ICEB 2008 Proceedings

International Conference on Electronic Business (ICEB)

Fall 9-30-2008

Determining Personal Evolving Topic-need to Support Information Search Activities

I-Chin Wu

Determining Personal Evolving Topic-needs to Support Information Search Activities

I-Chin Wu, Fu Jen Catholic University, icwu.fju@gmail.com

Abstract

With the growing amount of information in the organizational memories of knowledge-intensive work environments, knowledge workers are suffering increasingly from information overload. Hence, an important aspect of effective knowledge delivery is supporting task-relevant knowledge by considering the characteristics of tasks and the nature of workers' search behavior in organizations. The pilot research models in the information seeking (IS) research area show that workers' information seeking activities exhibit common patterns. Based on the observations of previous studies, this work investigates the issues involved in determining the variations in task-relevant topics to support the information search process. Specifically, we provide an overview of the ISP model and theory; propose an evolving topic-needs determination method to examine the variety of a worker's information needs for topics across task-stages; and identify a worker's task-needs precisely by interactively mapping his/her information needs to the specific level of topics in the taxonomy. We have conducted an evaluation in a research institute which has implications for assisting workers who search the relevance information while conducting a long-term research project.

Keywords: Evolving topic-needs, Information filtering, Information search process, Information seeking

1. Introduction

For professional projects in knowledge-intensive domains, improving the capability of knowledge retrieval functions to provide relevant information that meets users' information needs precisely is of the utmost importance. When executing a knowledge-intensive task, a worker requires lots of explicit or tacit knowledge to support the task's execution [2][7]. In organizations, intellectual content containing valuable explicit knowledge is usually codified in an explicit form to facilitate knowledge retrieval and reuse [2][6]. Generally, a worker uses documents to understand a task and solve a specific problem. When the worker begins a task, he/she may search the organization's knowledge repository for information that will help solve the problem at hand. The worker's search behavior results from the fact that there is a gap between his/her knowledge about the task and the perceived requirements of the task. The gap is called the information need and results in information seeking activities [3][11], i.e., a series of information retrieval activities.

Most commercial information retrieval systems rely on a keyword search method as the primary retrieval mechanism. However, knowledge workers are often unable to express their information needs precisely in short query terms [5][8]. In many cases, the worker may only have a general idea about a topic and may be uncertain about the information required for the task at hand. In recent years, several studies have stressed the importance of modeling users' interests or information needs for a specific work task incrementally in terms of topics, instead of as a set of weighted keywords or meta-data. For example, Sieg et al. (2004)[10] integrated user profiles and concept hierarchies to infer users' information contexts in order to enhance the original queries. In this way, IF systems learn users' current information needs from the relevance feedback and update the model for future information filtering. Thus, most user modelling approaches use profiling techniques to analyze changes in a user's topics needs in his/her daily work life, instead of considering the user's specific topics needs for long-term work tasks.

To date, researchers in the field of Information Retrieval (IR) have focused on representations of documents for the retrieval of documents, search strategies, and assessing the relevance of retrieved documents. Comparatively little attention has been paid to users' information needs and how to support their search activities. If IR was a tool for obtaining information during a problem-solving task, a single search session would not accurately reflect the changes in information needs during the task's execution. Information Seeking (IS) involves searching for and using information for a task when a person does not have sufficient prior knowledge. Several empirical studies have observed and analyzed workers' successive searches and connected them to the task complexities, relevance judgments, and situation of the subjects during the IS process [3][4][5][11][12][13][14]. Table1 lists the characteristics of existing research models of the Information Search Process (ISP). We also show the characteristics of our research

Table 1. Information Search Process (ISP) Models

Research	Kuhlthau' s model	Vakkari' s theory	Wang & Sogergel	This research
Search Process	Task Initialization Topic Selection Prefocus Exploration Focus Formulation Information Collection Search Closure	Pre-focus Pre-focus Pre-focus Focus Formulation Post-focus Post-focus	N.A N.A N.A Selecting Reading Citing	Pre-focus Pre-focus Pre-focus Focus Formulation Post-focus Post-focus
Observations	Feelings Thoughts Actions	Search terms Operator types Search tactics	Decision Criteria and rules	Changes in stages Evolving topic-needs
Information Types	Relevant Information Pertinent Information	General background information Faceted background information Specific information	Content information Ssituational information	General topic Specific topic
Data Collection	Observations Questionnaires	Observations Questionnaires	Observations Questionnaires	Automatic system tracking Interactive with humans
Research Purpose	Understand human search process	Understand human search process	Design intelligent document selection assistant	Supply relevant and pertinent knowledge Support information search activities

