

# Towards an Aspect-based Ranking Model for Clinical Trial Search

Soumyadeep Roy<sup>1</sup>, Koustav Rudra<sup>2</sup>, Nikhil Agrawal<sup>1</sup>, Shamik Sural<sup>1</sup>, and Niloy Ganguly<sup>1</sup>

<sup>1</sup> Indian Institute of Technology Kharagpur, India  
soumyadeep.roy9@iitkgp.ac.in, nikhilagrawal741@gmail.com,  
{shamik,niloy}@cse.iitkgp.ac.in

<sup>2</sup> L3S Research Center, Hannover, Lower Saxony, Germany  
rudra@l3s.de

**Abstract.** Clinical Trials are crucial for the practice of evidence-based medicine. It provides updated and essential health-related information for the patients. Sometimes, Clinical trials are the first source of information about new drugs and treatments. Different stakeholders, such as trial volunteers, trial investigators, and meta-analyses researchers often need to search for trials. In this paper, we propose an automated method to retrieve relevant trials based on the overlap of UMLS concepts between the user query and clinical trials. However, different stakeholders may have different information needs, and accordingly, we rank the retrieved clinical trials based on the following four aspects – **Relevancy**, **Adversity**, **Recency**, and **Popularity**. We aim to develop a clinical trial search system which covers multiple disease classes, instead of only focusing on retrieval of oncology-based clinical trials. We follow a rigorous annotation scheme and create an annotated retrieval set for 25 queries, across five disease categories. Our proposed method performs better than the baseline model in almost 90% cases. We also measure the correlation between the different aspect-based ranking lists and observe a high negative Spearman rank’s correlation coefficient between popularity and recency.

**Keywords:** clinical trial search · aspect-based ranking · biomedical information retrieval

## 1 Introduction

In recent years, the internet is being increasingly used as a source of health information by both medical (professionals, practitioners) and non-medical (consumers) users. Nowadays, various clinical database sources such as ClinicalTrials.gov, PubMed, Embase and Cochrane are publicly accessible, which makes health information search easier for the end-users. In this paper, we focus on the retrieval of clinical trials because clinical trials are crucial for the practice of Evidence-Based Medicine and are used for establishing the efficacy of new drugs and treatments. Apart from that, the need for different stakeholders varies a lot.

For example, medical practitioners may be interested in trials containing important information about new medicines and drugs. Users who want to participate in a clinical trial may look for the adversity score of a clinical trial (to judge the consequences of participation). It reveals that different users have various perspectives or aspects. The objective of this aspect-based trial search is two-fold. Firstly, it helps users in satisfying specific information needs, and secondly, it prevents users from taking wrong decisions. For example, it may be the case that a trial with a high number of adverse events, is ranked as highly relevant. However, it is not suitable from a participant point of view. If the user intends to participate in a trial, such results are disastrous. Based on the analysis of the data, we identify the following four essential aspects of the clinical trial search: (1) Relevancy, (2) Recency, (3) Adversity, and (4) Popularity. Goodwin et al. [8] have used aspect-based approaches (on a query) to improve the retrieval and ranking performance of clinical trials. In this work, we explore different notions of ranking aspects or relevance criteria for clinical trials search.

In this paper, we develop a novel and straightforward two-step method to retrieve and rank clinical trials. First, clinical trials are retrieved based on free-form text query given by ordinary users having less familiarity with medical terms. In the second phase, retrieved trials are ranked based on four aspects, as mentioned earlier (relevancy, recency, adversity, and popularity).

We develop a Synset Term Match-based Clinical Trial Retrieval model (STM) to retrieve relevant trials for the user given queries. For a given query, we first extract the UMLS concepts present in a query using *QuickUMLS* [26]. We are considering UMLS concepts instead of raw text to handle two issues. Firstly, users may give different variations of a single term, and secondly, users having limited knowledge of medical terminology may make spelling mistakes. QuickUMLS tool takes care of these issues. Next, we retrieve concepts from the clinical trials and retrieve the trials based on their match with the query related concepts. Finally, we rank the trials based on four different aspects. We evaluate our proposed method over 25 queries taken from five different disease classes - *Pathological Conditions Signs and Symptoms*, *Cardiovascular Diseases*, *Nervous System Diseases*, *Nutritional and Metabolic Diseases*, *Immune System Diseases*. However, we do not have any annotated set of ground truth labels for the queries of these disease classes. Hence, in this paper, we put extensive effort to prepare the ground truth set for these 25 queries (explained in Section 5.1). The complete set of data (queries and the set of annotated relevant trials for each query) and code files are publicly available in Github<sup>3</sup>.

