

國立交通大學

資訊管理研究所

碩士論文

透過分析資訊需求變動修正工作特徵檔
以提供工作相關資訊



Profile Adaptation for Providing Task-relevant Information
by Variation of Task Needs

研究生：張北晨

指導教授：劉敦仁博士

中華民國九十五年七月

透過分析資訊需求變動修正工作特徵檔
以提供工作相關資訊

Profile Adaptation for Providing Task-relevant Information
by Variation of Task Needs

研究生：張北晨

Student: Pei-Cheng Chang

指導教授：劉敦仁

Advisor: Duen-Ren Liu

國立交通大學

資訊管理研究所

碩士論文



Submitted to Institute of Information Management

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science in Information Management

July 2006

Hsinchu, Taiwan, the Republic of China

中華民國 九十五年 七月

透過分析資訊需求變動修正工作特徵檔 以提供工作相關資訊

研究生：張北晨

指導教授：劉敦仁 博士

國立交通大學資訊管理所

摘要

在以工作為基礎的企業環境中，如何有效提供工作相關資訊以滿足知識工作者的資訊需求為部署知識管理系統之重要課題。此外，知識密集工作環境中的工作通常具有需要參閱大量文件資料的特性；因此，透過資訊過濾相關技術分析並建構工作者的工作相關特徵檔，為有效提供工作相關知識之重要技術。本研究提出需求特徵化修正方法來分析工作者的動態資訊需求，該方法主要是分析工作者存取的文件及工作主題相關性，並考慮時間因素建構工作特徵檔。此外本研究分析工作者在工作主題之需求變化以找出具相似主題需求變化的相似工作者，並進一步依據相似工作者的主題需求變化進行合作式調整工作特徵檔並推論潛在資訊需求。最後，本研究進行實驗評估，以比較所提方法在提供工作相關資訊之成效。

關鍵字：資訊過濾、工作需求、適性化技術、工作特徵檔、工作主題變動、合作式工作特徵檔

Profile Adaptation for Providing Task-relevant Information by Variation of Task Needs

Student: Pei-Cheng Chang

Advisor: Dr. Duen-Ren Liu

Institute of Information Management

National Chiao Tung University

Abstract

In task-based business environments, an important issue of deploying Knowledge Management System (KMS) is providing task-relevant information (codified knowledge) to fulfill the information needs of knowledge workers. In addition, workers need to access lots of textual documents in conducting knowledge-intensive tasks. Accordingly, effective knowledge management relies on using information filtering (IF) techniques to model worker's information needs as profiles and provide relevant information based on the modeled profiles. This research proposes a novel adaptive task-profiling technique to model worker's information needs on tasks, i.e., task-needs. The proposed technique adjusts task profiles to model worker's dynamic task-needs based on the documents accessed by workers and the relevance on the task-based topic taxonomy. Generally, the more recent the document accessed the more important it is to reflect a work's current task needs. Thus, the effect of time factor is considered in profile adaptation. In addition, the proposed profiling technique adopts a novel collaborative profile adaptation approach to adjust task profiles. We analyze the variations of workers' task needs on the topic taxonomy to identify workers with similar variations of task needs on topics (i.e., topic needs) over time. Similar workers' variations of topic needs are used to predict the target worker's future variations of topic needs, and are used to adjust the target worker's task profile. The codified knowledge that is relevant to the current task can be retrieved based on the adjusted task profile to fit the worker's dynamic task needs. Empirical experiments demonstrate that the proposed approach models workers' task-needs effectively and helps provide task-relevant knowledge.

Keywords : Information filtering, task needs, adaptive profiling technique, self task profile, variation of task-needs on topics, collaborative profile

誌 謝

佈滿回憶的暑夏悄然到來，兩年的學生生活即將隨著時間化作點滴成為這回憶的一部份。下一段旅程！

首先感謝我的指導教授劉敦仁老師，這兩年來給予的許多指導與建議，讓我能夠順利完成碩士學位。也感謝口試委員陳彥良博士與李永銘博士，在口試時提供的許多寶貴意見，讓本論文得以更加完善。

感謝所有學長姐們在課業與生活上的照顧和幫忙；感謝同學們在課業上一同學習和研究、在論文撰寫期間一起加油打氣；感謝學弟妹們為研究室帶來歡樂氣氛。此外，特別感謝怡瑾學姐在研究過程中的指導，給了我很大的幫助與支持，著實使我獲益良多。

最後要感謝我的家人與朋友們，謝謝你們的關懷、支持與鼓勵，讓我能一直保有勇氣與希望。感謝所有幫助過我的人，願快樂常在你們心中。



Table of Content

1. Introduction.....	1
2. Literature review.....	4
2.1. Task-based knowledge management and retrieval.....	4
2.2. Information retrieval in a vector space model	5
2.3. User modeling by information filtering technique.....	6
2.4. Relevance feedback techniques	7
3. Overview of methodology	9
4. Modeling task needs	12
4.1. Capturing users' access behavior	12
4.2. Self profile adaptation.....	12
5. Measuring variations of user topic needs over Time	16
5.1. Variation measurement for a specific topic.....	16
5.2. Representation model: Topic-needs variation matrix	17
6. Collaborative profile adaptation.....	19
6.1. Identifying similar workers.....	19
6.2. Prediction of task needs	22
6.3. Result demonstrations.....	24
7. Experiments	28
7.1. Experimental setup.....	28
7.1.1. Experimental objective and design	28
7.1.2. Data and participants.....	29
7.1.3. Evaluation metrics	29
7.2. Experimental result and observations	30
7.2.1. Experiment 1-1: parameter selection for <i>TP-V</i> method	30
7.2.2. Experiment 1-2: parameter selection for <i>TP-D</i> method.....	31
7.2.3. Experiment 2: comparisons of self profile adaptation methods.....	32
7.2.4. Experiment 3: comparisons of various methods.....	34
7.2.5. Case inspections.....	35
7.2.6. Discussions	36
8. Conclusion and future research issues	38
Reference	39

List of Tables

Table 4.1 Definitions of symbols and parameters used in the equations.....	13
Table 6.1 The definitions of symbols used in the identification algorithm.....	19
Table 7.1 Effectiveness of <i>TP-V</i> method under various δ_v values	31
Table 7.2 Effectiveness of the <i>TP-D</i> method under various δ_D values.....	32
Table 7.3 Comparison between self profile adaptation methods	33
Table 7.4 Comparison between methods	34



List of Figures

Fig. 3.1 Overview of proposed profile adaptation	9
Fig. 4.3 Example of modeling task needs	15
Fig. 5.1 An example of topic-needs variation matrix.....	18
Fig. 6.1 Algorithm of identifying similar workers.....	21
Fig. 6.2 Sample of trimming the variation matrix of the target worker.....	21
Fig. 6.3 Sample of calculating <i>SimVM</i>	22
Fig. 6.4 Sample of calculating <i>SimTP</i>	22
Fig. 6.5 Event information	25
Fig. 6.6 Similarity with each topic.....	25
Fig. 6.7 Variation of topic needs over time	26
Fig. 6.8 Similar Workers	26
Fig. 6.9 Prediction Result.....	27
Fig. 7.1 Result of knowledge support of <i>TP-V</i> under various δ_v value.....	31
Fig. 7.2 Result of knowledge support of <i>TP-D</i> under various δ_D value.....	32
Fig. 7.3 Trends of retrieval effectiveness of the four methods under various top-N document support.....	33
Fig. 7.4 Trends of retrieval effectiveness of the four methods under various top-N document support.....	35
Fig. 7.5 Four experimental cases	35
Fig. 7.6 Trends of retrieval effectiveness of the three cases (case 1, 2,3) under various top-N document support	36

1. Introduction

Knowledge management systems (KMSs) are important and useful tools for an enterprise to manage business knowledge effectively. In addition, KMSs provide adequate support for improving decision making and gaining competitive advantage. In an organization, workers are assigned to carry out various tasks, and workers need to apply their experience and professional knowledge to complete the tasks to achieve organizational goal [12][13]. For this reason, an organization has the urgency to make use of the knowledge assets which are emerging from organizational operations and management activities to increase its profitability and productivity with the support of KMSs. For KMSs, Information Technology (IT) plays the role of facilitating the access, reuse and sharing of knowledge assets within and across an organization to assist knowledge workers in executing their tasks [9][17]. In addition, workers need to access lots of textual documents in conducting knowledge-intensive tasks. Accordingly, Information Filtering (IF) techniques are often employed to model users' information needs as profiles and provide relevant information based on the modeled profiles.

Effective knowledge management relies on understanding workers' information needs on tasks, for brevity, task-needs. As the operations and management activities of enterprises are mainly task-based, KMSs focus on providing task-relevant knowledge to workers engaged in knowledge-intensive tasks [1][2][11][12][13]. The Kabiria system supports knowledge-based document retrieval in office environments, allowing users to conduct document retrieval according to the operational context of task-associated procedures [8]. The KnowMore system maintains task specifications (profiles) to enumerate the process-context of tasks and associated knowledge items [1]. Context-aware delivery of task-specific knowledge can then be facilitated based on the task specifications and current execution context of the process. The above works provide an appropriate perspective for designing task-based knowledge support. However, they focus on specifying the process-context of a task to support context-aware or process-aware knowledge retrieval, rather than on a systematic method for constructing a task profile that models a worker's task needs. Moreover, very few researches address the issue of profile adaptation to track workers' dynamic information needs.

This research proposes a novel adaptive task-profiling technique to model worker's information needs on tasks, i.e., task-needs, as follows.

First, a *task profile* specifies the key concept terms of a worker's current task (task at hand), and models the information needs of the worker during the task's execution. The proposed technique adjusts task profiles to model workers' dynamic task-needs based on the documents accessed by workers without considering user feedback (positive or negative opinion).

Second, in order to model users' task needs, a proper task-based topic taxonomy is used to conceptualize the domain information of organizational activities. Note that the main subjects of organizational activities defined by domain experts and previously (executed) representative tasks form the task-based topic taxonomy. The topics and their corresponding topic profiles are used as references to adjust task profiles according to their relevance (similarity) to the documents accessed by the workers.

Third, generally, the more recent the document accessed the more important it is to reflect a work's current task needs. Thus, the effect of time factor is considered in profile adaptation.

Fourth, the proposed profiling technique adopts a novel collaborative profile adaptation approach to adjust task profiles. Knowledge workers usually require a substantial amount of time to accomplish knowledge-intensive tasks. For such long-term tasks, the information needs of the workers may vary according to their progress during the performance of tasks. For example, a graduate student is seeking adequate knowledge documents for her research. Her research topics may vary as the following: "Event detection" => "Mining event change" => "Mining Patent change" => "Patent Mining", where the symbol '>=>' denotes that the left-hand side occurs before the right-hand side. Without explicitly specifying the change of information needs directly by workers, we try to capture the variations of workers' information needs through those documents accessed by the workers. This work uses the variation of workers' topic needs to model the variation of the workers' information needs. Conventional user profiling approaches, which are based on relevance feedback on documents, can only reflect the information needs accumulated, and lack consideration of possible change of information needs. This work measures the variation of workers' topic needs according to their knowledge activities (e.g. access documents). Workers with similar variations of topic needs over time are identified. A novel collaborative profile adaptation approach is proposed to adjust task profiles via

using similar workers' variations of topic needs to predict the target worker's future possible variations of topic needs.

The proposed approach enhances knowledge retrieval through collaboration from similar workers. The codified knowledge that is relevant to the current task can be retrieved based on the adjusted task profile to fit the worker's dynamic task needs. Empirical experiments demonstrate that the proposed approach models workers' task-needs effectively and helps provide task-relevant knowledge.

The rest of this thesis is organized as follows. Section 2 surveys the related work. The overview of proposed methodology is described in Section 3. The approach of task needs modeling is detailed in Section 4. The details of measuring and maintaining variation of topic needs over time are presented in Section 5. Section 6 details how to identify similar workers with similar variations of topic needs and the proposed collaborative profile adaptation approach. The experiments of our proposed approaches are illustrated in Section 7. The conclusion and future works are concluded in Section 8.



2. Literature review

2.1. Task-based knowledge management and retrieval

Managing knowledge within and across organizations is considered as an important tactic for gaining competitive advantages in nowadays business environments. Knowledge Management (KM) activities generally include creation, management and sharing and all the activities make up a cycle or repeated processes [9][24]. Organizations make use of the knowledge assets emerging from organizational operations and management activities to increase their profitability and productivity with the support of Knowledge Management Systems (KMSs) [16][24]. Contemporary KMSs employ Information technology (IT) such as document management and workflow management to facilitate the access, reuse and sharing of knowledge assets within and across organizations [9][17].

Generally, IT mainly focuses on explicit and tacit dimension in knowledge management to support knowledge management activities [17]. In order to manage the explicit knowledge, there are four primary resources being utilized: repositories, refineries, organization roles, information technologies [42]. The repository of structured and explicit knowledge, especially in document form, is a codified approach to manage knowledge [9][15]. Nevertheless, with the growing amount of information in an organization, KMSs face the challenge to fulfill users' information needs.

