知識流探勘與文件推薦之整合研究(1/3)

Research on the Integration of Knowledge Flow Mining and Document Recommendation (1/3)

計劃編號：NSC 96-2416-H-009-007-MY3
執行期間：96 年 8 月 1 日至 97 年 7 月 31 日
主持人：　劉敦仁　　交通大學 資訊管理研究所
計劃參與人員：賴錦慧、陳韋孝、陳宇軒、黃健誠、邱璇

摘要

知識是企業組織最重要的資產，也是獲取競爭優勢的來源。在面對快速變動的環境中，組織必須使用有效的方法來管理知識，以幫助知識工作者找尋工作之相關知識，促進組織中的知識保存、分享與再利用。因此本研究主要探討如何根據知識工作者的歷史工作記錄以發掘知識流，進而了解工作者的工作需求，以提供相關知識支援。計畫執行進度包括：(1) 運用資料探勘與資訊擷取技術發掘個人知識流；(2) 分析知識工作者資訊需求與參考知識文件的特性，設計以知識流為基礎之文件推薦方法；(3)實驗評估驗證所提方法有效發掘知識流與提供工作相關知識支援。

關鍵詞：知識流探勘、知識分享、文件推薦、協同式過濾、序列規則探勘

Abstract

Knowledge is a critical resource that organizations use to gain and maintain competitive advantages. In the constantly changing business environment, organizations have to exploit effective and efficient methods of preserving, sharing and reusing knowledge in order to help knowledge workers find task-relevant information. Hence, an important issue is how to discover the knowledge flow (KF) of workers from their historical work records. The objectives are to understand the knowledge workers' task-needs and the ways they reference documents, and then provide adaptive knowledge support. The research progress of this project includes the following. (1) Using data mining and information retrieval techniques to discover workers' knowledge flows; (2) Analyzing the information needs and referencing behavior of workers, and designing knowledge flow-based recommendation methods; (3) Empirical evaluations were conducted to demonstrate that the proposed methods provide a basis for effective knowledge flow discovery and knowledge support.

Keywords: Knowledge Flow Mining, Knowledge Sharing, Document Recommendation, Collaborative Filtering, Sequential Rule Mining

1. Background and research objective

As the most important resource in an organization, knowledge can be used to create core competitive advantages and achieve commercial success in a constantly changing business environment. Hence, organizations need to adopt appropriate strategies to preserve, share and reuse such a valuable asset, as well as to support knowledge workers intelligently [10]. Knowledge and expertise are generally codified in textual documents, e.g., papers, manuals and reports, and preserved in a knowledge database. This codified knowledge is then circulated in an organization to support workers engaged in management and operational activities. Because most of these activities are knowledge-intensive tasks, knowledge management plays a key role in preserving and sharing organizational knowledge. Consequently, the effectiveness of knowledge management depends on providing task-relevant documents to meet the information needs of knowledge workers.

In such task-based business environments, knowledge management systems (KMS) can facilitate the preservation, reuse and sharing of knowledge, e.g. *KnowMore* system [1] and *K-support* system [9]. However, previous research on task-based knowledge support did not analyze and utilize the flow of knowledge among various types of codified knowledge (documents) to provide effective recommendations about task-relevant documents.

Knowledge flow (KF) research focuses on how KF can transmit, share, and accumulate knowledge when it passes from one team member/process to another. In a workflow situation, work knowledge may flow among workers in an organization, while process knowledge may flow among various tasks [16]. Thus, KF reflects the level of knowledge cooperation between workers or processes and influences the effectiveness of teamwork/workflow. However, to the best of our knowledge, there is no systematic method that can flexibly identify KF in order to understand the information needs of workers. Furthermore, conventional KF approaches do not analyze knowledge flow from the perspective of information needs and recommend relevant documents based on the discovered KF.

In this work, we propose two KF-based recommendation methods for recommending task-related codified knowledge. To adaptively provide relevant knowledge, collaborative filtering (CF), the most frequently used method, predicts a target worker's preference(s) based on the opinions of similar workers. However, the target worker's referencing behavior may change over the period of the task's execution, because his/her information needs may vary. Traditional CF methods only consider workers' preferences for codified knowledge. They neglect the effect of the time factor, i.e., workers' referencing behavior for knowledge. To overcome this limitation, we propose a KF-based sequential rule recommendation method (KSR) that recommends codified knowledge by utilizing the KF-based sequential rules. However, the method is based on the target worker's referencing behavior without considering the opinions of his/her neighbors who may have similar preference for documents. Therefore, to take advantage of the merits of typical CF and KSR methods, we propose hybrid recommendation methods that combine CF and KSR methods to enhance the quality of document recommendation. Such hybrid methods not only consider a workers' preferences for codified knowledge, but also their knowledge referencing behavior in order to predict topics of interest and recommend task-related knowledge.

