

1 Publishable summary

1.1 Khresmoi Context and Objectives

The Khresmoi project addresses the challenges of searching through large amounts of radiology data, including Magnetic Resonance (MR) and Computed Tomography (CT), in hospital archives, as well as general medical information available on the internet. For the latter, it addresses the issues of trustworthiness and comprehensibility levels of the documents. The project consortium, consisting of twelve partners from nine European countries, is developing a multilingual multimodal search and access system for health information and documents. The system will allow text querying in several languages, in combination with image queries. It will return translated document summaries linked to the original documents. Khresmoi started on the 1st of September 2010 and runs for four years. In summary, the objectives are Khresmoi are:

- 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 (MR, CT), and 4D (MR 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 is flowing into several open source components, which are integrated into an innovative open architecture for robust and scalable biomedical information search.

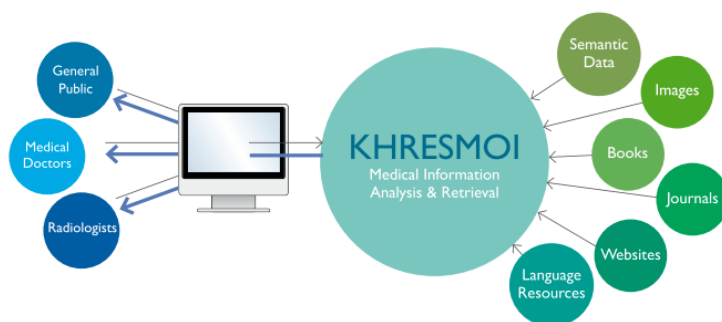


Figure 1: Khresmoi global view

1.2 Khresmoi Use Cases

Khresmoi is being evaluated in challenging use cases involving the following target user groups:

- **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 tools 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 of images are produced per day. The huge archives of radiology images available in hospitals (in anonymized 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

1.3.1 Khresmoi Prototype

The Khresmoi prototype integrates all technology developed in Khresmoi. The prototype now runs on the Khresmoi Cloud, a private cloud made up of nine servers with one Terabyte of RAM and 28 Terabytes of storage, leading to better performance of the prototype and higher reliability.

The Khresmoi prototype was clearly divided into three *faces*, with each face meeting the requirements of one of the target groups of end users. The three faces are:

- **Khresmoi for Everyone:** This face presents a straightforward search interface aimed at members of the general public. It also has features specific to the medical domain developed in Khresmoi, such as medicine-specific machine translation and automated estimated of the trustability and readability levels of documents. This face is shown in Figure 2. The red or green bar to the left of each result in the result list indicates the estimated readability level, while the scale to the right of each result presents the estimated trustability level of the website. Translation and filtering options are available on the right of the window.
- **Khresmoi Professional:** This face, shown in Figure 3, is aimed at medical professionals. The interface is more comprehensive, and allows results to be stored in a personal library, rated and shared with colleagues. Support for medicine-specific machine translation and image search based on visual similarity are also available. Various facets classifying the results are shown on the left of the window.
- **Khresmoi Radiology:** This face, shown in Figure 4, makes available the advanced visual search capabilities required by radiologists. It allows search by visual similarity through 3D images (CT, MRI, ...) stored in a hospital PACS, as well as through 2D images in the medical literature. A region of an image can be chosen (on the left in Figure 4), and the system will present the most similar images from the PACS (on the right in Figure 4). Search results and associated radiology reports can be viewed. Analyses of the texts in the radiology reports accompanying the search results allow the most commonly mentioned pathologies in the radiology reports to be identified.



