Cross-Language and Cross-Terminology Query Suggestion for Health Information Retrieval

Carla Teixeira Lopes (University of Porto / INESC TEC)
E-Health Research Symposium
February 6th, 2015
My academic background

5-years degree in Informatics and Computing Engineering

2-years MSc in Information Management

PhD in Informatics and Engineering
My professional background

Since May 2014  Senior researcher at INESC TEC
Since Sep 2013  Assistant Professor at DEI-FEUP
Oct 2009 – Jul 2013  Lecturer at DEI-FEUP
Mar 2003 – Sep 2009  Lecturer at ESTSP
Aug 1999 – Feb 2003  Project Manager at several companies

DEI-FEUP: Department of Informatics Engineering of the Faculty of Engineering of the University of Porto;
ESTSP: School of Allied Health Sciences of the Polytechnic Institute of Porto
My research interests

My research interests lie at the intersection of information retrieval and human-computer interaction.

I am interested in studying information search behaviour and in developing tools that help people search more successfully.

Lately, I have been focused in exploring how context can help improve the experience of health consumers searching the Web.
My PhD research

Context-Based Health Information Retrieval
Introduction

Motivation

• “the Internet can help us to level the disparity between the physician and the patient, the infallible and the uninformed” (Gigerenzer, 2003)

• Popularity of health web searches by consumers
  – 59% of American adults (Fox and Duggan, 2013)
  – 72% of American Internet users (Fox and Duggan, 2013)
  – 26% of Portuguese (Espanha et al., 2012)

• Search activities take place within a context that can be useful to improve them
Research Overview

Comparative Evaluation of Search Engines
Data Certification Impact
Context Effect on Query Formulation and Subjective Relevance

User Experiment 1
User Experiment 2
User Experiment 3

Query suggestion system

Analysis by User Language Proficiency
Analysis by health literacy and topic familiarity
Query Translation Studies

Context prediction
User Experiment 1

- 41 participants
  - 27 females; 14 males
  - with a mean age of 27.2 years
- 5 simulated work tasks
  - of different specialties, types of clinical questions and severity degrees
  - each containing 4 Information Needs (IN)
- 7 search engines (SE)
  - 4 generalist and 3 health-specific
- Each user chose 2 IN and 4 SE
- For each IN, users
  - Answered an initial questionnaire
  - Formulated a query and submitted it to the SE
  - Assessed the relevance of the top-30 retrieved documents in a 3-graded scale
  - Answered a final questionnaire

Your brother, 42 years old, had kidney stones some years ago and, recently, is having similar pains. Besides these pains, he also bleeds when he pees. You need to know more about kidney stones. For example, you need to know (IN5.1) if the bleeding is related to the kidney stones, (IN5.2) what are the symptoms, (IN5.3) causes and (IN5.4) major problems.
Comparative evaluation of search engines

Overview
Comparative evaluation of search engines

Main conclusions

• Google is the preferred SE
• Google with significantly higher precision than other SE
• Generalist SE surpass health-specific ones
• Overview clinical questions tend to have higher precision
• Gynecology and psychiatry better than dermatology
  – Is it easier to discuss these type of topics online?
  – Most popular specialties in terms of number web pages and searches
• Severe conditions have higher precision in both types of SE
Context effect on query formulation and situational relevance

Overview
Main conclusions

• Traditional methods of IR evaluation may be improved through the incorporation of additional measures.
  – Situational and motivational relevance are not always in harmony.

• Query expansion methods could explore the use of
  – English terms (beware of users proficiency)
  – Advanced operators
  – Medical terms (beware of results’ readability)
User Experiment 2

- 40 participants
- 8 information situations
  - To find a treatment for a particular disease or condition
  - Each associated with 4 queries formulated by the researchers
    - PT/MS, PT/Lay, EN/MS, EN/Lay
- Google as a black-box SE
  - Entire collection -> WebSys
  - HONCode certified pages -> HONSys
- Top-30 results were collected for each query and system
- 8 different tasks per user
- Readability – SMOG
- English Proficiency (EP) – quiz
- Health Literacy (HL)
  - “capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” (Kutner et al., 2006)
- Topic Familiarity (TF)
  - Users’ assessments: task familiarity, previous searches on the topic and knowledge on the topic’s medico-scientific terminology
- Medical accuracy
  - Medical evaluation of users’ answers in terms of correct and incorrect contents
- Motivational relevance
  - “I believe I succeeded in this search task”
Data certification impact

Overview
Data certification impact

Main conclusions

• Users value the diversity provided by generalist SE even if this means including non-certified documents

• In generalist SE, the medical accuracy may be at risk if users don’t understand documents or if the session has documents containing unreliable information