2. Related work

2.1. Knowledge flow

To generate ideas and solve problems, humans create knowledge that is both abstract and dynamic. Such knowledge can then flow among people and processes to facilitate knowledge sharing and reuse. The concept of knowledge flow has been applied in various domains, e.g., scientific research, teamwork, industry, and organizations [4, 12, 15]. To improve the efficiency of teamwork, Zhuge [17] proposed a pattern-based approach that combines codification and personalization strategies to design an effective knowledge flow network and develop strategies for managing knowledge.

2.2. Information retrieval and task-based knowledge support

Information retrieval (IR) deals with the representation, organization, and storage of knowledge, and facilitates access to, information items [5]. In the IR domain, the vector space model [13] is typically used to represent documents as vectors of index terms, where the weights of the terms are measured by the *tf-idf* approach. The *tf-idf* weight is then used to estimate the importance of a term in a document stored in the corpus. Information filtering with a similarity-based approach is often used to locate knowledge items relevant to the task-at-hand. The discriminating terms of a task are usually extracted from a knowledge item/task to form a task profile, which is used to model a worker's information needs [9].

2.3. Rule-based recommendation

Association rule mining [2] is a widely used data mining technique that generates recommendations in recommender systems. An association rule describes the relationships between items, such as products, documents, or movies, based on patterns of co-occurrence across transactions. The *Apriori* algorithm is usually employed to identify association rules in transactions. The discovered rules should satisfy two user-defined requirements: minimum support and minimum confidence.

Cho *et al.* [8] proposed a sequential rule-based recommendation method that considers the evolution of customers' purchase sequences. The method applies sequential rules to keep track of customers' preferences. A sequential rule is expressed in the form C_{T-l+1}, …, C_{T-1} => C_T, where C_T represents the customers' purchase behavior in period T. If a target customer's purchase behavior prior to period T was similar to the conditional part of the rule, then C_T is used to recommend products to the target customer.

2.4. Collaborative filtering recommendation

Collaborative filtering (CF), the most successful recommendation approach developed thus far, is used in many applications. CF is based on the concept that if like-minded users like an item, then the target user will probably like it as well [6]. The CF approach involves two steps: neighborhood formation and prediction. The neighborhood of a target user is selected according to user similarity, which is computed by Pearson's correlation or the cosine measure. Either the k-NN (nearest neighbor) approach or a threshold-based approach is used to choose n users that are most similar to the target user. Here, we use the k-NN approach. In the prediction step, the predicted rating is calculated from the aggregated weights of the selected k nearest neighbors' ratings.

3. Recommendation process based on knowledge flow model

Our methods consist of two phases, a knowledge flow mining phase and a KF-based recommendation phase. The first phase identifies the worker's knowledge flow from the large amount of knowledge in the worker's log data. Then, the second phase recommends codified knowledge to the target worker by using the proposed recommendation methods.

In the knowledge flow mining phase, KFs are identified from the task requirements and the referencing behavior of workers recorded in their logs.

As tasks are performed at various times, each knowledge worker requires different kinds of knowledge to achieve a goal or complete a task. The phase involves three steps: document profiling, document clustering, and knowledge flow mining.

To enhance the quality of recommendations, we propose hybrid recommendation methods that combine a KF-based sequential rule (KSR) method with collaborative filtering (CF). The KSR method is regarded as the core process of the proposed hybrid methods. In the KSR method, workers with similar KFs to that of the target worker are deemed neighbors of the target worker and their knowledge referencing behavior patterns are identified by a sequential rule mining method. Based on the discovered sequential rules and the KFs of the neighbors, relevant topics and codified knowledge are recommended to the target worker to support the task-at-hand. Moreover, by considering workers' preferences for codified knowledge, the CF method makes recommendations to the target worker based on the opinions of similar workers. Two approaches are used to find similar workers to the target worker. The preference-similarity-based CF method (PCF) chooses workers with similar preferences, while the KF-similarity-based CF method (KCF) chooses workers with similar KFs. To adaptively and proactively recommend codified knowledge, we consider workers' referencing behavior as well as their preferences for codified knowledge. Therefore, two hybrid recommendation methods are used in the KF-based recommendation phase: 1) a hybrid of PCF and KSR (PCF-KSR), and 2) a hybrid of KCF and KSR (KCF-KSR). Further details are given in Section 5. Through the hybrid recommendation methods, the top-N documents with highest predicted ratings are recommended to the target worker.

