

A Multimodal Dialogue Mashup for Medical Image Semantics



Daniel Sonntag and Manuel Möller

German Research Center for Artificial Intelligence, Saarbrücken and Kaiserslautern, Germany
 daniel.sonntag@dfki.de, manuel.moeller@dfki.de

Introduction

In contemporary, daily hospital work, clinicians can only manually search for “similar” images using outdated desktop search applications. After considering the relevant categories of similarity, they subsequently apply one filter after the other.

In addition to the fact that this approach is quite time-consuming, it is neither possible to formulate complex and semantically integrated search queries in a convenient way, nor can a radiologist easily annotate images with new anatomy or disease information.

Hence, the need exists for a seamless integration of medical images and different user applications by direct access to image semantics.

Adequate (multimodal) user interfaces play a significant role in achieving this goal.

Our Approach

We define a **mashup** as a Web application that combines data and functionality from two or more sources into a single integrated application.

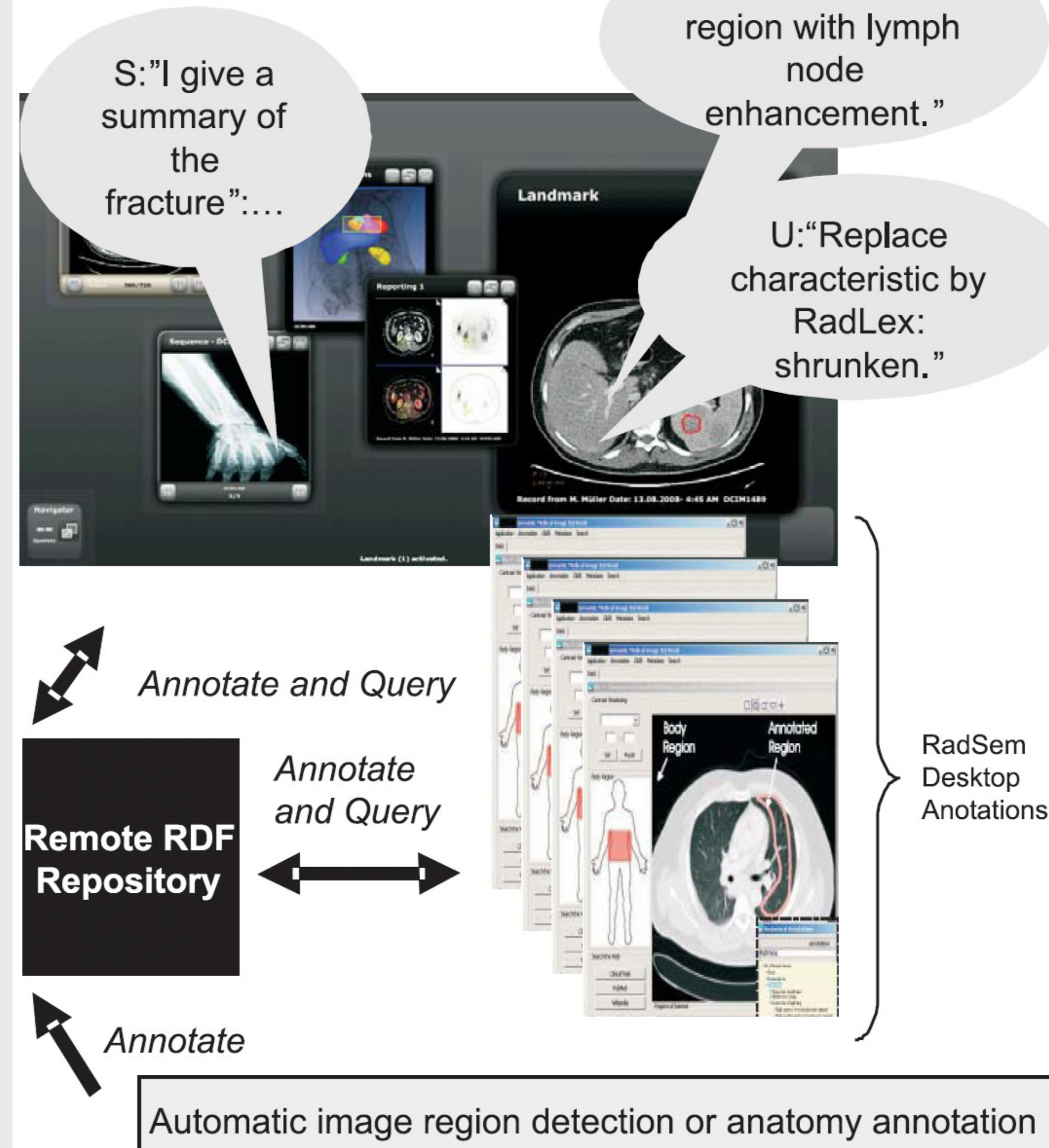
We focus on the HCI aspect of the integrated application when addressing advanced dialogical interaction with semantic (medical) image repositories. In particular, we address the knowledge acquisition bottleneck problem by concerning ourselves with the question how to mash-up