model, which is based on classic models, such as Kuhlthau's model, Vakkari's task-based performance model, and Wang-Soergel's document selection and decision model of different problem stages. The Kuhlthau's search process model (1993) [5] differentiates a task into six stages with their associated characteristics. It describes the information search process from the user's perceptive as being experienced in six stages of thoughts, feelings, and actions. The objective is to observe how users locate and interpret information to form a perspective on a topic. Kuhlthau's study (1993) observed students involved in information seeking for a certain period of time, whereas Vakkari's studies [11] considered a user's information seeking activities as the execution of a task progressed (e.g. writing a proposal or completing a project). Based on Kuhlthau's model, Vakkari divided information seeking activities into pre-focus, focus-formulation, and post-focus stages [5][11][12]. Overall, a user's search activities involve forming a perspective on the topic of interest based on the derived information, and then interpreting and presenting the information in a meaningful personal ontology. Wang & Soergel (1998) [13]and Wang & White (1999)[14] proposed cognitive models of document usage during research projects conducted in 1992 and 1995, respectively. The results show that, during a research project, document usage is a decision-making process in which decisions are made at three points or stages: selecting, reading, and citing. The above studies could contribute to research on the applications of intelligent information retrieval systems, and enhance the use of knowledge retrieval functions to support task execution by professionals.

Following the trends of information seeking (IS) studies in the information retrieval (IR) domain, we now introduce our research model, as shown in Table 1, to design our information seeking and retrieval model. Our goal is to deliver relevant and pertinent information that will facilitate the execution of professional tasks. Therefore, we track changes in the user's problem stages and variations in topic-needs when performing long-term knowledge-intensive tasks. Based on the observations of previous studies (see Table 1), the research model simultaneously considers the user's problem stage and the relevant topics in the research domain, which influence the user's behavior in selecting and using documents while information needs are evolving. In this work, we propose *a topic-need variation determination method* that captures contextual information derived from a domain ontology across task-stages. Since a user may only have a general idea about topics associated with performing a task, we introduce a multi-level domain topic taxonomy to determine variations in the worker's topic needs. This structure allows us to observe and record workers' feedback behavior when they use the proposed system (i.e., their search behavior and relevance feedback on knowledge items) and then determine their information needs. A change in task-stages is inferred by analyzing the correlation between the task's temporal profile and the worker's consecutive transactions. Evolving topic-needs across problem stages can be identified by indicators of "generality" and "specificity". Both indicators have different functionalities in different task-stages, which influence the interactive topic-needs identification process. Consequently, the system can achieve effective document supply in the long-term based on the proposed topic-needs variation determination method and interactive topic identification process.

The remainder of the paper is organized as follows. The process of examining evolving topic-needs across stages is proposed in Section 2. In Section 3, we describe the method to identify the workers' evolving topic-needs interactively. The experiment design and preliminary experiment results are presented in Section 4. Finally, we present our conclusions and indicate the direction of future works in Section 5.

2. Process of Identifying Evolving Topic-Needs across Stages

We propose a *topic-need variation determination method* across task stages. To conceptualize the domain information of an organization, we require a pertinent ontology; hence we configure the task-based information as a multi-level structure. The topic taxonomy is extracted from a set of documents and expressed as a hierarchy of topics and subtopics, and the variation in topics is determined by the indicators of "generality" and "specificity" defined according to the topic taxonomy. We discuss the steps of this process below, and illustrate the overall process of knowledge supply based on topic-needs across task stages in Figure 1.

Tracking user feedback behaviour patterns: To gather data about workers' search behavior, i.e., usage information, we observe and record workers' feedback behavior when they use the proposed system (e.g., search behavior and relevance feedback on knowledge items), and thereby determine their specific information needs.

Identifying changes in task-stages: We propose an on-line task stage identifier that determines a worker's task stage by analyzing his/her access pattern. The task's temporal profile in each timeframe is the basis for identifying the worker's task stage. A change in task stages is inferred by analyzing the correlation between the consecutive transactions detailed in the worker's temporal profile for the task. Further details are given in the Section 3.2.

Determining variations in topic-needs: In this work, we use a topic-based information search method to overcome the limitations of keyword search methods. We believe it is easier to express information needs by topic identification rather than by keyword-based queries [8]. Since our objective is to identify variations in a worker's topic-needs, we determine the level of his/her topic-needs with indicators of "generality" and "specificity" to show how a level influences the process of profile adaptation. Further details are given in the Section 3.3.