4. Knowledge flow model

A worker's KF is identified by analyzing a worker's knowledge referencing behavior from his/her historical work logs, which contain information about previously executed tasks, task-related documents and when the documents were accessed. Formally, we define knowledge flow as follows.

Definition 1 Knowledge Flow (KF): Let a worker's personal knowledge flow be $KFlow_w=\{TKF_w, CKF_w\}$, where TKF_w is the topic-level KF of the worker w for a task, and CKF_w is his/her codified-level KF for the task.

Definition 2. Codified-Level KF: A codified-level KF is a time-ordered sequence arranged according to the access time of the documents it contains. Thus, it is defined as $CKF_w = < d_w^{t_1}, d_w^{t_2}, \cdots, d_w^{t_f} >$ and $t_1 < t_2 < \cdots < t_f$ where $d_w^{t_j}$ denotes the document that the worker w accessed at time t_j for a specific task. Each document can be represented by a document profile, which is as an n-dimensional vector containing several terms that indicate the key content of the document.

Definition 3. Topic-Level KF: The topic-level KF is a time-ordered topic sequence derived by mapping documents in the codified-level KF into the corresponding topics. Thus, the topic-level KF is defined as $TKF_w = < TP_w^{t_1}, TP_w^{t_2}, \cdots, TP_w^{t_f} >$, $t_1 < t_2 < \cdots < t_f$, where $TP_w^{t_j}$ denotes the corresponding topic of the document that worker w accessed at time t_j for a specific task. Each topic can be represented by a topic profile, which is an n-dimensional vector containing terms that indicate the key content of the topic.

The codified-level KF is extracted from the documents recorded in the worker's work log. The topic-level KF is derived by mapping documents in the codified-level KF into corresponding clusters and is represented by a topic sequence. In the previous step, documents with similar content were grouped into clusters. We use the result of document clustering [5] to map the documents in the codified-level KF into topics (clusters) to compile the topic-level KF.

5. KF-based recommendation phase

The KF-based recommendation phase consists of two hybrid recommendation methods: 1) PCF and KSR (PCF-KSR), and 2) KCF and KSR (KCF-KSR). To adaptively recommend documents, both the traditional CF method and the KCF method select neighbors based on the similarity of preferences (document ratings). The two methods differ in the way they compute the similarity between workers' preferences to select the target worker's neighbors. The PCF method (traditional CF) uses preference ratings to compute the similarity, while the KCF method uses workers' KFs to derive the similarity. However, workers' referencing behavior may change over time, because their information needs may vary. These two methods do not consider the neighbors' referencing behavior. Thus, to recommend topics of the codified knowledge, we propose a KF-based sequential rule recommendation (KSR) method, which traces workers' knowledge referencing behavior by using the KF-based sequential rules. However, the opinions of the target worker's neighbors who have similar preference on the codified knowledge are not considered. To take advantage of the merits of the KSR, CF and KF methods, we propose two types of hybrid recommendation methods, namely PCF-KSR and KCF-KSR, to improve the quality of document recommendations.

5.1. Identifying similar workers based on their knowledge flows

To find a target worker's neighbors, his/her topic-level KF is compared with those of other workers to compute the similarity of their KFs. The resulting similarity measure indicates whether two workers' KF referencing behavior is similar. Since the KFs are sequences, the sequence alignment method [7, 11], which computes the cost of aligning one sequence with another sequence, can be used to measure the similarity of two sequences. Based on this concept, we propose a hybrid similarity measure, comprised of the KF alignment similarity and the aggregated profile similarity, to evaluate the similarity of two workers' KFs, as shown in Eq. 1.

$$sim(TKF_i, TKF_j) = \alpha \times sim_a(TKF_i, TKF_j) + (1-\alpha) \times sim_P(AP_i, AP_j) \quad (1)$$

where $sim_a(TKF_i, TKF_j)$ represents the KF alignment similarity, $sim_p(AP_i, AP_j)$ represents the aggregated profile similarity, and α is a parameter used to adjust the relative importance of these two types of similarity. By linearly combing these two similarities, we can balance the tradeoff between KF alignment and the aggregated profile.

