Pseudo-Relevance Feedback Combining Statistical and Semantic Term Extraction for Searching Arabic Documents

Maryamah Maryamah[1]   Agus Zainal Arifin[1]*   Riyanarto Sarno[1]
Rarasmaya Indraswari[2]   Rizka Wakhidatus Sholikah[3]

*[1]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology,
Institut Teknologi Sepuluh Nopember (ITS), Indonesia*
*[2]Department of Information Systems, Faculty of Intelligent Electrical and Informatics Technology,
Institut Teknologi Sepuluh Nopember, Indonesia*
*[3]Department of Information Technology, Faculty of Intelligent Electrical and Informatics Technology,
Institut Teknologi Sepuluh Nopember, Indonesia*
* Corresponding author's Email: agusza@if.its.ac.id

Abstract: Pseudo-relevance feedback (PRF) is an effective query expansion method for searching candidate terms based on top-ranked documents. PRF uses a statistical approach that relies heavily on term-based document retrievals. It causes a semantic gap between query and document that causes vocabulary mismatch problems, such as synonyms and polysemy. In this paper, we proposed pseudo-relevance feedback that combines statistical and semantic term extraction using Term Frequency–Inverse Document Frequency (TFIDF), Bidirectional Encoder Representations from Transformers (BERT), and Yet Another Keyword Extraction (YAKE) for searching Arabic documents. Pseudo-relevance feedback will produce top-ranked documents that are relevant to the query. Term extraction on top-ranked documents using TFIDF, BERT, and YAKE was conducted to obtain candidate terms related statistically and semantically, thus understanding the context of the query sentence. The term extraction results will be re-ranked using Borda ranking to ensure that the Top-K candidate terms are relevant and used for searching the final document with the original query. From the experimental results, the proposed method obtains Precision 5 (P5), P10, Recall, Mean Reciprocal Rate (MRR), and Success Rate (SR@K) of 21%, 14%, 42%, 42%, and 58%, respectively. These results indicate that PRF with a combination of statistical and semantic term extraction for searching Arabic documents can improve the relevant document results.

Keywords: Pseudo-relevant feedback, Statistical feature, Semantic feature, Searching arabic documents.

1. Introduction

Information retrieval (IR) related to creating systems that were used to find information or documents [1]. IR is the extraction process of the most relevant documents from large datasets [2]. Queries submitted from user to information retrieval systems usually contain insufficient words [3]. It because the user often has difficulties expressing the desired information into query words [4]. One method that can be used to solve this problem is query expansion. Query expansion (QE) in information retrieval is carried out by adding relevant terms for obtaining relevant documents [5]. Query expansion divided into three, which are statistical (based on documents analysis), semantics (based on the relation of knowledge structures), and hybrid (combination statistical and semantic-based) [6].

Pseudo-relevance feedback (PRF) is one of query expansion based on a statistical approach [7]. The statistical methods highly depend on a term-based document retrieval to retrieve the user interests from a collection of documents [8]. PRF method using top-ranked documents to identify and obtain new candidate of query terms [9]. PRF assumes that the top-ranked (Top-K) documents that have the highest similarity with the query in the first-pass retrieval

process is relevant to the original query [10]. PRF only focuses on statistical features in obtaining Top-K documents, and statistical features are insufficient to solve vocabulary mismatch problems, such as synonyms and polysemy [11].

Vocabulary mismatch problems can occur if the query is well specified but does not match the vocabulary of the collection being searched [12]. One of the solutions to solve this problem by using semantic approach. The method that is often used is BERT [13]. BERT is a method widely used in recent research because the improvement of results in performance. However, previous studies used BERT as document ranking method to replace the original PRF that using TFIDF term weighting [14–16]. The document results consist other contexts that have semantic similarities and ignore documents that are statistically similar. Statistical similar documents will get documents that have the same domain and the original PRF in the initial searching document is more relevant than replacing it with BERT.

Additional semantic terms are the right solution to overcome vocabulary mismatch problem. The additional term can be obtained from term extraction process. Semantic term can produce other context similar to the queries, such as synonyms, hypernyms, hyponyms, etc. These new terms play a role in increasing recall and relevance of documents. Based on these studies, the combination of term extraction that can generate statistical and semantic-based candidate terms, is required.

