A Novel Approach on Conducting Reviewer Recommendations Based on Conflict of Interest

Adi Setyo Nugroho
Department of Informatics, Faculty of Electrical Engineering And Intelligent Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
adisetyon.asn@gmail.com

Aizul Faiz Iswafaza
Department of Informatics, Faculty of Electrical Engineering And Intelligent Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
aizuliswafaza@gmail.com

Ratih Nur Esti Anggraini
Department of Informatics, Faculty of Electrical Engineering And Intelligent Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
ratih_nea@if.its.ac.id

Riyanarto Sarno
Department of Informatics, Faculty of Electrical Engineering And Intelligent Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
riyanarto@if.its.ac.id

***Abstract*— the reviewer's recommendation in accordance with the field of research is crucial where this is directly proportional to the results of the review on the research. In finding suitable reviewers, conflicts of interest (COI) are often found. In this paper, we propose an approach to reviewer recommendations with topic extraction and author extraction to prevent COI. First, we separate the process into 2, namely to do topic extraction using Latent Dirichlet Allocation (LDA), and also to do author extraction using Cosine Similarity. The next step is to combine the two results to rank the 10 recommended authors. The experimental results show that our approach has succeeded in getting 10 recommended reviewers to avoid COI.**

Keywords—Conflict of Interest, Recommendation, Topic Extraction, LDA, Cosine Similarity,

I. INTRODUCTION

Every article that is made, before being published, definitely requires review activities to ensure that the articles are made in accordance with the rules, content and quality [1]. Determining reviewers on an article is not easy. If we choose reviewers randomly, it can result in the results of article reviews being not optimal. So that we need a match between articles and reviewers with various aspects. Several aspects need to be considered to optimize article assignments, such as: areas controlled (relevance) by reviewers and conflict of interest (fairness). Several review service providers on the internet have tried to develop the automation of assigning a reviewer to an article [2]–[4].

Various ways or methods to align reviewers with articles based on topic relevance have been carried out, one of which is the bid method. In this method, the reviewer can select the article he wants to review. However, this method allows reviewers to choose articles that do not match their interests and choose articles with new themes. This can result in a review that is not optimal, because it can result in the processing time needed to be longer and not as detailed as when a reviewer gets an article according to their expertise. so that this method is not applied too much and prefers to do the division automatically according to the expertise of the reviewer [5]–[7].

Conflict of interest (COI) usually appears anywhere, not only in the business realm. All places that have the potential to provide benefits may arise a conflict of interest. On research [8] shows that with detailed reporting related to COI it is possible for the practice to be carried out to be more open and can facilitate other practices in the future. However, in certain cases, most conflicts of interest must be avoided. As in research [4] said that if the reviewer is a co-author in the same research group or the same affiliation then this condition should be avoided. Because with this relationship, the majority of the review results are not carried out professionally. So that in the end a Chairman also makes decisions in selecting reviewers in assigning a paper manually.

To resolve the conflict of interest problem that exists in assigning reviewers to an article, we propose to create an approach in providing reviewer recommendations based on aspects such as the suitability of the article and aspects that must be avoided and indicate a conflict of interest. Some aspects that we use on the suitability of articles such as titles, keywords and abstracts. The aspects used to avoid conflict of interest are the name of the author, organization, venue and publisher.

From these aspects, we use them as dimensions that are combined into one to produce a score for use in making appropriate reviewer recommendations. This research will be divided into two processes for the flow of making recommendations. First, a profile will be created to determine the relationship between authors and reviewers, then in the second process, we will match the results of relationships that have no conflict of interest with other aspects based on paper information. The proposed method is able to provide reviewer recommendations that are free from conflict of interest as one of the goals to achieve fairness and also in accordance with expertise based on information obtained as a goal to achieve relevance.

In this article, it is written in the following order, Section 2 we explain related work in accordance with this article. Then, Section 3 contains an overview of how the research flow and the proposed approach work. In Section 4 we explain how to implement the proposed method. In Section 5 we will discuss the results of the recommendations and evaluation results obtained from the proposed method. Section 6 contains conclusions and future work.

