

December 2004

# Organizational Knowledge Distribution: An Experimental Evaluation

Surendra Sarnikar  
*University of Arizona*

Leon Zhao  
*University of Arizona*

Akhil Kumar  
*Penn State University*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2004>

---

## Recommended Citation

Sarnikar, Surendra; Zhao, Leon; and Kumar, Akhil, "Organizational Knowledge Distribution: An Experimental Evaluation" (2004).  
*AMCIS 2004 Proceedings*. 276.  
<http://aisel.aisnet.org/amcis2004/276>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Organizational Knowledge Distribution: An Experimental Evaluation

**Surendra Sarnikar**  
Department of MIS  
University of Arizona  
sarnikar@eller.arizona.edu

**J. Leon Zhao**  
Department of MIS  
University of Arizona  
lzhao@eller.arizona.edu

**Akhil Kumar**  
Smeal College of Business  
Penn State University  
AkhilKumar@psu.edu

## ABSTRACT

Organizations need to leverage their knowledge assets by making them available to their members in an efficient and effective manner. This can be done either by distributing useful codified knowledge to the users as it becomes available or by having users retrieve it according to their needs. We call the first approach knowledge distribution, and the second approach as knowledge retrieval. This paper describes the design, development and validation of a knowledge distribution system designed to facilitate efficient distribution of relevant knowledge to interested users in an organization. The system implements and evaluates the dynamic grouping technique proposed by Zhao, Kumar and Stohr. Dynamic grouping is based on an organizational concept space consisting of user profiles to capture user interests and a network of terms that captures the relationships between the concepts in a domain. Our preliminary result indicates that organizational concept space can improve the precision and recall significantly under certain conditions.

## KEYWORDS

Knowledge distribution, information distribution, information filtering

## INTRODUCTION

We define information as meaningful textual information placed within a context and contained in messages, emails and documents. The potential of information is realized as knowledge, when it is matched with an appropriate user who values that information (Zack, 1999). We have used knowledge and information interchangeably in this paper as the word information is more appropriate in some contexts than knowledge. Matching information requirements with available information is an important problem for large organizations (Davenport and Prusak, 1997). Knowledge distribution mechanisms in organizations aim to distribute relevant knowledge to the appropriate users in a timely manner. The most commonly used mechanism to distribute relevant information to users is through mailing lists (Zhao, Kumar and Stohr, 2001). However, by following a plain mailing list approach, organizations risk either flooding the mailboxes of its members with irrelevant information or not delivering relevant information to members who are not subscribed to the list. New techniques designed to facilitate efficient distribution of information to interested users involve user profiling to capture user interests and hierarchical concept spaces or a domain specific network of terms that captures the relationships between different concepts in a domain.

## PREVIOUS WORK

Information distribution systems can be broadly classified in to collaborative systems and content based systems. Collaborative systems make decisions based on ratings supplied by a group of people, and is based on the assumption that similarity of historical preferences between users is a predictor of future interests of an individual user. Collaborative filtering techniques have been widely used for information filtering in usenet groups (Goldberg, Nichols, Oki and Terry, 1992; Resnick, Iacovou, Sushak, Bergstrom and Riedl, 1994). However collaborative systems require a critical mass of users with common interests to be efficient (Oard and Marchionini, 1996). Content based techniques rely on user profiles and information contained in the document to make decision on distributing the document. Most commonly used content based techniques for information filtering at a recent text retrieval conference include machine learning, rule based and vector space

based models (Robertson and Soboroff, 2002). Previous literature has proposed that user profiling can be used to filter messages and identify messages of interest to the users (Foltz and Dumais, 1992; Kindo, Yoshida, Morimoto and Watanabe, 1997; Stadnyk, and Kass, 1992). A variety of techniques to generate and maintain user profiles have been developed (Kuflik and Shoval, 2000). User profiles can be modified over time to reflect changes in user interests (Lam, Mukhopadhyay, Mustafa and Palakal, 1996). Also, techniques exist to automatically create domain specific conceptual networks from a set of documents by statistical co-occurrence techniques (Chen, Schatz, Ng, Martinez, Kirchoff and Lin, 1996) or by analyzing subsumption relationships (Sanderson and Croft, 1999). Zhao, Kumar and Stohr (2000) combine user profiling with hierarchical concept spaces to form an organizational concept space that can aid in efficient distribution of information to the members of the organization. Such a technique can also be integrated into an organization's workflow, enabling information distribution to become a routine organizational process (Zhao, Kumar and Stohr 2001).