In this paper, we proposed pseudo-relevance feedback by combining statistical and semantic term extraction for searching Arabic documents. Pseudo-relevance feedback will produce Top-K documents that are relevant to the query. Term extraction from the document using TFIDF, BERT, and YAKE to obtain candidate terms that are statistically and semantically related to the query. The term extraction results will be re-ranked using Borda ranking to ensure the Top-K candidate terms are relevant. The Top-K candidate terms will be used for searching the final document with the original query.

This paper is organized as follows. Section 2 presents the previous works related to this study. Section 3 describes the methodology of the proposed method. The experimental results and analysis are discussed in Section 4. Then the conclusion is presented in Section 5.

2. Related works

There are several previous studies on PRF, which are Neural pseudo-relevance feedback that combines PRF with Language Model [17], Document-based and Term based Linear Methods [18], Neural network architecture [19], Word embedding based [20], Graph-based Transformer (PGT) [21], and Context-aware BERT (Co-BERT) [22]. Other researchers have also suggested that PRF can use other knowledge, such as Wikipedia. The re-ranking of the PRF results is also carried out by other researchers using machine learning-classification methods [23], BERT [14-16], Multi-Stage using History, PRF original, and BERT [24].

Besides re-ranking the PRF document results, the extraction terms from the PRF document results also highly affect the relevant document results. This because the extraction terms from documents will produce candidate terms, which will be used for an additional search of the final document. Several previous studies have also used term selection methods, such as Kernel co-occurrence-based [25], Kullback–Leibler Divergence (KLD), Bose-Einstein1 (Bo1), and Information Gain (IG) [7], co-occurrence, chi-square, KLD, and Robertson selection value (RSV) methods [26], to obtain relevant candidate terms. Term selection is the process of selecting terms that are relevant to the query. Before term selection, provide candidate terms from the term extraction process is necessary to do.

YAKE [27] is an existing technique for terms extraction that outperforms other comparison methods, such as TFIDF, KP-Miner [28], Rake [29], TextRank [30], SingleRank [31], TopicRank [32], TopicalPageRank [33], PositionRank [34], MultipartiteRank [35], and ExpandRank [31]. From 20 datasets of the experiment, YAKE has higher results of 11 datasets compared to other methods. With these excellent results, YAKE is one of the term extractions that must be used to improve performance.

Based on the reviewed studies, some studies focus more on changing the original PRF architecture with other methods, both statistical and semantic approaches. Some studies also re-rank the results of the PRF first pass documents. The proposed method of previous studies aims to obtain documents that have statistical or semantic similarities or both. Meanwhile, the original PRF focuses on the statistical similarity of the document, and the result of the documents has the same domain context with query. Therefore, the original PRF was applied to obtain documents with the same domain context of the query, and statistical and semantic candidate terms were obtained in the next process.

The previous studies were conducted to obtain documents using different PRF architecture that would be used to obtain candidate terms and ignore the term extraction process. Terms extraction plays a significant role in generating candidate terms.

Several methods have been combined based on studies related to term extraction, but some previous studies only considered statistical features to obtain candidate terms. The addition of a term extraction method that considers semantic features is essential to improve the result. YAKE is term extraction with combined statistical and semantic features to calculate relevant terms. However, this study only focuses on combination term extraction based on statistical and semantic to obtain relevant candidate terms.

Some studies choose Top-K candidate terms based on similarity of the statistical term (TFIDF and cosine similarity) after combining term extraction results in the previous studies. This mechanism is biased where the result of term ranking has semantic terms. The ranking process is used to obtain the most appropriate term to be used as a candidate term. This ranking significantly affects the candidate terms selected as the relevant candidate terms and reduces the candidate terms noise. One of the ranking methods that is not biased towards selecting the two features (statistical and semantic) is the rank aggregation method. Therefore, this study proposed pseudo-relevance feedback by combining statistical and semantic term extraction to improve documents relevant for searching Arabic documents.

3. Methodology

In Fig. 1 there are several main processes carried out, which are documents search for candidate terms, term extraction, candidate terms ranking, and final searching of documents. The initial process carried out in this study was pre-processing. Pre-processing used to prepare documents, so there were not too many noise terms (stopwords) in the document. This pre-processing consists of several processes, which are normalization, stopwords removal, and stemming.

This study addresses the cross-lingual information retrieval (CLIR) problem, in which the input query uses Indonesian language while the retrieved document uses Arabic language. Translation using Google translate API were performed to equate the query language with the document before performing a document search. The following process did not need translation anymore because the candidate term search was processed using local documents (Arabic documents). The translation result was stored to use in the final searching of documents.

