Systematic Mapping of Automated Reviewer Recommendation Solutions

Mapeo Sistemático de las Soluciones Automatizadas de Recomendación de Revisores

Guillermo Omar Decoppet[1,2], Patricia Silvana San Martín[2]

[1] *Universidad Abierta Interamericana, Buenos Aires, Argentina*
[2] *Instituto Rosario de Investigación en Ciencias de la Educación (IRICE: CONICET- Universidad Nacional de Rosario), Rosario, Argentina*
decoppet@irice-conicet.gov.ar, sanmartin@irice-conicet.gov.ar

Abstract

The increase in scientific production has generated a recurring problem on a global scale in the recommendation of reviewers for scientific journals and academic events, incentivizing the emergence of a significant diversity of automated solutions. This article presents a systematic review of these reviewer recommendation solutions published in scientific journals and academic events in the period 2018-2023. Methodologically, the final selection focused on the analysis of twenty-five articles. It covered the reviewer recommendation solutions' domain, methods, factors, and the data sets utilized. The results systematize the diverse types of proposed solutions allowing us to observe the similarities between the different methods. It is estimated that the present mapping provides an original survey on this problem that provides well-founded comparative information to support future research on reviewer recommendations.

Keywords: Natural language processing, Peer Review, Recovery models, Selection process.

Resumen

El incremento de la producción científica ha generado una problemática recurrente a escala global en la recomendación de revisores de revistas científicas y eventos académicos, lo cual ha incentivado el surgimiento de una significativa diversidad de soluciones automatizadas. Este artículo presenta una revisión sistemática de dichas soluciones de recomendación de revisores publicadas en revistas científicas y eventos académicos en el período 2018-2023. En lo metodológico la selección final se enfocó en el análisis de veinticinco artículos. El mismo contempló el dominio de las soluciones de recomendación de revisores, sus métodos, factores y los conjuntos de datos utilizados. Los resultados sistematizan los tipos diversos de soluciones propuestas lo cual permitió observar las similitudes entre los diferentes métodos. Se estima que el presente mapeo aporta un relevamiento original sobre esta problemática que brinda información comparativa fundada para sustentar investigaciones futuras de recomendación de revisores.

Palabras claves: Modelos de recuperación, Procesamiento del lenguaje natural, Proceso de selección, Revisión por pares.

1. Introduction

The scalar global growth of the number of scientists and therefore of specialized production complicates the selection of appropriate reviewers for papers presented both in academic events and in scientific journals. Developments in Artificial Intelligence (AI) have contributed significantly to the creation of automatic reviewer recommendations.

Peer review is a central and important process of validation of scientific articles [1] allowing for verification of the contents and methodology in the manuscripts to be published, compliance with quality standards, validity, and clarity in the writing [2].

An initial survey indicates the existence of a highly diverse international corpus of publications focusing on the development of automated solutions for selecting experts to review scientific articles. Many different methodologies, models, strategies, techniques, algorithms, systems, indicators and approaches in finding the best solution are evident. Hence, the purpose of this systematic mapping was to collect and analyze the scientific productions associated with the problem of automatic reviewer recommendation based on the domain of reviewer recommendation solutions, their methods, the datasets used as a source of information. In this study, the interest is given in highlighting the different types of solutions provided by scientific knowledge on the subject. Considering their innovative nature, they are a contribution to science. The diversity of solutions observed in the selected corpus meets the requirements of rigor in the methodology used, an issue validated in its peer review. The successful or unsuccessful application of each is not the focus of the present article because systematic mapping has as its main objective to categorize and synthesize the available evidence in a research area, providing an overview of the scientific landscape in that field. Unlike a systematic review, systematic mapping does not focus on assessing the quality or impact of individual studies, but on mapping what has been investigated. For this purpose, the following questions were formulated, which have guided the present work and the consequent motivation behind them (Table 1).

Table 1 Research questions and motivation.

Research questions	Motivation
Q 1 - What general characteristics define the initial groupings within the set of selected scientific articles?	Extract and start classifying general information from the selected articles that will be further explored in the following questions.
Q 2 - What factors are included in the proposals for automatic assignment of reviewers?	Determine the aspects of the problem that are considered when proposing a solution.
Q 3 - What NLP models and techniques are used in the proposals?	Determine the degree of incorporation of Natural Language Processing (NLP) innovations in the proposals.
Q 4 - What are the methodologies employed for the different proposals?	Establish methods, techniques and alternative procedures that are implemented in each proposal.
Q 5 - What are the different data sources utilized in the different solutions?	Identify sources of information that may be useful for other work that seeks to address this problem.

A brief representative overview of the first decade of the 21st century is presented below. The characteristics of the implemented review method are then addressed. Subsequently, the results of the corpus analyzed within the period 2018-2023 are specified. The conclusions recapitulate some significant aspects that may be of interest for future developments.

2. Background and Related Work

Regarding the problem of automatic reviewer recommendation, an article based on self-distributed documents [3] and an iterative ranking method based on collaborative filtering [4] may be mentioned.

In 2005, an alternative proposal was presented using technology throughout the distribution and review process [5], which consisted of a peer review model that separates distribution from review and proposes the use of OAI (Open Archive Initiative) repositories for distribution and a review mechanism that is included in the OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting) protocol. A co-authorship network was proposed based on bibliographic data from the DBLP (Digital Bibliography & Library Project).

Rodriguez and Bollen [6], proposed a coauthorship network structure with a particle swarm algorithm but in a discrete form related to the random walker algorithms of Markov chain analysis and increasing the number of particles. The authors implemented the concept of negative energy particles to avoid conflict of interest. For the coauthorship network, the DBLP data were also used.

Balog, Azzopardi and Rijke [7] propose finding experts in different organizations with probabilistic language techniques useful for information retrieval. The recommendation was based on content filtering, proposing two models. One of these was to model an expert's knowledge based on the documents associated with it. The other classified documents into types of knowledge on certain topics and proceeded to search for experts according to their associated documents. The second model achieved better results.

Mimno and McCallum [8] tested several methods to define the correspondence of a reviewer to a document to be evaluated. Language models with Dirichlet smoothing were implemented to find experts instead of relevant documents. These models are Author-Topic (AT), and Author-Person-Topic (APT). In AT, the documents in the corpus are required to have a single author and in APT, the documents by each author are fragmented into one or more groups and each group has a separate distribution of topics. The best results were obtained in this last model.

