DEMIR at TREC-Medical 2011: Power of Term Phrases in Medical Text Retrieval

Okan Ozturkmenoglu, Adil Alpkocak

Dokuz Eylul University
Department of Computer Engineering
DEMIR Dokuz Eylul Multimedia Information Retrieval Research Group
Tinaztepe, 35160 Izmir, Turkey
okan.ozturkmenoglu@deu.edu.tr, alpkocak@cs.deu.edu.tr

Abstract. This paper present the details of participation of DEMIR (Dokuz Eylul University Multimedia Information Retrieval) research team to TREC 2011 Medical Records track. In this study, our aim is to index and retrieve medical terms and term phrases in medical text archives. We searched medical terms and term phrases with using UMLS which is a metathesaurus about medical. We evaluated the effects of terms and term phrases on retrieval system in TREC 2011 Medical Records track, considering terms and term phrases as medical entities. We improved results by examination of different weighting schemes for retrieved data.

Keywords: Information Retrieval, Text Retrieval, Medical Text Retrieval.

1   Introduction

This paper present the experiments performed by Dokuz Eylul University Multimedia Information Retrieval (DEMIR) Group[1], Turkey, in the context of our participation to the TREC 2011 Medical Records track [1]. The goal of the Medical Records track is to provide content-based access to the free-text fields of electronic medical records.

The main focus of our experiments is to evaluate the power of medical entities such as term and term phrases in medical text retrieval. The system intends to use medical entities found in Unified Medical Language System (UMLS) [2], that is a set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability between computer systems.

Before indexing the document collection, we performed a preprocessing phase to select the indexing terms. This phase, any term phrase (i.e., practically bi-word) exists in UMLS is accepted as a medical entity and selected as an index terms for our system. In this step, we used Lexical Variant Generation (LVG) tool [3], version 2006p, which is available as UTS Web service framework to generate a set of string normalizations for the input string. Figure 1 shows the major components of our

---

[1] Dokuz Eylul Multimedia Information Retrieval (DEMIR) Group, http://demir.cs.deu.edu.tr

approach and their transitions. In this study, we only include and used "chief complaint" and "report" fields of documents. In order to normalize tokens, we analyzed each token with LVG tool. Then, we selected any token as indexing term if it exist in UMLS. Furthermore, we have also checked if bi-words appear in UMLS. If appears, we signed them as medical entities and we ignored the words if not. We used all these terms and term phrases (i.e., bi-words) as indexing terms and used feed into Terrier for retrieval.
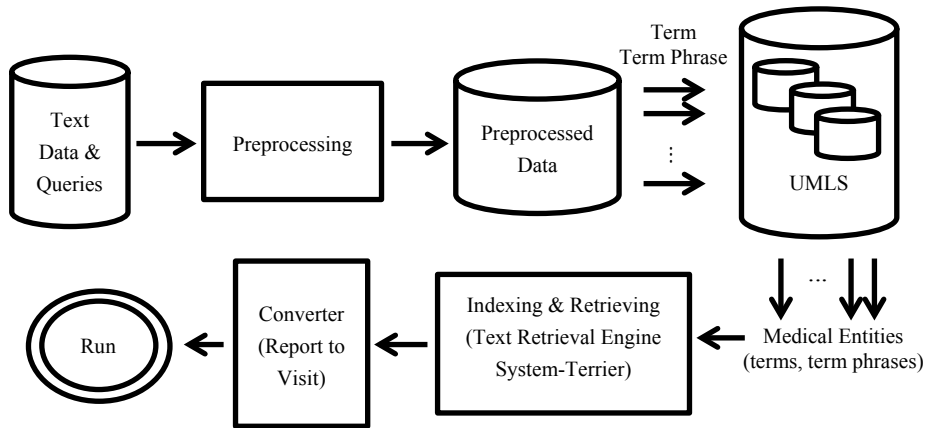

Fig. 1. Basic block diagram of retrieval system.

Rest of this paper is organized as follows: Section 2 presents the textual features we used in this study. Section 3 presents the runs we submitted to TREC 2011 Medical Records track data. Finally, section 4 concludes the paper by pointing out the open issues and possible avenues of further research in the area of make semantic relations between queries and documents and document expansion for medical text retrieval.

2   Textual Features

The data collection of TREC 2011 Medical Records track contains more than a thousand text documents, which are repository of de-identified clinical reports, are generated from multiple hospitals during 2007 by University of Pittsburgh BLULab and include these types: consults, discharge summaries, progress notes etc. Participants were given a set of 35 textual queries where each query may specify a particular disease/condition or treatment/intervention set. The retrieval system is expected to return a list of visits ranked by decreasing likelihood that the visit satisfies the specification.

