UNT Precision Medicine Information Retrieval at TREC 2017

Lokesh Kumar Viswavarapu[1,3], Jiangping Chen[2], Ana Cleveland[2], Haihua Chen[1,2]

[1] Intelligent Information Access Laboratory, College of Information
[2] Department of Information Science, College of Information
[3] Department of Electrical Engineering, College of Engineering
University of North Texas, Denton, TX 76203
Jiangping.Chen@unt.edu, Ana.Cleveland@unt.edu

ABSTRACT

This paper reports our participation in TREC 2017 Precision Medicine (PM) track. Based on our TREC 2016 Clinical Decision Support System, we implemented and tested five different query construction strategies: Query construction with disease weighted terms, with synonyms of disease terms, with Internet search results, with gene alias terms, and with Terrier logical query language. A re-ranking strategy that considered the occurrence of disease names, gene names, and treatment terms were applied to adjust the order of the retrieved documents. A new experimental module called *Relevance Judgment User System* (RJUS), which is an augmentation to the information retrieval platform Terrier[1] v4.2, was designed to facilitate the design of query construction and result re-ranking strategies. We submitted 5 runs applying the five query construction strategies respectively combining with pseudo relevance feedback and the re-ranking approach. The query construction with Terrier logical query language achieved the best performance among our submitted results. Our future study will investigate the use of topic modeling for query construction and effective approaches for finding relevant clinical trials.

KEYWORDS

Precision Medicine Retrieval, Re-ranking, Relevance Judgment, Query Expansion

1. INTRODUCTION

Precision Medicine is an approach in healthcare which deals with evidence-based treatment and prevention of diseases customize to the individual patient. Text Retrieval Conference[2] (TREC) Precision Medicine (PM) track[3] specifically deals with the challenges in developing an information retrieval (IR) system that retrieves relevant cancer treatment literature using patient's genomic and demographic information. In 2017, participants were challenged to design a system to perform two IR tasks, one to retrieve relevant cancer prognosis, treatment or prevention based on articles from a collection of scientific abstracts and the other task to retrieve eligible clinical trials from a collection of clinical trials for a given common set of 30 patient cases which are referred as topics[4]. These topics mainly contain information about the type of cancer, genomic and demographic information.

In this paper, we discuss our participation, experimental design and methodologies implemented to perform mainly information retrieval task on scientific abstracts. For the *scientific abstracts retrieval task,* the target document collection was MEDLINE's abstracts, abstracts obtained from the American

[1] http://terrier.org/
[2] http://trec.nist.gov/
[3] http://www.trec-cds.org/2017.html
[4] http://www.trec-cds.org/topics2017.xml

Association of Cancer Research[5] (AACR), and the American Society of Clinical Oncology[6] (ASCO) proceedings. The document collection for the *clinical trials task* was the clinical trials obtained from Clinicaltrials.gov[7]. The topics provided by TREC for query construction consisted of 30 patient cases with information of cancer type, gene/variant, demographics and other health problems of the patient.

To retrieve relevant documents from the scientific abstracts/clinical trials for the 30 topics, we have improved our previous IR system designed for 2016 TREC Clinical Decision Support[8] (CDS) track (Viswavarapu, Cleveland, Chen & Philbrick, 2016). In the new system, we introduced a new experimental module called *Relevance Judgment User System* (RJUS), which is an augmentation to the IR platform Terrier[9] v4.2, designed to assess and fine tune the IR strategies.

This paper is organized into the following sections. Section 2 discusses the general structure of the document collection and the topics provided by TREC. Section 3 summarizes the literature from TREC CDS 2016 track. Section 4 presents the experimental design and the methodologies implemented for 2017 TREC PM track. Section 5 describes the runs submitted, and section 6 presents the evaluation results and discussion. The paper concludes with a summary, and an outlook of future research.

2. THE DOCUMENT COLLECTION AND THE TOPICS

2.1 The Document Collections

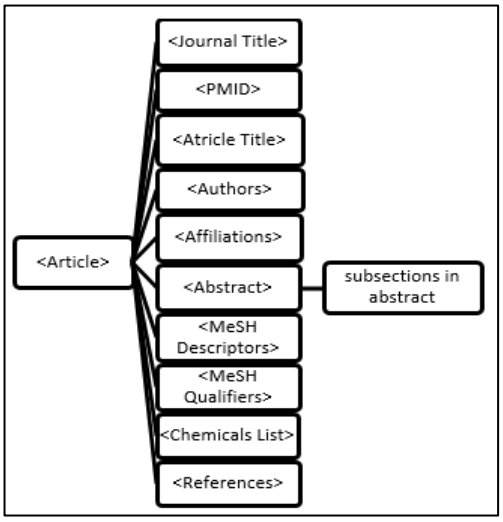
2.1.1 Scientific Abstracts

The scientific abstract collection is a combination of abstracts compiled from three different sources. MEDLINE, proceedings of AACR, and proceedings of ASCO. The content and structure of the documents vary. For example, the MEDLINE abstracts contain a number of metadata elements along with the abstract while AACR and ASCO proceedings abstracts have minimal metadata information. For our experiment, we worked with the XML version of MEDLINE abstracts and selected a subset of the metadata elements for indexing; As for AACR and ASCO proceedings abstracts, we took their original texts. The structure of the MEDLINE abstracts is shown in Figures 1 (a) and 1(b). A sample document from the proceedings of AACR is presented in Figure 2.

2.1.2 Clinical Trials

The clinical trials are clinical studies which mainly focus on experimental observation of effect of new treatment. The target document collection for retrieving eligible clinical trials task for given patient cases is a snapshot of clinical trials from ClincalTrials.gov. All the documents in this collection are in XML format with rich metadata about the clinical trials conducted.

