ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230715000112

Original scientific paper
Received: 2023-07-15, Accepted: 2024-01-02

Identification of LSA Data Retrieval Method and Temporal Graph for Document Retrieval

Shahla Rezvani, Nader Naghshineh*, Ahmad Khalilijafarabad

Abstract: The field of expert finding has seen a large number of approaches proposed both in universities and in industries, using a variety of new techniques in relevant data fields. This study tends to identify information retrieval method of latent semantic analysis and temporal graph for document retrieval. In this study, citation occurrence and author occurrence are independent variables and scales of expert author finding are dependent variables. The method used to evaluate judgment of document and author relevance in the test set formation phase is more similar to survey methods. Library method is used to study theoretical foundations and judge literature. This study has three populations: a) test set documents; b) people who make queries and judge relevance of retrieved documents; c) people who judge relevance of the retrieved experts. To measure judgments of document relevance, a method similar to peer tests is used. Among the retrieved results, repeated results are placed to determine accuracy and reliability of the judge. The degree of correlation obtained in this method is very high (0.98), indicating the reliability of the results. Regarding the results of the current study on application of latent semantic indexing (LSA) information retrieval model, which was ultimately used to retrieve expert authors, the performance of LSA-based retrieval model outperformed the baseline model. This was evident from the obtained metrics, including precision at the top 5 results (p@5) with a value of 0.895, mean average precision (MAP) of 0.839, and mean reciprocal rank (MRR) of 0.909. The improved retrieval performance can be attributed to the superior performance of the dimension reduction method compared to keyword matching.

Keywords: expert finding; expert retrieval; information retrieval; information systems; temporal graph; test set

1 INTRODUCTION

Latent semantic indexing (LSA) was one of the solutions that emerged to solve problems of vector space model. Vector space model is one of several methods for detecting similarity between two documents, which was developed by Salton in 1975 [1]. Vector space model treats dissimilar terms as irrelevant items. This, i.e. mismatch of terms, is the main problem of vector space model. The second drawback of these models is formation of a large and scattered phrase-document matrix, occurring in large document sets which requires a large storage space, and its processing and calculation time is also long. To solve these problems, one of the branches of this model called LSA is used, which has become popular in recent years [2, 3]. Latent indexing was first used by a group of scholars, Deerwoster et al. at Bell Corporation, for information retrieval [4] and was called latent indexing. The reason why scholars use the word ‘latent’ is that new statements that represent semantic information are not found directly from documents, but are the result of examining the set of documents and using the mathematical method called singular value decomposition (SVD). LSA assumes that usually the entire semantic content of a text such as paragraph, abstract or entire document is approximately equal to sum of the meanings of its words. Stable meaning of word representations can also be obtained from a large document set by considering each text as a linear equation of the entire set of documents in the form of a system of concurrent equations [5-7]. Yih et al. [8] showed that LSA improves vector space models by preferring the semantic concept of evidence over its words. In LSA, unlike document vector models that assume independent terms and words, different levels of correlation, dependence or connection are considered for them, and these connections between terms are characterized by formation of a new set of statements using statistical SVD method. In latent semantic space, a question and a document can also have a great cosine similarity even if they do not have a common statement. Because their statements are semantically similar, while this is not possible in vector models. The most important strength of LSA is retrieval efficiency based on user question, which is achieved through matrix calculation. As by using this method, relevant documents are retrieved even if the words of document content do not match with each other [9, 10].

Temporal information retrieval is presently a subject of study in the realm of information retrieval. Given the extensive amount of data available on the internet and the significant impact of time on document contents, procuring pertinent information becomes a formidable task. Traditional information retrieval approaches that are based on thematic similarity are not sufficient for searching the set of temporal documents. Time dimension of the documents available in the documents should be associated with the ranking of the documents for efficient retrieval. The objective of an information system is to detect documents that are relevant to a given query. However, the challenge lies in the fact that documents are influenced by time and continue to accumulate, resulting in a substantial number of irrelevant documents in the retrieved set. Consequently, users are compelled to invest more time in searching for documents that fulfill their information requirements. With the continuous influx of information in the digital realm, the incorporation of time as a crucial factor becomes significant for a wide range of searches [11, 12]. Kanhabua and Nørvåg [13] analyzed query logs and showed that a significant fraction of queries is temporal, that is, time-dependent relevance, and temporal queries play an important role in many fields such as digital libraries and document archives. There are two categories of temporal queries: 1) Those that provide a certain time criterion; and 2) those that are presented without time criterion. In this particular case, candidates who have retired or are no longer alive will not be retrieved. The introduction of the concept of graph and the subsequent development of graph theory are considered groundbreaking accomplishments in the fields of mathematics and various disciplines over the past few centuries. There are innumerable and numerous applications for graphs across a wide range of scientific disciplines, ranging from maths to computer science to chemistry to biology, to name a few. These domains inherently present significant problems that can be effectively represented and explored through the use of graphs. A temporal graph, as defined by [14], is a data structure comprising nodes and edges that are associated with time labels. There are two types of temporal graphs: those that undergo changes over time and those that do not. The temporal graph can be thought of as one in which time is discretized, in which only the relationships between participants may change instead of the different types of entities [15]. In the current study, the candidates who have retired may no longer have any publications in the considered field, and as a result, updated information is not available to the users. Considering the rapid changes that occur in the world of information and the passing of a few years reduces useful life of information, it is necessary to retrieve candidates who have up-to-date publications on the subject needed by users. This also applies to non-living candidates.