Another proposal [9] mentioned the concept of collaborative intelligence using Wikipedia. It divided the problem into parts, starting with domain modeling. Classifying the proposal and representing it as a multinomial probability distribution based on its keywords. Then, it performed expert matching using WCN (Wikipedia Concept Network),

measuring the semantic relationship between the proposal and expert publications. Finally, it composed and augmented the EKD (Expert Knowledge Database). In this sense, it relied on Wikipedia to model the domains and classify the proposal into related domains.
In summary, these historical articles show the possibility of having different perspectives in finding a solution to the reviewer recommendation problem (RAP). Considering the current developments in AI and more specifically in NLP, it is estimated that its methods and techniques may be useful in the present methodological proposals.

3. Review methods

This systematic mapping of the literature follows the methodology proposed by [3]. A protocol was established under these guidelines, considering three basic stages: 1) selection of the sources of primary academic articles, 2) determination of the search strings for finding academic articles, and 3) application of inclusion and exclusion criteria for scientific articles that serve as the basis for this work.

3.1. Selection of data sources

The main repositories and search engines for scientific articles worldwide were selected, determining that the most prominent are: ACM Digital Library, Google Scholar, IEEE Xplore and ScienceDirect.

3.2. Definition of terms

Using the methodology presented in [3], the search terms were defined. First, systematic mappings were searched and only the article by Aksoy, Yanik and Amasyali [4] was found, in which they selected 103 scientific productions according to the defined criteria. Then, several words that may be present in most of the scientific articles on this topic were collected. From this information, the following terms were selected: paper, reviewer, system, recommender, assignment and problem. Alternative search words were also included, such as peer, algorithm and methodology.
A generic search was then performed in all the selected repositories. Given the diversity of the results obtained, it was decided to perform different search strings adapted to each repository to obtain the best possible result (Table 2).

Table 2 Search strings.

Repositories and search engines	Search strings
Google Scholar	"paper reviewer" AND "assignment problem" AND (system OR algorithm OR methodology).
ACM Digital Library	("paper" OR "peer") AND "reviewer" AND "assignment problem" AND (system OR algorithm OR methodology).
IEEE Xplore	paper AND reviewer AND (assignment OR recommender) AND (problem OR system).
ScienceDirect	"paper reviewer" AND (assignment OR recommender) AND (problem OR system).

3.3 Inclusion and Exclusion Criteria

From the articles obtained in the search process in each repository, it was necessary to make a selection of those closest to the proposed objectives. For this purpose, different inclusion and exclusion criteria were defined to perform this task systematically.
Inclusion criteria:
- Articles in English
- Articles published in scientific journals or academic events with peer review
- Articles published between 2018 and 2023
- Full-text scientific articles that directly or indirectly answer the research questions

Exclusion criteria:
- Studies in languages other than English, duplicates, and not accessible.
- Articles published without peer review
-Articles previous to 2018
- Material that is not published in scientific articles will not be considered
- Not related to the research questions
- Associated only with the workload and/or coverage assignment restriction process
- Articles that do not have an adequate development of their methodology.

4. Search results

In order to select the appropriate articles for the study, the protocol procedure was followed. Using the search strings, 524 papers were obtained and the inclusion and exclusion criteria were applied to these papers in a two-stage iterative procedure: 1st and 2nd filter. In the first stage, the titles and abstracts were examined, resulting in 74 selected articles. And when the 2nd filter was applied by reading the entire content of the article, 25 articles were selected.

5. Synthesis of extracted data

The results are presented below following the research questions already mentioned.

5.1. What general characteristics define the initial groupings within the set of selected scientific articles?

Based on the 25 articles for the period 2018-2023, Fig. 1 shows the number of papers registered per year.

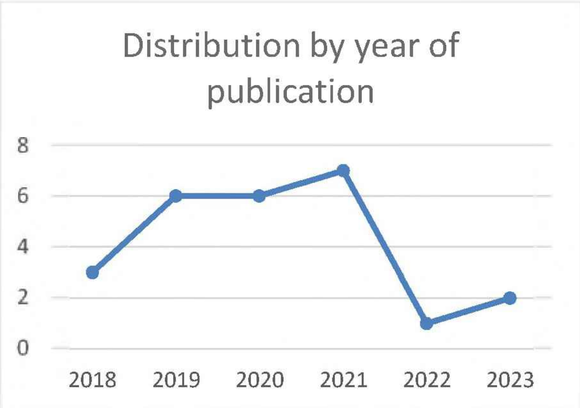

Fig. 1 Papers registered (2018-2023).

Distribution according to where they were published: 11 articles in scientific journals and 14 in academic events.

The reviewer recommendation problem with automatic methods may be initially divided between the extraction stage with representation and the assignment stage. In the extraction stage with representation, the characteristics that represent the documents to be reviewed and the possible reviewers are considered. The assignment stage refers to the reviewer selection process. Although the 25 articles address the problem, not all of them do the assignment in the same way. There are several forms of addressing extraction with representation. The assignment may be classified into two different types: 1) unrestricted and 2) with restriction.

In unrestricted assignment, the selection process finds reviewers for a specific scientific article. The identical procedure is executed for the following articles sequentially.

In assignment with restriction, there is a group of reviewers who have to be assigned to a certain number of scientific articles. This process, in contrast to the previous case, is a multiple process of several assignments at the same time and is common in scientific events. In general, reviewer workload and coverage are considered as constraints. Workload refers to number of articles to be assigned simultaneously to each reviewer. Coverage, in general, refers to the number of reviewers that should be assigned to each article.

All selected articles address extraction and assignment Some develop methodologies with unrestricted assignment [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28] and others with restrictions [29, 30, 31, 32, 33, 34, 35, 36].

There are 17 articles of type 1 and 8 of type 2. The analysis presented in this paper focuses on the processes of extraction and unrestricted assignment. The extraction processes with restricted assignments do not correspond to the objective of this research.

5.2. What factors are included in proposals for automatic assignment of reviewers?

All proposed methodologies consider various types of information to make the recommendation. This categorized information will be designated as "factors." The treatment of each factor varies depending on its type. The intervening factors, according to the survey conducted on the selected corpus, are the following: experience, conflict of interest, authority, diversity, researcher interest, seniority and reviews.

In the different proposals, diversity is observed in the forms of naming and there are some factors integrated into the calculation of others. Despite this, it is possible to recognize each of them in the different articles, although sometimes they are treated with another name.

The seven factors are conceptualized below:

Experience: refers mainly to the texts of the scientific productions of the candidate reviewers, collected from different sources and in different forms by each methodology. It is also employed as research topics expressly declared by the researchers as in [24]. In the text representation of scientific productions, some articles expressly clarify which parts of the text are significant in representing the experience. For example, considering the whole text as in [15]; keywords in [23]; abstracts in [26] and [17]; titles and abstracts in [31] [35] [12] [12] [16] [13] [28] and in the proofs in [36]; titles, abstracts and keywords in [33] and [21]; titles, abstracts and introduction in [20] and [25]; titles, abstracts, keywords, introductions and conclusions as in [34]. Only titles in [27] or dividing all scientific articles into five parts: title, keywords, abstracts, references and the rest of the text in [30].

