Supporting Scholarly Search by Query Expansion and Citation Analysis

Shah Khalid
School of Computer Science and Communication Engineering, Jiangsu University, China and NUST, Islamabad, Pakistan
shahkhalid@ujs.edu.cn

Shengli Wu
School of Computer Science and Communication Engineering
Jiangsu University, China
swu@ujs.edu.cn

Abstract−**Published scholarly articles have increased exponentially in recent years. This growth has brought challenges for academic researchers in locating the most relevant papers in their fields of interest. The reasons for this vary. There is the fundamental problem of synonymy and polysemy, the query terms might be too short, thus making it difficult to distinguish between papers. Also, a new researcher has limited knowledge and often is not sure about what she is looking for until the results are displayed. These issues obstruct scholarly retrieval systems in locating highly relevant publications for a given search query. Researchers seek to tackle these issues. However, the user's intent cannot be addressed entirely by introducing a direct information retrieval technique. In this paper, a novel approach is proposed, which combines query expansion and citation analysis for supporting the scholarly search. It is a two-stage academic search process. Upon receiving the initial search query, in the first stage, the retrieval system provides a ranked list of results. In the second stage, the highest-scoring Term Frequency–Inverse Document Frequency (TF-IDF) terms are obtained from a few top-ranked papers for query expansion behind the scene. In both stages, citation analysis is used in further refining the quality of the academic search. The originality of the approach lies in the combined exploitation of both query expansion by pseudo relevance feedback and citation networks analysis that may bring the most relevant papers to the top of the search results list. The approach is evaluated on the ACL dataset. The experimental results reveal that the technique is effective and robust for locating relevant papers regarding normalized Discounted Cumulative Gain (nDCG), precision, and recall.**

Keywords-academic search; query expansion; citation analysis; pseudo relevance feedback; user relevance feedback

I. INTRODUCTION

The rate in publications is about 2.5 million per year [1]. This large increase in the number of scholarly publications makes finding relevant papers with a few keyword query a challenging task [2]. This can be caused by several reasons: First, the problem of synonym and polysemy [3, 4], i.e. the query terms submitted, can be related to multiple topics due to which the search results list may not contain the intended papers. Second, the query terms can be too short making hard to discover what papers the user wants. This may be a matter of habit: searchers usually formulate very short queries (e.g. the average size of query terms is 2.4 words [5, 6]). Third, a new researcher has limited knowledge and often is not sure about what is looking for until the results are displayed. Even if searchers know what they are looking for, they are unable to formulate the search query for increasing accuracy and completeness of the search results. These issues obstruct scholarly retrieval systems in locating highly relevant publications for a given search query. To deal with these issues, a lot of research has been done in recent years, and the output has been presented in different styles, including research papers [7], books [8, 9], doctoral dissertations [10, 11], test collections [12], retrieval evaluation events [13], etc. To the best of our knowledge, no technique has been investigated that uses both query expansion (QE) by PRF and citation analysis in scholarly search.

In this study, we emphasize the use of QE methods by PRF and citation analysis to support academic searchers in finding the most relevant papers on their queries. In our recent research work, we incorporated QE by considering user's relevance feedback with citation analysis [7]. It was assumed that a user's search query if expanded by considering interesting terms from the initial few top-ranked results as pseudo relevance feedback (PRF) besides the citation graph may provide better performance [7]. Now we believe that this assumption may support academic searchers faster because the system does not wait for the user feedback. In this paper, we look at Pseudo Relevance Feedback (PRF)-based QE methods in greater detail. PRF-based QE technique augments the original user's query with terms generated from the initially retrieved results list. This technique has advantages and disadvantages. One of its main benefits is that it performs QE without the user's interaction. In contrast, scholarly retrieval systems also inherit the fundamental issues of search engines, i.e. at first the initial retrieval list may be a mixture of papers about different subjects and documents. Second, document retrieval systems normally provide a long list of documents ranked by their relevance to the user's query. To address these issues while expanding the original query, we utilize citation analysis and the concept of interesting terms from a top few papers [7].

In summary, the key contributions of this paper include: (1) We formulated the index to facilitate the design of QE by PRF and the results in re-ranking strategies. (2) We developed an algorithm that uses QE (PRF) and citation analysis in

Corresponding author:Shengli Wu

extracting and weighting QE terms. (3) We evaluated the proposed framework on the ACL data set using standard evaluation metrics, i.e. nDCG, Recall, and Precision, to demonstrate the comparative analysis.

II. RELATED WORK

Note that here we will not concentrate on the well-known text retrieval models such as BM25, VSM, Jaccard Index, n-gram string matching, etc, since the information about these models are easily available, and they rely on word matches. The aim of this study is to investigate how QE by PRF and citation graph may support academic searchers [2]. For this only the recent relevant literature in the domain of scholarly retrieval systems was taken into account in order to demonstrate our proposed technique.