In this paper, we describe the implementation of the knowledge distribution mechanism described in (Zhao, Kumar and Stohr, 2001) and present an analysis of the preliminary experimental results. We first provide a brief description of the model implemented, then outline the experimental procedure, and finally present the experiment results. A probabilistic model of recall under various practical conditions is also provided before we conclude the paper.

### ORGANIZATIONAL CONCEPT SPACE

Organizational concept space extends conceptual clustering techniques by integrating organizational information with concept hierarchies. An organizational concept space consists of an interest matrix and a similarity network (Zhao, Kumar, and Stohr, 2001). The interest matrix is a two dimensional matrix with users along one dimension and topics along another dimension. The entries in the matrix specify the interest of a particular user in a particular topic. A similarity network consists of a hierarchical network of concepts. Similar concepts and synonyms are clustered together in a similarity set. Parent-child relationships between different similarity sets are defined resulting in a network of related sets, called similarity network. A membership value describes the degree of membership of a particular topic to a similarity set, and an association value describes the strength of association between related similarity sets. A simplified version of the similarity network model, where association and membership values were assumed to be uniform was used in this study.

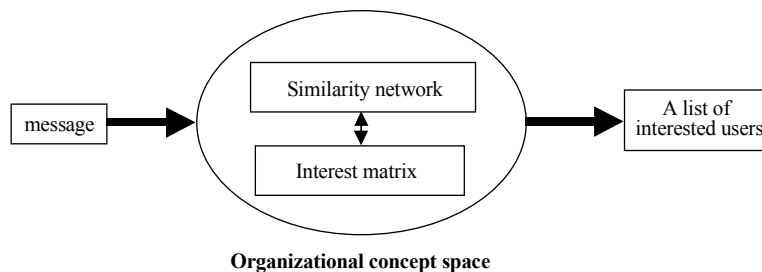


Figure 1. The organizational concept space approach (adopted from Zhao, Kumar and Stohr, 2000).

Given a message and an organizational concept space, it is possible to identify the users who would be interested in that message. The process involves identifying the set of concepts representative of the message and extending this set with related concepts that are determined with the help of a similarity network. The extent to which the set is expanded can be controlled by matching level. A low matching level selects concepts closest to the original concepts, while a higher matching level selects concepts that are further away from the original concept in the similarity network. The expanded set of concepts can then be super imposed on the interest matrix to identify interested users.

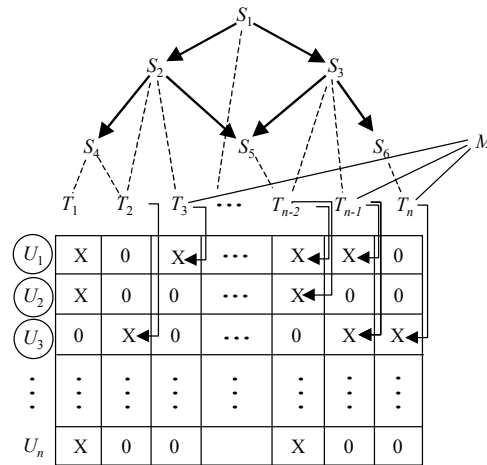


Figure 2. Extended matching via interest matrix and similarity network (adopted from Zhao, Kumar and Stohr, 2000).

A diagrammatic representation of the matching process is shown in figure 2. The X marks represent non zero interest values in the interest matrix. The concepts  $T_i$  extracted from message  $M$  are extended using the similarity sets  $S_i$  and the similarity network. A projection of extended concepts from the similarity network over the interest matrix reveals the set of users interested in the message.

**DATA COLLECTION METHOD**

**User Profiles**

Ten faculty members from the MIS departments of six different universities were selected for the experiment. The users were selected such that their research interest represented a wide range of specialties. The users were selected based on the availability of their research interests on their websites and the clarity of their research profile. The profiles of the users were obtained from faculty and university websites. Individual topics of interest were identified from the user profiles. The length of the profiles ranged from 15 to 60 words and varied in format from list of topics to a short paragraph describing research interests. In the case where the research profile was a short paragraph, a list of topics mentioned in it was extracted. A sample user profile is given in figure 3. The list of topics thus defined for each user ranged from 5 topics to 20 topics. Based on these topics a table of topics was created. Additional topics closely related to the list of topics were added to the topics table. The list of topics was analyzed and groups of similar topics were clubbed together in to similarity sets and parent, child relationships among them were defined, resulting in a similarity network.

A User Profile from the Source	Topics Extracted
Software evaluation and characterization, software development processes, software engineering education and practice, application of information technology to education	software evaluation software characterization, software development processes software engineering + education + practice information technology + education

Figure 3. Sample User Profile.