978-1-6654-4059-2/21/$31.00 ©2021 IEEE

II. Related Work

Research conducted by [9] offers a large multi-tiered case study model to engage a wider segment of the scientific community to support decision making on science issues. In doing so it is divided into 2 levels where level 1 has 5 panels, and the second level has 2 panels. And invite reviewers to register according to each panel's choice. With this model, we can find out whether there are reviewers in the research review who don't want to be involved, the diversity of panels can be managed properly.

Research conducted by Al-Zboon [10] which resolves the conflict of interest (COI) problem with link prediction (neo4j graph database), and link open data for reviewer recommendations in the academic field. Both of them are doing the same research data extraction from paper headers, in the case of link prediction they store article data in graph form. Then from the research title, LDA is used to choose from 10 suitable authors to be used as reviewers with Link prediction algorithm. If the LPA value is equal to 0 then the reviewer is suitable. Then further research, conduct COI management with linked open data. This will be done using LDA topic modeling to find reviewers who do not have COI. And can also recommend authors who have COI so that the two results will be reduced and produce authors with appropriate fields without COI.

In the research conducted [1], proposes to create profiles to record each reviewer's experience level information. It aims to reduce the time required by the reviewer. Then the author uses the Ordered Weighted Averaging (OWA) function to summarize the submitted paper information and then ranks the candidate reviewers. The results of the function proposed by the author, from the total of 106 papers submitted and a total of 46 reviewers. The ratio of each number of papers per reviewer is 2.3. In addition, the author also made a comparison when using the random assignment method, the results found that the proposed method had the highest match value between the author and the assigned paper and when using random-assignment, there were 8.6% discrepancies between the paper and the reviewer.

III. Research Methodology

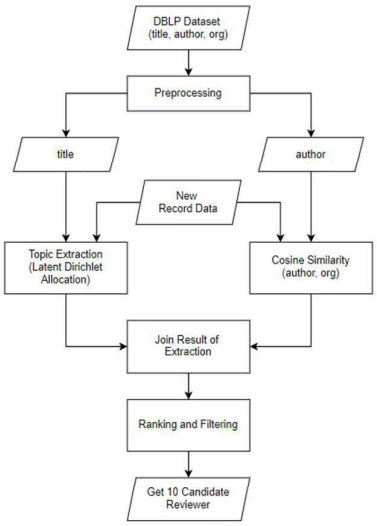

Fig. 1 Research Methodology

This research was conducted to make recommendations for reviewers to conduct reviews by considering the similarity of fields and also to prevent conflicts of interest. The data used in this study is a dataset related to paper citations. An overview of the research methodology can be shown in Figure 1.

The flow of the proposed system in this study will be divided into two, namely topic extraction and author extraction. Topic extraction is done to get the closeness of the meaning of the articles entered, while the author extraction aims to find the authors who are observed from the closeness of the organization, publisher and venue. Then the results of the data extraction from the two processes will be combined and then ranking and filtering will be carried out to obtain a sequence of reviewer recommendations that avoids conflict of interest and has a level of match between the topics entered.

A. DBLP Dataset

The dataset that we use in this study is a public dataset that contains bibliographies of research on journals and proceedings in the field of computer science. This dataset has various versions ranging from version 1 to the most recent version, version 12. We chose to use the dataset version 12 which was formatted in JSON, containing about 4,894,081 publications and 45,564,149 citation counts [11]–[16]. Details of the metadata in this dataset can be seen in Table 1.

In our case, from all record in the dataset, we take 2 thousand record dataset as experiment. Also we take title, authors_name, and author_org that we highlighted in Table 1 to be processed.

