IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 February 2023, accepted 1 April 2023, date of publication 10 April 2023, date of current version 19 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3266004

RESEARCH ARTICLE

Proximity-Aware Clinical Passage Retrieval Framework by Exploiting Knowledge Structure

KEEJUN HAN[1] AND JUNGEUN KIM[2]
[1]School of Computer Engineering, Hansung University, Seoul 02876, Republic of Korea
[2]Department of Software and CMPSI, Kongju National University (KNU), Cheonan 31080, Republic of Korea

Corresponding author: Jungeun Kim (jekim@kongju.ac.kr)

This work was supported by Hansung University under Grant 2022091393.

ABSTRACT Clinicians have minimal time to search for and absorb the information needed while performing the duties of their medical practice. Their time-pressured situations requires the relevant information in search to be retrieved and presented in a more succinct form, such as in a short passage, rather than a whole page or document. In this context, clinical decision support (CDS) searches are beneficial when used to retrieve critical medical passages that can assist the practice of medical experts by offering appropriate medical information relevant to the medical case at hand. We present a novel CDS search framework designed for passage retrieval in order to support clinical decision-making using laboratory test results by incorporating proximity information. To do so, we use a knowledge structure that graphically visualizes key concepts and the corresponding relationships in a specific domain where nodes denote with their associative relationships. Furthermore, unlike previous studies that exploit knowledge structures during a re-ranking step, only dealing with initially highly retrieved passages, we utilize a knowledge structure for the purpose of query expansion. By doing so, our approach can unveil passages that are not retrieved during the initial retrieval process by including latent terms in the query list. We compared two models with/without edge pruning to capture a more latent relationship between terms. Our experiment showed that the embedded-based knowledge structure outperformed previous knowledge structure building approaches and other proximity-aware state-of-the-art models.

INDEX TERMS CDS search, clinical decision support, knowledge structure, passage retrieval.

I. INTRODUCTION

Clinicians have minimal time to search for and absorb the information they need while performing medical practice [1], [2]. Such a time-pressured situation requires the relevant information in search to be retrieved and presented in a highly succinct form, such as in short passages, rather than whole pages or documents. However, most existing CDS search studies [3], [4], [5] focus on improving the performance of retrieving full-length relevant documents.

To develop a potential passage retrieval strategy for CDS, considering and incorporating the concept of term proximity, which is a distance information between terms, is currently considered to be an effective probabilistic retrieval model [6]. Given a query (a set of terms users input), one fundamental underlying assumption with regard to proximity is that the more compact the query terms, the more likely they are to be closely related. Accordingly the more potentially relevant the documents will be to the topic represented in the particular set of user queries. Further improvement can be realized for passage retrieval by incorporating proximity information as it imposes a proximity constraint on the matched query terms within a passage level.

However, previous probabilistic proximity studies do not fully reflect the genuine relationships between the terms. For instance, the current best proximity measure is *MinDist*, reported in [7], referring to the smallest distance of all pairs of unique matched query terms. We can consider the following two terms as an example: one that occurs at the end of a paragraph, and the other that occurs at the beginning of the next paragraph. The *MinDist* of the two terms is 1; thus, the relationship is considered to be significant but because

The associate editor coordinating the review of this manuscript and approving it for publication was Weiping Ding.

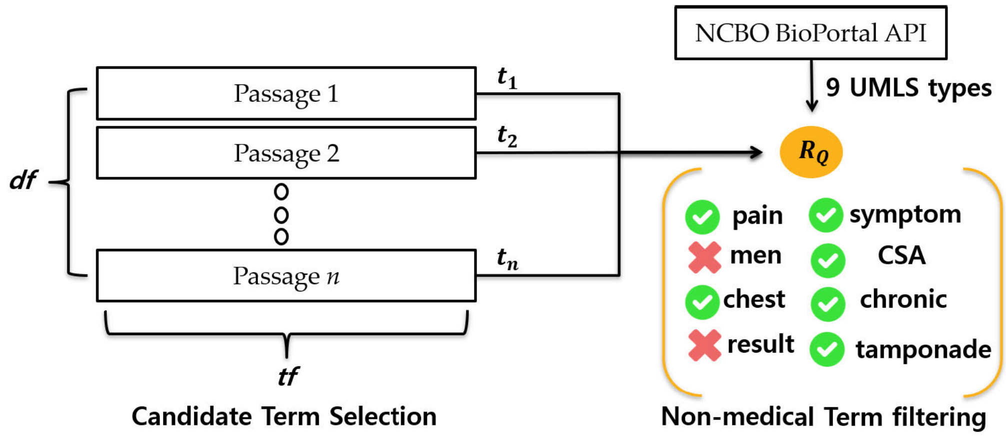


FIGURE 1. Proposed proximity-aware query expansion framework by using knowledge structure.

a paragraph is semantically segmented, it may be that the two terms are not actually cohesive.

To offset the aforementioned limitation, we propose a new proximity approach that exploits what is termed a knowledge structure initially conceptualized in the field of educational psychology [8]. A knowledge structure is different from declarative knowledge [9] or procedural knowledge, and it plays a principal role in converting knowledge into tangible products through an externalization process [10]. The knowledge structure of the learner becomes more elaborate in the process of learning. In this context, knowledge acquisition indicates learning about concepts and defining relationships between such elements. Leaning can be seen as a consecutive interacting process of creating and changing the knowledge structure of a specific domain. Figure 1 illustrates the clinical term selection process, which is then used for building a clinical-specific knowledge structure.

To this end, we develop and propose a framework of passage retrieval in support of clinical decision making, which expands the query by analyzing a knowledge structure based on the passages most commonly retrieved by the initial retrieval. The knowledge structure is constructed by utilizing a set of concepts and distance scores from these initially 'top-retrieved' passages.

In summary, the contributions of this paper are as follows:

- We formalize the process of constructing a knowledge structure from multiple passages.
- We develop a proximity-aware query expansion framework supporting clinical decision-making by investigating proximity information from the knowledge structure.
- We show that the proposed approach outperforms other CDS approaches by comparing it with several query expansion techniques and proximity probability retrieval models.

The article is organized as follows. The Related Work section provides a background of knowledge structure and its applications for information retrieval. The Methods section presents an overview and details of the proposed model. The Results section presents the experiment results, and Concluding Remarks concludes the paper.

II. RELATED WORK

In this section, we introduce some of the existing studies regarding knowledge structure. By exploiting various computerized methods, knowledge structure can be used as a tool to represent a certain domain's important concepts and their relations.

A. KNOWLEDGE STRUCTURE

Goldsmith et al. [8] presented an approach to extract the knowledge structures of individuals. It examined individuals' knowledge structures using three steps: knowledge elicitation, knowledge representation, and an evaluation. During the elicitation step, an individual was asked about relatedness between the concepts selected to represent the domain. In the representation step, scale scores were measured to express a network graph of concepts, pruned by a pathfinder algorithm [11]. This pruning process relieved noise in the proximity matrix. Finally, during the evaluation process, the derived knowledge structures were used as a gold standard, known as the expert's knowledge structure.

Several studies [12], [13] formalized the representation of the knowledge structure of a specific domain in a learning situation. In one study [14], participants were able to construct a more precise domain knowledge structure during the learning period. Acton et al. [15] compared knowledge structures created by an instructor, other experts, intermediate experts, and beginners. In the domain of an intermediate-level programming course, their experiment results showed that 1) the instructor-based knowledge structure was not significantly better than those of other experts, 2) There were variances between knowledge structures by the experts, and 3) knowledge structures derived from intermediate experts and beginners were still somewhat valid, though those of the experts were more valid with regard to the gold standard.

