

# Applying Metasearch Technique to Medical Literature Retrieval for Evidence-Based Medicine

Sungbin Choi  
College of Medicine  
Seoul National University  
Seoul, South Korea  
wakeup06@empal.com

Borim Ryu  
College of Medicine  
Seoul National University  
Seoul, South Korea  
borim@snu.ac.kr

Sooyoung Yoo  
Seoul National University  
Bundang Hospital  
Gyeonggi-do, South Korea  
yoosoo0@snu.ac.kr

Jinwook Choi  
College of Medicine  
Seoul National University  
Seoul, South Korea  
jinchoi@snu.ac.kr

## ABSTRACT

Evidence-based medicine has been a highly emphasized concept in the medical domain. To facilitate clinicians' practice of evidence-based health care, current best evidence, which is relevant to the clinical question and also have methodologically high quality, should easily be found. We hypothesized that by counting these two different aspects in ranking algorithm, search engine can automatically retrieve articles which are relevant to clinical question; and also have valid evidence. We approached this problem with combining methodologies. After working out document's query-relevance score and methodological quality score respectively, we combined them using various metasearch methods. For correct evaluation, we built a test collection utilizing preexisting reliable database; Cochrane Reviews.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - *Retrieval models*

## General Terms

Experimentation, Algorithms, Performance

## Keywords

Evidence-Based Medicine, Ranking, Classification

## 1. Introduction

EBM has been widely recognized as an important concept in medical domain. Evidence-based health care is the conscientious use of current best evidence in making decisions about the care of individual patients or the delivery of health services. Current best evidence is up-to-date information from relevant, valid research about the effects of different forms of health care.

Copyright is held by the author/owner(s).  
*SIGIR Workshop on "entertain me": Supporting Complex Search Tasks*,  
July 28, 2011, Beijing, China.

However, practicing EBM in daily clinical care might be challenging though, considering clinician's time scarcity and inadequate search skills. EBM entails appraising step, critically evaluating article's evidence to decide if it is reliable and robust. Searching for relevant article, plus assessing validity of them, must be a complex search task.

We hypothesized that by counting these two different aspects in ranking algorithm, search engine can automatically retrieve articles which is relevant to clinical question; and also has valid evidence. We approached this problem as an information retrieval task with two distinct priorities, finding enough research articles relevant to the clinician's question, and also valid from the perspective of EBM principled methodological criteria. Using various metasearch algorithms, we combined relevancy and methodological quality scores into single ranking.

In this paper, firstly we built test collection using preexisting sources. Secondly, we used a probabilistic retrieval model and a machine learning classifier to work out a document's query relevancy score and a quality score respectively. Finally, we applied various metasearch techniques to rerank documents. Experimental results show that there are significant improvements over baseline (Ranked by quality score only) with our reranking process.

## 2. Method

### 2.1 Test collection

We utilized *Cochrane Reviews* to make our test collection. *Cochrane Reviews* publish systematic reviews of primary research in human health care and health policy. On each review article, objectives are described explicitly, for example, "To assess the effects of donepezil in people with mild cognitive impairment but no diagnosis of dementia". Reviewers, who are domain experts, perform a comprehensive search to find all potentially relevant articles for given topic. When reviewing these retrieved articles, they assess methodological quality for each article, excluding studies not satisfying their predefined criteria, to draw sound conclusion.

We utilized 2009 MEDLINE®/PubMed® Journal Citations, having 17 million MEDLINE documents, as our corpus.

*Cochrane Reviews* topic (e.g. "Donepezil for mild cognitive impairment") was adopted as a search query. Reference lists included in the review were taken as gold standard for each query. On average, there were 11 target documents for each query. We prepared 145 queries, 100 queries randomly assigned to the training set, remaining 45 queries to held-out test set.

## 2.2 Design of our ranking strategy

### 2.2.1 Overall strategy

We organized our ranking strategy as a 3 step process. Our overall strategy is illustrated in Figure 1.

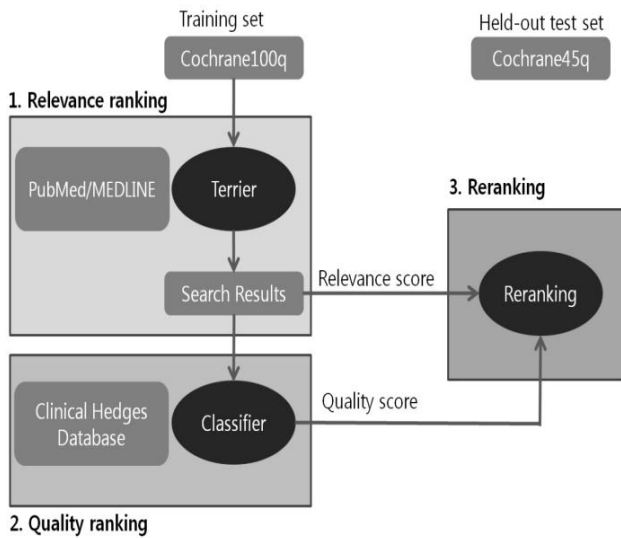


Figure 1. Overall strategy in this study

Firstly, we worked out relevance score using probabilistic retrieval model (Okapi BM25). Title, abstract, Medical Subject Headings (MeSH), publication type fields were extracted indexed.

Secondly, we used machine learning classifier (Naive Bayes, SVM) trained on Clinical Hedges Database, to compute quality score (The value of decision function was used as quality score). We generated various sets of models by trying different classifier and parameter combinations. We tried to find the best classifier model.

Finally, we combined relevance score and quality score using various metasearch methods. Mean Average Precision (MAP) was chosen as our evaluation metric.

### 2.2.2 Reranking

With relevance score and quality score computed, we combined those two scores with various reranking methodologies.

We used a number of simple combination methods referring to [1], and SVM<sup>rank</sup> [2], which used SVM algorithms for prediction of rankings.

## 3. Results

Results on held-out test sets are summarized in Table 1.

Borda-fuse, Weighted-Borda-fuse, Multiplicative combination, Weighted multiplicative combination showed significant increase in MAP (p-value < 0.01) compared to Baseline. Weighted linear combination and SVM<sup>rank</sup> also showed some improvements (p-value < 0.05).

Table 1. Reranked results on held-out test set

Reranking method	MAP %
Relevance ranking	7.4
Quality ranking (Baseline)	8.2
Linear Combination	13.0
Multiplicative Combination	16.4
Borda-fuse	19.6
Weighted Linear Combination	14.7
Weighted Multiplicative Combination	16.0
Weighted Borda Fuse	16.0
SVM <sup>rank</sup>	14.5

## 4. Conclusion

Confronted with difficulty of complex search task in medical domain, we tried to build effective search system, by counting relevance and quality aspects together in the ranking algorithm. Reranking process improved search performance impressively. We hope to make further progress in the future study.

## 5. ACKNOWLEDGMENTS

Use of the Clinical Hedges database was made possible through a collaboration agreement with R B Haynes and N L Wilczynski at McMaster University, Hamilton, Ontario Canada. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MEST) (No. 2009-0075089).

## 6. REFERENCES

- [1] E.A. Fox and J.A. Shaw. Combination of multiple searches. Proceedings of the 2nd Text REtrieval Conference (TREC-2), National Institute of Standards and Technology Special Publication 500-215, pp. 243-252, 1994.
- [2] T. Joachims. Training Linear SVMs in Linear Time, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2006.