Identifying user's topics of interest interactively: We identify a user's topic-needs precisely by interactively mapping his/her information needs to the specific level of topics in the taxonomy. Initially, the system only shows the portion of the topic taxonomy based on the value of the indicators for the associated topics and the current task stage. The basic algorithm is shown in Table 2, Section 3.4

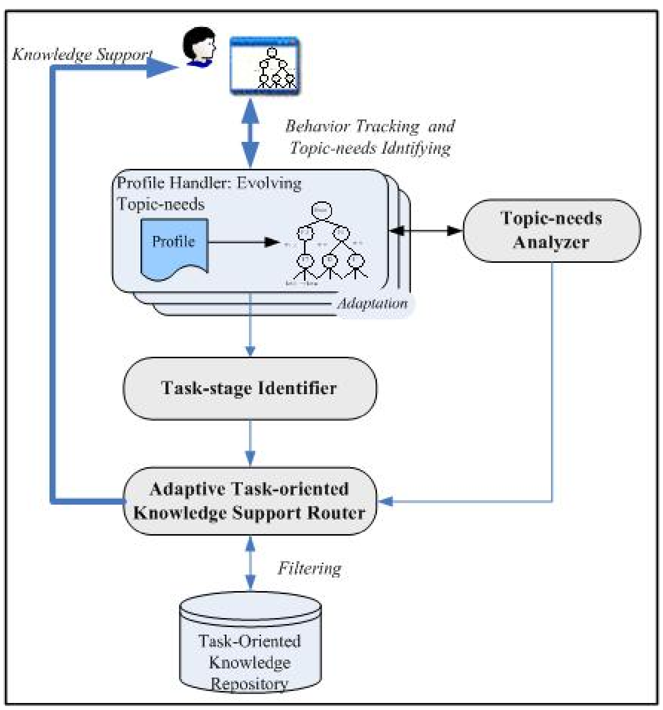


Figure 1. Process of topic-needs determination across task stages

3. Supporting Personal Work-Task Search Activities

3.1 Personal Ontology Formulation

The evolution of a worker's task-needs can be discovered by examining the variety of topics selected from the topic-based taxonomy. Once the top task-relevant topics with the associated weights (denoted as *TRTWs*) have been identified, it would be intuitive to formulate a user's personal topic ontology for a specific transaction. Each worker's information needs are represented by a topic ontology, Ψ_u. *Definition 1* defines the personal topic profile of a user *u*. We represent the worker's information needs in terms of topics in the research domain, instead of using keyword sets. The rationale is that topics provide a more expressive and less abstract means of representing information needs.

Definition 1: The topic ontology of a user *u*, denoted by $\Psi_u = \{<topic_j, w_p(topic_j)>\}$, contains a set of topics (field- or task-level nodes in the taxonomy) with associated degrees of relevance to the target task in a specific time period; $w_p(topic_j)$ represents the relevance degree of $topic_j$ to the target task at time *p*, from the aspect of u. The associated degree of relevance indicates a similarity measure between a topic and the target task in a specific time period. Let *FS* denote the set of topics in the field level and *TS* denote the set of topics in the task level. An ontology threshold value δ can be defined by a worker to generate his/her personalized ontology for the target task by filtering out irrelevant fields or tasks with relevance degrees below the threshold value. Accordingly, $\Psi_u = \{< topic_j, w_p(topic_j)> \mid w_p(topic_j) \geq \delta \text{ and } topic_j \in FS \cup TS\}$. The result forms a user's personalized topic ontology for the target search task. Figure 2 shows a user's personal topic ontology in a transaction based on Example 1.