Naveh-Benjamin et al. [12] proposed a new method by which to infer students' cognitive structures. The method involves the design of a new form of knowledge structure that strengthens significant relationships between concepts derived from the memory of a student towards domain-specific knowledge. By conducting a rigorous study involving 255 participants, the results showed that the level of proficiency is strongly correlated with the level of achievement for the domain. They also revealed that four measures - grouping, the hierarchical structure, the directionality of the structure, and similarity to the instructor - were related to academic achievement. McLaughin et al. [16] also revealed that the number of expert-type concepts in the knowledge structure is associated with increased odds of diagnostic success. A scheme used by small-group preceptors is associated with an increased number of expert-typed concepts in the knowledge structure.

Lastly, Samiee and Chabowski [17] examined the underlying forces that shaped the international marketing (IM) field using three methods: explanatory factor analysis (EFA), hierarchical cluster analysis (HCA), and multidimensional metric scaling (MDS). They applied these techniques to evaluate the knowledge structure of IM-related publications for the period of 1999-2008 and to concurrently provide a supplemental examination of the findings for the period of 2009-2010.

The knowledge structures of domain experts regarding a specific domain are known to be similar to one another [8] implying that knowledge can be alternatively extracted from a document pertaining to the correct domain. By representing each paragraph as a knowledge structure that preserves the proximity information among the terms, the relevant passages can be reached more easily using a descriptive set of queries.

Unlike theoretical studies of knowledge structure, there has been little effort to create a computerized knowledge structure. An initial effort was made by Kim and Yi [18], in which detects key concepts and their inter-relationships were detected using a bag-of-words approach and the co-occurrence score, respectively. Once the initial proximity matrix was constructed, weak relationships were removed through a pathfinder algorithm [11]. Their experiment showed that an automatically generated knowledge structure was very similar to a knowledge structure manually created by domain experts. Although their method of creating a knowledge structure was intuitive, only a single document was used. However, in the context of the CDS search, multiple passages should emerge and form a unified knowledge structure.

Bae et al. [19] further applied the automatic knowledge-structure extraction method to develop a mobile application recommendation model. This model collected app usage history metrics, such as app installation and usage times. Based on their data, they constructed a knowledge structure of a mobile application, consisting of nodes of a mobile application and the corresponding relationships defined as the number of co-uses per session. They experimentally demonstrated that the model outperformed existing mobile application recommendation algorithms. Kim et al. [6] also exploited a knowledge structure for movie retrieval. They highlighted how the query intent in the movie retrieval domain is more narrative, implying that proximity information is a significant indicator of differentiating movies for users. Once the knowledge structure for each movie was constructed, proximity scores between query terms of a candidate movie's knowledge structures were measured to re-rank the initial search result. Furthermore, variants of graph-based approach, which can be adapted for building variant forms of knowledge structures, for retrieving documents have been introduced [20], [21], [22], [23]. Although these studies [6], [19] showed promising results in their efforts to exploit knowledge structures, to the best of our knowledge, no current study utilizes a knowledge structure in a CDS search.

Therefore, in this paper, we exploit the knowledge structure for the purpose of query expansion based on the assumption that such a method can outperform existing query expansion techniques, such as Pseudo Relevance Feedback (PRF), as it can consider the importance of terms and the proximity information between the given query and candidate terms. Especially during the passage retrieval process, proximity information becomes more noticeable as the context of the paragraph becomes more cohesive.

B. CLINICAL DECISION SUPPORT SEARCH

In recent years, a significant body of studies has been conducted on CDS search with regard to various resources on the Web since traditional information retrieval techniques cannot be directly applied to biomedical information retrieval due to domain characteristics of biomedical literature: A longer length of a query with narrative structure [30]. The mainstream of current biomedical literature retrieval consists of three modules: Concept extraction, query expansion, and retrieval. Table 1 summarizes those studies which are broken into each module.

In the early studies [24], [25], concept extraction has not been actively conducted. Rather those focused on query expansion and retrieval techniques. Although some of the studies achieved a significant improvement without concept detection, it is clearly noted that those approaches are easily exposed to query drifting as the number of terms for expansion is increased, potentially causing unexpectedly worse search effectiveness since non-medical terms are likely to be expanded as well [25], [26].

For concept extraction, manual efforts by doctor [29], [34] to choose medical terms with the Unified Medical Language Systems (UMLS) Metathesuarus concepts showed a promising performance on retrieval. Thanks to the National Library of Medicine (NLM), MetaMap [35] was then developed to automatically map biomedical text to the Unified Medical Language Systems (UMLS) Metathesuarus concepts. Currently, MetaMap is actively employed

TABLE 1. Literature summary for CDS search.

Name	Concept extraction	Query expansion	Retrieval model
Yu et al. [24]	-	Explicit user feedback	MySQL
Abdou and Savoy [25]	-	PRF	Language Model (LM)
Cohen et al. [26]	-	PRF	Lucene[1]
Zhang et al. [3]	-	-	A variant of TF-IDF and BM25
Liu et al. [4]	DBPedia[2]	Google	Terrior[3]
Jo et al. [5]	DBPedia and UMLS	PRF	Solr[4]
Wang and Fang [27]	cTAKES	MediLexicon[5]	Indri[6]
Gurulingappa et al. [28]	MetaMap	UMLS, PRF	Learning to Rank
You et al. [29]	BioPortal	MeSH	Indri
Soldaini et al. [30]	MetaMap	PRF	Lucene
Palotti and Hanbury [31]	MetaMap	PRF	Terrior
D'hondt et al. [32]	MeSH	PRF	Solr
Chen et al. [33]	MetaMap	-	BM25
Hersh et al. [34]	UMLS	Manual Expansion with UMLS	TF-IDF

for concept detection in most various biomedical tasks [26], [28], [30], [33]. One of the characteristics of MetaMap is that it generates variants of each phrase. The biomedical literature texts are mapped to clinical concepts with UMLS identifier tags, and additionally, their synonyms, acronyms, abbreviations, and deviational variants are found.

Among techniques to enhance the performance of biomedical literature retrieval, query expansion is one of the most common techniques. The goal of query expansion is finding and adding synonyms, and other related terms to increase recall of relevant documents [36]. The expansion techniques adopted in biomedical IR so far are mainly classified into two categories: the use of external resources and the pseudo-relevance feedback technique.

As external recourses, UMLS [28], [34], Google [4], MediLexicon [27], and MeSH [29] are examples of lexical knowledge resources and actively adopted in finding relevant terms. In acquiring implicit relevance feedback, Pseudo-relevance feedback (PRF) is the fundamental technique of query expansion, usually done with initial expansion using external resources. It is well known to find potentially relevant terms by first querying the index and looking for new relevant terms from high-ranked documents [36].

Nonetheless, to the best of our knowledge, those approaches yet mainly have focused on finding the best combination among the applications from concept extraction to query expansion, in order to acquire retrieval effectiveness. Also, most of those adopted either variants of traditional retrieval model [3], [25], [33], [34] or open source retrieval applications such as Lucene, Indri, Solr, or Terrior. Gurulingappa et al. [28] attempted a semi-supervised approach to re-rank the initially retrieved document by exploiting Learning-to-rank and achieved a statistically significant improvement; however, it challenges system feasibility as it requires much computation time when data keep increases.

III. METHODS

In this chapter, first we briefly introduces the passage-retrieval framework for clinical literature. Then, we present details of the individual steps of the model.

A. OVERVIEW

Retrieving passages imposes a proximity constraint on the matched query terms at the passage level. In fact, in [6], a new proximity measure for exploiting a knowledge structure was suggested. The authors presented a new movie title retrieval model that effectively searches for movie titles by leveraging the knowledge structures extracted from movie plots. The experiment results from the study revealed that the proposed model outperforms other state-of-the-art proximity-aware retrieval models.

We developed their approach further in our clinical decision support framework while the original authors only used their approach to create a movie retrieval service. In fact, there is a difference between movie retrieval and clinical decision support. In movie retrieval, the target resource is a movie that contains a synopsis and other textual information describing the movie. In this case, the terms are mostly general and are thus reachable by the initial user query, meaning that general users can quickly formulate a successful query for most movies. Meanwhile, expert knowledge is not required for information retrieval. However, in the CDS search, medical knowledge is meant to be highly specific and difficult to interpret. Furthermore, there exist many synonyms and abbreviations related to a particular concept. This indicates that proximity information should be used such that the initial query can be enhanced by either alternating the query to be more appropriate or expanding the initial query to contain more variants for the identical concepts.

