Garima Mishra, Jyoti / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 2,Mar-Apr 2012, pp.1503-1507

Optimised Research Paper Recommender System Using Social Tagging

Garima Mishra
Mtech Research Scholar
JSS Academy Of Technical Education
Noida, India

Jyoti
Department Of Computer Science and Engineering
JSS Academy Of Technical Education
Noida, India

Abstract:

In this paper we proposed the method for recommending the research papers to the users. The proposed methodology involves two main steps: obtaining the database from the Google followed by the filtering of research papers on the basis of social tagging. We used Neighbor Weighted Collaborative filtering algorithm. Social tagging is used to provide the optimal results to the user.

Keywords: - Neighbor Weighted Collaborative Algorithm, Social Tagging, Google database.

1. Introduction

Search Engines are becoming important tools to facilitate document searches. The usages of Internet search engines are growing rapidly. The Internet and World Wide Web (WWW) provide various ways to create and store information. Search engines are designed to return the search results according to the queries of the users. Nowadays, social tagging has been widely adopted by various web services such as social bookmarking systems. The systems provide different functions that allow users to share content with one another. There are many works employing search engines for research papers and investigating literature reviews such as CiteULike [1]. Such systems enable the scientists and the researchers to store, organize, share and discover links to academic research papers. The Information Retrieval (IR) research community is highly interested in coming up with accurate and meaningful quality evaluations of search engines. There are two categories of effectiveness measures of search engines: the system-based approach and the user-based approach. The system-based approach relies on objective measurements of system performance, like precision and recall; while user-based approach emphasizes user's subjective evaluations of the search engines, such as perceptions and satisfaction.

Many Recommender Systems have been designed and implemented. The research work in recommender system mainly focuses on improving the accuracy of the recommendations. The main goals are trust, user satisfaction and transparency. The recommender systems also give the explanations to the user that why they may like the recommended research paper. The research paper recommender systems provide an increasingly fragmented and inhomogeneous body of data to the users. Social Tagging has become a very popular tool for categorizing knowledge items over the last decade. Social tagging is of interest to researchers because it is possible that with a sufficiently large number of tags, useful folksonomies will emerge that can either augment or even replace traditional ontologies. As a result, social tagging has created a renewed level of interest in manual indexing. In order for researchers to understand the benefits and limitations of using user-generated tags for indexing and retrieval purposes, it is important to investigate to what extent community influences tagging behavior, characteristic effects on tag datasets, and whether this influence helps or hinders search and retrieval.

The World Wide Web is a huge unstructured corpus of information. Various search engines crawl the WWW from time to time and index the web pages. However, it is virtually impossible for any search engine to have the entire web indexed. Most of the time a search engine can index only a small portion of the vast set of web pages existing on the Internet. Each search engine crawls the web separately and creates its own database of the content. Therefore, searching more than one search engine at a time enables to cover a larger portion of the World Wide Web.

2. Related Previous Work

During the past five years, a group of researchers from many countries around the world conducted research projected examining various aspects on social tagging. Denis Parra-Santander analyzed tags and found that tag suggestion function on any social tagging system could help suggesting "meaningful" tag. Pijitra Jomsri [6] explored communication tags in social tagging system. Siripun [3] looked at the tagging behavior of CiteULike users. Forsati introduced a term - tagsplanations – which are the explanations-based community tags. Jomsri improved research paper searching by using social tagging.