Sofia Search is a well-known example in the domain of academic search for identifying relevant articles [14], which starts from the initial set of papers and follows both the in-links and out-links of the papers repeatedly up to a given depth or when a desired numbers of candidates is found. However, in the growing rate of research papers, the use of Sofia Search is limited. It needs seed papers while all the in-links and out-links are not equally relevant [15]. Most of the approaches that use citation graphs do so in combination with content-based approaches. Examples include academic search engines such as Google Scholar, PubMed, and CiteSeerX which use the links between scholarly articles provided by citation network analysis for documents ranking. CB method processes the textual content of the papers, which can be title, abstract, keywords, and main content. The text-based methods weigh the relevant articles by the frequency and position of the terms in the article. Based on the term weight, several techniques have been developed to estimate the relatedness of articles. PubMed is a popular scholarly retrieval system, primarily designed for biomedical literature [16]. It is maintained by the US NCBI (National Center for Biotechnology Information) with over 28 million articles. It reflects many factors of the scholarly article for indexing and retrieval including (a) stemming, (b) number of terms in the article (TF), (c) position of terms (i.e. title, abstract, body-content), (d) weight of the terms in the article, and (e) key terms of the article in a domain-specific database (e.g. MeSH Database). Recently, PubMed enriched its search architecture by considering two stages. In the first stage, it retrieves articles that match a user query using standard Information Retrieval (IR) weighting function BM25. In the second stage, it re-ranks the top 500 articles using learning to rank (L2R) method. Web search query is an integral part of the IR, and it is generally accepted that searchers habitually pose short queries to search engines [5, 6]. Many research works have shown the effectiveness of QE by adding new words to original queries [7, 18]. Authors in [19] proposed synonym weighting strategies for biomedical retrieval. They took into account 'gene synonym QE' in biomedical information retrieval and presented the effectiveness of the technique in retrieving relevant information from biomedical literature. Likewise, Article Retrieval for Precision Medicine (ARtPM) has recently been proposed for relevant article retrieval using query formulation and expansion [20]. In ARtPM, the searcher has to specify a query consisting of the disease, contextual medical condition, and genetic mutation. It then uses several external resources to formulate and extend a query for effective article retrieval.

Several other approaches have been practiced to make scholarly search engines more effective [14, 16, 21-24]. Among these, the Explicit Semantic Ranking (ESR) is a well-known and recent representative approach that uses knowledge graph embedding in ranking scholarly documents [24]. It uses Semantic Scholar corpus, query log, and freebase for building an academic knowledge graph. Its knowledge graph considers concept entities and their descriptions, context correlations, relationships with authors and venues, and embedding trained from the graph structure. It uses L2R to query and represent documents (as entities in the knowledge graph) in the embedding space. Another popular approach is to transform the concept of keywords into key queries [22]. Key phrases are extracted from input documents for formatting key quires with the objective of finding more relevant documents. Liu introduced another technique called CCSE (Core Content Similarity Estimation) for retrieving scholarly articles with similar core content [16]. For a given article, CCSE recommends those articles that share similar core content terms with it. It has two interesting features: (1) it works on article titles and abstracts only, which are freely available on the Web and (2) it improves inter-article relation estimation by considering the core contents of the article, which include the research goal, background (problem description), and conclusion of the article. Recently, we used QE through URF and citation networks analysis for relevant retrieval of scholarly articles [7]. However, to the best of our knowledge, until now no published effort has investigated the use both QE by PRF and citation analysis for academic search in identifying relevant papers.

III. PROPOSED FRAMEWORK

The proposed technique enhances academic search not only by QE via PRF but by performing citation analysis while ranking scholarly articles. The citation network analysis can play a vital role in identifying influential papers [25]. Figure 1 demonstrates how the QE works in our proposed approach.

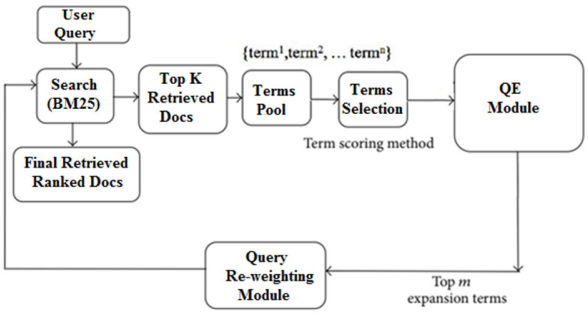

Fig. 1. The proposed QE model.

We have used BM25 similarity measure and citation analysis for selecting an initial set of retrieved documents, which is more efficient than the traditionally used cosine similarity measure.