In this article, we approach proximity-aware retrieval as a query expansion problem as this has been a common means of resolving syntax variations in the CDS searches. One of the critical observations related to query expansions in the CDS searches is that retrieved passages are semantically segmented [37] as they come from different sources stemming from the initial retrieval. Figure 2 presents a list of passages by an initial query. Current state-of-the-art proximity-aware retrieval models, including knowledge-structure-based models [6], cannot be used in this case, as they consider a set of sentences as semantically related in the same context. However, in the initial search result, the

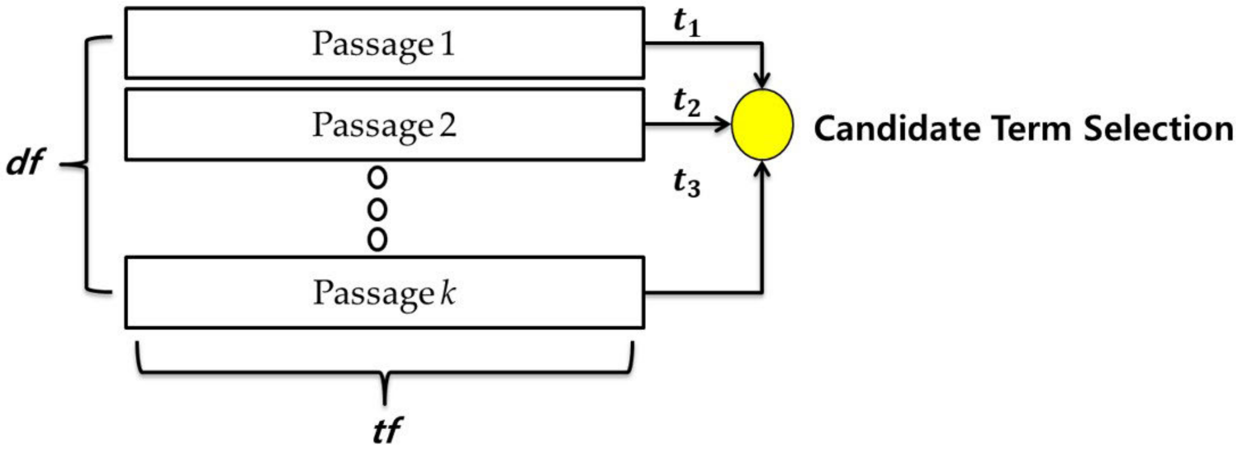


FIGURE 2. Passages are semantically segmented.

sources of retrieved passages vary meaning that the contexts of the passages differ from one another unless some of the passages come from the same page of the same source.

It is worth noting that the proposed method of exploiting proximity information using a knowledge structure is different from the previous approaches [6], [18], [19] as follows:

First, the earlier methods are limited to capture all the latent relationships among the multiple sources. Compared to creating a knowledge structure from a single document, the length of the passage is much shorter than the length of the document, causing incomplete shapes of the knowledge structure, as shown in Figure 3 (a). Meanwhile, Figure 3 (b) presents a unified knowledge structure from multiple sources. The aforementioned approach [6] did not consider creating knowledge structures from multiple sources, indicating that each knowledge structure from a single passage should be merged in a contrived approach, implying the need to develop a more sophisticated approach to creating a unified knowledge structure by scanning multiple passages.

Second, the earlier methods removed minor edges when building the knowledge structure by calculating the triangle inequality based on the Minkowski r distance theorem [18]. This can help to visualize the knowledge structure, as it removes insignificant edges between nodes. However, with regard to proximity, there exists a latent relationship, even when two nodes may not seem close. This also implies that we should further develop the initial approach further to discover the latent relationships among the terms and their relationships in the knowledge structure.

To overcome the aforementioned limitations, first we propose a new knowledge structure creation approach that relies on the use of a word embedding approach instead of counting the term frequency and the co-occurrences of words. The underlying idea of word embedding is that 'a word is characterized by the company it keeps' [38]. More recently, various NLP techniques have regarded this theory as a guide for quantifying the semantic similarities between linguistic items based on their occurrence in a large corpus.

Furthermore, it is significant to determine important terms related to the query expansion effort, as adding minor terms increases the noise in the initial query intention. We approach the query expansion as a mean of measuring node importance in the knowledge structure by various network analysis methods. As these are unsupervised methods, we reserve all of the edges of the knowledge structure to capture latent information between terms (nodes on the knowledge structure), as opposed to removing edges that do not meet a certain thresholds as calculated via triangle inequality.

Lastly, based on the created knowledge structure, we suggest a proximity-aware passage retrieval model according to the knowledge structure in Figure 4. It consists of three steps: Candidate term selection, final term selection, and the query expansion. After initially retrieved n passages are inputted, word-embedding is performed on the passages to create an embedding matrix. Nouns on the matrix are pruned to keep medical-related topics, which are Unified Medical Language System (UMLS) terms only. In the second step, we create a knowledge structure where the nodes are medical-related terms and edges are weights between two terms in the matrix. The final terms are extracted from the knowledge structure by node analysis technique to be expanded in the initial query set. In the last step, the final retrieval results are obtained by a set of expanded queries.

B. STEP 1: CANDIDATE TERM SELECTION

The knowledge structure with word embedding effectively captures the latent relationships among candidate terms of the knowledge structure. In the subsequent sections, we introduce word embedding and knowledge structure building based on the word embedding results.

An effective way to understand the context between words for NLP is to map each word unit to an appropriate vector space of numbers. There are two types of well-known vectors for NLP in the literature: one-hot representation and word vectors. The latter is a distributed representation of words in

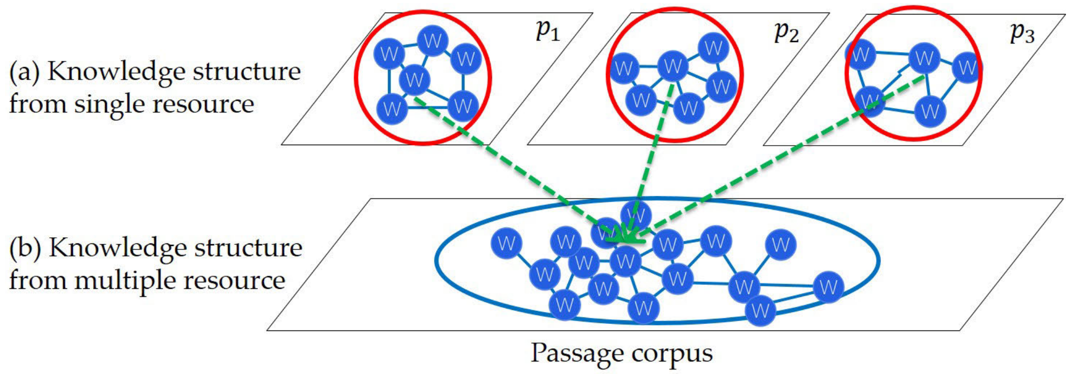


FIGURE 3. Limitations of existing knowledge structure fitted for query expansion.

