

Extracting Intention from Web Queries – Application in eHealth Personalization

George Drosatos
Avi Arampatzis and Eleni Kaldoudi

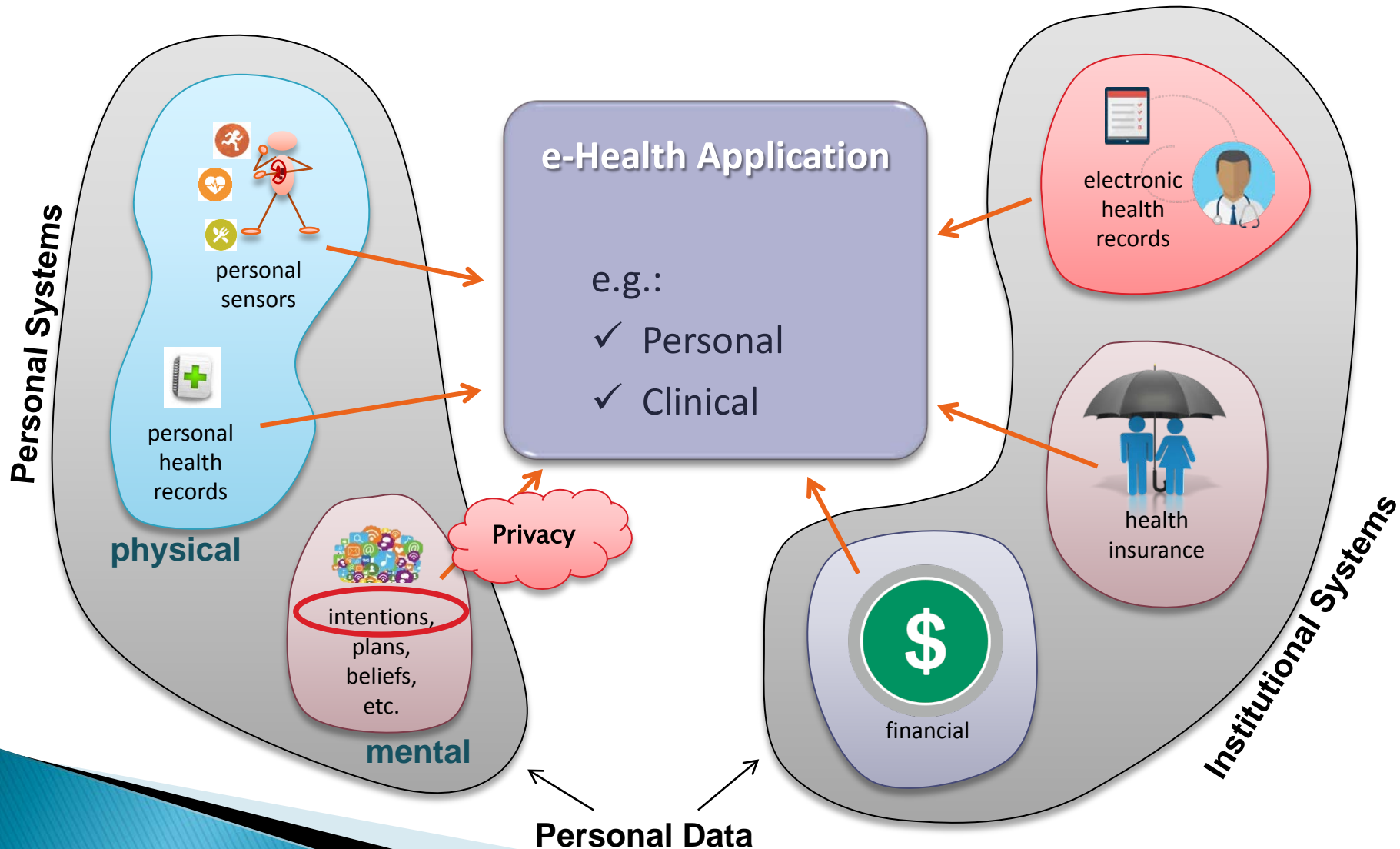
School of Medicine
Dept. of Electric and Computer Engineering
Democritus University of Thrace



