UNT MEDICAL INFORMATION RETRIEVAL AT TREC 2016

Lokesh Kumar Viswavarapu[1] , Jiangping Chen[2], Ana Cleveland[2], Jodi Philbrick[2]

[1]Intelligent Information Access Laboratory, College of Information
Department of Electrical Engineering, College of Engineering
University of North Texas, Denton, TX 76203

[2]Department of Information Science, College of Information
University of North Texas, Denton, TX 76203

ABSTRACT

This paper provides a description of a project to design and evaluate an information retrieval system for clinical decision support track. The target document collection for retrieval consisted of 1.25 million biomedical related documents taken from the Open Access Subset of PubMed Central. The topics provided by TREC for query construction consisted of 30 patient narrative cases, each of which includes a Note section, a Description section, and a Summary section. The PMCID, title, abstract, keywords, subheadings of article body, introduction and conclusion paragraphs were extracted from the documents. Terrier was used as the platform for indexing and retrieval. Several models, including the LemurTF_IDF weighting model with pseudo relevancy feedback, were applied to retrieve and rank relevant documents. Out of the five runs submitted, two runs were performed by merging the retrieval results of top five individual weighting models, and the remaining three runs were obtained through passing three types of queries constructed, manually and automatically, using the Note and the Summary sections. The automatic runs are observed to receive a better performance than the manual runs. The automatic run using the Note section for query construction performed better than other runs. Overall performance of the system is around the median when compared to all TREC 2016 CDS Track submissions .

KEYWORDS

Biomedical information retrieval, clinical decision support, open-source information retrieval platform

1. INTRODUCTION

The growing reliance on biomedical information in a digital format for clinical decision-making is prompting research on novel methodologies for information retrieval. Text Retrieval Conference[1] (TREC) Clinical Decision Support (CDS) track[2] encourages and provides test collections for research in this area. The tasks in TREC CDS 2016 were more focused on retrieving results based on the Note section, which consisted of the History of Present Illness (HPI) of a patient, as it simulated a more realistic clinical decision-making scenario.

In this paper, we discuss our experimental design and methodologies implemented towards achieving our goal to develop a baseline medical information retrieval system. The target document collection for

[1] http://trec.nist.org/
[2] http://www.trec-cds.org/

retrieval consisted of 1.25 million biomedical related articles taken from Open Access Subset of PubMed Central (PMC). The topics provided by TREC for query construction consisted of 30 patient narrative cases which include a Note section (a note on the HPI of the patient); a Description section (a general description of the patient's problem); and a Summary section (a short summary of the patient's problem).

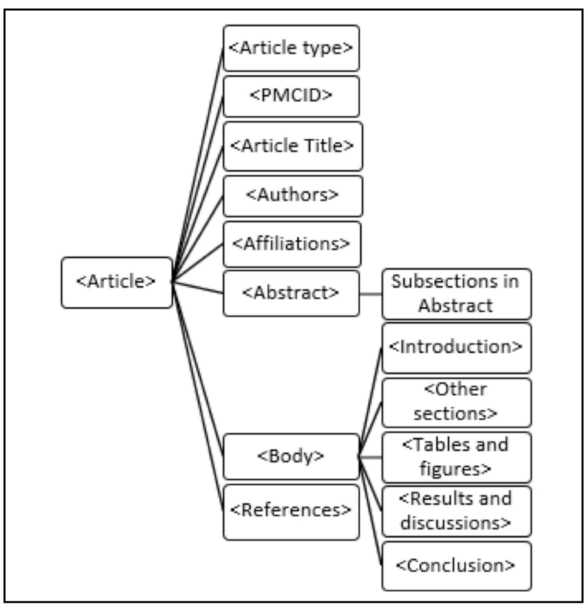
To retrieve relevant documents for the 30 topics, we developed an experimental system using an open source information retrieval (IR) platform Terrier v4.1 configured with pseudo relevancy feedback. We used the Note and Summary sections for query construction.