Detailed experiment over five different disease classes reveals the efficacy of our proposed *Synset based term match* model (STM) over the baseline. For *relevancy* aspect, STM is performing better than the baseline in 90% cases and achieves a precision@5 value of 0.56 as compared to 0.12 by the baseline model. Extracting concepts from the queries and trials help in improving the coverage because users provide different variations of text as input. Side by side, we observe high statistically significant negative correlation exists between *recency*

<sup>3</sup> [https://github.com/nikhil741/COCTR\\_multidimensional\\_ranking](https://github.com/nikhil741/COCTR_multidimensional_ranking)

and *popularity*. This result is quite evident since a paper published earlier is likely to get a higher number of citations.

The rest of the paper is organized as follows. We perform the literature review in Section 2. Section 3 provides details about the clinical trial dataset. We discuss our proposed method in Section 4. Experimental results are elaborated in Section 5. Finally, we discuss the limitations of this work, future directions and conclude the paper in Section 6.

## 2 Related Work

Health-related information mining has gained much attention nowadays. Researchers focused on different kinds of tasks, such as biomedical information retrieval, clinical trial search, and query formulation strategies. In this section, we provide a brief overview of these tasks.

**Biomedical information retrieval and Health information search:** Nowadays, an abundant source of clinical repositories such as PubMed, Embase, ClinicalTrials.gov are publicly available to satisfy user requirements. Clinical trials are crucial for the practice of Evidence-Based Medicine and are used to establish the effectiveness of new drugs and treatments. Systematic reviews are time-consuming and usually take 9 to 12 months to complete. Hence, results are often outdated, and it is necessary to develop automated techniques to generate such reviews [30]. This field of research is greatly enhanced by the introduction of new datasets like MeSH-based retrieval set [13] and biomedical retrieval challenges such as bioCADDIE [1], BioASQ [2], and BioCreative Precision Medicine Track [5]. Recently, TREC [22] started a precision medicine track to deal with clinical trial retrieval challenges.

Previous works have explored and provided a taxonomy of the type of consumer health questions asked [20] and the complaints made by patients [12]. Patel et al. [19] cover the kind of clinical trial information a user searched. Generally, ordinary users face difficulty in formulating effective queries because it may involve complicated medical terminologies.

**Aspect based information extraction in medical domain:** Felix et. al [9] explored two different aspects (effectiveness and side-effects) of the drug reviews. Cavalcanti et. al. [7] proposed a syntactic tree based review analysis to classify reviews into four different aspects - *Condition, Side Effects, Dosage, Effectiveness*. Prior works used eligibility criteria to develop information management systems [15,11], correlate adverse events [25], and cluster trials [27].

**Clinical trial search systems:** Several recent studies focus on searching clinical trials to improve the automation strategies for the generation of systematic review [30]. Online trial search interfaces such as ClinicalTrials.gov, WHO ICTRP, EmergingMed.com, SearchClinicalTrials.org, and UK Clinical Trials Gateway, provide users with an option to search for their requirements. Most of the commercial trial search engines such as eTACTS [17], Antidote<sup>4</sup> are disease-specific and allow users to search for trials related to specific disease types [10]. In the context of consumer health search, Zuccon et. al [31] recently provided

<sup>4</sup> <https://www.antidote.me>

certain empirical insights to aid the development of future knowledge base based health information retrieval framework.

**Aspect based approaches in clinical trial search:** Goodwin et al. [8] used six aspects – disease, genetic, demographic, precision medicine, treatment and other medical problem aspects, for developing an information retrieval system for PubMed and clinical trials. For TREC 2017, MedIER team used query expansion strategies and leveraged medical ontologies [29], while in TREC 2018, they developed a system based on query generation and document re-ranking [4].

### 3 Dataset

We collected the dump of clinical trials from ClinicalTrials.gov on 12/01/2019. The total number of clinical trials present in XML format was 294,679. However, all the trials do not have a corresponding PubMed entry and Medical Subject Headings (MeSH) terms [14]. MeSH is “*the National Library of Medicine’s controlled vocabulary thesaurus, used for indexing articles for the MEDLINE or PubMed database.*” Hence, we only consider trials for which have an associated PubMed entry and at least one MeSH term. After this step, we are left with 35,204 trials. Next, we map these trials into 26 disease categories as reported in the MeSH database.