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

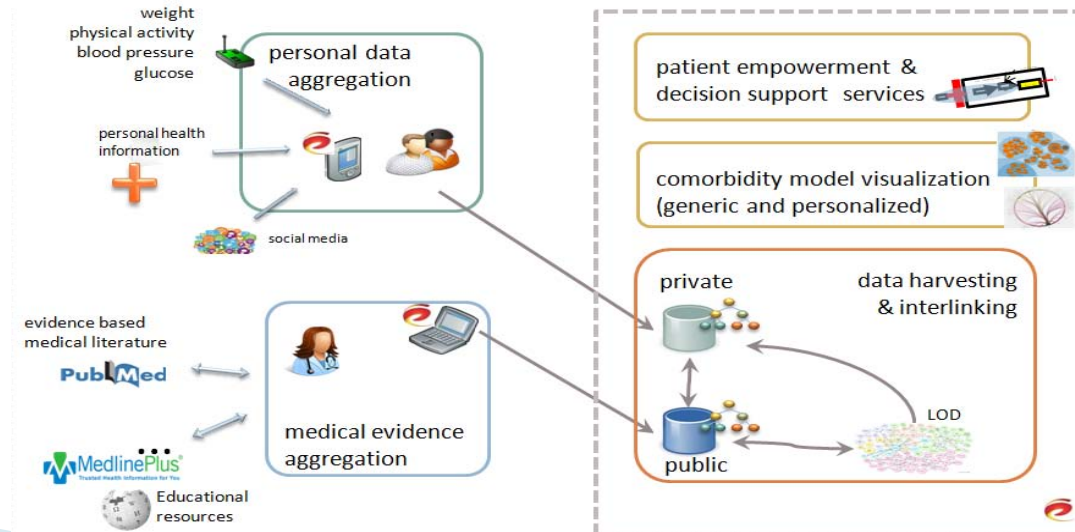


Personalized e-Health System



CARRE Project

- ▶ 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

**The Pew Research Center. (2013) Health Online 2013.*

<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

What Revealing Search Data Reveals

AOL posted, but later removed, a list of the Web search inquiries of 658,000 unnamed users on a new Web site for academic researchers. An interview with one of those unnamed users, Thelma Arnold, combined with her data reveal what she was searching for, why and on which Web sites.

A sample of Thelma Arnold's search data released by AOL

4617748	swing sets	2006-04-24	15:39:30	4	http://www.buyswingset.com
4617748	swing sets	2006-04-24	15:39:30	9	http://www.buychoice.com
4617748	swing sets	2006-04-24	15:39:30	10	http://www.creativeplaythings.com
4617748	swing sets	2006-04-24	15:39:30	5	http://www.childlife.com
4617748	swing sets	2006-04-24	15:39:30	6	http://www.planitplay.com
4617748	that do not shed	2006-04-28	9:55:54	2	http://www.gopetsamerica.com
4617748	dog who urinate on everything	2006-04-28	13:24:07	6	http://www.dogdiaryusa.com
4617748	walmart	2006-04-28	14:07:32	1	http://www.walmart.com
4617748	women's underwear	2006-04-28	14:12:28	10	http://www.bisrate.com
4617748	jcpenny	2006-04-28	14:16:05		
4617748	jcpenny	2006-04-28	14:16:49	1	http://www.jcpenny.com
4617748	tortois and turtles	2006-04-29	13:12:47		
4617748	manchester terrier	2006-05-02	9:05:31	1	http://www.manchestertierrier.com
4617748	della	2006-05-02	11:45:26		
4617748	fingers going numb	2006-05-02	17:35:47		
4617748	dances by laura	2006-05-02	17:59:32		
4617748	dances by lori	2006-05-02	17:59:57		
4617748	single dances	2006-05-02	18:00:18	1	http://singlesingles.com
4617748	single dances in atlanta	2006-05-02	18:01:13		
4617748	single dances in atlanta	2006-05-02	18:01:50		
4617748	dry mouth	2006-05-06	16:49:14	2	http://www.mayoclinic.com
4617748	dry mouth	2006-05-06	16:49:14	8	http://www.wrongdiagnosis.com
4617748	thyroid	2006-05-06	16:55:36		
4617748	thyroid	2006-05-06	16:55:44		
4617748	competitive market analysis of homes in liburn	2006-05-14	12:14:52		
4617748	competitive market analysis of homes in liburn	2006-05-14	12:16:17		
4617748	competitive market analysis of homes in liburn	2006-05-14	12:16:43		