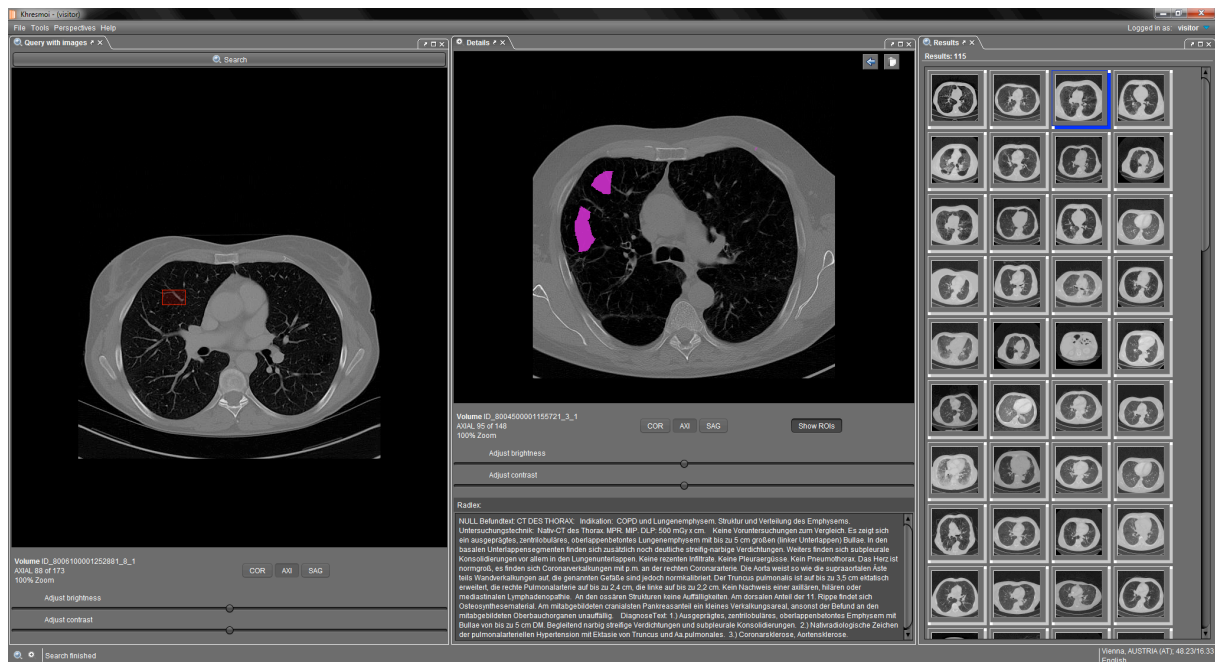


Figure 4: Khresmoi Radiology

1.3.2 Khresmoi Technology

The software that Khresmoi is built upon has undergone significant advancement through work in Khresmoi. The software is listed below, along with the advances achieved in Khresmoi:

- **GATE** (<https://gate.ac.uk/>): The General Architecture for Text Engineering (GATE) is used to annotate at word, section and document levels. Through work in Khresmoi, its capabilities for annotating medical documents have been expanded. The use of cycles of human correction to improve the automatic annotation has also been extensively tested.
- **Mimir** (<https://gate.ac.uk/mimir/>) uses GATE annotations to perform semantic search. The Khresmoi Mimir Interface (KMI) has been developed to allow more user friendly querying of Mimir from Khresmoi. A semantic typeahead service and corresponding interface has also been developed to allow straightforward semantic querying.
- **ezDL** (<http://ezdl.de/>) is a framework for interactive search applications. New features have been added, including drop down options for query specification, and automatic translation of non-English query terms if too few results are returned. It has also been made more stable and efficient. Two front-ends are now available for ezDL: the original Java Swing interface and a new web interface.
- **ParaDISE** is a new visual search engine developed in Khresmoi as a successor to the GNU Image Finding Tool (GIFT). It is more scalable than GIFT and contains state-of-the-art image features and visual similarity calculation.
- The **MOSES** statistical machine translation software (<http://www.statmt.org/moses/>) has been further adapted to machine translation in the medical domain by extensive training on domain-specific texts in English, German, French and Czech. For texts in the medical domain, experiments have shown that the translation results are now better than those obtained using Google or Bing translation services.
- The **OWLIM** semantic repository (<http://www.ontotext.com/owlim>) has received performance and functionality upgrades, and has also had its medical knowledge base expanded through the addition of new medical vocabularies and new links between the medical vocabularies.

1.3.3 Evaluation

The three faces of Khresmoi were evaluated in user-centered evaluations for all three end user groups targeted by Khresmoi. Specifically, they were performed by 28 members of the general public, 19 physicians and 17 radiologists. The evaluations for all three user groups involved the users carrying out well-defined search tasks on the appropriate prototype, while a large amount of data was gathered about their interaction with the system, including search logs, sound and video recording and mouse interactions. A detailed analysis of the user-centered evaluation data was done, and results from the evaluations guided further development of the prototypes.

A meta-analysis of all evaluation results at component level and user level was carried out, allowing an objective image of the progress of the Khresmoi project to be created. Khresmoi also organized the search task in the CLEF eHealth evaluation campaign in 2013. Nurses created questions that were likely to be asked by patients after looking at their hospital discharge letter, and participants in the evaluation campaign tried various approaches to finding relevant information to answer these questions from the collection of Khresmoi websites.

1.4 Expected Final Results and their Potential Impact and Use

Medical Impact: With *Khresmoi Professional*, Khresmoi aims to 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. Furthermore, *Khresmoi Radiology* aims to convert the flood of radiological image data into a boon instead of a curse. This will be done by allowing the data to be searched using visual similarity criteria, hence providing radiologists with a tool to assist them in diagnoses.

Scientific Impact: Khresmoi will release large-scale data sets and realistic task-based scenarios on which to assess new technologies in the medical domain, addressing the current lack of these resources. It will also make available cutting edge techniques implemented in open source software.

Industrial Impact: Khresmoi is improving 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 *Khresmoi for Everyone* search engine through the Health on the Net website, leading to better understanding of the trustability of online health information.

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.