In general, the operations in an organization are planned around tasks. As knowledge is embedded during the execution of tasks, providing task-relevant knowledge to fulfill the information needs of knowledge workers to complete their tasks is important. Knowledge retrieval is considered as a core component in supporting the workers to perform knowledge-intensive task in a business environment [11]. Recently, Information Retrieval (IR) technique has been greatly exploited in workflow management systems to support the knowledge workers to obtain task relevant knowledge. Furthermore, it is combined with workflow management systems to proactive deliver task-specific knowledge to users [2][11].

For complex and knowledge-intensive tasks, the collaboration among knowledge workers is helpful when there are some knowledge workers interested in similar problems or they share common interests. Sharing knowledge with peer groups is a

superior method in knowledge management [13]. The Computer-Supported Cooperative Work (CSCW) and Recommender Systems also give something additional splendor on collaboration [28][29]. CSCW emphasizes on the power of computer system to help groups of people perform the tasks in a shared environment [29]. Recommender systems employ content-based filtering and collaborative filtering to recommend web pages, movies, books and so on [14][26][28]. However, it is more difficult to provide task-relevant knowledge during the progress of execution of complex and knowledge-intensive tasks because such tasks often consist of several smaller tasks.

2.2. Information retrieval in a vector space model

The key contents of a codified knowledge item (document) can be represented as a term vector (i.e., a feature vector of weighted terms) in n -dimensional space, using a term weighting approach that considers the term frequency, inverse document frequency, and normalization factors [32]. The *term transformation* steps, including case folding, stemming, and stop word removal, are performed during text pre-processing (Salton et al., 1971; Poter, 1980; Witten et al., 1999). Then, *term weighting* is applied to extract the most discriminating terms [3]. Let d be a codified knowledge item (document), and let $\vec{d} = \langle w(k_1, d), w(k_2, d), \dots, w(k_n, d) \rangle$ be the term vector of d , where $w(k_i, d)$ is the weight of a term k_i that occurs in d . Note that the weight of a term represents its degree of importance in representing the document (codified knowledge). The well-known *tf-idf* approach, which is often used for *term (keyword) weighting* (Poter, 1980), assumes that terms with higher frequency in a document and lower frequency in other documents are better discriminators for representing the document. Let the term frequency $tf(k_i, d)$ be the occurrence frequency of term k_i in d , and let the document frequency $df(k_i)$ represent the number of documents that contain k_i . The importance of k_i is proportional to the term frequency and inversely proportional to the document frequency, which is expressed as Eq. 2.1:

$$w(k_i, d) = \frac{1}{\sqrt{\sum_i (tf(k_i, d) \times \log(N/df(k_i) + 1))^2}} tf(k_i, d) \times \left(\log \frac{N}{df(k_i)} + 1\right) \quad (2.1)$$

where N is the total the number of documents. Note that the denominator on the right-hand side of the equation is a normalization factor that normalizes the weight of

a term.

Similarity measure: The cosine formula is widely used to measure the degree of similarity between two items, x and y , by computing the cosine of the angle between their corresponding term vectors \vec{x} and \vec{y} , which is given by Eq. 2.2. The degree of similarity is higher if the cosine similarity is close to 1.

$$\text{sim}(x, y) = \text{cosine}(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} \quad (2.2)$$

2.3. User modeling by information filtering technique

Information retrieval (IR) and information filtering (IF) technologies applied to document management is generally the first step of knowledge management activities, since textual data such as articles, reports, manuals, know-how documents and so on are treated as the valuable and explicit knowledge within organizations [24]. In addition, IR and IF are considered as the core technologies to help organizations collect and process documents, to reduce the problem of information overload, and to provide relevant and needed information for knowledge workers to accomplish their tasks [6][19][34].

IF systems are similar to conventional IR ones, but rather than focusing on facilitating users' short-term information needs, IF systems lay emphasis on personalization to support long-term information needs of users [3][5][22][23][39]. Accordingly, maintaining and learning users' profiles is an important issue in order to support long-term information services.

Various approaches for learning users' interests or preferences from textual documents or web pages have been proposed [4][21][22][23][27]. The well-known approaches in Information Retrieval or Information Theory are modified and then employed to model or capture user's dynamically changed interests, for example, Rocchio algorithm, information gain theory, Bayesian classifier. Sieg et al. (2004) integrate user profiles and concept hierarchies to infer users' information context to enhance original queries. Widiantoro et al. (2001) use a three-descriptor model to learn user's multiple interest dynamics, which maintains a long-term descriptor to capture the user's general interests and a short-term descriptor to keep track of the user's more recent, faster-changing interests. Moreover, an auto weight-adjusted mechanism is employed to adjust the weight of positive and negative descriptors to make the short-term descriptor react to a drastic change in interest faster. Notably, all

these approaches require users' relevance feedback, including explicit one (users' linguistic rating) and implicit one (users' access behavior) to reach this goal. Relevance feedback effectively improves search effectiveness through query reformulation. Various studies have demonstrated that relevance feedback applied in the vector model is an effective technique for information retrieval [30][33]. Consequently, the IF systems learn users' current information needs from the relevance feedback and update the model for information filtering in the future. Such kind of learning approaches can maintain the users' profiles adequately once the systems receive the feedback, hence the learning approaches are regarded as the incremental learning technique.

In addition to relevance feedback, the characteristic of knowledge retrieval activities in the working environment is also needed to be taken into consideration to support the workers more precisely. The characteristic of knowledge retrieval activity is that the worker's information needs is always associated with the executing task at hand. Generally, a worker uses documents to understand a task, solve the encountered problem, or result in another search behavior to find a solution. Accordingly, several empirical studies focus on how documents are selected and used by workers during executing task [36][37]. Furthermore, κ -Support System takes the characteristics of task stage into account and employs domain ontology to provide dynamic knowledge support [19].

Though the success of the IF techniques, few consider that the effect of variation processes of users' information needs and collaboration between the workers who have similar variation processes. So an enhanced profile adaptation approach taking the variations of topic needs over time of workers into account in advance and keeping them rather than just reflecting the information needs accumulated is required. Such can help the workers take advantage of the past experience of the other workers and enable them to have more chances to retrieve task-relevant knowledge.

2.4. Relevance feedback techniques

Relevance feedback (RF) improves the search effectiveness through query reformulation [33]. The RF technique reformulates or expands the original query based on partial relevance judgments, i.e., feedback on part of the evaluation set. Relevant documents with positive feedback have a positive influence on the weight of terms, while irrelevant documents with negative feedback have a negative influence

on the weight of terms. A refined query vector can be generated by adding the term weights of relevant documents and subtracting the term weights of irrelevant documents. Eq. 2.3 illustrates the Standard_Rocchio method designed by Rocchio. A modified query vector \vec{q}_m is derived using the relevance of documents (as feedback) to adjust the query vector \vec{q} .

$$\text{Standard_Rocchio: } \vec{q}_m = \alpha \vec{q} + \beta \frac{1}{|D_r|} \sum_{\forall d_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_n|} \sum_{\forall d_j \in D_n} \vec{d}_j \quad (2.3)$$

where D_r denotes the set of relevant documents and D_n represents the set of irrelevant documents according to user assessments. $|D_r|$ and $|D_n|$ represent the number of documents in the sets D_r and D_n respectively; and α, β, γ are tuning constants.



3. Overview of methodology

In this research, we propose a profile adaptation technique to adjust task profiles and enhance knowledge retrieval. Fig. 3.1 shows the proposed methodology.

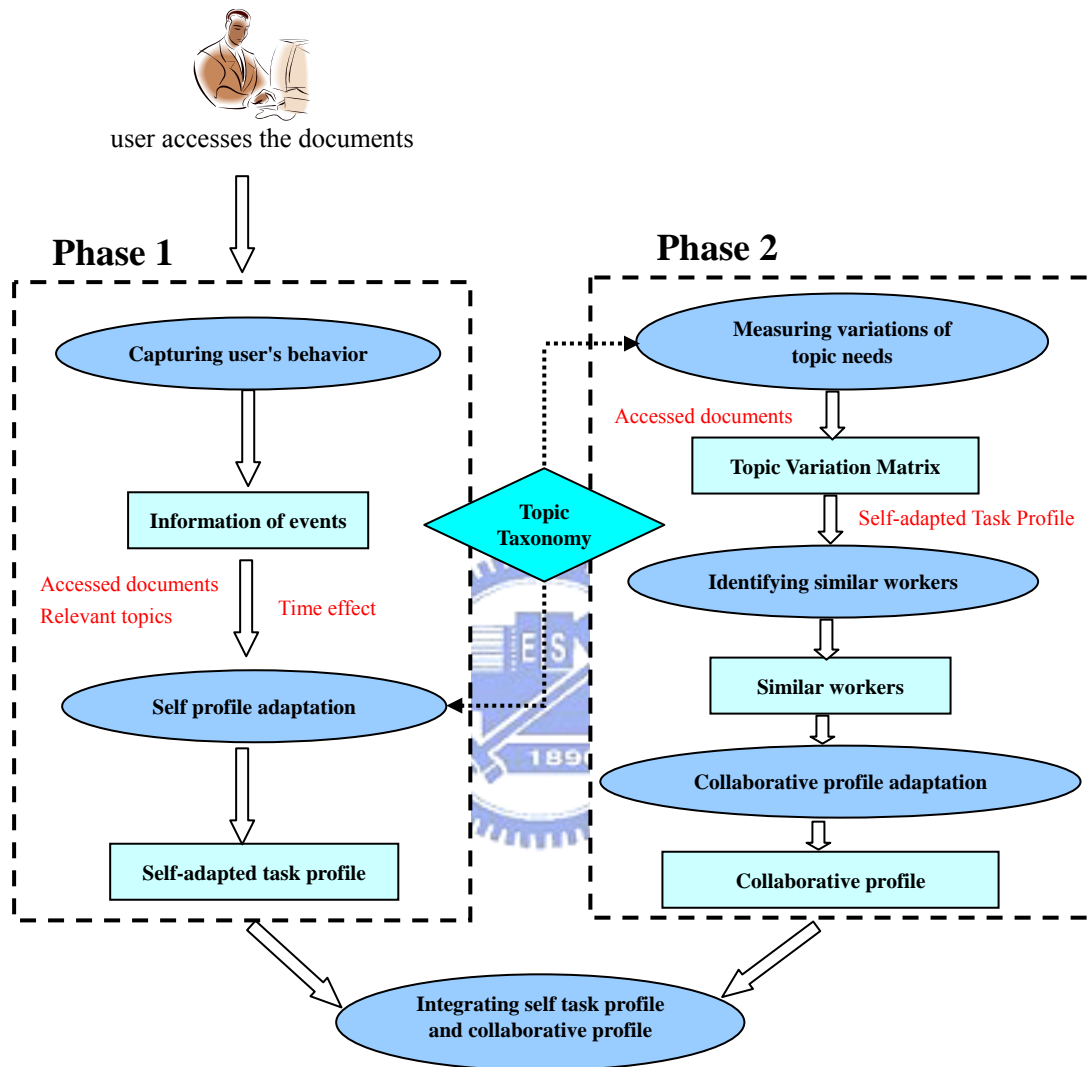


Fig. 3.1 Overview of proposed profile adaptation

The proposed adaptive task-profiling technique for modeling worker’s dynamic task-needs mainly comprises two phases, the self adaptation phase and collaborative adaptation phase. The self adaptation phase incrementally adjusts task profiles based on the documents accessed by workers, relevant topics and the time effect. The collaborative adaptation phase uses similar workers’ variations of topic-needs to adjust task profiles. The profiling approach uses an event-based approach to trigger the profile adaptation process, where an event occurs when a worker accesses a document at specific time.

(1)Phase 1 : Self profile adaptation

Whenever workers access a knowledge document, the information about the user's behavior is captured. The self adaptation phase considers the effect of time factor and the user's behavior (document accessed) to adjust the corresponding task profile with the aid of task-based topic taxonomy without considering user feedback (positive or negative opinion). In order to model users' task needs, a proper task-based topic-taxonomy is used to conceptualize the domain information of organizational activities. Some topics in the topic taxonomy are defined by the domain experts according to the main subjects of applied organizational activities. Some topics are defined using the previously executed tasks, which are considered as important tasks in an organization. Task-based topics play as important references of past experience to provide workers with task-relevant knowledge, since the task at hand is generally related to some previously executed tasks in an organization [9]. Our previous work shows that relevant topics play as points of reference and are very helpful to generate task profiles, especially when very few documents are accessed in the earlier phase of task executions. The topics and their corresponding topic profiles are used as references to adjust task profiles according to their relevance (similarity) to the documents accessed by the workers. Note that a topic profile specifies the weighted concept terms of a topic.

Previous task profile is adjusted to generate a new self-adapted task profile that represents a worker's future task needs according to the task-relevant topics and the accessed documents incrementally. Moreover, the adaptation takes the time factor into consideration. The rationale is that the more recent the document accessed the more important it is to reflect a work's current task needs. Thus, the contribution of the user's previous task needs (profile) towards adjusting the task profile needs to be reduced properly according to the proportion of passed time. The details of the proposed self profile adaptation approach are described in Section 4.2.