3.1 Pseudo-relevance feedback (PRF)

Pseudo-relevant feedback (PRF) is a searching process for documents originating from local datasets based on top-ranked documents and is used to search for candidate terms. This searching was carried out by calculating the highest similarity between queries and documents. Top-K documents which have the highest similarity were considered the most relevant documents. PRF performs document search with TFIDF term weighting calculation on Eq. (1) and cosine similarity Eq. (2).

$$TFIDF_{ij} = tf_{i,j} \times \log \frac{N}{Df_i} \qquad (1)$$

$TFIDF_{ij}$ on Eq. (1) used to count the number of words in the document. This method combines term frequency (TF) and inverse document frequency (IDF).$tf_{i,j}$ is the frequency of terms emergence (i) in the document (j). The more words a document emerges, the weight $tf_{i,j}$ will be higher. N is the total of all documents and Df_i is the total document containing term (i).

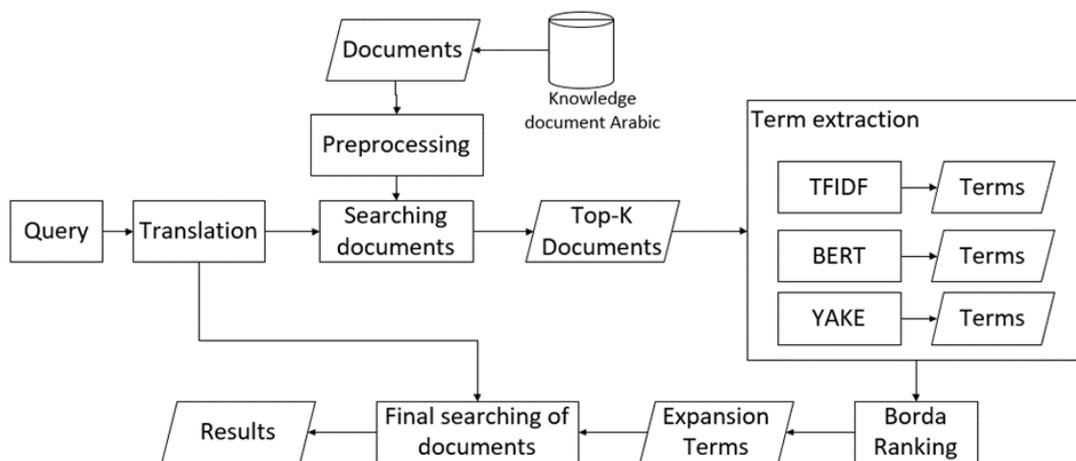$$Cosim_{d,q} = \frac{d.q}{||d|| \times ||q||} \qquad (2)$$

Figure. 1 Proposed method

$Cosim_{d,q}$ on Eq. (2) is the cosine similarity calculation between documents (d) and query (q). $||d||$ is the Euclidean norm of vector $d = \sum_{i=0}^{n} d$ and same value with $d = d_1, d_2, \dots, d_n$ that is defined as $\sqrt{d_1^2 + d_2^2 + \cdots + d_n^2}$. $||q||$ is the Euclidean norm of vector q. After obtaining the first pass document PRF the next process is term extraction. Terms extraction plays a significant role in generating candidate terms. Term extraction is a process to identify important terms based on statistical and semantic proximity to the query. The term extraction process does not consider the position result of first pass PRF documents.

3.2 Term extraction

Term extraction uses to identify the important terms of the PRF document related to the query. This process is significant because taking the optimal term from the PRF results will help to improve the search results for relevant documents. In this study, term extraction was carried out by combining statistical and semantic term extraction methods on the documents. Several methods were used in this study according to the desired statistical and semantic term relations. The term extraction is a combined method of TFIDFBERTYAKE.

