



Design of Research Buddy: Personalized Research Paper Recommendation System

Deepti Sharda¹, Punam Dawgotra²

¹Assistant Professor, DCSA, MCM DAV College for Women, Chandigarh

²Assistant Professor, DCSA, MCM DAV College for Women, Chandigarh

ABSTRACT

Abundance of the information available across the web accumulated over the years has resulted in a large repository of information. Extracting the information of interest has become a challenge for the users especially researchers who look for precise, accurate and quick data for their research work. There is an urgent need for specialized recommender systems to boost the productivity of the researcher by saving the researcher's time spent in searching for latest developments made in a specific field. This paper is an effort to design a personalized research paper recommender system: Research Buddy to provide a personalized assistance in finding the best research paper. The proposed design is based on the concept of creating separate user profiles for each researcher, using the information gathered through user profiles and creating clusters of researchers working in the same field and interests. It uses fusion hybrid recommender technique to recommend new research papers in a specific field or subject and maintains the history of research papers created, downloaded, read, viewed and liked by the user in the past. It also takes long time and short time research needs of the user into consideration while recommending.

Keywords- Clustering, Hybrid recommender, Personalization, User profiling

I. INTRODUCTION

The abundance of information available on the Web and in numerous Digital Libraries meant for academicians and researchers, in combination with their dynamic and heterogeneous nature, has made the search of a suitable research paper a difficult task over the years. Researchers find it extremely difficult to find the appropriate research work when it is needed urgent the most.

It has been observed that researchers spend countless valuable hours searching for the right information they need for their work. If they could spend these hours in actual research work and the difficult task of digging right information from huge repository of data can be handed over to a recommender system, then it could speed up their research. This paper will introduce the design of a personalized research paper recommender system: Research Buddy. It is meant to search, evaluate and recommend research papers to the researchers, strictly according their specific areas of interest by saving their efforts, energy and time.

An informal study was conducted among researchers pursuing different fields of research. The aim was to study and analyze the techniques of search adopted by them and the problems faced during the process of finding the most appropriate papers. After this study it was found that researchers spend many hours daily to locate the research papers, evaluate them for usefulness, assess their authenticity and match their needs. This becomes more difficult since most of the commonly used the search engines by these researchers lack enhanced search



options and end up producing the same list of research papers even if the keywords are redefined and altered. In this study researchers who are physically nearby say in a department or a lab spend a considerable amount of time independent of each other in searching published articles, papers or books relevant to their common project. Despite having similar field of interests, they conduct independent and time-consuming search that leads to wasted effort and time. When they have to consolidate their work they end up having same search results since they were unable to leverage previous and their colleagues' search results.

The proposed model Research Buddy is an effort to handle all these issues thus providing a valuable and time efficient support to researchers. It is designed to save time by recommending research paper matching the current interest of the user. It is personalized system to maintain users' profile so that future recommendations match the previous searches and the ratings given by user. It also maintains detailed record of all the people who are working in the same field and provides the required information on query.

II. RELATED STUDY

Recommender systems have gained immense popularity in commercial and business applications especially the online ones. The benefits of recommender systems have motivated researchers and developers to propose, design and create recommender systems for academic research. This is evident since there is a significant emergence of research papers presented at many conferences and available in various reputed journals. Some of the applications of research paper recommender systems are also included in this paper.

J. Lee et al.[8], discuss about a personalized academic research paper recommendation system. It recommends articles relevant to the research field of its users. It is based on the assumption that users like the articles written by them and so all the papers that match the contents and work in the papers written by a recommender system user are recommended since their relevance score is higher for the user. It uses text similarity techniques to determine the similarity between two research papers and collaborative methods to recommend the items.

B.S. Oladapo [18], designed a research paper recommendation system which used a content-based filtering technique as the recommendation technique. In this paper, Jaccard similarity coefficient is suggested to compute the similarity between user's query (user's attributes) and the attributes of the papers. The recommendations suggested by the system were sent via email to the intended users.

Nascimento et al.[14] provide another example of a content-based recommender system for scientific articles. They point out that most of the recommender systems approaches assume availability of a large collection of scientific papers beforehand. It is true for some digital libraries like IEEE Xplore, but it does not hold correct for many other situations. Their proposed solution depends on publicly available scientific metadata. Instead of using user defined keywords, they generate keywords from a particular article written and submitted by a certain user.

