Khresmoi Professional: Multilingual Semantic Search for Medical Professionals

Niraj Aswani[i], Thomas Beckers[g], Erich Birngruber[k], Célia Boyer[j], Andreas Burner[k], Jakub Bystroň[h], Khalid Choukri[d], Sarah Cruchet[j], Hamish Cunningham[i], Jan Dědek[h], Ljiljana Dolamic[j], René Donner[k], Ondřej Dušek[h], Sebastian Dungs[g], Ivan Eggel[b], Antonio Foncubierta[b], Norbert Fuhr[g], Adam Funk[i], Alba Garcia Seco de Herrera[b], Arnaud Gaudinat[b], Georgi Georgiev[e], Julien Gobeill[b], Lorraine Goeuriot[f], Paz Gomez[c], Mark A. Greenwood[i], Manfred Gschwandtner[l], Allan Hanbury[a], Jan Hajič[h], Jaroslava Hlaváčová[h], Markus Holzer[k], Gareth Jones[f], Blanca Jordan[c], Matthias Jordan[g], Klemens Kaderk[k], Franz Kainberger[k], Liadh Kelly[f], Sascha Kriewel[g], Marlene Kritz[l], Georg Langs[k], Nolan Lawson[j], Johannes Leveling[f], David Mareček[h], Dimitrios Markonis[b], Iván Martinez[c], Vassil Momtchev[e], Alexandre Masselot[j], Hélène Mazo[d], Henning Müller[b], Michal Novák[h], Johann Petrak[i], João Palotti[a], Pavel Pecina[h], Konstantin Pentchev[e], Deyan Peychev[e], Natalia Pletneva[j], Martin Popel[h], Diana Pottecher[c], Angus Roberts[i], Rudolf Rosa[h], Patrick Ruch[b], Alexander Sachs[l], Matthias Samwald[a], Priscille Schneller[d], Veronika Stefanov[a], Aleš Tamchyna[h], Miguel Angel Tinte[c], Zdeňka Urešová[h], Alejandro Vargas[j], Dina Vishnyakova[b]

[a] Vienna University of Technology, Austria
[b] University of Applied Sciences Western Switzerland
[c] ATOS, Spain
[d] Evaluations and Language Resources Distribution Agency, France
[e] Ontotext, Bulgaria
[f] Dublin City University, Ireland
[g] University of Duisburg-Essen, Germany
[h] Charles University in Prague, Czech Republic
[i] The University of Sheffield, United Kingdom
[j] Health on the Net Foundation, Switzerland
[k] Medical University of Vienna, Austria
[l] Association of Physicians in Vienna, Austria

ABSTRACT

There is increasing interest in and need for innovative solutions to medical search. In this paper we present the EU-funded Khresmoi medical search and access system, currently in year 3 of 4 of development across 12 partners. The Khresmoi system uses a component-based architecture housed in the cloud to allow for the development of several innovative applications to support target users' medical information needs. The Khresmoi search systems based on this architecture have been designed to support the multilingual and multimodal information needs of three target groups: the general public, general practitioners and consultant radiologists. In this paper we focus on the presentation of the systems to support the latter two groups using semantic, multilingual text and image-based (including 2D and 3D radiology images) search.

Categories and Subject Descriptors

J.3 [LIFE AND MEDICAL SCIENCES]: Medical information systems

Keywords

Multilingual, multimodal, medical search system

HSD 2013, August 1, 2013, Dublin, Ireland.

Figure 1 - The Khresmoi Concept

1. INTRODUCTION

The Khresmoi project[1] is developing a multilingual multimodal search and access system for medical and health information and documents. It addresses the challenges of searching through huge amounts of medical data, including general medical information available from various online sources via the internet, as well as 2D and 3D radiology images in hospital archives. The system

[1] http://khresmoi.eu/

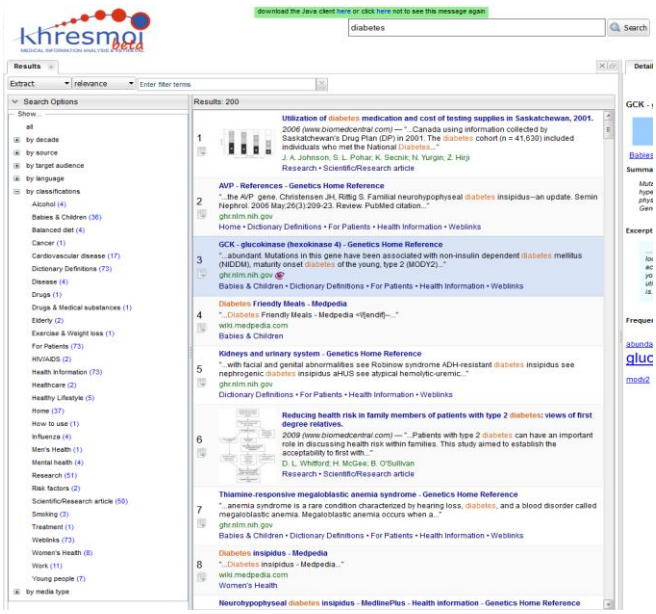

Figure 2: The web frontend

allows text querying in several languages with query translation for cross language searching, in combination with image queries. Extensive medical knowledge bases support semantic search. Results can be translated using a machine translation tool specifically trained on medical text.

