

Exploiting Structural Patterns for Health Search

Research Track - Project Proposal

Alberto Ueda
CS Dept, UFMG

ABSTRACT

Health related topics have become a common theme within Information Retrieval (IR). A number of premiere IR publication venues, including SIGIR, have dedicated workshops, tutorials or tracks dedicated to health search. For instance, the vast literature available for precision medicine can make it difficult to find the most appropriate treatment for the clinician's current patient. The ability to quickly locate relevant information for a current patient using IR can be an important tool for helping clinicians find the most up-to-date evidence-based treatment for their patients. In this project, we aim to propose novel strategies for clinical decision support by applying IR techniques such as similarity-based rankings, machine learning and feature engineering.

ACM Reference format:

Alberto Ueda. 2017. Exploiting Structural Patterns for Health Search. In *Proceedings of ACM Conference, Washington, DC, USA, August 2017 (Conference '17)*, 2 pages.

DOI: 10.1145/nnnnnnnn.nnnnnnnn

1 THE PROBLEM

The problem we tackle in this project is defined by the TREC 2015 Clinical Decision Support Track. As in previous editions, the main goal of this track is the retrieval of biomedical articles relevant for answering generic clinical questions about medical records [5]. More specifically, the participants of the track were challenged with retrieving for a given case report full-text biomedical articles that answer questions related to several types of clinical information needs. Retrieved articles are judged relevant if they provide information of the specified type that is pertinent to the given case.

A case report typically describes a challenging medical case, and it is often organized as a well-formed narrative summarizing the portions of a patient's medical record that are pertinent to the case. It provides information such as a patient's medical history, the patient's current symptoms, tests performed by a physician to diagnose the patient's condition, the patient's eventual diagnosis, and finally, the steps taken by a physician to treat the patient. The table 1 shows an example of the kind of case-based topic used in the track.

In the previous editions of TREC, several ranking models for medicine document retrieval were proposed [4]. Techniques from IR and NLP were employed by the most successful participant teams, including disease-centered document clustering and semantic word vectors using word embeddings [3], query expansion by identification of clinical intent types, negation-aware ranking models, and synonyms [1], and supervised learning-to-rank approaches based on document similarity [2].

Conference '17, Washington, DC, USA
2017. 978-x-xxxx-xxxx-x/YY/MM...\$15.00
DOI: 10.1145/nnnnnnnn.nnnnnnnn

Table 1: Example of input and expected output for the task.

Case Report Example	Relevant Articles
Type: Treatment.	1065341
Description: A 21-year-old female is evaluated for progressive arthralgias and malaise. On examination she is found to have alopecia, a rash mainly distributed on the bridge of her nose and her cheeks, a delicate non-palpable purpura on her calves, and swelling and tenderness of her wrists and ankles. Her lab shows normocytic anemia, thrombocytopenia, a 4/4 positive ANA and anti-dsDNA. Her urine is positive for protein and RBC casts.	1459118 1526641

2 OUR APPROACH

In some contexts, finding highly related documents is a desirable feature for IR systems. For instance, given a relevant medicine article with a treatment for a patient's disease, one can be interested in finding as many documents similar to this article as possible.

For such tasks, we can apply document clustering techniques. In particular, using disease-centered document clusters improve clinical document retrieval effectiveness [3]. Our proposal is to still consider the disease but on a treatment-centered network. Given a patient with a specific disease and a reference treatment, we aim to provide articles with similar treatments or clinical trials. To determine the reference treatment, we aim to use learning to rank methods based on the clinical ground-truths.

3 EVALUATION PROCEDURE

We aim to use both the data and the evaluation procedures provided by previous tracks of the TREC.¹ The highest ranked articles for each topic were pooled and judged by medical librarians and physicians from the Department of Medical Informatics of the Oregon Health and Science University. The articles were judged as either *definitely relevant*, *definitely not relevant*, or *potentially relevant*. To obtain a graded relevance scale, the performance of the retrieval submissions were measured using NDCG. As the previous participants' results are also available, we will be able to directly compare the effectiveness of our approaches against the top tier participants.

REFERENCES

- [1] Simon Greuter, Philip Junker, Lorenz Kuhn, Felix Mance, Virgile Mermet, Angela Rellstab, and Carsten Eickhoff. 2016. ETH Zurich at TREC 2016 Clinical Decision Support. In *TREC 2016*, Ellen M. Voorhees and Angela Ellis (Eds.).
- [2] Harsha Gurulingappa, Luca Toldo, Claudia Schepers, Alexander Bauer, and Gerard Megaro. 2016. Semi-Supervised Information Retrieval System for Clinical Decision Support. In *TREC*.
- [3] Seung-Hyeon Jo and Kyung-Soon Lee. 2016. CBNU at TREC 2016 Clinical Decision Support Track. In *TREC*.
- [4] Kirk Roberts, Matthew S. Simpson, Ellen M. Voorhees, and William R. Hersh. 2015. Overview of the TREC Clinical Decision Support Track. In *TREC 2015*.
- [5] Guido Zuccon and Bevan Koopman. 2017. SIGIR 2017 Tutorial on Health Search (HS2017): A Full-day from Consumers to Clinicians. In *SIGIR 2017 (SIGIR '17)*. ACM, New York, NY, USA, 1415–1418. DOI: <https://doi.org/10.1145/3077136.3082061>

¹<http://trec-cds.org/2015.html>