

# A Hybrid Knowledge Discovery System Based on Items and Tags

Winyu Niranatlamphong and Worasit Choochaiwattana\*

College of Creative Design and Entertainment Technology, Dhurakij Pundit University, Bangkok, Thailand

**Abstract:** Exponentially increasing knowledge in a management system is the main cause of the overload problem. Development of a recommender service embedded in the management system is challenging. This paper proposes a hybrid approach by combining an item-based recommendation technique (collaborative filtering technique) with a tag-based recommendation technique (content based filtering technique). In order to evaluate the performance of the proposed hybrid approach, a group of knowledge management system users are invited as participants in the research. Participants are asked to use the prototype of a management system embedded within the knowledge recommender service for four months, which guarantees that each interaction by participants with knowledge items are recorded. A confusion matrix is used to compute accuracy of the proposed hybrid approach. The results of the experiments reveal that the hybrid approach outperforms both item-based and tag-based approaches. The hybrid approach seems to be a promising technique for a recommender service in the knowledge management system.

**Keywords:** Collaborative Filtering, Content-based Filtering, Item-Based Recommendation, Tag-Based Recommendation, Knowledge Recommender Service.

## 1. INTRODUCTION

With the advancement of information technology, modern businesses exploit the usage of various information systems to improve their daily productivity and to gain competitive advantages over their competitors. Information and knowledge become the most important critical factors for the success of modern businesses. A knowledge management system helps the employees of a business organization to store, retrieve and disseminate their knowledge, and allows the business organization to maintain its core knowledge.

One of the core services in knowledge management is a knowledge retrieving service. When a tremendous collection of knowledge is stored in the knowledge management system, its users will encounter the problem of knowledge overload. Then the knowledge retrieving service becomes a major player to help its users to search for their desired knowledge. One way to overcome the problem of knowledge overload is to develop an automatic knowledge dissemination mechanism, or a knowledge recommender service.

This paper is organized as follows: section 2 provides a literature review on details on related work. Section 3 introduces a proposed hybrid recommendation mechanism. The experimental setting and evaluation are explained in a proposed hybrid mechanism. The results and discussion are revealed in

section 4 in an experimental setting. A conclusion and future work are described in section 5.

## 2. LITERATURE REVIEW

With a tremendous amount of information available, a recommender system plays a crucial role for filtering information that people may find of interest. Instead of asking for recommendations, the recommender system provides suggestions on any items which are likely to be interesting for a user (Bruke 2007; Resnick and Varian 1997; Ricci, Rokach, and Shapira 2011).

In 1992, Goldberg *et al.* (1992) introduced a recommender system, named Tapestry, and coined term "Collaborative Filtering". The principle idea of Collaborative Filtering (CF) is that, if two persons have the same behavior, for instance, buying similar items, they will act on other items similarly (Goldberg *et al.* 2001). The CF has been adopted and used to develop early generation of recommender systems, such as GroupLens, an online news recommender system (Resnick *et al.* 1994). Paul Resnick and his colleagues from MIT and University of Minnesota exploit the usage of user rating on news articles to provide news recommendation. Many commercial systems, such as Amazon.com, use this technique because of easy implementation which is highly effective (Despande and Karypis 2004; Linden, Smith, and York 2003).

The main drawback of the CF, however, is the cold-start problem. This problem occurs when there are few interactions available for items. Thus, recommender systems cannot make any recommendation at the beginning (Davoodi and Fatemi 2012). Besides the CF, Content-Based Filtering is another technique for

\*Address of correspondence to this author at the College of Creative Design and Entertainment Technology, Dhurakij Pundit University, Bangkok, Thailand; Tel: +662-954-7300 Ext. 786; Fax: +662-954-8651; E-mail: worasit.cha@dpu.ac.th

JEL: D83, I23, O31.

recommender systems. The principle idea of content-based filtering is that when each person interacts with any items, the content of those items will be recorded. A user profile of each person will then be created. A recommender mechanism in the content-based filtering technique will recommend any items based on a matched score calculation between the user profile and content of each item. The content-based filtering, however, will only be appropriate if each item in the recommender system has textual content.