TFIDF is a method that depends on the number of terms mentioned in the document (statistical), BERT considers the semantic of words according to the context in the sentence, and YAKE uses the five feature statistical term scores [27]. In the original paper experiment, BERT can understand the context of the word better than the previous methods (Embeddings from Language Models (ELMo) [36], OpenAI Generative Pre-trained Transformer (GPT) [37], and Global Vector (GloVe) [38]) because the input from BERT uses the whole sentence (left and right of the words). Previous research used sentence understanding based on co-occurrence sequential (left-to-right or right-to-left). The input of BERT is a sentence that would be processed into three embeddings, which are token embedding, segment embedding, and position embedding. Token embedding is the term in the sentence that added with [CLS] at the beginning of the sentence and [SEP] at the end of the sentence. The segment embedding is the term that belongs to the part or segment of sentence. For example, the first segment sentence is segment A, then all terms in the first sentence are marked with A, all the terms in the second segment are marked B, and so on. Position embedding is the position of the term in the sentence. This position is calculated from the term in the first to the last sentence in all documents. The prefix [CLS] and suffix [SEP] terms in each sentence calculate as one term.

$$YAKE\ (t) = \frac{T_{Related} \times T_{Position}}{T_{Case} + \frac{TF_{Normalized}}{T_{Related}} + \frac{T_{Sentence}}{T_{Related}}} \quad (3)$$

YAKE Eq. (3) calculates the term score based on five features: term relatedness ($T_{Related}$), term position ($T_{Position}$), term casing (T_{Case}), normalized term frequency ($TF_{Normalized}$), and term sentences ($T_{Sentence}$). Term relatedness calculates the co-occurrence of left and right related terms. The casing term assumes that the uppercase term is more relevant. The uppercase term is given higher weight than the lowercase term. Term position is considering the position of terms in the document. The position of the term at the beginning of the sentence will have a higher weight. Normalized term frequency calculates the number of terms in a document divided by the mean of frequency to avoid bias on long documents. Term sentences calculate the number of terms emerging in other sentences in the document. In document ranking, term frequency is the most important method [39]. The combination of the TFIDFBERTYAKE method will get the result of the candidate term based on statistical and semantic relevance to the query.

3.3 Borda ranking

Borda ranking is the rank aggregation method used to calculate the ranking of the candidate selection list [40]. Borda ranking calculates the ranking of terms based on the highest position [7]. Borda ranking Eq. (4) calculates rank from the highest term (t_i) position from the previous term extraction results (TFIDF, BERT, and YAKE). i is the index number of candidate terms, and the index number begins from 0. For example, the first term from the list of candidate terms has a value $i=0$. n is maximum index number of candidate terms. For example, total terms from the list of candidate terms are 100, then $n=99$. The term that has a higher position in those three term extraction methods will have a high ranking result.

$$Borda(t_i) = \max\left(\sum_{i=0}^{n}\left(TFIDF_{t_i}, BERT_{t_i}, YAKE_{t_i}\right)\right) \quad (4)$$

3.4 Final searching of documents

Final searching of documents was carried out using the same procedure in pseudo-relevant feedback, which are using TFIDF as term weighting

method and cosine similarity. This document searching was carried out using the original query and the resulting candidate terms as additional queries. The results of document search will be provided to the user. This search uses TFIDF term weighting because the document that wants to be provided to the user is a document with the same context as the query.

4. Result and discussion

In this study, the data used for the experiment is Question-Answer (QA) data obtained by scraping on the website piss-ktb.com. The total dataset of the experiment is 55 questions and 200 answer references. The language used in the question is Indonesian, and the answer reference is in the Arabic language. The dataset answer is a document reference source used to answer questions, and one question can consist of more than one document. The proposed method is not limited to the dataset Indonesian-Arabic. Other languages can use the proposed method by changing the preprocessing according to the language to be used. The problem domain of this study is the social-religious problem that is expected to help Indonesian citizens who are majority Muslim. Muslims need systems for searching for comprehensive documents reference.

There are five evaluation metrics used, namely Precision of 5 documents (P5) and 10 documents (P10), Recall, Mean Rate Reciprocal (MRR), Success Rate (SR@K). Precision (P) is calculating the total of relevant documents from system retrieve. Precision Eq. (5) calculates the total number of relevant documents retrieved by the system ($Documents_r$) divided by the total number of documents returned ($Total\ Documets_i$). The recall is calculating the total of relevant documents from all relevant collections. Recall Eq. (6) calculate from the total number of relevant documents retrieved by the system ($Documents_r$) divided by the total number of relevant documents in the dataset ($Total\ Documents_r$). MRR is the calculation of the highest-ranking position in the relevant document returned by the system. MRR Eq. (7) calculates the rank document ($rank_i$) of each question (q) by one divided with rank documents. This result is 0 when there is no relevant documents retrieve by systems. SR@K is the success rate calculation, where the system has successfully returned at least one document [41].

