A Study on the Use of Search Engines for Answering Clinical Questions

Andreea Tutos  Diego Mollá

Centre for Language Technology, Macquarie University

HIKM 2010, 22 January 2010
Contents

Introduction
Evidence Based Medicine; Question Answering for EBM

Corpus
Details of the corpus used

Method
Search engines; question processing; answer extraction

Results
Results of question processing; results of the systems tested; location of the answer

Conclusions
Conclusions; further work
Evidence Based Medicine

Evidence Based Medicine (EBM)

- Individual clinical expertise
- External evidence

Finding External Evidence

- Evidence-based summaries (e.g. UptoDate, the Cochrane Library)
- Clinical literature (e.g. PubMed)
Question Answering for EBM

Problems Facing the Practitioner

(Ely et al. 2002)
- Limited time available
- Difficulty in formulating the question
- Difficulty in synthesising multiple bits of evidence

Question Answering

1. Analyse the question
   - Classification (e.g. Evidence Taxonomy)
   - Query expansion (e.g. using MeSH terms)
   - Identify key information (e.g. PICO)

2. Select the documents (e.g. use a search engine)

3. Find the answer
Purpose of This Research

Goal
To assess the possibilities of using QA technology for EBM

Method
1. Gather a list of questions and answers
2. Send the questions to available search engines and QA systems
3. Manually search for the answers in the results
Evidence Taxonomy

(Ely et al. 2002)

I. Clinical (193 questions)
   A. General (141 questions)
      1. Evidence (106 questions)
         a. Intervention (71 questions)
            What is the drug of choice for epididymitis?
         b. No Intervention (35 questions)
            How common is depression after infectious mononucleosis?
      2. No Evidence (35 questions)
         What is the name of that rash that diabetics get on their legs?
   B. Specific (52 questions)
      What is causing her anaemia?

II. Non-clinical (7 questions)
    How do you stop somebody with five problems, when their appointment is only long enough for one?
Evidence Taxonomy

(Ely et al. 2002)

I. Clinical (193 questions)
   A. General (141 questions)
      1. Evidence (106 questions)
         a. Intervention (71 questions)
            What is the drug of choice for epididymitis?
         b. No Intervention (35 questions)
            How common is depression after infectious mononucleosis?
      2. No Evidence (35 questions)
         What is the name of that rash that diabetics get on their legs?
   B. Specific (52 questions)
      What is causing her anaemia?

II. Non-clinical (7 questions)
    How do you stop somebody with five problems, when their appointment is only long enough for one?
Corpus

- Parkhurst Exchange website
- http://www.parkhurstexchange.com
- Selected 50 medical questions
  - Criteria: select relatively simple questions
- Manually simplify the questions
10 things you should know about... Osteoarthritis

1. A quiet revolution has transformed our understanding of OA. We still don’t know quite what it is, but we no longer believe it’s caused by “wear and tear.” That’s why exercise — both weight-bearing and not — is now highly recommended. The benefit is greatest if there’s weight loss. Losing 10% of weight can reduce OA knee pain by 25%. Arrange patients that joint pain during exercise is not joint damage. ...
Example of Question and Answer

Title
Is watermelon allergenic?

Question
“A 16-year-old female patient had an urticarial reaction from watermelon. She now avoids eating it,” writes ABDULRAHEM LAFTA Hal, MD, of Watson Lake, Yukon. “What substance in watermelon would have caused the attack, and are there other related foods she should now stay away from?”

Answer
Watermelon does contain allergenic proteins that could provoke an IgE-dependent urticarial response. You can refer the patient for allergy skin testing to determine if this fruit was indeed the culprit. Watermelon belongs to a family of foods associated with ragweed pollen. These include cantaloupe, honeydew, zucchini, banana, cucumber and chamomile tea. Individuals suffering from ragweed allergic rhinitis may develop symptoms, often mild, after eating these foods. This is particularly true during or following hay fever season, when their IgE to ragweed is the highest.
Example of Question Simplification

**Original**

Regardless of any other recommendations, should family doctors be immunized with Pneumovax and Menactra or Mejugate?