Latent semantic analysis [4] converts queries and documents into a latent semantic space. Cosine similarities can be found between queries and documents (cosine of the angle between two vectors, the closer it is to 1, the smaller the angle between two vectors), even though they have no common term; but as long as their terms have semantic similarity, this applies. Latent semantic space has less dimensions than latent space [16, 17]. Therefore, LSA is a method to reduce dimension. A temporal graph for document retrieval is presented in this study in order to identify the LSA information retrieval method.

2 LITERATURE REVIEW

Graphs easily present a series of objects and a set of dual relationships between them. These graphs commonly display bidirectional relationships, accompanied by supplementary details. As an illustration, consider a graph that represents a collection of cities interconnected by roads, where each edge (C1, C2) includes information on the average travel time from city C1 to C2. Similarly, in a graph designed to establish connections between atoms within a molecule, edges may contain additional data on bond order or bond strength. Such scenarios can be modeled using weighted, or more generally, labeled graphs, where each edge (and sometimes node) is assigned values from specific domains, such as the set of natural numbers. An example of a well-studied and extensively explored type of labeled graph is graph-coloring regions [18-20, 39]. A temporal graph is also known as a dynamic and evolutionary graph and can be informally described as a graph that changes with time. In terms of modeling, they can be considered as a special case of labeled graphs in which the labels include some time measurements.

Numerous applications and areas of research have identified the potential benefits associated with the development of a comprehensive collection of results, tools, and techniques for temporal graphs. Various types of networks, including but not limited to information and communication networks, social networks, transportation networks, as well as several physical systems, lend themselves naturally to being represented as temporal graphs.

A graph includes two sets; a non-null set of nodes or vertices and a set of edges that connect the vertices. Assume cities of a country as vertices and the roads between them as edges of a graph. A name is assigned to each vertex or each edge of the graph. A null graph is a graph that contains only vertices and its set of edges is empty, that is, it has no edges. A graph can be directed or undirected. A directed graph is a graph in which direction of each edge is determined. In a directed graph, order of the vertices in each edge is important, and the edges are drawn with arrows from the first vertex to the end vertex. In an undirected graph, one can move between vertices in both directions, and order of the vertices does not matter (Fig. 1).

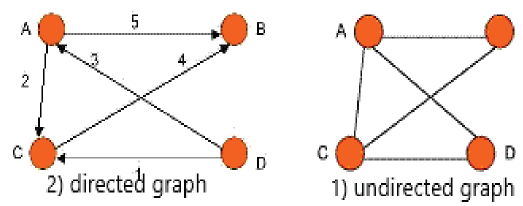

Figure 1 Types of graphs

Maximum number of edges in a simple directed graph with n vertices is $n \times (n - 1)$.

Maximum number of edges in a simple undirected graph with n vertices is equal to $n \times (n - 1)/2$.

2.1 Expert Finding

If technology is to strengthen the effective use of a wide range of knowledge, organizations should be able to not only use access to open and documented knowledge, but more importantly, they can also take advantage of tacit knowledge that other people have [35-38, 40-43]. By increasing the visibility and traceability of such knowledge, the knowledge can be analyzed and shared for strengthening the formation and sustainability of virtual organizations and companies, scientific communities, expert networks, etc. Yimam-Seid and Kobsa [21] identified two motives for expert finding, which are: 1) expert as a source of information; 2) expert as someone who is able to perform organizational or social role. Different situations where an expert is searched as a source of information are as follows:

1) Access to undocumented information. All the information within the organization is not fully documented; most of important information is gained through internships, experiences, and informal conversations. In many new situations, documented information is rarely helpful. Sometimes, the required information is not available to the public due to various economic, social and political reasons.
2) Specification of information needs. The information that users need is often not clear and specific. Therefore, it is necessary for them to consult with an expert to determine exactly what their information needs are.

3) Relying on expertise of others. Users often tend to spend less effort and time to find the information they need. This makes them use experts to select useful information from the huge amount of information.
4) Needs interpretation. In many cases, users are not interested in the information itself, but rather in its applications or interpretations. In some cases, they are not able to understand the retrieved information. In this situation, they turn to experts to interpret or understand the information.
5) The need to socialize. Users may raise their information needs with humans instead of interacting with documents and computers.