$$Precision = \frac{Documents_r}{Total\ Documets_i} \qquad (5)$$

$$Recall = \frac{Documents_{r,q}}{Total\ Documents_r} \qquad (6)$$

$$MRR_q = \frac{1}{|q|} \sum_{i=1}^{|q|} \frac{1}{rank_i} \qquad (7)$$

This study conducted three experiments to test the performance of the proposed method. The first experiment is the method comparison based on term extraction used without using Borda ranking: TFIDF, YAKE, BERT, BERTTFIDF, YAKETFIDF, and the proposed method. The second experiment re-ranking candidate terms using Borda ranking on term extraction BERTTFIDF, YAKETFIDF, and the proposed method. It uses to determine the effect of the re-ranking using Borda ranking on the result of term extraction. The third experiment is the optimal total candidate terms used for the final searching of documents.

Table 1 shows the experiment of several term extraction methods used without using Borda ranking. The results show that the proposed method has the four highest results (P5, P10, Recall, and MRR) of the five evaluation metrics used. It shows that the proposed method has more relevant document search results than the other compared methods. In comparing several term extractions, TFIDF has the one highest evaluation metrics SR@K compared to other comparison methods (except proposed methods). Candidate terms using TFIDF are selected based on the number of terms mentioned in the documents. The terms frequency in the documents proves that the more terms mentioned in a document, the more relevant the terms are. However, the frequency of these terms is very biased when the stopword does not implement. The list word of stopword is also a word which mentioned a lot in the document, but the word is not relevant. Therefore, the stopwords removal process is essential to use before TFIDF term extraction.

The highest SR@K results in this study are TFIDF, YAKE, and YAKETFIDF. These results prove that these methods (TFIDF, YAKE, and YAKETFIDF) can return more relevant documents for the queries used in the experiment. YAKE [28] is

Table 1. Comparison result

Method	P5	P10	Recall	MRR	SR@K
TFIDF	19%	13%	41%	41%	**62%**
YAKE [28]	21%	13%	40%	35%	**62%**
BERT [13]	17%	13%	41%	33%	60%
YAKETFIDF	20%	13%	40%	36%	**62%**
BERTTFIDF	19%	12%	36%	36%	53%
Reranking BERT	13%	10%	34%	28%	58%
Proposed Method	**21%**	**14%**	**42%**	**42%**	58%

one of the methods that has highest SR@K, which means YAKE can return more relevant documents compared with other method. YAKE also has the same value of P5 as the proposed method. YAKE can retrieve a more relevant document with five documents retrieved by systems compared with other comparisons (except the proposed method).

YAKE extracting term using five statistical features: term relatedness, term position, term casing, normalized term frequency, and term sentences. In all experimental cases, YAKE can get more relevant documents. However, YAKE has lower results on the evaluation of other metrics (P10, Recall, and MRR) than the proposed method. In some cases, YAKE can only return fewer total relevant documents than the proposed method. Term extraction using YAKE focuses on terms that are statistical and semantically similar based on word co-occurrence. The word co-occurrence term does not pay attention to the position of the word according to the context of the word in the sentence. The same word in a different sentence may have a different meaning according to the context of the word in the sentence. BERT has overcome this problem, and using BERT is very important for semantic understanding based on the context of words in sentences. However, in single term extraction, TFIDF and YAKE have better results than BERT. BERT alone is insufficient to provide candidates of relevant terms.

BERT [13] is a method that focuses on extracting terms that have a similar semantic context to the query. In this case, BERT ignores statistical features of documents. The statistical feature is very important to get context with the same domain as query, whereas semantics get context with similar meaning with query. This similarity of semantic context has the result of term extraction being a term with another domain. Combination with statistical terms such as TFIDF or YAKE can improve BERT's performance. The combination of BERT with other methods (BERTTFIDF) obtains a relatively high result compared to other methods. Single term extraction using BERT is not recommended because term extraction focuses only on semantic features.

Comparison of other BERT methods is re-ranking BERT which is almost the same in previous studies [14]. The initial stage of the method is to search for the initial document with the original PRF using BM25. Then, the results of the BM25 document are re-ranked using BERT, and term extraction on the document results using TFIDF is implemented to obtain candidate terms. Finally, the results of the candidate terms and the original query are used as the final document search. The results from BERT's ranking were 13%, 10%, 34%, 28%, 58% using P5,

P10, Recall, MRR, and SR@K, respectively. These results indicate that re-ranking using BERT documents has the lowest results compared to all methods compared.

