

Prototyping Semantic Dialogue Systems for Radiologists

Daniel Sonntag

DFKI–German Research Center for AI
Intelligent User Interfaces (IUI)
Stuhlsatzenhausweg 3, 66123 Saarbrücken
sonntag@dfki.de

Manuel Möller

DFKI–German Research Center for AI
Knowledge Management (KM)
Trippstadter Straße 122, 67663 Kaiserslautern
manuel.moeller@dfki.de

Abstract—In the future, speech-based semantic image retrieval and annotation of medical images should provide the basis for clinical decision support and help in computer aided diagnosis. In this paper, we describe today’s clinical workflow and interaction requirements and present a semantic dialogue system installation for radiologists. Our research focus is on the interaction design in combination with the implementation of our prototype system for patient image search and image annotation while using a speech-based dialogue shell and a big touchscreen in the radiology environment. Ontology modeling provides the backbone for knowledge representation in the dialogue shell and the specific medical application domain.

Keywords—user/machine dialogue; multimodality; semantic data model; clinical information system

I. INTRODUCTION

Clinical care and research increasingly rely on digitized patient information. There is a growing need to store and organize all patient data, including health records, laboratory reports, and medical images. Effective retrieval of images builds on the semantic annotation of image contents. 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. This means that with traditional user interfaces, users may browse or explore visualized patient data, but little or no help is given when it comes to the interpretation of what is being displayed. Semantic annotations should provide the necessary image information and a semantic dialogue shell should be used to ask questions about the image annotations while engaging the clinician in a natural speech dialogue simultaneously.

In this paper, we will provide an outline of the design phase for our prototypical semantic dialogue system for radiologists, including the discussion of clinical requirements, followed by an overview of our implementations of these requirements. We build upon the developments and implementations of the first phase (2008-2009) to achieve the objectives of the medical environment’s integration project. We will focus on the challenges, requirements, and possible solutions related to new multimodal interaction metaphors where the information access based on natural speech plays the major role during the patient finding process.

II. NEW RADIOLOGY INTERACTION REQUIREMENTS

The main task in diagnostic radiology is to interpret medical images from various modalities like computed tomography (CT) or magnetic resonance (MR) imaging. Modern radiology information systems automatically route images to the assigned radiologist immediately after the acquisition of the images. Since a single examination can result in hundreds or even thousands of images, the images are organized into series according to the Digital Imaging and Communications in Medicine (DICOM) standard. DICOM is the current standardized format used for storing all medical images.

A series, for example, contains individual 2D images (“slices”), acquired during one run of a medical imaging device, and these images make up a 3D volume of some body part. Figure 1 shows a screenshot of a desktop-based examination tool.

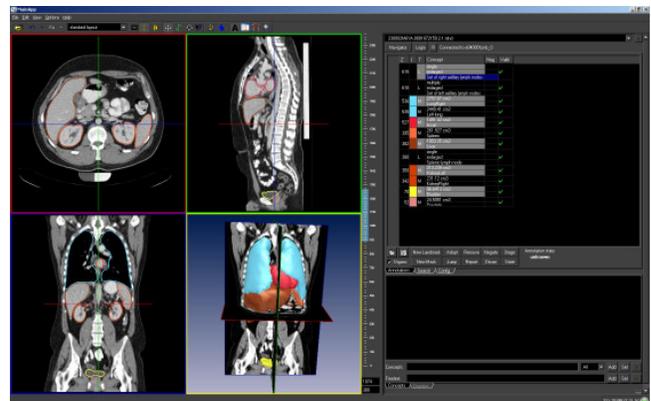


Figure 1. Traditional desktop-based 2D/3D image examination tool. The displayed images show 2D and 3D volumes of a patient’s chest and abdomen.

Typically, one imaging examination, referred to as a “study” in DICOM, consists of multiple series that are acquired using different machine settings, before or after administration of some contrast media. The series may also contain images from a variety of post-processing options (e.g., to enhance soft tissue contrast).

The process of *reading* the images is highly efficient when using a traditional desktop-based 2D/3D examination tool. While the radiologist views the images in each series essentially in sequential order, he uses a special mouse or keyboard to navigate and manipulate the images (e.g., to zoom, to change display settings, or to perform measurements) while he dictates the image findings that make up his report.

The problem is that he cannot directly create a structured report while scanning the images. In this *eyes-busy* setting, he can only dictate the finding to a tape-recorder. After the reading process, he can replay the dictation to manually fill out a patient's finding form. Another possibility is to have a clinical assistant complete the form. But since the radiologist has to check the form again, this task delegation does not save much time which is spent on one report. In addition, the form has to be manually transferred into a machine-readable report, which again is very time consuming and prone to errors.