To satisfy the second type of motivation, those experts are required who are able to perform a specific task or role in the organization. In these cases, there is a need for those who have a certain type of expertise necessary to play a role in special situations. This occurs when inclusion of an expert in a specific activity or continuous cooperation between him and the searcher is considered. Examples include:

1) Searching for a consultant, employee or contractor,
2) Searching for a colleague, team member, committee member, or judge of journal and conference articles,
3) Searching for an expert speaker, lecturer, researcher, and interviewee for the media.

Searching for experts as a source of information is considered "information need", and searching for experts for the purpose of entrusting them with a role or task is considered "expertise need".

Table 1 Relevant research studies

No.	Authors	Result
1	Karimzadehgan et al. [22]	An algorithm was proposed to enhance the performance of expert finding systems by considering not only the expertise of an individual but also the expertise of their peers. The experimental results demonstrated that incorporating this supplementary information resulted in improved retrieval performance for the expert finding system.
2	Macdonald [23]	Effective document retrieval approaches have a positive effect on performance of voting techniques. Finally, this thesis states that the proposed model can be used to search for other people such as bloggers in the environment of blogs, and to suggest judges for papers submitted to conferences.
3	Daud et al. [24]	Occurrence of topics (probabilistic semantic classification of words) and correlation changes throughout time, while meaning of a specific topic remains almost unchanged. The results significantly show superiority of subject time general modeling approach over modeling approach that was not based on effect of conference and information time.
4	Smirnova & Balog [25]	To test their proposed algorithm, researchers used a real test set created from interactions of employees at the Tilburg University level, and the results showed that substantial progress has been made in all retrieval measures in the basic approach.
5	Omidvar et al. [26]	Novelty of this study is to find semantic relevance of the posts using text mining technique and semantic similarity provided by using WordNet. Evaluation tests were used to calculate precision of the proposed method and compare with other methods, and the results showed that it performed better.
6	Attapur [27]	In document retrieval stage, performance of other models was not significantly different from each other except for the Hiemstra linguistic model, which had a weaker performance compared to other models. Therefore, documents retrieved by any of the vector space models, Dirichlet language model, Okapi BM25 probabilistic model, PL2 model, and DLH13 model can be used to extract and rank authors.
7	Lipani [28]	The research article focused on the aim of studying biases in information retrieval systems. The thesis emphasized the reduction of retrieval systems to filters or sampling processes as a means to systematically investigate these biases. By approaching retrieval systems from that perspective, the researchers aimed to uncover and understand the various biases present in these systems. That knowledge could lead to the development of more effective and unbiased retrieval systems in the future.
8	Wu et al. [29]	This research paper aimed to address the retrieval of interns from CQA (Community Question Answering) websites. To identify suitable candidates for internship programs in companies, it introduced the notions of generalist and shape of expertise. The researchers conducted experiments using three test collections extracted from StackOverflow, comparing the effectiveness of their models against several baseline approaches. They presented retrieval models and assessed their performance using specific measures, highlighting their effectiveness compared to baseline methods.

It is shown in Tab. 1 that all works examined estimate the relationship between search queries and supporting documents concerning expertise retrieval using the occurrence of query terms in those supporting documents for the search queries. These models are not capable of semantic relevance. By using these models, if there are candidates really related to query and query words are not used in their respective supporting documents, but they contain words synonymous with the query words, they will not be considered as expert in that query. Subject-based models tend to solve this problem. The goal of these models is to somehow include the concept of meaning in the retrieval process. Two people with different words may describe a subject. Because of this, different people are introduced as experts for a specific query raised by different people. Usually, to calculate the degree of relevance of a document to a query, frequency of occurrence of the words of a query in the document is considered. If words other than query words with the same concept are used in the text, they are not considered as relevant words. This has a negative effect on calculation of relevance of a document to the query. As a result, it is necessary to consider multiplicity of meanings (one word with multiple meanings) or synonyms (different words with the same meaning) for words when calculating the degree of relevance of a document to a query. These descriptions show the challenges that exist in the expert finding process based on subject.

In this study, the supporting documents are the articles published by the authors, and the time factor or the date of publication of these articles is also considered as an important feature, which can be similar to giving different weights to different sources of expertise; of course, a time graph was used to include it.

3 RESEARCH METHODOLOGY

In this study, citation occurrence and author occurrence are independent variables and expert finding scales are dependent variables. In a typical classification, experimental research is divided into three types: real, preliminary and semi-experimental. The current study is also consistent with characteristics of a single-stage case study of preliminary experimental research subsets; it can be considered preliminary experimental research [27]. The method used to evaluate judgement of document and author relevance in the test set formation phase is more similar to survey methods. Library method is used to study theoretical foundations and literature.

This study has three populations: a) test set documents; b) people who make queries and judge relevance of the retrieved documents; c) people who judge relevance of the retrieved experts. The test set consists of a huge set of documents and has three components: set of documents, a set of subject elements (queries), and a set of relevance judgements [30]. Document set that makes up the main body of the test set is the first population. These documents consist of English articles in the field of information science and librarianship that have been indexed under the subject of information science and librarianship in the Web of Science database from 1989 to 2018 (N = 126924). Queries created by users were presented to all these articles and no sampling was done among the above articles. In order to make queries and make relevant judgements about the retrieved documents and authors, people should be selected who have knowledge of the query subject area and expert authors and have the ability to make relevant judgements about the retrieved documents. For this reason, graduates and postgraduate students of information science and epistemology of Tehran University will be selected as one of the populations.