5.2. KF-based sequential rule recommendation method

The KF-based sequential rule recommendation method (KSR) considers the referencing behavior of neighbors whose KFs are very similar before time *T*, and then recommends documents at time *T* for the target worker. To determine the similarity of various topic-level KFs, the target worker's KF is compared with those of other workers by measuring their KF similarity, as discussed in Section 5.1. Workers with similar KFs to that of the target worker are regarded as the latter's neighbors and their topic-level KFs are used to predict the knowledge referencing behavior of the target worker by applying sequential rule mining techniques. Then, the discovered sequential rules with high degrees of rule matching are selected to recommend topics at time *T*.

Mining knowledge referencing behavior

The knowledge referencing behavior and information needs of knowledge workers may change over time. Based on the similarities mentioned in the previous sub-section, knowledge workers with similar referencing behavior (high similarities) are grouped together and regarded as neighbors of the target worker. Using the topic-level KF of each group member, we modify the association rule mining method [2] and sequential pattern mining method [3] to discover topic-level sequential rules from the neighbors' KFs. The extracted rules can be used to keep track of the referenced topics and the relationships among workers with similar referencing behavior. Let R_y be an sequential rule, as defined in Eq. 2.

$$R_y: r_{y,T-l}, \ldots, r_{y,T-1} \Rightarrow r_{y,T} \quad (Support_y, Confidence_y) \\ \text{where } r_{y,T-s} \in TKF;\ s=0 \text{ to } l \quad (2)$$

The conditional part of the sequential rule is $<r_{y,T-l},\ldots,r_{y,T-1}>$, and the consequent part is $r_{y,T}$. The items that appear in the rules are topics extracted from the neighbors' topic-level KFs (*TKF*). The support and confidence values, $Support_y$ *and* $Confidence_y$, are used to evaluate the importance of rule R_y. We use the support and confidence scores to measure the degree of match between the referencing behavior and the conditional part of a rule for a target worker. Note that if the knowledge referencing behavior of the target worker is similar to the conditional part of R_y, then the topic predicted for him/her at *T* will be $r_{y,T}$.

Identifying the knowledge referencing behavior of the target worker

This step identifies the target worker's knowledge referencing behavior by matching his/her KF with the sequential rules discovered in the previous step. Specifically, the rules are matched with the topic-level KF of the target worker to predict the topics required at time *T*. We set a knowledge window on the KF before time *T*. The size of the window is determined by the user. Let $KW_u = < TP_u^{T-s}, TP_u^{T-s+1} \cdots, TP_u^{T-1} >$ be the knowledge window for the topic-level KF of a target worker *u* before time *T*. Note that TP_u^{T-s} is the topic referenced by *u* at time *T-s*, *s=1…l*. The knowledge window KW_u covers several topics previously referenced by the target worker and arranged in time order. Given the similarity matching scores of all topic subsequences extracted from a sequential rule, we choose the subsequence with the highest score to compute the matching degree of the rule.

Document recommendation

The KSR method predicts a document rating based on sequential rules derived from the KFs of a target worker's neighbors. Let KNB_u be a set of neighbors of target worker *u*, selected according to the KF similarity (using Eq. 1). The sequential rules derived from KNB_u with high degrees of rule matching are selected to recommend topics for the target worker at time *T*. However, the referencing behavior of some workers in KNB_u does not match the selected sequential rules. Therefore, we apply the sequential rule matching method discussed in Section 5.2 to compare the KFs of workers in KNB_u with the selected sequential rules. Let $KNBR_u$ denote the neighbors in KNB_u whose KFs are highly similar to that of the target worker and whose referencing behavior matches the selected sequential rules.

Let *RTS* be set of recommended topics derived from the consequent part of recommended sequential rules, and let τ be the topic of a document *d*. Based on the KF in $KNBR_u$, the predicted rating of a document *d* belonging to topic τ for the target worker *u* is calculated by Eq. 3:

$$\hat{p}_{u,d} = \begin{cases} \bar{R}_{u,\tau} + \dfrac{\sum_{y \in KNBR_u} (R_{y,d,\tau} - \bar{R}_{y,\tau})}{|KNBR_u|}, & if\ \tau \in RTS \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $\bar{R}_{u,\tau}$ / $\bar{R}_{y,\tau}$ is the topic rating of the target worker *u*/worker *y*, derived from the worker's average rating of documents in topic τ; $R_{y,d,\tau}$ is the rating of document *d* belonging to topic τ and is given by worker *y*; and $|KNBR_u|$ is the number of workers in $KNBR_u$. If the target worker *u* does not rate any documents in τ, then $\bar{R}_{u,\tau}$ is replaced by the average rating of his/her complete documents. Note that if τ (the topic of document d) $\notin RTS$, the predicted rating of the document *d* will be 0.