The main difference between collaborative filtering and content-based filtering is that collaborative filtering only uses user-item interactions to make predictions and recommendations, while the content-based filtering used the extracted feature of each user and each item for recommendation (Si and Jin 2003). However, both techniques can suffer when the system has a huge number of users and items, and a few user-item interactions, which will lead to the problem of Sparsity (Jain *et al.* 2015). In order to overcome the limitations of other techniques that were previously discussed, a hybrid recommender technique combines collaborative filtering and content-based filtering. It exploits the usage of item content and user-item interaction to create item recommendations (Prasad and Kumari 2012).

As mentioned in the previous section, the users of the knowledge management system will encounter the problem of knowledge overload when a tremendous collection of knowledge is stored. The knowledge retrieving service could derive benefit from applying collaborative filtering or the content-based filtering to provide an automatic knowledge dissemination mechanism or a knowledge recommender service (Aryal, Dutta, and Morshed 2013; Choochaiwattana 2015; Huang *et al.* 2012; Li, Liu, and LV 2006; Liang, Cai, and Zhao 2007; Si and Jin 2003; Vizcaino *et al.* 2009; Zhao, Wang, and Lui 2009). There are only a few published papers that have focused on a combination of collaborative filtering and content-based filtering to develop the knowledge recommender service. Hence, this paper investigates how well the combination of collaborative filtering, called item-based recommendation, and content-based filtering, called tag-based recommendation, contributes to the task of automated knowledge dissemination.

### 3. PROPOSED HYBRID MECHANISM

The proposed hybrid recommendation mechanism combines the item-based recommendation mechanism

with the tag-based recommendation mechanism. In order to explain the proposed mechanism, the two recommendation mechanisms are described briefly.

As illustrated in Figure 1, the item-based recommendation mechanism considers user-item interactions for calculating similarity measurements among users. The users, who interact with similar knowledge items, will be placed in the same group. The user-item interactions of all users in the same group can then be identified. The item-based recommendation mechanism will select knowledge items and disseminate the items to all the users in that group, as illustrated in Figure 2.

The tag-based recommendation mechanism, as illustrated in Figure 3, on the other hand, considers user-item interactions for extracting knowledge content, which is knowledge for this paper. The extract knowledge tags will be placed in a set of user's knowledge tags, which represents the knowledge interest of each user. The tag-based recommendation mechanism will select new or unvisited knowledge items and disseminate the items by calculating similarity measurements between content of the set of user's knowledge tags and content of new or unvisited knowledge items.

In order to implement the proposed hybrid recommendation mechanism, there are six main components – set of users, set of interaction with knowledge items, set of users' knowledge tag, set of tags from new/unvisited knowledge items, similarity measurement, and knowledge corpus. Let  $N_u$  be the number of users and  $N_k$  be the number of knowledge items in a knowledge management system. Let  $U$  be a set of users that contains all users in the knowledge management system;  $U = \{u_1, u_2, u_3, \dots, u_n\}$ , and  $K$  is a set of knowledge items and contains all knowledge items in the knowledge corpus;  $K = \{k_1, k_2, k_3, \dots, k_m\}$ . Let  $M_{uk}$  be the  $N_u \times N_k$  association matrix between users and knowledge items.  $M_{uk}(u_x, k_y)$  will be equal to 1 when user  $u_x$  bookmarks a knowledge item  $k_y$ .

Thus, each row, or  $UK_i$ , in  $M_{uk}$  represents user interaction with knowledge items. In addition, for each user  $u_x$ , let  $UTK_x$  be a set of user's knowledge tag derived from  $M_{uk}$ ;  $UTK_x = \{ \langle u_x, t_p \rangle \mid u_x \in U \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 1 \}$  and  $NTK_y(u_x)$  be a set of tags from new or unvisited knowledge items derived from  $M_{uk}$ ;  $NTK_y(u_x) = \{ \langle k_y, t_p \rangle \mid k_y \in K \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 0 \}$ .

