A Task-oriented Search Engine for Evidence-based Medicine

Bevan Koopman
CSIRO
Brisbane, Australia
bevan.koopman@csiro.au

Guido Zuccon
Queensland University of Technology
Brisbane, Australia
g.zuccon@qut.edu.au

Jack Russell*
CSIRO
Brisbane, Australia
jackrussell1996@gmail.com

ABSTRACT

Evidence-based medicine (EBM) is the practice of making clinical decisions based on rigorous scientific evidence. EBM relies on effective access to peer-reviewed literature — a task hampered by both the exponential growth of medical literature and a lack of efficient and effective means of searching and presenting this literature. This paper describes a search engine specifically designed for searching medical literature for the purpose of EBM and in a clinical decision support setting.

CCS CONCEPTS

•**Information systems** → *Expert search;*

ACM Reference format:
Bevan Koopman, Guido Zuccon, and Jack Russell. 2017. A Task-oriented Search Engine for Evidence-based Medicine . In *Proceedings of ACM SIGIR conference, Tokyo, Japan, July 2017 (SIGIR'17),* 6 pages.
DOI: 10.475/123_4

1 PROBLEM AND TARGET USERS

While there are mature resources for searching medical literature (the PubMed digital library being a widely used example), these are primarily focused on retrieving literature for research purposes, not for clinical decision support. Research on how clinicians (doctors, nurses or other health professionals) search in a clinical decision support setting [2] has shown that clinicians pose queries within three common clinical tasks: i) searching for **diagnoses** given a list of symptoms; ii) searching for relevant **tests** given a patient's situation; and iii) searching for the most effective **treatments** given a particular condition. An effective search engine should facilitate interactions that support these three tasks. Doing so would lead to both improved retrieval effectiveness and a more economic interaction with the search engine and, ultimately, improved clinical decisions for patients.

2 TASK-BASED SEARCH ENGINE FOR EVIDENCE BASED MEDICINE

A task-oriented approach is at the core of our proposed search engine. Document representation, the retrieval method, and how results are presented to the user are all centred around the three tasks of diagnosis, test and treatment. Figure 1 shows the overall architecture of the system, which was developed using Elasticsearch.[1]

*Student internship while studying at Southern Cross University.
[1]http://www.elastic.co/products/elasticsearch.

SIGIR'17, Tokyo, Japan
 123-4567-24-567/08/06...$15.00
DOI: 10.475/123_4

We detail the indexing, retrieval and visualisation components in the following sub-sections.

2.1 Task-oriented indexing

In the indexing phase, medical articles are fed to a task extraction process that annotates mentions of diagnoses, tests and treatments. Task extraction achieved by first identifying mentions of medical concepts using a medical information extraction system [13]. The identified medical concepts are then mapped to higher level semantic types (e.g., the concept "Headache" belongs to the semantic type "Sign or Symptom"). Each semantic type can then be mapped to one of the three clinical tasks, diagnosis, treatment or test, by consulting the i2b2 challenge guidelines [15] which define a mapping between semantic types and clinical tasks. Once the tasks are identified, the original span of text from the article is annotated with details of the task type. A sample text, with annotated spans, is shown in Figure 2. The resulting annotated articles are indexed into an inverted index with separate fields for diagnoses, tests and treatments.

2.2 Task-oriented retrieval

When a clinician poses a clinical query, they would typically be provided with a long list of search results. In the task-oriented approach, it is desirable to provide the clinician with a summary of the significant diagnoses, tests and treatments. This allows them to quickly gain an understanding of what they might expect to find when examining the search results. In addition, when these summaries are interactive (e.g., the searcher can drill-down on specific tests or treatments) then they are provided with an easy mechanism to navigate the information space. To facilitate such interactions we implement the following retrieval strategy. Given a set of search results, we estimate significant diagnoses, tests and treatments. This is done by scoring each mention of a diagnosis, test or treatment (which can comprise of more than one terms) according to its frequency of appearance within the set of search results (foreground probability) vs. the frequency it appears within the collection as a whole (background probability). The top five diagnoses, tests and treatments are displayed to the searcher (along with the regular search results for that query). Given an individual document within the search results, we also estimate the significant diagnoses, tests and treatments in that document according to IDF and display those to the user. As the underlying retrieval model, we adopted the default Elasticsearch BM25 model.

