

Patient selection for clinical trials based on concept-based retrieval and result filtering and ranking

Johannes Leveling

teckro

The Bank Building
63 O'Connell Street
Limerick, Ireland
joleveling@gmail.com

Abstract. For our participation at the clinical trials task in the TREC 2017 Precision Medicine track 2017, we investigated retrieving and matching clinical trial documents with patient information based on text and concept annotations of the text, filtering results for demographic information such as gender and age, and re-ranking results based on patient information. Experimental results show a competitive precision at high ranks for our least complex approach.

1 Introduction

The TREC evaluation campaign has run a number of tracks in the medical domain in previous years. Teckro¹ participated in task 2 of the Precision Medicine track at TREC 2017², where the participants were challenged with finding relevant clinical trials from ClinicalTrials.gov for which a patient is eligible. The task represents the process of selecting candidate patients for a study on experimental treatments, e.g. if prior treatments have been ineffective for a patient. A more detailed overview of the track and its tasks can be found in the track overview paper [7].

The rest of this paper is organized as follows: Section 2 describes related work. Section 3 introduces the motivation and questions around the research described in this paper. Section 4 gives a brief overview of documents, topics, and relevance assessments for the task. Document and query processing for our experimental system are outlined in Sections 5 and 6, respectively, before the retrieval approach is described in Section 7. Section 8 details the official submissions and additional experiments and their results (see Table 2).

¹ The research described in this paper was conducted while Dr. Leveling was employed at teckro.

² <http://trec-cds.appspot.com/2017.html>

2 Related Work

The clinical trials task at TREC 2017 resembles the problem of identifying clinical studies for a patient. A patient can be characterized by his/her demographic group (e.g. gender, age, ethnicity), medical history and conditions, and other factors. To support the patient selection process, clinical trials usually define from several up to several hundred criteria for the inclusion and exclusion of patients. For example, a trial may be restricted by age, gender, or previous conditions (e.g. “pregnant women”).

Previous tracks at TREC have focused on several different use cases from the medical domain. A similar track was concerned with searching for potential participants in trials based on electronic health records [3].

Researchers have also previously tried to parse and analyze eligibility criteria for clinical trials to facilitate finding suitable clinical studies for patients. Much of the work has focused on documents from the ClinicalTrials.gov web site [11], the same web site that provided documents for the TREC 2017 task.

Ash et al. [2] describe a system to automatically encode and evaluate eligibility criteria to find appropriate clinical trials for patients. The system uses standard vocabularies to represent concepts and employs Bayesian networks to infer missing information. An evaluation using clinical trial protocols from NCI’s Physician Data Query shows high agreement (0.84 kappa) between the proposed system and an independent physician in protocol selection.

Bhattacharya and Cantor [4] aim for a more portable and standardized representation of eligibility criteria. They analyze a number of full-text protocols from Pfizer and documents from ClinicalTrials.gov to cluster criteria and find templates in language use.

Similarly, Ross et al. [9] try to analyze the complexity of eligibility criteria in clinical trials. They classify criteria in trials from ClinicalTrials.gov with respect to complexity, semantic patterns, clinical content, and data sources. Their findings include that 85% criteria have significant complexity and 40% have temporal references.

Samson et al. [10] propose the ERGO annotation as a formal representation for eligibility criteria to capture the semantics of eligibility criteria in documents from ClinicalTrials.gov. They test this approach to search for trials and screen patients for eligibility.

For the TREC 2017 clinical trials task, generating a standardized representation of eligibility criteria might have proved helpful as well. However, much of the standardization efforts aim at easier matching of numeric values or values with a fixed domain (e.g. “intraocular pressure < 30 mmHg” or “cholesterol level = high (> 350mg/dl)”). In contrast, the TREC 2017 topics do not seem to include lab tests or references to numeric relationships as constraints.

3 Research questions

We had initially planned to investigate the use of word embeddings and deep learning for this task. Due to the lack of adequate training data and due to time constraints, we abandoned this approach.