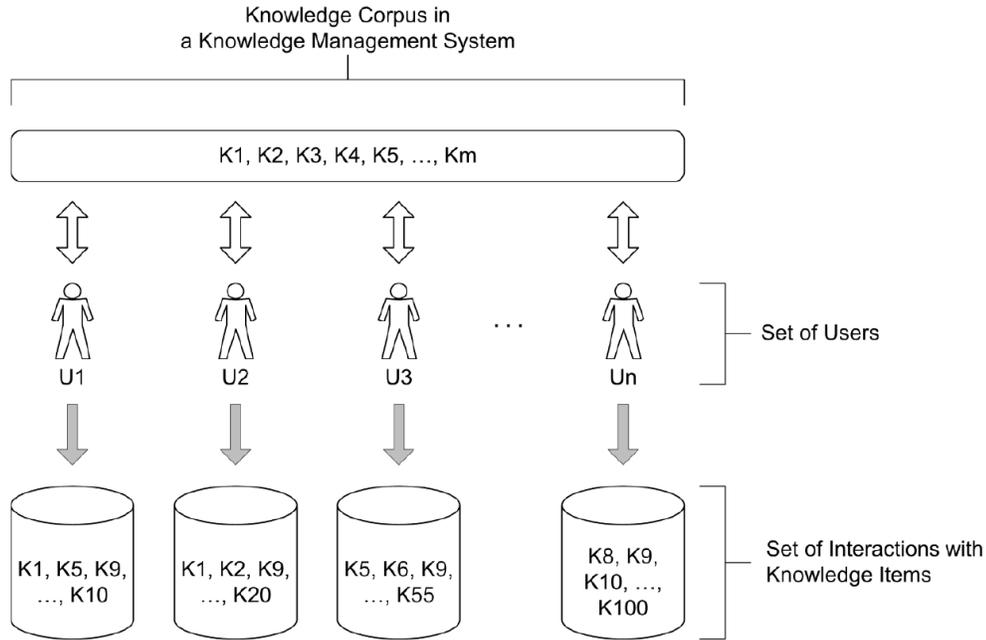


Figure 1: Concept of the Item-Based Recommendation Mechanism.

A similarity measurement between each user is then performed. A group of users, who interact with similar knowledge items, can be identified if the similarity score, as illustrated in Equation (1), is equal to or greater than a predefined threshold ( $\alpha \geq 0.55$ , for this particular study). Let  $GroupU_x$  be a set of users who interact with similar knowledge items, and  $GroupKU_x$  be a set of knowledge items of every users in  $GroupU_x$ ;  $GroupKU_x = \{ \langle u_i, k_j \rangle \mid k_j \in K \wedge u_i \in GroupU_x \wedge M_{uk}(u_i, k_j) = 1 \}$ . Thus,  $GroupKU_x$  can be used in the item-base recommendation mechanism, as illustrated in Figure 2.

$$USim(UK_x, UK_y) = \frac{\sum_{j=1}^t UK_{xj} \cdot UK_{yj}}{\sqrt{\sum_{j=1}^t UK_{xj}^2 \cdot \sum_{j=1}^t UK_{yj}^2}} \quad (1)$$

Since  $UTK_x$  represent the knowledge interest of user  $x$  and  $NTK_y(u_x)$  represents the content of new or unvisited knowledge, a similarity measurement between  $UTK_x$  and  $NTK_y(u_x)$  can be calculated as illustrated in Equation (2). Let  $KU_x$  be a set of new or unvisited knowledge items, where  $KSim(UTK_x, NTK_y(u_x))$  is equal to or greater than a predefined threshold ( $\alpha \geq 0.55$ , for this particular study), which can be used in the tag-based recommendation mechanism, as illustrated in Figure 3.

$$KSim(UTK_x, NTK_y(u_x)) = \frac{\sum_{j=1}^t UTK_{xj} \cdot NTK_{yj}(u_x)}{\sqrt{\sum_{j=1}^t UTK_{xj}^2 \cdot \sum_{j=1}^t NTK_{yj}(u_x)^2}} \quad (2)$$

As mentioned previously, the proposed hybrid recommendation mechanism combines the item-based and tag-based recommendation mechanisms. The proposed hybrid approach takes the results from  $GroupKU_x$  and  $KU_x$  into account when predicting the recommended list of knowledge items for each user.

Let  $RecList1(u_{x5})$  be a set of recommended knowledge items for user  $u_i$ , which shows up in both  $GroupKU_x$  and  $KU_x$ ;

$$RecList1(u_i) = \{ \langle u_i, k_j \rangle \mid (k_j \in GroupKU_x \wedge M_{uk}(u_i, k_j) = 0) \wedge (k_j \in KU_x) \}$$

and  $RecList2(u_{x5})$  be a set of recommended knowledge items for user  $u_i$ , which show up in either  $GroupKU_x$  or  $KU_x$ ;  $RecList2(u_i) = \{ \langle u_i, k_j \rangle \mid k_j \in GroupKU_x \wedge M_{uk}(u_i, k_j) = 0 \}$ .