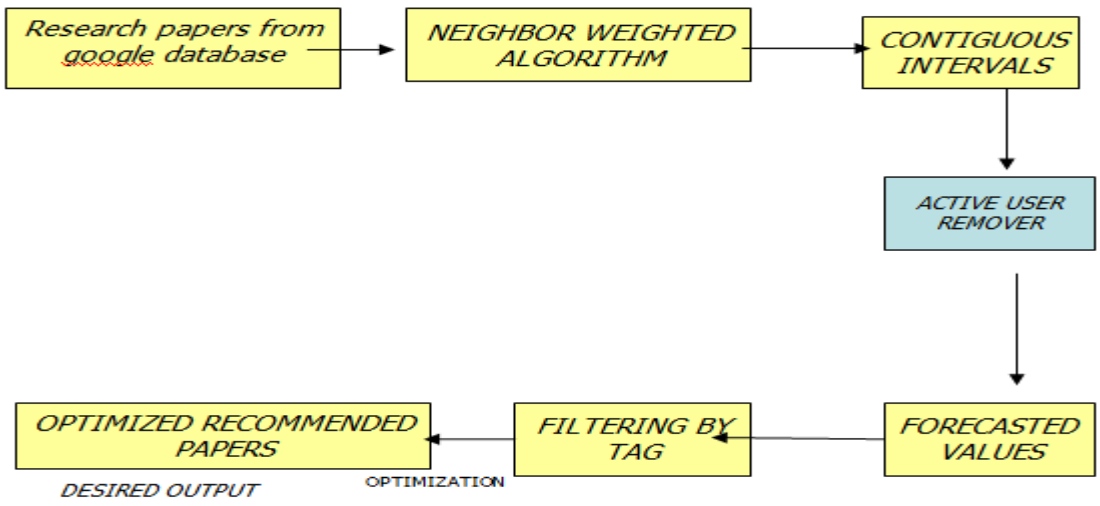
In recent years, many studies of community-based search engines have been carried out. The main techniques involving community-based search engines include recommendation, relevance feedback, personalization, and any of their combinations. The pioneering work on social recommendation was by Peter. Additionally, L. Ardissono[5] studied the origins of personalization. Researchers who presented a framework for community based search engines include: R.Furnari[5] , who introduced standardized cross-site personalization frameworks for the web; A.Goy suggested a Web search framework enhanced by social networks, and studied the mechanisms for content publishing and location in social networks, a framework that led to considerable improvements in effectiveness.

Worasit Choochaiwattana [6] who presented a "semantic annotation approach" to support search in a social network. Researchers who studied and improved folksonomies or social tagging are: Suchanek , who analyzed tags and found that tags are "meaningful" where the tagging process is influenced by tag suggestions.

Walter Christian[8] explored tags for communication in social tagging; Budura presented HAMLET to promote an efficient and precise reuse of shared metadata in highly dynamic content where tags are scarce; M.Segnan used the self organizing characteristics of SOM neural networks to classify popular tags within the "delicious" website; Choochaiwattana[9] examined the use of social annotations to improve the quality of web searches; Karsten[4] compared the information retrieval value of the cloud format tags and the tag words themselves as found in the Library Thing catalog. G.Petrone utilized association rules mining to analyze and structure folksonomies. The results can be used for ontology learning and supporting emergent semantics. Many researchers tried to improve the recommender system: Maximilian [4] introduced ''tagsplanations'', which are the explanations-based community tags. They also examined which types of tags are the most useful for tagsplanations.

3. Research Paper Recommendation System and Community-based Search Engine

Web searches based on social bookmarking, which let users specify their keywords of interest, or tags on web resources, have become increasingly popular. One of the most famous social bookmarking websites in academia is CiteULike, a web-based social bookmarking service and traditional bibliographic management tool. It has been available as a free web service since November 2004. Like many successful software tools, CiteULike has a flexible filing system based on tags. Tags provide an open, quick and user-defined classification model that can produce interesting new categorizations. Benefits from these innovative and time-saving scholarly bookmarking services includes : one-click extraction of bibliographic references**,** tag and rate user references on what others are reading and for sharing resources with peers**,** export libraries in various citation formats**,** import existing libraries**,** access from any computer with Internet connections**.** It can be easily seen that CiteULike is more than just your personal research library. The components of CiteULike are registered and add bookmarks. Collaborative filtering (CF) is the most common method for research paper recommendation.