Our system for the participation at TREC 2017 was developed with the following considerations in mind:

- This track offers the traditional challenges related to natural language processing and information retrieval, i.e. how to deal with synonyms or related terms. A typical solution would be to annotate text with concepts from a standard vocabulary and exploit relationships between concepts from this vocabulary in an ontology-guided search. For example, including child concepts in a query can help improve recall as synonyms and subordinated concepts in documents would match and increase the result set size.
- In the ClinicalTrials.gov documents, inclusion and exclusion criteria are merged in the *criteria* field in the XML documents. Parsing the criteria field and differentiating between inclusion and exclusion criteria would help improve precision by not matching concepts related to exclusion criteria.
- Individual eligibility criteria can be seen as logical constraints, where exclusion criteria can be transformed into inclusion criteria by negation. For example, the exclusion criterion “age > 60” can be transformed into an inclusion criterion “age ≤ 60” by negating the relation.
- Medical texts often contain negated concepts or negative findings, e.g. “no findings for tumor”. For retrieval, negating concepts (i.e. concepts in a negative scope) might prevent mismatches and improve precision. For example, the concept for “tumor” should be represented as a negated concept, e.g. “not-tumor” if found within a negation scope.

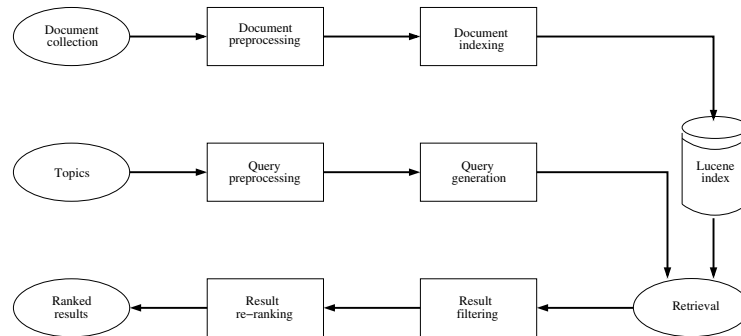


Fig. 1. Data flow and processing steps in the clinical trials retrieval system.

Our system for this task comprises a combination of approaches aiming to address the challenges mentioned above. The data flow and processing steps in

the system are illustrated in Figure 1. We will briefly describe the individual approaches in the following sections.

4 Documents, Topics, Relevance Assessments

The document collection for the TREC 2017 task consists of data from an April 2017 snapshot of ClinicalTrials.gov.³ The collection contains 241,007 trial documents. Sixteen training topics were provided, of which seven had one or two documents assessed as relevant. These topics were used to test our system’s functionality and perform some plausibility checks. Thirty official topics were provided for the task. For each topic, a set of up to 1000 ranked documents was retrieved by the participants. Relevance of documents was assessed by experts at the Department of Medical Informatics of the Oregon Health and Science University (OHSU), judging relevance for a set of pooled documents for each topic, using simple depth-15 pools. More details can be found in the track overview paper [7]. We submitted the allowed maximum of five runs. Before describing the experiments, we provide an overview over the experimental setup.

5 Document Processing

In our system, document processing comprises the following steps:

1. “parsing” document content into separate fields (e.g. inclusion and exclusion criteria from *criteria*),
2. annotating text with concepts from a medical ontology, and
3. detecting negation scope.

We briefly outline the document processing steps in the following subsections.

5.1 Parsing eligibility criteria

The clinical trial documents contain textual information in different fields, of which the *criteria* field lists both inclusion and exclusion criteria. In order to distinguish between positive and negative criteria for inclusion, we analyzed the *criteria* field in the documents. The main idea for the analysis is to support differentiating factors that make a patient eligible for a trial and ones that make him/her ineligible. For example, a patient with “breast cancer” would be eligible for cancer treatment trials, which will list “cancer” as an inclusion criterion, whereas other studies might explicitly exclude patients with cancer. Both criteria would be mentioned in the *criteria* field, and parsing the content of this field aims at distinguishing inclusion and exclusion criteria, so as not to mix matches and mismatches in the result set.

We found that while the original document XML does not differentiate between inclusion and exclusion criteria with XML markup, some text markers

³ <https://clinicaltrials.gov/>

```
<criteria>
  <textblock>
    Inclusion Criteria:

    1. Subjects must be healthy male nondependent recreational drug users

    2. Subjects must be 18 to 55 years old, inclusive.

    3. Subjects must have greater than or equal to 10 lifetime nontherapeutic experiences
       with central nervous system (CNS) stimulants (e.g., amphetamines, cocaine,
       methylphenidate), greater than or equal to 1 nontherapeutic use of prescription
       stimulants within the 12 months prior to Screening, and greater than or equal to 1
       nontherapeutic use of a CNS stimulant within the 12 weeks prior to Screening.

    Exclusion Criteria:

    1. Subjects that are deemed medically unsuitable or unlikely to comply with the study
       protocol for any reason.

    2. Subjects who do not pass Qualification Phase criteria to be eligible for the
       Treatment Phase.
  </textblock>
</criteria>
```