TABLE I. DBLP Metadata Description

Field Name	Description
Id	Article ID
title	Article title
Authors_name	Article Authors
Author_org	Authors Organization
Author_id	Author ID
Venue_id	Venue Conference/Publish ID
Venue_raw	Venue Details
Year	Article Publish
Keywords	Article Keyword
Fos_name	Field of study
For_w	Field of study
References	Article References
N_citation	Total Citation
Page_start	Article Page Start Number
Page_end	Article Page End Number
Doc_type	Article Type (Paper/Conference)
Lang	Article Language
Publisher	Publisher Name
Volume	Article Volume
Issue	Article Issue
ISSN	Article ISSN
ISBN	Article ISBN
DOI	Article DOI
PDF	Article format
URL	Link to publication
Indexed_abstract	Abstract in word

B. Preprocessing

At this stage the DBLP dataset was standardize and normalized. The goal is to get the attributes we need and to make it easier to process. The attributes required for this recommendation system are already explained in previous section. In the DBLP dataset, each data row has a structure as shown in Table 2.

TABLE II. SAMPLE DATA DBLP DATASET

Id	Author_Name	Title
1	[{'id': 2108140589, 'name': 'Marcus Kaiser'}, {'id': 2195580174, 'name': 'Stan Matwin', 'org': 'Sch. of Inf. Tech. & Eng., Univ. of Ottawa, Ottawa, ON, Canada'}, {'id': 2203347233, 'name': 'Jae-Nam Lee'}]	Extending adaboost to iteratively vary Design of a Network Protocol Learning Tool its base classifiers

Each row has represent 1 publication or article detail such as author name, organization, abstract, keyword, etc. Because the data used is in JSON format, each attribute has a dynamic value. As shown in Table 2, each row of data in the Author Name attribute will contain different number of object array data, besides that not all data in the author attribute will have an organization key (org). To make it easier to extract author, each data was broken down per author. So that the data will be as shown in Table 3.

TABLE III. SAMPLE DATA AFTER PREPROCESSING

Id	Author_Name	Title
1	Marcus Kaiser	Extending adaboost to … Tool its base classifiers
2	Stan Matwin	Extending adaboost to … Tool its base classifiers
3	Jae-Nam Lee	Extending adaboost to … Tool its base classifiers

As in table 3, each data will be easier to find out the relationship when testing data is entered to extract both topic and author. The next is to preprocessing the title field using common natural language preprocessing. To perform the preprocessing we did several step that shown in Fig 2.

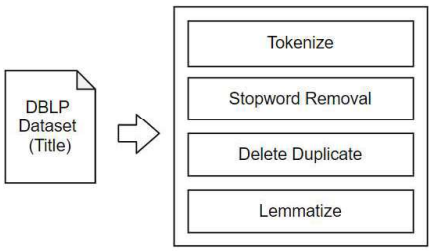

Fig. 2 Preprocessing Title field

As shown in Fig 2, there are 4 step to do preprocessing. First is to tokenize each sentence into a word. Next is to remove every stopwords that didn't use in process. Delete duplicate is important to reduce complexity. Because, from previous step there is so many duplicated data. And the last is lemmatize text to return to basic word.

C. Cosine Similarity

At the author extraction stage, a search for the closeness value between the testing data and training data is carried out. At this stage we use cosine similarity. Cosine similarity is a method to compare the similarity between two vectors. To perform this calculation, several values are needed, such as the result of the scalar multiplication between the query (q_{ij}) and the document (d_{ij}), then the multiplication of the length of the document by the length of the query that has been squared and the result will be the square root. We then form the two values in division so that they become a scalar calculation divided by the calculation of the length of the document and the query. Equation 1 show how to calculate cosine similarity.

$$cs = \frac{\sum(d_{ij} \times q_{ij})}{\sqrt{\sum {d_{ij}}^2 \times \sum {q_{ij}}^2}} \quad (1)$$

Our goal is to determine the closeness of each author so that if the proximity value is greater, it indicates that the person has COI. To calculate the detail proximity value, it is carried out with the procedure as written in the pseudocode in Table 4.

This stage will calculate the cosine of each candidate by comparing whether the author organization in an article is the same or has similarities. Then it will be sought whether the candidate has ever been a co-author, otherwise a cosine similarity calculation will be carried out against the author and candidate organizations. From this step, each candidate will have a closeness value derived from the calculation of cosine which is then used to rank.