Prior studies have looked at this problem: Recently, structured reporting was introduced that allows radiologists to use predefined standardized forms for a limited but growing number of specific examinations. Structured reporting enables the capture of radiology report information so it can later be retrieved and reused, i.e., there are sections for specific anatomical parts and disease annotations, and the vocabulary must be standardized. However, radiologists feel restricted by these forms and fear a decrease in focus and eye dwell time on the images [2, 11]. This means that the structured reporting must be done during the reading process and since the radiologists cannot easily put an eye off the image sequences, they simply cannot use the structured template while reading the images.

As a result, the acceptance for structured reporting is still low among radiologists. In contrast, referring physicians and hospital administration staff in general are supportive of structured standardized reporting since it eases the communication with the radiologists and can be used more easily for further processing (statistics, quality control, alerts, reminders, etc.). Therefore, we strive to overcome the limitations of *structured reporting*:

- 1) Content-based information should be automatically extracted from medical images.
- 2) In combination with dialogue-based radiology image reporting, radiologists should no longer fill out forms but focus on the images while either dictating the image annotations of the reports to the dialogue system or refining existing annotations.¹

¹ If, for example, a radiologist detects a stenosis in a coronary artery, he would simply point to the stenosis, dictate "moderate stenosis," which would be acknowledged by the dialogue system as "moderate stenosis in proximal segment of the right coronary artery." This would additionally make use of the automatic analysis capabilities of

Content-based extraction of information for medical images is currently restricted to the detection of anatomical location. Anatomical location such as liver, heart, kidneys, spleen, bladder, or prostate can be detected with high accuracy [5]. Anatomical annotations which are needed to complete a finding process (cf. section 4 on ontology models for medical applications), however, cannot be detected automatically (e.g., capsule of spleen). The same applies to the much more critical disease annotations, especially in the lymphoma context: marginal changes in tissue structure cannot be detected automatically with state-of-the-art technology. As a consequence, while automated image parsing remains incomplete, *manual image annotation* continues to be an important complement.

Our system is the only one of several other research projects [2, 11] to integrate manual image annotation in the reporting workflow of radiologists while using a speech-based dialogue system. Although some other speech-recognition systems for radiology exists (cf. [7]), they do not use multimodal interaction in a real dialogue scenario, cannot deal with ontological image annotations, and do not directly annotate image regions according to the ontological image model.

Currently, clinical system users can manually add semantic image annotations by selecting or defining anatomical landmarks or arbitrary regions/volumes of interest via speech dialogue. In this paper, we will focus on the multimodal interaction design issues and the specific multimodal dialogue-based interaction sequence that makes our prototype unique.

III. RESEARCH QUESTIONS

To address the challenges of advanced medical image search while using a dialogue shell, the following four HCI research questions arise:

- 1) What kind of ontological information is relevant for completion of his daily tasks and at what stage of the workflow should selected information items be offered or asked for? (section 4)
- 2) Which usability methods are adequate in order to support the prototype implementation stage? (section 5)
- 3) Can the challenges which concern the knowledge retrieval and acquisition process be addressed by a semi-automatic knowledge extraction process based on clinical user interactions with a dialogue system? (section 6)
- 4) How can the usability of the implemented system be evaluated? (section 7)

image contents which allow automatic detection of anatomic locations [5].

IV. ONTOLOGY MODELS FOR MEDICAL APPLICATIONS

The system architecture uses a comprehensive and multi-layered ontology. This ontology hierarchy is used to represent medical domain knowledge as well as specify the format of image annotations and patient metadata. Using the same representation formalism to represent domain knowledge and annotations allows us to formulate cross-modal and language-independent search queries. During the execution of these queries, the background knowledge from different medical ontologies such as the Foundational Model of Anatomy ontology (FMA, see [6]), RadLex [3], and the International Classification of Diseases (ICD-10)² is used to perform query expansion to retrieve images which are annotated with semantically similar concepts. Further details on the ontology hierarchies are covered in [9]. More information about the knowledge representation in the dialogue shell for the clinical reporting process can also be found in [4].