**Mapping trials to diseases:** For mapping a clinical trial into disease classes, we perform the following steps. Firstly, we extract the MeSH terms from a trial. Finally, we map the MeSH terms to different diseases using the disease trees of MeSH thesaurus. In general, we find out that all the diseases are present at the root of the trees. The clinical trials are mapped into 26 different disease classes, and a clinical trial may be mapped to more than one disease class. For example, we observe that a clinical trial (*NCT00000106*) is mapped to *Musculoskeletal Diseases* and *Skin and Connective Tissue Diseases* because the trial consists of MeSH terms about both *Rheumatic Diseases* and *Collagen Diseases*. We rank the diseases based on the number of trials they contain. Finally, we consider the top five diseases for this study – (1) Pathological Conditions, Signs and Symptoms (12826 trials); (2) Cardiovascular Diseases (7293 trials); (3) Nervous System Diseases (6172 trials); (4) Nutritional and Metabolic Diseases (5240 trials); (5) Immune System Diseases (5016 trials).

## 4 Method

As mentioned in Section 1, the objective of this work is to retrieve and rank the trials across four different aspects — (1) Relevancy, (2) Adversity, (3) Recency, and (4) Popularity. In this section, we describe our ranking method in details.

### 4.1 Clinical trial retrieval

In this section, we describe our clinical trial retrieval framework. There are two components in the retrieval framework – (1) Concept extraction, and (2) Match based retrieval. After retrieving relevant trials, we rank them across four different aspects (described in the next part).

**1. Concept extraction:** We extract UMLS medical concepts from a trial’s *brief title* and *brief summary* using an unsupervised, scalable medical concept

extraction tool, *QuickUMLS* [26]. We also represent a query in terms of its extracted UMLS medical concepts, following the same methodology.

**2. Match based retrieval:** After the concept extraction step, each query is represented by a set of UMLS concepts. For a given query  $q$ , we retrieve all the clinical trials whose brief title contain all the UMLS concept ids which are present in the query  $q$ .

#### 4.2 Aspect based ranking of clinical trials

**1. Ranking based on Relevancy:** After retrieving clinical trials, we rank the clinical trials based on relevancy. We derive three different relevancy based ranking measures, as explained below.

**1.1 PageRank (PGR):** For ranking, we first create an undirected graph  $G(V, E)$  where vertices are the clinical trials that we have retrieved for a given query. For providing edge weights between  $(V_i, V_j)$  vertices, we measure Simpson similarity between clinical trials in terms of UMLS concepts extracted from *brief title* and *brief summary* fields of a clinical trial. ‘*Simpson similarity*’ between 2 sets, is defined as the ratio of their cardinality of intersection and the cardinality of the smaller set. Next, we apply PageRank [18] algorithm on the graph  $G$  to compute the importance of each trial. Finally, the clinical trials are ranked based on their PageRank score.

**1.2 Exact term match (ETM):** In the retrieval phase, we only focus on the UMLS concepts present in the query or the title of a trial. However, we observe that sometimes other important terms are also present in the query for which is not mapped to any UMLS concept. We analyze a lexicon containing 1440 commonly used queries by patients and observe that for 15% of the queries, QuickUMLS is unable to extract medical concepts. We also focus on such terms for ranking along with our concept based retrieval and ranking. First, we remove stopwords and perform stemming of the remaining words present in the query. The similar preprocessing technique is also applied to the *brief summary*, *brief title* and *official title* fields of a given clinical trial. We compute the count of such processed query words present in the three fields of a clinical trial. Finally, we rank the trials based on term frequency instead of the standard TF-IDF, because we assume each of the remaining terms of the query to be of equal importance. We observe that most of these terms are either a part of the extracted UMLS concept or strongly affect the meaning of the query (the length of a query is four words on average). We rank the trials in the following manner – (1) trials are ranked based on the frequency of those terms in the brief summary, (2) trials for which terms are not present in the brief summary, we measure their frequency in the official title field of those trials, (3) in case of mismatch in both brief summary and official title, trials are ranked based on the count in brief title, (4) trials for which terms are absent in all the three fields, they are ranked based on their PageRank score.

