

User Profile Based Paper Recommendation System

Buket Kaya*¹

Accepted : 23/05/2018 Published: 29/06/2018

Abstract: With the spread of science and the increasing number of researchers working in academic fields, there has also been a significant increase in the number of academic publications. Researchers always follow new works published for keeping their knowledge up to date. However, due to thousands of academic publications published every day from many academic sources, academics are not always able to find publications about their subjects. Today, almost all of online academic databases employ a recommendation module that only considers the studies similar to the paper that the user looked at. However, a recommendation system based on the information of a single article is often not enough. In this study, the proposed method recommends by considering user's publications, user's co-authors and co-authors' papers. Therefore, meta-data of the articles published by the researcher in the past are evaluated as time-awareness by the method we proposed. In this way, the most relevant articles to the user's profile can be found by using the proposed method in the data repository created from the exact contents of hundreds of thousands of academic works. The method uses TF-IDF frequency-based similarity analysis method. In the evaluation phase, the performance of the proposed method was examined. The accuracy of the method was measured by several different tests. The results are very promising and demonstrate that the method can produce accurate and quality results.

Keywords: Paper recommendation system, Profile based recommendation system, TF-IDF

1. Introduction

Researchers in all science areas want to reach fastly the publications they are interested in. In the past, people often used physical libraries, printed publications and books to study and access information. Nowadays, with the widespread use of computers, access to information has become easier. With the development of technology, academic search engines such as online digital libraries and Google Scholar are being widely used by researchers every day. For example, IEEE Explorer¹ site, which is one of the leading organizations in Electrical and Electronics Engineering and Computer Science, enters approximately 240,000 new publications each year.

Researchers usually write key words in search terms when searching for articles or presentations. They try to access the related articles by expressing the problem with a few keywords. Often they also want to follow the publications of people who have worked in same area. In particular, authors who are working on the same field and whose studies are cited by many people are frequently followed by other researchers.

The purpose of this study is to recommend the academic publications specific to the user's profile by looking at user related metadata such as published articles, the publication year of these articles, and number of citations to articles. There are some paper recommendation systems in some digital libraries. However, they usually look at only single paper's data. For example, when IEEE Explorer finds similar articles for users, it considers only the content of the article; it does not take into account the profile of the author.

In previous studies, the articles that are similar to the reviewed article are listed and the same articles were recommended for

everyone who were examining the reviewed article. In this study, many different data such as the number of articles in the profile of each user, the number of citations of the articles, the year in which the articles are published, article key words are taken and the values calculated from these variables are sorted by the scores of the articles in the data repository. Finally, n articles with the highest value are presented to the researcher. In the experiments, Sobiad Citation Index² that contains more than 200,000 academic works is used as dataset. The one of the most important reasons of using the Sobiad Citation Index is that it has already created a profile for many users by system and/or users. Three different models were used in order to test the proposed method. First, the results obtained by using user's profile have been sent to the user by email and the number of clicks on the e-mail has been counted. The second method is to make a short survey by showing the results to the users who visited the website. Finally, the success rate was measured by comparing the results of the academic publication recommendation with the existing data set.

The rest of the paper are as follows: Section II contains the related work, the advantages and disadvantages of studies similar to this study are discussed. In Section III, the data collection is first explained and then the architecture of the study is given. Finally, the proposed approach is described. Section IV gives the experimental results and evaluates them. Section V includes the conclusions and the future work.

2. Related Works

The popularity of the Internet, the increase in the number of academic publications and the popularization of online academic search and retrieval processes have made academic recommendation systems more important. Although there have been many studies on the tagging recommendation system, the