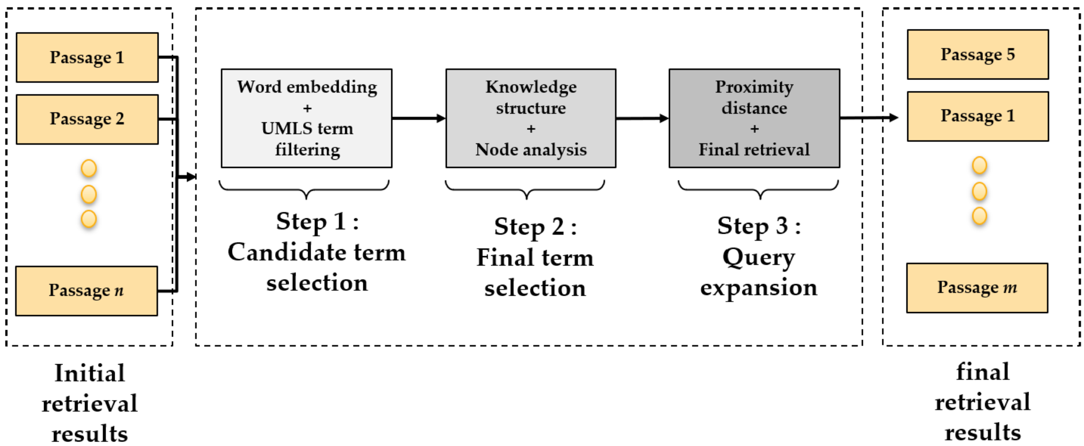


FIGURE 4. Proposed proximity-aware query expansion framework by using knowledge structure.

text based on the following statistical language model:

$$p(w_1^T) = \prod_{t=1}^{T} p(w_t|w_t^{t-1}) \tag{1}$$

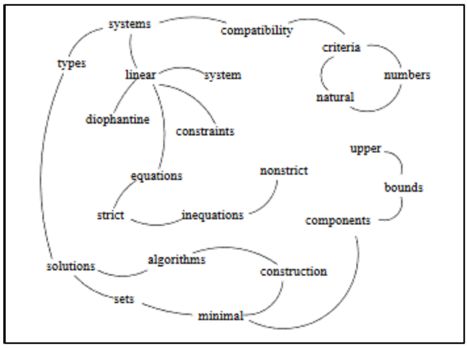
where w_t represents the tth word, and $p(w_1^T) = (w_1, w_2, \cdots, w_{T-1}, w_T)$ represents the word sequence before w_t. This general approach, referred to as word embedding, has performed well when used to generate the numerical word vector representations in various NLP tasks. In particular, the continuous bed-of-words (CBOW) and continuous Skip-gram models, as proposed in earlier work [39], have revealed significant advantages in terms of efficiency when training neural network models for word representation. The main idea of CBOW is to predict the representation of a target word that appears in the middle of other words by combining the representations of the surrounding words in a sentence or a document. In comparison, the training objective of the Skip-gram model is to predict the surrounding words using the word representation in the middle. By adopting well-trained CBOW and Skip-gram models, we can obtain a numeric vector to represent the word w_i in documents D by mapping it onto the vector $vec(w_i)$ as follows:

$$w_i \rightarrow vec(w_i)^K \tag{2}$$

where K denotes the dimension of the word vector, which is usually suggested as a value between 50 and 200. Hereafter, we use $W(D)$ to denote all the word vectors for the words in D.

We then need to reduce the number of non-clinical terms to achieve relevant candidate terms only for the query expansion process. The root set R_Q is obtained in descending order of term frequency (*tf*) normalized by document frequency (*df*) from top-retrieved passages by an initial query. As the R_Q contains general terms such as *men* and *result*, those terms may impact negatively on the furthering query expansion process, implying that those should be removed from the R_Q. In the biomedical domain, clinical terms are predefined as UMLS terms. The clinical terms detected by NCBO BioPortal API,[1] returning UMLS terms only from the initial set of terms, remain in R_Q. Let $t_i, t_j \in R_Q$ denote clinical terms in the root set R_Q according to the initial query. At this

[1] https://bioportal.bioontology.org/



Sample knowledge graph

by TF and co-occurrence:

('systems', 4), ('set', 3), ('solutions', 3), ('minimal', 3), ('types', 3), ('linear', 2), ('algorithms', 2), ('constructing', 1), ('numbers.', 1), ('considered.', 1), ('equations,', 1), ('given.', 1), ('inequations,', 1), ('solving', 1), ('system', 1), ('compatibility', 1), ('strict', 1), ('criteria', 1), ('supporting', 1),

by TextRank:

linear constraints; linear diophantine equations; natural numbers; nonstrict inequations; strict inequations; upper bounds

by human annotators:

linear constraints; linear diophantine equations; minimal generating sets; non−strict inequations; set of natural numbers; strict inequations; upper bounds

FIGURE 5. **Limitations of existing knowledge structure fitted for query expansion.**

point, the distance score between two terms t_i and t_j is defined as $S_{KS}(t_i, t_j)$, where t_i and t_j are from the word vector $W(D)$.

C. STEP 2: FINAL TERM SELECTION

As manual knowledge structure creation [9] measured the distance between two terms by involving human judges, for automatic knowledge creation, we also converted our initial distance score into a seven-point Likert Scale (1: strongly related, 7: Not related at all) as follows:

$$S_{KS}(t_i, t_j) \longleftarrow 7 - S_{KS}(t_i, t_j) \times 6 \tag{3}$$

It should be noted that, unlike the aforementioned previous approach [6], we did not apply a pathfinder algorithm [40] to reserve the latent relationship between two different terms.

We now run a variant of PageRank, TextRank, which has shown the best performance in [41], in order to obtain the importance score of all terms in the knowledge structure. Figure 5 shows a sample knowledge structure and each terms identified as important by three approaches: TF and co-occurrence [6], TextRank [41], and human annotators as golden standard, presenting that the results by TextRank are more similar to the results of the human annotators. Specifically, the term importance weight of t_i, $S_i(t_i)$, is defined as follows:

$$S_i(t_i) = (1-d) + d \cdot \sum_{t_j \in B_i} \frac{S_{KS}(t_i, t_j)}{\sum_{t_k \in L_j} S_{KS}(t_i, t_j)} S_i(t_j) \tag{4}$$

where B_i is the set of vertices connecting to t_i and L_j is the set of vertices connected to by vertex t_j connects to. d is a damping factor for implementing a random surfer model. In this paper, d is set to 0.85 as suggested in earlier work [42].

Once the term importance score is obtained, we calculate the proximity distance $S_p(t_i)$ between the term and queries in a knowledge structure.

$$\begin{aligned} S_p(t_i) &= average(\frac{distance_f(q_m, t_i)}{maxDistance(KS_p)}) \times n \\ &= \frac{2}{n-1} \frac{\sum_{q_m, t_i \in Q \bigcap P, q_m \neq t_i} distance_f(q_m, t_i)}{maxDistance(KS_p)} \end{aligned} \tag{5}$$

where $distance_f(q_m, t_i)$ is the distance score between the two terms q_m and t_i in the knowledge structure of the set of passages p, denoted as KS_p. n is the number of terms in the query set Q. For normalization, $maxDistance(KS_p)$ is the longest distance between any two terms in KS_p and n is multiplied to differentiate the score based on the length of queries. In the case that either of two terms does not occur in p, the distance between the terms is set to be $maxDistance(KS_p)$.

To reflect the proximity characteristic that a distance score drops rapidly when the distance between two terms is short while it does not change much as the distance becomes longer [7], we use a convex curve of which the first derivative is negative and the second one positive as follows:

$$S_p(t_i) \longleftarrow exp(-S_p(t_i) \times \alpha) \tag{6}$$

We used an exponential function to ensure a [0,1] range of the proximity score and to introduce α as the parameter that varies. As α becomes small, the proximity function becomes linear. We set α to 0.8 to match the best performing value in eariler work [6]. Finally, we combine this function with the node importance weighting, as follows:

$$S(t_i) \longleftarrow S_i(t_i) \cdot S_p(t_i) \tag{7}$$

D. STEP 3: QUERY EXPANSION

Once all of the weights have been determined, the terms in R_Q are ranked by their score; the top m terms not in the

Topic1 - Diagnosis
Note:
[l]78 M w/ pmh of CABG in early [**Month (only) 3**] at [**Hospital6 4406**]
(transferred to nursing home for rehab on [**12-8**] after several falls out
of bed.) He was then readmitted to [**Hospital6 1749**] on [**3120-12-11**]
after developing acute pulmonary edema/CHF/unresponsiveness?.
There was a question whether he had a small MI; he reportedly had a
small NQWMI. He improved with diuresis and was not intubated.
.
Yesterday, he was noted to have a melanotic stool earlier this evening
and then approximately 9 loose BM
w/ some melena and some frank blood just prior to transfer, unclear quantity.
Description:
[l]78 M transferred to nursing home for rehab after CABG. Reportedly readmitted with a small NQWMI.
Yesterday, he was noted to have a melanotic stool and then today he had approximately 9 loose BM
w/ some melena and some frank blood just prior to transfer, unclear quantity.
Summary
A 78 year old male presents with frequent stools and melena.