The hybrid approach of recommender system was used for recommending research paper by R. Torres [2]. Techlens recommends using different algorithms for recommending different kinds of papers. Various techniques of combining content-based filtering and collaborative filtering have been compared in this paper. It used a dataset of CiteSeerX. Techlens also points towards the fact that users with different levels of experience perceive recommendations differently.

B. Gipp et al. [17], presented hybrid recommendation system Scienstein in order to make a powerful alternative to academic search engines. Instead of solely relying on text mining, it combines citation analysis, explicit ratings, implicit ratings, author analysis, and source analysis to a recommender system on a user-friendly GUI. CiteSeerX also uses citations to find similar scientific papers[22]. Some other applications with citation recommendation are presented in [19], [20], [21].

S. Patil and P. M. B. Ansari [1], adopted the content-based technique to recommend research papers. TF-IDF and cosine similarity were used to determine how relevant a research paper is to user’s query. It used Keyword-based Vector Space model to depict the relationship between the research paper and user’s query.

K. Hong et al. [5], introduced Personalized Research Paper Recommender System (PRPRS) for keyword extraction and keyword inference. It considers the title and text as an argument of keyword and executes the algorithm. Whenever collected research papers by topic are selected, a updated UserProfile increases frequency of each domain, topic, and keyword. Each ratio of occurrence is recalculated and reflected in UserProfile. It uses Cosine Similarity to recommend the initial paper for each topic in information retrieval.

N. Agarwal et al. [3], introduced a subspace clustering approach for recommender system. It studies the reading habits of other researchers who are interested in similar concepts. It uses collaborative filtering approach to collect data from other researchers browsing patterns and avoids issues with the interpretation of content. It creates groups of people having similar interest. Such a group is represented by a subspace cluster. Finding these experts will ultimately help in finding research papers that form fields of interest.

J. Beel et al. [23], introduced Docear’s Research paper recommender system. It manages the user’s data (papers, references, annotations, etc.) using mind maps. It allows all the users to create their own mind maps which are used by the system to recommend the research papers. It is based on content based filtering technique.

III. PROPOSED SYSTEM

The proposed model “Research Buddy” aims to combine the already known concepts with new ones in order to create a holistic research paper recommender system. By combining different concepts, many disadvantages become obsolete. It tries to merge various techniques suggested earlier to create a personalized research paper recommender system that can overcome shortcomings of these techniques and provide the user with a recommender system that saves their effort and time spent on locating research papers.

The proposed components of Research Buddy are - User Interface, Extractor, Cluster Manager, Cluster Profile, Profile Manager, User Profile, Fusion Hybrid Recommender and Monitor as show in figure 1. It uses CiteseerX Repository to search for the desired research papers matching the keywords provided by the user.

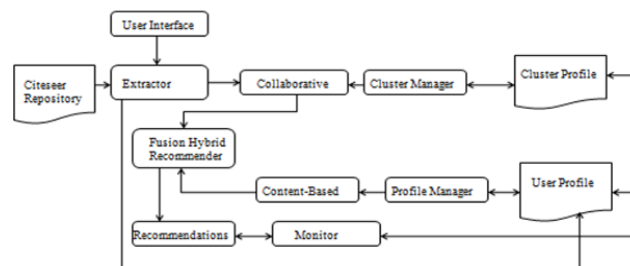


Figure 1 System Architecture of Research Buddy

IV. COLLECTING INFORMATION ABOUT THE USERS

All the users who intend to use Research Buddy will need to get registered. The registration process is the first interaction of a user with this recommender system and he/she must provide the personal and academic details when logging into Research Buddy. This information will be used by the system to create a distinct user profile for the user that will be used each time the user uses this system providing the personalized feel of the system. As a user keeps on using the Research buddy, all the activities done by the user will be used to update the user profile and better user experience for the future.

At the time of registration a user will provide his/her name, *email-id*, address, profession, area of interest, sub-areas, research papers published by the user, and his research preferences. A user profile will be created for the user using this information supplied by the user.

A User profile will be stored on the local machine of the user in a temporary folder. Whenever he or she logs in, the user profile stored on the local machine will be uploaded first. Research Buddy recommender system will use this information and provide recommendations about the research papers available matching the specified user data.