The remaining paper is organized as follows: Section 2 discusses the general structure of the document collection and the topics provided by TREC; Section 3 summarizes the literature from TREC CDS 2015 track; Section 4 presents the experimental design and the methodologies implemented; and Section 5 presents a description of the runs submitted and the evaluation results. The paper concludes with Sections 6 and 7, which include a discussion of our system performance, future research, and a summary.

2. THE DOCUMENT COLLECTION AND THE TOPICS

2.1 The Document Collection

The document collection for the task is the collection of biomedical related articles of various types, such as research articles, editorials, review articles, and meeting notes. Each article in the collection is an encoded xml file with a general hierarchical structure shown in Figure 1(a) and (b).




```
<article>
    <article-type>…</article-type>
    <pmcid>…</pmcid>
    <article-title>…</article-title>
    <authors>…</authors>
    <affiliations>…</affiliations>
    <abstract>…</abstract>
        <subsections>…</subsections>
    <body>
        <introduction>…</introduction>
        <other-sections>…</other-sections>
        <conclusion>…</conclusion>
    </body>
    <references>…</references>
</article>
```

Figure 1 (a): General Structure of PMC Articles

Figure 1(b): General XML Attributes of PMC Articles

2.2 Topics

TREC provided 30 case narratives of the patient's health problem, each containing three elements: note, description and summary. The topics of the case narratives were divided into three types: diagnosis, tests, and treatment, with 10 case narratives each. A sample topic of type diagnosis is show in Figure 2.

<topic number="1" type="diagnosis">

<note>

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.

</note>

<description>

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.

</description>

<summary>

A 78 year old male presents with frequent stools and melena.

</summary>

</topic>

Figure 2. Sample "Diagnosis" Topic from TREC CDS 2016 Case Narratives

3. REVIEW OF TREC 2015 TRACK METHODOLOGIES

The TREC CDS 2015 papers were reviewed to understand various methodologies implemented in developing biomedical information retrieval systems.

We observed that most systems had a document processing phase before indexing. In the majority of the methodologies, the title, abstract, and body of the article were extracted during the processing and passed on to indexing (Drosatos, Roumeliotis, Arampatzis, & Kaldoudi, 2015; Stöber, et al., 2015). Song, He, Hu, & He, 2015 noted that an additional extraction of age, gender, captions of tables and figures and references of the article increase the performance of the system. In the document processing phase, some systems performed semantic type and synonyms annotations to the free text in the document using external knowledge sources such as Unified Medical Language System (UMLS) and Medical Subject Headings (MeSH) (Hu, Wu, Mei, & Vydiswaran, 2015; Stöber, et al., 2015). Some other document processing techniques included applying language models such as SPUD language model (Cummins, 2015), unigram and bigram models (Nikolentzos, Meladianos, Liakis, & Vazirgiannis, 2015) and Himestra language model (Abacha & Khelifi, 2015).

For indexing and retrieval, it was observed that Apache Lucene, Indri, and Terrier were the most commonly used information retrieval (IR) platforms. By examining the retrieval configuration of 2015 TREC participants, BM25 and TF_IDF were noted to be the most commonly used weighting models. Whereas, some other models include TW-IDF, a retrieval model based on graph of words approach (Nikolentzos, et. al., 2015), and Unigram query likelihood model (You, Zhou, Peng, & Zhu, 2015).

Some systems had implemented re-ranking mechanisms such as Learning-to-Rank (Hu, et. al., 2015; Jiang, Guan, Su, Zhao, & Yang, 2015; Stöber et al., 2015; Song, et.al., 2015), ranking using co-occurrence network (Jiang, et.al., 2015), time based re-ranking (D'hondt, Grau, & Zweigenbaum, 2015) and CQT based re-ranking (D'hondt, Grau, & Zweigenbaum, 2015). Learning-to-rank is a machine learning application in ranking process. In this technique, a combination of multiple features, which are obtained from weighting models in case of query dependent systems (all systems developed to perform TREC CDS tasks), is used to minimize the loss function. An optimal loss function modeled after learning is used for ranking the relevant documents. The most popular used Learning-to-Rank machine learning algorithms were Support Vector Machines (SVM) (Jiang, et.al., 2015; Song, et.al., 2015) and Random Forest algorithm (Song, et.al., 2015).