[5] http://www.aacr.org/Pages/Home.aspx
[6] https://www.asco.org/
[7] https://clinicaltrials.gov/
[8] http://www.trec-cds.org/2016.html
[9] http://terrier.org/




```
<Pubmedarticle>
   <PMID >...</PMID >
   <Article Title>...</ Article Title >
   <Authors>...</Authors>
   <Affiliations>...</Affiliations>
   <Abstract>...</Abstract>
       <Subsections>...</subsections>
   <MeSH Descriptors>...</MeSH Descriptors>
    <MeSH Qualifiers>...</MeSH Qualifiers >
    <Chemical List>...</Chemical List>
    <References>.../References>
</Pubmedarticle >
```

Figure 1(a): General Structure of MEDLINE Abstract **Figure 1(b): XML Attributes of MEDLINE Abstract**

Meeting: 2012 AACR Annual Meeting
Title: Oncogenic extracellular domain mutations of ERBB2 in cancer

The ERBB2 receptor tyrosine kinase gene is frequently amplified and mutated in human cancer. Using publically-available sequencing datasets, we have found that extracellular domain mutations of ERBB2 located beneath the dimerization arm recur in lung, breast, and ovarian cancers……[cont.]

Figure 2: Sample Abstract from the Proceedings of American Association of Cancer Research

2.2 Topics

The topics given by TREC were 30 synthetic patient cases structured in XML format. For 2017 TREC PM topics, the information from each attribute of a topic is very specific without any narratives. Each topic contains type of cancer, gene/variant information, demographics of the patient and other health problems. A sample topic is shown in the Figure 3.

```
<topic number="1">
<disease>Liposarcoma</disease>
<gene>CDK4 Amplification</gene>
<demographic>38-year-old male</demographic>
<other>GERD</other>
</topic>
```

Figure 3: A Sample TREC 2017 PM Topic

3. REVIEW OF TREC 2016 TRACK METHODOLOGIES

A review of 2016 TREC CDS IR systems was conducted to observe the research methodologies followed by top performing teams. A semi-supervised learning to rank model for re-ranking was observed to be one of the unique re-ranking approaches which uses recently popular techniques like word embedding

(Gurulingappa, et al., 2016; Zhang & Liu, 2016) have evaluated the popular weighting models available in Terrier and observed that In_expC2 had performed the best. Also, usage of Medical Subject Headings[10] (MeSH) terms for synonyms expansion of query terms was observed as a common technique in most systems.

The review of the above discussed IR systems and their unique research methodologies were helpful in adopting some of their techniques and intuitive approaches suitable for the challenges in TREC 2017 PM tasks.

4. EXPERIMENTAL DESIGN

In 2017, our experimental design was an improved version of our TREC 2016 CDS baseline information retrieval system which performs document processing, indexing, query construction and retrieval. Specifically, we added a feedback module called *Relevance Judgment User System* (RJUS) to assess the performance of each run and fine-tune the queries to yield better retrieval results. It also helped us to design the re-ranking strategies. Below we describe each component in more detail. Figure 4 illustrates the experimental design.

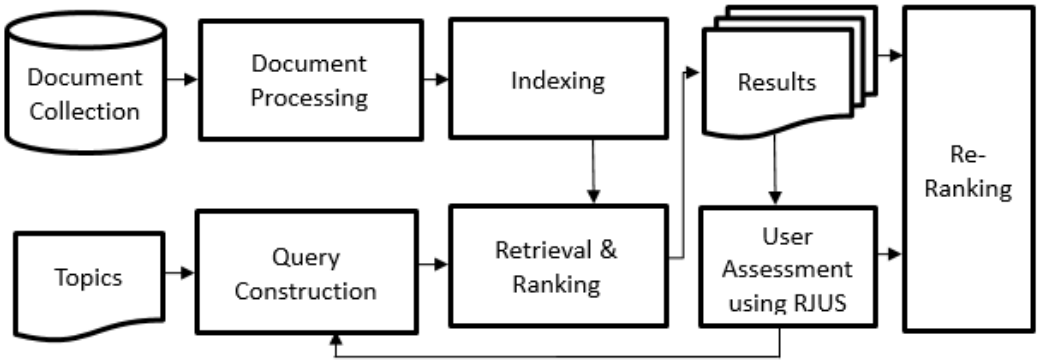


Figure 4: Workflow of our PM System

4.1 Document Processing

In the document processing stage for MEDLINE abstracts, journal title, PMID, article title, abstract, Medical Subject Headings (MeSH) descriptors, MeSH qualifiers and chemical list were extracted using python regular expressions. Similarly, the additional abstracts from proceedings of AACR and ASCO were processed to obtain the article title and abstract. The extracted desired content from all the scientific abstracts was structured and converted into Terrier supported TREC indexing format. The structure of TREC Indexing formatted document collection is shown in Figure 5.

