IGI PUBLISHING **ITJ 5280**
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

This paper appears in the publication, Journal of Cases on Information Technology, Volume 11, Issue 4
edited by **Mehdi Khosrow-Pour © 2009, IGI Global**

Pillars of Ontology Treatment in the Medical Domain

Daniel Sonntag, DFKI - German Research Center for Artificial Intelligence, Germany

Pinar Wennerberg, Externer Dienstleister der Siemens AG, Germany

Paul Buitelaar, DERI - National University of Ireland, Galway

Sonja Zillner, Siemens AG, Germany

EXCUTIVE SUMMARY

In this chapter the authors describe the three pillars of ontology treatment in the medical domain in a comprehensive case study within the large-scale THESEUS MEDICO project. MEDICO addresses the need for advanced semantic technologies in medical image and patient data search. The objective is to enable a seamless integration of medical images and different user applications by providing direct access to image semantics. Semantic image retrieval should provide the basis for the help in clinical decision support and computer aided diagnosis. During the course of lymphoma diagnosis and continual treatment, image data is produced several times using different image modalities. After semantic annotation, the images need to be integrated with medical (textual) data repositories and ontologies. They build upon the three pillars of knowledge engineering, ontology mediation and alignment, and ontology population and learning to achieve the objectives of the MEDICO project.

Keywords: *Data Integration, Health Care Infrastructure Knowledge Base, Information Engineering, Medical Imaging Systems, Multimedia Database Clinical Information System, Ontologies, Semantic Data Model, Semantic Matching, User/Machine Dialog*

INTRODUCTION

Clinical care and research increasingly rely on digitized patient information. There is a growing need to store and organize all patient data, such as health records, laboratory reports and medical images, so that they can be retrieved effectively. At the same time it is crucial that clinicians have access to a coherent view of these data within their particular diagnosis or treatment context.

With traditional applications, users may browse or explore visualized patient data, but little to no help is given when it comes to the interpretation of what is being displayed. This is due to the fact that the semantics of the data is not explicitly stated, which therefore remains inaccessible to the system and therefore also to the user. This can be overcome by the incorporation

DOI: 10.4018/jcit.2009072103

of external medical knowledge from ontologies which provide the meaning (i.e., the formal semantics) of the data at hand.

Our research activities are in the context of the THESEUS MEDICO project. MEDICO addresses the need for advanced semantic technologies in medical image and patient data search. The objective is to enable a seamless integration of medical images and different user applications by providing a direct access to image semantics. A wide range of different imaging technologies in various modalities exist, such as 4D 64-slice Computer Tomography (CT), whole-body Magnet Resonance Imaging (MRI), 4D Ultrasound, and the fusion of Positron Emission Tomography and CT (PET/CT). All these image modalities have the common property that their semantic contents include knowledge about human anatomy, radiology, or diseases.

One important requirement for advanced applications in semantic image retrieval, clinical decision support and computer aided diagnosis is the comparative exploration of similar patient information. For this purpose, we envision a flexible and generic image understanding software for which semantics of the images plays the major role for access and retrieval. However, currently, large amounts of medical image data are indexed by simple keywords to be stored in distributed databases without capturing any semantics.

The objective of MEDICO is to build the next generation of intelligent, scalable and robust search engines for the medical imaging domain, based on semantic technologies. With the incorporation of higher level knowledge represented in ontologies, different semantic views of the same medical images (such as structural aspects, functional aspects, and disease aspects) can be explicitly stated and integrated. Thus, the combination of formal semantics with image understanding helps building bridges between different but related domains that can be used for comparative exploration of patient data. MEDICO is a consortium research project funded by the German Federal Ministry of Economics with several R&D sites and the Erlangen University Hospital as a clinical partner. Visit http://theseus-programm.de/scenarios/en/medico.

Within the MEDICO project, one of the selected scenarios aims for improved image search in the context of patients that suffer from lymphoma in the neck area. Lymphoma, which is a type of cancer affecting the lymphocytes, is a systematic disease with manifestations in multiple organs. During the course of lymphoma diagnosis and continual treatment, image data is produced several times using different modalities. As a result, the image data consist of many medical images in different formats, which additionally need to be associated with the corresponding patient data. Hence, the lymphoma scenario is particularly suitable to demonstrate the strength of a semantic search engine as we envisioned in MEDICO.

To address the challenges of advanced medical image search, different medical resources need to be semantically integrated. Consequently, the following four research questions arise:

1. How is the workflow of the clinician, i.e.,
   a. What kind of information is relevant for his daily tasks?
   b. At what stage of the workflow should selected information items be offered?
2. What are the particular challenges and requirements of knowledge engineering in the medical domain?
   a. Can those challenges be addressed by a semi-automatic knowledge extraction process based on clinical user interactions?
   b. Can we embed the semi-automatic extraction process into the clinician's workflow?
3. How can different possibly overlapping data sources (i.e., ontologies) be aligned?
4. How can we learn and populate ontologies?