TABLE IV. OUR COSINE SIMILARITY PSEUDOCODE

```
x is formatted dataset
y is data new article
for candidate in x:
    if candidate.org is 'unknown'
        set cosine = 0.35
    get any author on author reviews
    if any authorXorg is null
        set cosine = 1
        continue
    for author in y
        if author name in candidate.partner
            set cosine = 1
            continue
        if author org in candidate org
            set cosine = 1
            continue
        if cosine = 0
            find cosine similarity by org
    find cosine similarity by name
set candidate cosine
```

D. Topic Extraction

Topic extraction is carried out to find groups or types of research that exist in the dataset. At this stage we use the Latent Dirichlet Allocation (LDA) method. LDA is a one of topic extraction model which has the ability to find topics that appear in the corpus in a natural language dataset [17], [18].

From Figure 3, α and β are matrix for calculating LDA method per document and per word. Θ is topic distribution

for M. Then M as the entire dataset that collects the title and N as a list of all publication title in the pre-processing of the title. The key of LDA process is w, which mean specific word. The upper box distribute every word into specific topic. And lower box distribute every topic into specific title (N). And the final result, we got every document in document, and also the specific word that represent the topic.

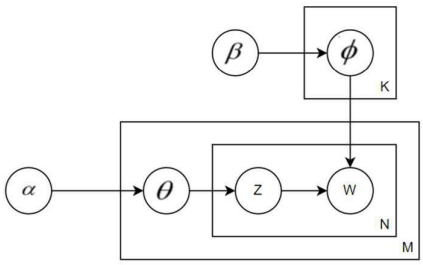

Fig. 3 LDA Process

In our case, topic extraction is not done by document, but by processing the entire research title on the training dataset to retrieve several topics. To increase the accuracy value of topic extraction, this study divides into 5 groups of topics as automatically based on the value generated by LDA. Figure 4 show how our case work.

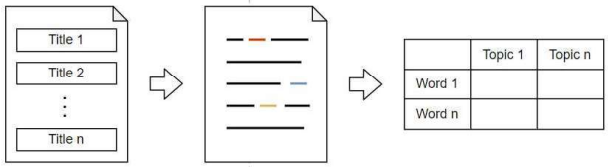

Fig. 4 Topic Extraction basic flow

From Fig 4 from the DBLP dataset we take only the title of article. After that we group all together to be analyzed with LDA. The LDA work with mapping each word to probability of possible topic that we determine before.

E. Joining, Ranking and Filtering

The ranking results generated in the extraction of the reviewer and topic are then formed into two sets A and set B. From these results we will form slices based on the similarity of topics and authors so that the possibilities we will get are shown in Figure 5.

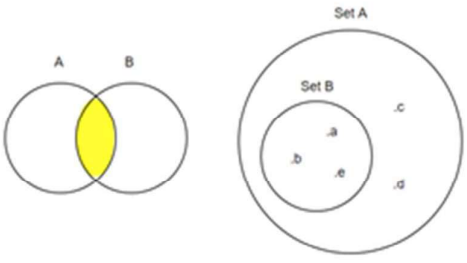

Fig. 5 Join 2 Result

From the two sets of forms, we will get the slices of the two sets derived from the results of topic extraction and cosine similarity. The results of the slice aim to filter the topics that are in accordance with the articles entered to get reviewer recommendations. After filtering, the final result will be ranked based on the highest cosine value and topic suitability. The final result will then be taken as the top 10 candidates as recommendations for the reviewer article that has avoided COI and has the same topic expertise in the test data used.