The system is aimed at three main end user groups: members of the general public, general practitioners and consultant radiologists (a group for whom image search is crucial). An outline of the Khresmoi concept is shown in Figure 1. In this paper we focus on innovative search functionality for general practitioners and consultant radiologists search.

2. KHRESMOI TECHNOLOGY

The Khresmoi system has been developed using a Service Component Architecture (SCA) supported by a cloud infrastructure [7]. Components in this system include a search component, a knowledge-base component, a machine translation component, a query disambiguation component, a spell-checking component, etc. An overview of how these components are combined in the architecture is shown in Figure 3. The sources of the documents indexed are shown on the left of the figure (in purple). Text is crawled from websites aimed at the general public and physicians, while images and text are extracted from open access medical journals. 3D images (MR/CT) and associated radiology report text are exported from a Picture Archiving and Communication System (PACS) and indexed to be accessed by radiologists working within the hospital in which this data is stored.

The components that process and index the data are shown in blue. Key components in this architecture are built upon open source software, developed by project partners, which has been significantly advanced by work in Khresmoi to meet the retrieval requirements of medical search systems.

GATE[2]: The General Architecture for Text Engineering (GATE) is used to annotate documents at word, section and document levels. Through work in Khresmoi, its capabilities for annotating medical documents have been expanded. The use of cycles of manual correction of the annotations to allow the automatic annotation software to improve the annotation results by learning to correct its errors has also been extensively tested.

Mimir[3] uses GATE annotations to perform semantic search. The latest version of Mimir (Mimir 4) includes the ability to rank returned documents.

ezDL[4] 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 due to the use of Hadoop/MapReduce, and contains state-of-the-art image features and visual similarity calculation.

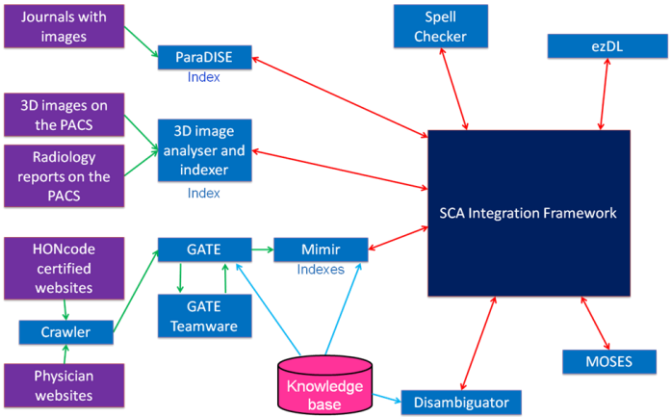

Figure 3: Overview of the Khresmoi architecture

Furthermore, the **MOSES** statistical machine translation software[5] has been trained on extensive collections of medical documents to obtain domain-adapted statistical machine translation in the medical domain. The **OWLIM** semantic repository[6] 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.

Finally, technology for analysing 3D CT and MR images is being developed. This allows structures in the images to be automatically identified and mapped to a standard vocabulary. It