The user profile will be updated back on the local machine whenever user accepts the recommendations or rates the recommendations received by him. The monitor will keep on monitoring user’s activities. It will update the user’s profile based on explicit and implicit rating made by the user. Recommendations rated by him directly or accepted by him through implicit actions will be added to the user profile so that it does not recommend the same research papers time and again.

V. CLUSTERING THE USERS

When a user profile is created, the user will be added to one of the clusters having users working in the same research field. If a matching cluster does not exist then a new cluster will be created and the user will be assigned this new cluster. Information about all the clusters is maintained in Cluster Profile. Cluster Profile will be stored and managed on the server hosting Research Buddy.

VI. BUILDING THE DATABASE

The database building is the most crucial aspect of Research Buddy which is done with the help of Extractor component.

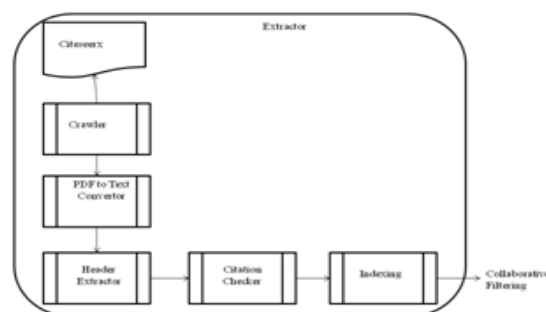


Figure 2 Extractor

Extractor is responsible for extracting the research papers written by the user with the help of crawler. These extracted research papers are then converted into plain text by pdf to text converter. Header information which



includes the title of the research paper, author(s) name, journal name, ISSN number etc. is extracted from these text files. Citation count of this paper is handled by the citation checker. Indexing is done of these extracted research papers. Indexed set of research papers are inputted to collaborative filtering component.

VII. GENERATING RECOMMENDATIONS

Research Buddy uses Fusion Hybrid Recommender to make recommendations. It applies Collaborative filtering and Content-based filtering independently and then the output of both is given as input to Fusion Hybrid Recommender. It merges the results of both the filtering algorithms together. The papers which are recommended by both filtering components are given higher ranking than the ones recommended by either of them.

The content-based filter takes input from the user profile that is updated regularly. The collaborative filter takes its input from the extractor and the cluster manager. Cluster manager handles all the clusters that are stored in cluster profile. Clustering helps the collaborative filter to get information about the research papers liked by other users present in the same cluster.

After generating recommendations through Hybrid Filtering, they are matched them with the paper recommendations available in the user's profile. It will filter out the already recommended research papers and will present the new recommendations to the user. So it reduces all the efforts at the end of the user to filter out the previously searched and used papers. Recommendations generated by Fusion Hybrid Recommender are further passed over to the user through the graphical user interface. It is designed to display ten recommendations at a time.

VIII. ANALYZING THE USER'S SATISFACTION

Once the users are presented with the recommendations, they are asked to explicitly rate these recommendations generated by the recommender. It helps the recommender to judge the user's preferences and hence improves the accuracy of recommending the research papers in future. The explicit rating can be helpful in improving user profile as well as the cluster profile benefitting other users in the cluster as well.

Apart from explicit rating, Research Buddy monitors the actions of the user to judge implicit rating. The monitor is responsible for monitoring and analyzing all the actions of the user. Monitor maintains the record of all the actions taken by the user by logging these actions and uses this information to update user and cluster profile.

IX. RESEARCH BUDDY: A RESEARCH PAPER RECOMMENDER SYSTEM

Research Buddy is a holistic recommender system which provides personalized research interface by maintaining user profile. It applies both collaborative and content-based technique to recommend the research papers and keeps a track on other users working in the same area. Research Buddy is intended to help a researcher by suggesting research papers of a specific interest area.

X. USER PROFILING

In order to recommend papers to the users, we need to keep track of user's interests. This can be done by maintaining user's profile. User Profile represents the user's tastes and opinions about the papers that are



recommended to him. It can vary from a person to person. In order to give personalized touch user profiling is a must.