3.1. *User Profiling*

User profiling is a process of modeling a user's preferences. Each user self-defined tag is then converted to a user profile in a tag vector format. A research paper sharing system provides users with new ways to share their research interests. They can post and comment on papers. They can also discover interesting papers posed by other users who share the same interests. This kind of system allows users to create their own keywords for

attaching to the posted papers. These keywords are known as tags. Tags provide user-defined terms for a paper. The interesting part is that these tags can be used to create a profile for each user. A research paper recommender mechanism could take advantage of this created user profile in research paper recommendations.

4. Approach

One of the first social tagging applications was Delicious, a social bookmarking platform. User-based collaborative filtering process consists of two steps. The first step is finding the neighborhood of the center user, i.e., a set of the most similar users. The second step consists of using this neighborhood to rank the items to be recommended, and recommend the Top N items. These items are taken from the set of items which the neighbors rated positively, and which the center user has not posted on her library.

4.1. *Classic Collaborative Filtering (CCF)*

In the classic CF model, the similarity between two users is calculated using the Pearson correlation over the ratings of their common items. ***The equations[2] used in Classic collaborative filtering are as follows : Similarity is given by-***

$$\frac{\sum_{i \subset CR_{u,n}} (r_{u_i} - \bar{r}_u)(r_{n_i} - \bar{r}_n)}{\sqrt{\sum_{i \subset CR_{u,n}} (r_{u_i} - \bar{r}_u)^2} \sqrt{\sum_{i \subset CR_{u,n}} (r_{n_i} - \bar{r}_n)^2}}$$

In the formula, *r* stands for rating, u denotes the center user and n a neighbor. *rxi* represents the rating given by the user *x* to the item *i*, and is the average rating of the user x over all her items. *CRu,n* denotes the set of co-rated items between *u* and *n*, being *i* an element in that set. Next, we rank the articles of these users to recommend to the center user, using the formula of predicted rating for user *u* with average adjusts.

$$\text{pred}(u,i) = r_u + \frac{\sum_{n \subset neighbors(u)} usersim(u,n).(r_{n_i} - \bar{r}_n)}{\sum_{n \subset neighbors(u)} usersim(u,m)}$$

4.2. *Neighbor-weighted Collaborative Filtering (NwCF)*

This method enhances the ranking step by taking into account the number of raters represented as nbr(i) . It is useful to filter out potentially noisy items, which have been rated by only one or at most two users. The equation that is used in neighbor weighted collaborative filtering is as follows:

$$\text{Pred}(u,i) = \log_{10}(1 + nbr(i)).pred(u,i))$$

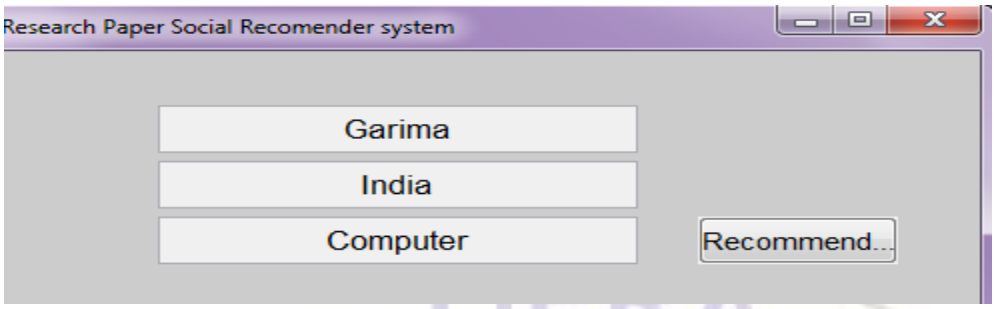
4.3. *Social tagging*

Social tagging has become very popular with the rise of Web 2.0, making it an important piece of the puzzle that makes up the web phenomenon. In contrast to more elaborate ways of organizing resources, such as taxonomies or ontologies, tagging is very easy to use and understand. Because of its simplicity tagging does not create explicit, formalized, structures. One of the first social tagging applications was Delicious, a social bookmarking platform. Many other Internet services such as Flickr, Citeulike or Bibsonomy, make use of tagging for structuring their content. But since the information is being categorized by many individuals who do not necessarily share a common understanding of all or even a part of the subject matter, the challenge of interpreting tags in large group or corporate contexts has been of developing interest. First of all social tagging follows very heterogeneous and individual usage patterns. Each user and each application has different characteristics. In general one can observe a long tail distribution of tag frequencies. This means that users tend to employ some tags very frequently and a huge number of tags very infrequently. Having many different tags leads to information scattering and therefore navigational interfaces based on unfiltered tags quickly become very inefficient.