To construct the terms pool, as shown in Figure 1, we first retrieve several top documents from the first retrieved document set for the query using a matching function and citation analysis. Once the top relevant documents are retrieved, all the unique terms of top documents are selected to form a term pool or candidate term set. The terms are ranked by the TF-IDF scoring scheme to rank the terms based on their appropriateness for QE. The following subsections illustrate the proposed approach.

A. Indexing Formulation

Index formulation parses the document collection into indices so that retrieval can be done accordingly. We used the Solr v7.2 IR platform for indexing and retrieval. Solr tokenizes the free text in the target document collection to index tokens and passes through a term pipeline, which removes stop words and performs stemming to the indexing terms. A predefined list of stop words and stemming algorithms have been configured in Solr schema to carry out all the steps in the pipeline. The tokens obtained after the term pipeline were used to generate indices. The document indexing is designed in a way to consider both QE via PRF and citation analysis, i.e. the indexing scheme keeps paper citation network record for feeding into citation networks analyzer to utilize it in the strategic ranking of the proposed approach.

B. Structural Overview

This section demonstrates the pictorial flow of the way the system refines the final results set by considering citation networks analysis besides QE in article ranking. Identifying the most cited papers using citation networks by means of the ranking algorithm has gained considerable attention [7]. Citation network is a graph model G_p = (P, E), consisting of nodes (P) and edges (E). Here, P is the set of nodes representing papers P= $\{P_1, P_2, P_{3...} P_n\}$, and E is the set of edges E= $\{e_{ij}, 1 \leq i \leq n, 1 \leq j \leq n, i 6= j\}$ that represent the links between papers. For example, if P_i cites P_j, then there should be an edge E_{ij} in between P_i and P_j. This edge/citation or formal reference shows a conversation between the paper's authors and exists when a published article cites an external source. Citation network codifies scholarly conversation and plays a vital role in ranking scholarly articles [2]. This conversation appears in each published article as a pointer to other published views under consideration in footnotes, endnotes, or bibliographies for several aims. For instance, if some authors want to contradict the arguments of another author, they will cite the works where those arguments appear. Likewise, if other authors rely on the conclusions of someone else's work or his previous work, they will cite it accordingly. The proposed framework uses citation analysis along with the BM25 retrieval model while computing the base-weight of articles for a given user query. The schema-independent view of how the system works is visually presented in Figure 1. In Figure 2, relevant and irrelevant papers to the user's query are indicated by red and blue rectangles respectively. The experimental analysis presented below reveals that QE by PRF and citation analysis can support academic searchers in a more nuanced way with a little computational overhead. The original titles of the documents with interesting terms to the user query are explicated in Figure 2 as a general retrieval scenario. The following section explains the way the proposed approach considers and expands the original query.

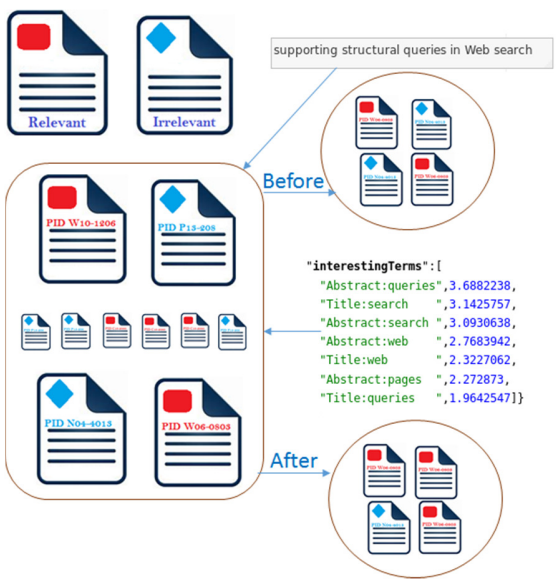

Fig. 2. General retrieval scenario.

C. PRF Documents Selection for Query Expansion

Query Expansion (QE) methods mainly depend on feedback documents and extracted interesting terms from those documents. The documents used by PRF query expansion methods might not be all relevant. Without loss of generality, relevant documents contain relevant terms to the query, and those terms are useful in finding more relevant documents when added to the query. However, non-relevant documents may contain noisy terms that are not relevant to the query, and choosing those noisy terms for QE will not help in finding more relevant documents, but it will fetch irrelevant documents instead. Thus, it is crucial to accept proper feedback documents and feedback terms for QE. Selecting feedback documents in the QE process for scholarly document retrieval was first described in [26]. In the process of choosing good feedback documents, our classification module as a local QE method [27] adjusts the query relatively to the documents that match by considering the query and the related documents in the form of Term Frequency and Inverse Document Frequency (TF–IDF). This approach selects the interesting terms from the title and abstract of a few top articles as a PRF from the first round search results list. The selected papers are then processed in Rocchio style [28] to revise the search query using the TF-IDF score of the top *k* words from title and abstract. In this way, the revised query and citation analysis are utilized to produce the final results list.