**Call for Papers**

The input for the system was 58 call for papers (CFP) from American and International journals and conferences, and call for book chapters. A typical call for paper includes the description of objectives of the conference, journal or a journal special issue, and a list of suggested topics. The topics for CFP’s were from a wide variety of disciplines related to information technology. The CFP were collected from the ISWORLD website. The call for papers were sequentially chosen from the list. Dead and outdated URLs were ignored. The portion of the CFP describing topics of interest was selected as the message. Contact and other information not related to the research topics were ignored. The call for papers were exhaustively studied and matched with user profiles to determine the relevance of each message to a user. A sample call for paper is shown below.

**A Sample Call for Paper**

- URL: <http://www2.cs.fau.de/GTVMTO2/SoSyM-cfp.html>

- Call for Papers:
- Software and System Modeling (SoSyM) Journal
- Special Section on Graph Transformations and Visual Modeling Techniques
- Guest Editors: Paolo Bottoni and Mark Minas

As diagrammatic notations become widespread in software engineering and visual end user environments, there is an increasing need of formal methods to precisely define the syntax and semantics of such diagrams. In particular, when visual models of systems or processes constitute executable specifications of systems, not only is a non-ambiguous specifications of their static syntax and semantics needed, but also an adequate notion of diagram dynamics. Such a notion must establish links (e.g., morphisms) which relate diagram transformations and transformations of the objects of the underlying domain. The field of Graph Grammars and Graph Transformation Systems has contributed much insight into the solution of these problems, but also other approaches (e.g., meta modeling, constraint-based and other rule-based systems), have been developed to tackle specific issues.

Following the successful workshop on Graph Transformations and Visual Modeling Techniques (<http://www2.cs.fau.de/GTVMT02/>) held in conjunction with the First International Conference on Graph Transformations, held in Barcelona in October 2002, a special section of the Software and System Modeling (SoSyM) journal (<http://www.sosym.org>) has been scheduled.

High quality papers are sought on different methodologies and approaches to problems such as diagram parsing, diagram transformation, integrated management of syntactic and semantic aspects, tool support for working with visual models. The paper focus should be on methodological aspects rather than on particular technical aspects.

Authors of papers presented at the workshop are solicited to submit revised and extended versions of their papers. Submissions related to visual modeling are welcome also from authors not previously attending the workshop. Each paper will be revised by 4 reviewers.

Please submit your papers in electronic format (.pdf or .ps.gz files) to the guest editors Paolo Bottoni ([bottoni@dsi.uniroma1.it](mailto:bottoni@dsi.uniroma1.it)) or Mark Minas ([Mark.Minas@UniBw-Muenchen.de](mailto:Mark.Minas@UniBw-Muenchen.de))

Important dates:

- 31st January 2003: Submission deadline
- 31st March 2003: Notification to Authors
- 15th June 2003: Revised version submission
- 1st August 2003: Final decision
- 1st Sept. 2003: Camera-ready version

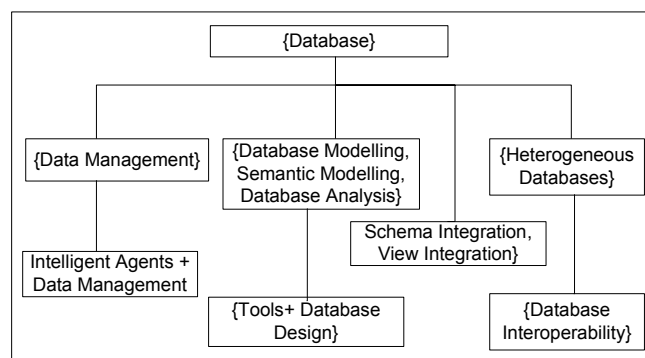


Figure 4. Sample Similarity Network

### Similarity Network

A basic similarity network was developed from topics in the subject areas of database systems, software engineering, E commerce, computer networks, artificial intelligence, human computer interaction and information systems. Additionally more topics were added when creating user profiles. An entry was created for every topic of interest listed by a user at an appropriate level and a relationship to other topics was defined. The resulting similarity network had four levels in its

hierarchy with about 130 topics. Figure 4 shows a subset of the similarity network under the topic of “database”.

**Interest Matrix**

An Interest Matrix of size 130 x 10 was created to represent the interests of the users in each of the topics. The research interests of the faculty users included areas like databases, software engineering, AI, computer networks, IT policy etc. Topics that were mentioned in the research profiles of the users were given a value of 1. Related topics were given a value of 0 to 0.9 based on the closeness of the topic to a mentioned topic of interest. The interest values were assigned depending on generality of the topic. A detailed topic describing a narrow area was assigned a high value of interest and more general topics were assigned a lower interest value. For example “intelligent agents for data management” would have a high interest value while “information systems” which is a more general topic would have a lower interest value.

	user1	user 2	user3	user6	user7
Software Development		1			
Data Management	0.7				
Distributed Computing			0.4		
E-Commerce				1	
E-Government					0.6
Information Policy				1	0.5

Figure 5. Sample Interest Matrix.