#### 4. EXPERIMENTAL SETTING AND EVALUATION

In order to evaluate the proposed hybrid recommendation mechanism, data from a knowledge management system at the Faculty of Information Technology, Dhurakij Pundit University, Bangkok, Thailand, were crawled and loaded into our prototype knowledge management system embedded with knowledge recommender services. This prototype was designed and developed to make the task of recommendation mechanism evaluation easier. The results from the item-based recommendation mechanism and from the tag-based recommendation mechanism were compared with the results from the

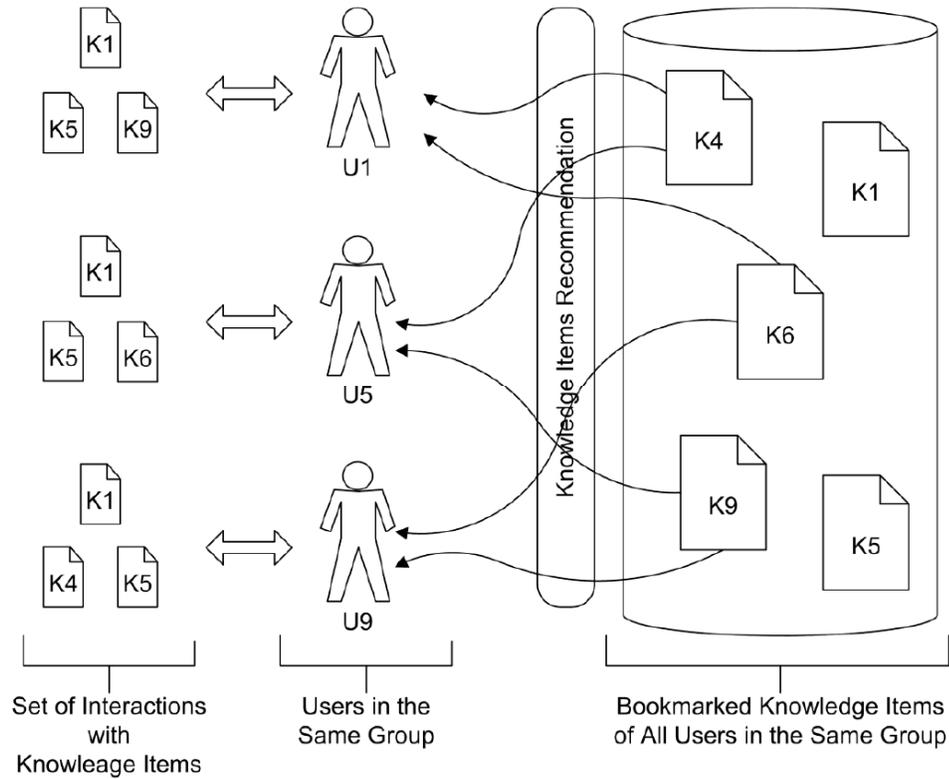


Figure 2: Example of Knowledge Items Recommendation Method.

