Addressing the Search Challenges of Precision Medicine with Information Retrieval Systems and Physician Readers

Kate Fultz HOLLIS[a,1], Kirk ROBERTS[b], Steven BEDRICK[a], William R. HERSH[a]

[a] *Oregon Health & Science University Department of Medical Informatics and Clinical Epidemiology, Portland, OR, USA*
[b] *The University of Texas Health Science Center, School of Biomedical Informatics, Houston, TX, USA*

Abstract. The Text REtrieval Conference (TREC), co-sponsored by the National Institute of Standards and Technology (NIST) in the US and US Department of Defense, was started in 1992. TREC's purpose is to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. In 2017, the TREC Precision Medicine (Roberts et al., 2017) track grew from the Clinical Decision Support track and focused on a narrower problem domain of precision oncology. After three years of computer runs being evaluated for relevance by physician readers, we provide a unique perspective of how to evaluate computer-generated articles and clinical trials pulled from PubMed and Clinicaltrials.gov to find relevant information on medical cases.

Keywords. Precision medicine, information retrieval

1. Introduction

Starting in 2017, the Text REtrieval Conference (TREC), co-sponsored by the US National Institute of Standards and Technology (NIST) and the US Department of Defense, produces as part of its large information retrieval research program a project devoted to information retrieval focused on precision medicine. The 2017 "Overview of the TREC 2017 Precision Medicine Track" reported on search challenges with "precision medicine." Many more treatment options become available with information tailored to the individual patient. The abundance of clinical information can easily overwhelm clinicians attempting to stay up to date with the latest findings and can easily inhibit a clinician's attempts to determine the best possible treatment for a particular patient. As a forerunner to the Precision Medicine Track, three consecutive years of the TREC Clinical Decision Support (CDS) track sought to evaluate information retrieval (IR) systems that provide medical evidence to the point-of-care. The CDS track then focused on the needs of precision medicine so IR systems can focus on this important issue [1].

The years 2017, 2018, and 2019 Precision Medicine Tracks focused on a single field, oncology, for a specific use case: genetic mutations of cancer. The main idea behind

---

[1] Corresponding Author, Kate Fultz Hollis; E-mail: fultzhol@ohsu.edu.

precision medicine is to use detailed patient information (largely genetic information in most current research) to identify the most effective treatments. Improving patient care in precision oncology then requires both (a) a mechanism to locate the latest research relevant to a patient, and (b) a fallback mechanism to locate the most relevant clinical trials when the latest techniques prove ineffective for a patient. In the first part, the track continues the previous Clinical Decision Support track [2], while in the second part the study expands the task to cover a new type of data (clinical trial descriptions). All of the TREC Precision Medicine tracks have used synthetic data to create topics of clinical cases for the computer search. All topics were created synthetically because it was difficult to obtain actual patient data for the project [3].

Detailed results from the TREC Precision Medicine Tracks are available in the overviews published on the TREC website. This paper focuses on the process of recruiting physicians to read the output from the computer runs, analyze the relevancy of the article or clinical trial for information about the topic, and help researchers provide information about the accuracy of the computer to find useful information for the physician.

2. Methods

A pooling strategy was used to determine the set of articles and trials to be judged. Slightly different pooling strategies were used for each year for each type of result, but the most common method is as follows. All articles and trials that were within the top 10 results for any research participant (i.e., the research groups searching for the article and clinical trial documents) were included in the pool. Additionally, a 10% sample of the results in position 11 to 100 were added to the pool. The pooled articles and clinical trials for each topic were judged by physicians who, in most cases, were graduates of the Oregon Health & Science University (OHSU) biomedical informatics graduate program. In the case of the clinical support track where physicians evaluated whether a clinical visit was relevant to the case, the challenges included terminology differences, negation, and time aspects having the most impact on retrieval inaccuracy. Edinger et al. describe in detail that these challenges also arise during cohort identification in medical text [4].

The cohort of graduates from OHSU ideally needed to have the following characteristics to become readers for the project: a certificate or master's degree in biomedical informatics, a medical or doctor of osteopathic medicine degree to evaluate relevancy appropriate in a clinical situation, and an ability to quickly review many documents from a website created for the project. The number of topics (topic examples are in Table 1), documents found (total number of abstracts from PubMed and clinical trials from ClinicalTrials.gov), and readers for each year in the precision medicine track are outlined in Table 2.

Table 1. Example topics (3)

Disease	Variant	Demographic
melanoma	BRAF (V600E)	64-year-old male
melanoma	high serum LDH levels	69-year-old female
medullary thyroid cancer	RET	45-year-old female

Table 2. Number of topics, documents and readers

|	2017	2018	2019
Number of Topics	30	50	40
Total Number of Documents	36081	36617	32162
Number of Physician Readers	18	26	19

2.1. Recruitment

Over three years of reader recruitment for the Precision Medicine track, we had to slightly change where we got readers for the project. In all three years, the chair of the OHSU Department of Medical Informatics and Clinical Epidemiology (DMICE) would send a letter to DMICE alumni roughly three months before the October 1 deadline. As the numbers in Table 2 indicate, we had many more topics in 2018 and so also sent the letter to chair people in other departments of biomedical informatics for their suggestions of clinical informatics fellows and doctorate students who might be interested in being readers for the project.

The incentives for readers to work on the project were to learn more about information retrieval and receive payment for reading. Noted in our recruitment letter for the work, we paid a set amount per hour. We paid for quantity of articles judged, not by time, but experience has shown that almost everyone can average one minute or less per article.