To avoid the existence of a race condition where the value of the result of the calculation of the cosine of the candidate's name, the name of the candidate's organization and the average have the same value, sorting will be done based on the results of the cosine of the candidate's organization name first, then the results of the cosine of the candidate's name and the average cosine value. By prioritizing the cosine results of the organization's name, it aims to avoid the same name but different organizations. In addition, this study also applies several rules such as:

- If there are data that have the same cosine value, one of them will be taken
- If the data taken is less than 10, then data will be taken from the nearest neighbor topic

IV. Result and Discussion

In conducting this research experiment, the environment used is Google Colaboratory. With this environment, it helps in providing powerful resources that can be used anywhere. To shorten the processing time, this study conducted a sampling of the DBLP dataset so that approximately 2000 data articles were obtained and named DBLP2K. After preprocessing, the amount of data is obtained as shown in Table 5. The increase in data was due to changes in the dataset structure. Which initially, each row represented data per article or publication with several author in it. Then the dataset are transformed into data per author. So, if there is one publication data with several author, there is transform into n data depend on how many author in the publication. Process of splitting each article into per author as shown in Table 2 and Table 3. The purpose of this solution is to make it easier to search for a co-author relationship with an article author in determining reviewers.

TABLE V. DBLP 2K Before and After Preprocessing

DBLP2K Before Preprocessing	**DBLP2K After Processing**
2334 rows XML	6090 rows XML

TABLE VI. An Example Data Test Used

Name	Organization	Title
Marcus Kaiser	-	Extending adaboost to iteratively vary Design of a Network Protocol Learning Tool its base classifiers
Stan Matwin	Sch. of Inf. Tech. & Eng., Univ. of Ottawa, Ottawa, ON, Canada	
Jae-Nam Lee	-	

One of the data used for testing uses new record data that contain all field like DBLP dataset. From that data, author extraction is carried out and values are obtained as Table 6. The cosine value is obtained by comparing each author of an article with the dataset that has been solved. This cosine value will be 0.35 if the organizational value is not found or known. We use this number so that incomplete data will not be selected for the next process. One if it has a co-author attachment to a particular research and results other than

these two values are obtained by determining the proximity of the organization's name using natural language processing.

To make this comparison, this study has several testing data as shown in Table 6. These data will be compared with the training data held to obtain the cosine value of the author's name and author's organization. The title variable will be used to perform calculations using LDA.

As in the Table 7, it can be seen that the candidate in the first row when the name and organization is compared with the test data shown in table 6 has an average cosine of 0.210, the average is obtained based on the results of the cosine of the candidate's name and the result of the cosine of the candidate's organization

TABLE VII. RESULT OF CALCULATE COSINE SIMILARITY

Name	Organization	Cosine by Name	Cosine by Organization	Average Cosine
Philippe Lognonné	Institut de Physique du Globe de Paris (IPGP),	0.18	0.24	0.210
Boualem Haddad	Laboratory of Image Processing and Radiation, ...	0.18	0.29	0.235

From the data in Table 7, several cosine values are obtained, these values have previously been converted so that the smaller the value, the COI or their proximity can be avoided. Then from the data, we also do it in parallel to group topics using LDA. From the LDA calculation, we get 5 topics as shown in Table 8. The data is then used for marking the dataset from the previous preprocessed results.

The five topics will then be used to map each existing dataset. The topics in the dataset are then compared with the topics in the dataset. In this experiment, data that has a different topic group from the topic of the test data will be eliminated, making it easier to determine prospective reviewers.

TABLE VIII. RESULT OF CALCULATE TOPIC EXTRACTION

Topic	Description
1	0.038*"network" + 0.027*"algorithm" + 0.024*"perform" + 0.023*"data" + 0.017*"mobile" + 0.017*"parallel" + 0.016*"sensor" + 0.016*"graph" + 0.015*"detect" + 0.012*"semantic"
2	0.034*"learn" + 0.028*"design" + 0.026*"recognition" + 0.020*"analysis" + 0.018*"speech" + 0.017*"network" + 0.014*"multi" + 0.013*"robust" + 0.013*"model" + 0.013*"problem"
3	0.045*"model" + 0.035*"time" + 0.023*"approach" + 0.019*"system" + 0.018*"real" + 0.016*"network" + 0.016*"manage" + 0.015*"simulate" + 0.013*"data" + 0.012*"interact"
4	0.026*"develop" + 0.026*"system" + 0.021*"object" + 0.020*"knowledge" + 0.020*"multi" + 0.018*"method" + 0.018*"model" + 0.018*"generate" + 0.017*"image" + 0.015*"robot"
5	0.032*"compute" + 0.025*"service" + 0.022*"machine" + 0.021*"image" + 0.021*"support" + 0.020*"dynamic" + 0.018*"study" + 0.018*"application" + 0.016*"software" + 0.016*"system"