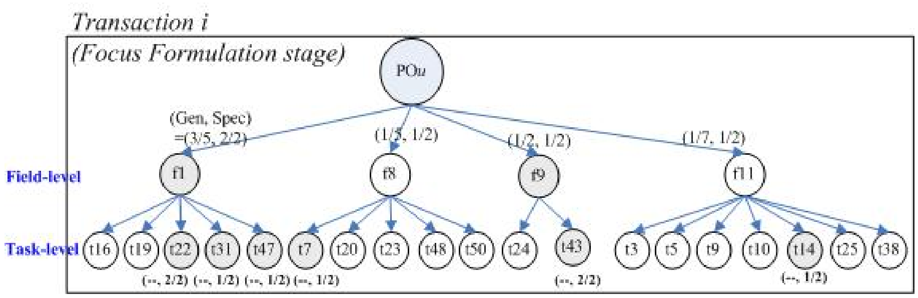


Figure 2. A worker's topics ontology in a specific transaction

3.2 Identifying changes in task-stages

A task-need pattern can be expressed as a set of topics and associated relevance degrees. The task-need pattern of a session *l* in transaction *i* , $Patt^{s_l}_{Trans_i}$, is expressed as a set of topics with their associated relevance degrees ($topic_j$, rd_j). Once the task-needs pattern has been derived, the correlation of the user's task-needs pattern across transactions can be calculated by *Pearson's correlation coefficient*. It is reasonable to assume that a worker's task-needs will not change dramatically during consecutive sessions of the same transaction, but they may change significantly between two different transactions. Thus, we calculate the correlation between the previous transaction, $Trans_{i-1}$, and the start session of the current transaction, $Trans_i^{s_1}$, as shown in the Equation (1):

$$corr_u(A,B) = \frac{\sum_{j \in topic\ set} (rd_j^A - \overline{rd^A})(rd_j^B - \overline{rd^B})}{\sqrt{\sum_{j \in topic\ set} (rd_j^A - \overline{rd^A})^2 \sum_{j \in topic\ set} (rd_j^B - \overline{rd^B})^2}} \quad (1)$$

Let *A* represent $Patt_{Trans_{i-1}}$ and *B* represent $Patt^{s_1}_{Trans_i}$; rd_j^A and rd_j^B are the relevance degrees of topic *j* in $Patt_{Trans_{i-1}}$ and $Patt^{s_1}_{Trans_i}$, respectively; and $\overline{rd^A}$ and $\overline{rd^B}$ are the average relevance degrees of the topics in $Patt_{Trans_{i-1}}$ and $Patt^{s_1}_{Trans_i}$, respectively. Table 1 shows an example of $Patt^1_{Trans_7}$ and $Patt_{Trans_6}$. The rationale behind the proposed *correlation analysis method* is that some task-relevant topics in the topics taxonomy may have a high degree of relevance to the temporal profile of the previous transaction; however, they may have a low degree of relevance to the temporal profile at the beginning of the current transaction. In addition, because the correlation values are within the range [-1, 1], it is easy to track the worker's access pattern based on the correlation value between transactions. Using a correlation analysis method, we took about one year to observe the behavior of workers when they accessed the knowledge repository in the presented task-based workspace. Based on the results of our sample analysis, we were able to identify a user's task stage[16].

3.3 Evolving Topic-needs across Stages

Next, we analyze the user's topic-needs within each task stage. An examination process with two indicators is designed to filter and extract the users' specific topic-needs and update the user's personal ontology. The examination procedure is a top-down process. First, we check the nodes at the field level to assess the generality of the topic-needs, and then check the nodes at the task level to determine the specificity of the topic-needs. The output of the discovery process expresses a worker's information needs for specific topics with "generality" and "specificity" indicators. Technically, the two indicators are clues to help the system reformulate the user's profile. Figure 2 shows an example of a user's personal ontology with two indicators in a specific transaction. The different functionalities of the indicators in each problem stage influence the result of profile adaptation while providing relevant documents.

Generality of relevant topics of worker's task-needs: The higher the generality value, the greater the user's interest in the topics of a specific field. We calculate the generality of task-needs topics across sessions within the same transaction. A field-level topic may include one or more task-level topics; therefore, the generality is the ratio of the top task-relevant topics *(TRTWs)* at the task level to all nodes in a specific field-level topic, as shown in Equation (2). Thus, if the generality value of a field node is equal to one, we know the user has a general interest in each task-level topic within that field.

$$Gen(f_l)_{field-level} = \frac{N_{TRT}^{f_l}(Trans_i)}{N^{f_l}} \quad \textbf{(2)}$$

where N^{f_l} denotes the number of task-level topics belonging to field l in the proposed topic-based taxonomy; and $N_{TRT}^{f_l}(Trans_i)$ is the number of distinct task-level topics belonging to field l and the *TRTW*s of transaction i.