[10] https://meshb.nlm.nih.gov/

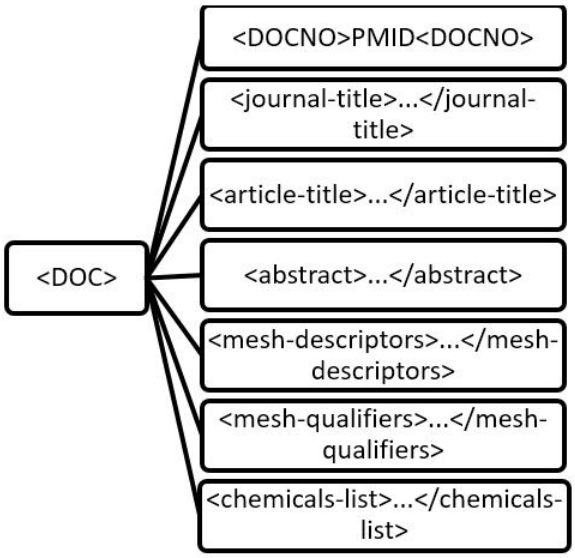


Figure 5: General Structure of TREC Indexing Formatted Scientific Abstract Document Collection

Similarly, for clinical trials, the text information in brief_title, official_title, brief_summary, primary_purpose, eligibility, keywords and MeSH terms attributes were extracted using Python regular expressions and formatted to Terrier supported TREC Indexing format as part of document processing. The structure of TREC Indexing formatted clinical trial document collection is shown in Figure 6.

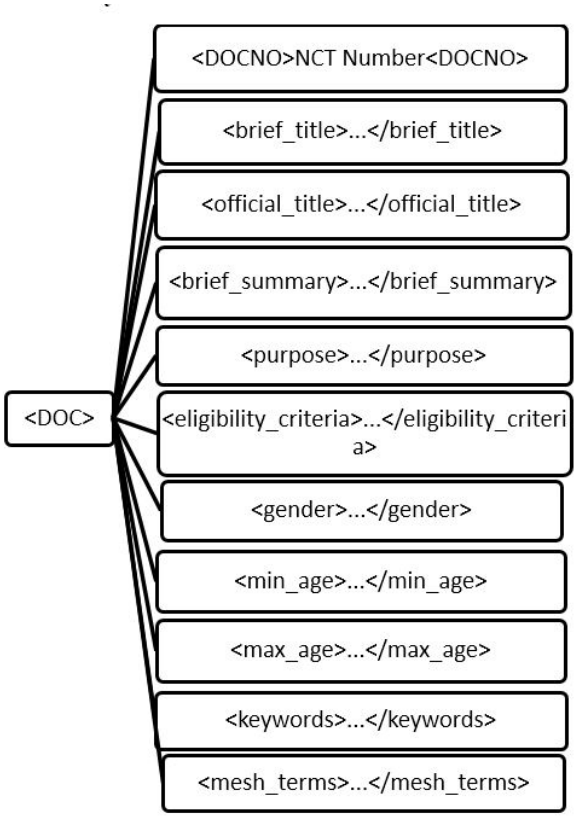


Figure 6: General Structure of TREC Indexing Formatted Clinical Trial Document Collection

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 v4.2 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

Our participation in TREC 2016 gave us a greater understanding that processing the topics and query formulation will have significant influence on the performance of IR system. For this year's tasks, all automatic queries were constructed in single line Terrier query format, a special query format which allows adding custom weights and perform logical operations on query terms. Five types of queries were constructed commonly for both tasks. Using topic 1 in Figure 3 as an example, descriptions of the queries constructed from topics file are as follows:

1) **Query construction with disease weighted terms**
   In this type of queries, all the terms in disease attribute were multiplied by 3 which means the original weight of disease terms were increased by 3.0 times. All the other information like gene terms, demographics were multiplied by 1.0 by default and few generic cancer treatment related terms, such as surgery, radiation therapy and chemotherapy were added to the query.

   Example: 1 Liposarcoma^3.0 CDK4 Amplification 38-year-old male Adult GERD Surgery Radiation therapy chemotherapy Targeted therapy treatment.

2) **Query construction with synonyms of disease terms**
   In this type of queries, along with the disease weighted terms; synonyms of disease terms (like soft tissue cancer in the below example) were obtained from Unified Medical Language System (UMLS) using the MetaMap tool. All the disease terms weights were multiplied by 3.0 times and gene and its variant terms were multiplied by 2.0 times; all other terms were multiplied by 1.0 by default and few generic cancer treatment related terms used in the previous type were repeated in this type of construction.

   Example: 1 Liposarcoma^3.0 soft tissue cancer Neoplasms CDK4^2.0 Amplification^2.0 38-year-old male Adult GERD Surgery Radiation therapy chemotherapy Targeted therapy treatment.

3) **Query construction with Internet search results**
   In this type of queries, along with disease weighted terms and their synonyms; treatment terms specific to the disease in a topic, such as external beam radiation therapy, intraoperative radiation therapy (specific treatments to Liposarcoma) were extracted from search results of the cancer.org[11] portal and were added to the query.

   Example: 1 Liposarcoma^3.0 Soft Tissue Cancer Neoplasms CDK4^2.0 Amplification^2.0 38-year-old male Adult GERD surgery amputation limb-sparing surgery neoadjuvant treatment palliative treatment adjuvant External beam radiation therapy Intraoperative radiation therapy

[11] https://www.cancer.org/

Brachytherapy chemotherapy chemo ifosfamide doxorubicin cisplatin targed therapy Olaratumab (Lartruvo) Pazopanib (Votrient) Imatinib (Gleevec).

4) **Query construction with gene alias terms**

In this type of construction, in addition to the previous type of query terms, aliases of gene terms extracted from PubMed gene[12] search results were added at the query. For instance, in the below example CMM3, PSK-J3 are alias names for CDK4.

Example: 1 Liposarcoma^3.0 Soft Tissue Cancer Neoplasms CDK4^2.0 Amplification^2.0 38-year-old male Adult GERD surgery amputation limb-sparing surgery neoadjuvant treatment palliative treatment adjuvant External beam radiation therapy Intraoperative radiation therapy Brachytherapy chemotherapy chemo ifosfamide doxorubicin cisplatin targed therapy Olaratumab (Lartruvo) Pazopanib (Votrient) Imatinib (Gleevec) cyclin dependent kinase 4 CMM3 PSK J3.