After the two extractions are obtained, the next step is to combine the two results into one and then perform ranking and filtering. From the combination, data such as Table 10. This merger is done by adjusting the cosine value of each reviewer candidate with the group topic value. So that later on the new dataset will contain information on how close the candidate reviewers are and the types of research topics that have been created. This data will then be sorted or known as the ranking method to get the largest cosine value with the appropriate topic.

TABLE IX. RESULT OF JOINED EXTRACTION

Name	Organization	Cosine by Name	Cosine by Organization	Cosine	Topic
Philippe Lognonné	Institut de Physique du Globe de Paris (IPGP),	0.18	0.24	0.210	4
Boualem Haddad	Laboratory of Image Processing and Radiation, ...	0.18	0.29	0.235	4

From the sample data, filtering is then carried out that is not in accordance with the topic of the test data used. Then the best 10 scores are taken as shown in table 10. These results show 10 reviewer recommendations that can be used to perform review tasks on the previously inputted testing data. The reason for taking the 10 best reviewer candidates is to summarize the reviewer candidates generated by the system created.

TABLE X. RESULT OF JOINING TOPIC AND COSINE SIMILARITY

No	Field Name	Description	Cosine
1	Cristina Alcalde	Dpt. Matemática Aplicada, Universidad del País...	0.083333
2	Jidi Zhao	Faculty of Computer Science University of New ...	0.088388
3	Pippin Barr	Pippin Barr	0.091287
4	Janusz Marecki	Comput. Sci. Dept., Univ. of Southern Californ...	0.117851
5	Ville Salo	TUCS --- Turku Centre for Computer Science, Fi...	0.117851
6	José P. Zagal	College of Computing, Georgia Inst. of Tech., ...	0.125
7	Ali-Akbar Samadani	Electrical and Computer Engineering Department...	0.125
8	Masayuki Takata	The University of Electro Communications	0.125
9	Vadim Bulitko	Dept. of Computing Science, University of Albe...	0.125
10	Pietro Baroni	Dip. Elettronica per l'Automazione, Univ. of B...	0.133631

V. CONCLUSION

Doing reviewer recommendations that avoid COI and in accordance with the topic being taught is a difficult thing to do. Several studies have carried out various methods to make

recommendations, such as using Neo4j technology to facilitate access to data in the form of objects such as those provided by DBLP. In this study, we propose an approach that is able to solve these problems by carrying out topic extraction and author extraction. the results of the extraction will be used to determine reviewers' recommendations that avoid COI and are in line with the topics that have been worked on by the authors of the article. From the results of the experiments carried out, they succeeded in extracting topics using LDA and grouping topics into 5 major groups or categories. then on author extraction, managed to get the order of reviewer candidates that avoided COI. The results of the two extractions are then combined and given the test data that has been provided. As a result, the system succeeded in providing 10 candidate reviewers who avoided COI and aligned with the topic of the test data provided.

VI. Future Work

From the research we conducted and the approach we propose is not as good as we expected, there are some limitations that we get, such as limited dataset processing resources so that we can only process less than 10% of the data we have. To solve this problem, we can increase the resources we have and break the dataset into small parts and then run in parallel. In addition, from several attribute data, it was found that each row contains dynamic objects. This is because the DBLP dataset documentation does not represent the real form of each received row. On several attributes such as keywords and the author's organization when searched, it turns out that the attribute is in an object attribute. Therefore, if we use the concept of this research, a process of solving and retrieving data contents that are important for experiments is needed. And the last, in the DBLP dataset not all attributes will be available in every data. Some samples that used attributes such as author organization were not found in some data lines and many were also in other attributes. The author's suggestion to overcome the data, maybe by providing a default value when solving data like this research did.