The third population consists of people who judge relevance about expert authors retrieved for each query. These people should have a comprehensive understanding of the query as well as experts in that subject. To judge relevance of the retrieved expert authors, ten queries are randomly selected from each of the queries and presented to eight people who are introduced by people of the second population.

Various methods are used to calculate reliability coefficient of the measurement instrument, including re-run (retest method), parallel method (peer), split-half and Kudrichardson's method and Cronbach's alpha. To measure judgments of relevance of documents, a method similar to peer tests is used. For this purpose, repeated results are placed within the retrieved results to determine precision and reliability of the judge. The degree of correlation obtained in this method is very high (0.98) and indicates reliability of the results. To measure the relevance judgment of expert authors, parallel tests (simultaneous judgment of several people) is used.

4 RESULTS

To select the required number of components that represent maximum information available in the matrix, scree plot diagram is used, which according to the article presented for 250 components is as shown in Fig. 2. According to elbow position in Fig. 2, n = 100 for the number of components will be statistically precise enough to separate the subjects.

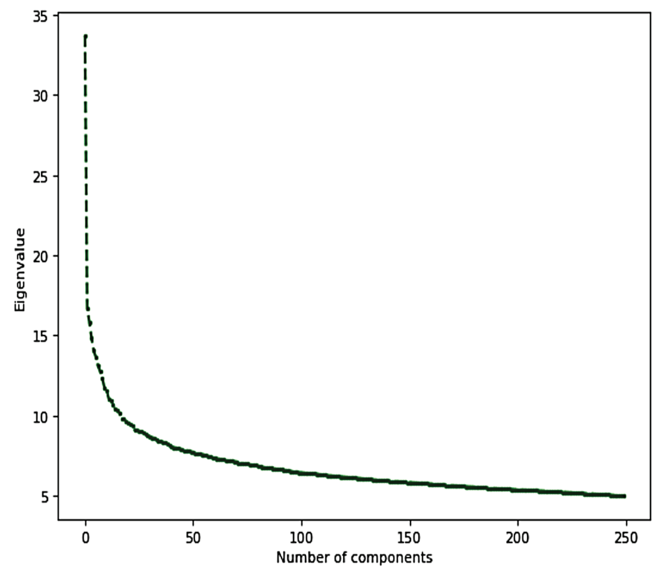

Figure 2 Scree plot for determining the number of components

The next step in this project is applying queries in the form of vector multiplication and finding relevant documents. In this method, the query is transformed into a vector with dimensions as long as the number of components of the above matrix based on the words in it and multiplied in each row of the result_transform matrix. The result gives us a numerical value, the larger it is, the closer it is to the relevant subject and, as a result, closer to content of that document (Tab. 2).

Table 2 Comparison of three measures in LSA model and the basic model

p@5 in LSA	MAP in LSA	MRR in LSA	p@5 in DLH13	MAP in DLH13	MRR in DLH13
0.895	0.839	0.909	0.887	0.567	0.903

After inserting the stage packages, including Numpy and Pandas, as well as JGraph package for drawing the graphs, root folder is determined, that is, address of the storage location, and then results file is read again in the form of a data frame by Pandas.

The "create-array" function converts the same variables into a list. Inside the parentheses is the query number. For example, if the number 100 is placed inside the parentheses, an array of author names, date of publication and citations of articles relevant to this query will be calculated.

In "get-names-year" function, main author name and year of publication are extracted from the references.

In "get-auth-names" function, the list of all authors of the article is separated because it is in the form of a string.

The "get-years" function extracts publication year from string mode and transforms it to list mode.

The "get-unique-name" function removes duplicate names from the list and provides a set of non-duplicate names.

The "draw_raph" function is for drawing graphs. Its sub-function i.e. "ith_graph" calculates the i^{th} function.

