevance by showing that a multidimensional model of information need is better than one that is uni-dimensional. . With this foundation, prior studies on topicality and novelty relationship (Xu & Yin, 2008) can be firmly justified. Moreover, it reconfirms Xu and Yin's (2008) study by showing that topicality and novelty are two important aspects of users' tacit information needs and they can be effectively integrated to boost system performance. Together with the prior study by Xu and Yin (2008), this study lays a theoretical foundation for future IR system designs to better capture the topicality and novelty aspects of users' information needs.

Acknowledgements

This study is supported by the School of Computing, National University of Singapore, research grant R253-000-057-112.

References

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1), 107-136.

Bateman, J. (1998). Changes in relevance criteria: A longitudinal study. In *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, 35, 23–32.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science*, 30(4), 205–214.

Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61–71.

Berkeley, University of California (2001). Web term document frequency form. Web Document: http://elib.cs.berkeley.edu/docfreq. Accessed on June 9, 2006.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913-925.

Boyce, B. (1982). Beyond topicality: A two stage view of relevance and the retrieval process. *Information Processing & Management*, 18(3), 105-109.

Brants, T., Chen, F., & Farahat, A. (2003). A system for new event detection. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Toronto, Canada, July 28 - August 1, 2003), ACM Press, New York, NY, 2003, 330-337.

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Melbourne, Australia, August 24-28 1998). ACM Press, New York, NY, 1998, 335-336.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management*, 36(4), 533-550.

Fidel, R. (1991). Searchers' selection of search keys: I. The selection routine. *Journal of the American Society for Information Science*, 42(7), 490–500.

Fitzgerald, M. A., & Galloway, C. (2001). Relevance judging, evaluation, and decision making in virtual library: A descriptive study. *Journal of the American Society for Information Science and Technology*, 52(12), 989-1010.

Froehlich, T. J. (1994). Relevance reconsidered: Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for Information Science*, 45, 124 –134.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 602-615.

Kirk, R. E. *Experimental Design*. Pacific Grove, CA: Books/Cole Publishing, 1995.

Kumaran, G., & Allan, J. (2004). Text classification and named entities for new event detection. *In the Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Sheffield, UK, July 25-29, 2004), ACM Press, New York, NY, 2004, 297-304.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9), 810–832.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, 106(4), 643–675.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In *ACM SIGIR Conference on Information Retrieval* (pp. 275–281). New York: ACM Press.

Robertson, S. E., & Spark-Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27, 129-146.

Rocchio, J. J. (1971). Relevance feedback in information retrieval. In Salton, G. (ed.), *The SMART Retrieval System - Experiments in Automatic Document Processing* (Chapter 14, pp. 313-323). Englewood Cliffs, NJ: Prentice-Hall.

Ruthven, I. (2003). Re-examining the potential effectiveness of interactive query expansion. In *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 213-220. New York, NY.

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion of information science. *Journal of American Society for Information Science*, 26(6), 321-343.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. *Advances in Librarianship*, 30, 3-71.

Schamber, L. (1994). Relevance and information behavior. In Williams, M. E. (ed.), *Annual Review of Information Science and Technology*, 29, 33-48.

Spink, A., Wolfram, D., Jansen, B. J. & Saracevic, T. (2000). Searching of the web: the public and their queries. *Journal of the American Society for Information Science*, 52(3), 226-234.

Vakkari, P. (2003). Task-based information search. In B. Cronin, (Ed.), *Annual Review of Information Science and Technology* (Vol. 37), *Information Today*, Medford, NJ, 2003, 413-464.

Xie, H. (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science*, 51(9), 841-857.

Xu, Y. (2007). The dynamics of interactive information retrieval behavior Part I: An activity theory perspective, *Journal of the American Society for Information Science and Technology*, 58(7), 958-970.

Xu, Y., & Chen, Z. (2006). Relevance Judgment – What Do Information Users Consider beyond Topicality? *Journal of the American Society for Information Science and Technology*, 57(7), 961 – 973.

Xu, Y., & Liu, C. (2007). The dynamics of interactive information retrieval Part II: An empirical study from the activity theory perspective, *Journal of the American Society for Information Science and Technology*, 58(7) 987-998.

Xu, Y., & Yin, H. (2008). Topicality and novelty in interactive information retrieval. *Journal of the American Society for Information Science and Technology*, 59(1), 201-215..

Yang, Y., Zhang, J., Carbonell, J., & Jin, C. (2002). Topic-conditioned novelty detection. In *the ACM International Conference on Knowledge Discovery and Data Mining* (Edmonton, Alberta, Canada 2002). ACM Press, New York, NY, 2002, 688-693.

Zhai, C., Cohen, W.W., & Lafferty, J. (2003). Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Toronto, Canada, July 28 - August 1, 2003), ACM Press, New York, NY, 2003, 10-17.