2.3 Task-oriented visualisation of results

A web-based interface provides the clinician with a means to search and interact with results. A screenshot of the user interface, presenting the results of a search for 'malaria', is shown in Figure 3. The interface provides a single input box where clinicians can enter a free text, keyword query. Retrieval results are displayed as a ranked

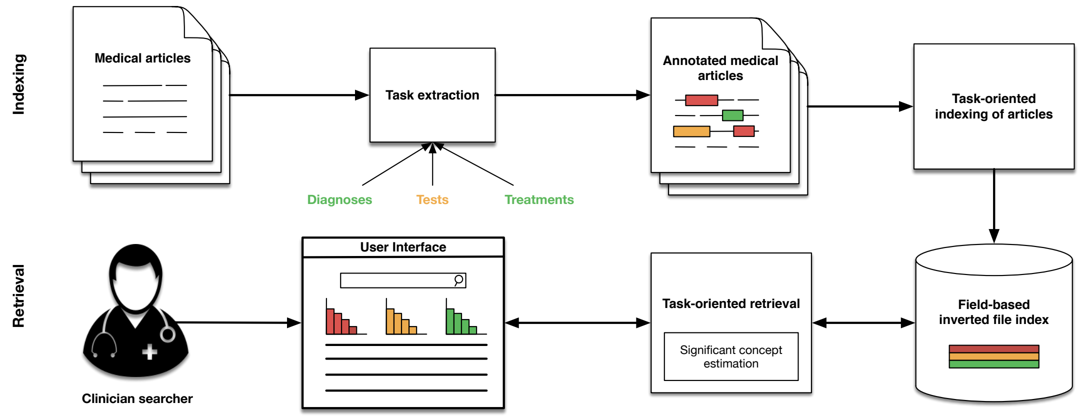

Figure 1: Overview of the task-oriented search approach.

```
Patients with a <test UMLSid="C2238079" title="blood smear">blood smear</test> found to be positive for <diagnosis
UMLSid="C0024530" title="malaria [disease/finding]">malaria</diagnosis> were often administered <treatment UMLSid="C0034414"
title="quinidine [chemical/ingredient]">quinidine</treatment>.
```

Figure 2: Text containing three task annotations: a test ("blood smear"), a diagnosis ("malaria") and a treatment ("quinidine").

list in decreasing order of relevance. Each result is comprised of the article title, journal title, publication date and a snippet.[2]

Three barplots provide an overview of the significant diagnoses (red), tests (orange) and treatments (green). These plots are interactive: clinicians can click on a particular diagnosis, for example, and the set of search results would be filtered to include only articles mentioning that diagnosis; multiple filters can be applied. The purpose of this interface was to allow the clinicians to, firstly, easily get an overview of the search results by inspecting the plots and, secondly, easily navigate the set of search results by applying various filters.

The clinician can view an article by clicking on its title. This opens a dialog displaying the full-text of the article with appropriate annotation displayed. A screenshot for a sample article showing the annotated diagnoses and treatments is shown in Figure 4.

3 COMPARISON WITH EXISTING METHODS

3.1 Qualitative comparison to other systems

While research on how clinicians search indicates that they pose queries according to three mains tasks (diagnoses, tests and treatment) [2], most systems for searching EBM resources do not take these tasks into account. However, structuring IR systems around different categories of informations is a common approach in IR – generally referred to as faceted retrieval [3]. Faceted retrieval reduces mental workload by promoting recognition over recall and by suggesting logical yet unexpected navigation paths to the user [16]. Meaningful facets have been found to support learning, reflection, discovery and information finding [8, 12, 16]. EBM-based search can be viewed as a complex search task [7]: clinicians have complex information needs and are often time pressured. Thus, an IR approach such as faceted retrieval, which reduces mental overhead, is desirable. In this paper, we test the hypothesis that faceted retrieval, which has shown benefits in general web search, can improve search for EBM (Section 3.2).

The importance of access to biomedical literature has resulted in many biomedical-specific retrieval systems [4]. While some systems mention different types of clinical queries (e.g., therapy, diagnosis, harm and prognosis) they typically did not integrate these into the retrieval method or in the way the searcher was presented with or interacted with the search results. Our system uses the clinical tasks as the bases for both retrieval and interaction. Finally, most methods for searching EBM resources were for research purposes, rather than clinical decision support. As such, recall was an important factor (i.e., finding all the relevant articles for a particular information need). In contrast, for clinical decision support, precision can be more important (i.e., finding the article that helps with the clinical task without reading many irrelevant articles). Our system bases the design of the IR system around improving precision via task-based filtering.