[2] https://gate.ac.uk/

[3] https://gate.ac.uk/mimir/

[4] http://ezdl.de/

[5] http://www.statmt.org/moses/

[6] http://www.ontotext.com/owlim

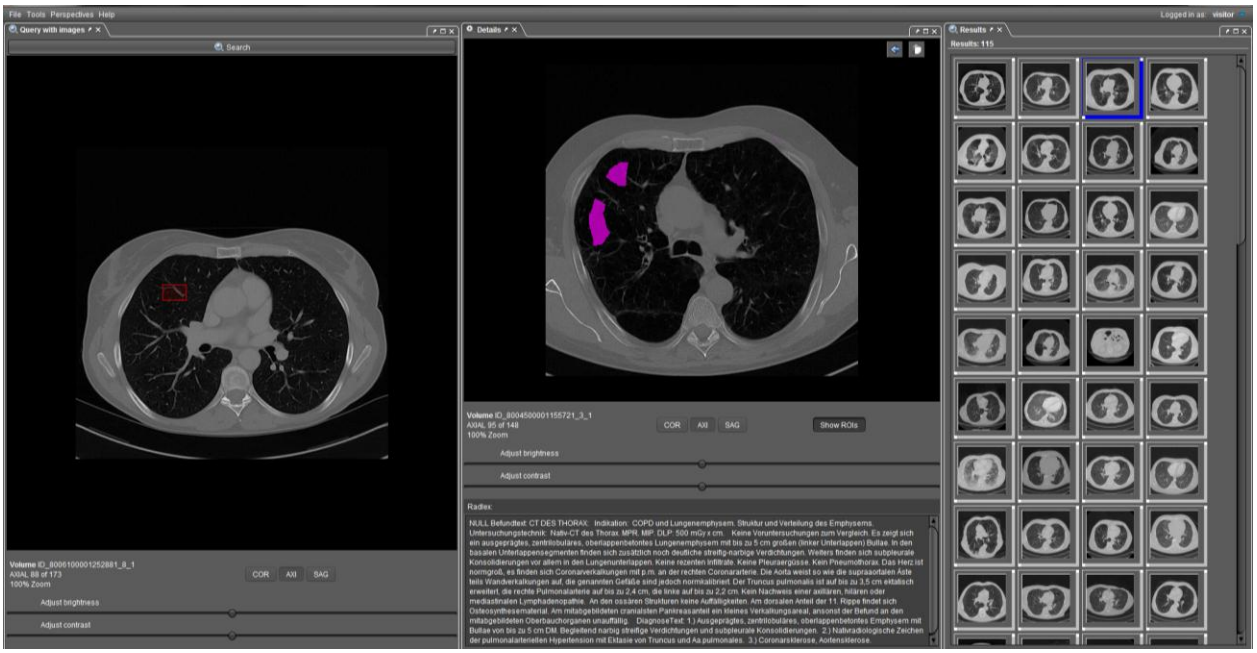

Figure 4 - Khresmoi interface for radiologists

also allows retrieval of images to be done based on the visual similarity between images.

The modular integration of multiple software technologies in the system architecture allows for easy development of the required medical search applications, such as the general practitioner and radiologist applications described next.

3. MEDICAL SEARCH FOR GENERAL PRACTITIONERS

The Khresmoi search prototype for general practitioners combines technologies developed within the Khresmoi project into an integrated platform. It currently provides two user interfaces. One is a browser based web application[7] written in GWT while the other is a Java Swing desktop application[8]. Both share a common backend service infrastructure also written in Java. A third user interface for Android devices is currently under development.

Crawled websites with trustworthy medical information targeted at practitioners are semantically annotated using GATE technology, and then indexed by Mimir, described in previous section. In addition, images from medical publications are also indexed (using ParaDISE, described in previous section).

The web frontend (see Figure 2) features basic functionality including running text-based searches, filtering and sorting of result sets. It also includes a facet explorer, which provides a means to quickly filter the result set using the available metadata attributes. The system offers spelling corrections and disambiguation suggestions while a user is typing a query. The system marks documents that the user has viewed with an eye icon (visible in the highlighted result in Figure 2) to allow the user to more easily keep track of the progress of the search process. Result sets may include images which can be used to trigger searches for visually similar images.

Users can store retrieved documents in a tray for later inspection. The personal library is a permanent storage for documents of various formats and is available to all registered users. Queries are recorded and can easily be reissued by utilizing a separate view in the interface.

The interface consists of several components each containing an aspect of the system's functionality, which are arranged in predefined layouts suitable for the most common user tasks. All components can be (un-)hidden from the perspective, re-sized and moved in the interface.

The Swing interface includes all features of the web prototype. In addition, the desktop client has collaborative features, which registered users can use to share documents with other users or user groups. For scientific work import and export of document Bibtex records is supported.

Both interfaces are fully internationalised for all Khresmoi project languages, including English, German, French, Spanish and Czech, as well as for Chinese and Vietnamese. The system also offers translation suggestions for search terms – for example, a German-speaking physician wishing to search in the English literature can type in the search terms in German and selected the proposed translations into English. The user may also machine translate the summaries of documents back into their selected language.

[7] http://khresmoi.is.inf.uni-due.de:8182/
[8] http://khresmoi.is.inf.uni-due.de/khresmoi.jnlp

The prototype for radiologists was created based on the Swing version of the interface. It shares the same technological basis and is described in the next section.