Specificity of relevant topics of worker's task-needs: The higher the specificity value of a field or task-level nodes, the greater the worker's focus on a specific topic node. The specificity of task-needs topics is derived by counting the frequency of the top task-relevant topics across sessions within the same transaction. Equations (3) and (4) show the specificity of topic f_l at the field level and the specificity of topic t_k at the task level.

$$Spec(f_l)_{field-level} = \frac{\sum_{session\ j} B_{i,j}^{f_l}}{S_i} \quad \textbf{(3)}$$

$$Spec(t_k)_{task-level} = \frac{\sum_{session\ j} B_{i,j}^{t_k}}{S_i} \quad \textbf{(4)}$$

where S_i is the number of sessions within a transaction i. $B_{i,j}^{f_l} = 1$ if f_l is a top relevant topic of session j in transaction i; otherwise 0. Similarly, $B_{i,j}^{t_k} = 1$ if t_k is a top relevant topic of session j in transaction i; otherwise 0. The summation of $B_{i,j}^{t_k} / B_{i,j}^{f_l}$ counts the number of sessions in which the topic (t_k or f_l) is a top relevant topic.

3.4 Identifying Personal Topic-needs Interactively

To interactively map a user's information needs to the specific level of topics in the topic-based taxonomy; the system presents general or specific topics to users based on the results reported in the previous section. Table 2 shows the procedure for interactively identifying a user's topic-needs by using stage information. The task-needs analyzer determines the user's task-needs in terms of topics in the domain topic taxonomy, and represents his/her information needs as a personal topic ontology, $\boldsymbol{\Psi_u}$.
The function of the identification procedure is twofold: (1) to support search activities by determining variations in topic-needs and task stages, as shown Steps 1 and 2 in Table 2; and (2) to update the user's personal ontology, $\boldsymbol{\Psi_u}$,

based on his/her feedback on the topics, as shown in Step 3 of Table 2. The variable, *I_Topic* is an array that stores the user's topics of interest derived from the user feedback on the topics during the interactive topic identification process. If the user is in the task pre-focus stage, the system checks whether he/she has general or specific topic-needs. The system will then display the appropriate part of the topic taxonomy to fulfil the user's information needs. It will update the user's Ψu, to help the user conduct future searches. The mechanism helps the user expand his/her topic-needs if the information needs are vague (i.e., in the task pre-focus stage), as shown in the Step 1 in Table 2. As the task progresses, the mechanism guides the user to identify specific topic-needs (i.e., Step 2 in Table 2). Figure 3 is an example of a user has specific field level topic-needs, but doesn't have specific task level topic-needs. The interface can assist the user to identify specific task-level topic needs interactively.

Table 2. Procedure for updating a user's topic ontology by interactive topic-needs identification

Input: (a) $Gen(f_l)$, $Spec(f_l)$: The user's personal ontology with "generality" and "specificity" indicators
(b) Ψ_u: The user' s personal ontology in the Transi
(c) ***Trans$_i$.stage*** : The task-stage of ith transaction
(d) δ, θ: : The threshold of generality or specificity

Output: (a) Show and update the Ψ_u to assist user feedback on relevant topics
(b) Update the user's task profile $\vec{S}_p$ to retrieve relevant documents

Begin

(1) If *Trans$_i$.stage* in "pre-focus stage" then
(1.1) **If** $Spec(t_k)$>= δ **then**
// The user has specific task-needs, but in the pre-focus stage
(1.1.1) **List** the topics in the task-level to help the user identify topics of interest, *I_Topic*
(1.1.2) **Set** all topics in the *I_Topic* = δ

(2) **Else if** *Trans$_i$.stage* is not in "pre-focus stage" **then**
(2.1) If ($Gen(f_l)$ >θ and $Spec(f_l)$ < δ) then
//The user does not have specific field level topic-needs.
(2.1.1) **List** the topics in the task-level of *fields l* to help the user identify topics of interest, *I_Topic*
(2.1.2) **Set** all topics in the *I_Topic* = δ
(2.2) **Else if** ($Gen(f_l)$ > θ and $Spec(f_l)$ >= δ) **then**
// The user has specific field level topic-needs, but not specific task level topic-needs.
(2.2.1) If there are no topics in *field l* with $Spec(t_k)$>= δ then
(2.2.1.1) **List** the topics in the task-level of *fields l* to help the user identify topics of interest, *I_Topic*
(2.2.1.2) **Set** all topics in the *I_Topic* = δ