The annotated image regions should be made available to the radiologist while the patient finding process at the finding workstation. He should be able to retrieve patient information and patient image information. In addition, he should be able to refine the semantic image annotations according to the FMA, RadLex and ICD-10 models. For this purpose, it is necessary to retrieve meta data from (1) the images obtained from the CT/MR imaging center (and stored as RDF-annotations of the images), and (2) the rudimentary image region annotation obtained from the automatic image parsing methods. Then, he should be able to refine the image region annotations by clicking on a respective image region and saying, e.g., “This lymph node here, annotate with Hodgkin-Lymph (RadLex ontology term).” Figure 2 shows the radiology environment into which the dialogue system is integrated.

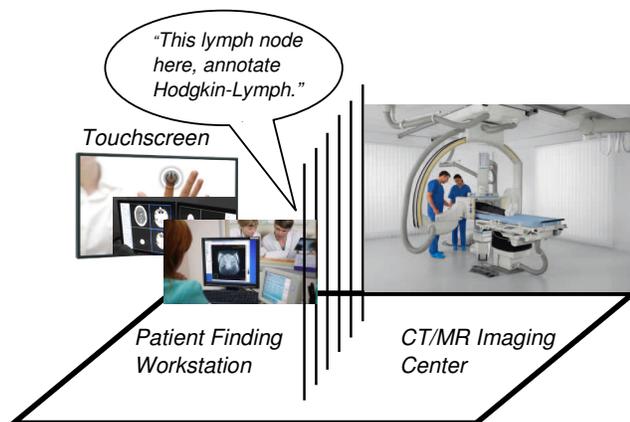


Figure 2. Radiology environment and integrated dialogue system

² See <http://www.who.int/classifications/apps/icd/icd10online>.

V. USABILITY AND INTERACTION DESIGN ISSUES

Usability applies to every aspect of a research prototype or product with which a person interacts. Every design and development decision made throughout the product cycle has an impact on that product's usability. Here we describe the usability at the prototype implementation stage.

Usability guidelines for the prototype development and implementation stage (see, e.g., [1]) consider five different planes. Every plane has its own issues that must be considered. From abstract to concrete, these are (1) the strategic plane, (2) the scope plane, (3) the structure plane, (4) the skeleton plane, and (5) the surface plane.

Defining the users and their needs on the strategic planes is the first step in the design process. It is useful to create personas that represent a special user group. On the scope plane you have to define the system's capacity (cf. *clinical reporting process*) and then the technical requirements. These two planes have already been discussed as new radiology interaction requirements (section 2), and research questions and ontology models for medical applications (sections 3, 4), respectively. In order to specify the strategic and scope plane issues, we used several usability methods: *cognitive walkthrough*, *observation of the (medical) user*, and *hierarchical task analysis*.

A cognitive walkthrough starts with a task analysis that specifies the sequence of steps or actions a user requires to accomplish a task, and the system's responses to those actions. The designers and developers of the software then walk through the steps as a group in dialogue, asking themselves a set of questions at each step. We used this to specify the example dialogue for our speech-based dialogue system.

Simply visiting the users to *observe them work* is an extremely important usability method with benefits both for task analysis and for the collection of information about the true field usability of installed systems. The observer's goal is to become virtually invisible to the users so that they will perform their work and use the system in the same way they normally do. We visited the radiology department four times in 2008. A team of five to ten radiologists are working in such a department. After each observation session, we collected their feedback according to improvements of the finding process if there were no restrictions to the employed technology. After two visits, it became clear to us that a speech-based system would best fit this environment where it is generally dark, quiet, and the users very much focus on the image sequences.

Hierarchical Task Analysis (HTA) breaks down the steps of a radiologist's task as performed by a medical user and describes the task as seen at various levels of detail. Each step can be decomposed into lower-level sub-steps, thus forming a hierarchy of sub-tasks (this corresponds to the information retrieval and annotation stages already explained). According

to the three usability methods, the structure, skeleton, and surface planes can be implemented (these correspond to the design and implementation of the concrete intelligent environment where the dialogue system/shell can be used).

The information design of the skeleton plane is represented by the ontologies we modeled in the context of the *clinical reporting process*. This means the skeleton plane is already pre-specified by the ontology models in the medical application domain. The design phase for the multimodal user interface (i.e., the dialogue shell) is restricted to the interaction design/information architecture storyboard on the structure plane and the speech and touchscreen interaction design on the surface plane. The touchscreen surface plane (figure 3) deals with the logical arrangements of the design elements. In the case of a multimodal dialogue system, the logical arrangement results in a user-system natural dialogue whereby the user input is speech and touch and the system output is generated speech or the generation of semantic interface elements (SIEs) which display windows for images, image regions, or other supported interaction elements.