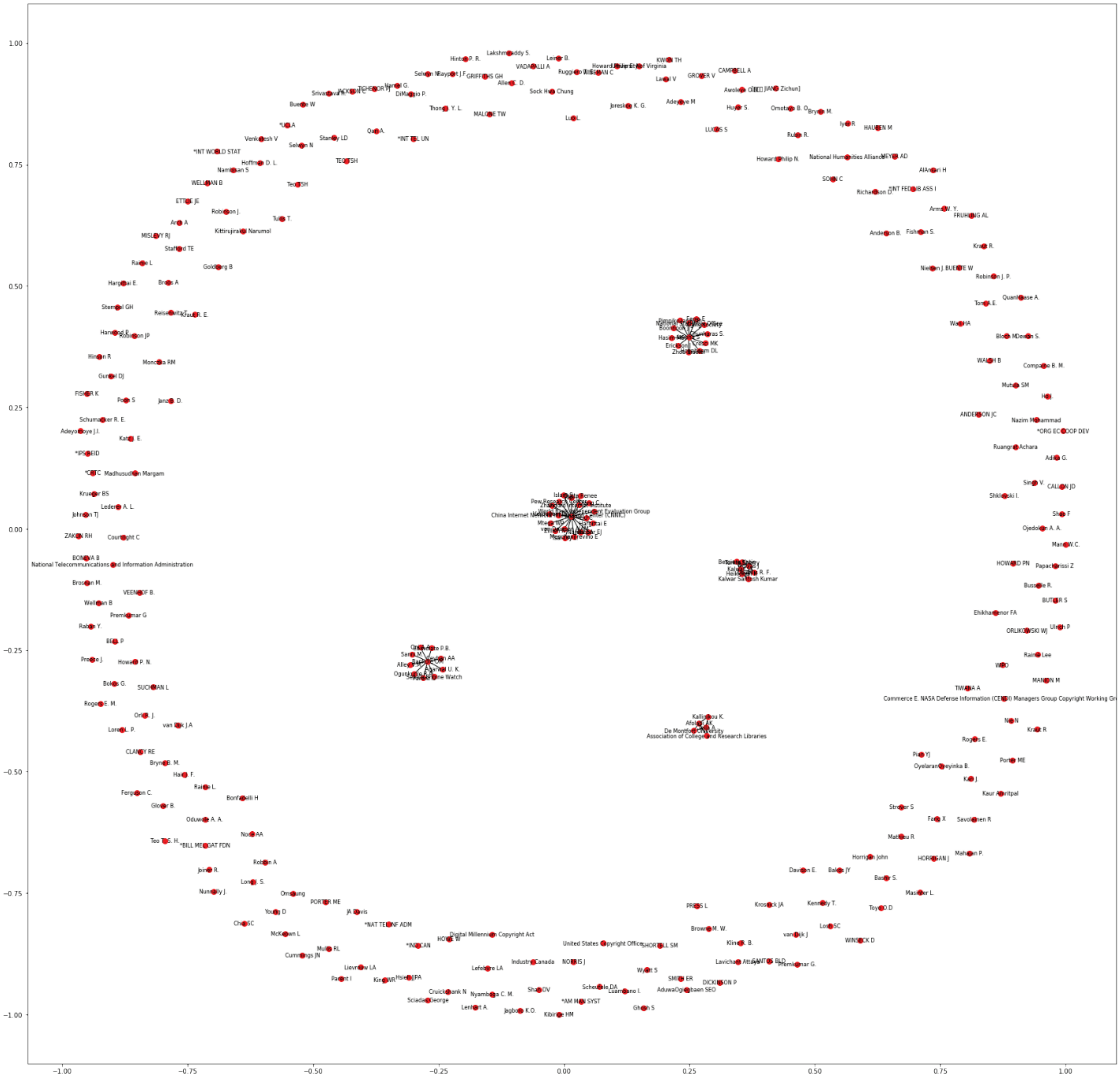

Figure 3 Temporal graph

As shown in Fig. 3, authors whose publication year was not in the period from 2008 to 2018 were removed from the graph.

In the following, authors should be retrieved whose works are published in the last ten years, because they lose their usefulness after ten or fifteen years, depending on the subject. Sciences that are mostly theoretical, such as mathematics, have a long half-life, and sciences that depend on new, up-to-date topics and technology, such as medicine, have a short half-life [44]. The minimum year is calculated from 2008. Then centrality measures, i.e. degree, closeness, betweenness and eigenvector centrality, followed by busy author, are calculated in the first hundred results.

The number of 10 queries from each study, a total of 20 queries, were randomly selected and the specified experts of each study were given a score of zero or one by the third statistical population, which was consisted of 8 people. The basic model was given value 4 by the 1^{st} person, value 5 by the 2^{nd} person, value 2 by the 3^{rd} person, value 4 by the 4^{th} person, value 4 by the 5^{th} person, value 5 by the 6^{th} person, value 8 by the 7^{th} person and value 5 by the 8^{th} person. Time inclusion and temporal graph were given value 8 by the 1^{st} person, value 7 by the 2^{nd} person, 10 by the 3^{rd} person, 9 by the 4^{th} person, 9 by the 5^{th} person, 9 by the 6^{th} person, 7 by the 7^{th} person and 8 by the 8^{th} person. The results of expert relevance judgment are shown in Tab. 3 and Fig. 4.