The KSR method only predicts ratings for documents that belong to the recommended topics; other documents are rated as 0. Unlike traditional recommendation methods, KSR recommends documents to the target worker according to the selected sequential rules and the document ratings.

5.3. The hybrid PCF -KSR method

The hybrid PCF-KSR recommendation method linearly combines the preference-similarity-based CF method (PCF) with the KSR method to recommend documents for a target worker. The PCF method is the traditional CF method that makes recommendations according to workers' preferences for codified knowledge. To recommend a document, the neighbors of a target worker are selected based on the similarities of the workers' preference ratings. Pearson's correlation coefficient is used to find similar workers based on the document rating vectors. Then, PCF-KSR predicts the rating of a document by linearly combining the predicted ratings calculated by the two methods. Because a worker's knowledge flow may change over time, the hybrid method considers the worker's preference for documents as well as topic changes in the worker's KF to make recommendations adaptively.

The predicted rating of a document *d* is derived by combining the PCF method with the KSR method, as defined in Eq. 4:

$$\hat{p}_{u,d} = \beta_{PCFKSR} \times \left[\bar{R}_u + \frac{\sum_{x \in PNB_u} PSim(u,x) \times (R_{x,d} - \bar{R}_x)}{\sum_{x \in PNB_u} PSim(u,x)} \right] + (1 - \beta_{PCFKSR}) \times \mathrm{KSR}_{u,d} \quad (4)$$

where $\bar{R}_u$ / $\bar{R}_x$ is the average rating given by the target worker *u* / worker *x* for documents; *PSim*(*u*, *x*) is the similarity between the target worker *a* and a neighbor worker *x*, derived from Pearson's correlation coefficient; PNB_u is the set of neighbors of target worker *u*, selected by *PSim*(*u*, *x*); $R_{x,d}$ is the rating of document *d* given by worker *x*; $KSR_{u,d}$ is the predicted rating for document *d* based on the KSR method; and $\beta_{PCF\text{-}KSR}$ is the weighting used to adjust the relative importance of the PCF method and KSR method.

According to Eq. 4, a document in a recommended topic has a higher priority for recommendation than those documents that are not in the recommended topics, since the predicted ratings of those documents derived by the KSR method will be zero. Documents with high predicted ratings are used to compile a recommendation list, after which the top-N documents are recommended to the target worker.

5.4. The hybrid KCF-KSR method

The hybrid KCF-KSR method linearly combines the KF-similarity-based CF method (KCF) with the KSR method to recommend documents to the target worker. The KCF method is very similar to PCF method (the traditional CF method), because it uses the KF similarity instead of Pearson's correlation coefficient to derive the similarity and select neighbors. The KCF method is based on the referencing behavior of neighbors with similar KFs, while the PCF method is based on the Pearson correlation coefficients to select neighbors with similar preferences. Like the PCF-KSR method, the predicted rating of a document is also derived by integrating two parts of the ratings. One part is obtained by the KCF method, while the other is obtained by the KSR method described in Section 5.2.

The KCF-KSR method predicts a document rating by using Eq. 5, and then determines which documents should be recommended.

$$\hat{p}_{u,d} = \beta_{KFKSR} \times \left[\bar{R}_u + \frac{\sum_{x \in KNB_u} sim(TKF_u, TKF_x) \times (R_{x,d} - \bar{R}_x)}{\sum_{x \in KNB_u} sim(TKF_u, TKF_x)} \right] + (1 - \beta_{KFKSR}) \times \mathrm{KSR}_{u,d} \quad (5)$$

where $\bar{R}_u$ / $\bar{R}_x$ is the average rating of documents given by the target worker *u* / worker *x*; $R_{x,d}$ is the rating of a document *d* given by worker *x*; TKF_u / TKF_x is the topic-level KF of the target worker *u* / worker *x*; $sim(TKF_u, TKF_x)$ is the KF similarity of worker *u* and worker *x*, which is derived by Eq. 1; KNB_u is the set of neighbors of the target worker *u*, selected according to their KF similarity; $KSR_{u,d}$ is the predicted rating of document *d* based on the KSR method; and $\beta_{KCF\text{-}KSR}$ is the relative weighting used to adjust the importance of the KCF method and the KSR method.