The resulting clinical dialogue/workflow is best explained by example. Consider a radiologist (**R**) at his daily work of the *clinical reporting process* with the speech-based semantic dialogue shell (**S**): The potential application scenario includes a radiologist which treats a lymphoma patient; the patient visits the doctor after chemotherapy for a follow-up CT examination. The speech-based dialogue begins with the selection of a patient record (figure 3) using speech.

R: “Show me my patient records, lymphoma cases, for this week.”

S: Shows corresponding patient records. Then the dialogue progresses as shown in figure 4.

S: Shows corresponding patient image data according to referral record (2).



Vorname	Nachname	Geburtsdatum	Datum der Abt.
Gerda	Maier	12.02.1958	06.02.2008
Max	Madermann	18.08.1932	25.02.2003
Wolfgang	Schubert	27.02.1949	17.07.2002
Heinz	Maier	02.12.1908	14.12.2002
Herbert	Maier	04.02.1981	10.07.2008

Figure 3. Touchscreen surface plane

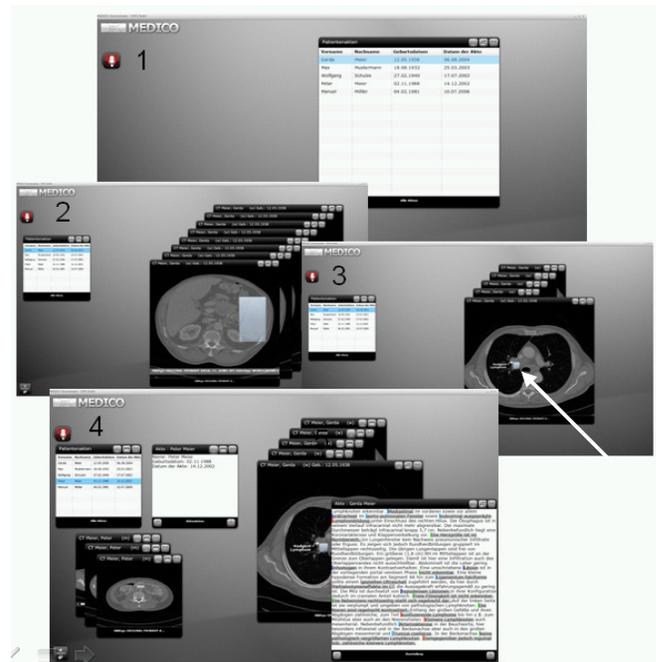


Figure 4. Speech-based multimodal dialogue with the clinician

The presentation planner of the dialogue system rearranges the semantic interface elements (SIEs). The top-most picture frame, showing the patient information in the header, is interactive; when touching it, special image regions and region annotations are highlighted (2).

R: Switches to the first image and clicks on a specific region (automatically determined) (2).

R: “Open further images, internal organs: lungs, liver, then spleen and colon of this patient (+ pointing gesture on highlighted region).”

S: The system rearranges the semantic interface elements (SIEs) to signalize that the dialogue focus is on regions. (3)

R: “This lymph node here (+ pointing gesture), annotate Hodgkin-Lymphom.” (see arrow in (3)).

S: Annotates the image with RDF annotations (cf. highlighted pathological part) and displays a label for the recognized ICD-10 term (see arrow in (3)).

The dialogue finishes after providing a complete overview of patient information and related cases on the touchscreen display (4). MEDICO displays the search results in the record table (also see screenshot (1)) ranked by the similarity and match of the medical terms that constrain the semantic search (left) and opens the first hit, Peter Maier, the record, and his images that correspond to the search. The system rearranges the SIEs for the two patients for a comparison. **R:** “Also get the findings of this patient.” The complete overview is again illustrated in figure 5.

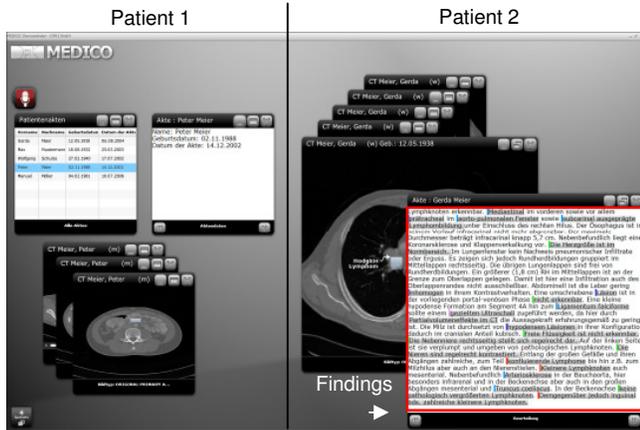


Figure 5. Complete overview for patient comparison

The most important step, however, is the annotation of the specific image region as shown again in figure 6. After saying “This lymph node here (+ pointing gesture), annotate Hodgkin-Lymphom,” the radiologist gets direct feedback of the annotation step which delivers the RadLex term annotation Hodgkin lymphoma – RID3842³ according to the ontological image model [4].