**SYSTEM ARCHITECTURE AND ALGORITHMS**

A system was implemented using java and oracle database on a Windows 2000 platform to test the organizational concept space. The algorithms were implemented using java programming language and SQL statements. Since this is a proof-of-concept implementation, we do not consider computational efficiency in this particular study. In a full-scale implementation, the SQL statements will be replaced with more efficient algorithms. The similarity network, similarity sets and the user interest matrix were implemented as relational tables. The system extracts keywords from a message, identifies similar concepts and generates users interested in the selected concepts. The major components of the system include a message parser and an OCS module. The architecture of the system is depicted in figure 6. The message parser extracted topics from the message that were also contained in a topics table. The topics table consisted of all the topics contained in the user interest matrix and the similarity network.

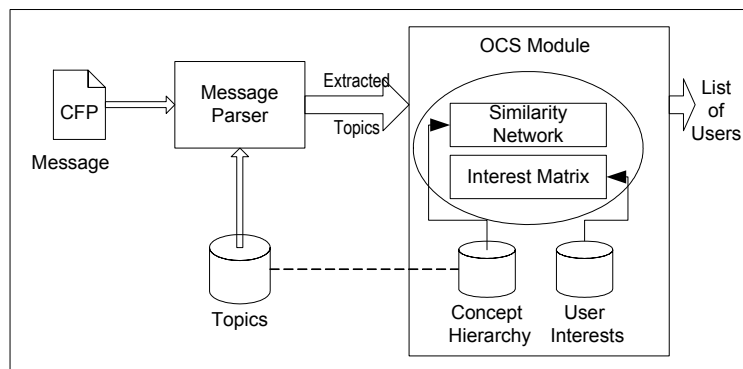


Figure 6 System Architecture.

Two different word matching algorithms, strict matching and relaxed matching were used to identify topics contained in the messages. Strict matching involved searching for whole words in the document, while relaxed matching searched for words embedded anywhere in the document. The OCS module was a software implementation of the organizational concept space. After known topics were extracted from the message, they were passed on to the OCS module. The OCS module expanded the extracted topics by including all concepts related to the extracted set of topics. The related concepts were identified by traversing the concept hierarchy up to three nodes away from the initial concept found in the message.

The core algorithm for the knowledge distribution system, simply the KDS Algorithm, is given in Figure 7. A list of users was identified based on the match level selected. Four different match levels are possible. In match-level zero or direct matching, only those users are selected who have mentioned specific interest in the exact topics mentioned in the message. In

match-level one or similarity sets matching, users interested in topics that are similar to those extracted from the message, thus contained in the same similarity set as the topics in the message, are also selected. Users selected via match-level two or similarity network matching, are those who are interested in topics that are closely related to the original topics found in the message. In match-level three, users interested in topics that were two and three nodes away from the original concepts found in the message were also selected. The above procedure was repeated for different threshold values, to study its effect on system performance. The threshold value for each user was varied from 0 to 0.9. After the first run, improvements were made to the organizational concept space to determine the effect of its quality on system performance. The changes incorporated into the system were, removal of an incorrect association in the similarity network, addition of five new topics and two synonyms and modifications to the interest values in three cells of the interest matrix. The same procedure was repeated with the improved organizational concept space.

In total, four repetitions of the above procedure, with two varieties of organizational concept space and two different word matching techniques were conducted. The users selected by the system for each matching level (direct, similarity sets, similarity network level two and similarity network level three) and threshold level was recorded. From this information, the total number of users selected, the number of users correctly selected, the number users incorrectly missed and the number of users incorrectly selected were calculated. This information was then summarized for each of the threshold and matching levels.

#### Notation

$T$  = All Topics, i.e., all topics contained in similarity network.  
 $T^*$  = Selected Topics, i.e., a list of topics selected for matching  
 $T_i$  = Individual Topic, i.e., a single selected topic  
 $N_{ij}$  = Interest in topic  $i$  for user  $j$ .  
 $U_i$  = User  $i$   
 $H_i$  = Threshold value for User  $i$   
 $T_m$  = Message Topics, i.e., topics extracted from the message.  
 $T_s$  = Similar Topics, i.e., topics selected from similarity sets where Message Topic in similarity set.  
 $T_p$  = Parent Topics, i.e., topics selected from similarity network where Message Topic is child concept.  
 $T_c$  = Child Topics, i.e., topics selected from similarity network where Message Topic is parent concept.  
 $T_b$  = Sibling Topics, i.e., topics selected from similarity network where  $T_p$  is parent concept and topic is not Message Topic.  
 $T_{gp}$  = GrandParentTopics, i.e., topics selected from similarity network where  $T_p$  is child concept.  
 $T_{gc}$  = GrandChildTopics, i.e., topics selected from similarity network where  $T_c$  is parent concept.