FIGURE 6. Example of a Topic for the CDS16 dataset (Topic 1).

original query are added to Q, and m is set to be 35 as was done in earlier work [37]. We adapted Terrior to execute the combination by defining a customizing weighting function. Lastly, each term in the reformulated query is used for the final passage retrieval step.

To retrieve initial relevant passages, we ran multiple IR models, in this case, TF-IDF, BM25, and LM (Language Modeling). The score of passage P_i for query Q, denoted as $S(P_i)$, is calculated by combining the scores from the IR models used as follows:

$$S(P_i) = \sum_{N} \frac{S(j,k)}{\sum_{M} S(j,k)} \tag{8}$$

where $S(j, k)$ is the score of passage P_i using the IR model k, N is the total number of used IR models, M is the total number of retrieved passages. Note that N and M were set as 3 and 100, respectively.

IV. RESULTS

Our evaluation followed a standard evaluation process of CDS search introduced in [30], initially inspired by standard TREC evaluation procedures for ad-hoc retrieval tasks [36]. We also performed CDS search simulation imitating the simulation process in [43] to see the proposed CDS framework is well performed in a practical clinical setting.

A. DATASETS

In this experiment, we have used two datasets: the CDS16 dataset [36] and the CST17 dataset [37]. The CDS16 dataset is a snapshot of the open-access subset of Pubmed Central (PMC), a database of full-text biomedical literature. This dataset was used in various CDS searches, ensuring the reproducibility of the proposed experiment. It contains 733,138 articles, provided as documents. There are 30 medical topics in the snapshot. Figure 6 shows an example of a topic in the dataset. For this experiment, the note and summary topic types were used for two reasons: first, the note is well suited for simulating the CDS search process, as a query is a set of sentences describing a patient's laboratory test results. Second, a summary was proven to perform best among the three topic types [36].

On the other hand, Table 2 shows a partial list of the CST17 dataset [37], which includes 30 cases of patients and its number of syntax variations based on the patients' laboratory test results. The cases can be classified into normal and abnormal cases to determine if the proposed model can capture false-negative cases. In practice, a robust CDS search should perform fairly between normal and abnormal cases. Relevant clinical passages were annotated with each case. In total, 28,581 variants of textual reports from 30 cases were used for the experiment.

B. EVALUATION METRICS

To evaluate the ranking effectiveness metric, we used the average precision (AvgPr), R-Precision (R-prec), the precision at the top k documents (Pr@K), and recall at the top k documents (Re@K). Pr@K is the proportion of top-k ranked documents that are relevant. Re@K is the proportion of relevant documents that ranked in the top-k positions. In general, recall is not considered to be an important measure when evaluating the search performance as it is assumed that there are a sufficient number of relevant resources in a corpus [30]. Nevertheless, the recall metric was chosen for our experiment to support the claim that query expansion is better at detecting unreached potentially relevant passages than result re-ranking.

C. OVERALL PERFORMANCE

Table 3 reports the overall ranking performances of the two proximity and bag-of-words baselines (PRM and BM25, respectively), the co-occurrence based knowledge structure retrieval model (CKS) presented in the literature [6]. The best performance among all models is in boldface font. Here, * denotes the statistically significant difference between for BM25. ** denotes the statistically significant for the BM25 and PRM models, and † and ‡ are the significance test results of EKS - PF + PX over EKS + PF and EKS - PF, respectively. The statistical significance test used here is a two-tailed

TABLE 2. Example of CST17 dataset.

Case text (translated)	Type	The number of unique variants
Blood test result is normal.	Normal	3,741
AFP (alpha fetoprotein) is normal.	Normal	1,009
Vitamin D (25OH-vitamin D) is deficit.	Abnormal	826
White blood cell in urine is positive.	Abnormal	598

TABLE 3. Retrieval performance of compared retrieval models on TREC CDS16 on summary and note topics.

Dataset	Approach	Model	AvgPr	PR@10	PR@20	PR@30	R-prec	Re@10	Re@20	Re@30
Task1 (Summary)	QE	CBNU	-	0.340	-	-	0.115	-	-	-
	QE	UDEL	-	0.336	-	-	0.152	-	-	-
	RR	PRM	0.266	0.276*	0.245*	0.229*	0.130*	0.346	0.430	0.485*
	RR	BM25	0.261	0.265	0.238	0.218	0.092	0.341	0.424	0.477
	RR	CKS	0.281*	0.301**	0.266**	0.239**	0.128*	0.357**	0.453**	0.501**
	QE	EKS + PF	0.313	0.333	0.281	0.250	0.164	0.361	0.445	0.508
	QE	EKS - PF	0.314	0.338	0.280	0.250	0.167	0.364	0.461†	0.515
	QE	EKS - PF + PX	**0.327‡**	**0.376‡**	**0.313‡**	**0.277‡**	**0.177‡**	**0.415‡**	**0.513‡**	**0.673‡**
Task3 (Note)	QE	CBNU	-	0.246	-	-	0.093	-	-	-
	QE	UDEL	-	0.183	-	-	0.066	-	-	-
	RR	PRM	0.147	0.209*	0.146	0.128*	0.059	0.234*	0.305*	0.368*
	RR	BM25	0.141	0.198	0.138	0.119	0.047	0.233	0.297	0.343
	RR	CKS	0.164**	0.220**	0.151	0.128	0.101**	0.230	0.296	0.345
	QE	EKS + PF	0.171	0.235	0.167	0.144	0.066	0.247	0.323	0.377
	QE	EKS - PF	0.178†	0.232†	0.164†	0.139†	0.108	0.240	0.312	0.369†
	QE	EKS - PF + PX	**0.202‡**	**0.256‡**	**0.179‡**	**0.152‡**	**0.118‡**	**0.265‡**	**0.341‡**	**0.399‡**

paired t-test with $p < 0.05$. We also report the performance outcomes of three variants of the proposed model, which is an embedding-based knowledge structure model (EKS), on two topics sets (the 30 summaries and the 30 notes in TREC CDS16) to observe the effect of reserving the latent edges. It should be noted that each models can be classified into two different CDS search approaches: query expansion and results re-ranking, denoted as QE and RR, respectively.

1) REFERENCE METHOD (CBNU AND UDEL)

As a reference, we included the results of UDEL and CBNU, which were the third and fourth-best performing models in terms of PR@10 in the TREC CDS2016 task. We could not find the first performance report, and the second did not report performance outcome in notes. UDEL [27] utilized the UMLS ontology to detect medical-related terms and CBNU [5] applied Word2Vec to the medical topics by UMLS. A direct comparison of UDEL and CBNU and other models in the table was unfair owing to the different problem settings used. Nevertheless, we observed two following implications: First, the UDEL'S PR@10 outcomes of 0.336 and 0.183 for the summary and note, respectively, demonstrates that detecting medical terms is a crucial step in achieving higher CDS search performance outcomes. Second, the CBNU's PR@10 scores of 0.340 and 0.246 for the summary and note, respectively, implying that utilizing word embedding would have a positive impact on the search retrieval outcome. Overall, we also observed that there is a performance drop on all measures from summary to note.

2) BM25+PROXIMITY (CKS) VS BASELINES

First, we first compared the CKS and baseline models by with BM25 and PRM without query expansion techniques on the CDS16 dataset.