User Profile can be helpful to find out long-term interests and short-term interests of the user. Long term interests can be judged by monitoring all the research papers that he has read and downloaded till now. Short term interests can be judged by the research papers recently read or downloaded. Short term interest can change from time to time. While recommending the research paper it is necessary for Research Buddy to differentiate between these interests categories based on the duration. For example, in near past a user was interested in 'image processing, but currently he is looking for works done in the 'semantic web'. Recommender system should be flexible and adaptive so that it must recommend research papers related to 'image processing' for the current search actions of the user.

The user can change his/her area of interest at any time. This change is also updated in user profile so that next time research papers related to the current area of interest should be recommended. Research Buddy allows keeping more than one areas of interest.

Research Buddy applies implicit rating along with the explicit rating. Sometimes a user may select a group of research papers which appears interesting to him but later he may ignore many of these recommended papers. It is the responsibility of the monitor to record all the actions taken by the user and update the profile accordingly.

The user actions that can be monitored are enumerated in Table 1.

Read	Edit Document Details
Download	Highlight passages in PDF
Read Abstract	Create Bookmark within PDF
View Document Details	Add Annotations
Bookmark Document	Add Ratings
View Annotations	Add Links
View Ratings	Add Classifications
View Links	Send/Recommend to friend
View Classifications	Print
View Bibliography	Follow Recommendations
View Related Documents.	Reference Document

TABLE 1 Actions Monitored for Implicit Ratings

XI. CLUSTERING

Research buddy uses the clustering technique to group researchers so that search performed and ratings given by one researcher in cluster can help others in a cluster. This technique is suggested in [3]. This approach is helpful to find groups of users who share a common interest in a particular field. This technique is quite helpful to collaborative filtering as it takes care of high dimensionality and scarcity, the two major problems faced filtering.

Research Buddy form clusters based on the information present in user profiles. The user profile maintains the record of the area of interests of all the users and the research papers downloaded and liked by the users belonging to one cluster. Using this information it creates various clusters based on the area of interest. It then monitors the research papers of all the users who are present in the same cluster. This helps to find out a group of research papers that are liked by all the users.

Forming such kinds of clusters can be helpful to a researcher to learn, coordinate and get support from other researchers working in the same field. One of the findings of our initial study with researchers was that despite

having similar interests, researchers belonging to same department generally conduct independent, time-consuming searches. The cluster based operation of Research Buddy will handle this problem of researchers.

XII. FUSION HYBRID RECOMMENDER

The three basic approaches used in the design of recommendation systems are collaborative filtering, content-based and hybrid approach. Collaborative filtering (CF) is one of the most successful techniques used in recommender systems. It has been used to recommend Usenet news[11], audio CDs[12], and research papers[13], among others. CF works by recommending items to people based on what other similar people have previously liked. CF creates neighborhoods of “similar” users (neighbors) for each user in the system and recommends an item to one user if her neighbors have rated it highly. CF is “domain independent” as it does not perform a content analysis of item but rather relies on user opinions about the items to generate the recommendation.

Systems implementing a content-based filtering (CBF) approach analyze a set of documents and/or descriptions of items previously rated by a user and build a model or profile of user interests based on the features of the objects rated by that user.

Hybrid recommendation systems [7] usually use a combination of content-based and collaborative filtering recommendation for recommending items. This combined approach deals with the drawbacks of the discussed filtering strategies, allowing for an initial content-based recommendation in the case of a cold start (lack of user profiles) [14]. The collaborative filtering recommendation can improve the results by adding context-related information to the content-based approach.

Fusion hybrid recommender was suggested by R. Torres et al. [2]. It runs the collaborative recommender and content-based recommender in parallel and generates a final recommendation list by merging the results together. As mentioned above, in the case of Research Buddy, collaborative recommender uses cluster profile as well as other research papers extracted from Citeseerx repository and then making the final recommendations. Content-based recommender uses information available in the user profile to make recommendations.

Fusion Hybrid recommender first gets the recommendations by both the recommenders. It checks out the recommendations that are present in the both the lists and the ranking provided. It sums up the rank of both the lists and terms it as a score. Lower the score, higher will be its ranking in the final list. Recommendations that are not present in other lists will be appended at the end of the final list.