#### KDS Algorithm:

```

INPUT Message m;
SET matchlevel = x; where  $x \in \{0,1,2,3\}$ 
PARSE(Message) {
  for each  $T_i \in T$  { /* for each topic in topic table */
    if ( $T_i \in$  Message) { $T_m += T_i$ ; } /* if topic belongs to message, add to message topics */
  }
}

SELECTUSERS(matchlevel) {
  if (matchlevel == 0) { /* Direct Matching */
     $T^* = T_m$ ; findUser( $T^*$ );}
  if (matchlevel == 1) { /* Sets Matching */
     $T^* = T_m + T_s$ ; findUser( $T^*$ );}
  if (matchlevel == 2) { /* Network Matching 2 levels*/
     $T^* = T_m + T_s + T_p + T_c + T_b$ ; findUser( $T^*$ );}
  if (matchlevel == 3) { /* Network Matching 3 levels*/
     $T^* = T_m + T_s + T_p + T_c + T_b + T_{gp} + T_{gc}$ ; findUser( $T^*$ );}
  findUser( $T^*$ ) {
    for each  $T_i \in T^*$  {
      if( $N_{ij} > H_i$ ) { select  $U_i$ ; } }
  }
}

```



**Figure 7. The Core Algorithm of the Knowledge Distribution System.**

**EXPERIMENT RESULTS**

A summary of the results obtained is presented in table 1. The precision was calculated as the percentage of correctly selected users out of the total number of users selected. The recall was calculated as the percentage of correctly selected users among the actual number of users interested in a message summed over all the messages.  $P = \sum_{i=1}^{58} \frac{C_i}{T_i}$ ,  $R = \sum_{i=1}^{58} \frac{C_i}{A_i}$ . Where  $C_i$  is the

number of users correctly selected by the system for message  $i$ ,  $T_i$  is the total number of users selected by the system for message number  $i$  and  $A_i$  is the actual number of users interested in the message.

Note that the baseline values,  $C_i$  and  $A_i$  are computed after carefully analyzing the user interests against each CFP. Since we are dealing with relatively small data sets, we were able to derive the perfect matching results. That said, in the future, when we deal with large data sets, we will conduct user studies to derive the baseline values.

		Strict Matching		Relaxed Matching	
Enhanced Network	Best Precision	P = 0.96	R = 0.4	P = 0.87	R = 0.71
	Improvements	-0.04	0.24	0	0.4
	Best Recall	P = 0.79	R = 0.74	P = 0.67	R = 0.90
	Improvements	0.04	0.2	0.06	0.3
Simple Network	Best Precision	P = 0.92	R = 0.19	P = 0.75	R = 0.38
	Improvements	-0.08	0.05	0.06	0.13
	Best Recall	P = 0.70	R = 0.46	P = 0.54	R = 0.54
	Improvements	0	0.05	0.03	0.08
		Strict Matching		Relaxed Matching	
<ol style="list-style-type: none"> <li>1. P = Precision, R = Recall</li> <li>2. Best precision/recall is the best of different precision/recall levels achieved by the system at different threshold values under a particular network and matching algorithm.</li> <li>3. Improvements are the absolute difference in precision and recall levels achieved by the system with and without using similarity sets.</li> </ol>					

Threshold	Level 0(Direct)		Level 1(Sets)		Level 2(Network)	
	Precision	Recall	Precision	Recall	Precision	Recall
0	0.612903	0.603175	0.678571	0.904762	0.504132	0.968254
0.1	0.612903	0.603175	0.678571	0.904762	0.504132	0.968254
0.2	0.612903	0.603175	0.678571	0.904762	0.504132	0.968254
0.3	0.612903	0.603175	0.678571	0.904762	0.504132	0.968254
0.4	0.666667	0.507937	0.736111	0.84127	0.504132	0.968254
0.5	0.78125	0.396825	0.833333	0.793651	0.504132	0.968254
0.6	0.869565	0.31746	0.865385	0.714286	0.504132	0.968254
0.7	0.866667	0.206349	0.868421	0.52381	0.495726	0.920635
0.8	0.866667	0.206349	0.868421	0.52381	0.495726	0.920635
0.9	0.866667	0.206349	0.868421	0.52381	0.495726	0.920635