Table 3 Comparison of relevance judgement of the retrieved experts in the current model and basic model

Current model	*N*	Basic model	*N*
1^{st} person	7	1^{st} person	4
2^{nd} person	8	2^{nd} person	5
3^{rd} person	10	3^{rd} person	2
4^{th} person	9	4^{th} person	4
5^{th} person	9	5^{th} person	4
6^{th} person	9	6^{th} person	5
7^{th} person	7	7^{th} person	8
8^{th} person	8	8^{th} person	5
Sum	67	Sum	46

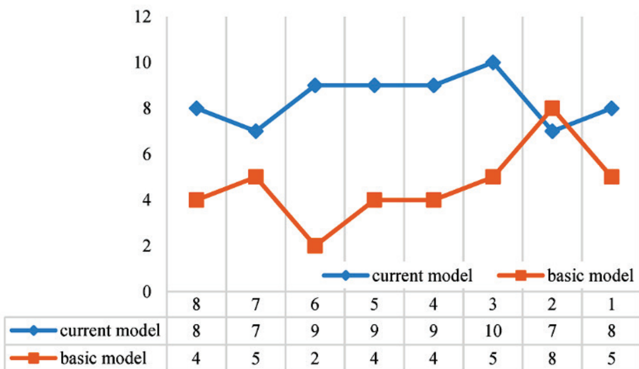

Figure 4 Comparison of the retrieved expert judgment in two models

According to the diagram, the retrieved experts have a higher relevance judgment in the current model, except for one point.

5 CONCLUSION

This study tended to identify LSA information retrieval method and temporal graph for document retrieval. According to the findings, LSA retrieval model outperformed the basic model in terms of p@5 (0.895), MAP (0.839) and MRR (0.909). The reason for the improved retrieval performance using dimension reduction methods compared to keyword matching lies in the use of latent semantic indexing (LSI), which is a type of conceptual indexing that utilizes statistical methods such as least squares estimation. The aforementioned indexing is extracted by employing this statistical approach. As we know, there are various ways to express a word (synonyms), so it is possible for query words to not be matched with document words. Additionally, many words have multiple meanings (polysemy), making information retrieval based on the concept and meaning of a document a better approach. LSI assumes that there are hidden structures in the usage of words that are partially captured by selecting diverse words. Singular value decomposition (SVD) is used to estimate these structures. The vectors obtained statistically enhance the representation of meanings more than individual words. Other research results also indicate that document retrieval using keyword matching is weaker compared to other methods. Furthermore, the performance of proposed model in retrieving documents is more pronounced in larger document sets compared to smaller ones [31]; however, this was not investigated in the current study. The results showed LSA has a better performance [32, 33].

One of the limitations of this study was that full text of the test set documents was not reviewed. The presence of full text of articles may have a positive effect on performance of information retrieval models. However, Bogers, Kox & Van Den Bosch [34] showed that performance of information retrieval models is better when indexed documents contain only abstracts than when indexed documents contain full text. This greatly reduces the problem of negative effect of not using the full text. The lack of a thesaurus in the document retrieval system is not a limitation due to the nature of LSA retrieval model explained above. In the document retrieval stage, relevance of the documents was judged based on title, abstract and keywords of the articles, which are common in most information retrieval systems. Since the present study tended to use the retrieved documents in the next stage to find expert authors, presenting the author names might lead to bias of the judges. For this reason, author names were not included in the relevance judgment files. However, information retrieval systems can consider the effect of the presence of these items on relevance judgment of users.

Based on the results of this research, the incorporation of temporal factors in expert finding and the utilization of social network indicators significantly enhance the performance of the method. The use of temporal factors to prevent the retrieval of individuals who are no longer active or have not published relevant works for an extended period has shown a significant improvement in the performance of the expert finding method compared to the baseline model. These findings can contribute to the enhancement of expert finding methods in the field of library and information science. Additionally, the use of social network centrality measures such as degree centrality, betweenness centrality, closeness, and eigenvector centrality as determining factors has also demonstrated a considerable improvement in the performance of the expert finding method compared to the baseline model. In the present study, ten queries were designed and sent to eight selected participants from the research population, and the results indicated that the use of the temporal graph and expert finding methods incorporating the factor of the highest number of published relevant works and the factor of social network centrality indicators yielded better performance.

Developing an organizational expert finding system that relies on expert profiles involves leveraging the organizational hierarchy [45]. This approach allows the prediction of relationships between managers, subordinates, and peers, as well as the identification of members with limited or no available information. The algorithm employed in this system takes into account not only the expertise of the individual members but also the expertise of their peers.

6 REFERENCES

[1] Salton, G., Wong, A. & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613-620. https://doi.org/10.1145/361219.361220

[2] Rezvani, S., Naghshineh, N. & Khalilijafarabad, A. (2023). Implementation of Experts' Retrieval Model Using Latent Semantic Indexing (LSA) Method and Temporal Graph. *Library and Information Science Research*.