Query construction and query expansion are two crucial functionalities of an IR system, which have a major impact on the system performance. These functionalities were implemented using various mathematical and intuitive approaches. Discriminative query model (DQM) (Cummins, 2015) is one such mathematical approach, which uses P´olya distribution for query modeling (Cummins, Paik, & Lv, 2015). The Summary section of the topics was used for query construction in most of the systems as this section contained relatively fewer and more precise terms in explaining the patient's health problem (Drosatos, Roumeliotis, et.al., 2015; Hu, Wu, Mei, & Vydiswaran, 2015; Jiang, et.al., 2015; Palotti & Hanbury, 2015).

Query expansion is a technique that enhances the free-text in the query with the domain specific terminology. From the review, Pseudo Relevancy Feedback (PRF) was the most commonly implemented query expansion technique. Along with PRF, query expansion using UMLS and MeSH were also more likely implemented (Drosatos, et.al., 2015; Hu, et.al., 2015; Jiang, et.al., 2015; You, et.al., 2015). Few innovative expansion techniques observed were expanding the terms using search results from Wikipedia and Google (Jo, Seol, & Lee, 2015; Song, et.al., 2015), ontology based query terms annotation (Audeh & Beigbeder, 2015), and personalized page rank technique (Zhang, He, & Fan, 2015).

The review of papers published by 2015 participants of TREC CDS track enabled us to achieve a good understanding of methodologies and specific tools and platforms used in the implementation of a medical information retrieval system.

4. EXPERIMENTAL DESIGN

Based on the literature review, we developed our experimental design as a four-stage model that follows the flow of a typical information retrieval system for performing the assigned task. Figure 3 presents our experimental design along with its functional flow. As illustrated in Figure 3, our system contained four stages: document processing, indexing, query construction and retrieval and ranking. Each stage is further explained below.

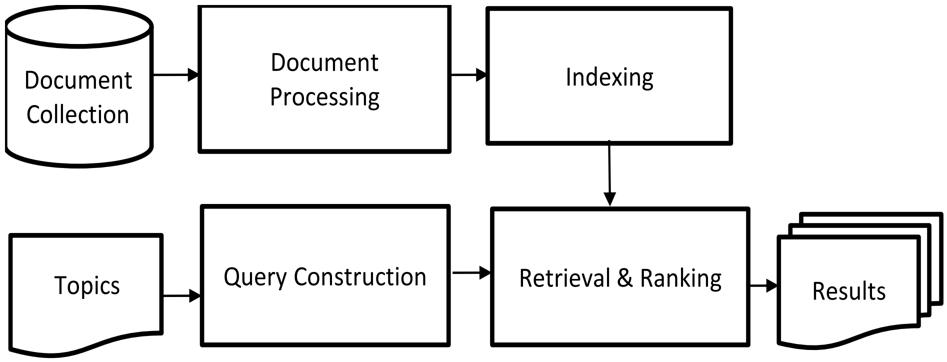


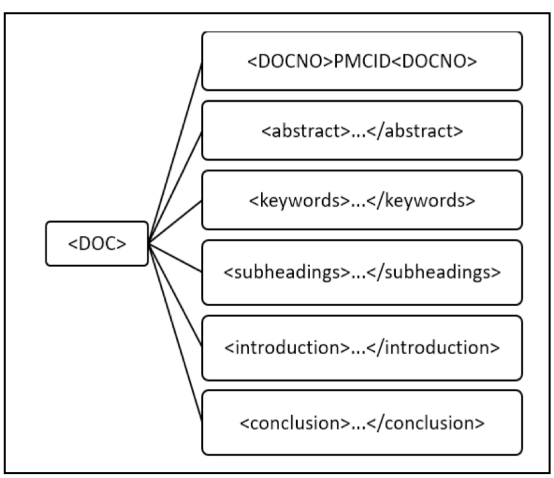
Figure. 3: Information Retrieval System

4.1 Document Processing

The document processing stage extracted the most informative sections of the document, which included PMCID, title, abstract, keywords, subheadings in body of the document, introduction, and conclusion. The document collection consists of 1.25 million articles, which was an Open Access Subset of PubMed Central (PMC).