(2)Phase 2 : Collaborative profile adaptation

Knowledge workers usually require a substantial amount of time to accomplish knowledge intensive tasks. For such long-term tasks, the information needs of the workers may vary according to their progress during the performance of tasks. Self-adapted task profiles can only reflect the information needs accumulated, and lack consideration of possible changes of information needs. The rationale of collaborative profile adaptation is that workers with similar changes of information needs in the

past are likely to have similar changes of information needs in the future; and thus a worker's possible change of information needs can be derived from similar workers' changes of information needs.

Without explicitly specifying the change of information needs directly by workers, we try to capture the variations of workers' information needs through those documents accessed by the workers. This work uses the variation of workers' topic needs over time to model the variation of the workers' information needs, as described in the following.

The profiling approach uses an event-based approach to trigger the profile adaptation process. An event-based topic needs at specific time is modeled as weighted topics, where the weight of a topic is derived by considering the similarity between the topic profile and the document profile of the document accessed by the worker at that time. A worker's topic-needs at specific time T is derived as weighted topics on the topic taxonomy by accumulating the event-based topic needs over time with the consideration of time decay. The derivation not only considers the event-based topic needs at time T , but also considers the event-based topic needs prior to time T . The variation of workers' topic needs can then be measured according to the difference of topic weights between the two time points.

Each worker's variations of topic needs over time are expressed as a topic-needs variation matrix, i.e., a time-period by topic-needs matrix. A similar worker identifying approach is used to find workers with similar variations of topic needs according to the derived topic-variation matrix and self-adapted task profiles over a time window. Note that we focus on identifying similar workers with similar variations of topic needs and similar task needs. Thus, self-adapted task profiles are also considered in the identification process. Once similar workers are identified, similar workers' task needs (at time $T+1$) are used to predict the target worker's task needs at time $T+1$ that is modeled as a collaborative profile. The derived collaborative profile is then combined with the self-adapted task profile to generate a new task profile representing the target worker's future task needs at time $T+1$. Two approaches are used to derive the collaborative profile. One is based on the variation of topic needs to derive the collaborative profile, in which similar workers' variations of topic needs from time T to $T+1$ are used to predict the target worker's possible variations of topic needs. The alternative is based on the documents accessed at time $T+1$, where similar workers' documents accessed at time $T+1$ are used to derive the collaborative profile.

4. Modeling task needs

Section 4.1 describes how our system captures and stores users' access behavior (document accessed). Section 4.2 illustrates the proposed self profile adaptation approach that considers the effect of time factor and the user's behavior (document accessed) to adjust the corresponding task profile with the aid of task-based topic taxonomy.

4.1. Capturing users' access behavior

Our \mathcal{K} -support system records workers' knowledge activities during the execution of their works in previous research [19]. Whenever a worker performs an action about accessing any document, the system creates a new record to store the information of corresponding knowledge activity. In the following, an example is shown to explain how to capture and store users' access behavior.

Example: Assuming that a worker "Mrkid" is searching for knowledge documents in \mathcal{K} -support system, and he finds that a document "Learning User Interest Dynamics with a Three-Descriptor Representation" may help his task. "Mrkid" performs a "reading" action at time "2005-10-31 21:05:03" accordingly, and with the help of \mathcal{K} -support system, the information about the "reading" action is recorded in the system.

In the above example, the stored information is {"Mrkid", "2005-10-31 21:05:03", "reading", "Learning User Interest Dynamics with a Three-Descriptor Representation"}. All attributes are converted properly into identifiable number except the 'time' attribute.

Hereafter, we use the word 'event' to denote an action performed by some user about accessing any document. In this research, only four kinds of event are adopted, including "download documents", "download reports of documents", "read documents on-line", and "upload documents".

4.2. Self profile adaptation

Whenever an event of worker's access behavior is detected, the system captures and records the document accessed by the worker. The event triggers the self profile adaptation process to adjust the worker's task profile according to the information of the corresponding event.

A document/task/topic profile specifies the weighted concept terms of the document/task/topic. Vector-based approach is adopted to represent a document/task/topic profile. A modified relevance feedback technique, adopted from the techniques proposed by Rocchio (1971), is used to adjust the workers' task profiles based on the profiles of documents accessed by the worker and the topic profiles of identified relevant topics. The adjustment considers the effect of time factor.

The proposed profiling technique is given in E.q. 4.1, and E.q. 4.2. The associated definitions of symbols and parameters used in the equations are listed in Table 4.1. Let T denote the index of the actual time when the worker performs the latest action of document access. \vec{S}_{T+1} denotes the worker's task profile generated at time T , which can be used to model his/her task-needs at time $T+1$. The equation includes the decay of previous task-profile \vec{S}_T , which models the worker's task needs at time T , and the current information needs derived from the document accessed at time T . $Decay(\vec{S}_T)$ represents the accumulated task needs from the beginning to the current time T by considering the time decay of previous task-profile.

$$\vec{S}_{T+1} = \alpha \times Decay(\vec{S}_T) + [\lambda \vec{O}_T + (1 - \lambda) \vec{D}_T] \quad (4.1)$$

$$\vec{O}_T = \beta \frac{1}{|O_{pos}^T|} \sum_{O_i \in O_{pos}^T} \vec{O}_i - \gamma \frac{1}{|O_{neg}^T|} \sum_{O_j \in O_{neg}^T} \vec{O}_j$$

$$Decay(\vec{S}_T) = \sum_{t=1}^{T-1} TW_{t,T} \times [\lambda \vec{O}_t + (1 - \lambda) \vec{D}_t] \quad (4.2)$$

$$TW_{t,T} = \frac{\text{the actual time for } t - ST}{\text{the actual time for } T - ST}$$

Table 4.1 Definitions of symbols and parameters used in the equations

T	the index of the actual time when the worker performs the latest action of document access
\vec{S}_{T+1}	the task profile generated at time T , which can be used to model the worker's task-needs at time $T+1$
\vec{D}_T	the document profile of the document accessed by the worker at time T
\vec{O}_T	the aggregate topic profile derived from the topic profiles of relevant topics and irrelevant topics

O_i	the topic i in the topic taxonomy
O_{pos}^T	the positive topic set derived at time T base on the relevance degrees of the topics to the document accessed at time T
$ O_{pos}^T $	the number of topics in O_{pos}^T
O_{neg}^T	the negative topic set derived at time T based on the relevance degrees of the topics to the document accessed at time T
$ O_{neg}^T $	the number of topics in O_{neg}^T
t	the index of an actual time when the user performs an action of document access; $t=0$ denotes the time when the task starts
$TW_{t,T}$	the time weight of the event that occurred at time t (with respect to time T)
ST	the starting time when the worker's task starts (in milliseconds)
α	the tuning parameter used to adjust the weight of previous task profile
λ	the tuning parameter used to adjust the weights of topic profile \bar{O}_T and document profile \bar{D}_T
β, γ	the tuning parameters used to adjust the weights of positive topic set and negative topic set

The self-adapted task profile \bar{S}_{T+1} is generated from previous task profile \bar{S}_T applied with a decay function and is refined by using the current information needs derived from the document accessed at time T . The current information needs consists of two parts: the *document profile* and the *aggregate topic profile*. The *document profile* \bar{D}_T intuitively is the profile (feature vector) of the document accessed at time T . Task-based topics play as important references of past experience to adjust task profiles according to their relevance (similarity) to the document accessed by the workers. The relevance degree of a topic O_i to the document D_T is obtained by calculating the similarity (cosine measure) between \bar{O}_i and \bar{D}_T . The *aggregate topic profile* is derived from the topic profiles of relevant topics in the positive topic set and irrelevant topics in the negative topic set. The positive topic set reflects the positive information needs of the worker, and is obtained by selecting the topics with relevance degree higher than a defined threshold. The negative topic set reflects the negative information needs of the worker on the topic taxonomy and is obtained by selecting the topics with relevance degree lower than a defined threshold. Our previous research shows that topic profiles are more important to adjust task profiles than document

profiles during the early phase of task executions [41], so a parameter λ is used here to adjust the weights of the document profile and the aggregate topic profile. Fig. 4.3 illustrates the given technique.

The profile adaptation also takes the effect of time factor into consideration. $Decay(\vec{S}_T)$ represents the accumulated task needs from the beginning to the current time T by considering the time decay of previous task-profile, as given in E.q. 4.2. \vec{S}_T denotes the previous task profile generated at time $T-1$, and plays the role of previous task needs. \vec{S}_T is the summation of aggregate topic profiles and document profiles derived from time ST to $T-1$. Generally, the more recent the document accessed the more important it is to reflect a work's current task needs. Thus, a time weight is employed to reflect the decay of previous aggregate topic profiles and document profiles towards their contribution to worker's current task needs. $TW_{t,T}$ is the time weight of an event occurred at time t with respect to T , and is defined as the ratio of time difference $t-ST$ to $T-ST$. Thus, $Decay(\vec{S}_T)$ can reflect the effect of previous task profile towards current task profile more accurately with TW than just using \vec{S}_T .

$$\vec{S}_{f+1} = \alpha \times Decay(\vec{S}_f) + [\lambda[\beta(\bar{O}_2 + \bar{O}_7 + \bar{O}_5) / 3 - \gamma\bar{O}_4] + (1-\lambda)\bar{D}_{46}]$$

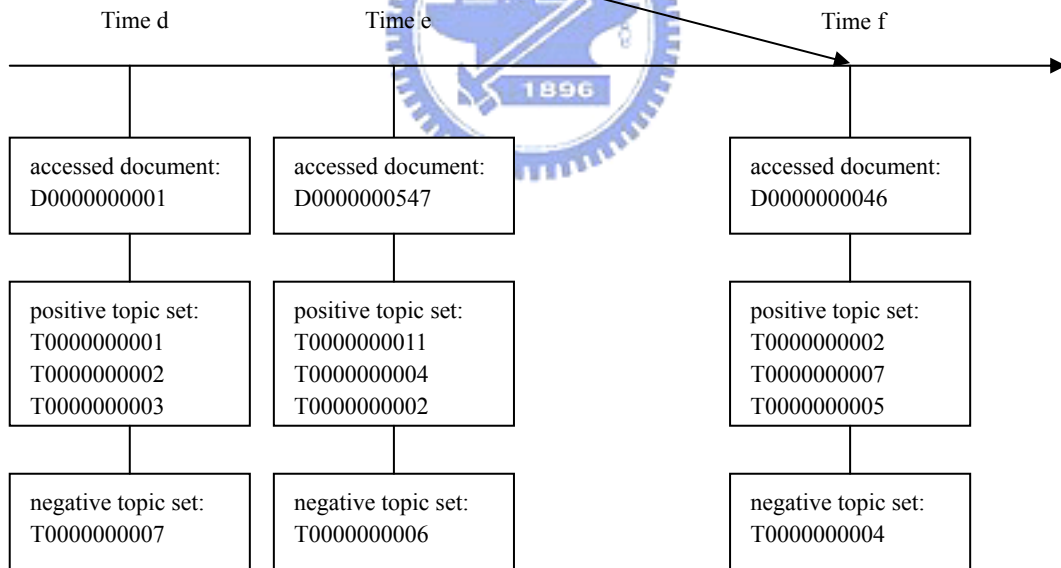


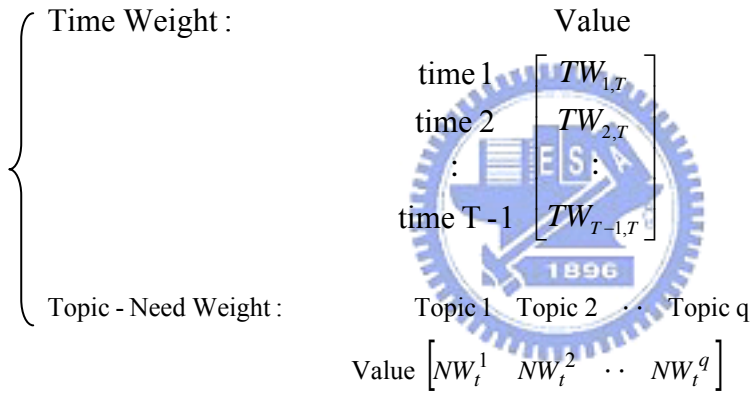
Fig. 4.3 Example of modeling task needs

5. Measuring variations of user topic needs over Time

Section 5.1 describes the proposed approach to measure the variation of the worker's topic needs, i.e., information needs reflecting on the topic taxonomy. The worker's variations of topic needs over time are represented as a topic-variation matrix, as described in Section 5.2.

5.1. Variation measurement for a specific topic

Herein, two factors are taken into account in measuring the variation of information needs on a specific topic i : one is the time factor, named the time weight $TW_{t,T}$, and the other is the relevance degree of topic i in the topic taxonomy, namely the topic-need weight NW_t^i , for each event occurred at time t . The following shows the concepts of time weight and topic-need weight over time.



where,

T	the index of an actual time when the latest event occurs
$TW_{t,T}$	the time weight of the event occurred at time t (with respect to time T)
NW_t^i	the topic-need weight on topic i for the event occurred at time t

$TW_{t,T}$ is the time weight of the event occurred at time t with respect to time T as described in Section 4.2. NW_t^i is the topic-need weight on topic i for the event occurred at time t and is obtained by calculating the similarity value (using vector-based cosine method) between the document accessed at time t and the profile of topic i . Whenever the worker performs an action of document access, the topic-need weight for the corresponding event is calculated.