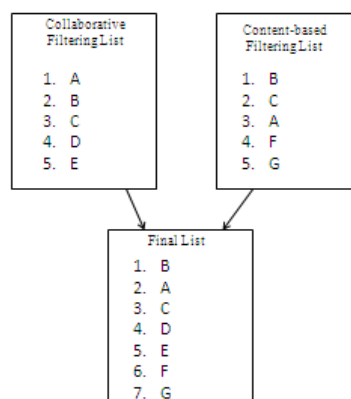


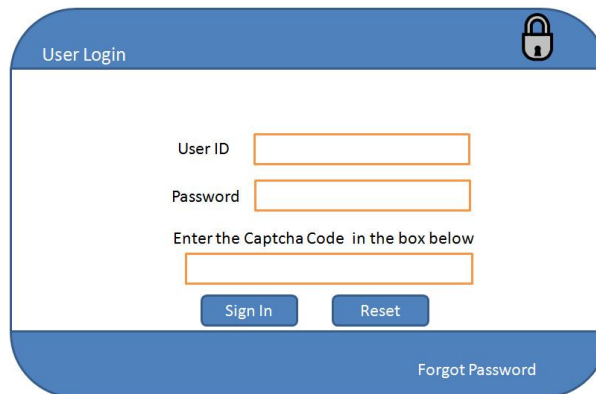
Figure 3 Fusion Hybrid Recommender

XIII. GRAPHICAL USER INTERFACE OF RESEARCH BUDDY

To support the user in managing the information, Research Buddy will offer a user-friendly GUI. It will help the users in locating the desired research papers. Following are the various screens included in Research Buddy:

FIRST SCREEN LOGIN SCREEN

This screen is used to enter username and password. If the user is new, then new user's sign in the facility is included in it along with forgot password and regeneration of password offered for existing users.



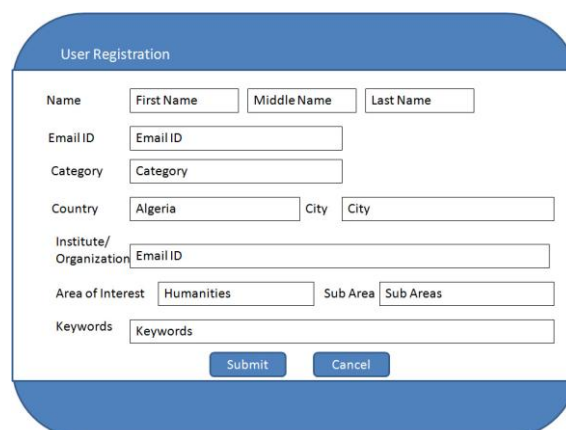
The screenshot shows a 'User Login' screen with a blue header and a lock icon. It contains three input fields: 'User ID', 'Password', and 'Enter the Captcha Code in the box below'. Below the fields are 'Sign In' and 'Reset' buttons. A 'Forgot Password' link is located at the bottom right.

Figure 4 Login Screen

USER'S SIGN IN SCREEN

In this screen, the user will be asked to enter information required to build the user profile. When user will enter his or her name then extractor will search for the research papers written by the user and will list all the research papers written. Provision to manually enter information related to research papers written by the user is also provided to quicken the profile generation activity.

Research Buddy judges the area of interests of the user through the list of research papers written by the user. Facility to manually enter the interest area is also provided to user to enter it afresh or update it whenever the user wants to. It accepts more than one area of interests. It asks the user to rank them. Through this information, his user profile will be created. This user profile will be updated in case of change of area of interest and ratings given by the user.



The screenshot shows a 'User Registration' screen with a blue header. It contains several input fields: 'Name' (split into First Name, Middle Name, Last Name), 'Email ID', 'Category', 'Country' (with a dropdown showing 'Algeria'), 'City', 'Institute/Organization', 'Area of Interest' (with a dropdown showing 'Humanities'), 'Sub Area', 'Sub Areas', and 'Keywords'. 'Submit' and 'Cancel' buttons are at the bottom.

Figure 5 User Profile/ Registration Screen

WELCOME SCREEN

After signing in the welcome screen appears giving users different options like whether the user wants to search for a research paper or persons working in the same area along with their demographic information.