**1.3 Synset based term match (STM):** We directly match the terms in case of our previous model, ETM. However, we observe that lots of variations are present in the given query and the retrieved clinical trials. Instead of performing exactly term-wise mapping, for a given query, we extract synset of those terms

from WordNet [16] before matching. Therefore, we now compute the total count of the presence of all the terms (present in a synset) in the brief summary, official title and brief title fields of a trial. Finally, we rank the trials in the decreasing order of brief summary count, official title count, brief title count, and PageRank score, as explained in the previous part.

**2. Ranking based on Adversity:** As mentioned in Section 1, *adversity* is another important aspect of the clinical trial search system. This aspect may be mapped to the ‘*safety events*’ category of patient complaints, where ‘adverse events’ is one of its sub-categories [20], which helps them to decide whether to participate in the trial. According to ClinicalTrials.gov, an adverse event is defined as “*any unfavourable change in the health of a participant, including abnormal laboratory findings that took place immediately or within a certain point of time after the study has completed.*” The adversity report of clinical trials is accessible in a publicly available database called Aggregate Analysis of ClinicalTrials.gov (downloaded on 03/02/2019). For this analysis, we mainly consider information present in the ‘Reported events’ table, which contains the adverse event information (for example, the number of participants affected) of each arm of a clinical trial. A clinical trial may have multiple arms who are given different or no treatment (an arm represents a specific group of trial participants).

After extracting relevant trials (discussed in Section 4.1), trials are ranked in the following manner – (1) trials are ranked in decreasing order based on *Subjects Affected* field. (2) the zero value of *Subjects Affected* field indicates that the trial does not have any adverse reported events. We place such trials at the end of the list in random order.

**3. Ranking based on Recency:** Users may have different objectives. Some may want to get enrolled in a clinical trial, and some are looking for new treatments or information. Sometimes, existing drugs or treatment methods do not work well for some patients. New inventions may help medical practitioners to handle such critical patients. Depending upon the study, the length of a clinical trial may vary; as a result, the information present may become outdated. A systematic review is a very time-consuming process and on average, takes 9-12 months. Hence, by ranking clinical trials based on completion date users can get information about new updated information about drugs, treatment, and therapies.

Completion date is reported for trials which are already completed; otherwise, the future date on which it is going to be completed is provided. The trials for which date information is missing, we consider the first day of the corresponding month. We observe that most of the trials which are going to be completed in future did not report any tested information or drugs. Hence, such trials are not useful for medical practitioners and discarded. We only consider completed trials for recency-based ranking.

**4. Ranking based on Popularity:** From July 2005, each of the completed trials has to make an entry in the PubMed article database to increase its vis-

ibility. This also, in turn, helps to improve the readability. It also provides an opportunity for researchers to cite past related trials. In this section, we try to measure the success of a clinical trial based on the *popularity* of its corresponding PubMed article. In general, the citation count of a paper may be considered as a proxy to determine the importance of it in the community [6]. Hence, we map the clinical trials with a PubMed entry to measure its popularity. After the mapping, we find out the number of PubMed articles that cited a given article. We use the rest API [24] service provided by ‘*NCBI E-utilities*’ to find out the citing articles. Finally, we rank the retrieved clinical trials based on decreasing popularity value (citation count). We break all ties based on the relevancy score (Section 4.2).

## 5 Experimental Setup and Results

We now describe our experimental setup and then evaluate our proposed retrieval and ranking technique for clinical trial search. We further compare the different aspect-based ranking and discuss the results. The source code and the data are available in Github<sup>5</sup>.

### 5.1 Experimental Setting

Here, we explain the formation of the queries, the evaluation metrics and baseline systems used for our experimental setup.

**Query Preparation:** To evaluate the performance of our proposed method, we prepare a set of five queries for each of the five disease classes. We follow the semantic-based query templates proposed by Patel et al. [19] to prepare the queries, the most frequent template being *disease or syndrome + research activity*. We prepare the queries based on the following templates - (1) (disease or syndrome) + (symptom or treatment), (2) disease + age group, (3) disease + safety information. We specifically do not consider location and gene information, which are also popular consumer query variants. We consult multiple patient or health-related lexicons such as MedDRA<sup>6</sup>, CLEF Consumer Health track [3], Reddit etc. to formulate the query terms.