To extract the desired sections of the document, Python script with regular expressions was used. The extracted fields were organized into TREC format to make them compatible with the indexing platform. Figure 4 provides the structure of the TREC documents.

The major challenge in this stage was the diverse format and structures of the documents in the collection. These increased the complexity of the document processing program. For example, the document type 'Meeting Notes' does not have an abstract, introduction, or conclusion sections in most of them.




```
<DOC>
   <DOCNO>PMCID</DOCNO>
   <TEXT>
   <article-title>…</article-title>
   <abstract>…</abstract>
   <keywords>…</keywords>
   <subheadings>…</subheadings>
   <introduction>…</introduction>
   <conclusion>…</conclusion>
   </TEXT>
</DOC>
```

Figure 4(a): Document Structure of TREC **Figure 4(b): XML Attributes of TREC**

4.2 Indexing

Indexing is a process of parsing the data in document collection to indices so that retrieval can be performed. We used Terrier v4.1 IR platform for both indexing and retrieval. Terrier tokenizes the free text in the target document collection to index tokens and passes through a term pipeline, which removes stop words and performs stemming to the indexing terms. A predefined list of stop words and stemming algorithm were configured in Terrier properties to perform the term pipeline. The tokens obtained after the term pipeline were used to generate indices.

4.3 Query Construction

Queries were constructed from the TREC CDS topics. In our experiment, we followed three approaches in constructing queries, which included:

- **Note Automatic**: The note section used for query construction was extracted using Python regular expressions. Irrelevant information like hospital admission is discarded.
- **Note Manual**: The key phrases in the note section were identified manually by a medical domain expert for query construction.
- **Summary Automatic**: The summary section used for query construction was extracted using Python regular expressions.

Three runs are performed using the queries constructed from the above approaches.

4.4 Retrieval and Ranking

In this stage, documents relevant to a given query were retrieved from the collection using a weighting model,which scored the degree of relevancy. For this task, we performed five retrieval runs; three out of which were obtained by passing the queries constructed using the three approaches described above. Two of the retrieval runs were obtained by merging the retrieval results from individual weighting models.

4.4.1 Query Expansion using Pseudo Relevance Feedback

Pseudo relevance feedback is an automatic relevance feedback technique where the system assumes the top 'n' documents retrieved in the initial run to be highly relevant and fetches the terms in those documents for query expansion. Relevance feedback has been shown to improve retrieval performance.

In our system, query expansion with pseudo relevancy feedback was applied to all the runs using Bose-Einstein model (Bo1) in Terrier with configuration of 20 terms in each of the top 100 documents retrieved in the initial run for expanding the queries.

4.4.2 Merging from Different Models

We evaluated the performance of individual weighting models available in Terrier using TREC 2015 relevance judgment results (see Section 4.5) and found that a weighing model's performance was high with respect to a particular measure and comparatively low for remaining measures. Intuitively, it might increase the overall performance of the system if retrieval results from the top performing weighting models were merged in an appropriate way. Therefore, we conducted two runs by passing the Note automatic and Summary automatic queries to five different weighting models and merged their individual results. Merging the individual results was performed using a merging technique called the Shadow Document method (Wu & Crestani, 2004). This method was one of the merging algorithms that optimally performs overall better. The five weighting models used were LemurTF_IDF, DLH, IFB2, PL2, BM25. the scores obtained from the weighting models were normalized and made comparable.

After normalization and comparison of two or more results, the algorithm below was applied to obtain the global score for each retrieved document.

$$\boldsymbol{global\_score(d)} = \sum_{i=1}^{m} \boldsymbol{s_i(d)} + \frac{\boldsymbol{k * (n - m)}}{\boldsymbol{m}} \sum_{i=1}^{m} \boldsymbol{s_i(d)}$$

(Wu & Crestani, 2004)

Where,

$\boldsymbol{k}$: A coefficient,which lies between 0 and 1. In our case k=0.5

$\boldsymbol{d}$: The documents retrieved by any one of the models

$\boldsymbol{s_i}$: The score of the document in the ith result file

$\boldsymbol{n}$: Number of result files available

$\boldsymbol{m}$: Number of result files in which the document $\boldsymbol{d}$ occurred

If the document $\boldsymbol{d}$ occurred in $\boldsymbol{m}$ result files out of $\boldsymbol{n}$ result files with a score $\boldsymbol{s_i(d)}$ (where $1 \leq \boldsymbol{i} \leq \boldsymbol{m} \leq \boldsymbol{n}$) then the global score of $\boldsymbol{d}$ was calculated using the above formula.