On the summary topics, for measures defined on precision, CKS achieved the best outcomes for AvgPR, and all Pr@K (K ∈ 10, 20, 30 measures. Paired t-test results showed that the improvements made by CKS on top of other baselines were statistically significant in most cases, except for AvgPr and R-Precision with on PRM. For measures defined on recall, the performances of CKS were best among all the methods compared here.

Overall, CKS delivered the best performance on note topics among all methods compared here. The main difference between a summary and a note was the topic length, implying that it was more likely for noisy terms to be increased with note topics. However, CKS was still robust in this case without the medical term identification step.

To summarize, we observed the consistently better performance of the CKS model over the BM25 and PRM models. This result was also consistent in the literature [6]. Exploiting proximity based on a knowledge structure demonstrated clear superiority over BM25 for both topic cases. Without the knowledge structure, PRM achieved better performance on AvgPR, PR@K, and R-Precision. However, in terms of Re@10 and Re@20, PRM could not match BM25. We attribute this to the limited ability of PRM to measure the shortest distance between two queries in context. As PRM does not consider the segmentation of sentences, it may not

TABLE 4. Evaluation of the updated ranking in the simulation of the CDS search framework.

Dataset	Iteration	Model	AvgPr	PR@10	PR@20	PR@30	R-prec	Re@10	Re@20	Re@30
CST17 dataset	1	Baseline	0.305^{*}	0.394^{*}	0.356^{*}	0.341^{*}	0.221^{*}	0.414^{*}	0.448^{*}	0.482^{*}
		PRM	0.341^{*}	0.440^{*}	0.389^{*}	0.361^{*}	0.238^{*}	0.438^{*}	0.497^{*}	0.535^{*}
		CKS	$\underline{0.469}$	$\underline{0.476}$	$\underline{0.418^{*}}$	$\underline{0.390^{*}}$	$\underline{0.264^{*}}$	$\underline{0.472^{*}}$	$\underline{0.525^{*}}$	$\underline{0.572^{*}}$
		EKS	**0.477**	**0.481**	**0.455**	**0.433**	**0.281**	**0.531**	**0.576**	**0.596**
	2	Baseline	0.443^{*}	0.457^{*}	0.434^{*}	0.421^{*}	0.252^{*}	0.565^{*}	0.596^{*}	0.522^{*}
		PRM	0.544^{*}	0.496^{*}	0.468^{*}	0.451^{*}	0.272^{*}	0.609^{*}	0.652^{*}	0.691^{*}
		CKS	$\underline{0.551^{*}}$	$\underline{0.565^{*}}$	$\underline{0.515^{*}}$	$\underline{0.491^{*}}$	$\underline{0.319^{*}}$	$\underline{0.664^{*}}$	$\underline{0.706^{*}}$	$\underline{0.747}$
		EKS	**0.575**	**0.599**	**0.544**	**0.520**	**0.348**	**0.687**	**0.733**	**0.759**
	3	Baseline	0.512^{*}	0.517^{*}	0.501^{*}	0.492^{*}	0.303^{*}	0.605^{*}	0.622^{*}	0.636^{*}
		PRM	0.535^{*}	0.536^{*}	0.515^{*}	0.503^{*}	0.323^{*}	0.630^{*}	0.660^{*}	0.671^{*}
		CKS	$\underline{0.589^{*}}$	$\underline{0.581^{*}}$	$\underline{0.568^{*}}$	$\underline{0.560^{*}}$	$\underline{0.348^{*}}$	$\underline{0.710^{*}}$	$\underline{0.747^{*}}$	$\underline{0.774^{*}}$
		EKS	**0.682**	**0.606**	**0.599**	**0.584**	**0.377**	**0.737**	**0.782**	**0.802**

reflect the genuine relationship between two terms. On the other hand, overall, CKS outperformed the baselines on all compared metrics, indicating that the knowledge structure is a more efficient means of exploiting proximity information among queries. It should also be noted that the knowledge structure can be easily visualized, allowing clinicians readily to interpret proximity between two terms in the knowledge structure.

We also observed that the CKS model could not match the CBNU and UDEL models in terms of Pr@10 and R-precision, although it has an aspect of proximity due to the knowledge structure. We attribute this to the limitation of the result re-ranking approach, as this approach cannot discover relevant documents and does not include the initial query. Especially in a clinical setting, users use abbreviations and various synonyms, causing ranking models to discover only a limited number of relevant documents that matched the initial query. Next, we discuss proximity-aware ranking models based on the query expansion method or denoted as the word embedding-based knowledge structure model (EKS).

3) EKS WITH/WITHOUT PATHFINDER (PF) AND PROXIMITY (PX)

Since the proposed approach detected important terms through a node analysis, we compared two cases. The first was to build a knowledge structure with edge pruning, as denoted in earlier work [40], whereas the second did not use edge pruning. The knowledge structure whas been proposed as a visualization tool [18] in the educational field. Thus, edge pruning was conducted to enhance the visualization quality by reserving core edges only. Although this appeared to tweak the knowledge structure, it still reserved latent relationships between two terms.

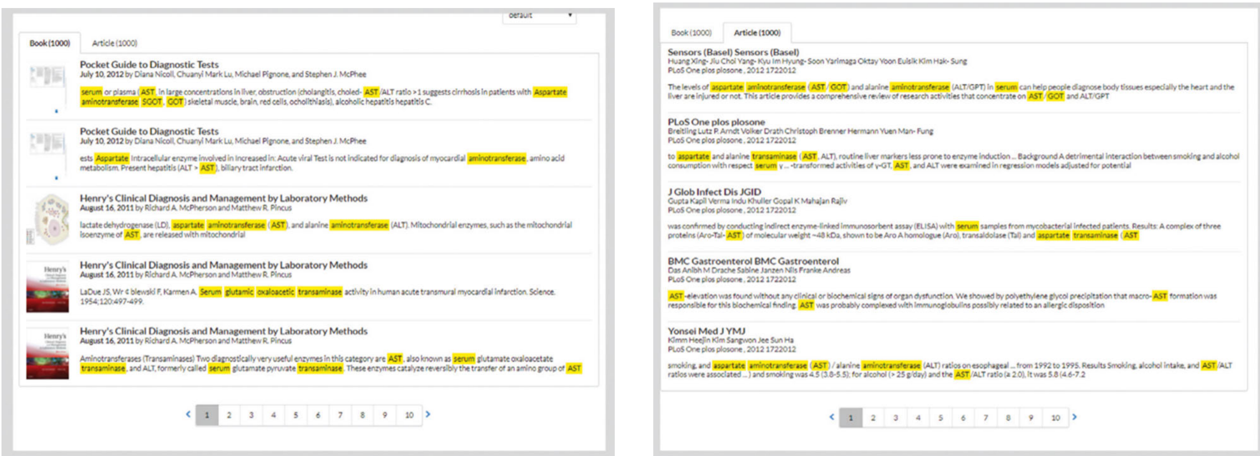
As reported in Table 3, overall reserving all edges overall in the knowledge structure tended to be a winning strategy on all metrics, or at least on the note topics. We argue that this is a promising result to be potentially applicable in a practical CDS search, as the query tends to be a set of sentences, such as a patient description. Furthermore, the number of edges in the knowledge structure only increased by 30%, indicating only an endurable burden of the computation time for a node analysis.

Furthermore, we observed that the performance gap between the two WKS models and baselines (CBNU and UDEL) was much smaller than that of the result ranking models. We attribute this to the effect of detecting medical-related terms by UMLS before creating the knowledge structure. In a CDS search, filtering non-UMLS terms out is a success factor for retrieving relevant documents as relevant documents for a given topic share high commonality in terms of clinical terms [43].