The effect of threshold on precision and recall was as expected. Recall decreases as the threshold increases, and there was an increase in precision with an increase in threshold value (Figure 8 and 9). An increase in recall was noticed as the level of the similarity network used increased from direct concept matching to matching parent and grandparent concepts. The change in recall was in same direction for each of the users (Figure 10). The precision and recall for individual users was calculated as  $P = \frac{C_j}{T_j}$  and  $R = \frac{C_j}{A_j}$ , where  $C_j$  is the number of messages correctly selected by the system as relevant to user  $j$ ,  $T_j$  is the total number of messages selected by the system as relevant to user  $j$  and  $A_j$  is the actual number of messages relevant to the user. We have omitted the recall graph for user 4 and 5 as no relevant documents were found for those users. However the data for both the users is reflected in the overall precision and recall for the system. The threshold value signifies the minimum interest level of a user in a topic to be used in the relevance score for identifying interested users.

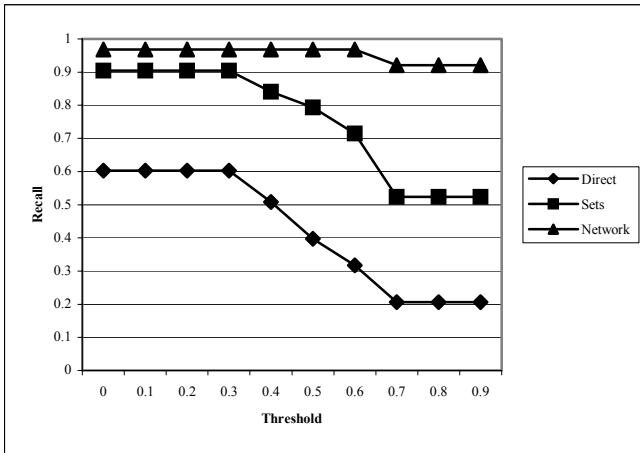


Figure 8. Impact of matching methods on recall.

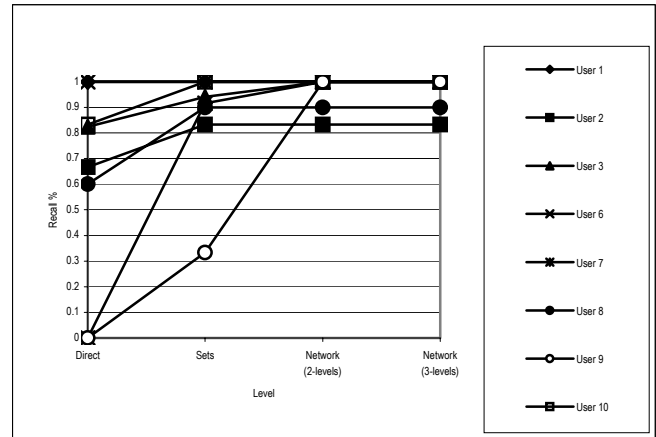


Figure 10. Impact on recalls for individual users.

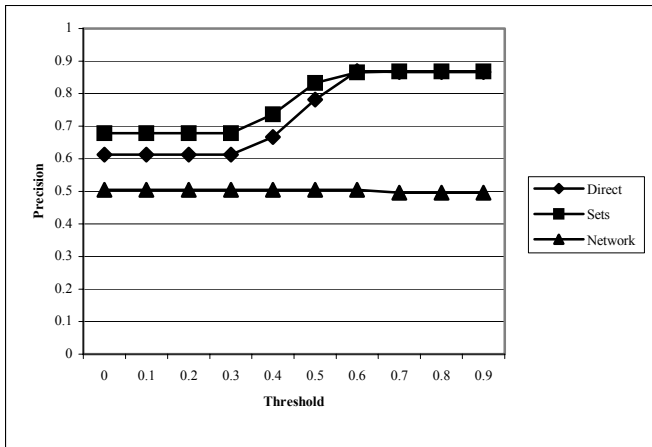


Figure 9. Impact of matching methods on precision.