- a multimodal interface for speech-based annotations (manual and semi-automatic annotation),
- a semantic image annotation tool RadSem for annotations on a desktop computer typically performed by medical students (manual annotation),
- statistical image region annotation (automatic annotation).

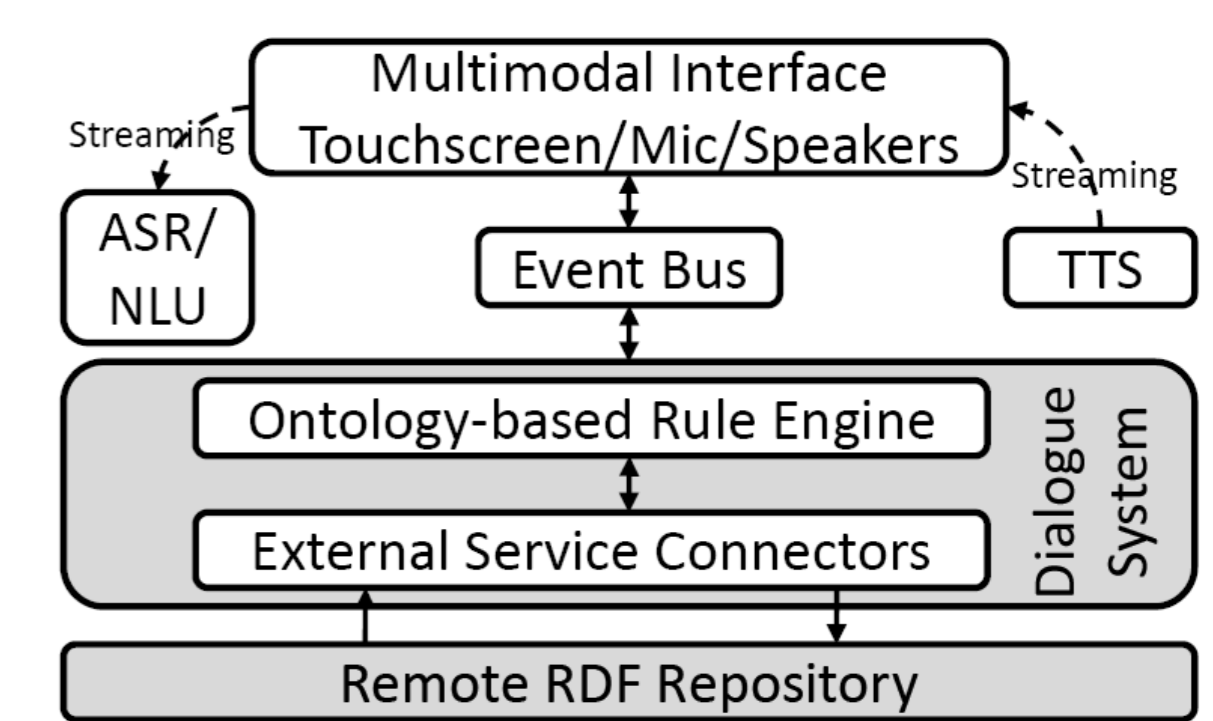
A remote RDF repository which stores the semantic image information and connects the annotation and querying task into a **common framework makes the mashup unique**.

Common Mashup Framework

Multimodal touchscreen dialogue shell Installation for the radiologist.