We then compared proximity-aware EKS with the other two EKS baselines (EKS with/without PF). The result reported in Table 3 shows that taking advantage of proximity information between queries and candidate terms was also a clear winner on all metrics. We conducted significant tests by comparing the results of proximity-aware EKS against the other EKS models. These tests clearly showed that the difference was statistically significant on most measures. As was observed earlier in the comparison result between the PRM and CKS models, the knowledge structure was a reliable tool for capturing proximity information between terms at abstract and feasible levels to implement a CDS search application.

4) PERFORMANCE DIFFERENCES ON 30 SUMMARY AND 30 NOTE TOPICS

Lastly, compared to the summary topics, performance degradation was observed on note topics. We reason that a summary consists of well-defined sentences created by manual efforts about the given medical case. As the topic length of topics increases, more medical terms are detected for potential query terms, indicating a greatest possibility of including terms to the original query intention. In fact, on summary topics, the average number of UMLS topics was 8.33, while 37.18 UMLS topics were detected on note topics. The result again highlighting the importance of exploiting proximity information between the candidate



(a) List of passage snippets

(b) List of retrieved articles

FIGURE 7. **Interface snapshots of the proposed search framework.**

terms and initial query terms. Indeed, considering the proximity between the terms and query prevents the newly formulated query from deviating from the original intention of the query.

D. SIMULATION OF CDS SEARCH

The goal of a CDS search is to find all relevant passages in a clinical setting [30]. Clinicians will refer to the relevant passages from the top of the ranked passages in order to support their clinical decision-making. Hence we evaluate the ranking in the setting of the CDS search process in an iterative manner. The simulation begins with initially retrieved passages from the initial query. We set a batch size (B) and assume that the queries are expanded from the retrieved passages. The obtained queries from the passages are appended to the initial query to update the CDS search ranking results. We run three iterations, with each iteration is based on the updated ranking result of the previous iteration.

We evaluate the three proximity-based ranking approaches and previously reported baseline [37] to evaluate the effect of the proximity aspect on the overall CDS search performance. We set the batch size B to be 20. That is, during each iteration, the top 20 ranked candidate passages in the current ranking were used to build the knowledge structure.

Table 4 reports the evaluation results. For each iteration, the top 20 ranked passages are retrieved and expanded queries from the passages are appended to the seed query to update the ranking. The best performance is in boldface and the second best is underlined. * denotes a statistically significant difference compared to the best outcome according to a paired t-test with $p < 0.05$. As observed in Table 4, the EKS model clearly outperformed all others on the dataset with all measures. The improvement over the other models is statistically significant in nearly all comparisons. Another observation is that the second-best performer is not always the CKS model; though it consistently outperforms on AvgPr and P@K. The PRM model outdoes CKS on a few Re@K metrics. Compared to the precision-related metrics, the recall performance increases more quickly. This occurs due to the fact that there are only 200 relevant passages for each topic, allowing the outcome to converge up to the maximum recall level as more relevant passages are included in the top-k ranking results.

As shown in Table 4, the overall performance is increased compared to that in the CDS16 dataset. A possible explanation is that the experimental setting is different with different topic sets and with a different corpus. We also find that the importance of this dataset is that its passage corpus. Proximity in the passage is much stronger than the proximity in the document, as the passage usually has more condensed sentences sharing the same context. In particular, the CDS search framework utilized here consists of PRF steps on multiple corpora; hence we claim that it is a promising result due to the consistent improvement observed after each iteration.

It should be noted that the parameters for the experiment were set to be 60 and 25 with regard to the number of feedback documents and the expansion terms, respectively. This was conducted with the AvgPr as it was shown to be a robust metric on both the CDS 16 and CST17 datasets. As the number of feedback documents increases, the AvgPr increases in the beginning. However, it soon begins to decrease as the possibility of including non-relevant passages increases as well. Similarly, the use of more than 25 expansion terms may have a negative effect on the dataset, as it also increases the possibility of including noisy terms, which are not relevant to the initial query intent.

In this simulation, when more relevant passages become available, we simply append them to the seed query to form a new larger query set for the CDS search. We compute half of the average weights on the previous queries to weight these queries.

Figure 7 presents snapshots of the interface of the proposed search framework. Figure 7(a) shows that a list of snippet for relevant passages is returned along with corresponding thumbnail images and tables. The concepts detected from the proposed framework are highlighted. If the retrieved passages contain some images or tables, its thumbnails are also given. On the other hand, Figure 7(b) shows that the "article" tab provides a list of relevant PMC articles. Each article is linked to its PMC location for browsing.

V. CONCLUDING REMARKS

In this paper, we developed the proposed framework further by expanding latent relevant terms from the knowledge structure. The main advantage of this framework is that it enhances the passage retrieval performance by incorporating proximity information among the terms. Furthermore, unlike previous studies that exploited knowledge structures during the re-ranking process, only dealing with initially highly retrieved passages, here we utilize a knowledge structure for query expansion. Thus, our approach can unveil passages not retrieved during the initial retrieval process by including latent terms in the query list.

Overall, we believe that our sequential research sheds light on a new approach for CDS searches for passage retrieval based on laboratory test results. Evidence-based medicine, such as laboratory medicine, maintains sophisticated and highly involved with decision support systems, which see heavy use, due to their dramatic increases in the needs arising in their clinical practice. We acknowledge that more advanced word-embedding approaches, that fully reflect the weights of the newly added terms in a more advanced manner, can improve CDS performance outcomes further. We leave this as our future work.

REFERENCES

[1] M. Sarrouti and S. O. El Alaoui, "A passage retrieval method based on probabilistic information retrieval model and UMLS concepts in biomedical question answering," *J. Biomed. Informat.*, vol. 68, pp. 96–103, Apr. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1532046417300503

[2] R. Chen, H. Lin, and Z. Yang, "Passage retrieval based hidden knowledge discovery from biomedical literature," *Exp. Syst. Appl.*, vol. 38, no. 8, pp. 9958–9964, Aug. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417411002405

[3] Y. Zhang, P. Hu, and F. Jian, "CCNU at TREC 2016 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016. [Online]. Available: http://trec.nist.gov/pubs/trec25/papers/CCNU2016TREC-CL.pdf

[4] H. Liu, Y. Song, Y. He, Y. Wang, Q. Hu, and L. He, "ECNU at TREC 2016: Web-based query expansion and experts diagnosis in medical information retrieval," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016. [Online]. Available: http://trec.nist.gov/pubs/trec25/papers/ECNU-CL.pdf

[5] S. Jo and K. Lee, "CBNU at TREC 2016 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016. [Online]. Available: http://trec.nist.gov/pubs/trec25/papers/cbnu-CL.pdf

[6] S. Kim, K. Han, M. Y. Yi, S. Cho, and S. Kim, "Exploiting knowledge structure for proximity-aware movie retrieval model," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manag.*, New York, NY, USA, Nov. 2014, pp. 1847–1850, doi: 10.1145/2661829.2661949.

[7] T. Tao and C. Zhai, "An exploration of proximity measures in information retrieval," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, New York, NY, USA, Jul. 2007, pp. 295–302, doi: 10.1145/1277741.1277794.

[8] T. Goldsmith, P. Johnson, and W. Acton, "Assessing structural knowledge," *J. Educ. Psychol.*, vol. 83, pp. 88–96, Mar. 1991.

[9] F. Davis and M. Yi, "Improving computer skill training: Behavior modeling, symbolic mental rehearsal, and the role of knowledge structures," *J. Appl. Psychol.*, vol. 89, p. 509, Jul. 2004.

[10] E. A. Day, W. Arthur, and D. Gettman, "Knowledge structures and the acquisition of a complex skill," *J. Appl. Psychol.*, vol. 86, pp. 1022–1033, Oct. 2001.

[11] R. Schvaneveldt, F. T. Durso, and D. W. Dearhold, "Network structures in proximity data," in *Psychology of Learning and Motivation*, vol. 24. Cambridge, MA, USA: Academic Press, Dec. 1989.