¹Department of Electronics and Automation, Firat University, Turkey

* Corresponding Author: Email: bkaya@firat.edu.tr

number of studies addressing academic publications is relatively small. One of the prominent studies on the tagging academic article recommendation method was proposed by Choochaiwattana [1] and the other is Bahulikar's [2] work. Choochaiwattana has created an index of publications by tagging academic publications, and has recommended making a recommendation system using these tags. However, it is not possible to find a tag for every publication in systems where the tags are listed. Ravi et al. [3] proposed a model that analyzes text by using Recurrent Neural Networks (RNN), and recommended academic publications accordingly. Ontology-based academic publications have also been recommended [4][5]. Hybrid approaches are also available within studies that recommend academic publications. Lee et al. [6] presented a new model combining two different approaches, content-based and graph-based, while Bancu et al. [7] proposed an approach that combines content-based and collaborative filtering methods. Apart from these, Zhao et al. [8] combine two different sets of data and use it in methods that use it as an academic publication recommendation method. It is seen from these studies that the success rates of hybrid methods are relatively high. West et al. [9] ranked the citations that are represented by nodes using Eigenvector algorithm. Nodes are clustered using Map Equation. Then the most important nodes were found in different steps of the hierarchy of clusters. An algorithm similar to PageRank was used while this node is found. The most important node found is presented as an article recommendation. Xia et al. [10] conducted a study that recommended articles with co-authors in articles containing more than one author. Zhou et al. [11] proposed a PaperRank algorithm similar to the PageRank algorithm, unlike all other studies, and constructed a recommendation system that tried to extract the confidence index of each academic publication on the data obtained there. One of the major problems encountered in implementing the PageRank algorithm on the academic area is that publications generally decline in value as they go past, while the number of citation increases in the past.

In another study, a recommendation system based on references of academic publications is available in [12]. The emphasis of this work, unlike the number of citations, is to score the citations by looking at the information, such as how many different sources and how many different academics have come from, in terms of the number of main cited citations. Finding similarity of different publications with keywords can be called a tag-based similarity computation algorithm. Although many approaches have been proposed by researchers for this method, the most simple and easily integrated TF-IDF method has been used more often. There are many studies [13] [14] [15] [16] [17] that use this method for what we use.

Many social network analysis methods, such as degree distributions, diameter, clustering coefficient [18]–[21], are intended to find user-based impact values at macro level. The majority of studies on social impact are also seeking a qualitative value [22], [23]. There are similar recommender systems proposed using bipartite graphs [24][25][26]. Ohta and colleagues [27] developed an academic publication recomender model using bipartite graphs. A similar study was done by Jeon and Jeon [28]. The researchers proposed a model named PRPRS (Personalized Research Paper Recommendation System) based on the user profile. This model also designed an algorithm to extract the key words according to the theme for each user profile and ranking was done according to the frequency of the key. Cosine equality was used to calculate the closeness of the articles to the subject.

Xue et al. [29] and Chen et al. [30] did the other recent studies. They analyzed the user's previously published studies to determine

the user's interest. The greatest original feature of Xue and his colleagues' work is the application of a real online academic database algorithm.

This paper uses a keyword extraction model similar to that of Jeon and Jeon [28]. These authors searched key words using the text processing methods between the title and the abstract. In the proposed study, Sobiad already keeps keywords and other metadata separate. This was done with the publishers, not with the program. Also in the previous study, the keyword was categorized in the subject heading. In the proposed study, each keyword was evaluated independently. Zhao et al. [8] performed their work by combining two different sets of data. The proposed study used several different data sets like this one. The previous study, on the other hand, used the same type of data sets. They have combined and used multiple data sets containing publications. In this study, data sets used for different purposes are handled together. One of these is the dataset in which the publications are located and the other is the dataset in which authors and author relations (such as institutional information) are found. As in Chen and colleagues [12], cited value has also played an important role in this study. It is stated that TF-IDF is used in many methods proposed in this area. TF-IDF was used in this study as well. Unlike other studies, TF-IDF was not implemented individually, but an Apache Lucene3 based architecture was indexed and TF-IDF was applied to this architecture. Using a Lucene-based application, the operation of the system has been accelerated very seriously.