Else No Action

(3) **Update** Ψ_u and $\vec{S}_p$

End

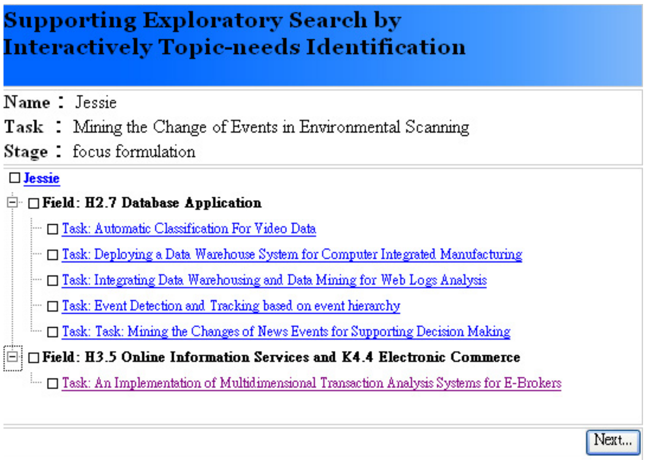


Figure 3. Identifing topic-needs interactively

3.5 Profile Adaptation

The system delivers task-relevant documents based on the results of topic identification, after which the new task profile of the target task, denoted as $\vec{S}_{p+1}$, is generated by Equation (5). The equation considers the worker's task stage as well as the "*generality*" and "*specificity*" of task-need topics, which influence the relevant part of the equation.

$$\vec{S}_{p+1} = \alpha\vec{S}_p + \lambda\ \vec{R} + (1-\lambda)\vec{Temp}_{u,p}$$
$$\vec{R} = w_{gene}\sum\nolimits_{\forall t_j \in Gene.\ topic} Gene(topic_j) \times w_{p+1}(topic_j)\ \overrightarrow{topics}_j + \qquad \textbf{(5)}$$
$$w_{spec}\sum\nolimits_{\forall t_j \in Spec.\ topic} Spec(topics_j) \times w_{p+1}(topic_j)\ \overrightarrow{topics}_j$$

where $\vec{S}_p$ denotes the task profile of the target task at time *p.* The relevant feature vector is derived from the task corpora of relevant tasks sets, while the temporal file, $\vec{Temp}_{u,p}$, is generated according to feedback analysis. In addition, $w_{p+1}(topic_j)$ denotes the associated relevance degree of task t_k or field f_l to the target task. $Gene(topic_j)$ and $Spec(topic_j)$ are derived from task-needs topic analysis, as mentioned earlier. α, and λ are tuning constants. The parameter α is the correlation values between transactions, $corr_u(A,B)$,as shown in the Equation (1). The parameter λ is used to adjust the relative importance of relevant tasks and the temporal profile. Note that the task-stage influences the relative importance of general and specific topics, w_{gene} and w_{spec}. For example, in the early stage of a task, we only consider general topics and ignore specific topics; therefore, w_{gene}=1, and w_{spec}=0. As a task progresses, the content of the temporal profile will be more important than the relevant topics.

4. Preliminary Experimental Setup and Results

4.1 Experimental Setup

The experiments evaluated whether examining a knowledge worker's various information needs for specific topics in each task-stage could help him/her retrieve more relevant information than the traditional incremental learning method based on the relevance feedback algorithm[9][15]. We evaluated four methods: the *Liner-0*, *Linear-0.5*, *Stage-Topic (G, S)* and *Interactive-Topic(G, S) methods*. The *Linear-0 (baseline) method* is an incremental learning process that learns a user's current information needs from feedback about the recommended information (i.e., documents), and updates the user model for future information filtering. It is based on the traditional relevance feedback method in the vector space model [9][15]. The *Linear-0.5* method is similar to the traditional *incremental learning technique* in that it also considers a worker's feedback on relevant topics, i.e. the parameter λ is set to 0.5. The *Stage-Topic (G, S)* and *Interactive-Topic (G, S)* methods are our new methods. The *Stage-Topic (G, S)* method integrates the user's task stage and degree of topic-needs variation into the incremental learning process to deliver task-relevant knowledge. The *Interactive-Topic (G, S)* method is based on the presented topic-needs variation determination method and the interactive topic identification process. After the user profile has been generated, the system retrieves relevant documents from the task-oriented knowledge repository based on the descriptions of the user profile.

Data and Participants: Task-relevant codified knowledge consists of documents in an organization's knowledge repository that have been accessed via the executed task set, i.e., historical tasks. In this work, the tasks were related to writing research papers or conducting research projects, so we selected evaluation subjects who were engaged in a current task. Since performing a task spans a long time period, we chose the subjects according to the task-stage they were in, i.e., the pre-focus, focus formulation, or post-focus task stages. The roles of the subjects were project leader, system analyzer, or technical reporter. The subjects conducted different projects, such as a survey of information technology service management, text analysis for business intelligence, product recommendation, and deployment of a knowledge management system. The system traced and recorded their search behaviors while using the presented system and delivered documents based on their feedback behavior.