PRF in the first pass using BERT for the document ranking have irrelevant results. The document results consist other contexts that have semantic similarities. For example, when the query context is "*shalat*", and the results of the document searching using BERT are "*tayamum*" or tayammum, "*wudhu*" or ablution, and other contexts. The document results have high context semantic similarity, but this causes documents with the same context (about "*shalat*") not being returned. Top-K documents result on the first pass using BERT are not used. It proves that BERT can understand the semantic of words well, however it is not suitable to be used for searching documents in the same domain. BERT must be combined with statistical methods to obtain optimal results.

Re-ranked the data is unnecessary because the process after obtaining the first pass document PRF is term extraction. Terms extraction plays a significant role in generating candidate terms. The term extraction process does not consider the position result of first pass PRF, and the re-ranking process is not necessary. Term extraction is a process to identify important terms based on statistical and semantic proximity to the query.

Table 2 shows the impact of Borda ranking on term extraction. This experiment shows that BERTTFIDF has an increase in P5 and MRR. It shows that Borda ranking can increase the relevant documents ranking higher in the results compared with without re-ranking using Borda ranking. The YAKETFIDF method also increases the MRR value. The increased of MRR value is a proof that adding Borda ranking after term extraction can improve the relevant document ranking results on results. However, the results of Recall and SR@K decrease on YAKETFIDF because the system failed to return the relevant documents in two cases that were previously successfully returned without Borda ranking.

The final searching document used all of the Top-K terms as candidate terms from the term extraction

Table 2. Borda ranking results

Method	P5	P10	Recall	MRR	SR@K
BERTTFIDF	21%	13%	38%	38%	56%
YAKETFIDF	21%	13%	39%	39%	**58%**
Proposed Method	21%	**14%**	**42%**	**42%**	**58%**

Table 3. Comparison Top-K QE terms

Method	P5	P10	Recall	MRR	SR@K
10 terms	21%	**14%**	**42%**	**42%**	58%
20 terms	21%	13%	39%	40%	**60%**
30 terms	21%	13%	38%	36%	58%

results without Borda ranking. By using Borda ranking, there are only 20 terms with the highest-ranking used as candidate terms. Terms that have a low ranking in the result of Borda ranking are failed two cases from all total experiments case compared with other methods that have the higher result of SR@K (TFIDF, YAKE, and YAKETFIDF). The usage of all candidate terms from the term extraction results also affects the results, and it is important to know the optimal number of the used terms for the final searching of documents. More candidate terms total will lead to longer execution time needed for the document search. Therefore, an experiment to determine the optimal Top-K candidate terms needs to be done.

Table 3 shows an experiment on the Top-K terms uses for final searching of documents. This experiment used top 10, 20, and 30 terms from the term extraction results. The experiment result is that using top-10 terms gives better performance than using top-20 and top 30-terms. It shows that more term does not improve system performance, but on the contrary, it decreases the system performance. The ten terms for total candidate terms are the most optimal terms used for the QE candidate terms. The candidate terms result used for the final searching of documents.

The semantic combination used in this study is term extraction using BERT to produce semantic candidate terms. The combination of statistical and semantic terms shows an increase in the performance of searching relevant documents. This result can improve with combination of the semantic terms come from other reliable external knowledge such as Wikipedia. The addition of this external knowledge can be done with other hybrid QE methods to process the knowledge and obtain relevant candidate terms. For future works, hybrid QE semantic-based by using other knowledge to improve relevant documents can be implemented. Another suggestion of future works is to combine more term extraction with different statistical approaches.

5. Conclusion

In this paper, we proposed pseudo-relevance feedback that combines statistical and semantics term extractions for searching Arabic Documents. Term extractions use TFIDFBERTYAKE to extract the statistical and semantics features of the terms. Borda ranking uses to ensure the candidate terms have the highest ranking from all the term extraction result and are relevant to the query. The proposed method obtains the highest value of 4 metrics out of the five evaluation metrics used in this experiment. The results of the proposed methods are 21%, 14%, 42%, 42%, and 58% using P5, P10, Recall, MRR, and SR@K, respectively.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Maryamah contributed to the formulation of the proposed method, the architecture of the proposed method, making information retrieval systems, experiment results, and analysis, also writing draft preparation. Agus Zainal Arifin supervised the conceptualization, problem reformulation, and review the draft preparation. Riyanarto Sarno supervised the conceptualization, problem reformulation, and review the draft preparation. Rarasmaya Indraswari and Rizka Wakhidatus Sholikah review and revise the paper draft.