3. The Proposed Method

3.1. Data Collection

Sobiad, Social Sciences Citation Index, is an academic index, which covers more than 1,000 journals and more than 220,000 academic articles. All the articles include full text. Some of the journals come back from 1980s. At the same time, Sobiad has more than 200,000 researcher profiles. These profiles include information such as academic scholarship, citation information, and interaction with academics at the institution, as well as researchers in Google Scholar. Because Sobiad actively used by more than 73 universities in Turkey now it takes nearly 1000 visitors per day. In 2017, more than 70,000 citations were searched. Sobiad save the words and phrases that users use for search in open-source PIWIK analytical tool. Thus, a lot of information is stored, including information about how long the user has been on which page. Although Sobiad does not work with the membership system, Sobiad keeps track of which pages users spend on which pages, and which pages they visit in order to keep all the records on the sessions. The user is remembered on the next visit with his/her cookies placed on his/her computer during his/her first visit and recording of user movements with the same session recording continues. However, since the user does not have a membership entry, the entry remains anonymous. Only the IP address can be used to identify the institution and this information is also recorded.

Users use Sobiad to see citations to their publications, to search for citations for a topic, and to find similar publications for their own work.

4. System Architecture

This work, which aims to recommend an article specific to a user's profile, consists of 5 main steps / modules.

4.1. Listing user's published papers

The first step is to find all the works published by the selected user

and extract the user's profile properties based on these works. All publications in Sobiad have the author name part. The selected user is searched in the author names section, listing all publications of the author.

4.2. Keyword-specific scoring of each academic publication

The second step is the scoring part of the academic publication. The score of an academic publication is calculated with the combination of the citation received in the year of publication, the journal in which it is published, and, if so, the score of co-authors. It has also been used in data from user actions, such as how many times and how long it has been reviewed. When the year score is calculated, the most recent year is multiplied by the coefficient of 1. 2000 was considered the lowest year and was multiplied by 0.1 with the score of the academic publication. All numbers between 2000-2018 affected the rating of the publication to increase iteratively. Thus, it is ensured that the score of the work done in the near today is higher. Although more recent work is recommended, the yearly coefficient is used to compensate for this advantage, given that past publications have received more citations.

For example, let's use the values u and v as the current year and the year the article was published, respectively. Then, the year-based weight of the publication is found as follows:

$$W^{(P_{(u \rightarrow v)})} = 1/(u-v) \quad (1)$$

In order to calculate the score according to the keyword, it is taken into account how often the keyword is passed in the related work and where it passes. The TF-IDF method was used to find this score in the study. The formula 2 shows how TF-IDF is applied in practice.

$$(tf \text{ [idf]})_{(t,d)} = [(tf)]_{(t,d)} \cdot [(idf)]_{(t,D)} \quad (2)$$

where t is keyword (term) and d is paper and D is the sequence of papers. The study also analyzed the documents for each term as shown in formula 3. If there is more than one term, the result value is obtained by summing the score of all the term sequence by the following formula, where q is term sequence in the paper.

$$\text{Score}(d) = \sum_{(t \in q)} [(tf \text{ [idf]})_{(t,d)}] \quad (3)$$

4.3. Calculation of the researcher score

This value is calculated independently from the key word according to the selected user profile, increasing according to the user's publications in recent years and the number of citations received.

The total score of the researcher is given in Equation 5, while the score of the researcher (F) in each publication (P_i) is shown in Equation 4.

$$F^{P_i} = \sum_{(x=1)^k} W^{(P_{(u \rightarrow v)})} \cdot f^{(P_{(u \rightarrow v)})} \quad (4)$$

$$P_{\text{User}} = W^{(P_{n \rightarrow 1})} + F^{P_1} + W^{(P_{n \rightarrow 2})} + F^{P_2} + \dots + W^{(P_{n \rightarrow n})} + F^{P_n} \quad (5)$$

The representation of the method expressed by the formulation is depicted in Figure 1.

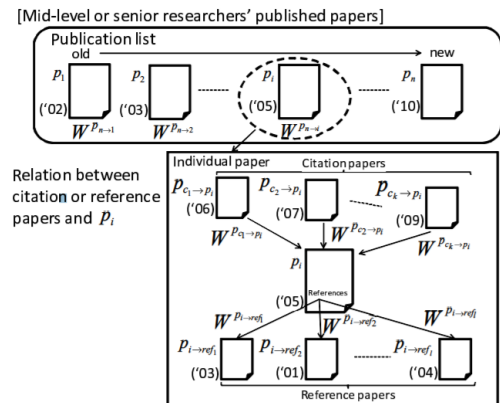


Figure 1. Calculation of researcher score

4.4. Finding scores of keywords

Keywords in all publications of the selected researcher are extracted and these keywords are evaluated by the score of the publication they are in. Keywords are generally expected to cross more than one article. In such a case, the keyword is aggregated with the scores of the articles in it and the overall score is obtained. Keyword discovery can be viewed as weighted TF-IDF. Where t is the term and d document and q is the sequence of terms. The study also analyzed the documents for each term as shown in formula 6.