Fig. 2. Sample *criteria* field for document NCT02144415.

```

<inclusion_criteria>
  Inclusion Criteria:

  1. Subjects must be healthy male nondependent recreational drug users

  2. Subjects must be 18 to 55 years old, inclusive.

  3. Subjects must have greater than or equal to 10 lifetime nontherapeutic experiences
     with central nervous system (CNS) stimulants (e.g., amphetamines, cocaine,
     methylphenidate), greater than or equal to 1 nontherapeutic use of prescription
     stimulants within the 12 months prior to Screening, and greater than or equal to 1
     nontherapeutic use of a CNS stimulant within the 12 weeks prior to Screening.
</inclusion_criteria>
<concept_inclusion_criteria>
  C28822 C54627 C20197 C47795 C28981 C646 C12438 C48261 C62006 C84368 C80153
  C15419 C84366 C62045 C120945
</concept_inclusion_criteria>
<concept_child_inclusion_criteria>
  C79869 C77844 C77843 C75029 C77846 C77845 C74454
  ...
</concept_child_inclusion_criteria>
<exclusion_criteria>
  Exclusion Criteria:

  1. Subjects that are deemed medically unsuitable or unlikely to comply with the study
     protocol for any reason.

  2. Subjects who do not pass Qualification Phase criteria to be eligible for the
     Treatment Phase.
</exclusion_criteria>

```

Fig. 3. Sample *inclusion_criteria*, *exclusion_criteria* and concept annotations for document NCT02144415.

(usually lines) and formatting (line breaks) often indicates where a list of inclusion or exclusion criteria starts. Figure 2 shows the content of the *criteria* field for a sample document. For example, most of the original formatting (i.e. line breaks) seems to have been retained in the XML, and a line such as “Eligibility criteria:” would indicate that a list of criteria follows in the text. We computed the frequency for all lines in the criteria field for all documents and manually assigned a category to lines occurring with a frequency of 10 or higher to indicate whether they start a list of inclusion (I) or exclusion criteria (E). Based on these categories, we built a simple parser, which iterates over the content of the *criteria* field line by line and assigns a category to each line based on the manually annotated data.

For example, lines marking the start of exclusion criteria in the *criteria* field include “Exclusion criteria:”, “A person will be excluded from the study if he/she:”, “Patients with any of the following conditions are excluded:”, and “Subjects to whom any of the following applies will be excluded from the study:”. Markers indicating the start of inclusion criteria include “Inclusion Criteria:”, “A person is eligible for inclusion in the study if he/she:”, “ELIGIBILITY CRITERIA:”, and “Participants will be able to enroll if they:”.

After assigning a category to each line, we concatenated the content for each category to form a field for *inclusion_criteria* and a field for *exclusion_criteria*. Figure 3 shows the result of parsing content of the *criteria* field for the sample document from 2. There are, of course, documents, where the distinction between inclusion and exclusion criteria is less obvious or more complex. For example, the content might contain the sentence “Pregnant women are excluded from this study”, which defines an exclusion criterion, but not as an item in a list of criteria. To deal with these cases, a more sophisticated parsing approach would be required.

Text in the following XML fields was concatenated into a single *text* field: *brief_title*, *brief_summary*, *condition*, *condition_browse*, *criteria*, *description*, *detailed_description*, *intervention_browse*, *intervention_name*, *keyword*, *mesh_term*, *official_title*, and *other_name*.

5.2 Concept annotation

We selected the NCI thesaurus⁴ (version Thesaurus_17.04d) as the basis for annotating concepts in text from clinical trial documents. The NCI thesaurus comprises more than 120k concepts with information including preferred terms, synonyms, semantic types, child and parent concepts, and definitions. Table 1 shows a sample concept from the NCI thesaurus in a readable format.

To alleviate the potential problem of ambiguity and to remove annotations that were deemed not helpful for the task at hand, we filtered out some concepts from the ontology. We discarded concepts marked as “retired” concepts in the NCI thesaurus. Concepts were also removed based on their semantic type,

⁴ <https://ncit.nci.nih.gov/>

Table 1. Sample concept from the NCI thesaurus showing concept ID, parent IDs, synonyms, and semantic type for “Myocardial_Infarction”. The parent concept is “Myocardial_Disorder”.