Figure 6. Speech-based annotation step of medial images

As clinical users depend more and more on automatic medical software to get their jobs done (cf. automatic organ detection) and use automatic computer systems in more critical use case scenarios (i.e., the *clinical reporting process*), usability can be the critical factor in ensuring that the intelligent system is accepted in the clinical (radiology) environment. Our approach, the described multimodal dialogue interface, provides some answers to the usability issues we identified in the beginning of this section.

³ The RadLex tree browser is available at <http://radlex.org/>, and, e.g., <http://radlex.org/RID3842> shows the Hodgkin lymphoma concept.

VI. DIALOGUE SYSTEM

Within a multimodal dialogue system two or more user input modes, such as speech, gestures, and other input modalities are proceed in a coordinated manner. The various input modalities can be combined. Our dialogue system is based on the Ontology-Based Dialogue Platform, ODP, which provides a lightweight open architecture for the flexible integration of multimodal dialogue processing components (a commercial version is available at <http://www.semvox.de/en.html>). The generic architecture of our multimodal dialogue system is illustrated in figure 7. It consists of components for the following tasks:

- Recognition of multimodal input, i.e., automatic speech recognition and image region clicks;
- The interpretation of the multimodal input including modality fusion (the click on the touchscreen must be interpreted according to the dialogue context);
- The dialogue and interaction management for the system behavior (essentially manages the question feedback);
- The semantic access to the backend application and services, including interactive semantic mediation and the access to the medical repositories;
- The presentation planning and realization (displays of medical SIEs and rearrangement for patient’s comparison);
- And the fission of the output modalities (The speech output has to be coordinated with the appearance of SIEs for answers on the graphical surface).

Input and output components can be attached to the generic system. Such components include a speech recognizer (ASR) and a speech synthesis (TTS) module. Our approach relies on a flexible toolbox of generic and configurable dialogue shell building blocks.

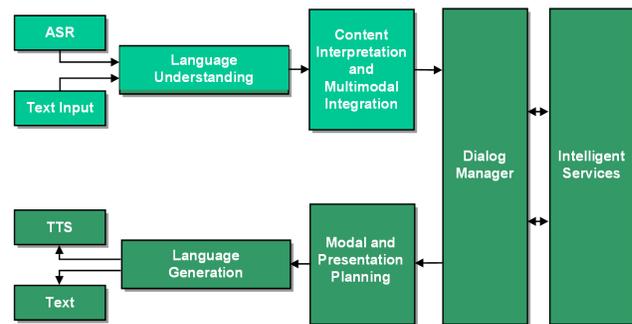


Figure 7. Architecture of our multimodal dialogue system

VII. FINDINGS FROM USABILITY STUDIES

The user studies we conducted evaluated the design of the dialogue system and its potential to speed up the patient finding process while delivering semantic annotations that can be directly used for image retrieval.

In intensive discussions with clinicians, we analyzed how the use of semantic technologies can support the clinician's daily work tasks, apart from the fact that in daily hospital work, clinicians can only manually search for *similar* images. After this initial period, we implemented our proposed solution, the semantic dialogue shell for radiologists. This pre-study involved three radiologists and eight medical experts who were responsible for the ontology models and automatic image annotations provided as input. The study reveals that all medical experts consider the image region annotation step for refined anatomy (FMA) and the disease (RadLex) as the real major knowledge acquisition bottleneck in this domain.