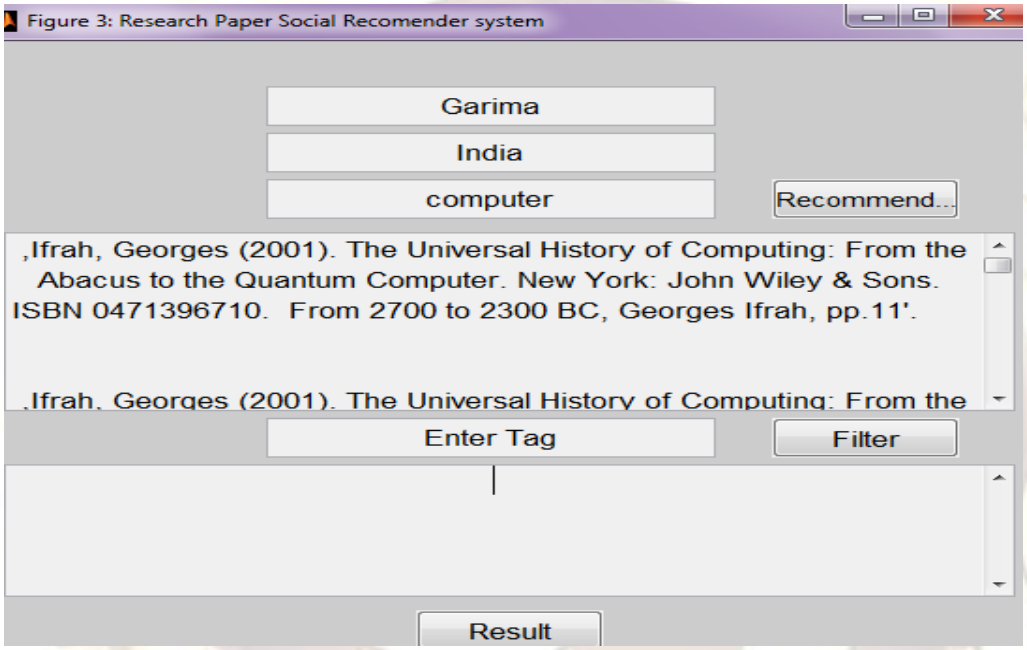
5. Experimental results

The algorithm that has been used in proposing the research papers to the users is as following:
PHASE 1:
The user enters the research area in which he has to obtain the research papers. The research papers are obtained from the Google database and then neighbor weighted collaborative algorithm is applied on the title of all the

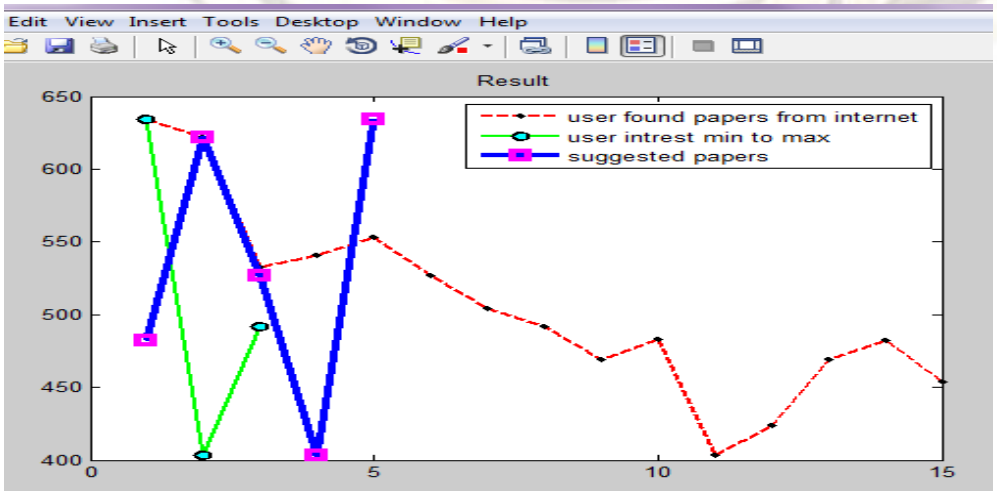
papers. The research papers are then ranked and a ranking matrix is prepared. The research papers are then displayed to the users according to the ranking matrix.