After obtaining all the task-need weights for the events of the worker, the variation of topic needs over time can be measured, as given in Eq. 5.1. Given two time index: d and e , where $d < e$. Let $NV_{d,e}^i$ denote the variation of the worker's information needs on a specific topic i between time d and time e . $NV_{d,e}^i$ is the difference of accumulated topic needs IN_e^i at time e and IN_d^i at time d . The accumulated topic needs at time e is the summation of $TW_{t,e} \times NW_t^i$, for $t = 1$ to e .

$$NV_{d,e}^i = IN_e^i - IN_d^i = \sum_{t=1}^e TW_{t,e} \times NW_t^i - \sum_{t=1}^d TW_{t,d} \times NW_t^i \quad (5.1)$$

where,

$NV_{d,e}^i$	the variation of the worker's information needs on a specific topic i between time e and time d ($d < e$)
IN_d^i	the worker's accumulated information needs on topic i at time d
IN_e^i	the worker's accumulated information needs on topic i at time e

The relevance degree of topic i is different at time d and e , and thus, NW_t^i is exploited to take such situation into consideration. The time weight is used to reflect the effect of time decay of topic needs between time d and e . The measurement considers the accumulated topic needs over time; thus the events occurred before time d and time e are also considered.

5.2. Representation model: Topic-needs variation matrix

A representation model is defined to represent the variations of topic needs over time for the given target worker. A vector-based model is adopted to represent the variations of information needs on topics. Each worker's variations of topic needs over time are expressed as a time-period by topics (in the topic taxonomy) matrix, i.e. topic-needs variation matrix, and the matrix consists of several topic-needs variation vectors.

Given any two time indexes: d and e , where $d < e$. Let $NV_{d,e}$ denote the variation vector of the worker's topic needs between time e and d . $NV_{d,e}$ are defined as E.q. 5.2. Notably, the measurement of the variation on a specific topic i , i.e., $NV_{d,e}^i$ is described in Section 5.1. Assume that there are q topics in the topic taxonomy.

$$NV_{d,e} = \langle NV_{d,e}^1, NV_{d,e}^2, \dots, NV_{d,e}^i, \dots, NV_{d,e}^q \rangle \quad (5.2)$$

Eq 5.2 defines the topic-needs variation vector between two time points, i.e., a specific time period. The variations of topic needs over consecutive time indexes are expressed as a time-period by topic matrix VM . An element $VM_{p,i}$ in the matrix represents the variation on topic i during time-period p (e.g. from time d to e). A row of the matrix, $VM[j]$ denotes a variation vector of topic needs. An example of topic-needs variation matrix is shown as follows:

	topic 1	topic 2	topic 3	topic 4	topic 5
2003 - 09 - 24 09 : 57 : 00	$\begin{bmatrix} 0.066 & -0.013 & -0.024 & -0.066 & 0.065 \\ 0.447 & 0.328 & 0.014 & 0.074 & 0.078 \\ 0.06 & 0.026 & 0.031 & 0.111 & 0.080 \\ 0.04 & 0.015 & 0.036 & 0.109 & 0.044 \\ 0.05 & -0.019 & 0.019 & 0.033 & 0.064 \end{bmatrix}$				
2003 - 10 - 07 18 : 25 : 42					
2003 - 10 - 13 14 : 13 : 00					
2003 - 10 - 13 14 : 18 : 00					
2003 - 10 - 13 14 : 25 : 00					
2003 - 10 - 14 14 : 49 : 30					

Fig. 5.1 An example of topic-needs variation matrix

The variation matrix shown in Fig. 5.1 is a 5x5 matrix. The variations of topic needs represented by this variation matrix spans from 2003-09-24 09:57:00 to 2003-10-14 14:49:30, and the value of each element in the variation matrix represents the variation on corresponding topic. For example, the value 0.447 represents the variation of topic needs on topic 1 from 2003-10-07 18:25:42 to 2003-10-13 14:13:00. Note that each time point shown in Fig. 5.1 represents the time when an event occurs, hence there are six events involved in the variation matrix of the worker.

Let the time index of 2003-09-24 09:57:00 be denoted as $t1$, and the time index of 2003-10-07 18:25:42 be denoted as $t2$. $NV_{t1,t2}^1=0.066$, which represents the variation of topic needs on topic 1 from time $t1$ to $t2$. $NV_{t1,t2} = \langle 0.066, -0.013, -0.024, -0.066, 0.065 \rangle$, which represents the variations of topic needs from time $t1$ to $t2$. The variations of topic needs over time are represented as set of topic-needs variation vectors.

6. Collaborative profile adaptation

This section illustrates the proposed collaborative profile adaptation approach. Section 6.1 describes the approach to identify similar workers according to the derived topic-variation matrix and self-adapted task profiles over a time window. Section 6.2 explains how to use similar workers' task needs to predict the target worker's potential task needs modeled as a collaborative profile. Section 6.2 also describes the integration of the derived collaborative profile with the self-adapted task profile to generate a new task profile representing the target worker's future task needs. The demonstrations of system implementation are described in Section 6.3.

6.1. Identifying similar workers

Each worker's variations of topic needs over time are expressed as a topic-needs variation matrix, i.e., a time-period by topic-needs matrix. Similar workers of the target worker are identified according to the workers' topic-needs variation matrices and self-adapted task profiles over a time window. Note that we focus on identifying similar workers with similar variations of topic needs and similar task needs. Thus, the self-adapted task profiles of the events involved in the variation process are considered to find workers with task needs similar to that of the target worker.

The details of the identification approach are illustrated as follows. Figure 6.1 shows the algorithm to identify similar workers. Table 6.1 lists the definitions of symbols used in the algorithm.

Table 6.1 The definitions of symbols used in the identification algorithm

q	the number of topics in the topic taxonomy
$\text{row}(VM)$	the number of rows in a variation matrix VM
$\text{col}(VM)$	the number of columns in a variation matrix VM ; Apparently, $\text{col}(VM)=q$
$VM^T(u_a)$	the variation matrix of the target worker u_a generated at time T
$VM^T(u_x)$	the variation matrix of the compared worker u_x generated at time T
W	the sliding window whose size is w , where $w \leq \text{row}(VM^T(u_a))$
$WM^T(u_a)$	the $w * q$ variation matrix generated from $VM^T(u_a)$ which only keeps the latest w variation vectors of u_a
$WM^T(u_x)$	the variation matrix of the compared worker u_x generated according

	to $VM^T(u_x)$ and the sliding window W ; T' is the latest time index within the window W
$WM^T(u_x)[j]/WM^T(u_a)[j]$	a variation vector of topic needs in $WM^T(u_x)/WM^T(u_a)$
$TP(u_x)[k]/TP(u_a)[k]$	a self-adapted task profile involved in $WM^T(u_x)/WM^T(u_a)$
$SimScore(u_x)$	the similarity score of u_x to the target worker u_a
$SimVM / SimTP$	the similarity calculated based on the variation matrix / task profile

Input:: $VM^T(u_a)$, W

Output:: SimilarWorkerList // the list of similar workers

```

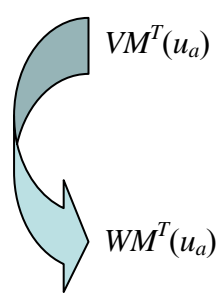
function FindSimilarWorker( $VM^T(u_a)$ ,  $W$ ){
1   Trim  $VM^T(u_a)$  to a  $w * q$  variation matrix  $WM^T(u_a)$ 
   // which only keeps the last  $w$  variation vectors;
2   foreach compared worker  $u_x$  {
3       Set  $SimScore(u_x) = 0$ 
4       Sliding the window  $W$  on  $VM^T(u_x)$  to derive  $WM^T(u_x)$ 
       from row 1 to row ( $row(M_y^T) - w + 1$ ), do {
5           Let  $WM^T(u_x)$  be the variation matrix of  $u_x$  generated according to  $VM^T(u_x)$ 
           and the sliding window  $W$ , when  $W$  is moving on  $VM^T(u_x)$ ;  $T'$  is the latest
           time index within the window  $W$ 
6           for the variation matrix  $WM^T(u_x)$  covered by  $W$ , do {
7               Set  $SimVM = 0$ ,  $SimTP = 0$ 
8               foreach variation vector  $WM^T(u_x)[j]$  of  $WM^T(u_x)$  do {
                   Let  $WM^T(u_a)[j]$  be the corresponding variation vector of  $M^T(u_a)$ 
9                    $SimVM = SimVM + similarity(WM^T(u_x)[j], WM^T(u_a)[j])$ 
10              }
11               $SimVM = SimVM / w$ 
12              foreach self-adapted task profile  $TP(u_x)[k]$  involved in  $WM^T(u_x)$  do{
                   Let  $TP(u_a)[k]$  be the corresponding task profile of  $WM^T(u_a)$ 
13                   $SimTP = SimTP + similarity(TP(u_x)[k], TP(u_a)[k])$ 
14              }
15               $SimTP = SimTP / (w+1)$ 
16              if ( $(\eta * SimVM + (1 - \eta) * SimTP) > SimScore(u_x)$ ) then{
17                   $SimScore(u_x) = \eta * SimVM + (1 - \eta) * SimTP$ 
18                  Set  $WM^T(u_x)$  as the candidate (similar) variation matrix of  $u_x$ 
19              }
20          }
21      }
22  }
23  Add the workers with top- $N$   $SimScore$  to SimilarWorkerList;
24  return SimilarWorkerList;

```

Fig. 6.1 Algorithm of identifying similar workers

First, the variation matrix $VM^T(u_a)$ of the target worker is trimmed to a $w*q$ variation matrix $WM^T(u_a)$ which keeps only the latest w variation vectors to be employed to identify similar workers that have similar variation matrix and task profiles. For each compared worker u_x , the sliding window W is employed accordingly to locate the part of variation matrix of u_x that is similar to $WM^T(u_a)$. $WM^T(u_x)$ is the variation matrix of u_x generated according to $VM^T(u_x)$ and the sliding window W . T' is the latest time index within the window W . In line 3 ~ line 21, the detailed procedure to find the candidate similar variation matrix for each compared worker u_x is described. The candidate variation matrix with the highest similarity score among all candidates of u_x is selected as the most similar variation matrix of u_x . The calculations (line 16 ~ line 19) of the similarity *SimScore* between u_x and u_a contains two parts, the calculation of similarity *SimVM* based on topic-needs variation vectors and the similarity *SimTP* based on the self-adapted task profiles. A parameter η is used here to balance the relative importance of *SimVM* and *SimTP*. In our application, we set $\eta=1/2$. That is, the similarity between variation vectors and self-adapted task profiles are equally important. The compared workers with top- N ranked similarity scores are selected as the similar workers of u_a . The value of N should be properly set according to the application domain

Figure 6.2 describes the trimming the variation matrix of the target worker according to the sliding window. Fig 6.3 and 6.4 illustrate the calculation of similarity score based on the variation matrices and task profiles within the sliding window.



	topic 48	topic 18	topic 27	topic 7	topic 47
2004 - 01 - 17 01 : 31 : 57	0.1089	0.031	0.5433	0.0083	0.1176
2004 - 01 - 17 01 : 32 : 03	0.3307	0.1126	0.5011	0.0247	0.1333
2004 - 01 - 17 01 : 38 : 43	- 0.497	- 0.1899	0.2381	- 0.0642	- 0.2211
2004 - 02 - 20 14 : 19 : 21	0.0662	- 0.0093	0.6623	0.0472	0.0948
2004 - 02 - 24 10 : 11 : 50	0.0298	- 0.0256	0.4398	0.0608	0.1446
2004 - 03 - 02 11 : 14 : 00					

	topic 48	topic 18	topic 27	topic 7	topic 47
2004 - 01 - 17 01 : 32 : 03	0.3307	0.1126	0.5011	0.0247	0.1333
2004 - 01 - 17 01 : 38 : 43	- 0.497	- 0.1899	0.2381	- 0.0642	- 0.2211
2004 - 02 - 20 14 : 19 : 21	0.0662	- 0.0093	0.6623	0.0472	0.0948
2004 - 02 - 24 10 : 11 : 50	0.0298	- 0.0256	0.4398	0.0608	0.1446
2004 - 03 - 02 11 : 14 : 00					