• To assure documents comprehension, the SE must consider their readability and guarantee that document terminology is adjusted to users’ knowledge

• In health tasks, it is advisable to incorporate the medical certification in the set of criteria currently used by the SE

• The classifications “for patients” and “for health professionals” may be useful to personalize the search experience
Research Overview

Comparative Evaluation of Search Engines
Data Certification Impact
Context Effect on Query Formulation and Subjective Relevance

User Experiment 1
User Experiment 2
User Experiment 3

Analysis by User Language Proficiency
Analysis by health literacy and topic familiarity
Query Translation Studies

Query suggestion system
Context prediction
Analysis by user language proficiency

Overview
Analysis by user language proficiency

Main conclusions

• English web health content
  – Has a larger proportion of health-certified documents
  – Is more suited to disseminate health information
  – Is associated with less HTTP errors

• English queries
  – Have higher precision [all EP levels]
  – Have higher medical accuracy [elementary and good EP]

• Despite the higher precision, with English queries, low EP users have less comprehension, obtain less accurate knowledge and feel less satisfied

• Readability is important for comprehension if the user is totally proficient

• Users with higher proficiency have more tendency to use English queries
  – Can EP be inferred from past behaviors?

• Cross-lingual assistance personalized to users English proficiency
  – Can improve non-English consumer health retrieval
  – Can be educational
  – Triggers new search strategies
Analysis by health literacy and topic familiarity

Overview
Analysis by health literacy and topic familiarity

Overview
Analysis by health literacy and topic familiarity

Main conclusions

• A personalized query suggestion system would improve the information retrieval experience
  – Giving access to different documents
  – Fostering terminology learning that can be used in future queries

• Personalization
  – Lay suggestions -> Inadequate HL users and the ones unfamiliar with the topic
  – Medico-scientific suggestions -> users with higher levels of HL and TF
Analysis by health literacy and topic familiarity

Main conclusions

• Readability
  – Essential for a document to be, at least, partially relevant;
  – More important if the document has medico-scientific terminology
  – Crucial in the lower satisfaction levels but not in the higher ones
  – Should be incorporated into ranking algorithms

• Relevance highly depends on its comprehension

• Medico-scientific terminology
  – Rarely used, even by who knows the scientific term
  – Used more often by users in higher levels of HL and TF

• Formulation of health queries is harder for inadequate HL users than for good HL users
Research Overview

Comparative Evaluation of Search Engines

Data Certification Impact

Context Effect on Query Formulation and Subjective Relevance

Exploratory Studies

User Experiment 1

User Experiment 2

User Experiment 3

Query suggestion system

Analysis by User Language Proficiency

Analysis by health literacy and topic familiarity

Query Translation Studies

Context prediction
Query suggestion system

Overview
Query suggestion system

Evaluation
Query suggestion system

User experiment 3

- 40 participants
- 8 simulated situations
- 2 retrieval systems
  - with and w/o suggestions
  - With a log mechanism for every action on the system
- Users performed 8 tasks
  - 3 iterations – (re)definition of a query and assessment of top-10 docs
- English Proficiency
  - Instrument from the European Council and validated by the Faculty of Arts

- Health Literacy
  - Medical Term recognition Test (METER)
- Familiarity
  - Self-assessed in a 5-level scale
- Medical accuracy
  - Medical committee defined a list of correct answers for each query
  - 30% of the answers assessed by two judges
  - Analysis of the knowledge acquired in the session
Query suggestion system

Main conclusions

• Suggestions used more, and more often, in the initial stages of a session
• Suggestions had a good acceptance
• English suggestions tend to be preferred to Portuguese ones
• Medico-scientific suggestions tend to be preferred to lay ones in higher levels of HL
• Retrieval system with suggestions and no personalization
  – Tends to be better in precision, correctness and incorrectness
  – Tends to be slightly worse in terms of motivational relevance
  – Incorrectness difference is significant
Query suggestion system

Main conclusions

• Best performance of the suggestion tool when
  – Is used a suggestion tool
  – All terms from suggestions are used

• Still, using suggestions as sources of terms, improves medical accuracy

• A suggestion system with personalization by EP and HL outperforms a system without personalization
  – Personalization bias users to the suggestions more beneficial to them

<table>
<thead>
<tr>
<th></th>
<th>EP1</th>
<th>EP2</th>
<th>EP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>-</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>PT</td>
<td>✔</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HL1</th>
<th>HL2</th>
<th>HL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lay</td>
<td>✔</td>
<td>✔</td>
<td>-</td>
</tr>
<tr>
<td>MS</td>
<td>-</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
Questions?

Carla Teixeira Lopes (University of Porto / INESC TEC)
E-Health Research Symposium
February 6th, 2015