[3] Ramezani, M., Shahryari, M. S., Feizi-Derakhshi, A. R. & Feizi-Derakhshi, M. R. (2023). Unsupervised broadcast news summarization; a comparative study on maximal marginal relevance (MMR) and latent semantic analysis (LSA). In *The 28th IEEE International Computer Conference, Computer Society of Iran (CSICC)*, 1-7. https://doi.org/10.1109/CSICC58665.2023.10105403

[4] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASI1%3E3.0.CO;2-9

[5] Cui, J., Wang, Z., Ho, S. B. & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 1-42. https://doi.org/10.1007/s10462-022-10386-z

[6] Landauer, T. K., Laham, D. & Derr, M. (2004). From paragraph to graph: Latent semantic analysis for information visualization. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5214-5219. https://doi.org/10.1073/pnas.0400341101

[7] Ahmad, S. N. & Laroche, M. (2023). Extracting marketing information from product reviews: a comparative study of latent semantic analysis and probabilistic latent semantic analysis. *Journal of Marketing Analytics*, 1-15. https://doi.org/10.1057/s41270-023-00218-6

[8] Yih, W. T., Zweig, G., & Platt, J. C. (2012). Polarity inducing latent semantic analysis. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1212-1222.

[9] Börner, K., Chen, C. & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, *37*(1), 179-255. https://doi.org/10.1002/aris.1440370106

[10] Duong, Q. H., Zhou, L., Meng, M., van Nguyen, T., Ieromonachou, P. & Nguyen, D. T. (2022). Understanding product returns: A systematic literature review using machine learning and bibliometric analysis. *International Journal of Production Economics*, *243*, 108340. https://doi.org/10.1016/j.ijpe.2021.108340

[11] Ma, K., Tan, Y., Tian, M., Xie, X., Qiu, Q., Li, S. & Wang, X. (2022). Extraction of temporal information from social media messages using the BERT model. *Earth Science Informatics*, *15*(1), 573-584. https://doi.org/10.1007/s12145-021-00756-6

[12] Chekuri, S., Chandrasekar, P., Banerjee, B., Park, S. H., Masrourisaadat, N., Ahuja, A., ... & Fox, E. A. (2023, June). Integrated digital library system for long documents and their elements. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL2023)*, 13-24. https://doi.org/10.1109/JCDL57899.2023.00012

[13] Kanhabua, N. & Nørvåg, K. (2010, September). Determining time of queries for re-ranking search results. In *International conference on theory and practice of digital libraries* (pp. 261-272). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15464-5_27

[14] Lightenberg, W., Pei, Y., Fletcher, G. & Pechenizkiy, M. (2018). Tink: A temporal graph analytics library for apache flink. In *Companion Proceedings of the Web Conference 2018*, 71-72. https://doi.org/10.1145/3184558.3186934

[15] Michail, O. (2016). An introduction to temporal graphs: An algorithmic perspective. *Internet Mathematics*, *12*(4), 239-280. https://doi.org/10.1080/15427951.2016.1177801

[16] Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 50-57. https://doi.org/10.1145/312624.312649

[17] Hu, X., Ma, W., Chen, C., Wen, S., Zhang, J., Xiang, Y. & Fei, G. (2022). Event detection in online social network: Methodologies, state-of-art, and evolution. *Computer Science Review*, *46*, 100500. https://doi.org/10.1016/j.cosrev.2022.100500

[18] Elumalai, A. (2020). Graph coloring and labeling applications in computer science. *Malaya Journal of Mathematic, S*(2), 2039-4041.

[19] Molloy, M. & Reed, B. (2002). *Graph colouring and the probabilistic method* (Vol. 23). Springer Science & Business Media. https://doi.org/10.1007/978-3-642-04016-0

[20] Vinutha, M. S. & Arathi, P. (2017). Applications of graph coloring and labeling in computer science. *International Journal on Future Revolution in Computer Science and Communication Engineering*, *3*(8), 14-16.

[21] Yimam-Seid, D. & Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, *13*(1), 1-24. https://doi.org/10.1207/S15327744JOCE1301_1

[22] Karimzadehgan, M., White, R. W. & Richardson, M. (2009). Enhancing expert finding using organizational hierarchies. In *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, Proceedings 31*, Springer Berlin Heidelberg, 177-188. https://doi.org/10.1007/978-3-642-00958-7_18

[23] Macdonald, C. (2009). The voting model for people search. *Doctoral dissertation*, University of Glasgow. https://doi.org/10.1145/1670598.1670616

[24] Daud, A., Li, J., Zhou, L. & Muhammad, F. (2010). Temporal expert finding through generalized time topic modeling. *Knowledge-Based Systems*, *23*(6), 615-625. https://doi.org/10.1016/j.knosys.2010.04.008

[25] Smirnova, E. & Balog, K. (2011, April). A user-oriented model for expert finding. In *European Conference on Information Retrieval* (pp. 580-592). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20161-5_58

[26] Omidvar, A., Garakani, M. & Safarpour, H. R. (2014). Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis. *Information Technology and Management*, *15*, 51-63. https://doi.org/10.1007/s10799-013-0173-x

[27] Attapur, H. (2015). The study of improving the efficiency of expert author finding model, focusing on the relationship between documents and people. *PhD Thesis*, University of Tehran, Tehran, Iran.

