Extracting Intention from Web Queries – Application in eHealth Personalization

George Drosatos
Avi Arampatzis and Eleni Kaldoudi

School of Medicine
Dept. of Electric and Computer Engineering
Democritus University of Thrace

This work was supported by the FP7-ICT project CARRE (No. 611140), funded in part by the European Commission.
Personalized e-Health System

- **Personal Sensors**
- **Personal Health Records**
- **Mental Intenions, Plans, Beliefs, etc.**
- **Physical Health Records**
- **Electronic Health Records**
- **Health Insurance**
- **Financial Data**

**e-Health Application**

- e.g.: ✓ Personal ✓ Clinical

**Privacy**

**Personal Data**
It is a **EU** funded project in the area of **cardiorenal** with focus to provide **personalized health**

Personal data: Sensor data (e.g. activity and blood pressure), PHR and patient’s **intentions** (travel, diet, diseases, etc)
Investigation of Patient’s Online Interaction

- Capture Personal Information
  - Goal: Detect intentions

- Possible Sources:
  - Social Media: Facebook, Twitter, etc
  - Browsing History
  - Web Searches

- **First choice**: Web searches to extract intentions
  - Good source to reveal user’s interests and intentions
  - Web search engines are one of the most popular uses of the web, e.g.
    >70% of internet users report looking online for health information

---

http://www.pewinternet.org/2013/01/15/health-online-2013/
Privacy & Legislation

- **What is privacy?**
  - “The right to be let alone” [Warren and Brandeis, 1890]
  - “The right of the individual to decide what information about himself should be communicated to others and under what circumstances” [Westin, 1970]
  - The right to informational self-determination [1983]

- **Personal Data**: Any information that refers to a person

- **Related Legislation**: e.g. EU Data Protection Directive 95/46/EC
  - Indicative principles:
    - Reported and transparent processing
    - Finality & Purpose Limitation
    - Personal data quality
    - Security
    - Personal data traffic outside EU

---

<table>
<thead>
<tr>
<th>What Revealing Search Data Reveals</th>
<th>Why the search</th>
</tr>
</thead>
<tbody>
<tr>
<td>A sample of Thelma Arnold’s search data released by AOL</td>
<td>&quot;I was thinking about my grandchildren&quot;</td>
</tr>
<tr>
<td>Searching for 'sex'</td>
<td>&quot;I was looking for some.&quot;</td>
</tr>
<tr>
<td>Searching for 'a woman was in the public bathroom copying'</td>
<td>&quot;A woman was in the public bathroom copying. She was going through a divorce, I thought there was a place called Dances by Lori, for singles.&quot;</td>
</tr>
<tr>
<td>Searching for 'my house was worth'</td>
<td>&quot;I wanted to find out what my house was worth.&quot;</td>
</tr>
</tbody>
</table>

---

G. Drosatos, IUPESM – World Congress 2015: 5
Privacy-friendly Architecture

main principle: preserve the patients’ privacy
(1) Offline, initialization process
- predefine query categories (~250)
- create index of documents = the collection of top most related documents to each category from a set representative of the entire web (ClueWeb 09_b)

(2) Real-time repetitive process
- run user query in the index
- based on the results, associate user query with predefined categories

privacy preserving: step #2 process is performed on user-side

Our Implementations

*(Open Source)*

- **Query Detector** as a browser extension
  - Firefox
  - Chrome
- **User Intention Extractor** as a Java application
  - Platform independent
Our Implementations

*(Open Source)*

- **Query Detector** as a browser extension
  - Firefox
  - Chrome
- **User Intention Extractor** as a Java application
  - Platform independent
Our Implementations
(Open Source)

- Query Detector as a browser extension
  ◦ Firefox
  ◦ Chrome
- User Intention Extractor as a Java application
  ◦ Platform independent
Example of Detected Intentions
Conclusions, Current & Future Work

Conclusions
- Provide a proof of concept
- Apply a privacy by design approach in our methodology

Work in Progress
- Improve the technique of query classification
- Determine the safe detected intentions based on classification technique (without a fixed limit, e.g. n=3)
- Perform a user study in the side of patients in order to determine the correctness of intentions

Future Work
- Detect intentions from other online activities (e.g. social media) of patient
- Investigate how to utilize the intentions in a Decision Support System (DSS)
Thank you!

- Slides & Reprints: [http://www.drosatos.info](http://www.drosatos.info)
- You can find binaries and source codes at:
  
Acknowledgement

This work was supported by the FP7-ICT project CARRE (No. 611140), funded in part by the European Commission.

CARRE Project: Personalized patient empowerment and shared decision support for cardiorenal disease and comorbidities.