5) **Query construction with Terrier logical query language**

These queries are constructed by modifying the above queries using Terrier query language, using which logical operations are defined between the terms[13]. The query terms include disease weighted terms, gene terms and their expansion, specific treatment related terms.

Example: 1 Liposarcoma^3.0 CDK4^2.0 cyclin dependent kinase 4 Amplification^2.0 38-year-old male Adult GERD Surgery Radiation Chemotherapy Targeted therapy treatment prognosis prognostic therapeutics survival DNA gene repair targeted therapy -diagnosis

Considering disease weighted terms query formation as a baseline query construction, all the other four types of queries are constructed based their performance on extra topics provided by TREC with few relevant documents for reference and the feedback from Relevance Judgment User System (RJUS).

The RJUS is a visual interface we developed that could present retrieval results for human inspection. Through this system, we identified the drawbacks of our current query construction and Terrier configuration. For example, assigning weighted to few terms in a query may reciprocate negative retrieval results, which can be detected only after examining the retrieval results. Based on our human inspection of some of the initial results, we adjusted our query construction algorithm so that important terms were assigned more weights in the constructed queries. These insights were also helpful in designing the re-rank algorithm performed after the retrieval.

The evaluation of the retrieval results in RJUS was performed based on the guidelines released by TREC organized for their internal human assessors. Even though the guidelines are specified only for relevant and irrelevant conditions, partially relevant condition is added in RJUS for our convenience of assessing the run's performance.

4.4 Retrieval and Ranking

In this stage, documents relevant to a given query were retrieved from the collection using a weighting model which scores the degree of relevancy. Terrier was configured to perform the retrieval using the in_expC2 (Inverse Expected Document Frequency model with Bernoulli after-effect and normalization 2) weighting model along with query expansion. In_expC2 is an inverse expected document frequency model for randomness, the ratio of two Bernoulli's process for first normalization, and normalization 2

[12] https://www.ncbi.nlm.nih.gov/gene/

[13] http://terrier.org/docs/v3.6/querylanguage.html

for term frequency normalization (Aghazade, Dehghani, Farzinvash, Rahimi, AleAhmad, Amiri, & Oroumchian, 2008). In_expC2 has been proven to be an effective retrieval model in previous medical information retrieval tasks (Hamdan, Albitar, Bellot, Espinasse, & Fournier, 2012; Cherdioui & Boubekeur, 2013; Zhang & Liu, 2016; Sankhavara & Majumder, 2016).

4.4.1 Query Expansion using Pseudo Relevance Feedback

Relevance feedback has been proved to improve retrieval performance. 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.

In our system, query expansion with pseudo relevance feedback was applied to all the runs using Bose-Einstein model (Bo1) which is also DFR model popularly known as Bose-Einstein model of randomness. Bo1 model is built in Terrier (Amati & Rijsbergen, 2002) with configuration of 20 terms in each of the top 3 documents retrieved in the initial run for expanding the queries. We converged to this configuration based on our inferences drawn from RJUS.

4.4.2. Re-ranking Scientific Abstracts Retrieval Results

After the retrieval results were obtained from Terrier, a re-ranking technique was applied to the retrieval results with the expectation of boosting the system performance through adjusting the order of the retrieved documents. In this technique, a re-arranging methodology was employed based on the availability of desired terms in the retrieval results.

From the relevance judgment guidelines draft from TREC 2017 30th June disease name, treatment, prevention or prognosis related terms and genome data are deciding factors for a document to be relevant. Based on these conditions, a re-ranking technique was designed to rearrange the ordered list such that the documents containing exact or more generic or more specific disease names, treatment/ prevention or prognosis related terms and genome information mentioned in the topic were given higher relevance score.

This re-ranking technique is a two-stage process where each document is assigned with a score in the first stage and ranks are assigned based on the scores given in the previous stage. Only top 100 retrieval results of each topic were processed for re-ranking. A description of each stage is as follows:

Assigning scores to retrieved documents. In this stage, a python script was written to assign scores to each retrieved document based on the availability of desired terms in the abstract. Three types of scores, 'd' score, 'g' score and 't' score, were assigned to a single document. The scores were assigned to each document based on the following process and rules:

- For a given document availability of exact, more generic or more specific form of disease names was searched using Python regular expressions in the first step.
- If the disease name was available in any of the desired form, 'd' is assigned with score '1' for that document.
- Next, availability of gene and variant terms in the abstract was searched and 'g' was assigned with score '2' if both gene and variant terms were in the abstract. 'g' was assigned with '1' if only gene terms were available or assigned with '0' if no gene terms were available.
- Next, treatment related terms were searched in the abstract where on availability of treatment 't' was assigned with score '1' or assigned with '0' if they were not available.

After this process, each document was assigned three scores, which helps to specify their relevance to the query. For example, a document with a combination like d=0, g=2, t=1 is considered irrelevant as the disease name does not appear in the abstract. Similarly, a combination like d=1, g=0/1, t=1 is considered partially relevant due to the absence of gene or variant information.

Re-arranging the scored documents. Upon obtaining the scores and relevant/partially relevant/irrelevant labels were assigned to all the documents. Then the documents were re-arranged in categories such that all the relevant documents were on the top followed by partially relevant documents and then irrelevant documents. In each of the category, such as relevant documents, sorting was performed based on the original relevancy score given by Terrier. Similarly, partially relevant documents were sorted based on their original relevance scores and ranked after the relevant documents and then followed by irrelevant documents.