Conflict of interest: represents situations in which reviewers may have interests of various kinds that may cause a loss of evaluative objectivity. The case of being the author of the article to be reviewed, co-author in another scientific article of the authors, among others. In some cases, it is checked if the authors have the same affiliation, as in [20] referring to the same institute, university and the PhD director of the author of the manuscript. In [21] affiliation is also controlled. In [16] the methodology verifies if the authors of the manuscript are not coauthors of the reviewers, in [32] add verification that they are not coauthors of coauthors. In [35] the researchers add the concept of conflicts of competence, which is the case in which the reviewer and the article to be reviewed share the same research area in a scientific event and could cause loss of objectivity. In [27] they add family relations.

Authority: refers to the academic prestige of the researcher. It can be understood as the academic impact of the researcher's production reflected in metric indexes such as number of citations or other bibliometric impact factors, which can be considered as an objective indicator of the academic community. In [34] [13] and [23] the authors use number of citations and h-index, in [16] the study uses the average h-index of the reviewers calculated on the relevant papers in relation to the manuscript and the average number of citations in these scientific articles. In [25] h-index is applied. In [21] the researchers use for authority, global authority unrelated to the subject of the manuscript and local authority on the subject of the manuscript to be reviewed, both are expressed by the number of references to the reviewer's papers. In [29] the quality score is utilized, which integrates the number of supervised PhDs, books and book chapters written, and articles published in journals and conferences in addition to the h-index. In [15] they employ the degree or level of study of the reviewer candidate.

Diversity: this factor is applied to the articles in different forms but refers to a fair, integral and equitable evaluation, trying to incorporate reviewers based on what constitutes diversity for each of the authors of the analyzed papers. In [16] the study defines diversity as the measure that ensures that the reviewers' experience is distributed to areas that are as different as possible. In [31] it includes background, location and seniority (which will be defined later). In the background, they consider that at least one reviewer must work in academia and one in industry. In location, they recommend including reviewers from different geographical locations. In [23] diversity is considered in the geographical location of their affiliations, referring to countries. In [13] random walk with restart (RWR) is applied to select reviewers, offering the potential to obtain a diverse group of reviewers.

Researcher interest: denominated in some articles as "freshness" or "recency", being associated with the selection of scientific articles within a certain recent time range, assigning more weight to articles that are closer in the time range. It is based on the fact that the researcher's topic of study may change over time, so the closest publications refer to the current research interests. The [14] and [15] articles utilize research interest but denominate it as experience. In [29] authors refer to the topic interest vector to represent the area of expertise of the reviewer but use the researcher impact factor as a recency score referring to the articles published in the last years of the reviewer and define it as the weighted average of articles published by the reviewer in a range covering recent years. In [13] the researchers consider the scientific articles closest in time with greater weight over the others, to make the network of reviewers, as in [16]. In [34] the study considers that the reviewer candidate must be active in research areas of the manuscript and a calculation is made considering the difference between the current year and the year of publication in the formula. In [25] they consider the recency of the reviewers' publications.

Seniority: refers to the academic category achieved by the researcher. In [31] this term is used within diversity as explained above and it is required that each set of reviewers contains at least one senior researcher. In [16] the methodology expresses as desirable that the review of each manuscript be performed by at least one senior and one junior researcher.

Reviews: consider the history of previous reviews of candidates. Only one of the articles [21] considers this by performing a calculation with three indicators.

In summary, the use of the factors in the corpus of selected articles is observed in Fig. 2. The most utilized is Experience and in decreasing order Conflict of interest, Authority, Diversity, Researcher interest, Seniority and finally Reviews.

There are 11 articles that only consider Experience. In the case of articles that use 2 factors, 5 combine Experience and Conflict of Interest, and one that includes Experience and Authority. For those that consider three or more, there is a diversity in the factors to include. Experience is the only factor that is present in all articles.

5.3. What NLP models and techniques are used in the proposals?

The objective of this research question is to determine if NLP techniques are implemented in the reviewer recommendation process, identifying which techniques are implemented in the different methodologies.

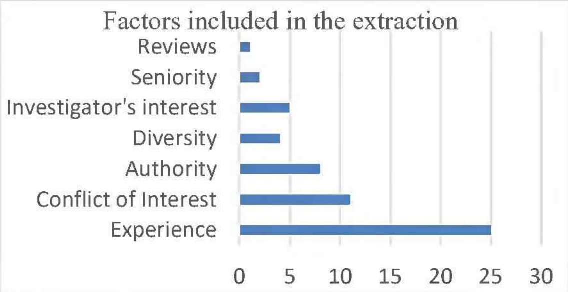

Fig. 2 Factors included in the extraction

The most commonly utilized technique in the articles is Latent Dirichlet Allocation (LDA) [37]. The model was implemented to process the experience factor in [29] [32] [32] [33] [33] [34] [18] [20] [25] [27]. Then follows the Term Frequency-Inverse Document Frequency (TF-IDF) technique being applied in [30] [31] [21]. In [17] TF-IDF or BM-25 is utilized [38]. In [26] they use BM-25. In [28] the authors use TF-IDF and a neural network algorithm such as Bidirectional encoder representations from Transformers (BERT) [39] or Convolutional Neural Network (CNN) [40] or Bidirectional Long Short Term Memory (BiLSTM) [41].

In [13] the study uses LDA and a variation applied to the TF-IDF reviewer recommendation that they denominate Term Frequency-Inverse Reviewer Frequency (TF-IRF). In [22] the methodology uses Author-Topic Modeling (ATM) [42]. In [14] the researchers apply Author-Subject-Topic Modeling (AST) [43]. In [15] they utilized Word2Vec [44] to extract the main semantic keywords. In [16] the authors employed in the selected scientific articles: TF-IDF, Doc2Vec [45] or BERT, and for the calculations they use TF-IDF and LDA. In [12] the article proposes a bidirectional two-level closed recurrent unit (GRU) [46] neural network with an attention mechanism. In [36] they utilize Probabilistic Latent Semantic Analysis (PLSA) [47]. In [35] the study applies the Prompt tuning technique [48]. In [19] a topic model is used but the technique is not specified. Two articles do not apply PLN techniques. In [23] the keywords of the reviewers' scientific articles and manuscripts are utilized. In [24] the keywords of the proposals and the research interests of the reviewers are based on the classifications available in a taxonomy. The following Table 3 details the grouping of techniques and models implemented.