Why the search

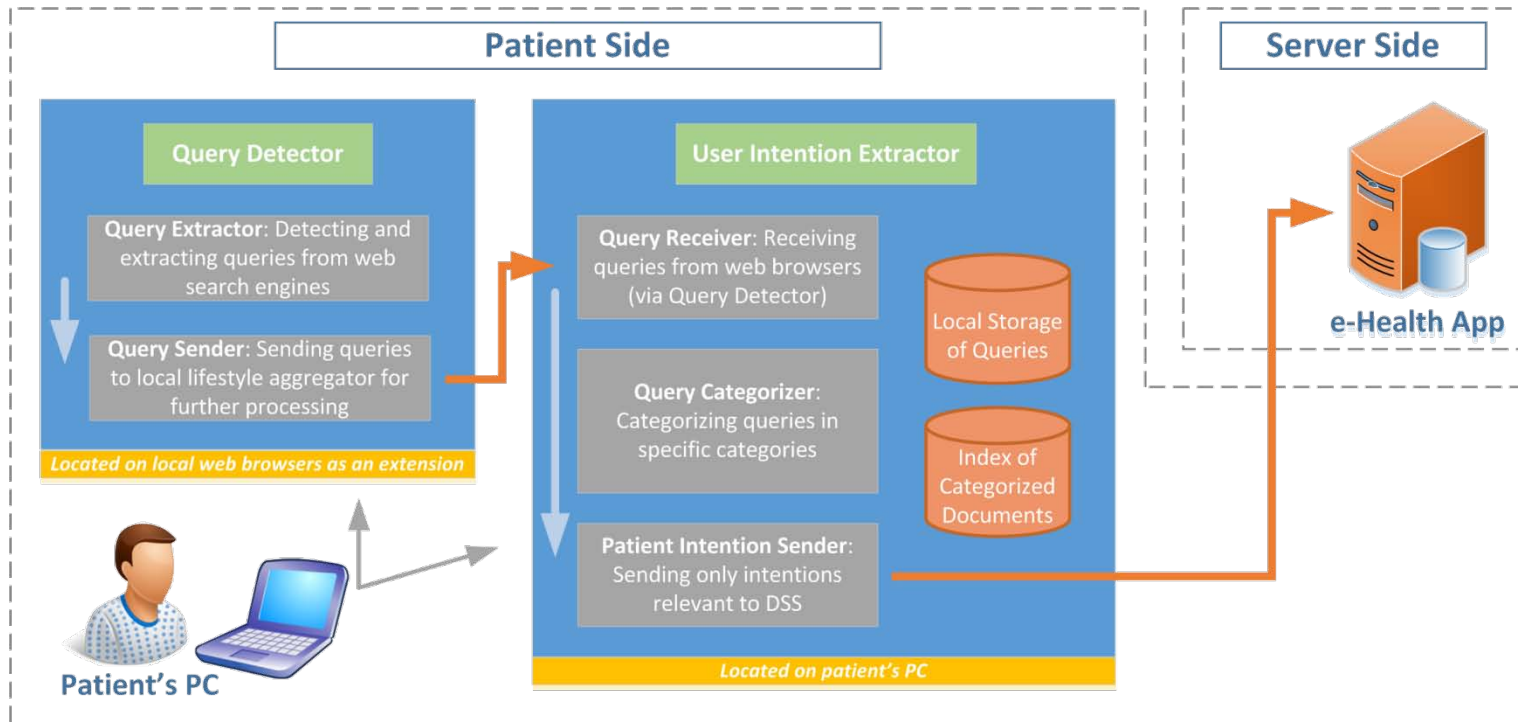
"I was thinking about my grandchildren"

"I was looking for some."

"A woman was in the [public] bathroom crying. She was going through a divorce. I thought there was a place called 'Dances by Lori,' for singles."

"I wanted to find out what my house was worth."