After re-arranging the top 100 documents IDs for each topic, the remaining 900 document IDs were appended as per the topic number. As the re-ranked list, doesn't hold the descending order of relevance score of the original result file, a normalized score from 1-0 was assigned to all the 1000 re-ranked results in the descending order.

5. OVERVIEW OF RUNS SUBMITTED

We have submitted 10 automatic runs, 5 for scientific abstracts and 5 for clinical trials retrieval tasks. An outline of the runs submitted is discussed in this section.

5.1 Scientific Abstracts Retrieval Task

In total 5 runs were submitted in this task, targeting to retrieve most relevant abstracts dealing with the treatment for cancer types mentioned in queries. An overview of the runs are as follows:

- **Run with queries using disease weighted terms (RunID: UNTIIADW)**
  This run was performed using queries constructed with disease weighted terms and retrieval was configured with query expansion through Pseudo Relevance Feedback (PRF) of 3 documents with 20 terms each and the retrieval results were further re-ranked using the above re-ranking algorithm. The final results file contains top 1000 relevant document IDs in the order of most relevant to least relevant.

- **Run with queries using synonymous terms of diseases (RunID: UNTIIASY)**
  This run was performed using queries constructed with synonyms of diseases in the topics and retrieval was configured with query expansion through Pseudo Relevance Feedback (PRF) of 3 documents with 20 terms each and the retrieval results were further re-ranked using the above re-ranking algorithm. The final results file contains top 1000 relevant document IDs in the order of most relevant to least relevant.

- **Run with queries using terms from Internet search results (RunID: UNTIIAIS)**
  This run was performed using the queries constructed with disease specific treatment terms extracted from Internet search results and retrieval was configured with query expansion through Pseudo Relevance Feedback (PRF) of 3 documents with 20 terms each and the retrieval results were further re-ranked using the above re-ranking algorithm. The final results file contains top 1000 relevant document IDs in the order of most relevant to least relevant.

- **Run with queries using gene alias terms (RunID: UNTIIAGA)**

This run is performed using the queries constructed with gene alias terms extracted from PubMed search results and retrieval was configured with query expansion through Pseudo Relevance Feedback (PRF) of 3 documents with 20 terms each and the retrieval results were further re-ranked using the above re-ranking algorithm. The final results file contains top 1000 relevant document IDs in the order of most relevant to least relevant.

- **Results with queries constructed using logical query language (RunID: UNTIIALQ)**
  For this run, queries were constructed using disease weighted terms and their synonyms, gene terms and their expansion, disease specific treatment terms with Terrier logical query language. Retrieval was configured with query expansion through Pseudo Relevance Feedback (PRF) of 3 documents with 20 terms each and the retrieval results were further re-ranked using the above re-ranking algorithm. The final results file contains top 1,000 relevant document IDs in the order of most relevant to least relevant.

5.2 Clinical Trials Retrieval Task

Similar to the scientific abstracts task, 5 runs were submitted in the clinical trials retrieval task, targeting to retrieve most eligible clinical trials for the given patient's cancer type, genome and demographic information mentioned in the queries. Below is the list of runs submitted.

- Run with queries using disease weighted terms (RunID: UNTIIACTDW)
- Run with queries using synonymous terms of diseases (RunID: UNTIIACTSY)
- Run with queries using terms from Internet search results (RunID: UNTIIACTIS)
- Run with queries using gene alias terms (RunID: UNTIIACTGA)
- Results with queries constructed using logical query language (RunID: UNTIIACTLQ)

Due to time constraints, the fine-tune or more research on clinical data and retrieval runs was not conducted, which negatively affects the performance of these experiments. Below we focus on reporting our results for the scientific abstracts retrieval.

6. RESULTS and DISCUSSION

The performance of TREC 2017 PM retrieval runs was evaluated using the measures Inferred Normalized Discounted Cumulative Gain (InfNDCG), Precision at 10 (P@10) and Precision at R(R_Prec), where "R" is number of relevant documents for a query "Q". The summarized evaluation results of our retrieval runs for scientific abstracts are listed in Table 1. It indicates UNTIIALQ achieved an overall best result, the reason is that this strategy constructed queries based on logical query language – a highly expressive language which contains rich semantic information (Beeri & Kornatzky, 1990).

Table 1: Evaluation Results for Scientific Abstracts Retrieval Runs

Run ID	Inf NDCG	P@10	R_Prec
UNTIIADW	0.3076	0.4467	0.2036
UNTIIASY	0.3110	0.4600	0.1989
UNTIIAIS	0.1183	0.1967	0.0733
UNTIIAGA	0.1248	0.2100	0.0674
UNTIIALQ	0.3163	0.5233	0.1790

We present our scientific abstract retrieval runs as compared with the maximum, minimum, and median values of P@10 for each topic in Figure 7 (a), (b), (c), (d) and (e). In Figure 7, "+" indicates our respective runs, and the "-" indicates the median score for that query. The figure shows that our best run (UNTIIALQ) has gained top performance for 6 topics based on P@10 measure, 12 queries whose P@10 is better than median, 3 queries whose P@10 is equal to median, and 9 queries whose P@10 is worse than median. Contrarily, the worst run (UNTIIAIS) only has 7 queries whose P@10 is better than median, 2 queries whose P@10 is equal to median, but 21 queries whose P@10 is worse than median. By exploring the possible reasons which cause this big difference, we found that the internet search-based query expansion method has produced too much redundant information while enrich the query, which negatively affected the performance. Thus, it may be better to consider the semantic distance between each original query term and generated web-based term to avoid irrelevant terms.