D. Why PRF?

It has been proved that URF-based techniques are better than PRF-based techniques for scholarly document retrieval [29, 30], but they require the cost of human interaction. In our recent work [7], we utilized URF besides citation analysis for supporting the scholarly search. In this paper, we take into account the advantage of PRF-based techniques, as they are

fully automated and do not require expensive outside inputs in the retrieval process while still performing well in IR. In the PRF technique, the system assumes the top *n* documents retrieved in the initial run to be highly relevant and fetches the terms in those documents for QE. In our system, QE via PRF is applied to all the runs by altering the Solr unsupervised feedback QE plugin with the configuration of 20 terms from the top 3 documents retrieved in the initial run for adjusting the query.

E. Query Expansion/Formulation

In the QE process, the terms having the highest TF-IDF are called interesting terms. The approach extracts these terms from the first results. Both of these mechanisms can be implemented as Solr plugins [31] by using the Rocchio algorithm. In our experiment, we employ the top 20 interesting terms after normalization in query reformulation through the Solr request handler. Our local-based QE procedure proceeds as follows: First, the base query is run for initial retrieval. The result is a ranked list of articles from the ACL collection, in descending order of predicted relevance. From this ranked list, we define the top-ranked R articles as the relevant set, $S_R = \{a_0, a_1, ..., a_R\}$, Each term a_i appears in a document S_R and in the reference section of the relevant papers after the citation networks analysis. This extraction creates an extended list of candidate terms, and their frequencies can then be used for QE weighting. A weighted query is then built with those top terms, and this query is combined with the original one to retrieve the final set of documents using citation network analysis. In our experiment, we take R=3 and consider the top 20 terms for QE.

F. Results Re-ranking after QE using Citation Analysis

After the retrieval results have been obtained from Solr, a re-ranking technique was applied to the results with the expectation of boosting the system performance by adjusting the order of the retrieved documents. In this technique, a rearranging methodology was employed based on the availability of the desired terms in the retrieval results using citation analysis. Besides citation analysis, the interesting terms are the deciding factors for determining a document's relevancy. Based on these conditions, a module is designed to rearrange the ordered list in such a way that the documents having the highest TF-IDF besides citation analysis were given a higher relevance score. A description of the approach is described below.

G. The Materialization of the Framework and Algorithm

The framework implements the above-mentioned task in three steps represented by three processing flow-lines, shown at the bottom of Figure 3. Blue arrows indicate the pre-processing, including information and citation extraction/ parsing of the research articles. The green lines show the actual search flow, i.e. how the approach performs initial retrieval. The brown lines represent QE and final retrieval. The system extracts interesting terms for QE from the top-three papers keeping in view both TF-IDF and citation score. Interesting terms are the ones having the highest TF-IDF, which is the most widely used technique for keyword extraction [32]. The system runs the revised query in the background for generating the final results list.

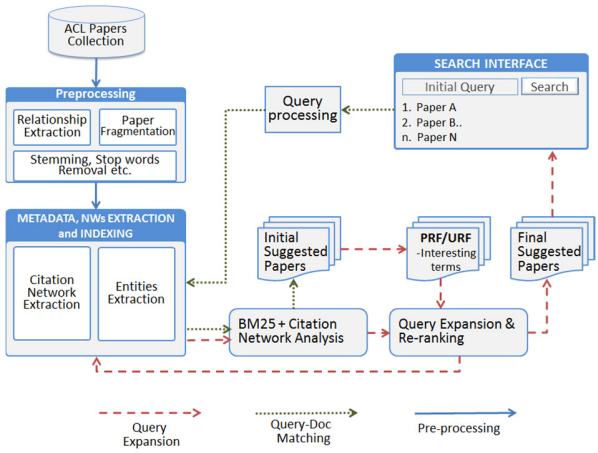

Fig. 3. Flow diagram of the proposed approach.

1) Metadata and Citation Networks Extraction

The system pre-processes the corpus to extract metadata (e.g. authors, publisher, date, incites, out-cites), to construct citation networks, and to fragment papers into different desired fields (title, abstract, and content). The extracted data are then indexed in a multi-core architecture for efficient retrieval.

2) Base Weight Computation, QE via PRF and Citation Analysis

In the second step, the framework computes the base weight of each paper for the initial query using BM25 and citation analyzer in terms of different indexed fields. The approach is described in the algorithm in Figure 4.