6. Experiment evaluations

In the experiments, we evaluate and compare the performance of three document recommendation methods, namely the hybrid PCF-KSR method, the hybrid KCF-KSR method, and the traditional CF method. We use the Mean Absolute Error (MAE), which is widely used in recommender systems, to evaluate the recommendation quality of our proposed methods. MAE measures the average absolute deviation between a predicted rating and user's true rating [14].

The experiment results show that the proposed methods improve the quality of document recommendation and outperform the traditional CF method. Additionally, using KF mining and sequential rule mining techniques can increase the accuracy of recommendations. The KF-based recommendation methods provide knowledge support adaptively based on the referencing behavior of workers with similar KFs, and also facilitate knowledge sharing among such workers.

7. Conclusions

The proposed methods integrate the concepts and methods of knowledge management, document management, information retrieval, and collaborative filtering to identify knowledge workers' KFs from the perspective of their information needs. KF reduces the difficulties that knowledge workers experience when performing unfamiliar tasks, helps other workers function more efficiently, and enhances knowledge sharing and reuse. In organizations, KF facilitates knowledge management by recognizing the information needs and referencing behavior of knowledge workers, and by providing knowledge support proactively and adaptively. The proposed methods can contribute to both academic research and practical knowledge management applications.

References

[1] A. Abecker, A. Bernardi, K. Hinkelmann, O. Kuhn, and M. Sintek, "Context-Aware, Proactive Delivery of Task-Specific Information: The KnowMore Project," *Information Systems Frontiers,* vol. 2, 3, pp. 253-276, 2000.

[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record,* vol. 22, 2, pp. 207-216, 1993.

[3] R. Agrawal and R. Srikant, "Mining sequential patterns," *Proceedings of the Eleventh International Conference on Data Engineering,* pp. 3-14, 1995.

[4] A. Anjewierden, R. de Hoog, R. Brussee, and L. Efimova, "Detecting knowledge flows in Weblogs," *13th International Conference on Conceptual Structures (ICCS 2005),* vol. University of Kassel, Kassel, pp. 1-12, 2005.

[5] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston: Addison-Wesley, 1999.

[6] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence,* pp. 43-52, 1998.

[7] K. Charter, J. Schaeffer, and D. Szafron, "Sequence alignment using FastLSA," in *Proceedings of The 2000 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS,2000),* 2000, pp. 239-245.

[8] Y. B. Cho, Y. H. Cho, and S. H. Kim, "Mining changes in customer buying behavior for collaborative recommendations," *Expert Systems with Applications,* vol. 28, 2, pp. 359-369, 2005.

[9] D. R. Liu, I. C. Wu, and K. S. Yang, "Task-based K-Support system: disseminating and sharing task-relevant knowledge," *Expert Systems With Applications,* vol. 29, 2, pp. 408-423, 2005.

[10] I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*: Oxford University Press, 1995.

[11] S. G. Oguducu and M. T. Ozsu, "Incremental click-stream tree model: Learning from new users for web page prediction," *Distributed and Parallel Databases,* vol. 19, 1, pp. 5-27, 2006.

[12] O. M. Rodriguez, A. I. Martinez, J. Favela, A. Vizcaino, and M. Piattini, "Understanding and supporting knowledge flows in a community of software developers," *International Workshop on Groupware (CRIWG),* pp. 52-66, 2004.

[13] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management,* vol. 24, 5, pp. 513-523, 1988.

[14] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth"," *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '95),* vol. Denver, Colorado, United States, pp. 210-217, 1995.

[15] H. Zhuge, "Discovery of knowledge flow in science," in *Communications of the ACM*. vol. 49: ACM Press New York, NY, USA, 2006, pp. 101-107.

[16] H. Zhuge, "A knowledge flow model for peer-to-peer team knowledge sharing and management," *Expert Systems with Applications,* vol. 23, 1, pp. 23-30, 2002.

[17] H. Zhuge, "Knowledge flow network planning and simulation," *Decision Support Systems,* vol. 42, 2, pp. 571-592, 2006.