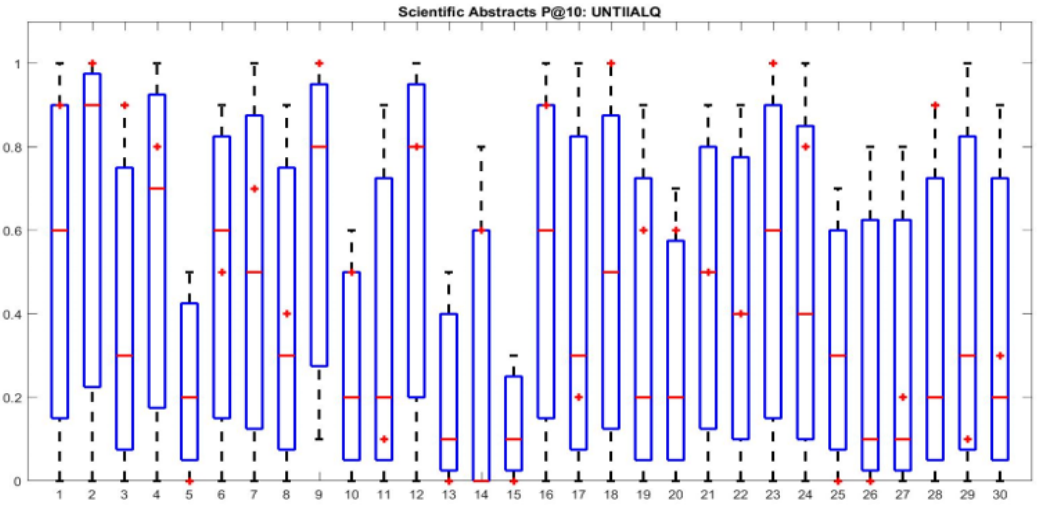


Figure 7(a): Box plot of UNTIIALQ run performance with respect to P@10

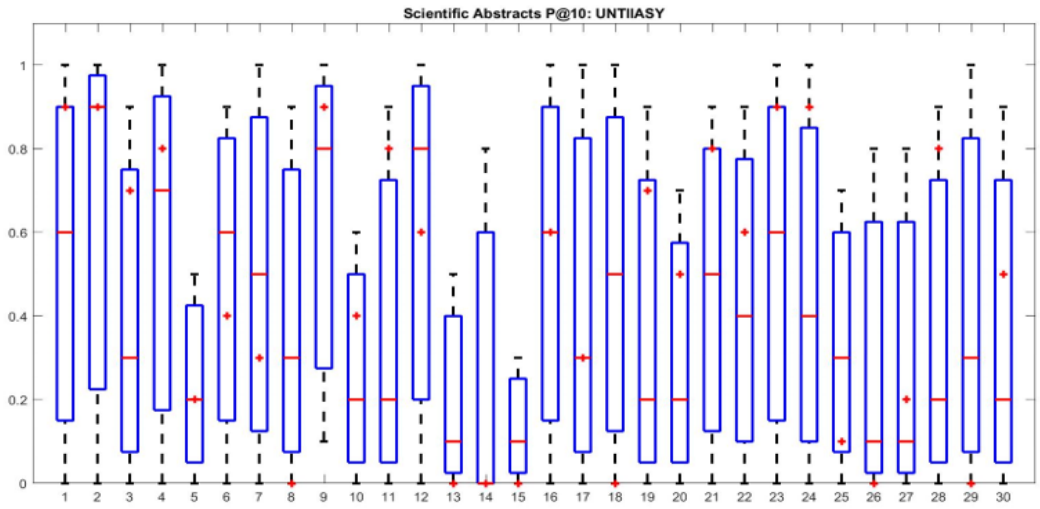


Figure 7(b): Box plot of UNTIIASY run performance with respect to P@10

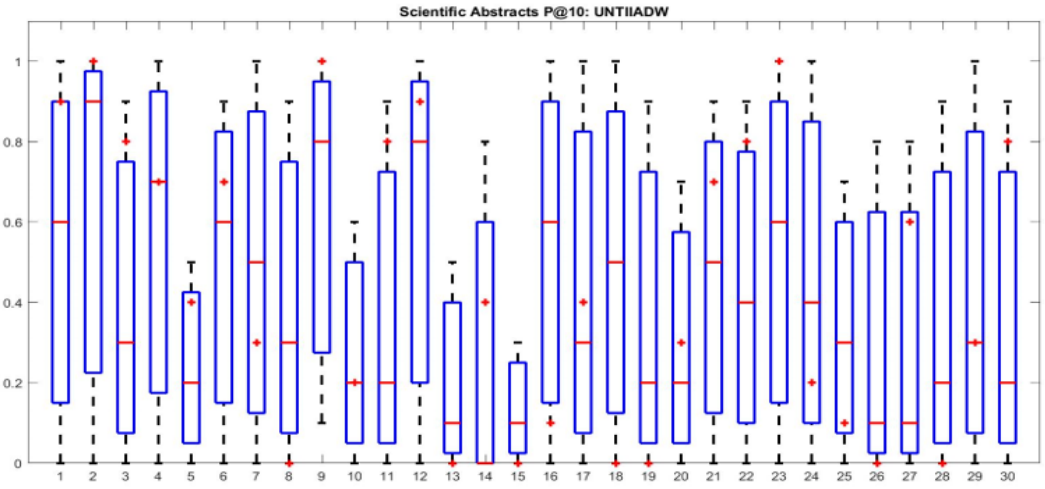


Figure 7(c): Box plot of UNTIIADW run performance with respect to P@10

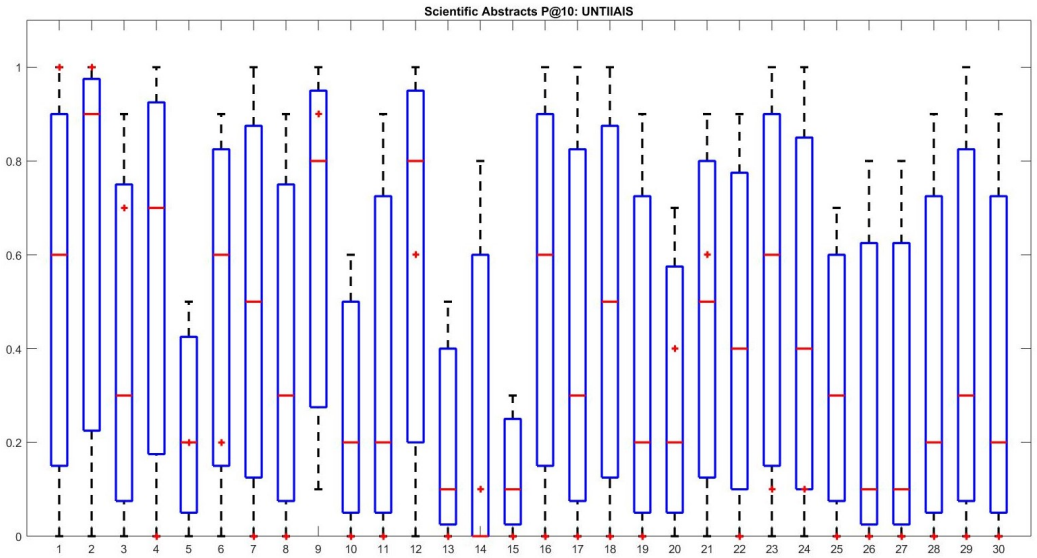


Figure 7(d): Box plot of UNTIIAIS run performance with respect to P@10

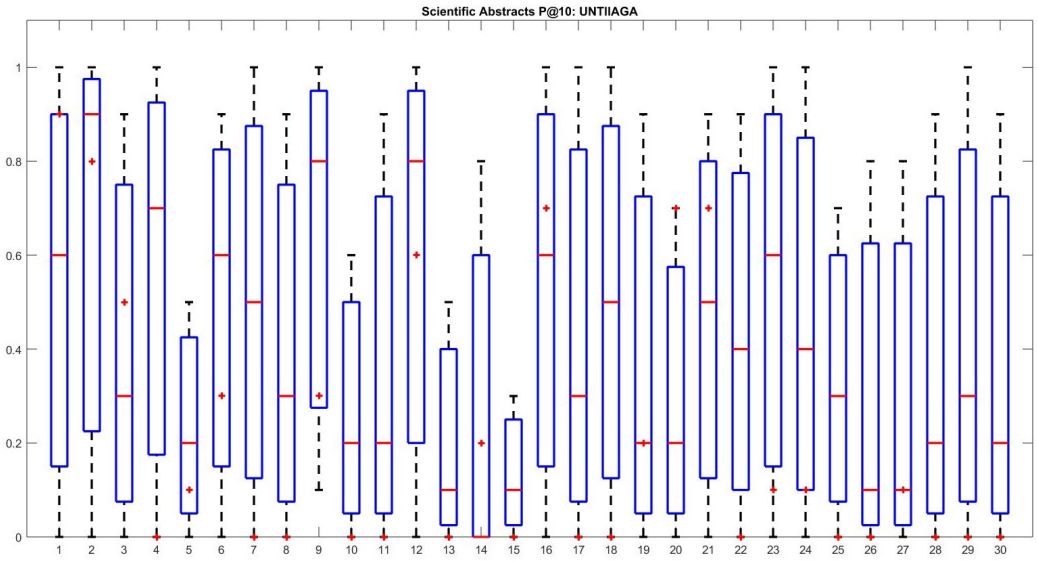


Figure 7(e): Box plot of UNTIIAGA run performance with respect to P@10

To compare performance on different topics, we analyzed the results of each topic over different runs. Clearly, all the query expansion strategies achieved good performance on eight topics (topic 1, topic 2, topic 3, topic 4, topic 9, topic 12, topic 21, and topic 23) and we analyzed the queries generated by these approaches. We found that these queries usually contain more disease-related terms and gene-related terms than other queries after queries expansion, indicating that disease and gene information contribute most to the performance of scientific abstracts retrieval. However, 14 topics (such as topic 5, topic 8, topic 13, topic15, topic 26, etc.) achieve very low performance across IR systems, as we checked the statistic performance over 125 TREC 2017 PM runs submitted. Even the median values of P@10 for these topics were extremely low. It is possible that these topics are more ambiguous and they may have fewer relevant documents.

Our analysis showed that the baseline run UNTIIADW using disease weighted terms improved the overall performance for all the queries. Some other strategies also had positive effects, such as the use of synonyms terms like stem cell growth for gastrointestinal stromal tumor in run UNTIIASY, the UMLS expansion of gene abbreviations (extend NF2 to neurofibromatosis), and adding custom weights to genome terms. Run UNTIIALQ is an intuitive modification of UNTIIASY with logical queries that discarded all the documents with term 'diagnosis' from retrieval results. This approach achieved an overall precision of 0.5233. It may indicate that constructing a logical query with higher level of complexity as well as containing some semantic information improves precision.

