# CBIA_VT at TREC 2015 Clinical Decision Support Track
## - Exploring Relevance Feedback and Query Expansion in Biomedical Information Retrieval

**Sihui Zhang[1], Bin He[2], Weiguo Fan[1]**
1. Center for Business Intelligence and Analytics, Pamplin College of Business
Virginia Tech, Blacksburg, VA, 24061, USA
2. Department of Computer Science, College of Engineering
Virginia Tech, Blacksburg, VA, 24061, USA
sihuiz07@vt.edu, binhe@vt.edu, wfan@vt.edu

## ABSTRACT
We present the description and results of our participation in the Clinical Decision Support track at TREC 2015. In this task, our goal was to use clinical narratives to retrieve biomedical articles. We compared the performance of pseudo relevance feedback, query expansion based on UMLS synonyms, and query expansion with personalized PageRank. In addition, we investigated the impact of different query formulation on retrieval performance. We also give out future work in this area in the end.

## KEYWORDS
biomedical information retrieval, clinical decision support, relevance feedback, query expansion, UMLS, personalized PageRank

## 1. INTRODUCTION
With the increasing prevalence of electronic health records (EHR) and the exponential growth in biomedical publications, automatically mapping clinical cases to relevant biomedical articles would greatly improve the efficiency and effectiveness of disease diagnosis and treatment. Thus the goal of the Clinical Decision Support Track (CDS) at TREC 2015 was to investigate information retrieval strategies to better assist physicians in retrieving biomedical literatures that are most relevant to patients' clinical cases.

Similar to CDS at TREC 2014 [1], participants were provided with clinical case narratives to simulate physicians' queries, and the task was to search for most relevant articles in a set of over 730,000 biomedical publications. Investigations in previous year's CDS track explored a wide range of techniques, and examples include various retrieval functions, re-ranking strategies, query expansion, relevance feedback, and machine learning techniques for article classification [2] - [5].

In this report, we describe the use of relevance feedback and several query expansion strategies in this task. The article is organized as follows: we outline our strategies in Section 2, we describe the data and technical details in Section 3, and we present the performance of our submitted runs in Section 4. Finally in Section 5 we conclude with insights gained from our participation and propose future directions.

## 2. EXPERIMENTAL DESIGN
We notice that compared with published biomedical literatures, clinical case narratives use a mixture of formal and informal languages to best suit their roles in clinical settings. However, when directly used as queries against scientific publications, there could exist vocabulary mismatches among them. Thus in our experiments, we focus on query refinement that uses pseudo relevance feedback, and investigate the effect of combining this with two query expansion strategies (**Figure 1**).

In query expansion, we take a knowledge-based approach, and use the rich information embedded in UMLS (Unified Medical Language System) at two different levels. First, we utilize the synonym relationships UMLS identifies. This method directly expands original queries with term variants for the same medical concept, thus potentially increases the recall of relevant documents.

In the second approach of query expansion, we use personalized PageRank to further exploit semantic relationships in a custom-built UMLS concept graph that represents the hierarchical relationships among multiple concepts (nodes). In such a graph, medical concepts are directly connected through parent-child and/or broader-narrower relationships, and are indirectly connected to non-immediate hypernym/hyponyms through intermediate concepts. These link structures provide the basis for us to use the PageRank algorithm to calculate concept relevance [6], [7]. The original PageRank algorithm assigns higher weights to nodes with higher number of links, thus will always identify general concepts that have many out-links as more important. These general concepts usually are not as informative as specific concepts in information retrieval. To alleviate this problem, we use personalized PageRank (PPR) that biases random jumps to a personalization vector that represents concepts in the original query [6], [7], and we further weigh these concepts against general concepts according to their scores in the basic PageRank algorithm. As a result of using PPR, we expand original queries with semantically related terms.

As the verbosity and formulation of queries affect retrieval, we further explore the effects of using versions of case narratives, including diagnosis, summary and description of the same topics.

