Semantic annotation of health information

Commercial use (Y/N): N
Related project (*if any – please specify funder and project name*): Khresmoi, EC FP7
Author: Angus Roberts
Organization: University of Sheffield, United Kingdom
Date: 22/12/2014

Use case description – *Describe the problem you are trying to address, the research area it affects, and the TDM application used. Please state all participants, whether and how the results were subsequently used.*

The Khresmoi project (http://www.khresmoi.eu/) was an EC FP7 project that is now being taken up by SMEs and commercialised via H2020 funding. Khresmoi developed a multilingual multimodal search and access system for biomedical information and documents. This was achieved in part by effective automated information extraction from biomedical documents, including improvements using manual annotation and active learning, and automated estimation of the level of trust and expertise of target users. Information extracted from unstructured or semi-structured biomedical texts and images was linked to structured information in knowledge bases. Khresmoi provided support of cross-language search, including multilingual queries, and returning machine-translated pertinent excerpts.

A full list of project partners can be found at http://www.khresmoi.eu/consortium/

Content sources – *List the targeted content sources and types of content to used.*

1. Web pages from health information providers

2. Specific health information sources:

     Cochrane reviews

     MEDLINE

3. Medical terminologies and ontologies

Targeted users – *Describe end users, their number and expertise.*

Number of users in Khresmoi demonstrators would be in the 100s. The Kconnect follow-on project aims to take this to commercial scale. Users fall in to two groups:

1. General medical practitioners. High level of domain expertise. Mixed level of expertise as regards technology, intellectual property and copyright issues.

2. General public. It should be assumed that this group has a generally low level of expertise.

Impact – *Describe all possible impacts of the use case. Specify any cross-border, societal and economic effects (if possible state any monetary benefits and market advantages) .*

1. Members of the general public frequently seek medical information online. This process is currently inefficient, unreliable and potentially dangerous. It is thus important that they are provided with **reliable and understandable medical information** in their own language.

2. Medical doctors need **rapid** and **accurate answers** – a search of MEDLINE takes on average 30 minutes, while doctors have on average 5 minutes available for such a search. Furthermore, over 40% of searches do not yield the information required.

Constraints – *Describe any legal, technical, economic, societal, organizational, cultural, multi-lingual or other limitations and you have encountered.*

The project faced copyright and licensing issues in some areas. Although many web pages are clearly unrestricted, some contain explicit copyright restrictions. In the case of USA sites, it could be argued that what Khresmoi was doing was fair use for research as stated by the United States copyright law. However, this would not be the case if Khresmoi were to be commercially exploited. Additionally, other non-USA sites contained restrictive copyrights.

Comments – *Other comments and any recommendations you may have.*

There is also an issue with third-party health information sources (such as the Cochrane Reviews). This is not limited to Khresmoi, and has been found in other projects. While getting a license and using Cohrane summaries is not a problem, the full text is by subscription only (although some countries do have national subscriptions), and it is not clear how it may be accessed for text mining outside of this license.

An additional problem arises from sites that publish personal details (e.g. of medical professionals), and the legality of reproducing these.