Concept ID:	C27996
Concept name:	Myocardial_Infarction
Parent IDs:	C35544
Synonyms:	Myocardial Infarction; Heart Attack; MI; Myocardial Infarct; Myocardial Infarction, (MI)
Definition:	Gross necrosis of the myocardium, as a result of interruption of the blood supply to the area, as in coronary thrombosis.
Concept status:	–
Semantic types:	Disease or Syndrome

i.e. excluding concepts with the semantic type “Quantitative Concept”, “Qualitative Concept”, “Idea or Concept”, “Temporal Concept”, “Research Activity”, “Geographic Area”, “Functional Concept”, “Spatial Concept”, “Molecular Biology Research Technique”, “Professional Society”, “Animal”, “Mammal”, “Fish”, “Vertebrate”, “Bird”, “Amphibian”, “Fungus”, “Plant”, “Language”, etc.

In addition, we filtered out concepts referring to adverse events (based on the concept label), as these concepts would be more relevant in later stages of a clinical trial. For example, the NCI thesaurus contains concepts for both “Tremor” and “Tremor_Adverse_Event”, which both have “Tremor” as a synonym and we keep only the first concept.

Concept annotation for a text is based on comparing all concept labels (i.e. for preferred and alternative concept labels) in their tokenized and stemmed form with the stemmed token sequence for the text. To maintain high precision for the concept annotation, we used the light “s”-stemmer [6]. Concept labels are matched from left to right, keeping the longest match only, without allowing overlapping matches.

Annotation information includes IDs for the matching concepts, and start and end offsets for the matching concept label in the text. The *criteria*, *inclusion_criteria*, *exclusion_criteria*, and *text* field are annotated with concepts, storing the corresponding concept IDs in a separate field (i.e. with a *concept_* prefix for the field name). The main idea behind concept annotation is to allow easier matching with morphological variants of the concept labels and to alleviate the problem of synonyms or related concepts. For example, “heart attack” vs. “myocardial infarct” would be annotated with the same concept ID.

For each annotated concept, all of its child concept IDs in the NCI thesaurus were determined and aggregated - together with the original concept ID - in a separate document field (i.e. with *concept_child_* prefix). The idea behind indexing child concepts is to allow matching between a more generic and a more specific (or subsumed) concept. For example, the topic might contain a more generic term such as “cancer”, while the document mentions the more specific “breast cancer”.

5.3 Negation detection

Negation has long been identified as a potential problem when analyzing text in the medical domain, see e.g. [1]. For example, “no findings for cancer” indicates that the presence of cancer cannot be confirmed, so an annotation of this text fragment with the concept for “cancer” would be incorrect. We employed the NegEx algorithm [5] to identify the scope of negations in the clinical trial content. Concepts found within a negative scope are negated, i.e. their concept ID is modified with a negation prefix so a search for the (un-negated) concept ID will not produce matches. Note that when a concept is negated, all its child concepts should be negated as well in the concept annotation step described in the previous subsection.

6 Query Processing

Figure 4 shows a sample topic with information in different fields: *disease* (disease under investigation), *gene* (genes of interest), *demographic* (patient’s gender and age), and *other* (other criteria such as medical conditions and findings).

```
<topics task="2017 TREC Precision Medicine">
  <topic number="1">
    <disease> Acute lymphoblastic leukemia </disease>
    <gene> ABL1, PTPN11 </gene>
    <demographic> 12-year-old male </demographic>
    <other> No relevant factors </other>
  </topic>
  ...
</topics>
```

Fig. 4. Sample topic.

For the generation of queries from topics, text from the *disease*, *gene* and *other* topic fields is annotated with concepts in the same way as for documents. Phrases such as “No relevant factors” are deleted from the topic text. Age and gender information are extracted from the topic’s *demographic* field for filtering out trials where the patient does not fall in the age or gender group. This query processing step allows for easier transformation of topics into queries for the retrieval engine.

7 Document retrieval and matching

Our system is based on a Lucene core⁵ to facilitate retrieval of indexed clinical trial documents. Retrieving and matching documents for a given topic involves the following steps:

- retrieving documents using the BM25 retrieval model,
- hard filtering of results with constraints such as gender and age, and
- re-ranking documents based on inclusion and exclusion criteria.

We briefly outline these steps in the following subsections.