References

[1] J. Nguyen, G. Sánchez-Hernández, N. Agell, X. Rovira, and C. Angulo, "A decision support tool using Order Weighted Averaging for conference review assignment," *Pattern Recognit. Lett.*, vol. 105, pp. 114–120, 2018, doi: 10.1016/j.patrec.2017.09.020.

[2] G. S. Daş and T. Göçken, "A fuzzy approach for the reviewer assignment problem," *Comput. Ind. Eng.*, vol. 72, no. 1, pp. 50–57, 2014, doi: 10.1016/j.cie.2014.02.014.

[3] D. Conry, Y. Koren, and N. Ramakrishnan, "Recommender systems for the conference paper assignment problem," *RecSys'09 - Proc. 3rd ACM Conf. Recomm. Syst.*, pp. 357–360, 2009, doi: 10.1145/1639714.1639787.

[4] L. Charlin and R. S. Zemel, "The Toronto Paper Matching System: An automated paper-reviewer assignment system," *ICML Work. Peer Rev. Publ. Model.*, vol. 28, 2013, [Online]. Available: http://cmt.research.microsoft.com/cmt/.

[5] D. Mimno and A. Mccallum, "Expertise Modeling for Matching Papers with Reviewerts," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 500–509, 2007.

[6] M. Karimzadehgan, C. X. Zhai, and G. Belford, "Multi-aspect expertise matching for review assignment," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 1113–1122, 2008, doi: 10.1145/1458082.1458230.

[7] Y. H. Sun, J. Ma, Z. P. Fan, and J. Wang, "A hybrid knowledge and model approach for reviewer assignment," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 817–824, 2008, doi: 10.1016/j.eswa.2006.10.021.

[8] A. Chan, F. Song, A. Vickers, K. Dickersin, H. M. Krumholz, and D. Ghersi, "Research : increasing value , reducing waste 4 Increasing value and reducing waste : addressing inaccessible research," 2014, doi: 10.1016/S0140-6736(13)62296-5.

[9] C. R. Kirman, T. W. Simon, and S. M. Hays, "Science peer review for the 21st century: Assessing scientific consensus for decision-making while managing conflict of interests, reviewer and process bias," *Regul. Toxicol. Pharmacol.*, vol. 103, no. August 2018, pp. 73–85, 2019, doi: 10.1016/j.yrtph.2019.01.003.

[10] S. A. Al-Zboon, S. K. Tawalbeh, H. Ai-Jarrah, M. Al-Asa'D, M. Hammad, and M. Al-Smadi, "Resolving Conflict of Interests in Recommending Reviewers for Academic Publications Using Link Prediction Techniques," *2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc.*, pp. 1–6, 2019, doi: 10.1109/ICTCS.2019.8923033.

[11] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "ArnetMiner: Extraction and Mining of Academic Social Networks," in *KDD'08*, 2008, pp. 990–998.

[12] J. Tang, L. Yao, D. Zhang, and J. Zhang, "A Combination Approach to Web User Profiling," *ACM TKDD*, vol. 5, no. 1, pp. 1–44, 2010.

[13] J. Tang *et al.*, "Topic Level Expertise Search over Heterogeneous Networks," *Mach. Learn. J.*, vol. 82, no. 2, pp. 211–237, 2011.

[14] J. Tang, A. C. M. Fong, B. Wang, and J. Zhang, "A Unified Probabilistic Framework for Name Disambiguation in Digital Library," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 975–987, 2012.

[15] J. Tang, D. Zhang, and L. Yao, "Social Network Extraction of Academic Researchers," in *ICDM'07*, 2007, pp. 292–301.

[16] A. Sinha *et al.*, "An overview of microsoft academic service (mas) and applications," in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 243–246.

[17] A. R. Baskara, R. Sarno, and A. Solichah, "Discovering Traceability between Business Process and Software Component using Latent Dirichlet Allocation," no. Icic, 2016.

[18] R. A. Priyantina, "Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation , Semantic Similarity and LSTM," vol. 12, no. 4, 2019, doi: 10.22266/ijies2019.0831.14.