Fig. 6.2 Sample of trimming the variation matrix of the target worker

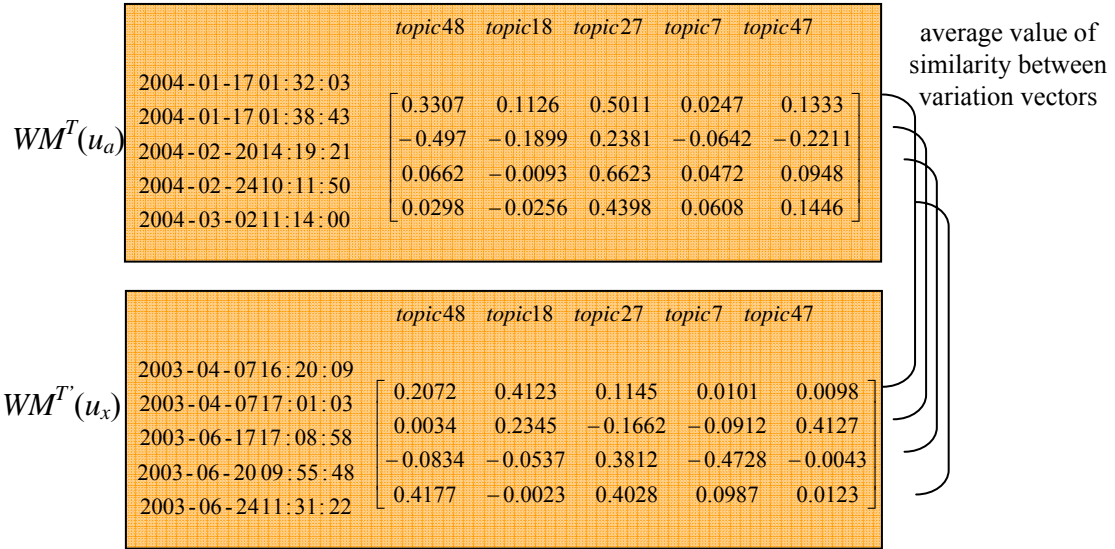


Fig. 6.3 Sample of calculating $SimVM$

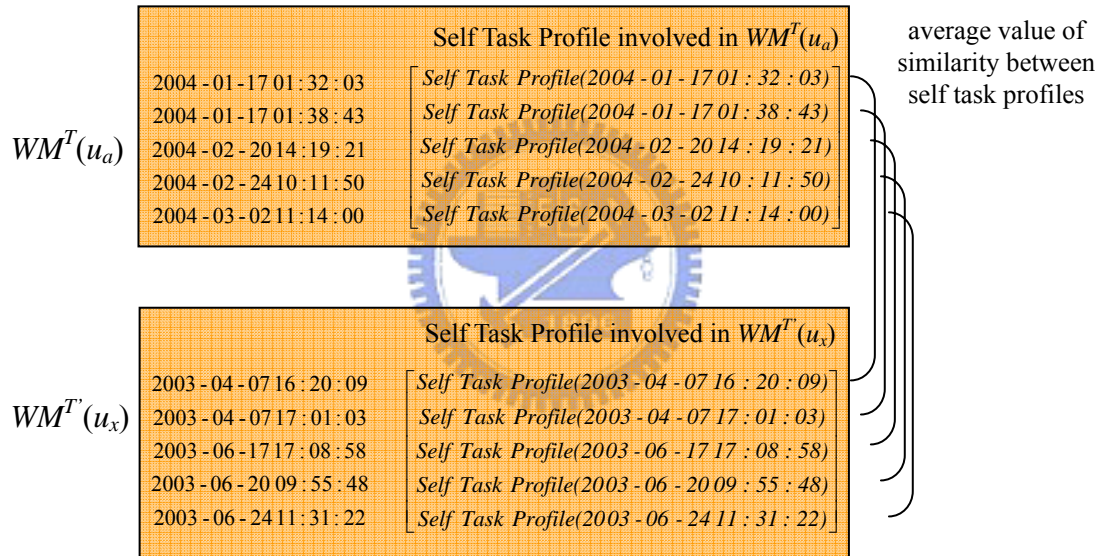


Fig. 6.4 Sample of calculating $SimTP$

6.2. Prediction of task needs

Once the similar workers of the target worker have been identified, their variation matrixes, i.e., candidate similar variation matrixes identified by the algorithm, can be used to predict the target worker's potential task needs. Note that time T'_{ux} is the latest time index in the candidate similar variation matrix. The variation vectors right after time T'_{ux} in the candidate similar variation matrixes can be regarded as the possible changes of topic needs that the target worker will experience in the near future. These variation vectors of similar workers are used to derive a vector-based variation profile, which can then be integrated with the

self-adapted task profile to predict the target worker's potential task needs accordingly. Alternatively, the documents accessed by similar workers right after time T'_{ux} , i.e., $T'_{ux}+1$, can also be regarded as the future potential information needs of the target worker. Thus, two approaches are proposed to predict the worker's potential task needs modeled as a collaborative profile which is derived according to the similar workers' behaviors at time $T'_{ux}+1$. One is based on the variation of topic needs to derive the collaborative profile, in which similar workers' variations of topic needs from time T'_{ux} to $T'_{ux}+1$ are used to predict the target worker's possible variations of topic needs at time $T+1$. The alternative is based on the documents accessed at time $T'_{ux}+1$, where similar workers' documents accessed at time $T'_{ux}+1$ are used to derive the collaborative profile. A linear combination approach is used to integrate the derived collaborative profile with the self-adapted task profile to generate a new task profile \bar{P}_{T+1} representing the target worker's future task needs, as given below. The detailed approaches are as follows:

$$\bar{P}_{T+1} = \begin{cases} \delta_V \times \bar{S}_{T+1} + (1 - \delta_V) \times \frac{\sum_{u_x \in U_a} Sim_T(u_a, u_x) \times \sum_{i=1}^q NV_{T'_{ux}, T'_{ux}+1}^i(u_x) * \bar{O}_i}{\sum_{u_x \in U_a} Sim_T(u_a, u_x)} & (TP-V \text{ method}) \\ \delta_D \times \bar{S}_{T+1} + (1 - \delta_D) \times \frac{\sum_{u_x \in U_a} Sim_T(u_a, u_x) \times \bar{D}_{T'_{ux}+1}(u_x)}{\sum_{u_x \in U_a} Sim_T(u_a, u_x)} & (TP-D \text{ method}) \end{cases}$$

where,

\bar{S}_{T+1}	the self-adapted task profile generated at time T , described in Section 4.2
u_a / u_x	the target worker / similar worker
U_a	the set of similar workers of u_a
δ_V / δ_D	the tuning parameters used to adjust the relative weight between the self-adapted task profile and the collaborative profile
T'_{ux}	the latest time index of candidate similar variation matrix $WM^{T'}(u_x)$ of worker u_x
$Sim_T(u_a, u_x)$	the similarity between the target worker u_a and the worker u_x at time T , which is obtained from the $SimScore(u_x)$, described in Section 6.1
$NV_{T'_{ux}, T'_{ux}+1}^i(u_x)$	the variation degree on topic i from time T'_{ux} to time $T'_{ux}+1$ of u_x , as described in Section 5.2

$\bar{D}_{T_{u_x}'+1}(u_x)$	the document profile of the document D accessed at time $T_{u_x}'+1$ by the worker u_x
\bar{O}_i	the topic profile of topic i in the topic taxonomy

For the TP-V method, the predicted profile of the target worker is the weighted combination of the self-adapted task profile and the accumulated collaborative topic-variation profile derived from similar workers' variations of topic needs. A parameter δ_v is used here to adjust their relative weights. Each worker's topic-variation profile is obtained through the profiles of topics multiplied by the corresponding variation degrees $NV_{T_{u_x}', T_{u_x}'+1}^i(u_x)$. The individual topic variation profile represents the variation of topic needs of the corresponding worker, while the collaborative topic-variation profile represents similar workers' weighted topic-variation profile by considering their similarity to the target worker. The collaborative topic-variation profile is integrated with the self-adapted task profile of the target worker to generate predicted task profile representing his potential task needs in the near future. The TP-D method is similar to the TP-V method, except that the documents accessed at time $T_{u_x}'+1$ by similar workers are used to derive the collaborative document profile. The comparisons of these two methods are described in Section 7.

6.3. Result demonstrations

For the purpose of reviewing and evaluation, an interface is implemented to show the detailed information after each operation of proposed methodology mentioned from Section 4. The individual operation is demonstrated respectively thereafter.

(1)Interface one: Information of task profile and variation process

Fig. 6.5 shows the event information during the adaptation process of the self-adapted task profile of worker "Nancy" from 2003-09-24 09:45:00 to 2005-04-21 13:27:00. For event 2, it shows that Nancy accessed the document labeled as D0000000339 at 2003-09-24 09:52:00. Fig 6.6 shows the similarity with all topics for each event. The topics are properly labeled and the similarity is obtained by calculating the cosine value between the topic profile and the document profile. From the figure, which topic is interesting for the worker can be grasped roughly.

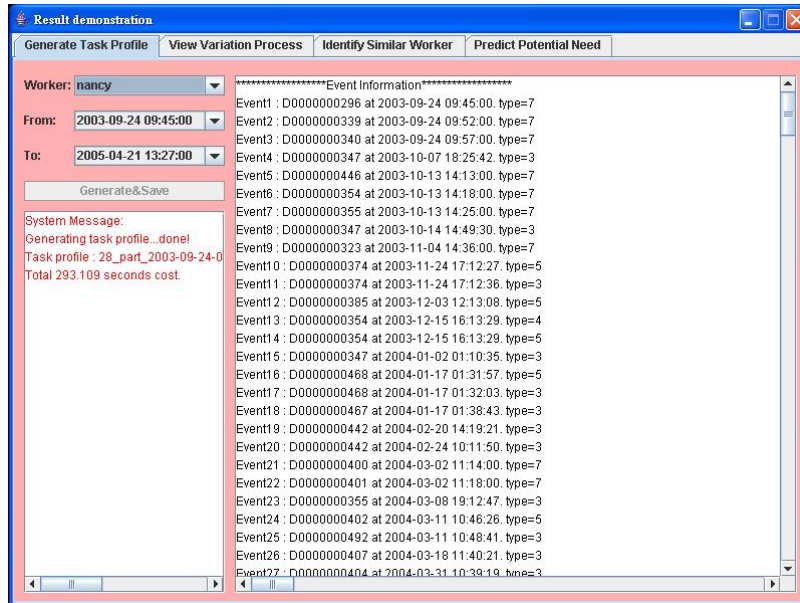


Fig. 6.5 Event information

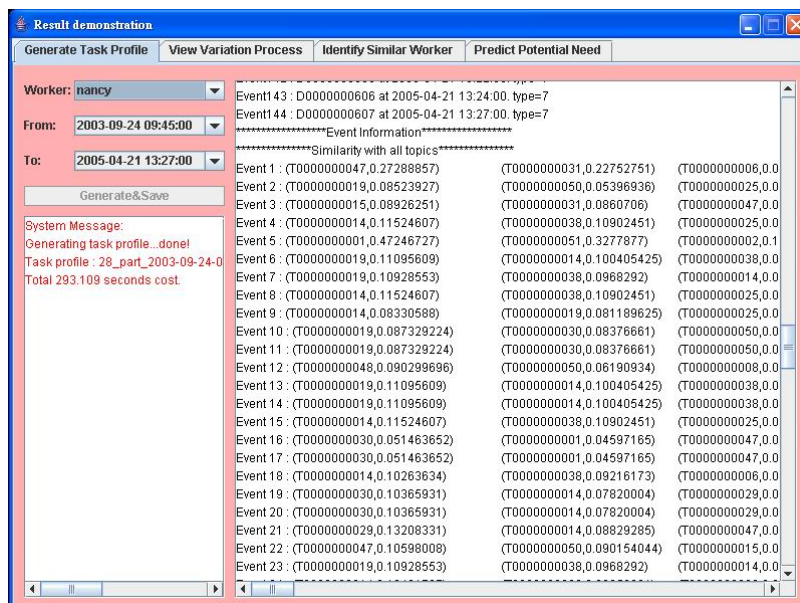


Fig. 6.6 Similarity with each topic

Fig. 6.7 shows the information of the variation process of Nancy from 2003-09-24 09:45:00 to 2005-04-21 13:27:00. Each row with the preceding words "Variation X" represents a variation vector, and the string in bracket represents the variation degree on corresponding topic during some period. From the result, the trend of variations on topics can be understood in advance.

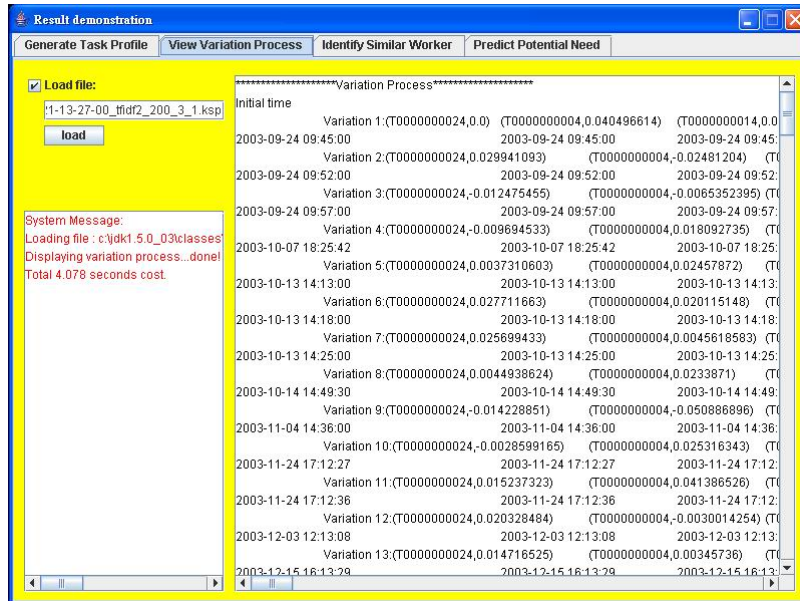


Fig. 6.7 Variation of topic needs over time

(2) Interface two: Identification of similar workers

Fig. 6.8 shows the information of similar workers after performing proposed identification approach. The top 2 similar workers are taken into account and are shown here. For the worker "Christine", the most similar variation process with what Nancy recently get interested around 2005-04-21 13:27:00 involves from event 74 to event 83 and it could be traced in advance to see the details by observing the variation process or event information mentioned above.