7.1 Retrieval

The standard retrieval function for Lucene was, until recently, a variant of *tf-idf* which, compared to state-of-the-art retrieval models, achieved lower retrieval effectiveness. The underlying retrieval engine for our TREC 2017 experiments is Lucene (version 6.1.0), which in this version employs the BM25 retrieval model [8] as the default ranking model. BM25 still offers competitive retrieval effectiveness, even without parameter optimization. We used BM25 “out-of-the-box”, i.e. without parameter optimization. Also note that Lucene’s BM25 implementation still suffers a loss in effectiveness compared to a faithful BM25 implementation due to a lossy encoding of document length.⁶ We did not apply the proposed fix to the Lucene BM25 model, which would potentially improve performance as it increases accuracy for document length normalization.

The query generated from a topic contains n-grams for the textual fields (i.e. *disease*, *gene*, *other*), which are searched as phrases. Terms and phrases for each field are syntactically grouped together. Terms for the *disease* field are treated as mandatory search terms, i.e. a result document **must** have at least one match with terms in the *disease* group.

For all topics, a maximum of 1500 documents are initially retrieved, to allow retrieval of 1000 documents after filtering out non-matching documents. Documents in the retrieved set are then filtered by matching age restrictions (i.e. minimum and maximum age in years in *minimum_age* and *maximum_age* document field) and gender (i.e. male, female, or no restriction in *gender* document field) from the *demographic* field in the topic.

For cases where less than 1000 documents are retrieved in the initial retrieval step (i.e. with mandatory matching of *disease* terms), a second query is generated as a fallback, dropping the restriction that *disease* terms are mandatory. Results from the second retrieval are appended to the initial results, discarding documents that have already been retrieved in the initial retrieval step. Scores for the second retrieval step are re-normalized, starting with 99% of the score for the lowest ranked document from the initial retrieval step to merge both result sets.

⁵ <https://lucene.apache.org/core/>

⁶ <http://searchivarius.org/blog/accurate-bm25-similarity-lucene>

Table 2. Parameter settings and results for official runs, re-runs,, and additional experiments.

Run Identifier	Parameters		Results				
	Text Field	Concept Field	MAP	NDCG	P@5	P@10	P@20
teckro1	<i>text</i>	<i>concept_text</i>	0.2430	0.5295	0.4276	0.4000	0.3207
teckro2	<i>criteria</i>	<i>concept_incl_criteria</i>	0.1653	0.4235	0.3172	0.3172	0.2508
teckro3	<i>text</i>	<i>concept_child_text</i>	0.2430	0.5295	0.4276	0.4000	0.3207
teckro3 re-run	<i>text</i>	<i>concept_child_text</i>	0.1884	0.4898	0.3724	0.3103	0.2466
teckro4	<i>criteria</i>	<i>concept_child_incl_criteria</i>	0.1653	0.4235	0.3172	0.3172	0.2500
teckro4 re-run	<i>criteria</i>	<i>concept_child_incl_criteria</i>	0.1054	0.3634	0.2000	0.1828	0.1397
teckro5	<i>text</i>	<i>concept_child_incl_criteria</i>	0.2435	0.5296	0.4276	0.4000	0.3224
teckro5 re-run	<i>text</i>	<i>concept_child_incl_criteria</i>	0.1895	0.4899	0.3724	0.3103	0.2466
teckro6	<i>text</i>	<i>concept_incl_criteria</i>	0.2436	0.5301	0.4276	0.4138	0.3234
teckro7	<i>text</i>	<i>concept_criteria</i>	0.1905	0.4910	0.3862	0.3172	0.2603
teckro8	<i>criteria</i>	<i>concept_criteria</i>	0.1065	0.3638	0.2069	0.1793	0.1414

7.2 Result filtering

Some demographic eligibility criteria for the trials are encoded in different document fields, e.g. *gender*, *minimum_age*, *maximum_age*. We parsed the content of these fields to facilitate matching with the same information from the *demographic* field in the topics. For example, the age fields contain the numeric value as well as the unit (e.g. year(s), month(s)). We separated out the unit into a separate field for the index (e.g. *minimum_age_unit*) and mapped the “All” value for gender to match both “male” and “female”. These restrictions were used to filter out any documents that did not match the gender or age criteria provided in the topics.