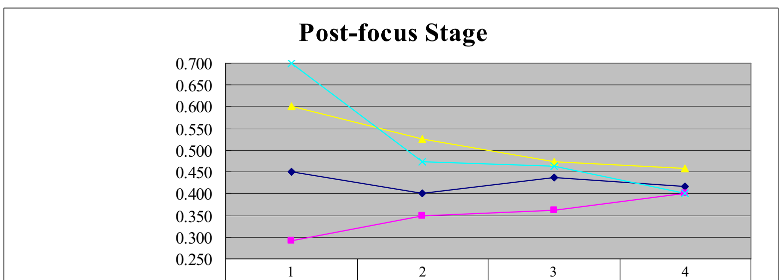
Evaluation metrics: We measured the effectiveness of knowledge support in terms of the precision as used in information retrieval research [1]. The precision rate is the percentage of retrieved items (tasks or documents) that are relevant compared to the total number of retrieved documents.

$$precision = \frac{|retrieved\ \ documents\ \ that\ \ are\ \ relevant|}{|total\ \ retrieved\ \ documents\ \ (TopN)|} \quad \textbf{(6)}$$

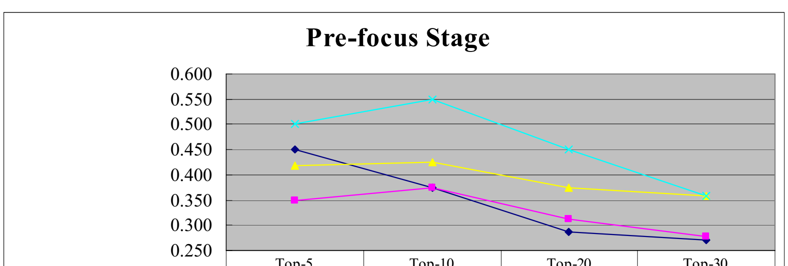
$$Where\ \ N = 5,\ \ 10,\ \ 20,\ \ or\ \ 30$$

4.2 Experimental Results

Observations under Various Levels of Top-N Supply across Stages: Figure 4(a) and 4(b) show the precision rates of the four methods for the pre-focus, and post-focus stages under various top-N documents. Figure 4(a) shows that the *Interactive-Topic (G, S)* method outperforms the other methods in the pre-focus stage under various numbers of document support, i.e. top-5, 10, 20, or 30. This suggests that the user can get effective support with the aid of the interactive topic identification process in the early stage of a task's performance, i.e., the user's topic selection phase. Similarly, in the post-focus stage, the *Stage-Topic (G, S)* and *Interactive-Topic (G, S)* methods perform better than the baseline methods, as shown in Figure 4(b). Notably, the *Interactive-Topic (G, S)* method achieves the best performance in terms of top-5 document support. It is clear that if the personal search factor (i.e., the user's problem stage) and the context factor (i.e., topic variations) are considered simultaneously, the retrieval results could be improved significantly in the early and late stages of a task's performance. Therefore, the proposed topic-needs variation determination method is more suitable for learning the worker's task-needs than the traditional incremental learning technique. In addition, when the worker has specific topic-needs, i.e., the worker in the task post-focus stage, the proposed methods are also effective in helping him/her retrieve task-relevant documents with the aid of topic identification process.



	1	2	3	4
Linear-0 Method	0.450	0.400	0.438	0.417
Linear-0.5 Method	0.292	0.350	0.363	0.400
Stage-Topic (G,S) Method	0.600	0.525	0.475	0.458
Interactive-Topic (G,S) Method	0.700	0.475	0.463	0.400



	Top-5	Top-10	Top-20	Top-30
Linear-0 Method	0.450	0.375	0.288	0.270
Linear-0.5 Method	0.350	0.375	0.313	0.278
Stage-Topic (G,S) Method	0.417	0.425	0.375	0.359
Interactive-Topic (G,S) Method	0.500	0.550	0.450	0.358

Figure 4(a) & (b): The average performance values under various levels of top-N Support in the pre-focus stage, and post-focus stag