Figure 6 Welcome Screen

RESEARCH PAPER RECOMMENDER SCREEN

This screen is divided into three parts. One part shows the name of the user and the area of interests selected by the user. The user will be able to change the area of interest on this screen as well. The second part of the screen shows the list of recommended research papers. It will be further divided into two parts. One showing new recommendations and the other showing the list of previously viewed research papers. This demarcation between old and new research papers will save the time of the user. Research papers which user views, bookmarks or downloads will automatically be shifted into previously viewed research papers. The third part of the screen will be used for showing the abstract of selected paper.

Facility of narrowing of search is also provided i.e. if the user further wants to select sub-area then it will be possible to do so. Another filtered list of recommendations will be provided. This feature can increase accuracy and satisfaction of the user.

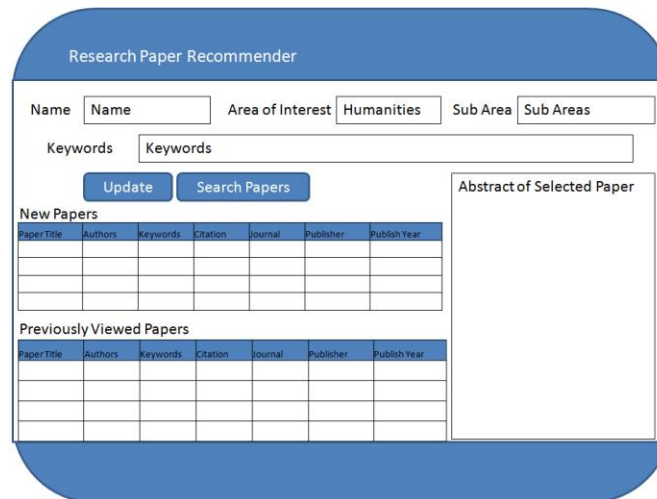


Figure 7 Research Paper Recommender Screen

EXPLICIT RATING SCREEN

Before logging off user will be asked to rate the recommended papers. The user will be asked to select the research papers which he wants to add to his profile. This explicit rating of the research paper will be helpful in learning the user's preferences and interests.

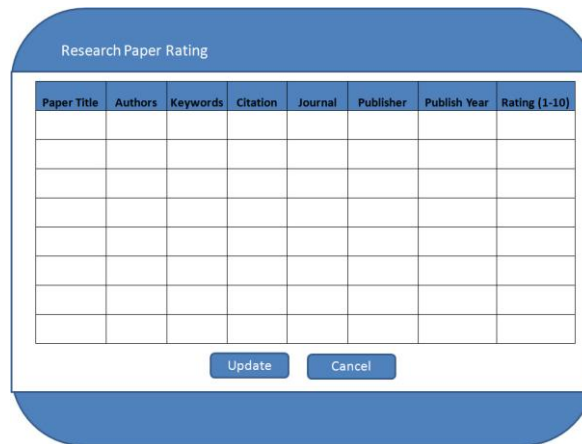


Figure 8 Paper Ranking Screen

XIV. CONCLUSION

This paper introduces the research paper recommender system: Research Buddy. The proposed model “Research Buddy” has combined already known concepts with new ones in order to create a holistic research paper recommender system. It uses fusion hybrid recommender to make recommendations, thus overcoming the shortcomings of both content-based and collaborative techniques. It maintains user’s profile, which helps in providing the personalized experience to the user. It also maintains the information about various people working in the same field by making clusters. Suggested Graphical User Interface is also very powerful catering all the requirements of users and still is simple and easy to use. This paper defines and highlights different component of Research Buddy in detail. It will provide a complete understanding of how Research Buddy works and handles various issues. The future plans for Research Buddy includes its detailed design, implementation and releasing the beta version.

REFERENCES

[1] S. Patil and P. M. B. Ansari, “User Profile Based Personalized Research Paper Recommendation System Using Top-K Query,” *vol. 5, no. 9*, pp. 209–213, 2015.

[2] R. Torres, S. M. McNee, M. Abel, J. a. Konstan, and J. Riedl, “Enhancing digital libraries with TechLens,” *Proc. 2004 Jt. ACM/IEEE Conf. Digit. Libr. 2004.*, pp. 228–236, 2004.

[3] N. Agarwal, E. Haque, H. Liu, and L. Parsons, “Research Paper Recommender Systems: A Subspace Clustering Approach,” *Lect. Notes Comput. Sci., vol. 3739*, pp. 475–491, 2005.

