

# SNUMedinfo at CLEFeHealth2013 Task 3

Sungbin Choi, Jinwook Choi

Medical Informatics Laboratory, Seoul National University, Seoul, Republic of Korea

wakeup06@empas.com, jinchoi@snu.ac.kr

**Abstract.** This paper describes the participation of the SNUMedinfo team at the CLEFeHealth2013 task 3. We submitted 7 runs in total: 1 baseline run using query likelihood model in Indri search engine; 3 runs using passage based language model; 3 runs using passage based language model with lexical query expansion. We tried to incorporate passage-based score into ranking model to reflect the degree of query term cohesion per each document.

**Keywords:** Passage based language model, Query expansion, Web document, Medical information retrieval, Indri

## 1. Introduction

In this paper, we describe the methods in participation of the CLEFeHealth2013 Task 3 – Information retrieval to address patients’ questions. For detailed task description, please see [1].

## 2. Methods

### 2.1 Baseline run

We submitted 1 baseline run (MEDINFO.1.3.noadd) using unigram language model with Dirichlet prior smoothing [2, 3]. Only title field is used as query. The queries are stopped at the query time using the standard 418 INQUERY stopword list, case-folded, and stemmed using Krovetz stemmer. Experimental results are described in Table 1.

**Table 1.** Baseline run result

<b>Runid</b>	<b>MAP</b>	<b>bpref</b>	<b>P10</b>
MEDINFO.1.3.noadd	0.3131	0.3779	0.4800

## 2.2 Passage based language model

We submitted 3 runs using passage based language model [4]. We combined max-scoring passage-based relevance score with unigram language model score. Many web pages contain hierarchical category menu or tables, which does not necessarily represent core topic information. We tried to incorporate passage-based score into ranking model to reflect the degree of query term cohesion. Different weighting parameter is applied on each run. In all 3 runs, only title field is used as query. Experimental results are described in Table 2.

**Table 2.** Passage based language model run result

<b>Runid</b>	<b>MAP</b>	<b>bpref</b>	<b>P10</b>
MEDINFO.5.3.noadd	0.2426	0.3368	0.4040
MEDINFO.6.3.noadd	0.2343	0.3332	0.3600
MEDINFO.7.3.noadd	0.2174	0.3250	0.3480

## 2.3 Passage based language model with lexical query expansion

In addition to section 2.2, we applied lexical query expansion method. UMLS concepts in queries are recognized using MetaMap [5], and then original query is expanded with UMLS preferred terms. Only terms occurring in the discharge summary is chosen for expansion. For MEDINFO.2.3.noadd, only title field is used as query. For MEDINFO.3.3.noadd, title and desc field is used as query. For MEDINFO.4.3.noadd, title, desc and narr field is used as query. Experimental results are described in Table 3.

**Table 3.** Passage based language model with lexical query expansion run result

<b>Runid</b>	<b>MAP</b>	<b>bpref</b>	<b>P10</b>
MEDINFO.2.3.noadd	0.2454	0.3389	0.3980
MEDINFO.3.3.noadd	0.2584	0.3434	0.4040
MEDINFO.4.3.noadd	0.2601	0.3457	0.4060

## 3. Conclusion

We submitted 6 runs all based on passage based retrieval model. Baseline retrieval model is shown to be quite effective. However, contrary to our intention, passage based retrieval score did more harm than good compared to our baseline. We hope to explore more effective method in the future study.

## 4. Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP)(2010-0028631). The Shared Annotated Resources (ShARe) project is funded by the United States National Institutes of Health with grant number R01GM090187.

## 5. References

1. Suominen, H., et al., *Overview of the ShARe/CLEF eHealth Evaluation Lab*, in *CLEF 2013*, Springer Valencia, Spain.
2. Strohman, T., et al. *Indri: A language model-based search engine for complex queries*. in *Proceedings of the International Conference on Intelligent Analysis*. 2005.
3. Zhai, C. and J. Lafferty, *A study of smoothing methods for language models applied to Ad Hoc information retrieval*, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*2001, ACM: New Orleans, Louisiana, USA. p. 334-342.
4. Callan, J.P. *Passage-level evidence in document retrieval*. in *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. 1994. Springer-Verlag New York, Inc.
5. Aronson, A.R. and F.-M. Lang, *An overview of MetaMap: historical perspective and recent advances*. *Journal of the American Medical Informatics Association*, 2010. **17**(3): p. 229-236.