Some IR systems use diagnosis, test and treatment information as features in a retrieval model. A common approach here is to map all queries and documents being searched to medical concepts according to an external domain-knowledge resource; matching is then done at the concept level, comparing a query concept with a document concept [9, 10, 14]. Although concept retrieval using tasks has proved effective, the tasks were simply used as features within the retrieval model and never exposed to the clinician [9, 10]. In this study, we attempt to make the task-based information explicit in the way the clinician interacts with the system, as well as the basis for the underlying retrieval model.

In summary, while other studies attempt to extract detailed, structured information from medical articles, we adopt a lightweight approach by considering only diagnoses, tests and treatments. These three tasks were treated in a facet-based approach, which has proved effective in improving search interactions in other domains. The tasks-oriented information is used not only as

[2]We used the default snippet generation provided by Elasticsearch.

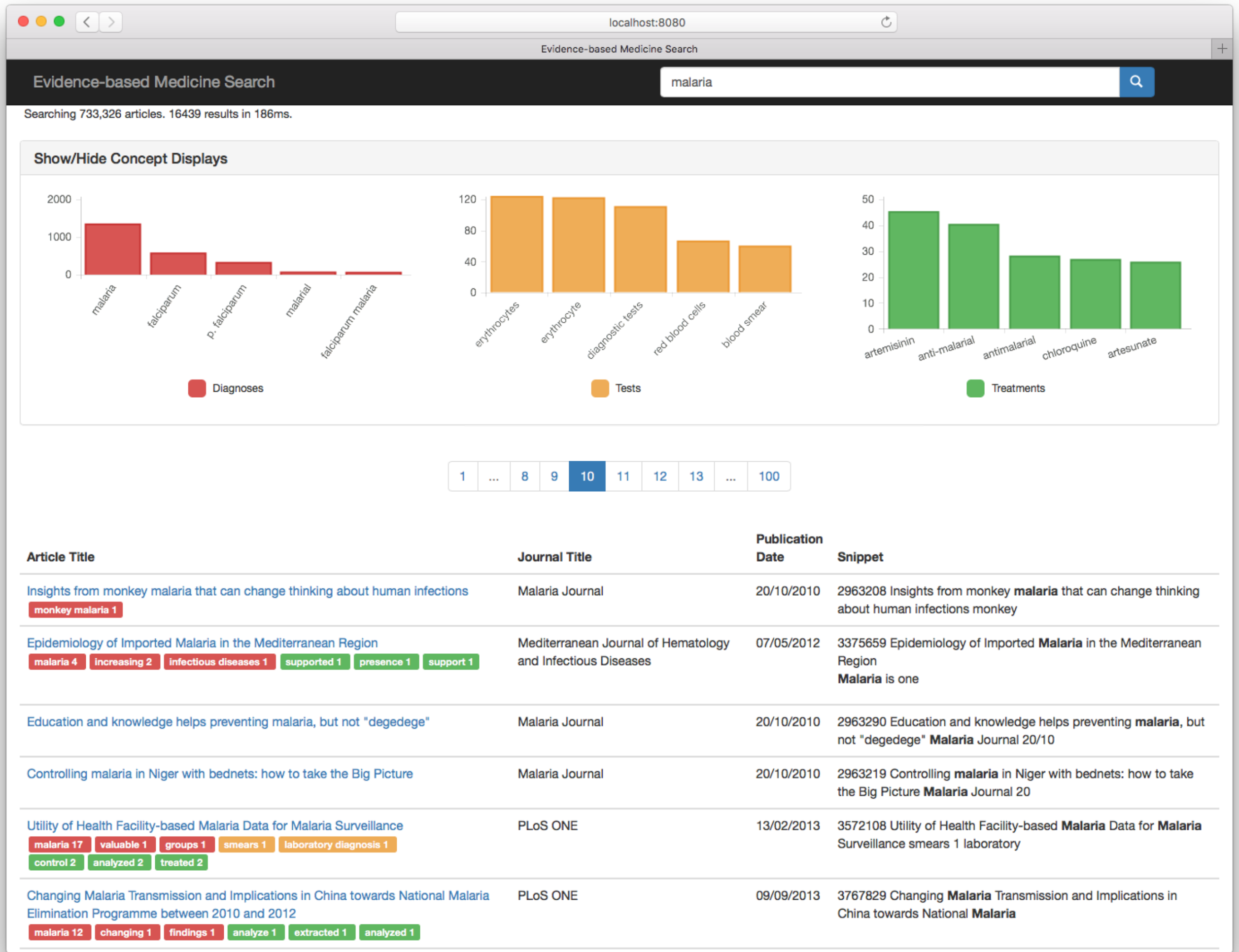

Figure 3: Screenshot showing the results of a search for 'malaria'. Three barplots provide an overview of the significant diagnoses (red), tests (orange) and treatments (green). Individual search results are shown below the barplots.