$$(tf \text{ [idf]})_{(t,d)} = [(tf)]_{(t,d)} \cdot [(idf)]_{(t,D)}, t \in q, d \in D \quad (6)$$

For example; Let the score of X paper be 1.3 and it contain three keywords, a1, a2, a3. Let the Y article's score be 0.8 and it include the keywords, a1, a4. Finally, let the Z article with score 0.3 contain a1, a3, a5 keywords. In this case, as shown in Table 1, the keyword a1 has the highest score.

Table 1. An example of scoring keywords

Keyword	Academic Publications and Scores			Total
	X (1.3)	Y (0.8)	Z (0.3)	
a1	1.3	0.8	0.3	2.4
a2	1.3			1.3
a3	1.3		0.3	1.6
a4		0.8		0.8
a5			0.3	0.3

Within these keywords, the highest n values are taken and the repository containing all the articles is scanned. Keywords are scanned in WordNet for English before being processed, and in the Zemberek NLP library for Turkish. In these libraries, which are used as a dictionary, the conjunctions and prepositions of the keyword candidates are removed and the first n highest-ranked keywords are retrieved.

4.5. Selection of related academic publications according to keywords from data repository

In order to find the closest publication to the searched words in the articles, this study utilized the widely used TF-IDF method. The highest ranked key words selected are compared to the key words of all academic works in the repository. The paper whose keywords are mostly matched to the searched ones is handled as the highest score. The k academic publications with the highest match are identified and listed.

For the k academic publications received from the repository, the operations in the second module are repeated. As the main source

in the calculation, the match between the keywords in the user profile and the articles in the repository is considered. The highest rated academic publications are recommended to the user if it is not one of his own works. Figure 2 shows a flow chart describing the operating principle of the system.

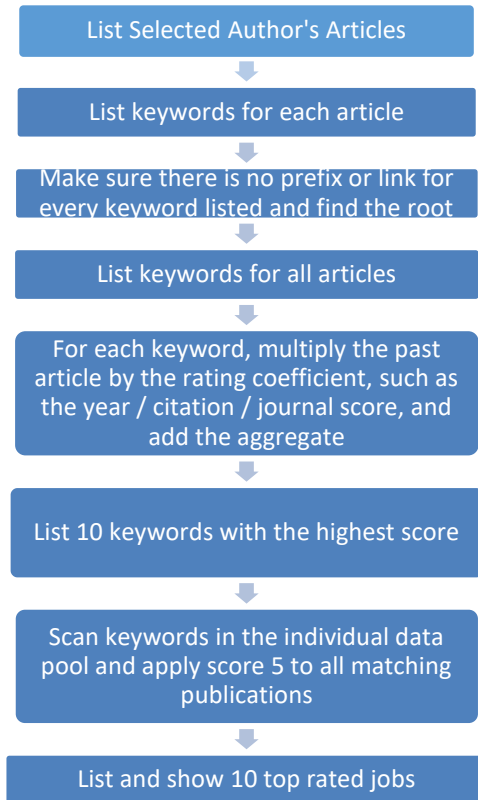


Figure 2. Flow diagram of the operation of the system

The study was implemented with a simple interface. The name of the researcher whose profile properties are to be extracted is entered into the name form. The system, in turn, finds the researcher's publications, explores the highest rated keywords, and lists the recommended publications. A screenshot that finds expertise by keywords or subject is given in Figure 3.



Figure 3. Screenshot of the form for researchers

5. Experimental Results

In this section, the obtained results of the recommendation system are presented to real users with three different methods and the user reactions are noted. The success rates of the results produced by the proposed method are evaluated.