Privacy-friendly Architecture

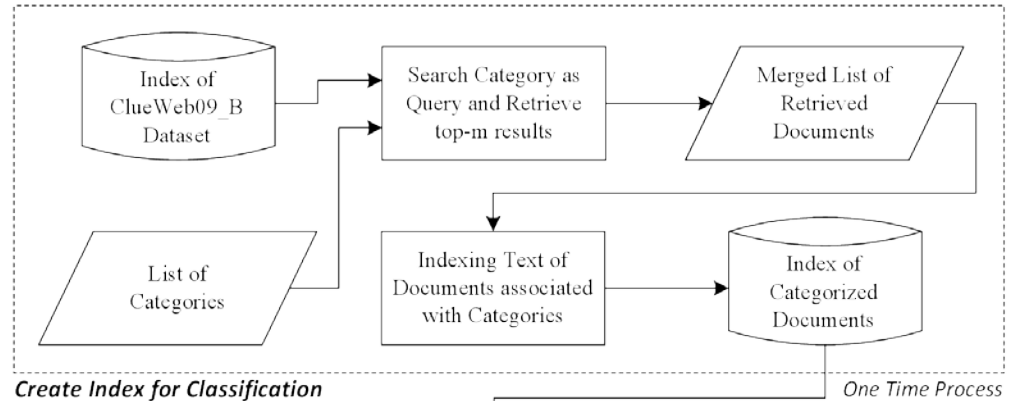


main principle: **preserve** the patients' **privacy**

Extract Users Intentions via Query Classification*

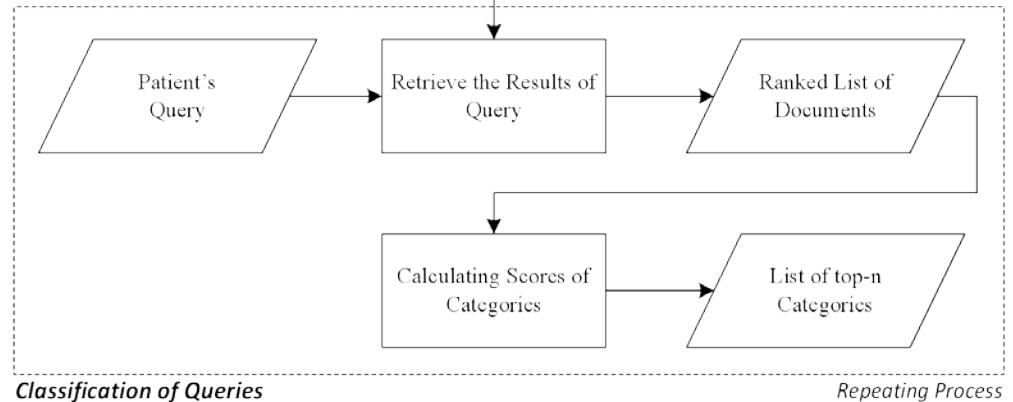
(1) Offline, initialization process

- 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



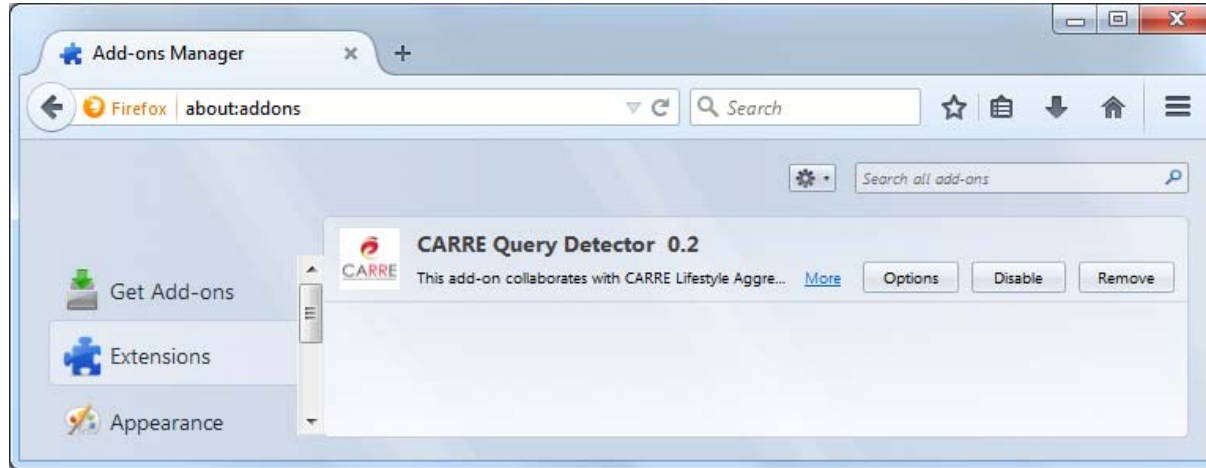
privacy preserving: step #2 process is performed on user-side