3. Results and discussion

The primary purpose for the physician readers in this project was to determine if each article and clinical trial found for a particular topic was 1) human medicine, 2) disease is relevant, partially relevant, or not relevant, 3) whether gene variants were discussed, and 4) whether the demographics (age and sex) matched (see figure 1 for screenshot of OHSU website for viewing articles and trials the computer selected). Details on the relevancy statistics for TREC Precision Medicine Tracks are in (1) and (2) and are forthcoming in 2019. The input of the physician readers has helped the project in many ways outlined in the following three years.
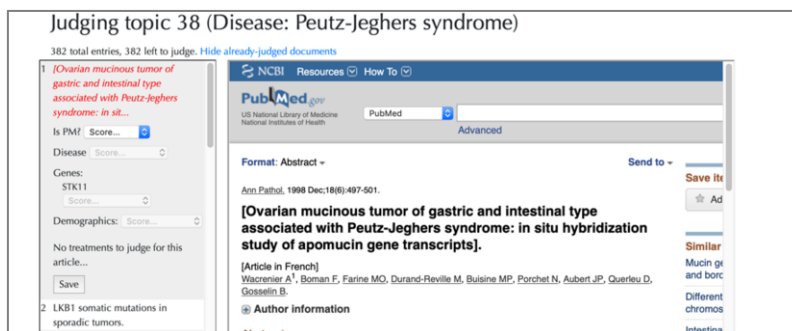
Figure 1. Example of OHSU website for reading articles captured in the computer runs.

3.1. 2017

The recruitment for readers began on August 17, 2017, and 18 readers finished 18 topics in 30 days. The remaining 12 topics were finished by October 8. In 2017, there were roughly 9% judged definitely relevant out of 22,642 literature articles, and 3% judged definitely relevant out of 13,441 clinical trials. We had and continue to have an international contingent of physician readers with many throughout the US and some in South America, Taiwan, and the Middle East. Readers mentioned that it was particularly difficult to find genes mentioned in the articles and usually would need to search in the full text of the article (the web interface we use shows article abstracts only). In terms of the web interface to read the abstracts and clinical trials, readers had, in some cases skipped articles and trials and had difficulty finding the particular missing article or trial in the web interface. Some of the time involved to get topics finished was due to skipped articles. By 2019, the website had been refined so that missing articles while working on the web interface was not a problem.

3.2. 2018

The second year of the Precision Medicine track included almost 50 topics, and it was necessary to recruit outside of the OHSU graduate program and in particular we found doctorate candidates in National Library of Medicine programs across the United States. The recruitment for readers began on August 22, 2018, and on average, 26 readers finished 25 topics in 21 days. The remaining 25 topics were finished by October 1. In 2018, there were roughly 15% judged definitely relevant out of 22,429 literature articles, and 6% judged definitely relevant out of 14,188 clinical trials. Readers from the doctorate programs would, in some cases, complete ranking topics faster than the physician readers and in other cases had the most questions and took more time. While we felt it was important to include informatics doctorate students without medical degrees, the physician readers had more skill in identifying articles and trials that would help for a particular topic.

3.3. 2019

At the time of this writing, the Precision Medicine Track reader analysis is still in operation (results due October 1, 2019). The recruitment for readers began on August 12, 2018, and on average 19 readers finished 19 topics in 16 days. The website took longer this year to produce, and there were more questions about how to match what the computer found as a genetic variant in articles and clinical trials.

   The challenge with working with physician readers is that most have full-time clinical responsibilities (one reader was on call for forest fire duty in California in addition to clinical duties), but all readers were able to finish on time and in the case of two readers, one reading up to 1212 combined abstracts and trials for one topic in 10 days and another did four topics in 2 weeks. 15 readers in 2019 have made all of the previous Precision Medicine Tracks and look forward to working with us again. Many of the readers provide us with questions to help make the search for precision medicine topics better: e.g., the computer runs were sometimes not bringing in any relevant articles and how do the research groups change the algorithm to bring in more relevant articles

and trials. Some readers tried to do the reading and analysis on mobile devices. That is not possible now but could be considered in the future to enable more readers to work with the TREC website from OHSU DMICE.

It is possible that duplicate judging might improve the process as readers thought that in some cases, they were not as familiar with the case and suggested that another expert review their cases. We used to do duplicate judging (in the CDS Track) with kappa analysis, but the results always seemed to be about the same (good but not great consistency). While readers noted that there seemed to them poor relevancy of the articles and clinical trials chosen by computer, there appear to be improvements from 2017 to 2018, and there were fewer comments by readers about how few relevant articles there were in 2019 (only one reader and he was new).

The TREC assessment exercise can improve how we look for cases that closely resemble a clinical problem in precision medicine. The data from the physician readers has allowed more work on computer analysis of medical cases and eventually help with better searching for clinical cases in electronic health records.

4. Acknowledgements

References

[1] K. Roberts, D. Demner-Fushman, E.M. Voorhes, W.R. Hersh, A.J. Lazar, Overview of the TREC 2017 Precision Medicine Track, https://trec.nist.gov/pubs/trec26/papers/Overview-PM.pdf, *Text Retr Conf*. (2017), 1–6.

[2] K. Roberts, D. Demner-fushman, E.M. Voorhees, W.R. Hersh, Overview of the TREC 2016 Clinical Decision Support Track, https://trec.nist.gov/pubs/trec25/papers/Overview-CL.pdf, *Text Retr Conf*. **2** (2016), 1–14.

[3] K. Roberts, D. Demner-Fushman, E.M. Voorhes, W.R. Hersh, A.J. Lazar, Overview of the TREC 2018 Precision Medicine Track, https://trec.nist.gov/pubs/trec27/papers/Overview-PM.pdf, *Text Retr Conf*. (2018), 1–6.

[4] T. Edinger, A.M. Cohen, S. Bedrick, K. Ambert, W. Hersh, Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540501/pdf/amia_2012_symp_0180.pdf, *AMIA Annu Symp Proc.*, 2012.