In a nutshell, if a document was present in more than one result, then all the scores of that document were added up to get the global score. In the case that the document was present in only one result, then the document's score was added to the coefficient value. The resulting merged result file was sorted in descending order of their scores and re-ranked 1-1000. The documents with rank greater than 1000 were discarded from the results.

4.5 Retrieval Model Evaluation with TREC 2015 Test Collection

To select a weighting model out of 16 models available in Terrier, we used TREC 2015 CDS track retrieval test collection to evaluate the individual weighting models performance with respect to the measures, infAP, infNDCG, R-Prec, and P@10. Since the document collection used for TREC 2016 was an expanded version of the one for TREC 2015, we assumed that the document collection would have a minimal effect of the performance evaluation. The evaluation results of these 16 weighting models with and without PRF are shown in Table 1(a) and (b) respectively.

Model	InfAP	infNDCG	R-Prec	P@10
LGD	0.0766	0.1546	0.0328	0.19
Hiemstra_LM	0.0854	0.1729	0.0396	0.2
BM25	0.0875	0.1766	0.0412	0.2067
DFRee	0.0773	0.1516	0.0329	0.2
BB2	0.0876	0.1802	0.0422	0.2133
InL2	0.0864	0.1708	0.0393	0.21
IFB2	0.0932	0.1871	0.0448	0.2033
PL2	0.0868	0.1785	0.0434	**0.2267**
DLH	0.0893	0.1814	0.0437	**0.2267**
DFR_BM25	0.0875	0.1766	0.0412	0.2067
TF_IDF	0.0865	0.1718	0.0399	0.2133
In_expB2	0.091	0.1859	0.0441	0.2067
DPH	0.0814	0.1623	0.0367	0.2133
DLH13	0.082	0.1649	0.0382	0.1967
LemurTF_IDF	**0.0996**	**0.1889**	**0.0464**	0.22
In_expC2	0.0944	0.1879	0.0455	0.2233

Table 1(a): Evaluation Results of Weighting Models Without PRF

Model	InfAP	infNDCG	R-Prec	P@10
LGD	0.0464	0.187	0.1007	0.2067
Hiemstra_LM	0.0524	0.2133	0.1017	0.16
BM25	0.0599	0.2204	0.1103	0.2367
DFRee	0.0499	0.1942	0.0978	0.2267
BB2	0.0619	0.2241	0.1127	0.2233
InL2	0.0584	0.2152	0.1097	0.24
IFB2	**0.0648**	0.2297	0.1171	0.2233
PL2	0.0578	0.2077	0.1046	0.2467
DLH	0.0595	0.2175	0.1118	**0.25**
DFR_BM25	0.0583	0.2161	0.109	0.2333
TF_IDF	0.0583	0.2135	0.1094	0.2467
In_expB2	0.0638	0.2282	0.1161	0.2233
DPH	0.0517	0.1983	0.0994	0.2333
DLH13	0.0567	0.206	0.1068	0.2367
LemurTF_IDF	0.0643	**0.2369**	0.1149	0.1933
In_expC2	0.0647	0.2259	**0.1183**	0.2

Table 1(b): Evaluation Results of Weighting Models With PRF

It was observed that LemurTF_IDF weighting model performed consistently better for all the four measures in case of queries without PRF (Table 1(a)), and the same model performed better with respect to infNDCG in case of queries with PRF (Table 1(b)). Based on this observation, the LemurTF_IDF model was chosen for three out of our five runs.

From the evaluation Table 1(b), it was observed that different models performed better with four different measures. Based on this observation, individual results of all the top performing models with respect to four measures were merged to obtain merged results.