Accordingly, we tried to factor in the benefits of the speech-based annotation step in the contemporary clinical workflow. In the experimental setting, the prototype was used to refine the anatomy and disease annotations of 10 image series of different patients (approx. 100 annotations). This annotation step can be compared to a desktop-based semantic annotation tool where the user is presented a top-down menu to select the ontology-based annotations [4]. The speech-based annotation system worked with the disease-relevant RadLex terminology (~6000 medical terms) and the dialogue competence as illustrated in the dialogue examples. The speech recognition accuracy is approx. 95% when using a commercial ASR component. (Please note that the speech grammar not only includes the medical terms, but also the complex expressions for patient retrieval and comparison.) The dialogue-based annotation can be done at a rate of approx. 6 annotations per minute (including the visual feedback phase) whereas the desktop-based annotation comes to a rate of approx. 3 annotations per minute. In addition, the desktop-based tool cannot be used to retrieve and compare complete patient records. For this purpose, we designed the bigger touchscreen installation. Most importantly, the prototype dialogue system delivers semantic annotations which are unavailable in the current clinical finding process at the partner hospitals and the radiologist can directly detect errors visually (cf. figure 6). In further studies, we will try to assess the benefit of semantic annotations in general terms of annotation accuracy/speed when compared to the process where the text-based form, which is currently used, has to be manually transferred into a machine-readable report.

VIII. CONCLUSION

We described the design and implementation of a semantic dialogue system for radiologists. Within a multimodal dialogue system, two or more user input modes, such as speech, gestures, and other input modalities, are processed in a coordinated manner. Our new generic architecture of a multimodal dialogue system follows [8, 10]. Input and output components can be attached to the generic system. Such components include a speech recognizer (ASR) and a speech synthesis (TTS) module. Our approach relies on a flexible toolbox of generic and configurable dialogue shell building blocks. We discussed the clinical workflow and interaction requirements and focused on the design and implementation of the multimodal user interface for image search and image region annotation and its implementation while using the speech-based dialogue system.

The overall semantic search architecture (<http://www.theseus-programm.de/anwendungsszenarien/medico/>) which includes our semantic dialogue shell will now be tested in a clinical environment. This also means that we will install a touchscreen and an instance of the dialogue shell with several technical input devices (i.e., Bluetooth microphone and conference room microphone) in the radiology department. Furthermore, the question of how to integrate semantic image knowledge with other types of data, such as textual patient data, is paramount. For clinical staging and patient management, the major concern is which procedure step has to be performed next in the treatment process. The textual patient data contains the necessary background information about the treatment process.

IX. ACKNOWLEDGMENTS

This research has been supported in part by the THESEUS program which is funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016. Thanks go out to Robert Nesselrath, Yajing Zang, Günter Neumann, Matthieu Deru, Simon Bergweiler, Gerhard Sonnenberg, Norbert Reithinger, Gerd Herzog, Alassane Ndiaye, Tilman Becker, Norbert Pflieger, Alexander Pfalzgraf, Jan Schehl, Jochen Steigner, and Colette Weihrauch for the implementation and evaluation of the dialogue infrastructure. The responsibility for this publication lies with the authors.

X. REFERENCES

- [1] Garrett, J.J. The Elements of User Experience. In: American Institute of Graphic Arts, New York, (2002).
- [2] Hall, Ferris M. The Radiology Report of the Future. *Radiology* 251, 2 (2009).
- [3] Langlotz, C. P. RadLex: A new method for indexing online educational materials. *RadioGraphics* 26, (2006).
- [4] Möller M., Regel S., and Sintek M. RadSem: Semantic Annotation and Retrieval for Medical Images, ESWC (2009).
- [5] Seifert S., Kelm M., Möller. M., Mukherjee S., Cavallaro A., Huber, M., and Comaniciu D. Hierarchical parsing and semantic navigation of full body CT data. *SPIE Medical Imaging*, (2010).
- [6] Rosse C. and Mejino, J.L. The foundational model of anatomy. *Anatomy Ontologies for Bioinformatics: Principles and Practice*, Volume 6, Springer, 59-117, (2007).
- [7] Pezzullo J.A., Tung G.A., Rogg J.M., Davis L.M., Brody J.M., Mayo-Smith W.W. Voice recognition dictation: radiologist as transcriptionist, *Digit Imaging*. (2008) Dec; 21(4):384-389.
- [8] Sonntag, D. *Ontologies and Adaptivity in Dialogue for Question Answering*. AKA and IOS Press, Heidelberg, (2010).
- [9] Sonntag, D. and Möller, M. Unifying Semantic Annotation and Querying in Biomedical Images Repositories. *Proceedings of KMIS, IC3K*, (2009).
- [10] Wahlster, W. (ed.) *SmartKom: Foundations of Multimodal Dialogue Systems*. Springer, Berlin, (2006).
- [11] Weiss, D. L. and Langlotz, C.P. Structured Reporting: Patient Care Enhancement or Productivity Nightmare? *Radiology* 249, 3 (2008).