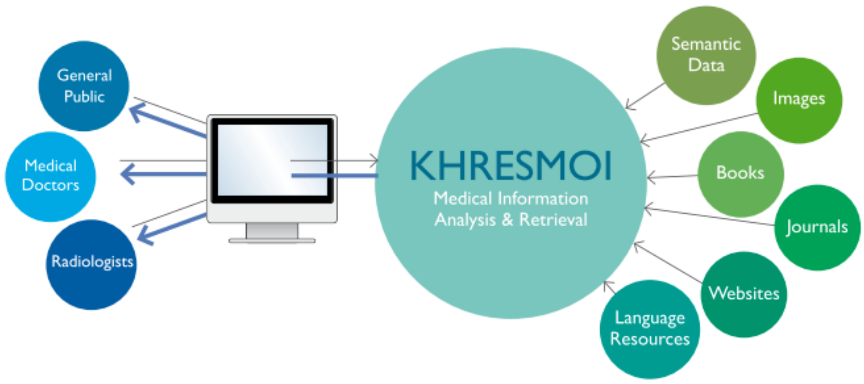
1 Publishable summary

1.1 Khresmoi Objectives

Khresmoi aims to develop a multilingual multimodal search and access system for biomedical information and documents. This will be achieved by:

- Effective automated information extraction from biomedical documents, including improvements using crowd sourcing and active learning, and automated estimation of the level of trust and target user expertise
- Automated analysis and indexing for medical images in 2D (X-Rays), 3D (MRI, CT), and 4D (MRI with a time component)
- Linking information extracted from unstructured or semi-structured biomedical texts and images to structured information in knowledge bases
- Support of cross-language search, including multilingual queries, and returning machine-translated pertinent excerpts
- Adaptive user interfaces to assist in formulating queries and display search results via ergonomic and interactive visualizations

The research will flow into several open source components, which will be integrated into an innovative open architecture for robust and scalable biomedical information search.



1.2 Khresmoi Use Cases

- **Members of the general public** want access to reliable and understandable medical information in their own language. At present, web search engines are the most-used tool for finding medical information on the internet, but the web pages returned are of varying quality, with no indication of the reliability of the information.
- **Clinicians and general practitioners** need accurate answers rapidly – a search on PubMed requires on average 30 minutes [1], while clinicians typically have 5 minutes available [2].
- **Radiologists are drowning in images** and need improved automated support for their analysis – at larger hospitals over 100GB (over 100,000 images) are produced per day. The huge archives of radiology images available in hospitals (in anonymised form) have a large potential to assist radiologists with diagnosis if search by visual similarity in these archives were possible.

Representative groups of end users are available for sizable evaluations within the project. These include access to the general public via a medical search engine with 11,000 queries per day (Health on the Net Foundation), a professional association of 2,700 physicians (Society of Physicians in Vienna), and two radiology departments with 175 radiologists (Medical University of Vienna and University Hospitals of Geneva).

1.3 Work Performed and Main Results

The main activity to be completed in the first year of Khresmoi was to elicit the current search practices of the target groups, as well as their requirements and wishes for future health and medical information search systems, as described in Section 1.3.1. Furthermore, work on developing the Khresmoi search technologies commenced, described in Section 1.3.2. Finally, Section 1.3.3 describes the collection of the data that will be used in the Khresmoi search engine.

1.3.1 End user surveys

Three surveys to elicit end user search practices and requirements were done. The first was an online survey aimed at members of the general public. The second survey was aimed at medical doctors, and was also performed online. A survey of the radiologists was done on a smaller scale, but included initial experiments on using eye tracking to determine the parts of an image that a radiologist concentrates on. The analysis of the results of these surveys will contribute to the specification of the Khresmoi system requirements in the second year of the project.