Zhang, Y., Callan, J., & Minka, T. (2002). Novelty and redundancy detection in adaptive filtering. In the *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tampere, Finland, August 11-15, 2002). ACM Press, New York, NY, 2002, 81-88.

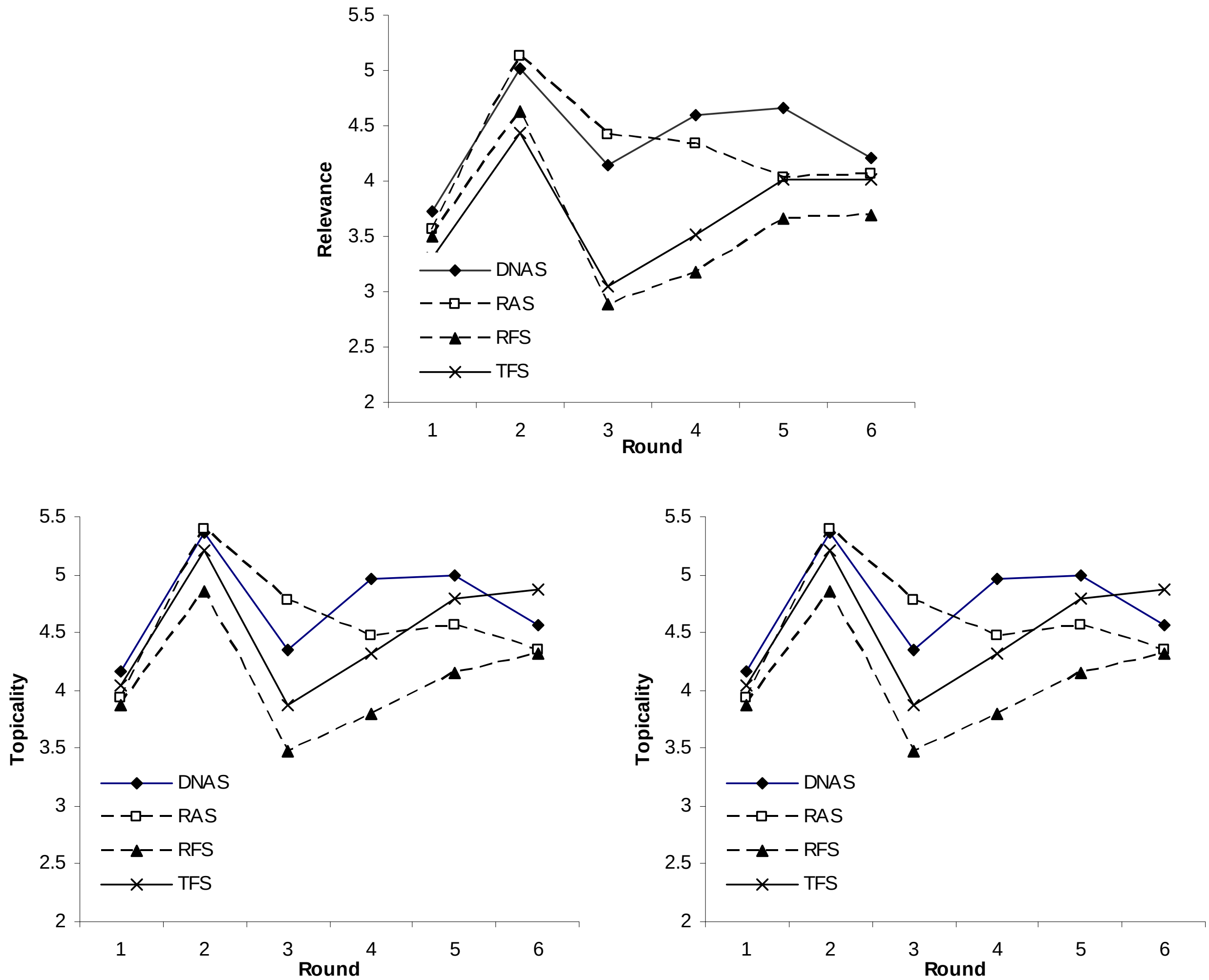

Figure 1. Performance of the four systems

Table 2. Comparison of the Four Systems

	Relevance feedback system	**Topicality feedback system**	**Directed novelty augmented system**	**Redundancy augmented system**
Information need dimensionality	uni-dimensional	uni-dimensional	multidimensional	multidimensional
Dimensions of information need	overall relevance	topicality	topicality, directed novelty	topicality, directed novelty, redundancy
Dynamics	dynamic	dynamic	dynamic	dynamic
User profile	one vector representing the overall relevance	one vector representing topicality of information need	one vector representing topicality, the other representing directed novelty	one vector representing topicality, one representing directed novelty, and one representing redundancy
Profile updating strategy	manual relevance feedback	manual topicality feedback	manual topicality feedback and F4 for novelty	manual topicality feedback, F4 for novelty, and MMR for redundancy
Relevance judgment	relevance judgment based on a monolithic overall relevance evaluation	relevance judgment based on topicality evaluation	stepwise judgment with topicality and directed novelty	stepwise judgment with topicality, directed novelty in a short term, and redundancy avoidance in a long term
Ranking of documents according to the above assumptions	documents ranked by their match to the relevance profile	documents ranked by their match to the topicality profile	documents ranked by topicality first, the top a few documents are then ranked by directed novelty	documents ranked by topicality first, the top a few documents are then ranked by directed novelty adjusted for redundancy

Table 3. MANCOVA for Hypothesis 1

a) Significance test.

	Intercept	Round 1	TFS or RFS	
H1	$F_{5,36}$	$F_{5,36}$	$F_{5,36}$	**p**
Relevance	2.77[a]	6.59[b]	0.56	0.73
Topicality	3.93[b]	11.45[b]	0.40	0.85