5.4. What are the methodologies employed for the different proposals?

This section includes an analysis of the different methodologies used. Grouping by factors used: First, those utilizing 3 or more, followed by those utilizing 1 or 2 factors.

Table 3 Grouping of NLP techniques and models.

Grouping	Qty.	NLP technique and/or models	Articles
Topic model	12	LDA, ATM, AST, PLSA	[29] [32] [33] [34] [36] [14] [18] [19] [20] [22] [25] [27]
Vector space models	5	TF-IDF, BM-25	[30] [31] [17] [21] [26]
Combination of different NLP models and techniques	1	LDA y TF-IRF	[13]
	1	(TF-IDF o Doc2Vec o BERT) y (TF-IDF y LDA)	[16]
	1	TF-IDF y (BERT o CNN o BiLSTM)	[28]
Models based on Word embeddings	1	Word2vec	[15]
Models with neural networks	1	BiGRU	[12]
LLM-based models	1	Prompt tuning	[35]
There is no model	2		[23] [24]

The first group includes eight articles whose methodology for calculating prioritization for assignment is based on the integration of multiple factors in a formula. When three or more factors are used, this type of solution is applied (Table 4).

In [23] keywords are extracted from the manuscript and the reviewers' scientific articles. Concepts such as article prestige (Pr), author impact (AI) and keyword impact (KI) are considered. Finally, the impact of a transaction (I) involves Pr, AI and KI. Using these data, two types of objective patterns are formulated. The first pattern is the researcher general topic pattern (RGP) includes only researchers and the other is the researcher-specific topic pattern (RSP) which is composed of combinations of researchers and a specific topic.

Table 4 Factors included in scientific articles (3 or +).

Art.	Experience	Conflict of interests	Authority	Diversity	Researcher Interest	Seniority	Revisions	Number of factors
[23]	Yes	No	Yes	Yes	No	No	No	3
[25]	Yes	No	Yes	No	Yes	No	No	3
[29]	Yes	Yes	Yes	No	Yes	No	No	4
[31]	Yes	Yes	No	Yes	No	Yes	No	4
[34]	Yes	Yes	Yes	No	Yes	No	No	4
[21]	Yes	Yes	Yes	No	No	No	Yes	4
[13]	Yes	Yes	Yes	Yes	Yes	No	No	5
[16]	Yes	Yes	Yes	Yes	Yes	Yes	No	6

In [25] they use the title, abstract, and introduction and process them with LDA to extract the most important topic of each document and find experts on each topic. They sort them in descending order based on the number of articles in that domain. They incorporate the researcher's interest factor associated with the recency of their publications and the authority based on the h-index, to find the most relevant expert in the domain of the manuscript.

In [29] the study creates a list of topics with LDA using the articles presented at the conference, and all the possible research topics of each reviewer are extracted from different sources and integrated into a common dictionary of terms, using an automatic assignment but with an expert evaluation. The researchers propose to calculate the coincidence between the topics covered by the manuscript and the experience of the reviewer, which is named "interest in the research topic" and add the authority factor named quality, the interest factor denominated weighted average impact factor and availability that is associated with the constrained assignment. The methodology verifies the conflict of interest.

In [31] the purpose of DiveRS is to extend the program committee of an academic event with new reviewers not included, by an automatic process. The assignment considers two interconnected tasks, the first identifies submissions that would not receive suitable reviewers using the current program committee and generates as a second task, the suggestion of new reviewers to expand it. To represent manuscripts and reviewer articles, it uses TF-IDF and applies cosine similarity. Additionally, it considers diversity considering background, location, and seniority. This proposal also verifies a conflict of interest.

In [34] the authors utilize the manuscripts and articles of the reviewers and process them with LDA to find all the topic domains to assign a set of reviewers to cover all topics in the manuscript. A similarity calculation is performed between the manuscript and candidate reviewers to find the most appropriate reviewers. The final assignment is made through a weighted sum of experience, authority, and interest of the researcher. For authority, they integrate several publications, h-index, and number of citations. For the researcher's interest, the study uses as a parameter a certain period of years from the current year and associated with the topic. In addition, they control the conflict of interest.

In [21] the authors use TF-IDF and cosine similarity calculation. For the assignment, they make a multi-criteria evaluation expressed by an indicator applying a weighted sum of different parameters related to their maximum values. For the authority, they use the Global Authority (GA) indicator considering the number of references of the reviewer over all references of the reviewers and the Local Authority (AL) indicator based on the number of references to the reviewers' articles on the topic of the manuscript. The review factor is calculated through the definition of quality of work in the role of reviewer with three indicators based on the number of years as a reviewer, days of review over total reviews and number of manuscripts where the editorial team was in agreement with the review of the total requested. Each indicator has a weight and after performing the calculation, reviewers are assigned. Conflict of interest is also verified.

In [13] the researchers propose a model called TCRRec with a multi-layer network that integrates three layers. Topic network layer: using publications from researchers and extracting topics with LDA, assigning greater weight to newer articles and grouping them by reviewer. They also capture the lexicographic content of each reviewer's publications by processing them with TF-IRF (variation of TF-IDF). The above two formulas are combined by summing their cosine similarities separately with weighted normalization with complementary weights. The relationship between the network of topics and citations is made by the topics of interest considering those that have the greatest weight in the manuscript. Citation network layer: considering that articles are related by similar citations and not by topics. The citation network is associated with the reviewer network because the set of associated reviewers are the authors of the candidate articles. Reviewer network layer: the reviewer candidate is identified using features such as h-index, citation, co-citation score, and some other features such as co-author similarity. For the assignment, they use a multilayer approach integrating the three networks.

The scoring mechanism eliminates less relevant articles and reduces computational complexity. Conflict of interest is verified. In this layer, random walk with custom reset is used to reduce the calculation, avoiding calculating the similarity of the document and all reviewers.

In [16] the researchers propose two stages: The first stage is divided into three parts: 1) performs the search for experts, using TF-IDF, BERT or Word2vec; 2) calculates the cosine similarity between the manuscript and the reviewers' papers and they are order them descending by score. Here, a first conflict of interest filter is considered by co-authors of authors and 3) to obtain the final list of reviewers a calculation is made using voting techniques (13 techniques were considered), with a threshold and using total score techniques. In the second stage, LDA and TF-IDF are implemented to represent the papers, join them and performs a cosine similarity calculation. Then, select the most similar reviewers and calculate experience, authority, diversity, research interest and seniority.
Table 5 shows that the remaining methods have only the Experience factor and others add the conflict of interest factor. There is no single grouping in these cases and there are few methodological coincidences that allow groupings, so they are treated as separate techniques.