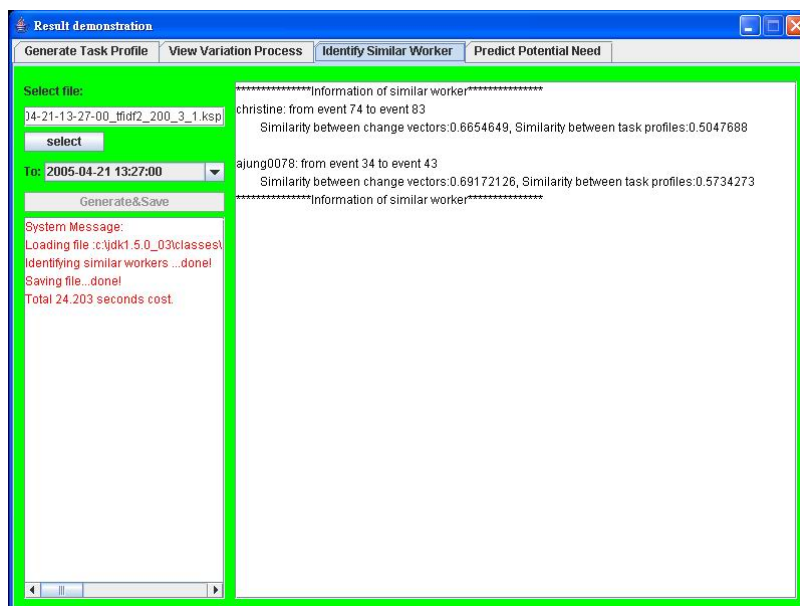


Fig. 6.8 Similar Workers

(3) Interface three: Prediction of potential task needs

Fig. 6.9 shows the prediction result of three proposed prediction approaches in this research. In Fig. 6.9, twenty knowledge documents are chosen in the order of similarity to be recommended for Nancy. These documents could be reviewed on-line by checking the titles to see if there are some documents helpful for the execution of task, and then the user may choose some to read in advance.

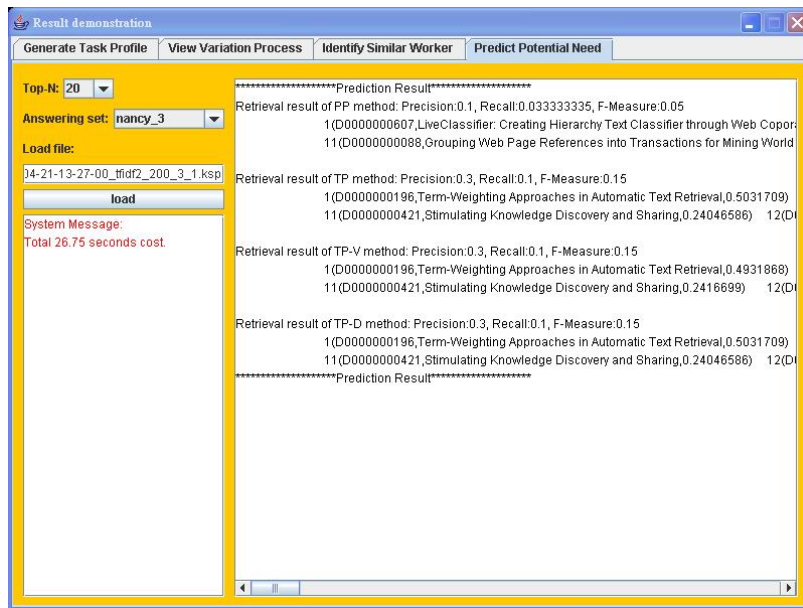


Fig. 6.9 Prediction Result

7. Experiments

Three experiments are conducted to evaluate the effectiveness of the proposed profile adaptation methods. Section 7.1 describes the experimental objective, design, resource and evaluation metrics. The experimental results and discussions are given in Section 7.2.

7.1. Experimental setup

7.1.1. Experimental objective and design

Experiments are conducted to evaluate the effectiveness of the proposed adaptive profiling techniques, including the *TP* method, *TP-V* method and the *TP-D* method. The *TP* method denotes the self profile adaptation method described in Section 4.2, which adjusts the worker's task profile based on the document accessed by the worker and the relevance on topic taxonomy. The effect of time factor is also incorporated into the profile adaptation process, as given in E.q. 4.1 and E.q. 4.2. The *TP-V* method denotes the collaborative profile adaptation method described in Section 6.2, which adjust the task profile by a weighted combination of the self-adapted task profile and the collaborative profile (topic-variation profile) derived from similar workers. The *TP-D* method is similar to the *TP-V* method, except that the documents accessed at time $T_{u_x}'+1$ by similar workers are used to derive the collaborative profile. Parameters δ_V and δ_D are used to adjust the relative weights of self profile and collaborative profile in *TP-V* and *TP-D* methods, respectively. Thereby, the parameters of δ_V and δ_D are determined in experiment one to select the best value of parameters in our application domain.

Experiment two compares the *TP* method with a baseline method, which is called the *PP* method - *primitive profiling* method, and two self profile adaptation methods, *PP-T* and *PP-P* methods. The *PP* method generates and adjusts the target worker's task profile based on the documents accessed, without considering topic profiles and the effect of time factor. The *PP-T* and *PP-P* methods act almost like the *PP* method, but they consider the effect of time factor and topic profiles respectively. The equation of the *PP*, *PP-T*, and *PP-P* methods are given as follows:

$$\begin{cases} \bar{S}_{T+1} = \alpha \bar{S}_T + \bar{D}_T & (PP \text{ method}) \\ \bar{S}_{T+1} = \alpha \text{Decay}(\bar{S}_T) + \bar{D}_T & (PP-T \text{ method}) \\ \bar{S}_{T+1} = \alpha \bar{S}_T + [\lambda \bar{O}_T + (1-\lambda) \bar{D}_T] & (PP-P \text{ method}) \end{cases}$$

where α is set to 1 in the experiments.

Experiment three compares *TP*, *TP-V*, and *TP-D* methods with the baseline method, *PP* method.

7.1.2. Data and participants

Experiments are conducted using a real application domain on conducting research tasks in a laboratory of a research institute. Knowledge workers usually require a longer time (e.g. one year) to accomplish knowledge-intensive tasks, so the sample size of the data and participants is restricted in the experiments.

In this work, the tasks concerned are writing research papers or conducting research projects. Thirteen workers were selected as the compared workers and ten workers were selected as the tested workers in the evaluation. Fifty five research tasks were collected, with 36 topics and 19 target tasks (the tasks at hand). Because the process of each task performance spans a long time period, the evaluation period of each target task is selected by the experts to examine the proposed methods.

We evaluate the four methods by examining the effectiveness of the retrieval result based on documents accessed by conducting tasks. Over 600 documents are collected during the period of 2002~2005. Each document contains an average of ninety distinct terms after information extraction, and document pre-processing (e.g. stemming, removing stop words, and indexing).

7.1.3. Evaluation metrics

The effectiveness of each method is measured in terms of precision, recall and F-measure, which are widely used measures in information retrieval [3].

Precision and Recall: Precision is the fraction of retrieved documents that are relevant, while recall is the fraction of known relevant documents that are retrieved. Known relevant documents of a worker u are those relevant documents selected by u in the test set. The definitions are given below:

$$precision = \frac{\text{number of retrieved documents that are relevant}}{\text{number of retrieved documents}}$$

$$recall = \frac{\text{number of relevant documents that are retrieved}}{\text{number of known relevant documents}}$$

F-Measure: For observing the relative importance of precision and recall, a combination metric F1-metric [29] is used to adjust the relative weight of precision and recall to balance the trade-off between these two indicators.

$$F_{\beta} = \frac{(1 + \beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$$

The function of β is to adjust the relative importance of the recall in comparison to the precision. If $\beta=0$, F_{β} coincides with precision, and if $\beta=\infty$, F_{β} coincides with recall. In this experiment, we set $\beta=1$ (precision and recall are equally important).

7.2. Experimental result and observations

7.2.1. Experiment 1-1: parameter selection for *TP-V* method

This experiment aims to determine the value of parameter δ_v in the *TP-V* method. *TP-V* method uses δ_v to adjust the relative weight of self task profile and collaborative topic-variation profile. When δ_v is set to 1, the *TP-V* method is equivalent to *TP* method, which takes only the self task profile into account. When δ_v is set to 0, it is equivalent to the collaborative topic-variation profile. The experiment was conducted by systematically adjusting the value of δ_v in an increment of 0.1. The *precision* metric was chosen as the performance measure to evaluate the effectiveness of the methods. The optimal parameter values with the best results (the highest precision values) were chosen as the parameter settings of the proposed equations. Table 7.1 shows the performance of *TP-V* with different δ_v value in terms of precision under various top-N supported documents.

Observation: Table 7.1 shows that the average precision value of *TP-V* method with $\delta_v = 0.7$ exceeds those with the other values. Meanwhile, while setting $\delta_v = 0.7$ in the given equation of *TP-V* method, it can achieve the best performance under Top-5,

Top-10, Top-15, or Top-20 document support. The $TP-V$ method is increasing dramatically from $\delta_v = 0$ to 0.7 and is decreasing slightly from $\delta_v = 0.7$ to 1, as shown in Fig. 7.1. The experimental result reveals that self task profile is more important than the collaborative topic-variation profile in the $TP-V$ method.

Table 7.1 Effectiveness of $TP-V$ method under various δ_v values

δ_v \ Top-N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Top-5	0.046	0.185	0.2	0.215	0.338	0.338	0.354	0.369	0.369	0.338	0.354
Top-10	0.062	0.138	0.192	0.292	0.323	0.362	0.385	0.408	0.392	0.385	0.385
Top-15	0.056	0.138	0.215	0.262	0.308	0.354	0.369	0.385	0.359	0.359	0.359
Top-20	0.065	0.146	0.208	0.258	0.304	0.362	0.354	0.358	0.327	0.323	0.319
Average	0.057	0.152	0.204	0.257	0.318	0.354	0.365	0.380	0.362	0.351	0.354

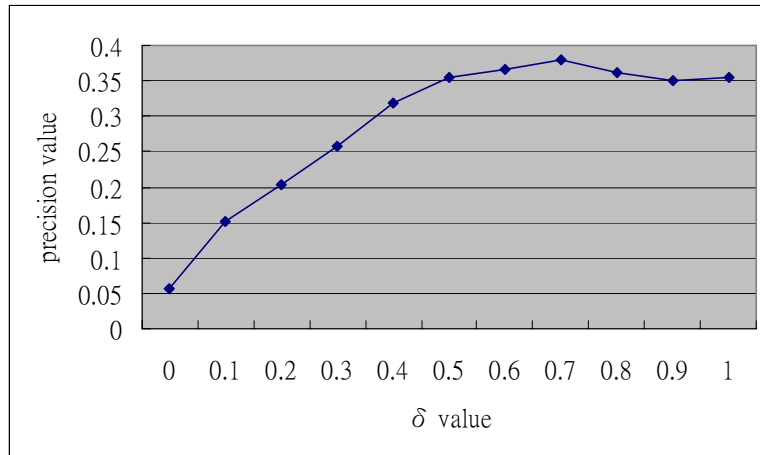


Fig. 7.1 Result of knowledge support of $TP-V$ under various δ_v value

7.2.2. Experiment 1-2: parameter selection for $TP-D$ method

This experiment aims to determine the value of parameter δ_D in the $TP-D$ method. $TP-D$ method uses δ_D to adjust the relative weight of self task profile and collaborative document profile. When δ_D is set to 1, the $TP-D$ method is equivalent to the TP method, which takes only the self task profile into account. When δ_D is set to 0, it is equivalent to the collaborative document profile. The experiment was also conducted by systematically adjusting the value of δ_D in an increment of 0.1. The *precision* metric was chosen as the performance measure to evaluate the effectiveness of the methods. The optimal parameter values with the best results (the highest precision values) were chosen as the parameter settings of the proposed equations.

Table 7.2 shows the performance of *TP-D* method with different δ_D value in terms of precision under various top-N supported documents.