Novelty	6.09[b]	18.50[b]	0.52	0.76

[a] $p<0.05$, [b] $p<0.01$.

b) Topicality, novelty and relevance over rounds

Round	Systems	Topicality		Novelty		Relevance	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
2	TFS	4.94	0.25	4.14	0.21	4.53	0.25
	RFS	5.14	0.22	4.42	0.19	4.51	0.23
3	TFS	3.54	0.28	2.58	0.26	2.83	0.27
	RFS	3.81	0.25	3.05	0.23	3.10	0.24
4	TFS	3.87	0.30	2.95	0.33	3.10	0.32
	RFS	4.27	0.27	3.21	0.29	3.58	0.28
5	TFS	4.22	0.28	3.57	0.33	3.58	0.32
	RFS	4.74	0.25	4.02	0.30	4.07	0.28
6	TFS	4.37	0.31	3.57	0.33	3.64	0.33
	RFS	4.83	0.27	3.82	0.29	4.07	0.30

Table 4. MANCOVA for Hypothesis 2

a) Significance test.

	Intercept	Round 1	DNAS or TFS	
H2	$F_{5,34}$	$F_{5,34}$	$F_{5,34}$	**p**
Relevance	1.92	7.01[b]	2.54[a]	0.05
Topicality	3.17[a]	6.08[b]	1.57	0.20
Novelty	1.44	10.63[b]	7.88[b]	0.00

[a] $p<0.05$, [b] $p<0.01$.

b) Topicality, novelty and relevance over rounds

Round	Systems	Topicality		Novelty		Relevance	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
2	TFS	4.97	0.30	4.35	0.24	4.74	0.30
	DNAS	5.26	0.28	4.15	0.22	4.92	0.28
3	TFS	3.57	0.32	2.75	0.27	2.97	0.33
	DNAS	4.26	0.30	3.65[a]	0.25	4.08[a]	0.31
4	TFS	3.89	0.38	3.12	0.33	3.26	0.37
	DNAS	4.88	0.35	3.96[a]	0.31	4.53[a]	0.35
5	TFS	4.26	0.35	3.74	0.32	3.76	0.34
	DNAS	4.90	0.32	3.99	0.30	4.58	0.32
6	TFS	4.40	0.30	3.75	0.28	3.78	0.31
	DNAS	4.49	0.28	3.45	0.26	4.14	0.29

[a] $p<0.05$.

Table 5. MANCOVA for Hypothesis 3

	Intercept	**Round 1**	**Round 2**	**DNAS or RAS**	
H3	$\mathbf{F_{5,35}}$	$\mathbf{F_{5,35}}$	$\mathbf{F_{5,35}}$	$\mathbf{F_{5,35}}$	**p**
Relevance	0.92	1.91	12.55[b]	1.27	0.30
Topicality	1.04	0.98	9.21[b]	1.87	0.14
Novelty	0.95	1.46	2.93[a]	0.19	0.94

[a] $p<0.05$, [b] $p<0.01$.

b) Topicality, novelty and relevance over rounds

Round	**Systems**	**Topicality**		**Novelty**		**Relevance**	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
3	RAS	4.39	0.25	4.75	0.25	4.07	0.29
	DNAS	4.17	0.24	4.36	0.24	3.97	0.28
4	RAS	4.29	0.24	4.45	0.26	4.34	0.27
	DNAS	4.64	0.23	4.98	0.25	4.30	0.26
5	RAS	4.02	0.26	4.57	0.27	4.13	0.29
	DNAS	4.67	0.25	4.98	0.26	4.33	0.27
6	RAS	4.05	0.24	4.35	0.26	3.77	0.27
	DNAS	4.23	0.23	4.56	0.24	3.76	0.26