Table 5 Reduction of factors included in scientific articles.

Art.	Experience	Conflict of interests	Authority	Diversity	Researcher Interest	Seniority	Revisions	Number of factors
[12]	Yes	No	No	No	No	No	No	1
[14]	Yes	No	No	No	No	No	No	1
[17]	Yes	No	No	No	No	No	No	1
[33]	Yes	No	No	No	No	No	No	1
[28]	Yes	No	No	No	No	No	No	1
[18]	Yes	No	No	No	No	No	No	1
[19]	Yes	No	No	No	No	No	No	1
[36]	Yes	No	No	No	No	No	No	1
[22]	Yes	No	No	No	No	No	No	1
[24]	Yes	No	No	No	No	No	No	1
[26]	Yes	No	No	No	No	No	No	1
[30]	Yes	Yes	No	No	No	No	No	2
[32]	Yes	Yes	No	No	No	No	No	2
[20]	Yes	Yes	No	No	No	No	No	2
[35]	Yes	Yes	No	No	No	No	No	2
[27]	Yes	Yes	No	No	No	No	No	2

The second group include methodologies that include 1 or 2 factors. The calculation of experience is always present and in cases where there is a control of conflict of interest it is considered as a restriction.

In [12] the authors use a two-layer bidirectional GRU network with a two-level attention mechanism. One works on the tokens to encode sentences, while the other works on the encoded sentences and encodes the document by taking the titles and abstracts to process them. The recommendation is the selection of the reviewer who has the most labels in the research field of the article to be reviewed, using a simple multi-label-based reviewer assignment (MLBRA).

In [14] the proposal employs the AST model, an extension of the TA, introducing a subject layer whose task is to supervise the generation of hierarchical topics and to allow sharing among authors. Two versions AST1 and AST2 are proposed. The first assigns to each document a topic distribution in the subject layer (soft clustering) and the second assigns to each document a subject label according to the category distribution (hard clustering). In the tests, AST1 performed better than AST2.

In [17] the researchers implemented the TF-IDF vector space model techniques, with variations in document term weighting and BM-25 to represent the documents from which the abstract was extracted. Cosine similarity was implemented to match the manuscript and reviewers. The best results were achieved with BM-25.

In [33] the study uses LDA to represent the manuscripts and scientific articles of the reviewers. The parts of the text that are used are the title, summary and keywords. The reviewer assignment is based on thematic similarity to the manuscript.

In [28] the authors use a predictive classification model with positive and negative cases generated on titles and abstracts of different articles. They propose a form of reviewer assignment based on sentence pair models (SPM-RA), modeling them by information supervision. First, the training set is constructed by the relationship between title and abstract. The methodology constructs a dataset using a TF-IDF sample and employs different neural network architectures, such as BERT, CNN, or BiLSTM, to build a sentence pair model to train the relationship between the title and abstract of the article (supervised model), and then to predict the similarity between reviewers and manuscripts.

In [18] the methodology consists of two components. In the first, the manuscripts and scientific articles of the reviewers are obtained and processed with LDA. To compute similarity, the Jensen-Shanon distance between reviewers and manuscripts is subtracted from unity. The other component calculates the similarity of references by constructing sets and comparing each reviewer against the article's references. The Jaccard index measure is employed to calibrate the similarity and diversity of the sample set. Then, the two previous formulas are unified, adding the similarity values of experience and references through weighted normalization with complementary weights. Conflict of interest is controlled.

In [19] the authors construct a fuzzy graph of the most important words extracted from the reviewers' papers and manuscripts, and apply centrality measures on this graph, creating fuzzy sets for the selected keywords and their weights. With Wordnet, the distance between the fuzzy sets represented by the authors' papers and the candidate reviewers' papers is calculated. Finally, the fuzzy extension principle is applied on the fuzzy sets to select the top three experts to review the proposal.

In [36] the publication employs a method that groups reviewers' publications into latent research areas. Addressing two subproblems: 1) identifying latent research areas in reviewer's publications and, 2) improving the match between reviewers and articles. The totality of reviewers' publications is clustered using k-means with cosine distance. Latent semantic indexing is also used to reduce dimensionality and improve clustering quality. Each reviewer is associated with a number of latent research areas to which their publications belong. PLSA is employed to extract topics. The objective is to assign to each article a group of reviewers that cover as many aspects of the article as possible.

In [22] the research considers the texts, selecting from the abstract to the conclusions, represented with the author-topic model which is an extension of LDA. This model adds a layer with authors over the distribution of topics. To select potential reviewers, authors use the Hellinger distance calculation based on the similarity of their topic probability vectors.

In [24] the researchers organize the proposal and reviewers into categorized tracks for evaluation according to a defined hierarchical research domain tree. Reviewers upload this information when they create their research interest profiles in the registry. The methodology calculates similarity using Jaccard's similarity coefficient between the reviewer's keywords and those of the proposal.

In [26] the methodology is divided into three parts. The first consists of searching and extracting reviewers' previous publications on the Internet, and obtaining their titles and abstracts. The second refers to the calculation of similarity factors through PLN techniques. Only vector space models were tested and BM-25 obtained the best results on abstracts. The third evaluates the accuracy of the similarity factors. It consists of the conversion of similarity indicators to levels, considering automatically determined levels of experience (high, medium, or low). In addition, it includes a Pearson correlation analysis between the automatically determined levels and those expressed by the reviewers.

In [30] the study uses TF-IDF to represent the text divided into the title, keywords, abstracts, references and the rest of the text. To determine the similarity, it calculates cosine similarity of each part, assigns a weight to it and concludes by calculating the total weighted similarity measure. Conflict of interest detection is also performed.

In [32] the reviewer experience score is calculated in several steps. With LDA the topics of each reviewer are extracted. The initial score is obtained by applying a vector space model and comparing this reviewer file with the manuscript. The score is obtained by scalar product between the reviewer's publications and the manuscript, after normalizing and smoothing the data. The experience declared by the reviewer is considered as ground truth and is combined with the initial score and a supervised prediction algorithm is utilized to obtain the final score. Co-authorship distance is calculated to detect conflict of interest. The final assignment maximizes topic similarity and minimizes conflict of interest.

In [20] the process has three steps: 1) Gibbs sampling with LDA to represent manuscript and reviewers' scientific articles, 2) topic analysis to find the relationship between parts of the article as supervisory content to interpret the research area of the article. The relationship of the topic, defined as research area, between the title and abstract is employed to train the model. This model is applied to calculate the similarity between a manuscript and the reviewer's publications and 3) relevance measure where the similarity index between each reviewer's publications and the article is calculated using the weights of the main topic, sorting them in descending order to obtain the five most relevant reviewers. Conflict of interest is also verified.