Observation: Table 7.2 shows that the average precision value of *TP-D* method with $\delta_D = 0.5$ has the best performance (i.e., precision value) than the other values. Interestingly, the result shows that if we set $\delta_D = 0.5$, $\delta_D = 0.8$, $\delta_D = 0.9$, or $\delta_D = 1$, they all have similar results. Thus, the curve of *TP-D* method shown in Fig. 7.2 is smooth and steady from $\delta_D = 0.4$ to 1. The result indicates the collaborative profile of *TP-D* method has no significant influence to this experiment. We may take a further analysis in the Experiment 2.

Table 7.2 Effectiveness of the *TP-D* method under various δ_D values

δ_D Top-N	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Top-5	0.138	0.231	0.292	0.323	0.323	0.338	0.338	0.338	0.354	0.354	0.354
Top-10	0.185	0.262	0.285	0.346	0.369	0.385	0.385	0.392	0.385	0.385	0.385
Top-15	0.169	0.231	0.303	0.354	0.379	0.369	0.354	0.354	0.359	0.359	0.359
Top-20	0.162	0.25	0.308	0.331	0.346	0.331	0.319	0.319	0.315	0.315	0.319
Average	0.163	0.243	0.297	0.338	0.354	0.356	0.349	0.351	0.353	0.353	0.354

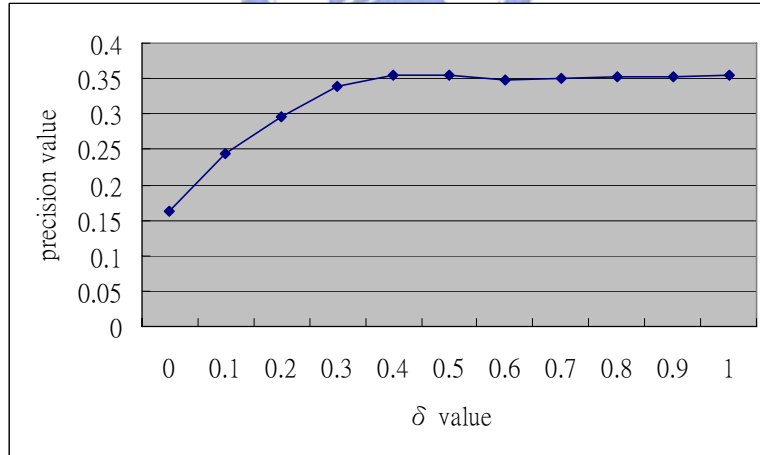


Fig. 7.2 Result of knowledge support of *TP-D* under various δ_D value

7.2.3. Experiment 2: comparisons of self profile adaptation methods

This experiment aims to compare the performance of task-relevant document support between the four methods: *PP*, *PP-T*, *PP-P*, and *TP* under various top-N supported documents. *PP* method is the baseline method as described in Section 7.1.1, which solely adjusts task profile based on the document profiles of documents accessed, but *PP-T* and *PP-P* methods consider the time factor and topic profiles respectively. The *TP* method is the self profile adaptation method which adjusts task

profile based on the document profile, topic profiles and time effect. Herein, Table 7.3 shows the performance of the four methods in terms of precision, recall and F-measure under various top-N documents.

Observation 1: Table 7.3 and Fig. 7.3 shows that the average values of precision, recall, and F-measure of proposed *PP-T*, *PP-P*, and *TP* methods are better than those of the baseline method, *PP* method, under various top-N retrievals. The result reveals that it is effective to consider the topic profiles and the time factor during the profile adaptation process.

Observation 2: Fig. 7.3 shows the curve of each method under various top-N supported documents. Notice that *TP* method is better than the *PP-T* and *PP-P* methods under various top-N supported documents. Consequently, the result implies that the *TP* method is more effective than the other three methods, by considering both the topic profiles and the time factor during the profile adaptation process.

Table 7.3 Comparison between self profile adaptation methods

method \ Top-N	PP method			PP-T method			PP-P method			TP method		
	Pre.	Re.	F.	Pre.	Re.	F.	Pre.	Re.	F.	Pre.	Re.	F.
Top-5	0.215	0.030	0.051	0.246	0.035	0.060	0.215	0.022	0.040	0.354	0.041	0.073
Top-10	0.215	0.055	0.085	0.269	0.068	0.105	0.262	0.055	0.090	0.385	0.092	0.145
Top-15	0.215	0.083	0.115	0.277	0.105	0.146	0.287	0.097	0.140	0.359	0.132	0.186
Top-20	0.204	0.102	0.129	0.277	0.138	0.175	0.277	0.125	0.166	0.319	0.155	0.200
Average	0.212	0.067	0.095	0.267	0.086	0.121	0.260	0.075	0.109	0.354	0.105	0.151

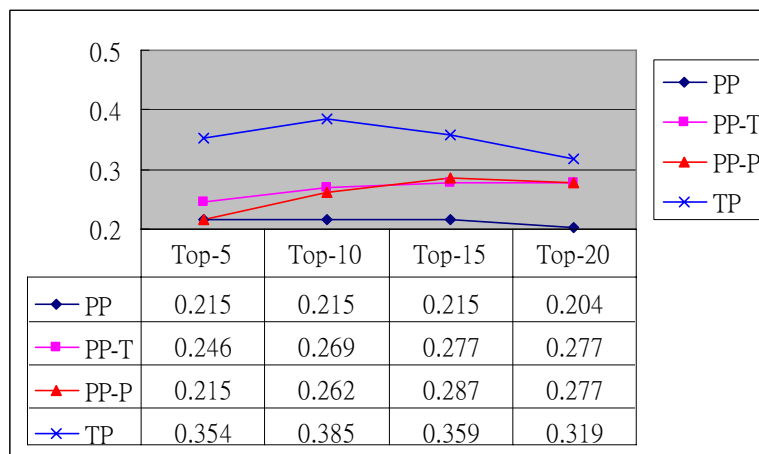


Fig. 7.3 Trends of retrieval effectiveness of the four methods under various top-N document support

7.2.4. Experiment 3: comparisons of various methods

This experiment aims to compare the performance of task-relevant document support between the four methods: *PP*, *TP*, *TP-V* (with $\delta_v = 0.7$), and *TP-D* (with $\delta_d = 0.5$) under various top-*N* supported documents. *PP* method is the baseline method as described in Section 7.1.1, which solely adjusts task profile based on the document profiles of documents accessed. The *TP* method is the self profile adaptation method which adjusts task profile based on the document profile, topic profiles and time effect. The *TP-V* method and *TP-D* method further consider the effect of collaborative profiles generated from similar workers. The parameters δ_v and δ_d are used to adjust the relative importance of the worker's self task profile and the collaborative profile in *TP-V* method and *TP-D* method, respectively. According to the result of experiment 1-1 and 1-2, the $\delta_v = 0.7$, and $\delta_d = 0.5$ can achieve the best performance. Herein, Table 7.4 shows the performance of the four methods in terms of precision, recall and F-measure under various top-*N* documents.

Observation 1: Table 7.4 shows that the average values of precision, recall, and F-measure of proposed *TP-V*, and *TP-D* methods are far better than those of the baseline method, *PP* method, under various top-*N* retrievals. The result reveals that it is effective to consider the collaboration from similar workers during the profile adaptation process.

Observation 2: Fig. 7.4 shows the curve of each method under various top-*N* supported documents. The *TP-V* method performs slightly better than the *TP* and *TP-D* methods. Collaborative adaptation of task profiles based on similar workers' topic variations is effective to improve the quality of document retrieval.

Table 7.4 Comparison between methods

method \ Top-N	PP method			TP method			TP-V method			TP-D method		
	Pre.	Re.	F.	Pre.	Re.	F.	Pre.	Re.	F.	Pre.	Re.	F.
Top-5	0.215	0.030	0.051	0.354	0.041	0.073	0.369	0.045	0.079	0.338	0.040	0.070
Top-10	0.215	0.055	0.085	0.385	0.092	0.145	0.408	0.098	0.153	0.385	0.091	0.143
Top-15	0.215	0.083	0.115	0.359	0.132	0.186	0.385	0.144	0.200	0.369	0.135	0.190
Top-20	0.204	0.102	0.129	0.319	0.155	0.200	0.358	0.181	0.229	0.331	0.160	0.206
Average	0.212	0.067	0.095	0.354	0.105	0.151	0.380	0.117	0.165	0.356	0.107	0.152

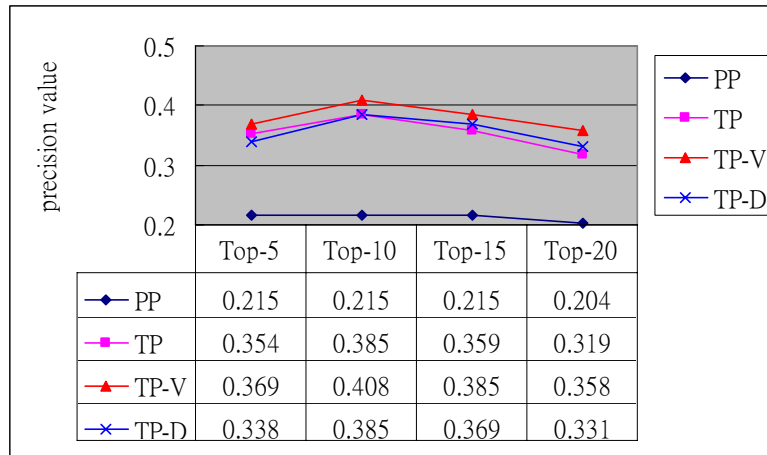


Fig. 7.4 Trends of retrieval effectiveness of the four methods under various top-N document support

7.2.5. Case inspections

The overall experimental results demonstrate that the *TP-V* method performs slightly better than the other methods with $\delta_V = 0.7$; that is, the self-adapted profile gets higher weight than the collaborative profile in the *TP-V* method. We conduct further inspections of each case and find that some cases have significant improvement of retrieval effectiveness. Fig. 7.5 shows three cases (case 1, 2, and 3) with large degree of topic-needs variation and one normal case (case 4) with small degree of topic-needs variation.

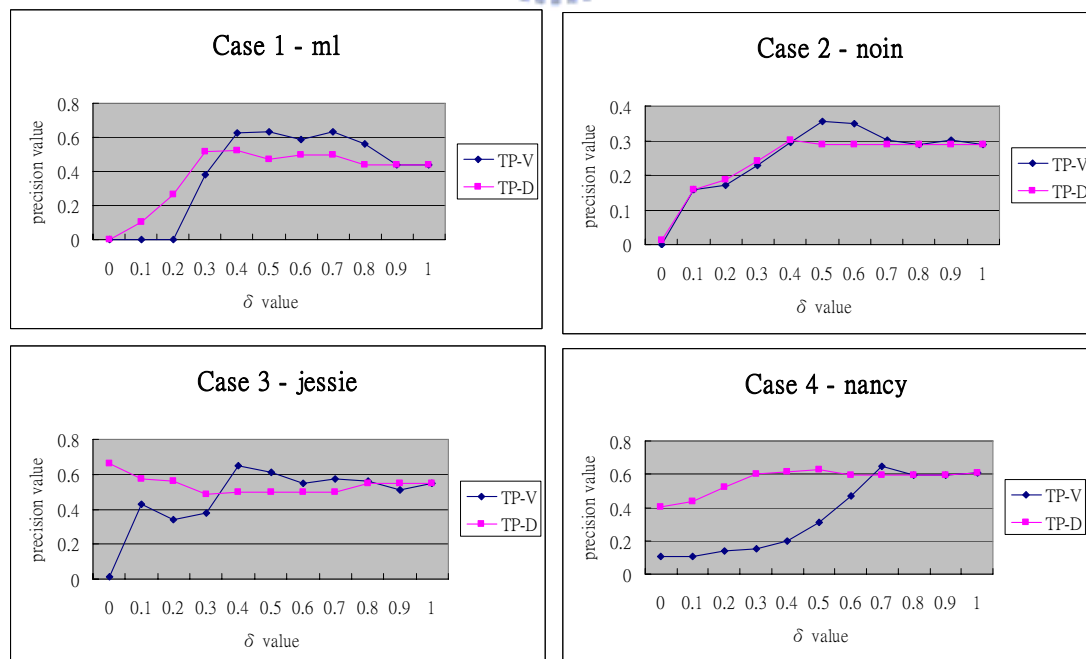


Fig. 7.5 Four experimental cases

Observation 1: Fig. 7.5 shows that the precision value of the four cases using proposed *TP-V* and *TP-D* methods under various δ values. For the *TP-V* method, cases 1, 2, and 3 perform better under $\delta = 0.4$ or $\delta = 0.5$; that is, the collaborative profile gets higher weight than the self-adapted profile ($\delta = 1$). The normal case (case 4) has better performance under $\delta = 0.7$ for *TP-V* method, and has better performance under $\delta = 0.3\sim 1$ for *TP-D* method. The result reveals that it is more effective to give higher weight on the collaborative profile during the profile adaptation process for cases with large degree of topic-needs variation, but more effective to give higher weight on the self-adapted profile for the normal case.