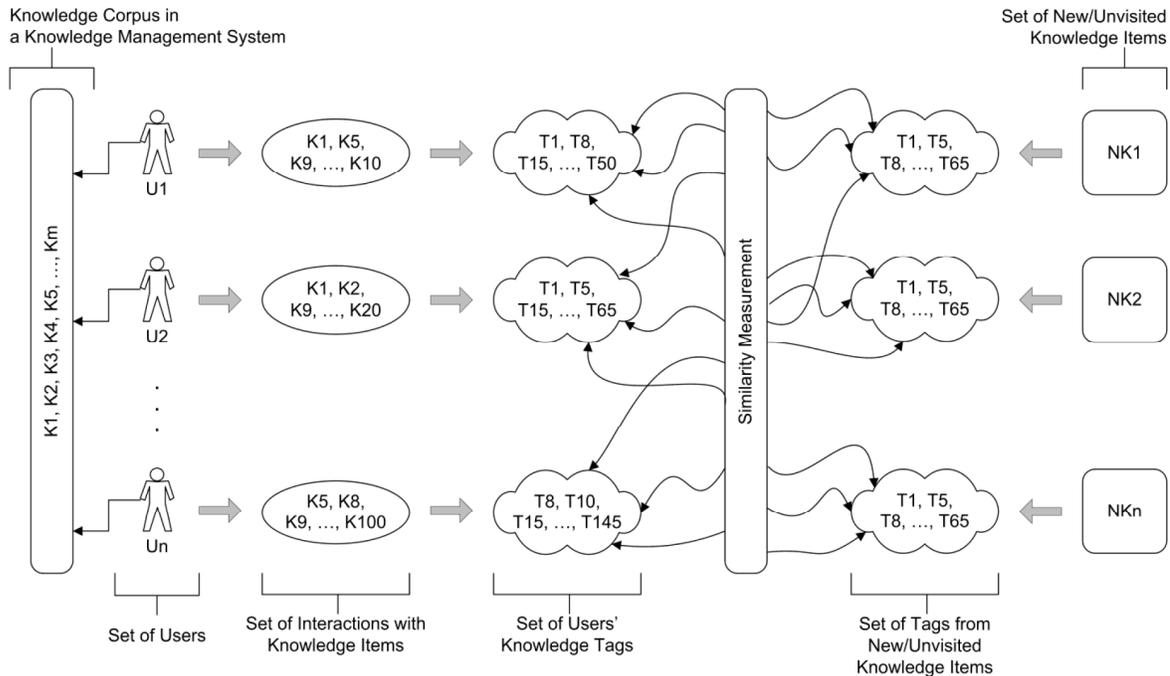


Figure 3: Concept of Tag-Based Recommendation Mechanism.

proposed hybrid recommendation mechanism. The percentage accuracy in knowledge items recommendation was an evaluation metric, as illustrated in Equation 3. Table 1 shows a confusion matrix used for calculating the percentage accuracy.

$$Percentage\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

Thirty subjects, consisting of fifteen graduate students and fifteen alumni, were recruited as

**Table 1: Confusion Matrix Used for Calculating the Percentage of Accuracy**

|                 | Predicted Positive   | Predicted Negative  |
|-----------------|--|---|
| Actual Positive | True Positive<br>(No. of Recommended User-Interested Knowledge Items)    | False Negative<br>(No. of Unrecommended User-Interested Knowledge Items)  |
| Actual Negative | False Positive<br>(No. of Recommended User-Uninterested Knowledge Items) | True Negative<br>(No. of Unrecommended User-Uninterested Knowledge Items) |

participants in the study. As a criterion, the participants were asked to use the prototype of a knowledge management system embedded with the knowledge recommender service for four months. This would guarantee that each participants' interaction with knowledge items could be recorded. During the four-month period, each participant was asked to evaluate a list of recommended knowledge items at the end of each month. The list of recommended knowledge items that were generated from the item-based recommendation mechanism, the tag-based recommendation mechanism, and the hybrid recommendation mechanism, were merged, and all duplicated knowledge items were removed. Before evaluating the list, each participant was informed that the list would be displayed in random order. The evaluation results provided by each participant were then associated with the original list produced by each recommendation mechanism. The percentage accuracy for each recommendation mechanism was then calculated.

## 5. DISCUSSION

In order to evaluate the effectiveness of each recommendation mechanism, the percentage accuracy was examined. Higher accuracy indicates greater effectiveness of the recommendation mechanism. Figure 4 provides a line chart of the percentage accuracy for the item-based recommendation mechanism, the tag-based recommendation mechanism, and the hybrid recommendation mechanism, during the four-month experimental period.

Figure 4 shows that the proposed hybrid recommendation mechanism outperformed the item-based and tag-based approaches. According to Figure 4, the performance of the proposed approach during the first month of experiment was lower than the performance of the tag-based approach. When the interactions between users and items were not enough, the tag-based approach performed better because the

set of users' knowledge tags, extracted from a set of interactions with knowledge items, represented actual users' interest.