In [35] using prompt tuning, the authors obtain the research domains for each article with the title and abstract. With the manuscripts and the reviewers' scientific articles, they calculate the percentage of coverage that corresponds to the percentage of research domains that the reviewers share. They also use a score from "The Toronto Paper Matching System" (TPMS). This score indicates the reviewers' experience and willingness to review. The authors combine the two previous formulas into a single formula, adding the coverage score and the trend score provided by TPMS with a weighted normalization with complementary weights.

In [27] the researchers use LDA to extract topics from articles, and then use a graph database to store the content. The data are stored as node graphs with the relationships between them. The 10 most relevant active authors on that topic are searched. The prediction algorithm is applied to find candidate reviewers with conflict of interest, remove them from the list and define the final list of recommended reviewers.

Table 6 presents a relationship between the NLP techniques and models used and abstraction in a general model of the process covering extraction, representation and assignment. Each of the models has a complexity and any abstraction is a generalization that leaves out important details of calculations, formulas and algorithms of the recommendation process. The generalization presented should be understood only as an orientation of the procedures followed in the general steps to obtain the solution.

The article [15] is not included in the previous groups because it is considered that the methodology may have different rules but in the example of the paper it includes only experience and authority. The methodology applies a text extraction and segmentation tool to select keywords from full-text records, based on the most important core words that were extracted using the Word2Vec algorithm. These rule sources are combined with different defined requirements to generate conclusions using rule-based reasoning. The procedure configures the rules engine that analyzes the semantic connections of keywords and makes inferences from the knowledge base and input data. The ontology of the manuscript and the reviewers is specified, which consists of a set of conceptual definitions, properties and relationships, then the algorithm and rules are applied to consider an article and a reviewer as input, execute the rule and generate the result as true or false.

Table 6 Relationship between NLP techniques and models, and general model.

PLN Technique	Qty.	Art.	General technique methodology
TF-IDF	3	[21] [31]	calculation of different factors
		[30]	vector space model
TF-IDF and neural network (BERT or other CNN, BiLSTM)	1	[28]	supervised sentence pair model (title and abstract)
AST model	1	[14]	layered topic model
AT model	1	[22]	Topic model
LDA	8	[29] [34] [25]	calculation of different factors
		[18]	combination of topic model and citation similarity
		[33]	topic model
		[20]	topic model and supervised model
		[27]	topic model, graph database and prediction model
		[32]	supervised model combining reported and calculated experience
LDA y TF-IRF	1	[13]	calculation of different factors
Not specified	1	[19]	fuzzy graph model on topic model and use of Wordnet
there is no technique	2	[23]	calculation of different factors
		[24]	similarity between declared keywords of reviewers and papers
PLSA	1	[36]	clustering with topic model
prompt tuning	1	[35]	large language model and TPMPS
neural network BiGRU	1	[12]	hierarchical label model (with neural networks)
TF-IDF or Doc2Vec or BERT selection and calculations: TF-IDF and LDA	1	[16]	calculation of different factors
TF-IDF or BM-25	1	[17]	vector space model
BM-25	1	[26]	model combining VSM model and reviewers' declaration
Word2Vec	1	[15]	rule-based model (similarity and constraints)

5.5. What are the different data sources utilized for each of these factors included in the different solutions?

Based on the following Table 7 and considering the sources of information utilized more than once, it is possible to observe that the most frequently employed source of information was DBLP. For articles and profiles, Aminer follows with four, then Google Scholar with three and Semantic Scholar and ResearchGate with two. As data sources for evaluations, NIPS is the most employed followed by SIGIR, then followed by other academic events and repositories. It is estimated that analysis could be useful to determine the sources of information for future methodological proposals.

Table 7 Grouping of the different sources of information from 2 to n.

	Data sources	**for evaluation**	**article and profile information**	**Qty.**
1	Digital Bibliography & Library Project (DBLP)	[34] [25]	[29] [31] [16] [17] [18] [20] [25] [26] [27]	10
2	Neural Information Processing Systems (NIPS)	[13] [18] [34] [20]		4
3	Aminer		[29] [13] [16] [32]	4
3	Special Interest Group on Information Retrieval (SIGIR)	[33] [35] [36]	[33] [35] [36]	3
4	Google Scholar		[30] [34] [23]	3
4	ResearchGate	[34]	[29] [20]	3
5	Arvix	[28] [18]	[28]	2
5	CiteSeer	[34]	[20]	2
	CompSysTech database	[17] [26]		2
5	Conference on Artificial Intelligence of AAAI	[34] [20]		2
5	Interspeech	[34] [20]		2
5	Semantic Scholar		[17] [26]	2

6. Conclusions

Based on an in-depth analysis of the state of knowledge, this article has addressed an updated systematic mapping of automated reviewer recommendation solutions for the period 2018-2023. It is possible to synthesize different aspects of the problem that are relevant both in terms of comparing methodologies and in terms of providing well-founded information for future methodological developments.

In principle, seven factors were recognized, with "Experience" being the factor present in all the articles, followed by "Conflict of interest" and "Authority". One aspect to consider is that only one proposal addresses the factor "Reviews" which, in the opinion of the authors of this work, is one of the keys necessary to evaluate the quality of the researcher's review trajectory. The factor "Seniority" seems to overlap with the factor "Authority" and in both cases, it is directly or indirectly associated with "Diversity". This factor has a wide range of interpretations on the part of the authors, which makes it notoriously difficult to establish a common conceptualization.

Use of PLN techniques and models: 12 articles include topic models and 5 use vector space models were utilized. Only in one case was a combination of topic and vector space models employed. More than 70% of the articles use these models. These models do not incorporate semantic information of the words separately, or contextualized in the text. In reference to the other techniques treated, Recurrent Neural Networks, Word embeddings and Transformers are observed. Only one article used Large Language Models, with the prompt tuning technique.

The different methodologies include from 1 to 6 factors. There is a direct relationship between the use of several factors and the methodology. Each factor is first treated independently with its own calculations and then, if it includes three factors or more, it is incorporated into an integral formula with the other factors. There is a high degree of differentiation in the integrations proposed in each article. The conflict of interest factor is considered as a restriction.

There is a wide dispersion of information sources, with 42 different, the most used being DBLP in 10 articles. This shows that there is no centralized database containing all the information worldwide. Thus, according to the needs for which the different methodologies were constructed, different sources of information were selected based on their usefulness about the purposes.