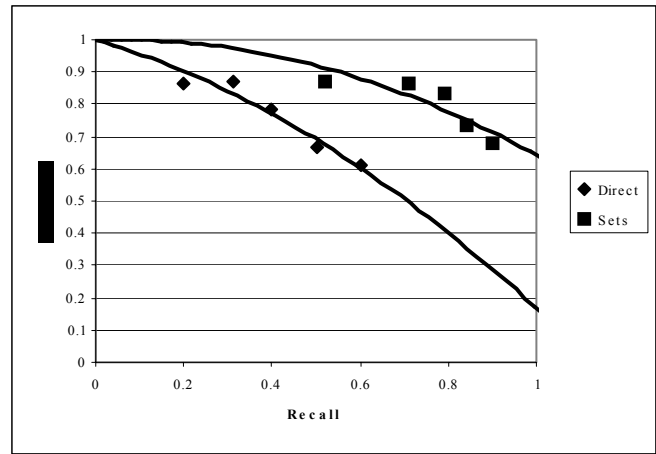


Figure 11. Precision vs. recall for two matching methods.

Analysis of the results shows that the use of similarity sets resulted in a higher precision and recall than direct matching as evidenced in the precision recall graph for direct matching and similarity set matching shown in figure 11. A precision-recall curve for similarity network matching could not be generated as no significant variation was noticed in recall or precision by varying the threshold when using similarity network matching. Also no difference in precision and recall was noticed for two level similarity network matching and three level similarity network matching. The reason for no difference is yet to be identified. Preliminary investigation of the data indicates the small size of the similarity network and insufficient modeling of the interest function due to non availability of set association values could have resulted in the negligible impact of the similarity network on system performance.

**PROBABLISTIC ANALYSIS OF PRACTICAL RECALL**

The high level of precision obtained in this experiment is due to the nature of the messages and queries and the ability of the OCS model to efficiently represent them. Both the research topics mentioned in the research profiles and the messages

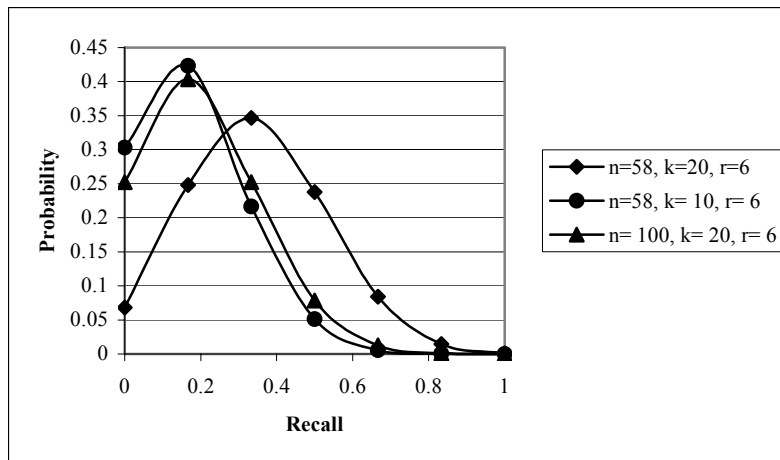
themselves are highly specific. However, because the messages are relatively short, the vocabulary problem is especially aggravated in such cases. As evidenced by the results the use of similarity sets can provide a solution to this problem.

Given a large set of documents, the probability that a user will correctly identify the most relevant documents depends on various factors including the user's cognitive ability to process a large number of messages, user availability, and time constraints, stress and user fatigue. For a set of  $n$  messages containing  $r$  messages relevant to a user, the probability that the user with an information processing capability  $k$  ( $k > r$ ), identifies all the relevant messages is given by the ratio of number of scenarios where all the relevant messages are contained in the messages processed by the user, to the number of all possible scenarios. Mathematically, the probability of 100% recall is given by  $P(R=1) = \frac{C(r,r) * C(n-r, k-r)}{C(n,k)}$ , where  $C(n,k) = \frac{n!}{k!(n-k)!}$ .

The probability of zero recall is given by  $P(R=0) = 1 - \sum_{i=0}^{r-1} \frac{C(r, r-i) * C(n-r, k-r+i)}{C(n,k)}$ . For example, in the current experiment, the

given set of documents had on an average six relevant documents per user. Assuming a user was emailed all 58 messages, and the user is able to process 20 messages before fatigue and disinterest sets in, the probability that the user will correctly identify all 6 relevant messages (100% recall) is 0.0009. For a 83% recall, the probability is 0.012.

Of course, when the number of messages is small, the practical recall should be high since most users will be able to go over all messages. However, when the number of messages becomes much larger than what the user can afford to read, the practical recall will decrease. In this case, automatic knowledge distribution will become very effective even if the recall is significantly less than 100% because it can increase the practical recall for the user by reducing the number of irrelevant messages.