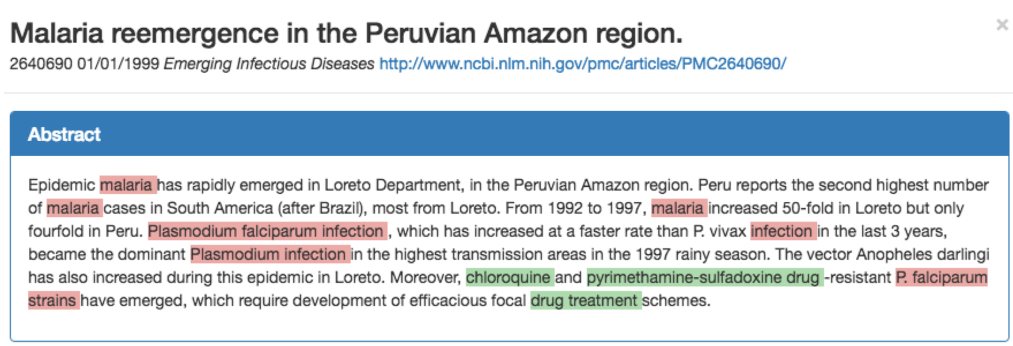

Figure 4: An sample medical article from the user interface showing annotated diagnoses (red) and treatments (green).

Table 1: Retrieval results for task-oriented search. All results showed statistical significance over 'No filter' baseline (paired t test, $p < 0.01$).

Task-oriented filtering	Prec.@10 (%Δ)	Recip. rank (%Δ)
No filter	0.2867	0.4349
Diagnoses	0.3250 (+13%)	0.5271 (+21%)
Tests	0.3283 (+15%)	0.5324 (+22%)
Treatments	0.3167 (+10%)	0.5113 (+16%)

a feature in retrieval but also as a means for improving the way clinicians might interact with the system.

3.2 Empirical evaluation of task-based filters

The task-oriented system was evaluated using the TREC Clinical Decision Support (CDS) test collection. Retrieval effectiveness of the system without filtering was compared with that of the system with a specific task filter. To evaluate the effectiveness of task-based filtering we conducted the following experiment. First, we issued each TREC CDS query topic to the retrieval system and, with no filtering, evaluated the corresponding precision @ 10 and mean reciprocal rank. We then simulated the clinician interacting with the results by selecting individual diagnoses, tests and treatments as filters. Specifically, we filtered the search results, one at a time, by each of the top-five diagnoses, tests and treatments; for example, filter with only the first treatment and evaluate the results, then filter with only the second treatment and evaluate the results, etc. Evaluation measures were calculated after each filter had been applied. Thus, the change in effectiveness between the first ('no filter') search and each of the subsequent task-oriented searches could be calculated. The retrieval effectiveness of the three different task types could be compared and contrasted.

The retrieval effectiveness are shown in Table 1. The results show that task-oriented filtering led to a statistically significant improvement in precision @ 10 and mean reciprocal rank. Filtering on tests exhibited the greatest improvement, followed by filtering on diagnosis and, finally, filtering on treatments.

4 IMPACT AND OUTLOOK

An important consideration for clinicians, who are often time-pressured, is any labour saving benefits that a system can provide. As well as improving retrieval effectiveness, our system can help reduce work load. Specifically, task-based filtering reduces the number of documents the clinician needs to view. A more detailed economic analysis, simulating a user applying various task-based filters, revealed cost savings when compared to not filtering [6]. Plainly put, the cost of choosing and applying a task-based filter is far less than reading even a single non relevant document. Thus, even for the same retrieval effectiveness, viewing less document offers benefits to clinicians.