It is possible to affirm that a significant percentage of proposals develop different methods for solving the reviewer recommendation problem. There is a very low use of the most advanced PLN techniques. There is also a direct impact between the number of factors included and the resolution methodology developed, although there is diversity in the development of each solution. The most complex and advanced methodologies are those that include multiple factors. In this sense, it is considered that this would be the

most appropriate orientation for the resolution of this diverse and complex problem because it considers each of the variables that influence a final recommendation.
It is important to note that although it is important to consider that the recommendation of automatic processes such as those described above is valuable information that may be accessed by the editorial team, it is their responsibility to make the final decision.

Funding

This work was partially supported by PUE 011/20 research grants from CONICET (Argentina)

Competing Interests

The authors have declared that no competing interests exist.

Authors' Contribution

The authors confirm contribution to the paper as follows: GOD and PSSM Conceptualization, Writing-Original draft preparation, Writing – review & editing; GOD Data curation, Formal Analysis, Methodology; PSSM Supervision.
All authors reviewed the results and approved the final version of the manuscript.

References

[1] G. S. Martínez, «La revisión por pares y la selección de artículos para publicación», *Rev. Colomb. Psicol.*, vol. 21, n.º 1, pp. 27-35, 2012.

[2] M. F. Abad, «El proceso de revisión por pares», en *Revistas científicas: situación actual y retos de futuro*, E. Abadal, Ed., Universitat de Barcelona, 2017, pp. 53-71. http://eprints.rclis.org/32131/

[3] T. Riggs & R. Wilensky, «An algorithm for automated rating of reviewers | Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries», JCDL '01: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, 2001, p. Pages 381-387. doi: doi.org/10.1145/379437.379731.

[4] P. Rigaux, «An iterative rating method: application to web-based conference management|Request PDF», ResearchGate. https://www.researchgate.net/publication/47800513_An_iterative_rating_method_application_to_web-based_conference_management

[5] M. Rodriguez, J. Bollen, & H. Sompel, «The Convergence of Digital-Libraries and the Peer-Review Process», *J. Inf. Sci.*, vol. 32, may 2005, https://doi.org/10.1177/0165551506062327

[6] M. A. Rodriguez & J. Bollen, «An Algorithm to Determine Peer-Reviewers», *Proceeding 17th ACM Conf. Inf. Knowl. Min. - CIKM 08*, p. 319, 2008, doi: 10.1145/1458082.1458127.

[7] K. Balog, L. Azzopardi, & M. de Rijke, «Formal models for expert finding in enterprise corpora», en *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, en SIGIR '06. Seattle, Washington, USA: Association for Computing Machinery, aug. 2006, pp. 43-50. doi: 10.1145/1148170.1148181.

[8] D. Mimno & A. McCallum, «Expertise modeling for matching papers with reviewers», en *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, en KDD '07. San Jose, California, USA: Association for Computing Machinery, aug. 2007, pp. 500-509. doi: 10.1145/1281192.1281247.

[9] K.-H. Yang, T.-L. Kuo, H.-M. Lee, & J.-M. Ho, «A Reviewer Recommendation System Based on Collaborative Intelligence», en *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, sep. 2009, pp. 564-567. doi: 10.1109/WI-IAT.2009.94.

[10] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, & S. Linkman, «Systematic literature reviews in software engineering – A systematic literature review», *Inf. Softw. Technol.*, vol. 51, n.º 1, pp. 7-15, jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[11] M. Aksoy, S. Yanik, & M. F. Amasyali, «Reviewer Assignment Problem: A Systematic Review of the Literature», *J. Artif. Intell. Res.*, vol. 76, pp. 761-827, mar. 2023, doi: 10.1613/jair.1.14318.

[12] Z. Dong, Z. Shu, D. Zhen, C. Jie, Z. Yanping, & T. Jie, «A Multi-Label Classification Method Using a Hierarchical and Transparent Representation for Paper-Reviewer Recommendation | ACM Transactions on Information Systems», 2020, doi: doi.org/10.1145/3361719.

[13] T. Pradhan, S. Sahoo, U. Singh, & S. Pal, «A proactive decision support system for reviewer recommendation in academia», *Expert Syst. Appl.*, vol. 169, p. 114331, may 2021, doi: 10.1016/j.eswa.2020.114331.

[14] J. Jin, Q. Geng, H. Mou, & C. Chen, «Author–Subject–Topic model for reviewer recommendation», *J. Inf. Sci.*, vol. 45, n.º 4, pp. 554-570, aug. 2019, doi: 10.1177/0165551518806116.

[15] Y. Yong, Z. Yao, & Y. Zhao, «A Framework for Reviewer Recommendation Based on Knowledge Graph and Rules Matching», en *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)*, mar. 2021, pp. 199-203. doi: 10.1109/ICICSE52190.2021.9404099.

[16] C. K. Kreutz & R. Schenkel, «RevASIDE: Assignment of Suitable Reviewer Sets for Publications from Fixed Candidate Pools», en *The 23rd International Conference on Information Integration and Web Intelligence*, en iiWAS2021. New York, NY, USA: Association for Computing Machinery, dec. 2022, pp. 57-68. doi: 10.1145/3487664.3487673.

[17] Y. Kalmukov, «Automatic Assignment of Reviewers to Papers Based on Vector Space Text

Analysis Model», en *Proceedings of the 21st International Conference on Computer Systems and Technologies*, en CompSysTech '20. New York, NY, USA: Association for Computing Machinery, aug. 2020, pp. 229-235. doi: 10.1145/3407982.3408026.

[18] A. N. Medakene, K. Bouanane, & M. A. Eddoud, «A New Approach for Computing the Matching Degree in the Paper-to-Reviewer Assignment Problem», en *2019 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS)*, dec. 2019, pp. 1-8. doi: 10.1109/ICTAACS48474.2019.8988127.

[19] K. Mittal, A. Jain, K. S. Vaisla, & D. K. Tayal, «A Novel Method for Reviewer Assignment Problem based on Reviewers' Profile and Fuzzy Graph Connectivity Measure», en *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, jun. 2020, pp. 386-391. doi: 10.1109/ICIEM48762.2020.9160042.

[20] A. H. Patil & P. N. Mahalle, «Fair and Accurate Review in Publication Process: A learning-based Proactive Approach for Assigning Reviewers to Manuscripts», *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 12, n.º 4, Art. n.º 4, abr. 2021, doi: 10.17762/turcomat.v12i4.455.

[21] V. Latypova, «Reviewer Assignment Decision Support in an Academic Journal based on Multicriteria Assessment and Text Mining», en *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, apr. 2023, pp. 1-4. doi: 10.1109/ITNT57377.2023.10139187.