The strategies applied to run UNTIIAIS and run UNTIIAGA negatively affected retrieval performance of the system. Both approaches were designed to increase medical terminology of the query terms so more relevant treatment and genome specific documents could be retrieved. They are plausible theoretically, but practically these approached failed to improve the retrieval performance. Subject Matter Experts (SME) may be needed to assess the medical terms that should be incorporated into the queries.

7. SUMMARY AND FUTURE WORK

In this paper, we presented our methodology in performing information retrieval from scientific abstract for TREC 2017 PM Track. We used the open source IR platform Terrier for indexing and retrieval for both the tasks. A special re-ranking technique in combination with the in_expC2 weighting model has been implemented on scientific abstracts retrieval runs to boost its performance. The evaluation results

showed that the overall performance of our scientific abstracts retrieval system is better than the previous version developed for TREC 2016 CDS Track. Our system produced one of the best results in reference with P@10 measure. We have used Terrier configuration of scientific abstracts task for clinical trials retrieval and submitted our results without additional re-ranking and adjusting.

We plan to continue our participation of TREC PM Track next year with the following directions. First, from our previous observation, demographic information is of little effect on the retrieval performance of scientific abstracts but rather important to the clinical trials retrieval. Therefore, we will focus on disease and gene information for query construction in scientific abstract retrieval; Second, we will investigate the role of topic modeling (Sankhavara & Majumder, 2016) for query expansion on pseudo-relevant document through training the topic distribution for each query rather than weighting each term intuitively; Third, we will explore reinforcement learning for re-ranking results, which takes the advantage of fully utilizing the conference evaluation metrics in the training phase. This has been proved to be effective for improving the ranking performance (Wei, Z., Xu, J., Lan, Y., Guo, J., & Cheng, X., 2017).

We will also investigate effective approaches for retrieving relevant clinical trials. Since the clinical trials often focus on general disease rather than a specific cancer, we may first map general disease with its related cancer then conduct matching. The specific type of cancer and its related gene terms will be expanded by appending various synonyms. In the meantime, demographic information maybe used as heuristic rules to filter retrieval results.

Acknowledgements

We would like to thank other members in the UNT Intelligent Information Access Lab for their valuable suggestions. We thank Swathi Kosuna for participating in relevance judgment. The RJUS system is adapted from a similar system developed for the MLIA Project[14], which was funded by the Institute of Museum and Library Services (IMLS) National Leadership Grant LG-06-13-0187-13. We also thank the Department of Information Science for supporting this research.

REFERENCES

Aghazade, Z., Dehghani, N., Farzinvash, L., Rahimi, R., AleAhmad, A., Amiri, H., & Oroumchian, F. (2008). Fusion of retrieval models at CLEF 2008 ad hoc Persian track. *In Workshop of the Cross-Language Evaluation Forum for European Languages* (pp. 97-104). Springer, Berlin, Heidelberg.

Amati, G., & Van Rijsbergen, C. J. (2002) Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Transactions on Information Systems 20(4):pp. 357- 389.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32(Database issue), D267–D270.http://doi.org/10.1093/nar/gkh061

Beeri, C., & Kornatzky, Y. (1990). A Logical Query Language for Hypertext Systems. *In ECHT* (Vol. 90, pp. 67-80).

Cherdioui, S., & Boubekeur, F. (2013). Information retrieval techniques for knowledge discovery in biomedical literature. *In Programming and Systems (ISPS), 2013 11th International Symposium on*

[14] http://iia01.ci.unt.edu/MLIA/

(pp. 137-142). IEEE.

Gurulingappa. H., Bauer. A., Toldo. L., Schepers. C., & Megaro. G. (2016). Semi-Supervised Information Retrieval System for Clinical Decision Support. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2016) Proceedings*. Paper presented at the Twenty-Fifth Text REtrieval Conference, Gaithersburg, MD.

How to Use MeSH, (2018, February 12.), Retrieved from http://research.library.gsu.edu/c.php?g=115556&p=753156

Hamdan, H., Albitar, S., Bellot, P., Espinasse, B., & Fournier, S. (2012). Lsis at trec 2012 medical track-experiments with conceptualization, a dfr model and a semantic measure. AIX-MARSEILLE UNIV (FRANCE).

Macdonald. C., McCreadie. R., Santos, R. & Ounis, I. (2002) From Puppy to Maturity: Experiences in Developing Terrier. *In Proceedings of the SIGIR 2012 Workshop in Open Source Information Retrieval*.

Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C. & Johnson, D. (2005) Terrier Information Retrieval Platform. *In Proceedings of the 27th European Conference on Information Retrieval*.

Sankhavara, J., & Majumder, P. (2016). Team DA IICT at Clinical Decision Support Track in TREC 2016: Topic Modeling for Query Expansion. *In TREC*.

Viswavarapu, L., Chen, J., Cleveland, A., & Philbrick, J. (2016). UNT Medical Information Retrieval at TREC 2016. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2016) Proceedings*. Paper presented at the Twenty-Fifth Text REtrieval Conference, Gaithersburg, MD.

Wei, Z., Xu, J., Lan, Y., Guo, J., & Cheng, X. (2017). Reinforcement Learning to Rank with Markov Decision Process. *In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 945-948). ACM.

Zhang, H., & Liu, L. (2016). NKU at TREC 2016: Clinical Decision Support Track. In E.M. Voorhees & A. Ellis (Eds.). *NIST Special Publication: The Twenty-Fourth Text REtrieval Conference (TREC 2016) Proceedings*. Paper presented at the Twenty-Fifth Text REtrieval Conference, Gaithersburg, MD.