[4] J. Beel and S. Langer, “Research Paper Recommender Systems: A Literature Survey,” *Under Rev. Pre-print available* <http://www. ...>, 2014.

[5] K. Hong, H. Jeon, and C. Jeon, “Personalized Research Paper Recommendation System using Keyword Extraction Based on UserProfile,” *J. Conver. Inf. Technol., vol. 8, no. 16*, pp. 106–116, 2013.

[6] J.L. Herlocker, J.A, Konstan, L.G. Terveen, and J. Riedl, “Evaluating Collaborative Filtering Recommender Systems,” *ACM Transactions on Information Systems, 22(1)*, pp. 5-53, 2004.

[7] M. Balabanovic and Y. Shoham, “Combining content-based and collaborative recommendation”, *Comm. ACM, vol. 40, no.3*, March 1997, pp. 66 -72

[8] J. Lee, K. Lee, and J. Kim, “Personalized Academic Research Paper Recommendation System,” *arXiv Prep. arXiv1304.5457*, pp. 1–8, 2013.



- [9] U. H. Security, "A Roadmap for Cybersecurity Research," *Dep. Homel. Secure Sci. Technol. Dir.*, vol. 2009, no. November, p. 126, 2009.
- [10] T. Bogers and A. Van Den Bosch, Collaborative and content-based filtering for item recommendation on social bookmarking websites, vol. 532. 2009.
- [11] Resnick, P., N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the ACM 1994 Conference on Computer Supported Collaborative Work (CSCW '94)*, Chapel Hill, NC, 1994, pp. 175-186.
- [12] Shardanand, U. and P. Maes. Social information Filtering: Algorithms for automating "word of mouth". In *Proceedings of the 1995 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 95)*, Denver, CO, 1995, pp. 210-217.
- [13] McNee, S., I. Albert, D. Cosley, P. Gopalakrishnan, S.K. Lam, A.M. Rashid, J.A. Konstan, and J. Riedl. On the Recommending of Citations for Research Papers. In *Proceedings of the ACM 2002 Conference on Computer Supported Cooperative Work (CSCW 2002)*, New Orleans, LA, 2002, pp. 116-125.
- [14] C. Nascimento, A. H. Laender, A. S. da Silva, and M. A. Gonçalves. "A source independent framework for research paper recommendation." In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, pp. 297-306. ACM, 2011.
- [15] K. Hoxha, A. Kika, E. Gani, and S. Greco, "Towards a Modular Recommender System for Research Papers written in Albanian," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 4, pp. 160–167, 2014.
- [16] K. Hong, H. Jeon, and C. Jeon, "Advanced Personalized Research Paper Recommendation System Based on Expanded UserProfile through Semantic Analysis," *Int. J. Digit. Content Technol. its Appl.*, vol. 7, no. 15, pp. 67–76, 2013.
- [17] B. Gipp, J. Beel, and C. Hentschel, "Scienstein : A Research Paper Recommender System," *Scienstein A Res. Pap. Recomm. Syst.*, no. January 2009, pp. 309–315, 2009.
- [18] B.S. Oladapo: A research paper recommender system, 2013 unpublished.
- [19] T. Strohman, W. B. Croft, and D. Jensen, "Recommending citations for academic papers.", In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '07*, pages 705–706, New York, NY, USA, 2007. ACM.
- [20] T. Rogers, and A. van den Bosch, "Recommending scientific articles using CiteULike", *Proceedings of the 2008 ACM conference on Recommender systems (2008)*, 287–290.
- [21] J. Tang and J. Zhang, "A discriminative approach to topic-based citation recommendation", In *T. Theeramunkong, B. Kijssirikul, N. Cercone, and T.-B. Ho, editors, Advances in Knowledge Discovery and Data Mining, volume 5476 of Lecture Notes in Computer Science*, pages 572–579. Springer Berlin / Heidelberg, 2009.
- [22] C. Giles, K. Bollacker, and S. Lawrence. "CiteSeer: An automatic citation indexing system.", *Proceedings of the third ACM conference on Digital libraries. ACM, 1998*.
- [23] J. Beel, S. Langer, M. Genzmehr, and A. Nürnberger, "Introducing Docear's Research Paper Recommender System," in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'13)*, 2013, pp. 459– 460