5.1. Academic Publication Recommendation by E-Mail

A similar approach to the model described in this study was not found in the literature earlier although there are some close studies. For example, Hassan [31] focused on publications only viewed by

users, and has not conducted a study on the publications of its users. Achakulvisut and his colleagues [32] made article recommendations in their work. However, there is no such thing as user profile creation in their work. TF-IDF was used instead. Therefore, no direct comparison was made with any of the previously proposed methods, and the tests were conducted using 3 different evaluation methods. In the first method, results obtained by the academic publication recommendation system were sent by e-mail to real persons. A link has been put in order to reach the details of the publications in E-mail. In order to understand whether the mass mailings have been read, an image with parameters per sent e-mail has been put. Thus, the number of people who read the e-mail was determined and the success rate was tried to be extracted from the rate of clicking on the link in the e-mail. It is thought that some e-mail clients do not automatically display pictures for safety reasons, and some researchers do not click on the recommended academic publication for different reasons. Therefore, this method of comparison alone does not give objective result. The results are given in Table 2, where the value of 'n' is the number of articles found by the author's key words, 'k' is the number of keywords selected for the data repository, and 'm' is the number of the last article in the repository. These tests are all measured from the responses of users to e-mails.

Table 2. Test Results For Type A Experiments

n Value	k Value	m Value	Total Outgoing Mail	Read Mail Number	Clicked Mail number	Reading Mail Number/ Total Outgoing Mail	Clicked Mail / Read Mail
5	5	10	21	15	8	0.714	0.533
5	10	20	53	26	10	0.491	0.385
5	20	30	156	41	23	0.263	0.561
10	5	10	26	13	8	0.500	0.615
10	10	20	65	25	9	0.385	0.360
10	20	30	253	129	60	0.510	0.465
20	5	10	36	11	3	0.306	0.273
20	10	20	92	39	15	0.424	0.385
20	20	30	469	216	65	0.461	0.301

The main reason for the low read and click rates in the table is that it is not a common and accepted method to send the papers / articles by e-mail and to access the papers by clicking the links in these e-mails. This is not only considered as a failure of the method, but it is also interpreted as the absence of much interaction with such e-mails. In the test, the values of 'n', 'k' and 'm' were changed.

Title	Authors	Journal	Year	
Doğu Anadolu Ağızlarında Ara Üniteler	İbrahim TOSUN	DİYALEKTOLOG Dergisi	2012	Detay
Arthur Lumley Davids'e göre XIX. yüzyılda Osmanlı Türkçesi Ottoman-Turkish at theXIX th century according to Arthur Lumley Davids	ERASLAN CANDAN	YÖK Ulusal Tez Merkezi - Ege Üniversitesi	2013	Detay
YABANCILARA TÜRKÇE ÖĞRETİMİNDE ZAMAN ALGISI	Çiğdem ŞAHİN , Serdar YAVUZ	Kesit Akademi Dergisi	2015	Detay
YABANCI DİL OLARAK TÜRKÇE ÖĞRENEN B1 DÜZEYİNDEKİ GÜRCÜ ÖĞRENCİLERİN YAZMA HATALARI	Mesut GÜN, Mehmet Dursun ERDEM, Mesut GÜN, Hamza SALDUZ, Betül KARATEKE	Journal of Educational Sciences	2015	Detay
TÜRKÇENİN YABANCI DİL OLARAK ÖĞRETİMİNDE BİLİŞİMDEN YARARLANMA: PARAGRAF ÇALIŞMALARI	Seda ARTUÇ	International Journal of Languages Education and Teaching	2013	Detay

Figure 4. An e-mail sample sent to the user

Since 4 different e-mails cannot be sent to the same person, e-mail groups are composed of different people. It has been found that the success rate is low in the e-mail sent with too little or too many keywords. The main reason for this is thought to be the fact that when working with too little or too many keywords or articles, it focuses on a very specific subject, or that very irrelevant subjects are added to the list. The ideal numbers were calculated as having 20 close articles for 10 keywords and totaling 30 articles from the repository. Figure 4 shows a screenshot of a sample mail to the user.

5.2. Academic Publication Recommendation by Online Survey

In Sobiad, where we use the dataset and servers, researchers were asked to evaluate whether they liked the publications recommended for the field where they could examine and edit their own profiles. The evaluation results of the persons participating in this evaluation are given in Table 3.