**Figure 12. Impact of information process capability on practical recall.**

Figure 12 plots the probabilities of practical recall for three different scenarios. The probability distribution shifts drastically towards lower recall rates in response to either an increase in the number of documents or a decrease in the information processing capability. This indicates that probability of zero recall increases with an increase in number of documents or decrease in information processing capability. The probability of zero recall for  $n = 58$ ,  $k = 20$  and  $r = 6$  is 0.06. However when the number of documents is increased to 100, there is a 25% chance that the user will not encounter any relevant message. The expected practical recall for  $n=58$ ,  $r=6$ ,  $k = 20$  is 34% and for  $k=10$ , the expected recall is 17%. Given that all the users in this experiment are MIS researchers and all 58 messages used in this experiment are from an MIS announcement board, the recall can be associated with a mailing list system without filter. Compared to the practical recall the proposed information distribution system delivers a 200%-400% increase in utility.

### LIMITATIONS OF STUDY

The above experiment validates the organizational concept space in a specific context of distributing call-for-papers to a relatively small sample size of ten users. However, a wide range of topics within the subject area were covered in the experiment. Since the experiment does not implement all aspects of the organizational concept space, it only serves as a preliminary validation of the core concept. We plan to incorporate all the features of the OCS in our next implementation and test its efficiency in a more complex scenario.

## CONCLUSION

In this paper, we conducted an experiment to evaluate the impact of organizational concept space (OCS) on the precision and recall of a knowledge distribution algorithm in the context of distributing call-for-papers to a set of interested users. We analyzed the specific impact of various ways of using the OCS. We observed that extended concept matching using similarity sets resulted in both higher precision and recall values as compared to direct concept matching. An increase in recall was observed as the level of similarity network used increased from direct matching to two and three level extended concept matching. However, we observed no variation in the precision and recall in response to variation in threshold when using two level and three level similarity network matching. In addition to the analysis of the experimental results, we have presented the key algorithms used in implementing the organizational concept space, and a theoretical framework to help evaluate the utility of the system.

In our future research, we plan to extend the research presented in this paper in several directions, including the development of more advanced techniques for extracting topics from the messages, for generating similarity networks based domain ontologies, and for maintaining user interest matrix, particularly the creation of interest values.

## REFERENCES

1. Chen, H., Schatz, B., Ng, T., Martinez, J., Kirchhoff, A. and Lin, C. (1996) A parallel computing approach to creating engineering concept spaces for semantic retrieval: the Illinois digital library initiative project, *IEEE Trans on PAMI*, 18, 8, 771-782.
2. Davenport, T. H. and Prusak, L. (1997) *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA
3. Foltz, P.W. and Dumais, S.T. (1992) Personalized information delivery: an analysis of information filtering methods, *Communications of the ACM*, 35, 12, 51-60
4. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992) Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35, 12, 61-70
5. Kindo, T., Yoshida, H., Morimoto, T. and Watanabe, T. (1997) Adaptive personal information filtering system that organizes personal profiles automatically, *Proceedings of the International Joint Conference on Artificial Intelligence*, Nagoya, Japan
6. Kuflik, T. and Shoval, P. (2000) Generation of user profiles for information filtering - research agenda, *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, Athens, Greece
7. Lam, W., Mukhopadhyay, S., Mostafa, J. and Palakal, M. (1996) Detection of shifts in user interests for personalized information filtering, *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, Zurich, Switzerland
8. Oard, D. and Marchionini, G. (1996) A conceptual framework for text filtering. *Technical Report CS-TR3643*, University of Maryland, College Park, MD
9. Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P. and Riedl, J. (1994) GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the Conference on Computer Supported Collaborative Work*. Chapel Hill, NC.
10. Robertson, S. and Soboroff, I. (2002) The TREC 2002 Filtering Track Final Report. In E. M. Voorhees and D. K. Harman, editors, *Proceedings of the Eleventh Text Retrieval Conference*, Gaithersburg, MD.
11. Sanderson, M. and Croft, M. (1999) Deriving concept hierarchies from text, *Proceedings of the 22<sup>nd</sup> Annual International ACM SIGIR Conference on Research and Development in Information retrieval*, Berkeley, CA.
12. Stadnyk, I. and Kass, R. (1992) Modeling users' interests in information filters, *Communications of the ACM*, 35, 12, 49-50
13. Zack, M. (1999) Managing Codified Knowledge, *Sloan Management Review*, 40, 4, 45-58
14. Zhao, L., Kumar, A. and Stohr, E. (2000) Workflow centric Information Distribution through Email, *Journal of Management Information Systems*, 17, 3, 45-72
15. Zhao, L., Kumar, A. and Stohr, E. (2001) A Dynamic Grouping Technique for Distributing Codified-Knowledge in Large Organizations, *Proceedings of the 10th Workshop on Information Technology and Systems*, Brisbane, Australia