**Evaluation Metric:** We have evaluated results based on two standard IR metrics, i.e. Precision and nDCG score. However, we can not measure recall due to unavailability of ground truth set of clinical trials for each query. Three annotators<sup>7</sup> manually annotated all the retrieved trials for each of the 25 queries. On average, 80 trials are annotated per query. They followed the ‘Definitely relevant’ annotation scheme for TREC Precision Medicine 2018 task [23].

**Baseline:** Most of the existing clinical trial retrieval systems focused only on a particular class of disease. Text REtrieval Conference Precision Medicine Track (TREC-PM) has a similar task of retrieving relevant clinical trials but focus only on oncology trials. However, the state-of-the-art systems of the “TREC-PM 2017 Task B” (clinical trial retrieval) have either not published their code-bases, or have used cancer-specific medical ontologies. This makes it very difficult

<sup>5</sup> [https://github.com/nikhil741/COCTR\\_multidimensional\\_ranking](https://github.com/nikhil741/COCTR_multidimensional_ranking)

<sup>6</sup> <https://www.meddra.org/patient-friendly-term-list>

<sup>7</sup> None of them is an author of this paper and has good knowledge of English.

for applying to other disease classes. We aim to develop a clinical trial retrieval model, which we may apply to multiple disease classes. Hence, we consider the system proposed by Ajinkya Throve [28] in the TREC 2017 Precision Medicine Track [21]. To the best of our knowledge, this is the only system which does not use any disease-specific knowledge bases and has made their codebase publicly available.

## 5.2 Performance Evaluation

We report the performance of our three relevancy based ranking methods (PGR, ETM, STM) and baseline in Table 1. We compute the mean precision values at 5, 10, 15 and 20 for all the 25 queries. We observe that **STM** outperforms all the other methods.

**Table 1.** Mean precision values for different relevancy based ranking methodologies across 25 queries

Method	P@5	P@10	P@15	P@20
Baseline	0.12	0.08	0.08	0.08
PGR	0.38	0.35	0.35	0.33
ETM	0.53	0.48	0.45	0.42
STM	<b>0.56</b>	<b>0.52</b>	<b>0.47</b>	<b>0.46</b>

We now provide a detailed comparison between STM and the baseline method (BAS). We observe that STM shows significant improvement for 10 such queries. However, STM achieves precision@10 value of less than 0.31 for 28 per cent of cases, which may be due to limitation during the retrieval stage. Next, we study the queries having precision@10  $\leq 0.3$ , with no or marginal improvement across PGR, ETM and STM. We observe that it may be because our model does not consider the prior history or eligibility criteria of trials into account. For example, in certain trials, we need users who already have a specific disease, like *treating people already having hypertension (CVD)*, *already having Celiac disease (NMT)*. Users may also search for trials with safety information, i.e. without having any subjects affected with any form of adverse effects, like *safe treatment for Alzheimer’s disease (NER)*, *hypercholesterolemia safe treatments (NMT)*.

We also report the nDCG scores of STM model in Table 3. We observe from Table 2 that BAS is only able to retrieve at least five trials, for 3 out of 25 queries and has all precision values as 1.0, outperforming STM in such cases. BAS is based on exact lexical matching between a query and brief title of a clinical trial, and therefore will always be relevant when retrieved. However, there is much variation in the query terms, and direct matching is not possible. In such cases, dealing with a UMLS concept is a useful option.

## 5.3 Comparison of different Aspect-based Rankings

In the previous section, we measure the performance of our relevancy based ranked search results. However, as mentioned in Section 1, the primary objective of this work is to provide users with a multi-dimensional ranked list of clinical trials. In this section, we compare the different list of trials based on different aspects and try to understand whether any form of relationship exists among them. Table 4 shows overlap score and Spearman’s rank correlation (SR) among different list pairs. Overlap score (OV) is computed as the total number of trials



**Table 2.** Performance evaluation of STM and Baseline (BAS). “-” indicates model fails to retrieve any trials. We represent the 5 disease classes — (1) PAT: Pathological Conditions, Signs and Symptoms; (2) CVD: Cardiovascular Disease; (3) NER: Nervous System Disease; (4) NMT: Nutritional and Metabolic Diseases; (5) IM: Immune System Diseases