The survey of the general public represents mostly the opinions of highly educated users occupied in areas of healthcare and IT. A total of 385 responses were collected. Overall, representatives from 42 countries around the world filled in the questionnaire with the top contributors being from France and Spain. One of the questions requested the identification of the most helpful tools in a search engine for medical information, for which the top five responses are shown in Figure 1:

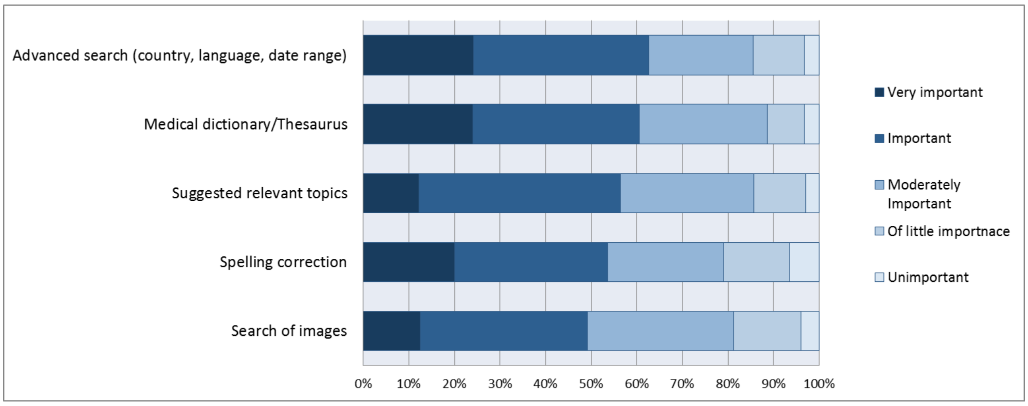


Figure 1: Most helpful tools for a medical search engine.

The survey of physicians received over 550 responses from 20 countries, with the majority of responses from Austria and Switzerland. A question on the importance of search tools resulted in the following tools being rated as the most important (numbers are the average of responses on an importance scale from 1 to 3), shown in Figure 2.

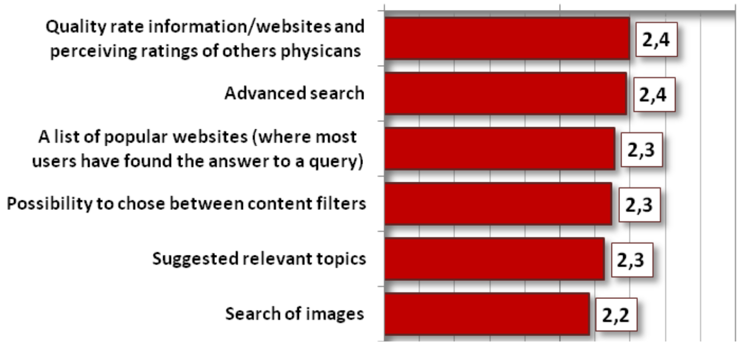


Figure 2: Important aspects for physicians searching for medical information.

The radiologist eye tracking experiments were performed on a clinical workstation where radiologists performed a diagnosis on a few example cases of different anatomic regions. Eye tracking allowed the identification of clear and well-defined regions of attenuation or interest. This is shown in Figure 3, in which red areas attracted the most attention of radiologists. Sometimes only one major region is viewed but for other images several well defined regions are viewed. The regions important for a diagnosis are often very small and linked to a specific organ, highlighting the need to identify anatomy and also a need for local retrieval and detection of anomalies. From the radiologist survey, it became clear that image search is a common activity, often for finding interesting cases or articles for a differential diagnosis. Sources are most often Google but also specialized search engines such as Goldminer and Yottalook were mentioned.