Table 3. Test Results For B Type Experiment

n Value	k Value	m Value	Survey Number	Completed Survey	Average Success Rate
5	5	10	13	10	0.769
5	10	20	25	12	0.480
5	20	30	62	20	0.322
10	5	10	6	4	0.666
10	10	20	35	20	0.571
10	20	30	86	48	0.558

Since multiple keywords failed in the practice described in the previous section, the 20-keyword series was not used in this test. It can be easily understood from the table that users are reacting more to filling in short questionnaires. Since less than 10 results are not considered, it is clear that successful results are found with low 'n', 'k' and 'm' values. As a consequence, it is based on more specific subject-oriented outcomes. The reason why it differs from the practice in section A is that 10 models are more suitable for clicking on curious and unexamined publications, whereas in the questionnaire scoring method, there is no question about known and previously read studies. Also, although the system is designed to exclude authors from their own work, there are also survey users who see their work in this section because of the abbreviations of non-English characters, title changes and names in the users' names. For this reason, these users have given high marks to their own work.

5.3. Comparing the Results of the Academic Publication Recommendation with the Existing Dataset

Another method of evaluation is to find the average of the similarity scores of the selected users of the recommended publications with their own studies. In contrast to other methods, this method allows a very high number of tests. However, the results are not expected to give a very different result because they are made by feeding the data, which is already tested. Information including the keyword consistency, text similarity, and the number of publications of the user's publications are provided in Table 4. The text correlations are found in each text with a mutual

match of at most 20 words in itself. When keyword correlations are calculated, not only the key words given by the author but also the system generated keywords are used.

Table 4. Test Results For C Type Experiment

n Value	k Value	m Value	Test Number	Keyword Correlation	Text Similarity Correlation
5	5	10	5000	0.789	0.891
5	10	20	5000	0.681	0.786
5	20	30	5000	0.598	0.725
10	5	10	5000	0.723	0.831
10	10	20	5000	0.621	0.734
10	20	30	5000	0.581	0.612
20	5	10	5000	0.723	0.765
20	10	20	5000	0.621	0.616
20	20	30	5000	0.569	0.589

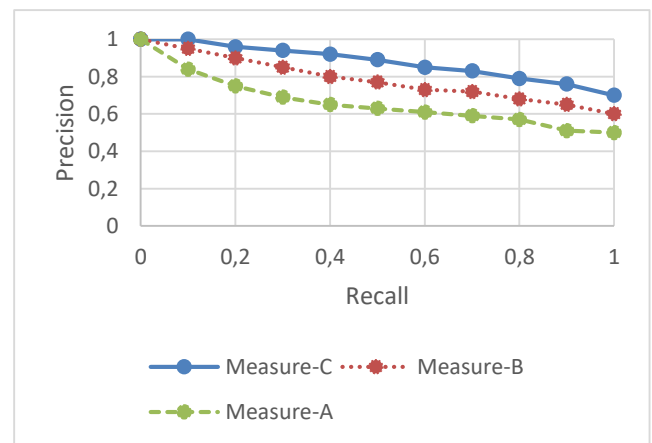


Figure 5. The precision-recall curve of the proposed measures

The results in Table 4 appear to be related to the results in Section A and the results in Section B. As the amount of keywords and data in all sections increases (as the number of different works made in different disciplines increases), the success seems to decrease. On the other hand, processing in a data repository containing only certain topics has provided more accurate results because the author matches publications with more accurate publications.

In the last experiment, we compared the performances of the proposed measures on 200 publications. The half of these publications was selected as positive instance and the other part was determined as negative instance. As the comparison criteria, we used the precision-recall curve. Figure 5 reports the results of this experiment. As can be easily seen from the figure, the measure C, that is, "comparing the results of the academic publication recommendation with the existing dataset" outperforms the other measures. The worst measure was obtained by the measure A (academic publication recommendation by e-mail).

Since the proposed method and the data set used are novel, many problems have been encountered. While some of these problems are being solved, some are overcome with assumptions. In this section, the details of the problems that are overcome with assumption will be expressed. In the final part, plans will be transferred on further studies.