Algorithm: Retrieval by considering QE & citations analysis
Input: Q ← initial user query, P ← Paper
Output: List of Papers
1 IntialResult ← call Procedure-1 (Q, P)
2 URF ← selected Papers from intialResult (automatic selection of top few papers (up to 3 papers))
3 Q_E ← URF/* it uses the top k terms having highest tf.idf score from title and abstract of the user selected relevant papers*/
4 Call Procedure-1 (Q_E,P)
5 **Procedure-1**
6 For each paper P_i *in N do*
7 **Using Equation 1**
8 End
9 P← W_i
10 While $W_i \neq 0$ do
11 For each W_i do
12 **Compute each paper score using** $PR_{(P_i)} = \frac{1-d}{N} + d\sum_{P_j \in in(P_i)} PR_{P_j} . W_{in(P_i,P_j)} . W_{out(P_i,P_j)}$
13 **End**
14 **End**
15 Return PL
16 **End Procedure-1**
17 Revised Result after user feedback

Fig. 4. The retrieval algorithm.

The algorithm shown in Figure 4 is the main algorithm that implements the proposed technique. It calls Procedure 1 in line

1 for the initial search query to retrieve the first set of search results. Procedure-1 computes the BM25 score of all the papers against the initial search query, which is combined with the citation analysis score to compute the final base weight of the candidate papers and ranks the initial search results in steps 6-14). The formula used in line 12 in Procedure 1 computes paper score *PR* of each paper P_i, in which d is a parameter usually set to 0.85 and N is the total number of research papers in the citation network, i.e. the set of all the nodes which in-link to paper P_i. $Wi_{n(Pi, Pj)}$ is the score of the link (P_i, P_j) that is estimated based on the number of incites of paper P_i and the number of incites of all reference papers of paper P_j. In line 2, the system selects the top 3 papers behind the scene for the QE process. In line 3, it uses these selected relevant results in selecting the most interesting terms as the revised query terms, i.e. used in QE for further processing. The algorithm, once again, calls Procedure 1 in line 4 to compute the base weights of all the newly matched candidate papers against the revised query. Based on the newly computed scores, the search results are re-ranked to the user in step-17. To demonstrate how the proposed approach work, we present an example-query in Figure 5: Let q be the search query "natural language processing technique." This initial query is represented by a list of terms {i.e. q_1, q_2… q_n}, and C is the list-of-candidate terms for QE, represented by {c_1, c_2 … c_k}. The initial set of C is selected out of all of the terms in the first m (a parameter to be set) selected papers, which include all terms found in the selected papers. The list of candidate terms C is then extended by all the terms that appear in the selected documents.

Fig. 5. Example query.

For QE, we devise a simple approach inspired by Rocchio's relevance feedback method [33]. An inverted index implementing BM25 is used initially to retrieve a ranked list of documents matching the original query. This list of documents acts as a data source for QE. Our inverted collection index allows accessing the TF-IDF weights of terms. The TF-IDF weights of every term (word) in the n top-ranked documents are summed up, and the terms are sorted by their accumulated weight. Finally, the first k terms of the sorted list are added to the original query. Figure 5 demonstrates the general retrieval and QE scenario of the proposed approach by considering the PRF with citation networks analysis. The first part of Figure 5 displays a few top-ranked papers from the initial retrieval after posing the query q. The Figure has four types of information: the user query, paper ID, title, and relevance score. The relevance score is displayed here from the human evaluators to demonstrate the effectiveness of the proposed approach. The top-ranked papers of the initial query are considered for QE, from which several terms are extracted. The new, expanded query is then re-run on the collection for the results set. This demonstrates that the proposed QE algorithm increases the precision of the scholarly retrieval system. As we see, the final retrieval has comparatively relevant papers because the documents which include the interesting terms, are listed at the top of the search results list. The following subsections describe the experimental methodology, and the data set used, the results, and performance comparison.

IV. Experimental Set Up and Results

The dataset consists of 23058 papers, 17695 authors, and 121137 citations, indexed in Solr [34]. For the evaluation, we use standard evaluation measures shown in Table I. For Recall, we took into account the assumption described in [22] that a human would often not consider many more than the top-50 results of a single query.

A. *Dataset and Query Formulation*

The ACL Anthology Network (AAN) [35, 36] dataset is used for the experimentation of the proposed technique. It is an IR dataset having 23058 research papers from the ACL anthology. The articles are from the field of computational linguistics. Some of the statistics of the data set are given in Table I.

TABLE I. SOME OF THE ACL DATASET STATISTICS

Item	Value
Papers	23,058
Authors	17,695
Venue	350

We evaluated the proposed approach on 60 queries (associated with each subject area following the method to capture the logic of underspecified queries to evaluate the effectiveness of the system accordingly). These queries were structured and formulated for experimental analysis with the help of three Ph.D. students, because the users' satisfaction can only be measured based on the experience of real users.