[12] M. Naveh-Benjamin, W. Mckeachie, Y.-G. Lin, and D. G. Tucker, "Inferring students' cognitive structures and their development using the ordered tree technique," *J. Educ. Psychol.*, vol. 78, pp. 130–140, Apr. 1986.

[13] T. Farrington-Darby and J. R. Wilson, "The nature of expertise: A review," *Appl. Ergonom.*, vol. 37, no. 1, pp. 17–32, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687005001262

[14] A. H. Schoenfeld and D. J. Herrmann, "Problem perception and knowledge structure in expert and novice mathematical problem solvers," *J. Experim. Psychol., Learn., Memory, Cognition*, vol. 8, no. 5, pp. 484–494, Sep. 1982.

[15] W. H. Acton, P. J. Johnson, and T. E. Goldsmith, "Structural knowledge assessment: Comparison of referent structures," *J. Educ. Psychol.*, vol. 86, no. 2, pp. 303–311, Jun. 1994.

[16] K. McLaughlin, S. Coderre, G. Mortis, and H. Mandin, "Expert-type knowledge structure in medical students is associated with increased odds of diagnostic success," *Teaching Learn. Med.*, vol. 19, no. 1, pp. 35–41, Feb. 2007, doi: 10.1080/10401330709336621.

[17] S. Samiee and B. R. Chabowski, "Knowledge structure in international marketing: A multi-method bibliometric analysis," *J. Acad. Marketing Sci.*, vol. 40, no. 2, pp. 364–386, Mar. 2012, doi: 10.1007/s11747-011-0296-8.

[18] H. W. Kim and M. Y. Yi, "Empirical validation of an automated method of knowledge structure creation from single documents," in *Proc. 9th Int. Conf. ICT Knowl. Eng.*, Jan. 2012, pp. 161–165.

[19] D. Bae, K. Han, J. Park, and M. Y. Yi, "AppTrends: A graph-based mobile app recommendation system using usage history," in *Proc. Int. Conf. Big Data Smart Comput. (BIGCOMP)*, Feb. 2015, pp. 210–216.

[20] D. Han, D. Kim, K. Han, and M. Yi, "Inductive graph-based knowledge tracing," in *Proc. Int. Conf. Big Data Smart Comput.*, Jeju, (South) Korea, Feb. 2023, pp. 91–99.

[21] C. Lee, D. Han, K. Han, and M. Yi, "Improving graph-based movie recommender system using cinematic experience," *Appl. Sci.*, vol. 12, no. 3, p. 1493, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/3/1493

[22] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, "Graph-based knowledge tracing: Modeling student proficiency using graph neural network," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Oct. 2019, pp. 156–163.

[23] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li, and Z. Zhang, "Deep graph library: A graph-centric, highly-performant package for graph neural networks," 2019, *arXiv:1909.01315*.

[24] H. Yu, T. Kim, J. Oh, I. Ko, and S. Kim, "RefMed: Relevance feedback retrieval system for PubMed," in *Proc. 18th ACM Conf. Inf. Knowl. Manag.*, New York, NY, USA, Nov. 2009, pp. 2099–2100, doi: 10.1145/1645953.1646322.

[25] S. Abdou and J. Savoy, "Searching in MEDLINE: Query expansion and manual indexing evaluation," *Inf. Process. Manag.*, vol. 44, no. 2, pp. 781–789, Mar. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457307000891

[26] A. Cohan, L. Soldaini, A. Yates, N. Goharian, and O. Frieder, "On clinical decision support," in *Proc. 5th ACM Conf. Bioinf., Comput. Biol., Health Informat.*, Newport Beach, CA, USA, Sep. 2014, pp. 651–652, doi: 10.1145/2649387.2660820.

[27] Y. Wang and H. Fang, "Extracting useful information from clinical notes," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, 2016, pp. 1–5.

[28] H. Gurulingappa, L. Toldo, C. Schepers, A. Bauer, and G. Megaro, "Semi-supervised information retrieval system for clinical decision support," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016. [Online]. Available: http://trec.nist.gov/pubs/trec25/papers/MERCKKGAA-CL.pdf

[29] R. You, S. Peng, S. Zhu, and Y. Zhou, "Fdumedsearch at TREC 2015 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2015. [Online]. Available: http://trec.nist.gov/pubs/trec24/papers/FDUDMIIP-CL.pdf

[30] L. Soldaini, A. Cohan, A. Yates, N. Goharian, and O. Frieder, "Retrieving medical literature for clinical decision support," in *Proc. Eur. Conf. Inf. Retr.*, 2015, pp. 538–549.
[31] J. R. M. Palotti and A. Hanbury, "TUW @ TREC clinical decision support track 2015," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2015. [Online]. Available: http://trec.nist.gov/pubs/trec24/papers/TUW-CL.pdf
[32] E. D'hondt, B. Grau, and P. Zweigenbaum, "LIMSI @ 2015 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2015. [Online]. Available: http://trec.nist.gov/pubs/trec24/papers/LIMSI-CL.pdf
[33] W. Chen, S. Moosavinasab, S. Rust, Y. Huang, S. M. Lin, A. Zemke, and A. Prinzbach, "Evaluation of a machine learning method to rank pubmed central articles for clinical relevancy: NCH at TREC 2016 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016.
[34] W. Hersh, S. Price, and L. Donohoe, "Assessing thesaurus-based query expansion using the UMLS metathesaurus," in *Proc. AMIA Symp.*, Feb. 2000, p. 344.
[35] A. Aronson, "Effective mapping of biomedical text to the UMLS metathesaurus: The metamap program," in *Proc. AMIA Symp.*, Feb. 2001, pp. 17–21.
[36] K. Roberts, D. Demner-Fushman, E. M. Voorhees, and W. R. Hersh, "Overview of the TREC 2016 clinical decision support track," in *Proc. 24th Text Retr. Conf.*, Gaithersburg, MD, USA, Nov. 2016. [Online]. Available: http://trec.nist.gov/pubs/trec25/papers/Overview-CL.pdf
[37] K. Han, H. Shim, and M. Y. Yi, "A new biomedical passage retrieval framework for laboratory medicine: Leveraging domain-specific ontology, multilevel PRF, and negation differential weighting," *J. Healthcare Eng.*, vol. 2018, pp. 1–19, Dec. 2018.
[38] J. R. Firth, "A synopsis of linguistic theory 1930-55," in *Studies in Linguistic Analysis* (Special Volume of the Philological Society), vol. 1952-59. Oxford, U.K.: The Philological Society, 1957, pp. 1–32.
[39] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119. [Online]. Available: http://dl.acm.org/citation.cfm?id=2999792.2999959
[40] S. Hauguel, C. Zhai, and J. Han, "Parallel PathFinder algorithms for mining structures from graphs," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 812–817.
[41] K. Han, M. Y. Yi, and J. Kim, "Search personalization in folksonomy by exploiting multiple and temporal aspects of user profiles," *IEEE Access*, vol. 7, pp. 95610–95619, 2019.
[42] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford InfoLab, Tech. Rep., 1999.
[43] G. E. Lee and A. Sun, "Seed-driven document ranking for systematic reviews in evidence-based medicine," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, New York, NY, USA, Jun. 2018, pp. 455–464, doi: 10.1145/3209978.3209994.

KEEJUN HAN received the B.S. degree in information technology from Australian National University, Canberra, Australia, in 2010, and the M.S. and Ph.D. degrees in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2013 and 2019, respectively. From 2019 to 2021, he was with Samsung Electronics, Suwon, Republic of Korea. From 2021 to 2022, he was a Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon. He is currently an Assistant Professor with the School of Computer Engineering, Hansung University, Seoul, Republic of Korea. His research interests include information retrieval, recommender systems, and smart factory.

JUNGEUN KIM received the Ph.D. degree in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST). He is currently an Assistant Professor with Kongju National University (KNU). Before joining KNU, he was a Senior Researcher with the Artificial Intelligence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI). His research interests include data mining, artificial intelligence, big data analysis with distributed processing platforms, and open data platforms.