Search engines for evidence-based medicine may particularly benefit junior doctors, who are still coming to grips with a large and evolving body of medical literature. It is this cohort of users that we hope to recruit as users of the system. While the empirical evaluation has shown improvements in retrieval effectiveness and cost savings in using the task-based system, the ultimate evaluation of the system is with real users, especially given the specialist domain of medical search. An A/B test with and without task-oriented filtering is planned to evaluate the system with real users.

In the current system, users explicitly initiate a search by entering ad-hoc queries via a free-text input box. However, in clinical practice there are situations where a search may be implicitly initiated by a user. A common scenario for this is when a clinician opens an electronic patient record — an effective system would retrieve relevant diagnosis, test or treatment oriented results based on the contents of the patient record. While the current system supports retrieval of such results, the process of generating an effective query from a verbose patient record is needed. Initial research on automatically generating clinical queries is underway [5].

Clinical practice that is informed by scientific evidence is known to improve quality of care [1]. A common means of integrating this evidence-based approach into clinical practice is through clinical decision support systems, which are also known to improve quality of care [11]. The system we describe in this paper provides a means for clinicians to access evidence-based literature in a clinical decision support setting. Improvements in retrieval effectiveness using task-based filters equate to improved access to evidence-based medicine resources. Coupled with the cost savings of using the system, there are good indications that the use of the system can lead to improved clinical decisions and, ultimately, patient care.

REFERENCES

[1] David M Eddy and John Billings. 1988. The quality of medical evidence: implications for quality of care. *Health Affairs* 7, 1 (1988), 19–32.
[2] J.W. Ely, J.A. Osheroff, P.N. Gorman, M.H. Ebell, M.L. Chambliss, E.A. Pifer, and P.Z. Stavri. 2000. A taxonomy of generic clinical questions: classification study. *British Medical Journal* 321, 7258 (2000), 429–432.
[3] Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee. 2002. Finding the Flow in Web Site Search. *Commun. ACM* 45, 9 (2002), 42–49.
[4] William Hersh. 2009. *Information retrieval: a health and biomedical perspective* (3rd ed.). Springer Verlag, New York.
[5] Bevan Koopman, Liam Cripwell, and Guido Zuccon. 2017. Generating Clinical Queries from Patient Narratives: A Comparison between Machines and Humans. In *SIGIR*. Tokyo, Japan.
[6] Bevan Koopman, Jack Russell, and Guido Zuccon. 2017. Task-oriented search for evidence-based medicine. *Inter. Journal of Digital Libraries* (2017), 1–13.
[7] Bevan Koopman and Guido Zuccon. 2014. Why Assessing Relevance in Medical IR is Demanding. In *MedIR, SIGIR*. Gold Coast, Australia.
[8] Barbara H Kwasnik. 2000. The role of classification in knowledge representation and discovery. *Library trends* 48, 1 (2000).
[9] Nut Limsopatham, Craig Macdonald, and Iadh Ounis. 2013. A Task-Specific Query and Document Representation for Medical Records Search. In *ECIR*. Moscow, Russia.
[10] Zhenyu Liu and Wesley W. Chu. 2007. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Information Retrieval* 10, 2 (2007), 173–202.
[11] E. V. Murphy. 2014. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. *Yale J Biol Med* 87, 2 (Jun 2014), 187–197.
[12] Dagobert Soergel. 1999. The rise of ontologies or the reinvention of classification. *JASIST* 50, 12 (1999), 1119.
[13] Luca Soldaini and Nazli Goharian. 2016. QuickUMLS: a fast, unsupervised approach for medical concept extraction. In *SIGIR MedIR Workshop*.
[14] Dolf Trieschnigg, Djoerd Hiemstra, Franciska de Jong, and Wessel Kraaij. 2010. A cross-lingual framework for monolingual biomedical information retrieval. In *CIKM*. ACM, 169–178.
[15] Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *JASIST* 18, 5 (2011), 552–556.
[16] Ryen W White. 2016. *Interactions with search systems*. Cambridge University Press.

REQUIREMENTS FOR PRESENTING THE DEMONSTRATION

Along with the wireless network access and poster mount provided, we kindly request the following resources:

(1) A large computer monitor with VGA connection.

Because our demo can be used as part of the TREC Clinical Decision Support task, we will provide attendees with some of the TREC topics, which they can read, formulate their own queries and use the system to attempt the task. A sample topic is shown below:

Topic# 2014-29
51-year-old smoker with hypertension and diabetes, in menopause, needs recommendations for preventing osteoporosis. **What are the best evidence-based treatments for this patient?**