B. *Evaluation Criteria*

Generally, for the relevance judgments of scholarly search systems, two different methods are used. The first method considers the reference list of the paper as ground truth to check that if the system can re-identify them or not. In the second method, the scholars' relevance judgments are taken into account. The first method is rather prejudiced toward citation networks analysis, but it does not address the use case we have in mind. Therefore, we adopted the second method and involved users to check the system's relevance.

C. *Performance Comparison*

We used a four-point scale for measuring nDCG: 3 for highly relevant, 2 for relevant, 1 for marginally relevant, and 0 for non-relevant. Moreover, as recall needs to have expected relevant papers and can be computed on any k, therefore, we took the hypothesis that a researcher would often not consider more than the top 40 related papers for a given single query. We compared the proposed technique's results with URF's (our previous technique) [7] and BM25's. To have a closer look and investigate the generalizability, we present the precision, recall, and nDCG results and graph of the three models in Table II and Figure 6.

TABLE II. PERFORMANCE COMPARISON

Method	P@5	P@10	P@20	R@15	nDCG@10
BM25	0.7148	0.6728	0.6335	0.2045	0.6812
URF	0.7600	0.7508	0.71	0.23	0.72231
PRF	**0.7534**	**0.7345**	**0.7074**	**0.2213**	**0.7115**

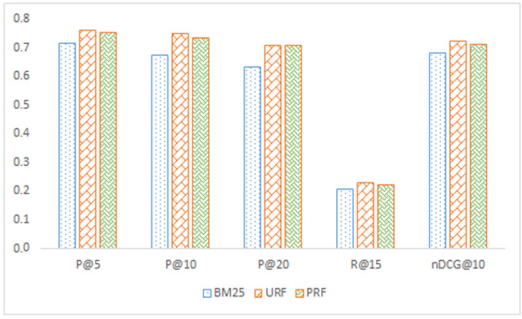

Fig. 6. Performance comparison.

The participants assessed the results using the 4-point scale. Each participant evaluated the proposed technique for 20 different input queries. For each query, the top 20 results were checked, i.e. in total each participant evaluated about 400 papers. The papers suggested by BM25, URF [7], and our proposed PRF approach vary. Out of all the evaluated papers, approximately 34% of the results were found highly relevant, 22% relevant, 19% marginally relevant, and 25% not relevant at all to the given input papers. The nDCG@10, Precision@5, 10 20, and Recall@15 of the proposed framework are presented in the fourth row of Table II. The results show that QE via PRF, along with a citation analyzer, improve precision, recall, and nDCG in all the cases. Overall, the graph in Figure 6 shows that the proposed technique can refine scholarly search effectively as compared to no query expansion (BM25). The URF [7] provides better results because the human can select more relevant results for QE. The benefits of the PRF approach includes saving computing resources and less human intervention. Since it assumes that the top papers returned by the initial user query are relevant, interesting terms are extracted from the top-ranked papers from the initial results to formulate a new query for a second retrieval cycle.

D. *Microscopic Analysis*

In this section, we evaluate the performance of the proposed technique at the query level. The analysis is given in Figure 7, which illustrates the fact that for good nDCG, the precision rate can be obtained at the top 5 and 10 results with the integration of both citation analysis and QE via PRF. The blue line in Figure 6 dominates the other line fluctuations. The precision rate at the top 5 results is higher for both the retrieval approaches and close in performance at the top 20. Overall, the microscopic analysis reveals our hypothesis that the incorporation of QE through PRF and citation network analysis can support academic searchers in today's colossal expansion of academic literature.

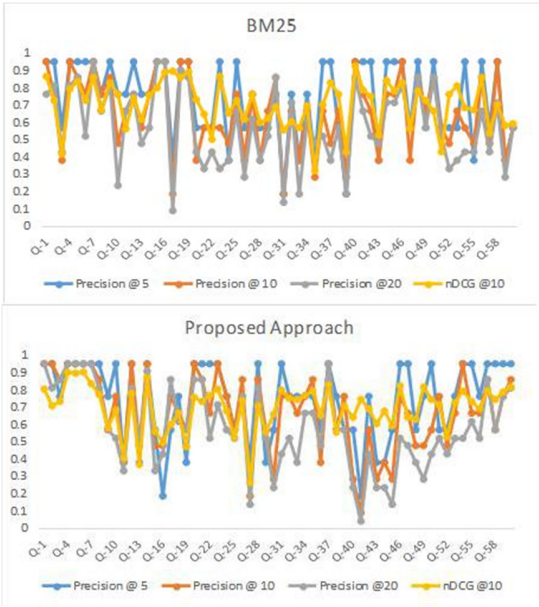

Fig. 7. Microscopic analysis.