*Agrawal R., Yu X et al. , Neural Information Proc., Springer, 7064 of LNCS, pp 148-157, 2011

Our Implementations

(Open Source)

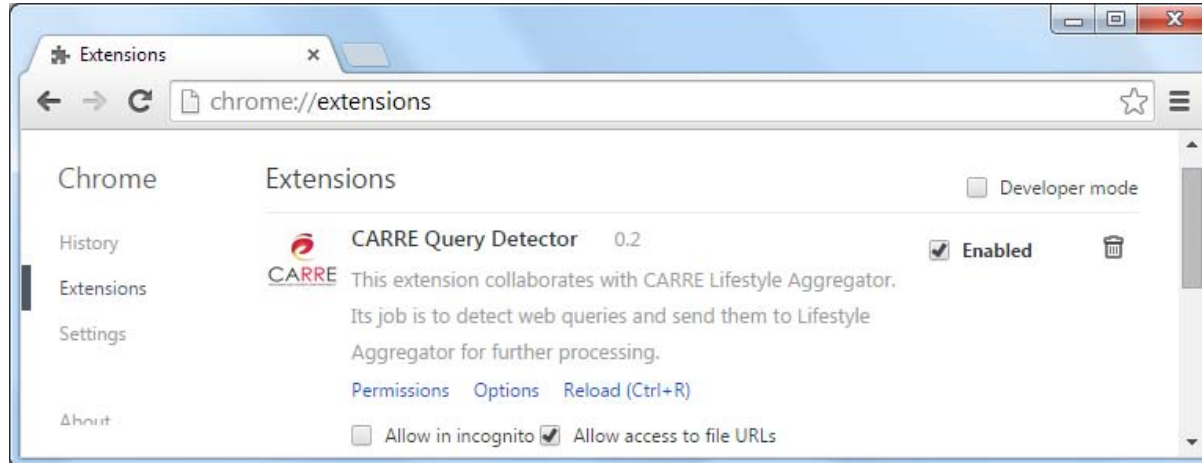
- Query Detector as a browser extension
 - Firefox ←
 - Chrome
- User Intention Extractor as a Java application
 - Platform independent



Our Implementations

(Open Source)

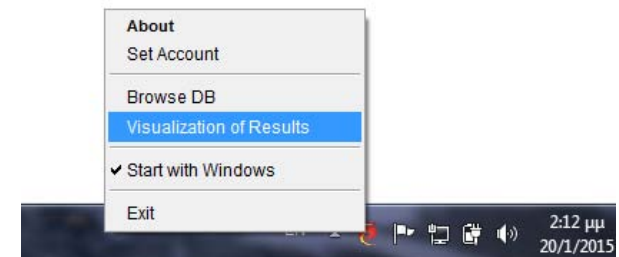
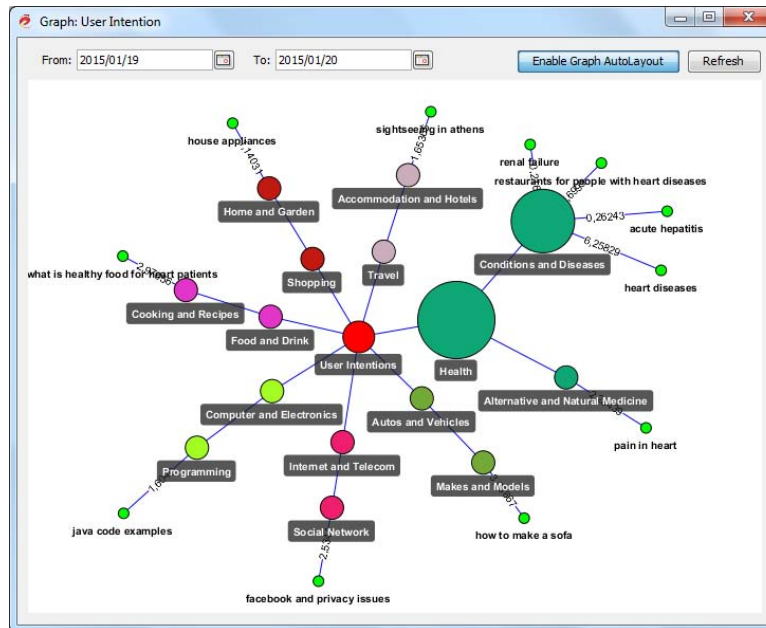
- Query Detector as a browser extension
 - Firefox
 - Chrome ←
- User Intention Extractor as a Java application
 - Platform independent