4. MEDICAL SEARCH FOR RADIOLOGISTS

Similar to the Khresmoi search for general practitioners, the Khresmoi search system for radiologists combines technologies developed within the Khresmoi project into an integrated platform. This system is for use by radiologists in medical institutions, allowing for the search and comparison of 2D and 3D radiology images. Given the sensitive nature of the medical imagining data of patients, the system is not publicly available. However, a demo can be viewed[9].

Figure 4 presents the interface instantiated for use by radiologists (also noticeable by the colour scheme adapted to the radiology requirements). Note that this is the same interface framework shown in Figure 2, but with different tools visible. Here the query is the selected area of the image slice shown in the left panel. The images in the panel on the right are returned based on their visual similarity to the region marked in the query. In the central panel, the selected image is shown with the region corresponding best to the query region highlighted. The associated radiology report is shown below the central image. For this application, only the images stored in the archives of the hospital in which the system is used are indexed. However, the possibility to do a visual search of 2D images from the medical literature is also provided.

A use case for this system is that a radiologist faced with an unusual or unknown structure in an image can query the hospital archives for images containing a similar structure, and use the (anonymised) radiology reports associated with these images to guide the reading of the image.

5. CONCLUDING REMARKS

Medical search systems are required by different classes of individuals – from members of the general public with differing levels of medical knowledge and a range of search and language skills, to numerous classes and types of medical professionals. The Khresmoi project satisfies this need through the development of different instantiations of the multilingual multimodal medical search systems for different classes of users. In this paper we present the Khresmoi system, with a particular focus on *Khresmoi Professional* which currently provides multilingual semantic, text and image based medical search applications for two classes of users: general practitioners and consultant radiologists.

These systems were developed in a holistic way, taking into consideration users' needs and requirements as determined by extensive questionnaires and analyses conducted within the Khresmoi project [3, 5]. Rounds of user-centered evaluation at both the interface component and interface system level have been, and continue to be, used for iterative system refinement [1, 2, 4]. These evaluations use general practitioners and radiologists to perform realistic tasks, to enable development of systems which function for the target users in the most useful ways.

In addition, the backend system components are empirically evaluated using document and image collections and generated search test collections from the Khresmoi project, which represent real users' information needs and querying behaviours [8]. As part of this analysis a novel global empirical evaluation is being conducted to measure the impact of components of the system on each other, and importantly how the performance of these components in isolation and combination impact on the information displayed to users and on the end user experience [6].

[9] http://youtu.be/UnPs7NSet1g

6. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 257528 (KHRESMOI).

7. REFERENCES

[1] Baroz F, Boyer C, Gschwandtner M, Goeuriot L, Hajic J, Hanbury A, Kritz M, Palotti J, Pletneva N, Ruiz de Castañeda R, Sachs A, Samwald M, Schneller P, Stefanov V, Uresova Z, Report on user tests with initial search system, Khresmoi Deliverable D10.1, khresmoi.eu/assets/Deliverables/WP10/KhresmoiD101.pdf (last accessed: May 2013)

[2] Beckers T, Dungs S, Fuhr N, Goeuriot N, Ignalski J, Jordan M, Kelly M, Kriewel S, Report on results of the WP3 first evaluation phase, Khresmoi Deliverable D3.2, khresmoi.eu/assets/Deliverables/WP3/KhresmoiD32.pdf (last accessed: May 2013)

[3] M. Kritz, M. Gschwandtner, V. Stefanov, A. Hanbury, M. Samwald, Utilization and Perceived Problems of Online Medical Resources and Search Tools among Different Groups of European Physicians, Journal of Medical Internet Research, 15(6):e122, 2013.

[4] Markonis D, Holzer M, Baroz F, Luis Ruiz De Castaneda R, Langs G, Boyer C, Mueller H, Report on the results of the initial user test of the radiology search system, Khresmoi Deliverable D10.2, khresmoi.eu/assets/Deliverables/WP10/KhresmoiD102.pdf (last accessed: May 2013)

[5] Müller H. Report on image use behavior and requirements. Khresmoi Deliverable D9.1, khresmoi.eu/assets/Deliverables/WP9/KhresmoiD91.pdf (last accessed: May 2013)

[6] Meta-analysis of the first phase of empirical and user-centered evaluations. Khresmoi Deliverable D7.2, August 2013 (to appear)

[7] Prototype and evaluation of the "Full Cloud Infrastructure" Khresmoi Deliverable D6.4.3, August 2013 (to appear)

[8] L. Goeuriot, L. Kelly, G. J. F. Jones, G. Zuccon, H. Suominen, A. Hanbury, H. Müller, J. Leveling, Creation of a New Evaluation Benchmark for Information Retrieval Targeting Patient Information Needs, Proc. 5th International Workshop on Evaluating Information Access (EVIA), 2013.