Finally, the approach is statistically more significant than the QE approach. The approach has a performance penalty: the query is executed twice. However, this can be partially mitigated through the intelligent cache, while it is known that the academic searchers usually spend more time in query reformulation to obtain the intended results [16]. Both Figures 6 and 7 show that the framework can provide more relevant papers for academic searchers at the top of the result list. We believe that our proposed approach is effective, and the integration of citation analysis and QE in the scholarly retrieval model can obtain comparatively more relevant papers.

V. CONCLUSION

Scholarly search engines attempt to reduce the manual efforts of researchers when searching for relevant papers. However, the huge expansion and the complexity of the academic web makes it a very hot and challenging area of research. In this paper, a new technique has been presented that uses Query Expansion (QE) via Pseudo Relevance Feedback (PRF) and citation analysis for supporting the scholarly search. The experiments were evaluated by standard evaluation metrics that demonstrate the usefulness of the proposed technique. The framework uses QE via PRF and citation network analysis and is capable of filtering and ranking relevant papers in the retrieval model and can help scholars locate relevant papers in scholarly published documents. The experimental results

illustrate that the framework can produce more nuanced ranking results. PRF provides a way of natural addition to the expansion process and helps, along with the citation network analysis, in mitigating the vocabulary mismatch issues that arise in IR [37], especially the approach that may solve the issue of different terms used for describing the same concepts. There are several possibilities for future work. In the near future, we would like to enrich the QE through user log and KG capabilities. One can also practice the clustering capability of KG for the extraction of interesting terms.

REFERENCES

[1] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," *International Journal on Digital Libraries*, vol. 17, pp. 305–338, 2016, doi: 10.1007/s00799-015-0156-0.

[2] S. Khalid, S. Khusro, I. Ullah, and G. Dawson-Amoah, "On The Current State of Scholarly Retrieval Systems," *Engineering, Technology & Applied Science Research*, vol. 9, no. 1, pp. 3863–3870, Feb. 2019.

[3] C. Carpineto and G. Romano, "A Survey of Automatic Query Expansion in Information Retrieval," *Acm Computing Surveys*, vol. 44, pp. 1–50, Jan. 2012, doi: 10.1145/2071389.2071390.

[4] P. Sharma and N. Joshi, "Knowledge-Based Method for Word Sense Disambiguation by Using Hindi WordNet," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3985–3989, Apr. 2019.

[5] A. Spink, D. Wolfram, J. Jansen, and T. Saracevic, "Searching the Web: The Public and Their Queries," *Journal of the American Society for Information Science and Technology*, vol. 52, pp. 226–234, Feb. 2001, doi: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1591>3.0.CO;2-R.

[6] J. Clement, "Average number of search terms for online search queries in the United States as of January 2020," *Statista*. https://www.statista.com/statistics/269740/number-of-search-terms-in-internet-research-in-the-us/ (accessed Jul. 21, 2020).

[7] S. Khalid, S. Wu, A. Alam, and I. Ullah, "Real-time feedback query expansion technique for supporting scholarly search using citation network analysis," *Journal of Information Science*, Jul. 2019, doi: 10.1177/0165551519863346.

[8] J. L. Ortega, *Academic Search Engines: A Quantitative Outlook*. Oxford, UK: Chandos, 2014.

[9] E. Amolochitis, *Algorithms and Applications for Academic Search, Recommendation and Quantitative Association Rule Mining*. Denmark: River, 2018.

[10] D. Mirylenka, *Towards structured representation of academic search results*. Italy: University of Trento, 2015.

[11] E. Amolochitis, "Algorithms for Academic Search and Recommendation Systems," Ph.D. dissertation, Aalborg University, Denmark, 2014.

[12] M. Kluck and M. Stempfhuber, "Domain-Specific Track CLEF 2005: Overview of Results and Approaches, Remarks on the Assessment Analysis," in *Workshop of the Cross-Language Evaluation Forum for European Languages*, vol. 4022, 2005, pp. 212–221.

[13] M. Kluck, "The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process," in *Workshop of the Cross-Language Evaluation Forum for European Languages*, vol. 3491, 2004, pp. 260–270.

[14] B. Golshan, T. Lappas, and E. Terzi, "Sofia search: a tool for automating related-work search," presented at the ACM SIGMOD International Conference on Management of Data, Scottsdale, Arizona, USA, May 2012, pp. 621–624.

[15] T. Chakraborty and R. Narayanam, "All Fingers are not Equal: Intensity of References in Scientific Articles," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, Nov. 2016, pp. 1348–1358, doi: 10.18653/v1/D16-1142.