5. Results

Five runs were submitted to TREC 2016. An overview of the runs is presented below:

i. **Note Automatic Run (Run ID: UNTIIANA)**: This run used the Note Automatic queries and a result file with the top 1000 relevant document IDs in the order of most relevant to least relevant was obtained.

ii. **Note Manual Run (Run ID: UNTIIANM)**: This run used the Note Manual queries and a result file with the top 1000 relevant document IDs in the order of most relevant to least relevant was obtained.

iii. **Summary Automatic Run (Run ID: UNTIIASA)**: This run was designed to use the Summary Automatic queries and the evaluation results were observed to be similar to the Note Automatic run. This run was performed with a configuration error, which invalidated the results. This would be further investigated in the future.

iv. **Results Merge Run with Note Automatic (Run ID: UNTIIANMERGE)**: This run was generated by merging the results of the top five highest performing weighting models using the queries constructed from the Note section automatically.

v. **Results Merge Run with Summary Automatic (Run ID: UNTIIASMERGE)**: This run was generated by merging the results of the top five highest performing weighting models using the queries constructed from the Summary section automatically.

The description and evaluation results of four runs are presented in Table 2 and Table 3.

Run ID	**Topic Section**	**Query Construction**	**Weighting Model**
UNTIIANA	Note	Automatic	LemurTF_IDF
UNTIIANM	Note	Manual	LemurTF_IDF
UNTIIANMERG	Note	Automatic	LemurTF_IDF+DLH+IFB2+PL2+BM25
UNTIIASMERG	Summary	Automatic	LemurTF_IDF+DLH+IFB2+PL2+BM25

Table 2: Description of Runs Submitted to TREC

Run ID	**infAP**	**infNDCG**	**R-Prec**	**P@10**
UNTIIANA	0.0153	0.1554	0.0951	0.2267
UNTIIANM	0.0144	0.1405	0.0880	0.1933
UNTIIANMERG	0.0132	0.1481	0.0819	0.2133
UNTIIASMERG	0.0113	0.1414	0.0841	0.1933

Table 3: Evaluation Results of Runs Submitted to TREC

We present three of our runs as compared with the maximum, minimum, and median values of infNDCG for each topic in Figure 5 (a), (b), and (c).

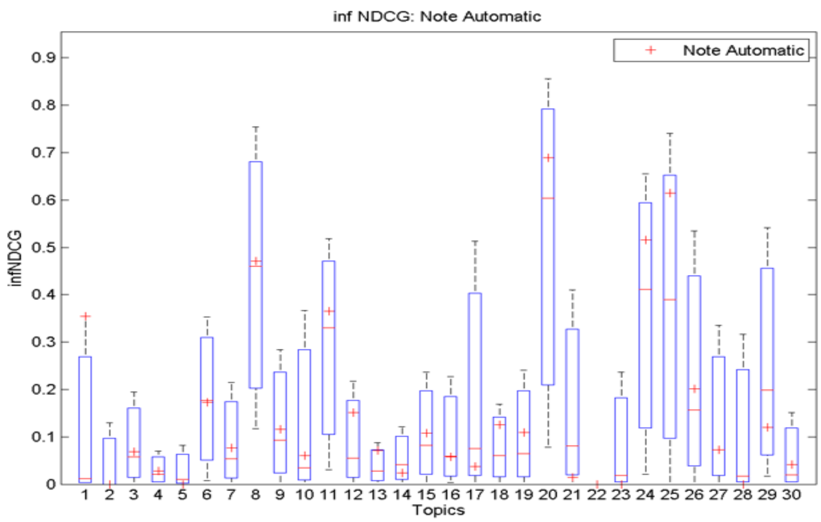


Figure 5(a): Box Plot of infNDCG for Note Automatic (UNTIIANA)