ANDREY BLITZER, MD, Thornhill, ON

**Simplified**

Should family doctors be immunized with Pneumovax and Menactra or Mejugate?
## Examples of Simplified Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is watermelon allergenic</td>
<td>No Intervention</td>
</tr>
<tr>
<td>When to introduce solids to infants</td>
<td>Intervention</td>
</tr>
<tr>
<td>Should family doctors be immunized with Pneumovax and Menactra or Menjugate</td>
<td>Intervention</td>
</tr>
<tr>
<td>Can cell phones cause cancer</td>
<td>No Intervention</td>
</tr>
<tr>
<td>How much folic acid — 400 g, 1 mg, 5 mg — is recommended before conception and during pregnancy</td>
<td>Intervention</td>
</tr>
<tr>
<td>How to beat recurrent UTIs</td>
<td>Intervention</td>
</tr>
<tr>
<td>How to recognize autism in adults</td>
<td>No Intervention</td>
</tr>
<tr>
<td>Does skin colour affect vitamin D requirements</td>
<td>No Intervention</td>
</tr>
</tbody>
</table>
Search Engines and QA Systems

**Search Engines**
- PubMed
- Google
- Google on PubMed

**QA Systems**
- MedQA (system often not working)
- Answers.com and BrainBoost
- OneLook
# Question Processing

## Method

If a question does not retrieve any relevant results in any of the systems, apply one level of processing until at least a result is returned by a system.

### Levels of Processing

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Original Question</th>
<th>Processed Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduce synonyms/hyponyms</td>
<td>infectious</td>
<td>bacterial</td>
</tr>
<tr>
<td>2</td>
<td>Replace abbreviations</td>
<td>BP</td>
<td>blood pressure</td>
</tr>
<tr>
<td>3</td>
<td>Introduce general medical terms</td>
<td>What is shoulder frozen</td>
<td>What is shoulder frozen syndrome</td>
</tr>
<tr>
<td>4</td>
<td>Eliminate additional terms</td>
<td>Are there any contraindications to dental office visits in pregnancy</td>
<td>Dental office visits in pregnancy</td>
</tr>
<tr>
<td>5</td>
<td>Express medical context</td>
<td>What is the evidence that antibiotics change the course of the disease in infectious conjunctivitis</td>
<td>Are antibiotics recommended for bacterial conjunctivitis</td>
</tr>
</tbody>
</table>
Answer Extraction

- Manual inspection
- Limit to 10 first links
- Abstracts only
- Judge if the abstract contains the answer
### Results of Question Processing

<table>
<thead>
<tr>
<th>Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use</td>
<td>5.41%</td>
<td>10.81%</td>
<td>10.81%</td>
<td>45.95%</td>
<td>27.03%</td>
</tr>
</tbody>
</table>

#### Legend

<table>
<thead>
<tr>
<th>Level</th>
<th>Processing level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Synonyms/hypernyms</td>
</tr>
<tr>
<td>2</td>
<td>Abbreviations</td>
</tr>
<tr>
<td>3</td>
<td>Add general medical terms</td>
</tr>
<tr>
<td>4</td>
<td>Eliminate additional terms</td>
</tr>
<tr>
<td>5</td>
<td>Rephrase using medical context</td>
</tr>
</tbody>
</table>

**Use** How often the level was applied
### Mean Reciprocal Rank of Systems

#### Source | MRR
--- | ---
**Intervention**
Google | 0.54
Google on PubMed | 0.35
PubMed | 0.24
Answers.com | 0.10
MedQA | 0.04
OneLook | 0.04

**No Intervention**
Google | 0.80
Google on PubMed | 0.41
Answers.com | 0.38
PubMed | 0.27
OneLook | 0.04
MedQA | 0.04

\[
MRR = \frac{\sum_{i=1}^{n} \frac{1}{\text{rank}_i}}{n}
\]
Location of the Answer

<table>
<thead>
<tr>
<th></th>
<th>Abstract</th>
<th>Results</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non Intervention</td>
<td>43.48%</td>
<td>17.39%</td>
<td>26.09%</td>
</tr>
<tr>
<td>Intervention</td>
<td>57.89%</td>
<td>10.53%</td>
<td>26.32%</td>
</tr>
<tr>
<td></td>
<td>Recommendations</td>
<td>Purpose</td>
<td>Methods</td>
</tr>
<tr>
<td>Non Intervention</td>
<td>0.00%</td>
<td>8.70%</td>
<td>4.35%</td>
</tr>
<tr>
<td>Intervention</td>
<td>5.26%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
Conclusions

- Main question processing errors are due to extra terms
- Answers to all questions found in the top 10 documents
- Google is best
  - Even better than Google on PubMed
- PubMed’s query expansion method didn’t help much
  - Google’s ranking outweights PubMed’s query expansion
Further Work

▶ Use more questions
▶ Use more complex question
  ▶ Is there a source of realistic questions with their answers?
▶ Test a re-ranked version of PubMed
  ▶ e.g. PubFocus
▶ Use other evaluation measures
  ▶ MRR doesn’t judge quality of the answer presented
Questions?

Introduction
Evidence Based Medicine; Question Answering for EBM

Corpus
Details of the corpus used

Method
Search engines; question processing; answer extraction

Results
Results of question processing; results of the systems tested; location of the answer

Conclusions
Conclusions; further work