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

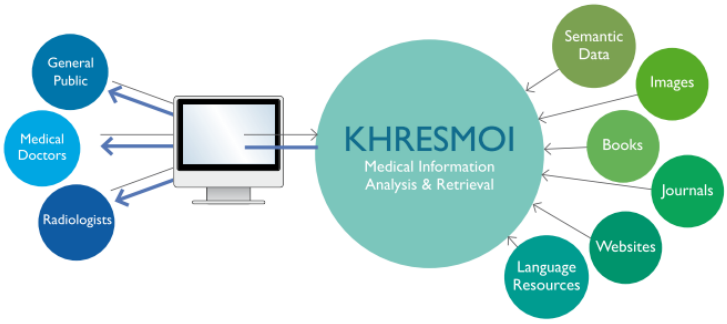

Figure 1: Khresmoi global view

1.2 Khresmoi Use Cases

Khresmoi will be 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 Technology

The open source software that Khresmoi is built upon has undergone significant advancement through work in Khresmoi. The open source 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 major achievement was the release of Mimir 4, including the ability to rank returned documents.
- **EzDL** (http://ezdl.de/) is a framework for interactive search applications. It has been extended with the capability to display image search results, as well as extensive tools to facilitate collaborative search, such as the ability to share documents and queries between users.
- **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.

Furthermore, the **MOSES** statistical machine translation software (http://www.statmt.org/moses/) has been adapted to machine translation in the medical domain. Finally, 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.

1.3.2 End User Requirements

Three surveys to elicit end user requirements of the Khresmoi target user groups have been done. The surveys for medical practitioners and the general public were carried out through online questionnaires, while the survey for 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. Furthermore, analysis of search logs from the Health on the Net search engine and the Goldminer search engines allowed the types of queries used by the target user groups to be determined.

The results of these surveys were used to specify the features that are to be included in the Khresmoi prototypes. An unexpected result of the surveys was the wish by medical professionals to be able to

perform collaborative search, through rating search results and sharing the search results. This led to a stronger concentration on collaborative search in Khresmoi.

1.3.3 Khresmoi Prototypes

Two integrated prototypes combining Khresmoi components have been created. These provide search of both text and image information. The following image shows the comprehensive search interface to the Khresmoi search system:

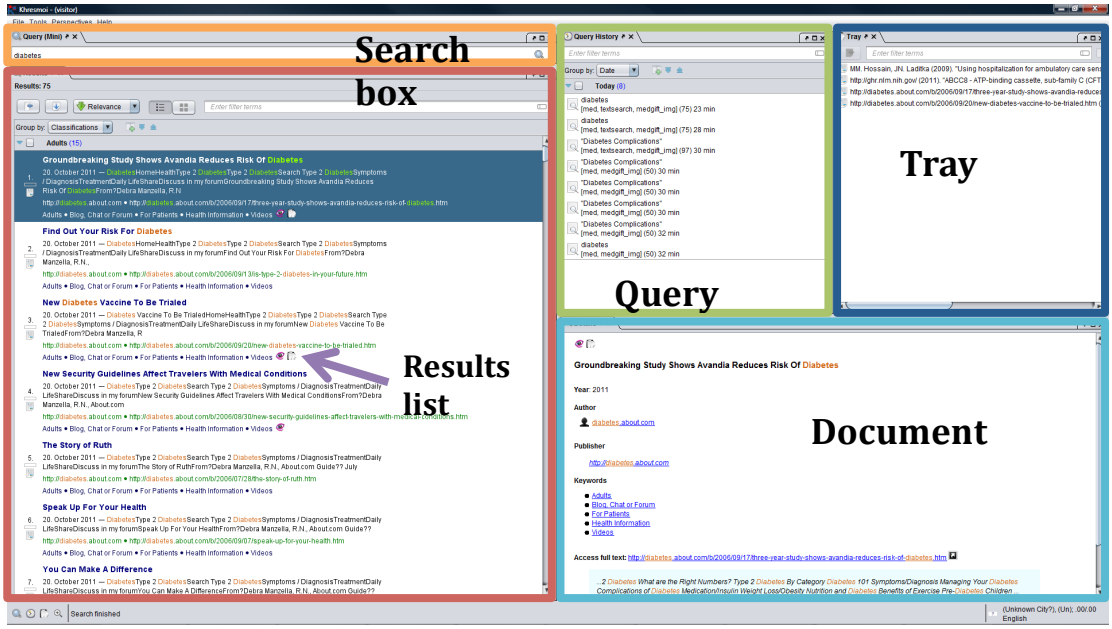

Figure 2: Search interface of the Khresmoi search system

This interface has the following features:

- **Search box:** The simple search box is shown. Advanced search is also available.
- **Result list:** A summary of each document is given. Controls at the top allow the results to be reordered or grouped by various criteria and filtered by terms. The icons indicated by the arrow show when a document has been viewed (eye icon) and moved to the tray (clipboard icon).
- **Tray:** Documents can be dropped here to be stored for future use. Logged in users have access to personal libraries, allowing more flexibility in organisation, including the capability to add tags to results and share results with other users.
- **Query history:** Lists details on all queries entered (including date and time) and allows queries to be repeated.
- **Document view:** All data available on the document selected in the results list are shown, including a link to the full document.

The 3D image retrieval prototype is shown in the following image:

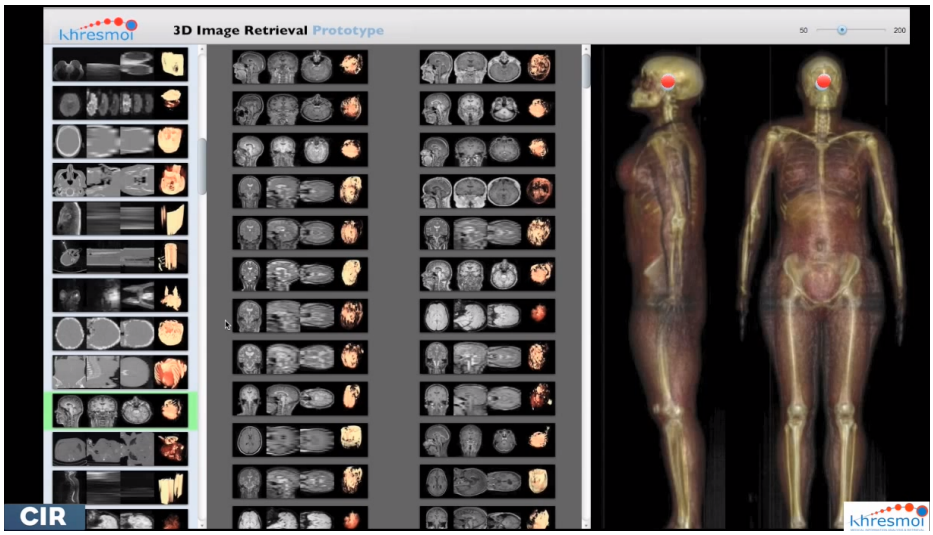

Figure 3: 3D image retrieval prototype

Various query volumes (3D images) are shown in the left pane of the interface. No information on the anatomical structures or the position in the body is known for these volumes. Clicking on a query volume shows the most similar volumes from the database in the middle pane. The right pane shows a position estimate in the human body suggested by the system for the query volume.

1.3.4 Evaluation Strategy

It is important to know how well the technology developed in Khresmoi is functioning. For this reason, an evaluation strategy has been developed in Khresmoi. Each of the components of Khresmoi has already undergone individual evaluation and the results have been published. The next step is to perform global evaluations, which will take two forms:

- The **global empirical evaluation** will measure how the components work together in the Khresmoi system. For the pipeline of components used in performing a query, the effect of the output of a component on the performance of the subsequent component in the pipeline will be examined.
- The **user-centered evaluation** will get end users from all three Khresmoi target user groups to perform tasks on the Khresmoi prototype. The effectiveness with which these tasks can be performed will be measured, and user feedback on the system will be gathered. The tasks that these users will be asked to perform have been created, based on the end user surveys and search log analyses.

1.4 Expected Final Results and their Potential Impact and Use

Medical Impact: 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 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 diagnosis.

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 Health on the Net search engine, improved by the Khresmoi technology.

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.