7.3 Reranking results

In a final step, the retrieved set of documents was reranked as follows:

- a document’s score is increased by 1% if it contains
 - i) a phrase in the inclusion text field,
 - ii) a concept in the inclusion concept field, or
 - iii) a negated concept in the exclusion concept field.
- a document’s score is reduced by half if it contains
 - i) a phrase in the exclusion text field,
 - ii) a concept in the exclusion concept field, or
 - iii) a negated concept in the inclusion concept field.

The motivation behind this step is to reduce the impact of terms from these fields – including the *other* field – from the original query, where the presence or absence of terms (or concepts) is not a meaningful indicator of relevance (or in the typical cases, of eligibility) for a clinical trial.

8 Results and Analysis

We submitted five runs, varying the parameter settings corresponding to which fields in the document index are searched. Parameters and results for the five official runs, re-runs, and additional experiments are shown in Table 2.

The three runs teckro3-5 were configured incorrectly (i.e. teckro1 is the same as teckro3) and we had to re-run these experiments with the corrected configuration. Results for three additional runs are also included (teckro6, teckro7, and teckro8). Evaluation results include mean average precision (MAP), NDCG, and precision at 5, 10, and 20 as reported by `trec_eval`.

We achieved the best results with our least complex approach (teckro1). Restricting concept matching to the inclusion criteria yields slightly higher results for some metrics such as P@10 (teckro7). The use of child concepts did not help obtain higher performance for any experiment. Precision at 5 documents for all runs likely suffers due to suboptimal settings for the re-ranking of documents.

The relatively high performance can be attributed mostly to document pre-processing (i.e. parsing criteria) and overcoming some problems of the vocabulary mismatch with the NCI thesaurus. Only one gene name in the topics was not covered by the thesaurus. Overall, precision values are competitive with results from other top teams.

In conclusion, the clinical trials task is not a simple information retrieval task, where treating the query and documents represented as a bag of words would result in high precision. Our least complex approach achieved the best performance compared to approaches including structural information for concepts.

In the future, our current system might form a baseline system to support comparison with and optimization of more complex approaches, including ones based on current research topics such as deep learning.

References

1. Shashank Agarwal and Hong Yu. Biomedical negation scope detection with conditional random fields. *Journal of the American Medical Informatics Association*, 17(6):696–701, 2010.
2. Nachman Ash, Omolola Ogunyemi, Qing T. Zeng, and Lucila Ohno-Machado. Finding appropriate clinical trials: evaluating encoded eligibility criteria with incomplete data. In *AMIA 2001, American Medical Informatics Association Annual Symposium*, pages 27–31. AMIA, 2001.
3. Steven Bedrick, Kyle H. Ambert, Aaron M. Cohen, and William R. Hersh. Identifying patients for clinical studies from electronic health records: TREC medical records track at OHSU. In Ellen M. Voorhees and Lori P. Buckland, editors, *The Twentieth Text REtrieval Conference Proceedings (TREC 2011)*, volume Special Publication 500-296, Gaithersburg, MD, 2011. National Institute for Standards and Technology.
4. Sanmitra Bhattacharya and Michael N. Cantor. Analysis of eligibility criteria representation in industry-standard clinical trial protocols. *Journal of Biomedical Informatics*, 46:805–813, 2013.

5. Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301 – 310, 2001.
6. Donna Harman. How effective is suffixing? *Journal of the American Society for Information Science*, 42:7–15, 1991.
7. Kirk Roberts, Diina Demner-Fushman, Ellen M. Vorhees, William R. Hersh, Steven Bedrick, Alexander J. Lazar, and Shubham Pant. Overview of the TREC 2017 Precision Medicine Track. In *TREC 2017 Notebook papers*. NIST, 2017.
8. Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Found. Trends Inf. Retr.*, 3(4):333–389, 2009.
9. Jessica Ross, Samson Tu, Simona Carini, and Ida Sim. Analysis of eligibility criteria complexity in clinical trials. *Summit on Translational Bioinformatics*, 2010:46–50, 2010.
10. Samson W. Tu, Mor Peleg, Simona Carini, Michael Bobak, Jessica Ross, Daniel Rubin, and Ida Sim. A practical method for transforming free-text eligibility criteria into computable criteria. *Journal of Biomedical Informatics*, 44(2):239–250, 2011.
11. Deborah A. Zarin, Tony Tse, Rebecca J. Williams, Robert M. Califf, and Nicholas C. Ide. The ClinicalTrials.gov results database – Update and key issues. *The New England journal of medicine*, 364:852–860, 2011.