[22] R. P. Kusumawardani & S. O. Khairunnisa, «Author-Topic Modelling for Reviewer Assignment of Scientific Papers in Bahasa Indonesia», en *2018 International Conference on Asian Language Processing (IALP)*, nov. 2018, pp. 351-356. doi: 10.1109/IALP.2018.8629124.

[23] M. I. M. Ishag, K. H. Park, J. Y. Lee, & K. H. Ryu, «A Pattern-Based Academic Reviewer Recommendation Combining Author-Paper and Diversity Metrics», *IEEE Access*, vol. 7, pp. 16460-16475, 2019, doi: 10.1109/ACCESS.2019.2894680.

[24] B. Y. Alkazemi, «PRATO: An Automated Taxonomy-Based Reviewer-Proposal Assignment System», *Interdiscip. J. Inf. Knowl. Manag.*, vol. 13, pp. 383-396, oct. 2018.

[25] A. H. Patil & P. N. Mahalle, «Multi-label Reviewer Profile Building and Ranking based on Expertise, Recency, Authority and h-index: Vital Module of Reviewer Paper Assignment», *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 12, n.º 6, Art. n.º 6, apr. 2021.

[26] Y. Kalmukov, «A Software Tool for Accuracy Evaluation of Calculated Similarity Factors between Papers and Reviewers», in *2020 7th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE)*, nov. 2020, pp. 1-5. doi: 10.1109/EEAE49144.2020.9279032.

[27] S. A. Al-Zboon, S. K. Tawalbeh, H. Ai-Jarrah, M. Al-Asa'd, M. Hammad, & M. Al-Smadi, «Resolving Conflict of Interests in Recommending Reviewers for Academic Publications Using Link Prediction Techniques», en *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, oct. 2019, pp. 1-6. doi: 10.1109/ICTCS.2019.8923033.

[28] Z. Duan, S. Tan, S. Zhao, Q. Wang, J. Chen, & Y. Zhang, «Reviewer assignment based on sentence pair modeling», *Neurocomputing*, vol. 366, pp. 97-108, nov. 2019, doi: 10.1016/j.neucom.2019.06.074.

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

[30] A. Q. Abduljaleel & E. Al, «Reviewer Assignment Using Weighted Matching and Hungarian Algorithm», *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 12, n.º 7, Art. n.º 7, abr. 2021, doi: 10.17762/turcomat.v12i7.2629.

[31] C. K. Kreutz, K. Balog, & R. Schenkel, «Diverse Reviewer Suggestion for Extending Conference Program Committees», en *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, en WI-IAT '21. New York, NY, USA: Association for Computing Machinery, apr. 2022, pp. 79-86. doi: 10.1145/3486622.3493931.

[32] D. K. Pradhan, J. Chakraborty, P. Choudhary, & S. Nandi, «An automated conflict of interest based greedy approach for conference paper assignment system», *J. Informetr.*, vol. 14, n.º 2, p. 101022, may 2020, doi: 10.1016/j.joi.2020.101022.

[33] C. Yang, T. Liu, W. Yi, X. Chen, & B. Niu, «Identifying expertise through semantic modeling: A modified BBPSO algorithm for the reviewer assignment problem», *Appl. Soft Comput.*, vol. 94, p. 106483, sep. 2020, doi: 10.1016/j.asoc.2020.106483.

[34] A. H. Patil, P. N. Mahalle, V. H. Patil, & S. A. Bhavsar, «PROACTIVE SYSTEM FOR REVIEWER PAPER ASSIGNMENT», *J. Data Acquis. Process.*, vol. 38, n.º 2, pp. 1569-1609, abr. 2023, doi: 10.5281/zenodo.776779.

[35] L. Xu, D. Zeng, J. Dai, & L. Gui, «Combining Coverage with TMPS for Reviewer Assignment», en *2022 International Conference on Intelligent Education and Intelligent Research (IEIR)*, dec. 2022, pp. 107-113. doi: 10.1109/IEIR56323.2022.10050060.

[36] M. Mirzaei, J. Sander, & E. Stroulia, «Multi-Aspect Review-Team Assignment using Latent Research Areas», *Inf. Process. Manag.*, vol. 56, n.º 3, pp. 858-878, may 2019, doi: 10.1016/j.ipm.2019.01.007.

[37] D. M. Blei, «Latent Dirichlet Allocation», p. 30.

[38] S. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, & M. Gatford, «Okapi at TREC-3.», jan. 1994, p. 0.

[39] J. Devlin, M.-W. Chang, K. Lee, & K. Toutanova, «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding». arXiv, may 2019. doi: 10.48550/arXiv.1810.04805.

[40] A. Krizhevsky, I. Sutskever, & G. E. Hinton, «ImageNet Classification with Deep Convolutional Neural Networks», en *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012.

https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[41] A. Sherstinsky, «Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network», *Phys. Nonlinear Phenom.*, vol. 404, p. 132306, mar. 2020, doi: 10.1016/j.physd.2019.132306.

[42] M. Rosen-Zvi, T. Griffiths, M. Steyvers, & P. Smyth, «The author-topic model for authors and documents», en *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, en UAI '04. Arlington, Virginia, USA: AUAI Press, jul. 2004, pp. 487-494.

[43] H. Mou, Q. Geng, J. Jin, & C. Chen, «An Author Subject Topic Model for Expert Recommendation», en *Information Retrieval Technology*, G. Zuccon, S. Geva, H. Joho, F. Scholer, A. Sun, & P. Zhang, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 83-95. doi: 10.1007/978-3-319-28940-3_7.

[44] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, & J. Dean, «Distributed Representations of Words and Phrases and their Compositionality». arXiv, oct. 2013. doi: 10.48550/arXiv.1310.4546.

[45] Q. V. Le & T. Mikolov, «Distributed Representations of Sentences and Documents». arXiv, may. 2014. doi: 10.48550/arXiv.1405.4053.

[46] K. Cho *et al.*, «Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation». arXiv, sep. 2014. http://arxiv.org/abs/1406.1078

[47] T. Hofmann, «Probabilistic Latent Semantic Analysis». arXiv, jan. 2013. doi: 10.48550/arXiv.1301.6705.

[48] Y. Yao *et al.*, «Prompt Tuning for Discriminative Pre-trained Language Models». arXiv, may. 2022. doi: 10.48550/arXiv.2205.11166.

Citation: G. Decoppet and P. San Martin. *Systematic mapping of automated reviewer recommendation solutions.* Journal of Computer Science & Technology, vol. 24, no. 2, pp. 168-181, 2024.
DOI: 10.24215/16666038.24.e16.
Received: April 1, 2024 **Accepted:** September 23, 2024.