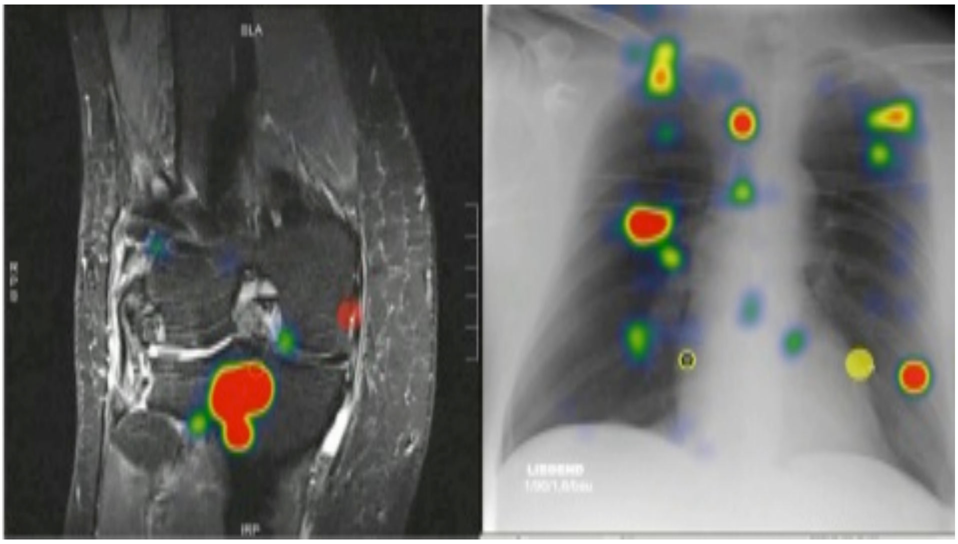


Figure 3: Example heat maps from the eye tracking experiments.

1.3.2 Khresmoi technologies

At the heart of the Khresmoi search engine will be a number of open source components, extended by the work done in Khresmoi. New components based on research results obtained in Khresmoi will also be included. Work on extending the existing components to meet the needs of the project has begun in year one. This has included the setting up of a basic query expansion service for Mimír[4], a search engine capable of indexing and searching over text, annotations, semantic schemas (ontologies), and semantic meta-data (instance data); as well as the development of a browser-based user interface and image display capabilities for ezDL[5], a framework for interactive search applications. Furthermore,

[4] http://gate.ac.uk/family/mimir.html

[5] http://ezdl.de/

basic research has led to innovative techniques for anatomical structure identification and localisation, visual-based anatomy retrieval and anomaly retrieval in 3D radiology images.

The first steps towards integrating the components that will make up the Khresmoi search engine were also taken this year. At the consortium meeting in March 2011, a focus was placed on familiarising the Khresmoi team with the components to be used. This led to the ambitious target of developing four prototypes illustrating different aspects of the Khresmoi vision by September 2011. These prototypes demonstrate: text search, machine translation, 2D radiology image and text search, and 3D radiology image search.

1.3.3 Data

A search engine is of little use if it does not access useful information. We have therefore paid attention to the information that will be indexed by the Khresmoi search engine. An initial knowledge base containing basic biomedical knowledge (Pubmed including UMLS MeSH, LinkedCT, Pubmed, Drugbank, Geonames) has been set up. The information accessed by the initial prototypes includes the set of HONcode-certified websites (the HONcode is the certification of medical websites managed by the Health on the Net Foundation) and a subset of MEDLINE abstracts. The Cochrane Collaboration has also made available the complete text of their systematic reviews for use in Khresmoi. Automatic annotation of entities in the texts is being done, coupled with manual correction of the annotations to flow back into improving the automated annotation. A group of annotators provided by sub-contractor Lighthouse is making use of the GATE Teamware software to perform the manual annotation that will help to produce a large gold standard to optimize existing tools.

For the radiology application, over three Terabytes of anonymised data, mainly Computed Tomography (CT) and Magnetic Resonance (MR) images were obtained. Part of these data were manually annotated by anatomical location (e.g. head, chest, hand) and anomaly (e.g. bone density in osteoporosis). These data will serve as a basis for semi-supervised learning in anatomy identification algorithms, as well as for anomaly retrieval.

1.4 Expected Final Results and their Potential Impact and Use

Medical Impact: Improve the access to medical information for doctors, so that they have more time to talk to and to treat patients, having all the information required for doing so more effectively.

Convert the flood of radiological image data into a boon instead of a curse.

Scientific Impact: Address the lack of publicly available large-scale data sets and realistic task-based scenarios on which to assess new technologies in the medical domain.

Make available cutting edge techniques implemented in open source software.

Industrial Impact: Improve existing open source products' stability, features and performance, and hence their attractiveness and suitability for wider deployment.

Public Impact: Members of the public will be using the Health on the Net search engine, improved by the Khresmoi technology, relatively early in the project.

1.5 Project Public Website

http://khresmoi.eu/

1.6 References

[1] W. R. Hersh, D. H. Hickam, How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review, Journal of the American Medical Association, Vol 280, No. 15, 1998

[2] A Hoogendam, A. F. H. Stalenhoef, P. F de Vries Robbé, A. J. P. M. Overbeke, Answers to Questions Posed During Daily Patient Care Are More Likely to Be Answered by UpToDate Than PubMed, J Med Internet Res, Volume 10, Number 4, 2008.