[28] Lipani, A. (2019). On Biases in Information retrieval models and evaluation. In *ACM SIGIR Forum*, 52(2), 172-173). https://doi.org/10.1145/3308774.3308804
[29] Wu, D., Fan, S. & Yuan, F. (2021). Research on pathways of expert finding on academic social networking sites. *Information Processing & Management*, *58*(2), 102475. https://doi.org/10.1016/j.ipm.2020.102475
[30] Sanderson, M. (2010). Test collection based evaluation of information retrieval systems. *Foundations and Trends® in Information Retrieval*, *4*(4), 247-375. https://doi.org/10.1561/1500000009
[31] Song, C., Guo, J., Gholizadeh, F. & Zhuang, J. (2022). Quantitative Analysis of Food Safety Policy—Based on Text Mining Methods. *Foods, 11*(21), 3421. https://doi.org/10.3390/foods11213421
[32] Berry, M. W., Dumais, S. T. & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM review*, *37*(4), 573-595. https://doi.org/10.1137/1037127
[33] Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(6), 683-692. https://doi.org/10.1002/wcs.1254
[34] Bogers, T., Kox, K., & van den Bosch, A. (2008, April). Using citation analysis for finding experts in workgroups. In *Proc. DIR*, 21-28.
[35] Nazari-Shirkouhi, S., Badizadeh, A., Dashtpeyma, M. & Ghodsi, R. (2023). A model to improve user acceptance of e-services in healthcare systems based on technology acceptance model: an empirical study. *Journal of Ambient Intelligence and Humanized Computing, 14*(6), 7919-7935. https://doi.org/10.1007/s12652-023-04601-0
[36] Noruzy, A., Dalfard, V. M., Azhdari, B., Nazari-Shirkouhi, S. & Rezazadeh, A. (2013). Relations between transformational leadership, organizational learning, knowledge management, organizational innovation, and organizational performance: an empirical investigation of manufacturing firms. *The International Journal of Advanced Manufacturing Technology, 64*, 1073-1085. https://doi.org/10.1007/s00170-012-4038-y
[37] Paksaz, A. M., Salamian, F. & Jolai, F. (2020). Waste collection problem with multi-compartment vehicles and fuzzy demands. *The 2nd National Conference on Industrial Engineering, Management, Economy and Accounting*. September 2020, Oslo, Norway. 1-12.
[38] Nazari-Shirkouhi, S., Keramati, A. & Rezaie, K. (2015). Investigating the effects of customer relationship management and supplier relationship management on new product development. *Tehnički vjesnik, 22*(1), 191-200. https://doi.org/10.17559/TV-20140623130536
[39] Neshatfar, S., Magner, A. & Sekeh, S. Y. (2023). Promise and Limitations of Supervised Optimal Transport-Based Graph Summarization via Information Theoretic Measures. *IEEE Access, 11*, 87533-87542. https://doi.org/10.1109/ACCESS.2023.3302830
[40] Khorsandi, H. & Bayat, M. (2022). Prioritizing operational strategies of saman bank. *International Journal of Health Sciences, 6*(S7), 1442-1453. https://doi.org/10.53730/ijhs.v6nS7.11548
[41] Gholami, M. H., Asli, M. N., Nazari-Shirkouhi, S. & Noruzy, A. (2013). Investigating the influence of knowledge management practices on organizational performance: an empirical study. *Acta Polytechnica Hungarica, 10*(2), 205-216. https://doi.org/10.12700/APH.10.02.2013.2.14
[42] Azimi Asmaroud, S. (2022). Preservice Elementary Teachers' Categorical Reasoning and Knowledge Transfer on Definition Tasks with Two Dimensional Figures. *Theses and Dissertations*. 1588. https://ir.library.illinoisstate.edu/etd/1588
[43] Zarei, P. & Dobakhti, L. (2023). An Exploration of Iranian Teachers' Professional and Institutional Identities and their Enactment of Critical Pedagogy. *Teaching English Language*, 1-25.
[44] Rasekh Eslami, Z. & Zohoor, S. (2023). Second Language (L2) Pragmatics and Computer Assisted Language Learning (CALL). *Technology Assisted Language Education, 1*(2), 1-17.
[45] Zandi, J. & Naderi-Afooshteh, A. (2015). LRBAC: Flexible function-level hierarchical role based access control for Linux. In *The 12th IEEE International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC2015)*, 29-35. https://doi.org/10.1109/ISCISC.2015.7387894

Authors' contacts:

Shahla Rezvani
Department of Information Science and Knowledge Management,
Faculty of Management, University of Tehran, Tehran, Iran
E-mail: shahla.rezvani.gi@ut.ac.ir

Nader Naghshineh, Associate Professor
(Corresponding author)
Department of Information Science and Knowledge
Management, Faculty of Management, University of Tehran,
Tehran, Iran
E-mail: Nnaghsh@ut.ac.ir

Ahmad Khalilijafarabad, PhD
Department of IT Management, Faculty of Management,
University of Tehran, Tehran, Iran
E-mail: Ahmad.khalili@ut.ac.ir