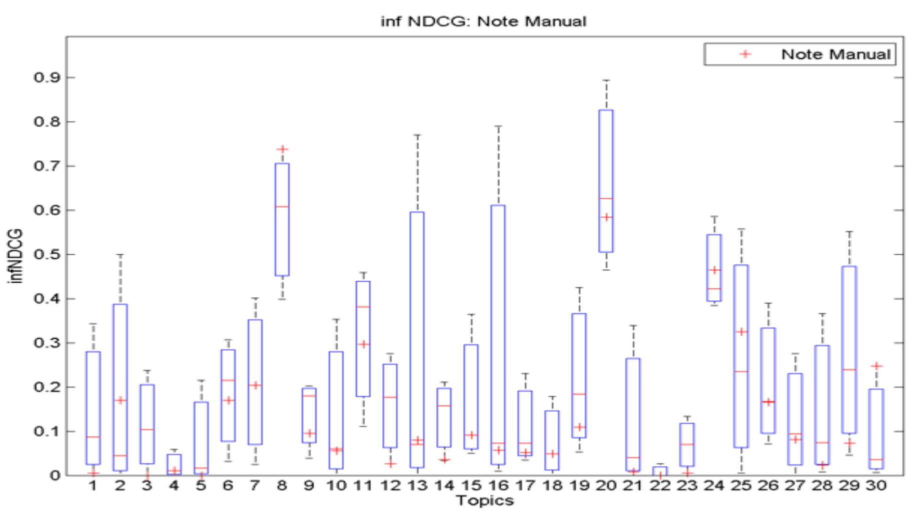


Figure 5(b): Box Plot of infNDCG for Note Manual (UNTIIANM)

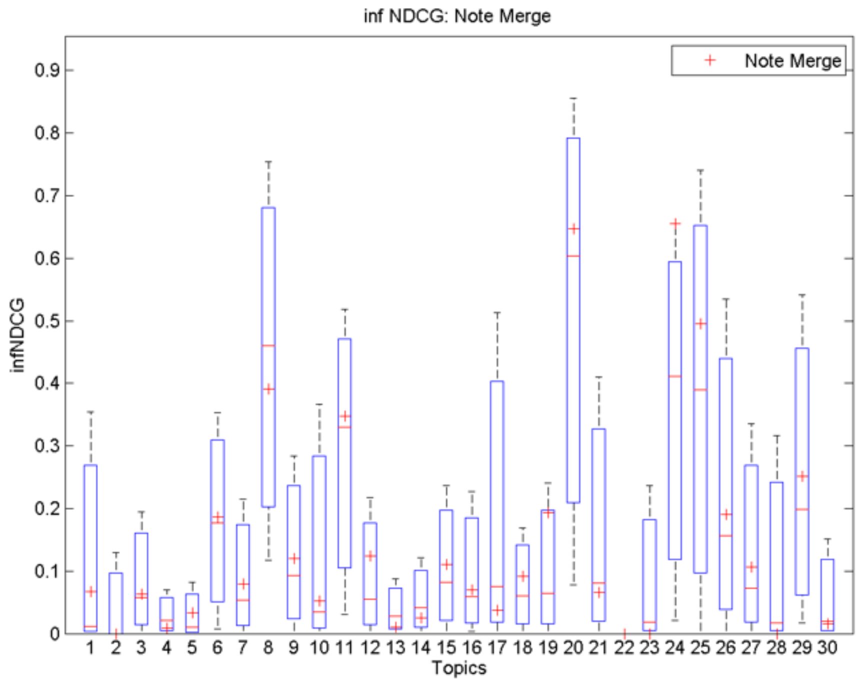


Figure 5(c): Box Plot of infNDCG for Note Merge (UNTIIANMERG)

6. DISCUSSION AND FUTURE RESEARCH

We successfully designed and implemented a baseline medical information retrieval system that could retrieve relevant medical documents for three types of topics within a short period of time. Processing was a complex automation challenge. At this stage, we extracted the PMCID, title, abstract, keywords, subheadings of body of the document, introduction and conclusion sections from the diverse document collection. For indexing and retrieval, Terrier IR platform was chosen, as it is one of the commonly used IR platforms and was used by many of the TREC 2015 participants.

Out of five runs submitted, automatic runs were observed to have a better performance than the manual runs and the Note Automatic run had a better performance with an overall inferred NDCG(infNDCG) of 0.1554 and steadily performed above or around the median for most of the 30 queries.