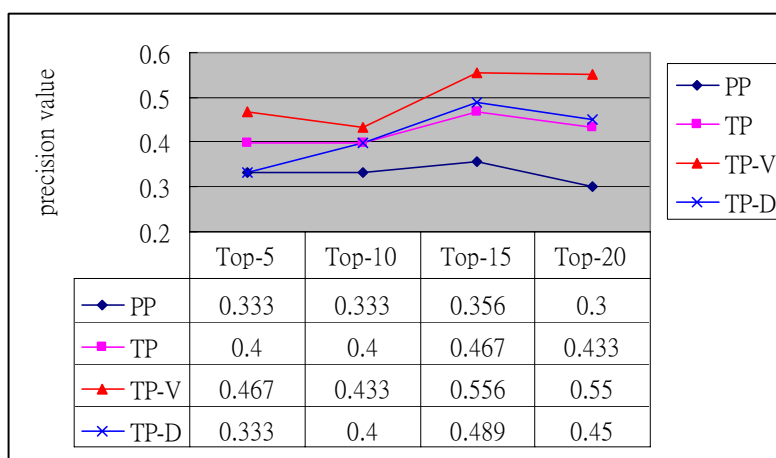


Fig. 7.6 Trends of retrieval effectiveness of the three cases (case 1, 2,3) under various top-N document support

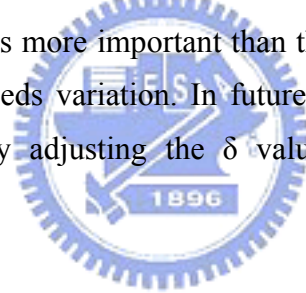
Observation 2: Fig. 7.6 shows the average retrieval effectiveness of the three cases (case 1, 2, and 3) under various top-N document supports. The curve shows that *TP-V* method performs better than *PP*, *TP* and *TP-D* methods under various top-N supported documents. Consequently, the *TP-V* method considering collaborative adaptation based on topic-needs variation is more effective than the other three methods, especially for the cases which have large degree of topic needs variation.

7.2.6. Discussions

The overall experimental results demonstrate that the proposed *novel adaptive task-profiling technique* is effective. Generally, the more recent the document accessed the more important it is to reflect a work's current task needs. Thus, the time factor is important for profile adaptation process. Experimental result shows that the proposed *adaptive task-profiling techniques* perform better than the baseline method - *PP* method which does not consider topic taxonomy and time effect.

Moreover, the proposed profiling technique adopts a novel collaborative profile adaptation approach to adjust task profiles. We analyze the variations of workers' task needs on the topic taxonomy to identify workers with similar variations of topic needs over time. Similar workers' variations of topic needs are used to predict the target worker's future variations of topic needs, and are used to adjust the target worker's task profile. Accordingly, the *TP-V* and *TP-D* methods are proposed to examine the effect of the profiling via collaboration. The result is interesting, and we summarized the results with the associated discussions below.

Generally, the *TP-V*, *TP-D*, and *TP* methods have similar results. That is, the similar workers' variations of topic needs or document access can be adopted to predict the target worker's future topic needs but can only slightly improve the retrieval effectiveness of self task profile (*TP* method). However, for workers with large degree of topic-needs variation, collaborative adaptation of task profile based on similar workers' topic-needs variation can gain significant improvement of the retrieval effectiveness. From the result demonstrated in Section 7.2.5, we can infer that the collaborative profile is more important than the self task profile for the cases with large degree of topic-needs variation. In future work, we plan to improve the *TP-V* and *TP-D* methods by adjusting the δ value according to the degree of task-needs variation.



8. Conclusion and future research issues

In this work, we propose a self profile adaptation approach to model the task needs of the workers by considering the knowledge activities of workers and the effect of time factor. Moreover, the variations of topic needs over time are also measured by the proposed approach. With the help of topic needs variation, we also propose approaches to find the similar workers by identifying similar topic needs variation over time, and to predict potential task needs with combination of the collaborative profile, i.e., the topic-variation profiles in *TP-V* method and the document profiles in *TP-D* method. In addition, an interface is implemented to show the detailed information after each operation of proposed methodology mentioned in this work, and two experiments are conducted. From the experimental results, we find that the proposed enhancement of adaptation approach is effective. The *TP* method which represents the proposed event-based self profile adaptation approach also performs well in the experiments. However, during the period of conducting experiments, we also find there are some limits in our methodology. As a result, these issues are addressed as follows to be further investigated in the future.

(1)Improvement of profiling approach: In this work, our proposed self profile adaptation approach is event-based approach in which a document access triggers the adaptation process. An event-based approach may result in overreaction because workers may not always access the documents they really need. In order to ease the situation of overreaction, a transaction-based profiling approach, which considers documents accessed within a period as a transaction, is worth to be explored. The time weight used to reflect the significance of time variance during the execution of task can be computed by adopting the time cancroids of transactions.

(2)Refinement of topic taxonomy: The topic taxonomy used in this work is inherited from previous research [19], which consists of thirty-six topics. Nevertheless, we find some cases conducting the opposite result to that in experiment 1-1. By using the interface to observe the details, we find that the similar workers identified by our proposed approach can not provide enough support for them, since their relevant topics are not contained in the topic taxonomy. For this reason, the refinement of topic taxonomy continuously with emerging topics is important to model workers' task needs.

Reference

- [1] A. Abecker, A. Bernardi, K. Hinkelmann, O. Kuhn, M. Sintek, "Context-aware, proactive delivery of task-specific information: the KnowMore Project," *Information Systems Frontiers*, 2(3/4), pp.253-276, 2000a.
- [2] A. Abecker, A. Bernardi, H. Maus, M. Sintek, C. Wenzel, "Information Supply for Business Processes: Coupling Workflow with Document Analysis and Information Retrieval," *Knowledge Based Systems*, 13(1), pp.271-284, 2000b.
- [3] R. Baeza-Yates, B. Ribeiro-Neto, "Modern Information Retrieval," *New York: The ACM Press*, 1999.
- [4] M. Balabanovi'c, "An Adaptive Web Page Recommendation Service," *Proceedings of the First International conference on Autonomous Agents*, Marina del Rey, CA, pp.378-385, February, 1997.
- [5] N.J. Belkin, W.B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Communications of the ACM*, 35(12), pp.29-38, 1992.
- [6] P. D. Bra, G. J. Houben, F. Dignum, "Task-Based Information Filtering: Providing Information that is Right for the Job," *Proceedings of INFWET97*, <http://wwwis.win.tue.nl/infwet97/proceedings/task-based.html>, 1997.
- [7] J.S. Brown, P. Duguid, "Organization Learning and Communities of Practice," *Organization Science*, 2(1), pp.40-57, 1991.
- [8] A. Celentano, M. G. Fugini, S. Pozzi, "Knowledge-based document retrieval in office environment: The Kabiria system," *ACM Transactions on Information Systems*, 13(3), pp.237-268, 1995.
- [9] T. H. Davenport, L. Prusak, "Working Knowledge: How Organizations Manage what they know," *Boston MA: Harvard Business School Press*, 1998.
- [10] J. Davies, A. Duke, A. Stonkus, "OntoShare: Using Ontologies for Knowledge Sharing," *Proceedings of the 11th International WWW Conference WWW2002*, Hawaii, USA, 2002.
- [11] K. D. Fenstermacher, C. Marlow, "Supporting Consultants with Task-Specific Information Retrieval," *Proceedings of The American Association of Artificial Intelligence*, Orlando Florida: AAAI Press, 1999.
- [12] K. D. Fenstermacher, "Process-Aware Knowledge Retrieval," *Proceedings of the 35th Hawaii International Conference on System Sciences*, Big Island, Hawaii, USA, pp.209-217, 2002.
- [13] G. Fischer, J. Ostwald, "Knowledge Management: Problems, Promises, Realities, and Challenges," *IEEE Intelligent Systems*, 16(1), pp.60-73, 2001.
- [14] D. Goldberg, B. M. Nichols, Oki, D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM*, 35(12), pp.61-70, 1992.
- [15] P. H. Gray, "The Impact of Knowledge Repositories on Power and Control in the Workplace," *Information Technology & People*, 14(4), pp.368-384, 2001.

- [16] J. Hahn, M. Subramani, " A Framework of Knowledge Management Systems: Issues and Challenges for Theory and Practice," *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Australia, pp.302-312, 2000.
- [17] A. Kankanhalli, F. Tanudidjaja, J. Sutanto, C.Y. Tan (Bernard), "The Role of IT in Successful Knowledge Management Initiatives," *Communications of ACM*, 46(9), pp.69-73, 2003.
- [18] J. Koh, Y.-G. Kim, "Knowledge Sharing in Virtual Communities: An e-Business Perspective," *Expert Systems with Applications*, 26, pp.155-166, 2004.
- [19] D.-R. Liu, I.-C. Wu, and K.-S. Yang (2005), "Task-based \mathcal{K} -Support System: Disseminating an Sharing Task-relevant Knowledge," *Expert Systems with Applications*, 29(2), pp.408-423, 2005.
- [20] D. W. McDonald, M.S. Ackerman, "Expertise Recommender: A Flexible Recommendation System and Architecture," *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work*, Philadelphia, PA, pp.231-240, 2000.
- [21] S. E. Middleton, N. R. Shadbolt, D. C. Roure, "Ontological User Profiling in Recommender Systems," *ACM Transaction on Information Systems*, 22(1), pp.54-88, 2004.
- [22] J. Mostafa, S. Mukhopadhyay, W. Lam, M. Palakal, "A Multi-level Approach to Intelligent Information Filtering: Model, System and Evaluation," *ACM Transactions on Information Systems*, 15(4), pp.368-399, 1997.
- [23] S. Mukhopadhyay, J. Mostafa, M. Palakal, "An Adaptive Multi-level Information Filtering System," *Proceedings of the Fifth International Conference on User Modeling*, pp.21-28, 1996.
- [24] I. Nonaka, "A Dynamic Theory of Organizational Knowledge Creation," *Organization Science*, 5(1), pp.14-37,1994.
- [25] J. Park, S. Hunting, "XML Topic Maps: Creating and using Topic Maps for the Web," *Boston MA: Addison-Wesley*, 2003.
- [26] M. Pazzani, J. Muramatsu, D. Billus, "Syskill & Webert: Identifying Interesting Web Sites," *Proceedings of the Thirteen National Conference on Artificial Intelligence*, Portland, Oregon: AAAI Press., pp.54-61, 1996.
- [27] M. Pazzani, D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," *Machine Learning* 27, pp.313-331, 1997.
- [28] P. Resnick, H.R. Varian, "Recommender Systems," *Communications of the ACM*, 40(3), pp.56-58, 1997.
- [29] C. J. v. Rijsbergen, "Information Retrieval," *Butterworths*, London, 1979.
- [30] J. J. Rocchio, "Relevance Feedback in Information Retrieval," in: G. Salton (Eds), *The SMART Retrieval System: Experiments in Automatic Document Processing*, Englewood Cliffs, NJ: *Prentice Hall*, pp.313-323, 1971.

- [31] T. Rodden, "A survey of CSCW Systems," *Interacting with Computers*, 3(3), pp.319-353, 1991.
- [32] G. Salton, C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, 24(5), pp.513-523, 1988.
- [33] G. Salton, C. Buckley, "Improving Retrieval Performance by Relevance Feedback," *Journal of the American Society for Information Science*, 41(4), pp.288-297, 1990.
- [34] F. Sebastiani, "Machine Learning in Automated Text Categorization: A Survey," *Technical report*, Istituto di Elaborazione dell'Informazione, C.N.R., Pisa, Italy, 1999.
- [35] A. Sieg, B. Mobasher, R. Burke, "Inferring User's Information Context: Integrating User Profiles and Concept Hierarchies," *Proceedings of the 2004 Meeting of the International Federation of Classification Societies*, Chicago, USA, 2004.
- [36] P. L. Wang, D. Soergel, "A Cognitive Model of Document Use during a Research Project. Study I. Document Selection," *Journal of the American Society for Information Science and Technology*, 49(2), pp.115-133, 1998.
- [37] P. L. Wang, M. D. While, "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*, 50(20), pp.98-114, 1999.
- [38] E. Wenger, R. McDermott, W. Snyder, "Cultivating Communities of Practice," *Harvard Business School Press*, 2002.
- [39] D. H. Widyantoro, T. R. Ioerger, J. Yen, "Learning User Interest Dynamics with a Three-Descriptor Representation," *Journal of the American Society for Information Science and Technology*, 52(3), pp.212-225, 2001.
- [40] I.H. Witten, A. Moffat, T.C. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd edn., Morgan Kaufmann Publishers, Los Alto, USA, 1999.
- [41] I-C. Wu, D.-R. Liu, W.-H. Chen, "Task-stage Knowledge Support Model: Coupling User Information Needs with Task Stage Identification," *Proceeding of the IEEE 2005 International Conference on Information Reuse and Integration (IRI 2005)*, Las Vegas, USA, Aug. 2005, pp. 19-24.
- [42] M. H. Zack, "Managing Codified Knowledge," *Sloan Management Review*, 40(4), pp.45-58, 1999.