The ranking matrix is obtained by finding the research paper having the maximum number of times the research area entered by the user. The result is then displayed to the user.



PHASE 2:
The user is then asked to enter the tag of his own interest. Then the concept of social tagging is applied on the result displayed to the users. So a précised number of research papers are recommended to the users.



6. Conclusion

Hence by the means of this combined approach a précised set of research papers will be recommended to the users. It will improve both the accuracy and the speed of the search engine. This paper presents a framework for the development of a research paper recommender system. The framework portrays the structural components and architecture for a research paper recommender system. The objective of this framework is to indicate what components and what modules they have to include in their systems. From this informal experiment setting, the recommender system illustrated encouraging results with an overall accuracy percentage up to 91.66%. Therefore, tagging created by users has a potential to be used as a representative of the user profile and paper in a recommendation. However, the number of subjects is considered to be small in the experiment. In order to confirm the findings, more subjects are needed in the experiments.

In the future, a more comprehensive formal evaluation will be conducted according to the proposed framework. The approach to an improved research paper recommender system will be determined.

7. References

[1] CiteULike. Available: http://www.CiteULike.org.

[2] Denis Parra-Santander, Peter Brusilovsky, 2010. **Improving Collaborative Filtering in Social Tagging Systems for the Recommendation of Scientific Articles**, *IEEE International Conference on Web Intelligence and Intelligent Agent Technology*, pp.136-142.

[3] F. M. Suchanek, M. Vojnovi´c, and D. Gunawardena, 2008"**Social Tags: Meaning and Suggestions**," *ACM Conference on tagging.*

[4] J. Gelernter, 2007. **A Quantitative Analysis of Collaborative Tags: Evaluation for Information Retrieval,** *International Conference on Collaborative Computing*, pp.376-381.

[5] L. Ardissono, R. Furnari, A. Goy, G. Petrone and M. Segnan, 2010. **A SOA-based Model Supporting Adaptive Web-based Applications**, *ACM Third International Conference on Internet and Web Applications and Services*, pp.708-713.

[6] Pijitra Jomsri, Siripun Sanguansintukul, Worasit Choochaiwattana, 2010. **A Framework for Tag Based Research Paper system.** *IEEE 24th International Conference on advanced Information Networking*, pp.103-108.

[7] R. J¨aschke, L. B. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme,2007. **Tag Recommendations in Folksonomies**, *In proceedings of PKDD 2007, volume 4702 of Lecture Notes in Computer Science, Springer Verlag*,,pp. 506–514

[8] Walter Christian Kammergruber, Maximilian Viermetz, Karsten Ehms, 2010. **Using Association Rules for Discovering Tag Bundles in Social Tagging Data**, *IEEE International Conference on Computer Information Systems and Industrial Management Applications,* pp. 414-419.

[9] Worasit Choochaiwattana, 2010. **Usage of Tagging for Research Paper Recommendation**, *3rd International Conference on Advanced Computer Theory and Engineering*, Vol.2, pp.439-442.

[10] Xiufeng Xia, Shu Zhang, Xiaoming Li, 2008. **A Personalized Recommendation Model based on Social Tags**, *Proceedings of the ACM conference on Recommender systems*, pp.712-716.