5. Conclusions

Most current user modeling approaches focus on the analysis of changes in a user's information needs in his/her daily work life, instead of considering the user's information needs for a specific long-term task. To address this issue, we propose a topic-needs variation determination method based on the refined information seeking and retrieval (IS&R) model. Specifically, the method examines variations in a worker's information needs for topics across task stages, and supports long-term exploratory searches. In contrast to the traditional incremental learning model, which relies on the implicit feedback algorithm (IRF) to identify relevant or irrelevant documents, the proposed model can track a worker's evolving topic-needs across task stages. We developed the Stage-Topic (G, S) and Interactive-Topic (G, S) methods to evaluate the effectiveness of the topic-needs variation determination method. The results of experiments show that the two methods, which consider the personal search factor (i.e., the problem stage) and the context factor (i.e., topic-needs) simultaneously, provide better knowledge support than the traditional incremental learning method. In the future, we will conduct an in-depth analysis of the interrelationship of users' information search behavior patterns, information needs, and different search tasks. The IS&R model will be enhanced by incorporating personal, task, and context factors in order to design a more comprehensive IS&R framework. Furthermore, the cognitive model of document use for a research project at three points or stages, i.e., document selecting, reading, and citing, proposed by Wang & Soergel (1998) and Wang & White (1999) will be incorporated into the IR system design process. Our work could contribute to research on the applications of intelligent information retrieval systems, and enhance the use of knowledge retrieval functions to support project/task execution by professionals.

Acknowledgement

This research was supported by the National Science Council of Taiwan Government under the grant NSC 97-2416-H-030-030-MY2.

References

[1] Baeza-Yates, R. & Ribeiro-Neto, B. *Modern Information Retrieval*, Reading. MA: Addison-Wesley, 1999

[2] Bloodgood, J.M. ,& Salisbury, W.D. "Understanding the influence of organizational change strategies on information technology and Knowledge Management Strategies," *Decision Support Systems*, 2001, 31(1), 55–69.

[3] Byström, K. & Järvelin, K. "Task complexity affects information seeking and use," *Information Processing and Management*, 1995, 31(2), 191-213.

[4] Ingwersen, P. and Järvelin, K. , *The Turn: Integration of Information Seeking and Retrieval Context*, Dordrecht, The Netherlands: Springer, 2005

[5] Kuhlthau, C. *Seeking Meaning: A Process Approach to Library and Information Services*, Ablex Publishing Corp., Norwood, NJ, 1993.

[6] Markus, M.L."Toward a theory of knowledge reuse: Types of knowledge reuse situation and factors in reuse success," *Journal of Management Information Systems,* 2001, 18(1), 57-94.

[7] Nonaka, I. & Konno, N. "The concept of "Ba": Building a foundation for knowledge creation," *California Management Review*, 1998, 40(3), 40-54.

[8] Pons-Porrata, A., Berlanga-Llavori, R. & Ruiz-Shulcloper, J. "Topic discovery based on text mining technique," *Information Processing and Management*, 2007, 43(3), 752-768.

[9] Salton, G. & Buckley, C. "Improving retrieval performance by relevance feedback," *Journal of the American Society for Information Science*, 1990, 41(4), 288-297.

[10] Sieg, R., Mobasher, B., & Burke, R. "Inferring user's information context: Integrating user profiles and concept hierarchies," *In Proceedings of the 2004 Meeting of the International Federation of Classification Societies,* Chicago, IL, USA, 2004

[11] Vakkari, P., Pennanen, M. & Serola, S. "Changes of search terms and tactics while writing a research proposal: A longitudinal case study," *Information Processing & Management,* 2003, 39(3), 445-463.

[12] Vakkari, P. "Growth of theories on information seeking: An analysis of growth of a theoretical research program on the relation between task complexity and information seeking," *Information Processing & Management*, 1998, .34(.2-3), 361-382.

[13] Wang, P. & Sogergel, D. "A cognitive model of document use during a research project. Study I. Document selection," *Journal of the American Society for Information Science and Technology*, 1998, 49(2), 115-133.

[14] Wang, P.L. & White, M.D. "A cognitive model of document use during a research project. Study II. Document at reading and citing stage," *Journal of the American Society for Information Science and Technology, 1999,* 50(20), 98-114.

[15] Widyantoro, H., Ioerger, T.R. & Yen, J. "Learning user interest dynamics with a three-descriptor representation," *Journal of the American Society for Information Science*, 2001, 52(3), 212-225.

[16] Wu, I.C., Liu, D.R. & Chang, P.C. "Toward incorporating a task-stage identification technique into the long-term document support process," *Information Processing & Management* , 2008, 44(5), 1649-1672.