There are academicians among the academicians whose name is partly or completely the same. Because the proposed model starts

the process by searching by author name, it causes some results to be wrong even if there are only a small number of publications that are not belonging to the researcher in the first step. Due to the way Sobiad communicates with publishers, in many publications there are missing keywords, missing metadata such as a list of authors containing only surnames. However, some publications are missing the year and / or reference. Such shortcomings can also negatively impact success rates.

In the proposed model, the number of citations is an important criterion affecting the rating of the publication. On the other hand, the addition of a model to the year in which the reference is taken and / or from several different publications / authors will increase the calculation success in citation.

Sobiad is planned as three different data repositories for science, social and health sciences in its architecture. Therefore, for the proposed model, it is first necessary to determine in which discipline the researcher is studying (science, social, health field). However, for researchers who study in multi-discipline, A model needs to be developed.

The keyword discipline has different forms of representation from different publisher. Some publishers use others commas. For these reasons, it is necessary to develop the keywords with RegEx.

Pairing of binary and triple keywords with other keywords is a problem at many points. For example, there are two keywords that are used as Ottoman Sultans and Ottoman Princes. If the keywords are not fragmented and computed in sentence, the two closest keywords given in the example will not match. If the key words are thought of as words and if the key words that are sent are broken up, perhaps a very important keyword that is used frequently will be missed. Because the key words that are sentences are not include in the dictionary, the type of object such as prepositions, conjunctions cannot be detected.

The number of publications of the researcher is important in the success rate of the model. Having more or less publication by the researchers is one of the factors that negatively affect the success rate. It is evident that when a researcher with more than one field work is considered to be the second and third author of the students who work in different fields, the researcher will be publishing in different topics. However, it is common for researchers who have been studying for many years to work on a different field in the first years of their career. In these cases, some erroneous results can be obtained. In order to prevent this situation, the researcher selects the most recent 40 publications or the publications he has made in the last 5 years.

The method of "comparing the results of the academic publication recommendation with the existing data set" expressed in the experimental results does not give a very high success value for the academicians who have their works in different subjects. Limiting the number of years or publications is done to prevent such situations.

6. Conclusion

This study proposed a novel system that recommends user-specific academic paper instead of the studies similar to a paper like many online academic databases. The system uses many different information such as the publications meta data and paper citation. As the study uses the users' published publication, the co-authors of the users and their works, they can make a more successful recommendation. This paper is the first study that uses Sobiad academic database. Since NoSQL and Big data methods were applied, it can be used in Sobiad in real time. The study is also the first paper recommendation system that handles the academic

publication scoring.

Experiments conducted on Sobiad database that is used by 73 universities in Turkey demonstrate that the proposed method obtained the promising results. In the future, the recommendation system is expected to improve with the some parameters such as the changing of the network over time and attractive academic topics.

Acknowledgment

I would like to thank to Sobiad (Social Science Citation Index-<https://atif.sobiad.com/>) for sharing their data and services.