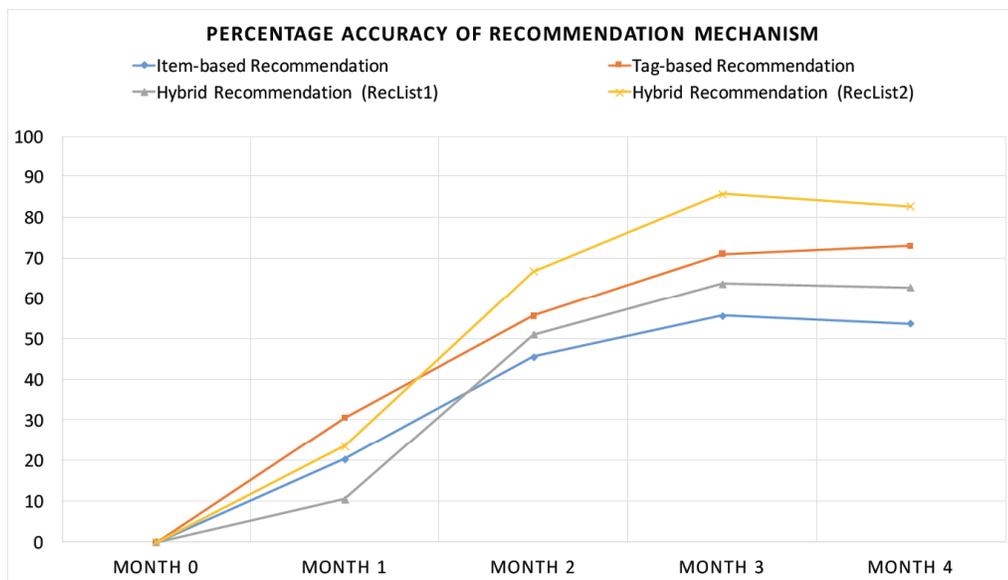
On the other hand, recommending knowledge items using the item-based approach suffered from the cold start problem and the false positive. When the information in the set of interactions with knowledge items was not enough, it impacted directly on the performance of the similarity measurement between users, resulting in assigning users to an inappropriate group. From the second to the fourth month of the experiment, the hybrid approach with ReList2 performed better than the other approach.

Exploiting the use of both the tag-based and item-based approaches, not only built a users' profile for their interest items, but also took opportunities in exploring potential items of interest by considering other users' interactions. It is not surprising that the item-based approach and the hybrid approach with ReList1 has a lower percentage accuracy compared with the other two approaches. It is possible that users might be interested in different pieces of knowledge, even though they have interacted with the same knowledge items.

From the results of the experiment, it can be concluded that both individual's item interactions and other people's item interactions should be considered when developing recommendation mechanisms. The proposed hybrid method can be applied not only to improve a recommendation service for a knowledge management system, but also to improve a recommendation service for systems that contain other content as well, such as a research paper recommender system.

## 6. CONCLUSION AND FUTURE WORK

This paper investigated how well a combination of collaborative filtering, called item-based recommendation, and content-based filtering, called tag-based recommendation, augments the task of



**Figure 4:** The percentage accuracy of recommendation mechanisms.

automated knowledge dissemination services in a knowledge management system. The evaluation results suggest that the hybrid recommendation mechanisms outperformed the item-based recommendation mechanism, a collaborative filtering technique and the tag-based recommendation mechanism, a content-based filtering technique.

Although the hybrid recommendation mechanism proposed in this paper relies on a simple combination of item-based and tag-based recommendation mechanisms, the effectiveness of the proposed mechanism is remarkable. Considering a group of people interacting with the same items, together with extracting content from individual's interaction with items, benefits the task of recommendation.

However, in order to improve the performance of the proposed mechanism, further analysis on how to combine a content-based filtering technique with a collaborative filtering technique needs to be performed. It appears that people's interest may change as time progresses. Thus, a more robust technique to keep track of people's interest should be investigated. We would also like to explore how to apply the proposed hybrid approach to build a recommender service on other domains such as a research paper recommender service, and news recommender service, among others.

## ACKNOWLEDGEMENT

The authors wish to thank Chia-Lin Chang and Michael McAleer for helpful comments and suggestions.