Disease	Query	P@5		P@10		P@15		P@20	
		BAS	STM	BAS	STM	BAS	STM	BAS	STM
PAT	Q1	-	1.0	-	0.8	-	-	-	-
	Q2	-	0.2	-	0.4	-	0.4	-	0.35
	Q3	-	1.0	-	0.5	-	0.33	-	0.3
	Q4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
	Q5	-	0.6	-	0.7	-	0.6	-	0.45
CVD	Q1	-	0.6	-	0.6	-	0.53	-	0.65
	Q2	-	0.0	-	0.0	-	0.07	-	0.1
	Q3	-	0.2	-	0.3	-	0.33	-	0.25
	Q4	1.0	0.6	1.0	0.8	1.0	0.8	1.0	0.8
	Q5	1.0	0.2	-	0.1	-	0.13	-	0.15
NER	Q1	-	0.2	-	-	-	-	-	-
	Q2	-	0.8	-	0.6	-	0.53	-	-
	Q3	-	0.6	-	0.5	-	0.4	-	0.4
	Q4	-	0.4	-	0.4	-	0.27	-	0.25
	Q5	-	0.8	-	0.4	-	0.27	-	0.2
NMT	Q1	-	0.4	-	0.3	-	0.2	-	0.35
	Q2	-	0.2	-	0.3	-	0.2	-	-
	Q3	-	0.0	-	0.2	-	0.4	-	-
	Q4	-	1.0	-	1.0	-	0.93	-	0.95
	Q5	-	0.6	-	0.7	-	0.67	-	-
IM	Q1	-	0.4	-	0.4	-	0.33	-	0.35
	Q2	-	1.0	-	0.8	-	0.73	-	0.6
	Q3	-	1.0	-	1.0	-	1.0	-	1.0
	Q4	-	0.6	-	0.4	-	-	-	-
	Q5	-	0.6	-	0.3	-	0.2	-	0.25

**Table 3.** nCDG scores for STM across 25 queries

Query No.	PAT	CVD	NER	NMT	IM
Q1	0.97	0.88	0.5	0.64	0.8
Q2	0.77	0.65	0.92	0.62	0.95
Q3	0.96	0.66	0.72	0.54	0.96
Q4	1.0	0.88	0.66	0.99	0.71
Q5	0.84	0.52	0.98	0.91	0.8

which intersect between the ranked list of two aspects (up to first 20 ranks). It is clear from Table 4 that high statistically significant negative correlation exists between ‘recency’ and ‘popularity’, which is quite obvious. We study the queries (Q3, Q12, Q17, Q18, Q20) which have an overlap@20 score of *greater than 15* (out of 20), among most aspect-based ranked list pairs. This is because on average, they retrieve only 18 trials.

**Application in a real-life setting:** Instead of four separate ranked lists, a single ranked list that considers all the aspects proves to be more useful in a real-life setting. A simple ranking scheme for combining them may be as follows – (1) rank them using STM model and fix the top-K (top 20, in our case) trials. (2) sort using popularity (higher the popularity score, higher up the list). (3) Perform stable sorting in non-decreasing order of adversity (lower the number of

‘subjects affected’, higher up the list) because only a few trials have a positive adversity score.

**Table 4.** Comparison of different aspect based rankings in terms of overlap (OV) and Spearman rank correlation(SR)

Aspects	Relevancy		Adversity		Recency		Popularity	
	OV	SR	OV	SR	OV	SR	OV	SR
<b>Relevancy</b>	20	1.0	11	0.14	8	-0.04	10	0.11
<b>Adversity</b>	11	0.14	20	1.0	9	0.03	10	0.12
<b>Recency</b>	8	-0.04	9	0.03	20	1.0	7	-0.41
<b>Popularity</b>	10	0.11	10	0.12	7	-0.41	20	1.0

## 6 Concluding Discussions

In this paper, we introduce the concept of the multi-dimensional ranking of clinical trials in terms of **adversity**, **popularity** and **recency**, along with **relevancy**, with the idea of addressing the different information needs of various stakeholders associated with clinical trial search. We follow a rigorous annotation scheme and create an annotated retrieval set for 25 queries, belonging to 5 different disease categories. Our proposed multi-dimensional ranking model, Synset based term matching model achieves a precision@5 value of 0.56 and outperforms the baseline in more than 90% cases. Further, we explore the limitations of our model by testing it over an oncology-related benchmark gold standard data and report those limitations with proper justifications.