References

- [1] W. Choochaiwattana, "Usage of tagging for research paper recommendation," in *ICACTE 2010 - 2010 3rd International Conference on Advanced Computer Theory and Engineering, Proceedings*, 2010, vol. 2.
- [2] S. Bahulikar, "Analyzing recommender systems and applying a location based approach using tagging," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017.
- [3] R. K. Mysore, M. Junichiro, and S. Ichiro, "Cross-Domain Academic Paper Recommendation by Semantic Linkage Approach Using Text Analysis and Recurrent Neural Networks," in *Management of Engineering and Technology (PICMET), 2017 Portland International Conference on*, 2017.
- [4] S. S. Weng and H. L. Chang, "Using ontology network analysis for research document recommendation," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1857–1869, 2008.
- [5] K. V. Neethukrishnan and K. P. Swaraj, "Ontology based research paper recommendation using personal ontology similarity method," in *Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on*, 2017.
- [6] Y. C. Lee *et al.*, "Recommendation of research papers in DBpia: A Hybrid approach exploiting content and collaborative data," in *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, 2017, pp. 2966–2971.
- [7] C. Bancu, M. Dagadita, M. Dascalu, C. Dobre, S. Trausan-Matu, and A. M. Florea, "ARSYS-article recommender system," in *Proceedings - 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2012*, 2012, pp. 349–355.
- [8] W. Zhao, R. Wu, and H. Liu, "Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target," *Inf. Process. Manag.*, vol. 52, no. 5, pp. 976–988, 2016.
- [9] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, "A Recommendation System Based on Hierarchical Clustering of an Article-Level Citation Network," *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 113–123, 2016.
- [10] F. Xia, H. Liu, I. Lee, and L. Cao, "Scientific Article Recommendation: Exploiting Common Author Relations and Historical Preferences," *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 101–112, 2016.
- [11] Q. Zhou, X. Chen, and C. Chen, "Authoritative scholarly paper recommendation based on paper communities," in *Proceedings - 17th IEEE International Conference on Computational Science and Engineering, CSE 2014, Jointly with 13th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2014, 13th International Symposium on Pervasive Systems, , 2015*, pp. 1536–1540.
- [12] C. H. Chen, S. D. Mayanglambam, F. Y. Hsu, C. Y. Lu, H. M. Lee, and J. M. Ho, "Novelty paper recommendation using citation authority diffusion," in *Proceedings - 2011 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2011*, 2011, pp. 126–131.
- [13] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.
- [14] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Proc. first Instr. Conf. Mach. Learn.*, pp. 1–4, 2003.
- [15] T. Kenter and M. de Rijke, "Short Text Similarity with Word Embeddings Categories and Subject Descriptors," *Proc. 24th ACM Int. Conf. Inf. Knowl. Manag. (CIKM 2015)*, pp. 1411–1420, 2015.
- [16] S. Albitar, S. Fournier, and B. Espinasse, "An Effective TF/IDF-Based Text-to-Text Semantic Similarity Measure for Text Classification," *Web Inf. Syst. Eng. - Wise 2014, Pt I*, vol. 8786, pp. 105–114, 2014.

- [17] C.-H. HUANG, J. YIN, and F. HOU, "A text similarity measurement combining word semantic information with TF-IDF method," in *Chinese Journal of Computers*, 2011.
- [18] S. H. Strogatz, "Exploring complex networks.," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [19] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [20] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," in *In SIGCOMM*, 1999, pp. 251–262.
- [21] R. Albert and A. L. Barabasi, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [22] D. Krackhardt, "The strength of strong ties: The importance of Philos in organizations," *Networks Organ. Struct. Form. Action*, vol. 216, pp. 216–239, 1992.
- [23] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [24] A. A. Müngen and M. Kaya, "A Novel Method for Event Recommendation in Meetup," in *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2017.
- [25] E. Gündoğan, B. Kaya, and M. Kaya, "Prediction of Symptom-Disease Links in Online Health Forums," in *IEEE/ACM International Conference on Advances in Social Networks and Mining (ASONAM, 2017)*, 2017.
- [26] S. Aslan and M. Kaya, "Link prediction methods in bipartite networks," in *Computer Science and Engineering (UBMK), 2017 International Conference on*, 2017.
- [27] M. Ohta, T. Hachiki, and A. Takasu, "Related paper recommendation to support online-browsing of research papers," in *4th International Conference on the Applications of Digital Information and Web Technologies, ICADIWT 2011*, 2011, pp. 130–136.
- [28] H. Jeon and C. Jeon, "UserProfile-Based Personalized Research Paper Recommendation System," *Comput. Netw. Technol. (ICCNT), 2012 8th Int. Conf.*, pp. 134–138, 2012.
- [29] H. Xue, J. Guo, Y. Lan, and L. Cao, "Personalized paper recommendation in online social scholar system," in *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014, pp. 612–619.
- [30] J. Chen and Z. Ban, "Literature recommendation by researchers' publication analysis," in *Information and Automation (ICIA), 2016 IEEE International Conference on*, 2016.
- [31] H. A. M. Hassan, "Personalized Research Paper Recommendation using Deep Learning," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17*, 2017, pp. 327–330.
- [32] T. Achakulvisut, DE Acuna, T. Ruangrong, K. Kording, "Science Concierge: A fast content-based recommendation system for scientific publications", *PLoS one* 11 (7), e0158423, 2016