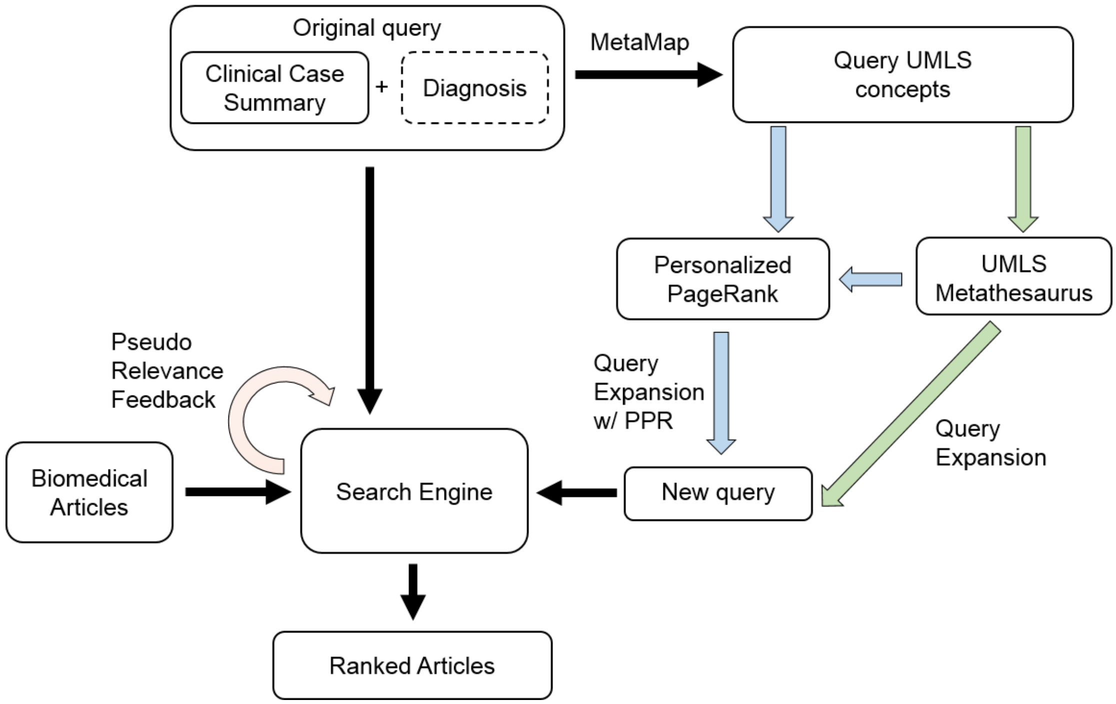


**Figure 1: System architecture for searching biomedical articles with clinical case narratives**

# 3. METHODS

## 3.1 Query and Document Description

*Queries:* thirty simulated medical case narratives (topics) in three categories: diagnosis, test, and treatment. The category label reflects doctors' information needs. In our experiments, we did not distinguish among the categories. For each topic, there are both descriptions and summaries of the case (for task A), and additionally doctor's diagnosis (for task B). As queries, we used summary texts for task A, and combined diagnosis with summary for task B.

*Documents collection:* biomedical literatures in the Open Access Subset of Pubmed Central (PMC). We obtained the PMC plain text collection as our source documents, and removed duplicates and extra documents not in the TREC CDS provided NXML files. As a result, a total of 732504 documents were indexed and used in the experiments.

## 3.2 Indexing and Retrieval

We used the open source search engine platform Terrier IR for indexing and retrieving documents. In document preprocessing, words were stemmed with Porter Stemmer, and stop words were removed according to the stop word list adopted by PubMed. During retrieval, we used BM25 as the ranking function for all of our runs.

## 3.3 Pseudo Relevance Feedback

We used Bo1 model for relevance feedback with terms extracted from three top ranked documents.

## 3.4 QE with UMLS Metathesaurus

As one of our query expansion strategies, we used MetaMap developed at NLM (National Library of Medicine) to recognize and extract biomedical concepts from query topics according to the Metathesaurus of UMLS. Specifically, we used data version 2014AB USAbase with the following sources: MSH (Medical Subject Headings), RXNORM (RxNorm Vocabulary), SNOMEDCT_US (Systematized Nomenclature of Medicine - Clinical Terms), and removed negated concepts.

The extracted CUIs (Concept Unique Identifiers) of each topic were converted to a set of terms that account for possible variants of the original concept, for example, synonyms.

## 3.5 QE with Ontology-Based PPR

As a second strategy of query expansion, we exploited the hierarchical relationship among concepts. To measure the relevance between UMLS concepts, we used personalized PageRank (PPR) on an ontology graph constructed with a subset of the UMLS concepts. In this ontology graph, nodes are UMLS concepts identified by CUI from MSH and SNOMEDCT_US sources, and edges represent relationships between concepts. Specifically, concepts include semantic types relevant to diagnosis, treatment, tests, and diseases, and the types of semantic relationship include broader/parent relationships. As regular PageRank algorithm assigns higher weights to more general/broad concepts, we biased random jumps by using query topic CUIs as a personalization vector. We further reduce the effect of general concepts by weighting scores from PPR with scores from regular PageRank. The resulting top concepts were converted to terms as in query expansion with UMLS Metathesaurus.

# 4. RESULTS

We submitted a total of six automatic runs for task A and task B (three runs for each). These runs differ in the combinations of retrieval methods and the types of queries used (**Table 1**).