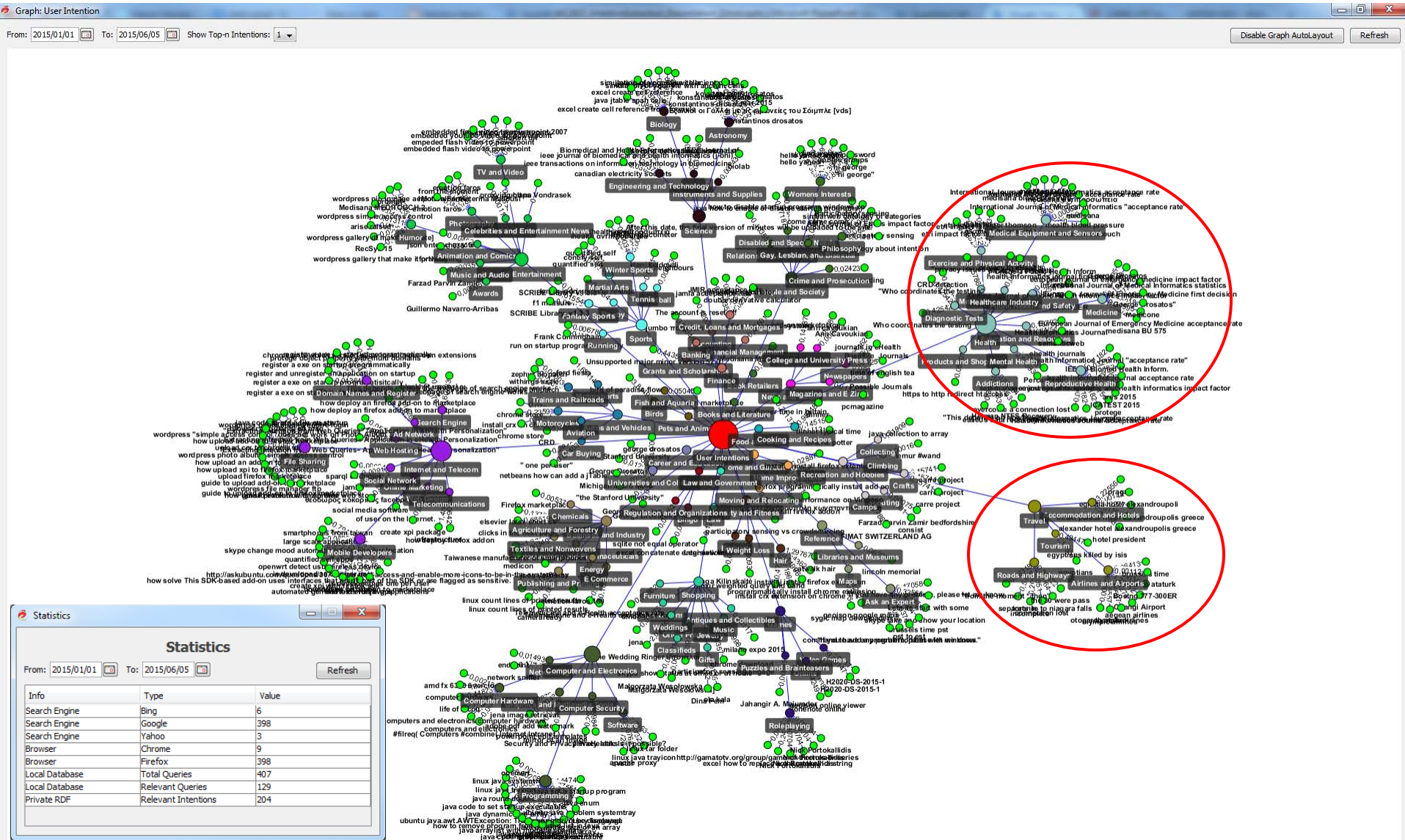
Our Implementations

(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>
- Online Demonstration: <http://youtu.be/IMHlIbwcDRY>
- You can find binaries and source codes at:

<https://www.carre-project.eu/innovation/web-lifestyle-data-aggregator/>

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.