The two automatic runs, Note Merge and Summary Merge, generated by merging the individual results obtained from five different models were expected to perform better than the other runs as the models selected for merging were the best performing models for TREC 2015 evaluation but observed to be inverse.

In order to improve our current system, we are going to investigate the reasons that the merging method did not work well. A literature review shall be conducted to understand the applications of Shadow Document method in information retrieval systems and its limitations. Also, we would like to understand systematically Terrier's weighting models and their scoring functions using TREC 2016 relevance judgment results.

We would like to explore the following strategies:

Query expansion using external knowledge sources: From the literature it was observed that both healthcare specific knowledge sources like UMLS, MeSH; and generic knowledge sources like Google, Wikipedia search results were used for enhancing the free-text with medical terminology. In our future research we would like to investigate the impact of these knowledge sources on the performance of an IR system.

Learning-to-Rank: Learning-to-rank is a machine learning application in ranking process. In our research, we would like to understand the feasibility of various 'Learning-to-Rank' models and the time and code complexities introduced into the system by those models.

7. SUMMARY

In this paper, we presented our approaches in performing the tasks required by TREC CDS 2016. We used a target document collection for retrieval consisting of 1.25 million biomedical related documents taken from the Open Access Subset of PubMed Central (PMC). In document processing we used regular expressions to extract the article tile, PMCID, abstract, keywords, subheadings of the document, introduction and conclusion paragraphs from the documents. Indexing and retrieval were performed by Terrier v4.1. We also experimented to merge individual results obtained by different weighting models and generated a new result. From the evaluation results received from TREC, the overall performance of our IR system is around the median when compared to all submissions of TREC 2016 CDS Track.

REFERENCES

Abacha, A. B., & Khelifi, S. (2015). LIST at TREC 2015 Clinical Decision Support Track: Question analysis and unsupervised result fusion. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Audeh, B., & Beigbeder, M. (2015). EMSE at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Cummins, R. (2015). Clinical decision support with the SPUD language model. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Cummins, R., Paik, J. H., & Lv, Y. (2015). A Pólya urn document language model for improved information retrieval. *ACM Transactions on Information Systems (TOIS), 33*(4), 21.

D'hondt, E., Grau, B., & Zweigenbaum, P. (2015). Laval University at TREC Dynamic Domain 2015: Combination of techniques for subtopics coverage. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Drosatos, G., Roumeliotis, S., Arampatzis, A., & Kaldoudi, E. (2015). DUTH at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Hu, F., Wu, D. T., Mei, Q., & Vydiswaran, V. V. (2015). Learning from medical summaries: The University of Michigan at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Jiang, J., Guan, Y., Su, J., Zhao, C., & Yang. J. (2015). HIT-WI at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Jo, S. H., Seol, J. W., & Lee, K. S. (2015). CBNU at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Nikolentzos, G., Meladianos, P., Liakis, N., & Vazirgiannis, M. (2015). AUEB at TREC 2015: Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Palotti, J., & Hanbury, A. (2016). TUW@ TREC Clinical Decision Support Track 2015. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Song, Y., He, Y., Hu, Q., & He, L. (2015). ECNU at 2015 CDS Track: Two re-ranking methods in medical information retrieval. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Stober, J., Heale, B.S.E., Fulghum, K., & Del Fiol, G. (2015). Concept based information retrieval for clinical case summaries. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Wu, S., & Crestani, F. (2004, March). Shadow document methods of results merging. In *The 19th ACM Symposium on Applied Computing*. Paper presented at the 2004 ACM Symposium on Applied Computing, Nicosia, Cyprus (p. 1067-1072).

You, R., Zhou, Y., Peng, S., & Zhu, S. (2015). FDUMedSearch at TREC 2015 Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.

Zhang, S., He, B., & Fan, W. (2015). CBIA VT at TREC 2015 Clinical Decision Support Track-Exploring relevance feedback and query expansion in biomedical information retrieval. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings*. Paper presented at the Twenty-Fourth Text REtrieval Conference, Gaithersburg, MD.