**Table 1: Combination of approaches used in the submitted runs**

| Run Name | Query Type | Retrieval Func. | Rel. Feedback | UMLS Expansion | PPR |
|---|---|---|---|---|---|
| QFB | summaries | BM25 | x | | |
| UMLS | summaries | BM25 | x | x | |
| PPR | summaries | BM25 | x | x | x |
| QFBdiag | summaries + diagnosis | BM25 | x | | |
| UMLSdiag | summaries + diagnosis | BM25 | x | x | |
| PPRdiag | summaries + diagnosis | BM25 | x | x | x |

**Table 2: Evaluation of retrieval results**

| Task | Run Name | infAP | infNDCG | R-prec | P@10 |
|---|---|---|---|---|---|
| A | QFB | 0.0611 | 0.2613 | 0.2097 | 0.4367 |
| A | UMLS | 0.0411 | 0.2089 | 0.1638 | 0.3500 |
| A | PPR | 0.0345 | 0.2002 | 0.1538 | 0.3167 |
| B | QFBdiag | **0.0778** | **0.3201** | **0.2472** | **0.5033** |
| B | UMLSdiag | 0.0629 | 0.2703 | 0.2223 | 0.4367 |
| B | PPRdiag | 0.0529 | 0.2493 | 0.2033 | 0.3967 |

As shown in **Table 2**, runs with relevance feedback only (QFB and QFBdiag) performed consistently better than the other methods, irrespective of the types of queries. This was also the case in our preliminary experiments comparing relevance feedback with no feedback. On the other hand, adding knowledge-based query expansion degraded the performance. In addition, we found that reduced verbosity in the query improved the retrieval performance, as concise topic summary alone was better than lengthier description in our preliminary results and in most of the 2014 CDS runs. Moreover, adding keys terms such as those in diagnosis boosted the performance over summary alone.

## 5. DISCUSSION

In our participation in the TREC 2015 CDS track, we investigated the effects of different query refinement methodologies. We found that pseudo relevance feedback and knowledge-based query expansion have opposing effects on retrieval performance. We think the reason lies in the sources of expansion terms the two approaches use: pseudo relevance feedback uses domain expert-written articles as a source for expanding original queries, whereas UMLS Metathesaurus provides additional terms solely based on synonym and/or ontology relationships. Due to this difference, (1) it is difficult for the UMLS-based query expansion to obtain terms that are not relevant in the sense of ontology, but are otherwise critical in conveying context-dependent domain knowledge. In the case of professionally generated clinical narratives, it is likely that the user already has ample knowledge of the medical language, but would like to gain additional insights from concepts that are previously unknown to be connected through published biomedical articles; (2) extensive use of the Metathesaurus bias the query towards the set of terms that are discovered by MetaMap, which may not be relevant to a user's information needs. On the contrary, adding concise yet informative terms, such as diagnosis of patient's conditions, gear retrieval to the correct direction.

Based on the above results, one promising direction is to device strategies that fully take advantage of the information contained in feed-back documents. For example, it has been shown that combining ranking function tuning and blind feedback greatly improve retrieval performance [8], thus it would be interesting to apply this strategy to biomedical information retrieval. Another direction would be to leverage recent development in deep learning, and use it to find semantically related terms specific to biomedical literature.

## 6. REFERENCES

[1] M. S. Simpson, E. M. Voorhees, and W. Hersh,"Overview of the TREC 2014 Clinical Decision Support Track," in The 23rd Text REtrieval Conference (TREC 2014) Proceedings, 2014.

[2] S. Choi and J. Choi, "SNUMedinfo at TREC CDS track 2014: Medical case-based retrieval task," in The 23rd Text REtrieval Conference (TREC 2014) Proceedings, 2014.

[3] F. Thakrar and S. Sarkar, "Fusing manual and machine feedback in biomedical domain", in The 23rd Text REtrieval Conference (TREC 2014) Proceedings, 2014.

[4] T. Xu, P. Mcnamee, and D. W. Oard, "HLTCOE at TREC 2014: Microblog and Clinical Decision Support," in The 23rd Text REtrieval Conference (TREC 2014) Proceedings, 2014.

[5] U. N. De Lisboa, "NovaSearch at TREC 2014 Clinical Decision Support Track," in The 23rd Text REtrieval Conference (TREC 2014) Proceedings, 2014.

[6] V. N. Garla and C. Brandt, "Semantic similarity in the biomedical domain: an evaluation across knowledge sources," BMC Bioinformatics, vol. 13, no. 1, p. 261, 2012.

[7] J. Z. Wang, Y. Zhang, L. Dong, L. Li, P. K. Srimani, and P. S. Yu, "G-Bean: an ontology-graph based web tool for biomedical literature retrieval," BMC Bioinformatics, vol. 15, 2014.

[8] W. Fan, M. Luo, L. Wang, W. Xi, and E. A. Fox, "Tuning Before Feedback: Combining Ranking Discovery and Blind Feedback for Robust Retrieval," in Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2004, pp. 138-145.