[16] R.-L. Liu, "Retrieval of Scholarly Articles with Similar Core Contents," *International Journal of Knowledge Content Development & Technology*, vol. 7, no. 3, pp. 5–27, 2017.

[17] N. Fiorini *et al.*, "Best Match: New relevance search for PubMed," *PLOS Biology*, vol. 16, no. 8, 2018, doi: 10.1371/journal.pbio.2005343, Art no. e2005343.

[18] O. A. Abass, O. Folorunso, and B. O. Samuel, "Automatic Query Expansion for Information Retrieval: A Survey and Problem Definition," *American Journal of Computer Science and Information Engineering*, vol. 4, no. 3, pp. 24–30, 2017.

[19] Y. Lu, H. Fang, and C. Zhai, "An empirical study of gene synonym query expansion in biomedical information retrieval," *Information Retrieval*, vol. 12, no. 1, pp. 51–68, Feb. 2009, doi: 10.1007/s10791-008-9075-7.

[20] L. Milliken, S. Motomarry, and A. Kulkarni, "ARtPM: Article Retrieval for Precision Medicine," *Journal of Biomedical Informatics*, vol. 95, Jun. 2019, doi: 10.1016/j.jbi.2019.103224, Art no. 103224.

[21] M. Dunaiski, G. J. Greene, and B. Fischer, "Exploratory search of academic publication and citation data using interactive tag cloud visualizations," *Scientometrics*, vol. 110, no. 3, pp. 1539–1571, Mar. 2017, doi: 10.1007/s11192-016-2236-3.

[22] M. Hagen, A. Beyer, T. Gollub, K. Komlossy, and B. Stein, "Supporting Scholarly Search with Keyqueries," in *European Conference on Information Retrieval*, vol. 9626, 2016, pp. 507–520.

[23] S. Liu, C. Chen, K. Ding, B. Wang, K. Xu, and Y. Lin, "Literature retrieval based on citation context," *Scientometrics*, vol. 101, no. 2, pp. 1293–1307, Nov. 2014, doi: 10.1007/s11192-014-1233-7.

[24] C. Xiong, R. Power, and J. Callan, "Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding," presented at the 26th International Conference on World Wide Web, Perth, Australia, Apr. 2017, pp. 1271–1279.

[25] A. Di Iorio, R. Giannella, F. Poggi, S. Peroni, and F. Vitali, "Exploring Scholarly Papers Through Citations," presented at the ACM Symposium on Document Engineering, New York,United States, Sep. 2015, pp. 107–116.

[26] J. Sankhavara and P. Majumder, "Biomedical Information Retrieval," in *Fire (Working Notes)*, 2017.

[27] J. Xu and W. B. Croft, "Quary Expansion Using Local and Global Document Analysis," *ACM SIGIR Forum*, vol. 51, no. 2, pp. 168–175, Aug. 2017, doi: 10.1145/3130348.3130364.

[28] B. He, "Rocchio's Formula," in *Encyclopedia of Database Systems*, L. Liu and M. T. Ozsu, Eds. Boston, Massachusetts: Springer, 2009, pp. 2447–2447.

[29] J. Sankhavara, "Biomedical Document Retrieval for Clinical Decision Support System," presented at the ACL Student Research Workshop, Melbourne, Australia, Jul. 2018, pp. 1–7, doi: 10.18653/v1/P18-3012.

[30] C. Lucchese, F. M. Nardini, R. Perego, R. Trani, and R. Venturini, "Efficient and Effective Query Expansion for Web Search," presented at the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, Oct. 2018, pp. 1551–1554.

[31] "relevancy-feedback-plugin," *GitHub*. https://github.com/topics/relevancy-feedback-plugin (accessed Jul. 21, 2020).

[32] Z. A. Shaikh, "Keyword Detection Techniques: A Comprehensive Study," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2590–2594, Feb. 2018.

[33] J. Rocchio, "Relevance feedback in information retrieval," in *The Smart Retrieval System-Experiments in Automatic Document Processing*, Prentice Hall, 1971, pp. 313–323.

[34] T. Grainger and T. Potter, *Solr in Action*. Shelter Island, New York: Manning, 2014.

[35] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The ACL anthology network corpus," *Language Resources and Evaluation*, vol. 47, no. 4, pp. 919–944, Dec. 2013, doi: 10.1007/s10579-012-9211-2.

[36] A. A. Jbara and D. R. Radev, "The ACL Anthology Network Corpus as a Resource for NLP-based Bibliometrics," 2013.

G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The vocabulary problem in human-system communication," *Communications of the ACM*, vol. 30, no. 11, pp. 964–971, Nov. 1987, doi: 10.1145/32206.32212.