**Limitations:** In this paper, we have analyzed the ranked result for five different diseases. However, in all these cases, we only rely on the search term given by the user. Currently, we test our approach using only 25 queries across 5 disease classes, which does not capture all acronyms and microtext variations of a query. In some cases (e.g., cancer-related diseases), users also provide criteria along with the search terms. For example, in CLEF task, we have found that queries contain gender and age information along with the search terms to find out relevant clinical trials for a patient. We also observe that in specific topics, a description is provided instead of the specific disease. In this paper, we only focus on the brief title and brief summary of trials. It is also necessary to look into the eligibility criteria of a trial which is composed of two parts. One is *inclusion criteria*, which specifies the requirements for a person to be eligible for a trial. The second one is *exclusion criteria*, which prevents a person from participating in the study. This may significantly improve the performance of our model.

**Future work:** We observe that Synset based term matching (STM) performs retrieval and ranking well at a lexical level but performs poorly when significant semantic information is required. We can leverage the publicly available generic knowledge bases like PreMedKB, PharmaGKB, LifeMap Integrated Knowledgebase, NCBI Human Gene Database and MalaCards Human Disease Database. In this paper, we propose different aspect based ranking lists. However, in a real-life setting, we require a single ranked list. In future, we will apply more sophisticated aspect fusion techniques [8] to produce a single combined ranked list or may model it as a multiple-criteria decision-making problem.

## 7 Acknowledgements

The work is supported in part by the SERB-funded project titled “IoTDiff: Service Differentiation for Personal Hubs”, the IMPRINT-funded project titled “Development of a Remote Healthcare Delivery System: Early Diagnosis, Therapy, Follow-up and Preventive Care for Non-communicable Diseases (Cardio-pulmonary)”, and the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No. 832921.

## References

1. Gururaj et al, A.E.: A publicly available benchmark for biomedical dataset retrieval: the reference standard for the 2016 bioCADDIE dataset retrieval challenge. *Database* **2017** (08 2017). <https://doi.org/10.1093/database/bax061>
2. Tsatsaronis et al, G.: An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. *BMC Bioinformatics* **16**(1), 138 (Apr 2015). <https://doi.org/10.1186/s12859-015-0564-6>
3. Suominen et al, H.: Overview of the clef ehealth evaluation lab 2018. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. pp. 286–301 (2018). [https://doi.org/10.1007/978-3-319-98932-7\\_26](https://doi.org/10.1007/978-3-319-98932-7_26)
4. et al, J.L.: Retrieving scientific abstracts iteratively: Medier at trec 2018 precision medicine track (2018), <https://trec.nist.gov/pubs/trec27/papers/MedIER-PM.pdf>
5. Islamaj Doğan et al, R.: Overview of the biocreative vi precision medicine track: mining protein interactions and mutations for precision medicine. *Database* **2019** (2019). <https://doi.org/10.1093/database/bay147>
6. Bornmann, L., Daniel, H.D.: What do citation counts measure? a review of studies on citing behavior. *Journal of documentation* **64**(1), 45–80 (2008). <https://doi.org/10.1108/00220410810844150>
7. Cavalcanti, D., Prudêncio, R.: Aspect-based opinion mining in drug reviews. In: Oliveira, E., Gama, J., Vale, Z., Lopes Cardoso, H. (eds.) *Progress in Artificial Intelligence*. pp. 815–827 (2017). [https://doi.org/10.1007/978-3-319-65340-2\\_66](https://doi.org/10.1007/978-3-319-65340-2_66)
8. Goodwin, T.R., Skinner, M.A., Harabagiu, S.M.: Utd hltri at trec 2017: Precision medicine track. In: *National Institute of Standards and Technology (NIST)* (2017), <https://trec.nist.gov/pubs/trec26/papers/UTDHLTRI-PM.pdf>
9. Gräßer, F., Kallumadi, S., Malberg, H., Zaunseder, S.: Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. In: *Proceedings of the 2018 International Conference on Digital Health*. pp. 121–125. *ACM* (2018). <https://doi.org/10.1145/3194658.3194677>
10. Jiang, S.Y., Weng, C.: Cross-system evaluation of clinical trial search engines. *AMIA Jt Summits Transl Sci Proc* **2014**, 223–229 (2014), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419768/>
11. Kang, T., Zhang, S., et al.: Eliie: An open-source information extraction system for clinical trial eligibility criteria. *Journal of the American Medical Informatics Association* **24**(6), 1062–1071 (2017). <https://doi.org/10.1093/jamia/ocx019>
12. Kilicoglu, H., Abacha, A.B., et al.: Semantic annotation of consumer health questions. *BMC bioinformatics* **19**(1), 34 (2018). <https://doi.org/10.1186/s12859-018-2045-1>
13. Kim, W., Yeganova, L., et al.: Mesh-based dataset for measuring the relevance of text retrieval. In: *Proceedings of the BioNLP 2018 workshop*. pp. 161–165 (2018). <https://doi.org/10.18653/v1/W18-2320>