## REFERENCES

- Aryal, Jagannath, Ritaban Dutta, and Ahsan Morshed. 2013 "Development of an Intelligent Environmental Knowledge Recommendation System for Sustainable Water Resource Management Using Modis Satellite Imagery." *Proceedings of the 2013 IEEE International Geoscience & Remote Sensing Symposium*. pp. 2204–2207. <https://doi.org/10.1109/igars.2013.6723253>
- Bruke, Robin. 2007 "Hybrid Web Recommender System." *The Adaptive Web*. Springer. Pp. 377-408.
- Choochaiwattana, Worasit. 2015. "A Comparison Between Item-based and Tag-based Recommendation on a Knowledge Management System: A Preliminary Investigation." *International Journal of Information and Education Technology*. 5:754-757. <https://doi.org/10.7763/IJNET.2015.V5.605>
- Davoodi, Fatemeh G. and Omid Fatemi. 2012. "Tag Based Recommender System for Social Bookmarking Sites." *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. pp. 934-940. <https://doi.org/10.1109/asonam.2012.166>
- Deshpande, Mukund and George Karypis. 2004. "Item-based Top-N Recommendation Algorithms." *ACM Transaction on Information Systems*. 22(1):143–177. <https://doi.org/10.1145/963770.963776>
- Goldberg, David, David. Nichols, Brian M. Oki, and Douglas Terry. 1992. "Using Collaborative Filtering to Weave an Information Tapestry." *Communications of ACM*. 35:61-70. <https://doi.org/10.1145/138859.138867>
- Goldberg, Ken, Theresa Roeder, Dhruv Gupta, and Chris Perkins. 2001. "Eigentaste: A Constant Time Collaborative Filtering Algorithm." *Information Retrieval*. 4(2):133-151. <https://doi.org/10.1023/A:1011419012209>
- Huang, Zhenxing, Xudong Lu, Huilong Duan, and Chenhui Zhao. 2012. "Collaboration-based Medical Knowledge Recommendation." *Artificial Intelligence in Medicine*. 55(1):13–24. <https://doi.org/10.1016/j.artmed.2011.10.002>
- Jain, Sarika, Anjali Grover, Praveen S. Thakur, and Sourabh K. Choudhary. 2015. "Trends, Problems and Solutions of Recommender System." *Proceedings of the International Conference on Computing, Communication and Automation*. pp. 955-958. <https://doi.org/10.1109/CCAA.2015.7148534>

- Li, Hong, Lu Liu, and Chenggong Lv. 2006. "Knowledge Recommendation Services Based on Knowledge Interest Group." *Proceedings of the International Conference on Service Systems and Service Management*. pp. 162–166. <https://doi.org/10.1109/icsssm.2006.320606>
- Liang, Kaichun, Shuqin Cai, and Qiankun Zhao. 2007. "Context-based Knowledge Recommendation: A 3-D Collaborative Filtering Approach." *Proceedings of the 5th IEEE International Conference on Industrial Informatics*. pp. 627–632. <https://doi.org/10.1109/indin.2007.4384846>
- Linden, Greg, Brent Smith, and Jeremy York. 2003. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering." *IEEE Internet Computing*. 7:76–80. <https://doi.org/10.1109/MIC.2003.1167344>
- Prasad, RVVSV and V. Valli Kumari. 2012. "A Categorical Review of Recommender Systems." *International Journal of Distributed and Parallel Systems*. 3(5):73-83. <https://doi.org/10.5121/ijdps.2012.3507>
- Resnick, Paul and Hal R. Varian. 1997. "Recommender Systems," *Communications of the ACM*, March, pp. 56-58. <https://doi.org/10.1145/245108.245121>
- Resnick, Paul, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews." *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. pp. 175-186. <https://doi.org/10.1145/192844.192905>
- Ricci, Francesco, Lior Rokach, and Bracha Shapira. 2011. "Introduction to Recommender System Handbook." *Recommender Systems Handbook*. Springer. Pp. 1-35.
- Si, Luo and Rong Jin. 2003. "Flexible Mixture Model for Collaborative Filtering." *Proceedings of the 20th International Conference on Machine Learning*. pp. 704–711.
- Vizcaino, Aurora, Javier Portillo-Rodriguez, Juan P. Soto, Mario Piattini, and Oliver Kusche. 2009. "A Recommendation Algorithm for Knowledge Objects Based on a Trust Model." *Proceedings of the IEEE International Conference on Research Challenges in Information Science*. pp. 93–102. <https://doi.org/10.1109/rcis.2009.5089272>
- Zhao, Wenzhao, Jun Wang, and Guannan Lui. 2009. "A Knowledge Recommendation Algorithm Based on Content Syndication." *Proceedings of the 4th International Conference on Computer Sciences and Convergence Information Technology*. pp. 742–745. <https://doi.org/10.1109/iccit.2009.153>

---

Received on 16-02-2017

Accepted on 13-05-2017

Published on 09-06-2017

DOI: <https://doi.org/10.6000/1929-7092.2017.06.32>

© 2017 Niranatlamphong and Choochaiwattana; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.