Acknowledgments

References

[1] B. Hjørland, "Information retrieval and knowledge organization: A perspective from the philosophy of science", *Inf.*, Vol. 12, No. 3, p. 135, 2021.

[2] D. Virmani and S. Taneja, "A Text Preprocessing Approach for Efficacious Information Retrieval", *Adv. Intell. Syst. Comput.*, Vol. 669, pp. 13-22, 2019.

[3] B. Xu, H. Lin, Y. Lin, L. Yang, and K. Xu, "Improving pseudo-relevance feedback with neural network-based word representations", *IEEE Access*, Vol. 6, pp. 62152-62165, 2018.

[4] M. Maryamah, A. Z. Arifin, R. Sarno, and Y. Morimoto, "Query expansion based on Wikipedia word embedding and BabelNet method for searching Arabic documents", *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 5, pp. 202-213, 2019.

[5] S. Jain, K. R. Seeja, and R. Jindal, "A fuzzy ontology framework in information retrieval

using semantic query expansion", *Int. J. Inf. Manag. Data Insights*, Vol. 1, No. 1, p. 100009, 2021.

[6] M. A. Raza, R. Mokhtar, and N. Ahmad, "A survey of statistical approaches for query expansion", *Knowl. Inf. Syst.*, Vol. 61, No. 1, pp. 1-25, 2019.

[7] I. Rasheed, H. Banka, and H. M. Khan, "Pseudo-relevance feedback based query expansion using boosting algorithm", *Artif. Intell. Rev.*, No. 0123456789, 2021.

[8] H. Almarwi, M. Ghurab, and I. A. Baltah, "A hybrid semantic query expansion approach for Arabic information retrieval", *J. Big Data*, Vol. 7, No. 1, pp. 1-19, 2020.

[9] C. Li, Y. Sun, B. He, L. Wang, K. Hui, A. Yates, L. Sun, and J. Xu, "NPRF: A neural pseudo relevance feedback framework for ad-hoc information retrieval", In: *Proc. of 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 4482-4491, 2020.

[10] A. Htait, S. Fournier, P. Bellot, L. Azzopardi, and G. Pasi, "Using Sentiment Analysis for Pseudo-Relevance Feedback in Social Book Search", *ICTIR 2020 - Proc. 2020 ACM SIGIR Int. Conf. Theory Inf. Retr.*, pp. 29-32, 2020.

[11] J. Abdul and I. Varlamis, "A knowledge-based semantic framework for query expansion", *Inf. Process. Manag.*, Vol. 56, No. 5, pp. 1605-1617, 2019.

[12] A. Montazeralghaem, H. Zamani, and J. Allan, "A Reinforcement Learning Framework for Relevance Feedback", In: *Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 59-68, 2020.

[13] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 2019.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv Prepr. arXiv1810.04805*, 2018.

[15] R. Padaki, Z. Dai, and J. Callan, "Rethinking query expansion for BERT reranking", In: *Proc. of European Conference on Information Retrieval*, pp. 297-304, 2020.

[16] Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and A. Yates, "BERT-QE: Contextualized Query Expansion for Document Re-ranking", *arXiv Prepr. arXiv2009.07258*, pp. 4718-4728, 2020.

[17] Y. Lv and C. X. Zhai, "Revisiting the divergence minimization feedback model", *CIKM 2014 - Proc. 2014 ACM Int. Conf. Inf. Knowl. Manag.*, pp. 1863-1866, 2014.

[18] D. Valcarce, J. Parapar, and Á. Barreiro, "Document-based and Term-based Linear Methods for Pseudo-Relevance Feedback", *ACM SIGAPP Appl. Comput. Rev.*, Vol. 18, No. 4, pp. 5-17, 2018.

[19] L. Wang, Z. Luo, C. Li, and B. He, "An end-to-end pseudo relevance feedback framework for neural document retrieval", *Inf. Process. Manag.*, Vol. 57, No. 2, p. 102182, 2020.

[20] A. El Mahdaouy, S. O. El Alaoui, and E. Gaussier, "Word-embedding-based pseudo-relevance feedback for Arabic information retrieval", *J. Inf. Sci.*, Vol. 45, No. 4, pp. 429-442, 2019, doi: 10.1177/0165551518792210.

[21] H. Yu, Z. Dai, and J. Callan, "PGT: Pseudo Relevance Feedback Using a Graph-Based Transformer", *arXiv Prepr. arXiv2101.07918*, pp. 440-447, 2021.