14. Lipscomb, C.E.: Medical subject headings (mesh). *Bulletin of the Medical Library Association* **88**(3), 265 (2000), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC35238>
15. MacKellar, B., Schweikert, C.: Patterns for conflict identification in clinical trial eligibility criteria. In: 2016 IEEE Healthcom. pp. 1–6 (Sept 2016). <https://doi.org/10.1109/HealthCom.2016.7749519>
16. Miller, G.A.: Wordnet: a lexical database for english. *Communications of the ACM* **38**(11), 39–41 (1995). <https://doi.org/10.1145/219717.219748>
17. Miotto, R., Jiang, S., Weng, C.: etacts: A method for dynamically filtering clinical trial search results. *Journal of Biomedical Informatics* **46**(6), 1060 – 1067 (2013). <https://doi.org/10.1016/j.jbi.2013.07.014>
18. Page, L., Brin, S., et al.: The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (November 1999), <http://ilpubs.stanford.edu:8090/422/>
19. Patel, C.O., Garg, V., Khan, S.A.: What do patients search for when seeking clinical trial information online? In: AMIA Annual Symposium Proceedings. vol. 2010, p. 597. American Medical Informatics Association (2010), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041375>
20. Reader, T.W., Gillespie, A., Roberts, J.: Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ Qual Saf* **23**(8), 678–689 (2014). <https://doi.org/10.1136/bmjqs-2013-002437>
21. Roberts, K., Demner-Fushman, D., et al.: Overview of the trec 2017 precision medicine track. NIST Special Publication pp. 500–324 (2017), <https://trec.nist.gov/pubs/trec26/papers/Overview-PM.pdf>
22. Roberts, K., Simpson, M., et al.: State-of-the-art in biomedical literature retrieval for clinical cases: a survey of the trec 2014 cds track. *Information Retrieval Journal* **19**(1), 113–148 (Apr 2016). <https://doi.org/10.1007/s10791-015-9259-x>
23. Roberts, K., et al.: Trec precision medicine 2018 track (2018), <http://www.trec-cds.org/2018.html>
24. Sayers, E.: A general introduction to the e-utilities. in: Entrez programming utilities help [internet]. betesda (md): National center for biotechnology information (us) (2010-), <https://www.ncbi.nlm.nih.gov/books/NBK25497/>
25. Sen, A., Ryan, P.B., et al.: Correlating eligibility criteria generalizability and adverse events using big data for patients and clinical trials. *Annals of the New York Academy of Sciences* **1387**(1), 34–43 (2017). <https://doi.org/10.1111/nyas.13195>
26. Soldaini, L., Goharian, N.: Quickumls: a fast, unsupervised approach for medical concept extraction. In: MedIR workshop, SIGIR (2016)
27. Surian, D., Dunn, A.G., et al.: A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates. *Journal of Biomedical Informatics* **79**, 32–40 (2018). <https://doi.org/10.1016/j.jbi.2018.01.008>
28. Thorve, A.: Team ajinkyathorve at trec 2017 precision medicine track (2017), <https://github.com/ajinkyathorve/TREC-2017-PM-CDS-Track>
29. Tong Yin, Danny Wu, V.V.: Retrieving documents based on gene name variations: Medier at trec 2017 precision medicine track (2017), <https://trec.nist.gov/pubs/trec26/papers/MedIER-PM.pdf>
30. Tsafnat, G., Glasziou, P., et al.: Systematic review automation technologies. *Systematic Reviews* **3**(1), 74 (Jul 2014). <https://doi.org/10.1186/2046-4053-3-74>
31. Zuccon, G., Koopman, B., et al.: Choices in knowledge-base retrieval for consumer health search. In: European Conference on Information Retrieval. pp. 72–85 (2018). [https://doi.org/10.1007/978-3-319-76941-7\\_6](https://doi.org/10.1007/978-3-319-76941-7_6)