First, we performed our experiments on sample test data set, produced by NIST [4] and OHSU [5]. We evaluated the variation of retrieval methods on textual information to gain the best result. Then, we first performed a series of experiments to choose the best weighting scheme for given data collection since the choice of the weighting model may crucially affect the performance of any information retrieval system.

We used Terrier [6] as retrieval engine in our experimentations on evaluating the relative merits and drawbacks of different weighting schemes. We started from a traditional bag-of-words representation of pre-processed texts, and used porter stemming [7] algorithm and removed stop words. Figure 2 shows the results of experimentations we performed. The results show that performance of all weighting schemes is so close to each other. However we have chosen DFR_BM25, because it gave the best result in ImageCLEF 2011 Medical Retrieval result [8]. Finally, we submitted these settings as our base-line retrieval.
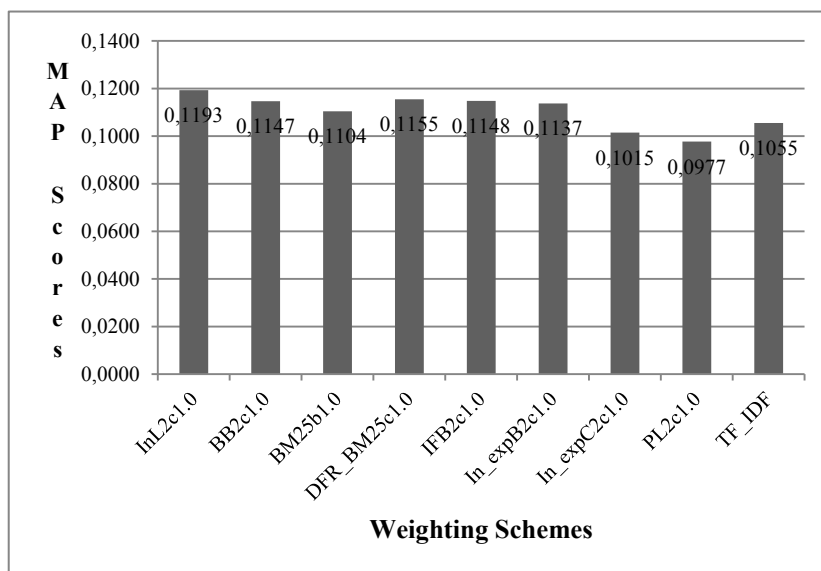

Fig. 2. MAP scores of weighting schemes for textual features

3   Experimentations

We submitted 3 runs to TREC Medical Records track. However none of our runs were judged since we didn't catch the deadline. In all runs we used Terrier as retrieval engine. The first run is our baseline retrieval, where we removed the stop words, used Porter stemmer algorithm and apply the DFR_BM25 weighting model on text retrieval engine system. In second run, we used only single terms exist in UMLS. In the last one, we added term phrases into second run's lexicon. Table 1 shows our runs and the performance measure we gained. Results we obtained shows that result of mixed term and term phrases are better than only term result. However, the results are below the base-line retrieval.

Table 1. Runs of DEMIR group in TREC 2011 Medical Records Track.

RunID	MAP	P10	P20	Rprec	bpref	rel_ret
1	0.3865	0.5500	0.4853	0.3921	0.4757	1524
2	0.3335	0.5029	0.4206	0.3476	0.4418	1489
3	0.3358	0.5353	0.4784	0.3518	0.4558	1493

4   Conclusion

In this study, we evaluated the effect of term-phrases on medical text archives and found that proper use of term phrases may improve the results. However, it should be combined with text other text retrieval methods. Furthermore, the role of proper weighting model selection is also crucial to the performance of text retrieval systems.

Directions for future work includes that retrieve the related semantic type documents matching with topics type. In addition, we also want to classify reports according to ICD-9 codes or admit/discharge diagnosis.

References

1. TREC 2011 Medical Records Track, http://trec.nist.gov/tracks.html
2. Unified Medical Language System (UMLS), http://www.nlm.nih.gov/research/umls/
3. The Lexical Variant Generation, The SPECIALIST Lexical Tools, http://lexsrv3.nlm.nih.gov
4. National Institute of Standards and Technology (NIST), http://www.nist.gov
5. Oregon Health&Science University (OHSU), http://www.ohsu.edu
6. The Terrier IR Platform, http://terrier.org/docs/v2.2.1/
7. Porter, M.F.: An algorithm for suffix stripping, Program: electronic library and information systems, vol. 14, iss. 3, pp. 130--137 (1980)
8. Alpkocak, A., Ozturkmenoglu, O., Berber, T., Vahid A.H., Hamed R.G.; , "DEMIR at ImageCLEFMed 2011: Evaluation of Fusion Techniques for Multimodal Content-based Medical Image Retrieval," 12th Workshop of the Cross-Language Evaluation Forum (CLEF), Amsterdam, Netherlands. (2011)