[22] X. Chen, K. Hui, B. He, X. Han, L. Sun, and Z. Ye, "Co-BERT: A Context-Aware BERT Retrieval Model Incorporating Local and Query-specific Context", *arXiv Prepr. arXiv2104.08523*, 2021.

[23] J. Lin, "The Simplest Thing That Can Possibly Work : Pseudo-Relevance Feedback Using Text Classification", *arXiv Prepr. arXiv1904.08861*, 2019.

[24] H. Al-Thani, B. J. Jansen, and T. Elsayed, "HBKU at TREC 2020: Conversational Multi-Stage Retrieval with Pseudo-Relevance Feedback", 2020.

[25] M. Pan, J. X. Huang, T. He, Z. Mao, Z. Ying, and X. Tu, "A simple kernel co-occurrence-based enhancement for pseudo-relevance feedback", *J. Assoc. Inf. Sci. Technol.*, Vol. 71, No. 3, pp. 264-281, 2020, doi: 10.1002/asi.24241.

[26] J. Singh, M. Prasad, Y. A. Daraghmi, P. Tiwari, P. Yadav, N. Bharill, M. Pratama, and A. Saxena, "Fuzzy LogicHybrid model with semantic filtering approach for pseudo relevance feedback-based query expansion", in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2018, pp. 1-7, doi: 10.1109/SSCI.2017.8280930.

[27] R. Campos, V. Mangaravite, A. Pasquali, and A. Jorge, "YAKE ! Keyword extraction from single documents using multiple local features", *Inf. Sci. (Ny).*, Vol. 509, pp. 257-289, 2020, doi: 10.1016/j.ins.2019.09.013.

[28] S. R. El-Beltagy and A. Rafea, "KP-Miner: A keyphrase extraction system for English and Arabic documents", *Inf. Syst.*, Vol. 34, No. 1, pp.

132-144, 2009, doi: 10.1016/j.is.2008.05.002.

[29] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents", *Text Min. Appl. Theory*, pp. 1-277, 2010.

[30] R. Mihalcea and P. Tarau, "TextRank: bringing order into texts", *Proc. 2004 Conf. Empir. methods Nat. Lang. Process.*, pp. 404-411, 2004, doi: 10.1109/ICECTA48151.2019.8959702.

[31] X. Wan and J. Xiao, "Single document keyphrase extraction using neighborhood knowledge", *Proc. Natl. Conf. Artif. Intell.*, Vol. 2, pp. 855-860, 2008.

[32] A. Bougouin, F. Boudin, and B. Daille, "TopicRank : Graph-Based Topic Ranking for Keyphrase Extraction", in *International Joint Conference on Natural Language Processing (IJCNLP)*, 2013, pp. 543-551.

[33] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Topical word importance for fast keyphrase extraction", *WWW 2015 Companion - Proc. 24th Int. Conf. World Wide Web*, No. 2, pp. 121-122, 2015, doi: 10.1145/2740908.2742730.

[34] C. Florescu and C. Caragea, "PositionRank: An unsupervised approach to keyphrase extraction from scholarly documents", *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, Vol. 1, pp. 1105-1115, 2017, doi: 10.18653/v1/P17-1102.

[35] F. Boudin, "Unsupervised keyphrase extraction with multipartite graphs", *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, Vol. 2, pp. 667-672, 2018, doi: 10.18653/v1/n18-2105.

[36] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations", *NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, Vol. 1, pp. 2227-2237, 2018, doi: 10.18653/v1/n18-1202.

[37] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training.", 2018, doi: 10.1007/978-3-319-91280-6_300915.

[38] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation", in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532-1543, doi: 10.1080/02688697.2017.1354122.

[39] S. Na and K. Kim, "Verbosity normalized pseudo-relevance feedback in information retrieval", *Inf. Process. Manag.*, Vol. 54, No. 2, pp. 219-239, 2018, doi: 10.1016/j.ipm.2017.09.006.

[40] J. Fraenkel and B. Grofman, "The Borda Count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia," *Aust. J. Polit. Sci.*, Vol. 49, No. 2, pp. 186-205, 2014, doi: 10.1080/10361146.2014.900530.

[41] W. Sakata, T. Shibata, R. Tanaka, and S. Kurohashi, "FAQ retrieval using query-question similarity and BERT-based query-answer relevance.", in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, pp. 1113-1116.