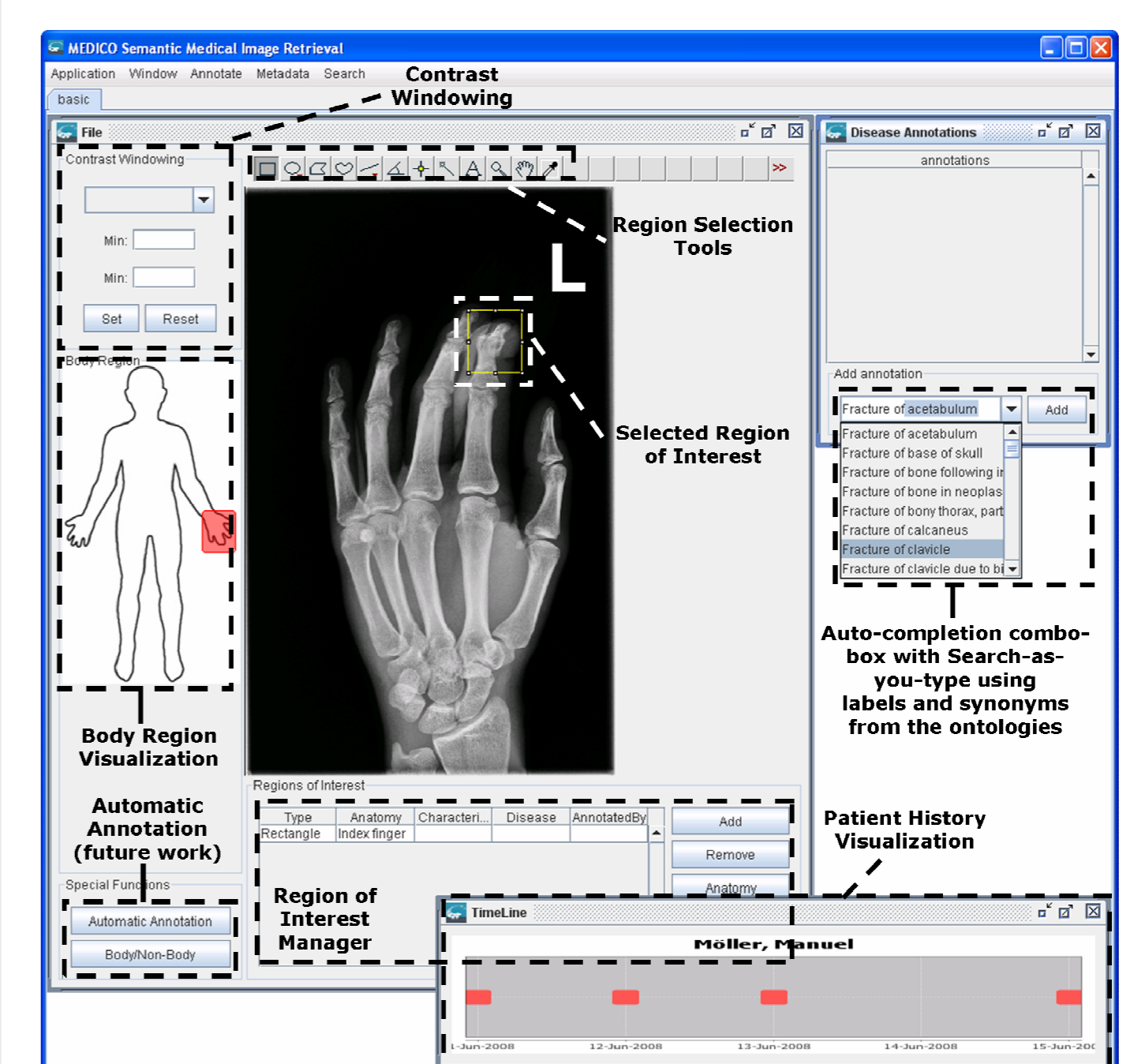


Speech Dialogue System



The generic framework follows a programming model which eases the interface to external third-party components (e.g., the automatic speech recognizer (ASR), natural language understanding (NLU), or synthesis component (TTS)).

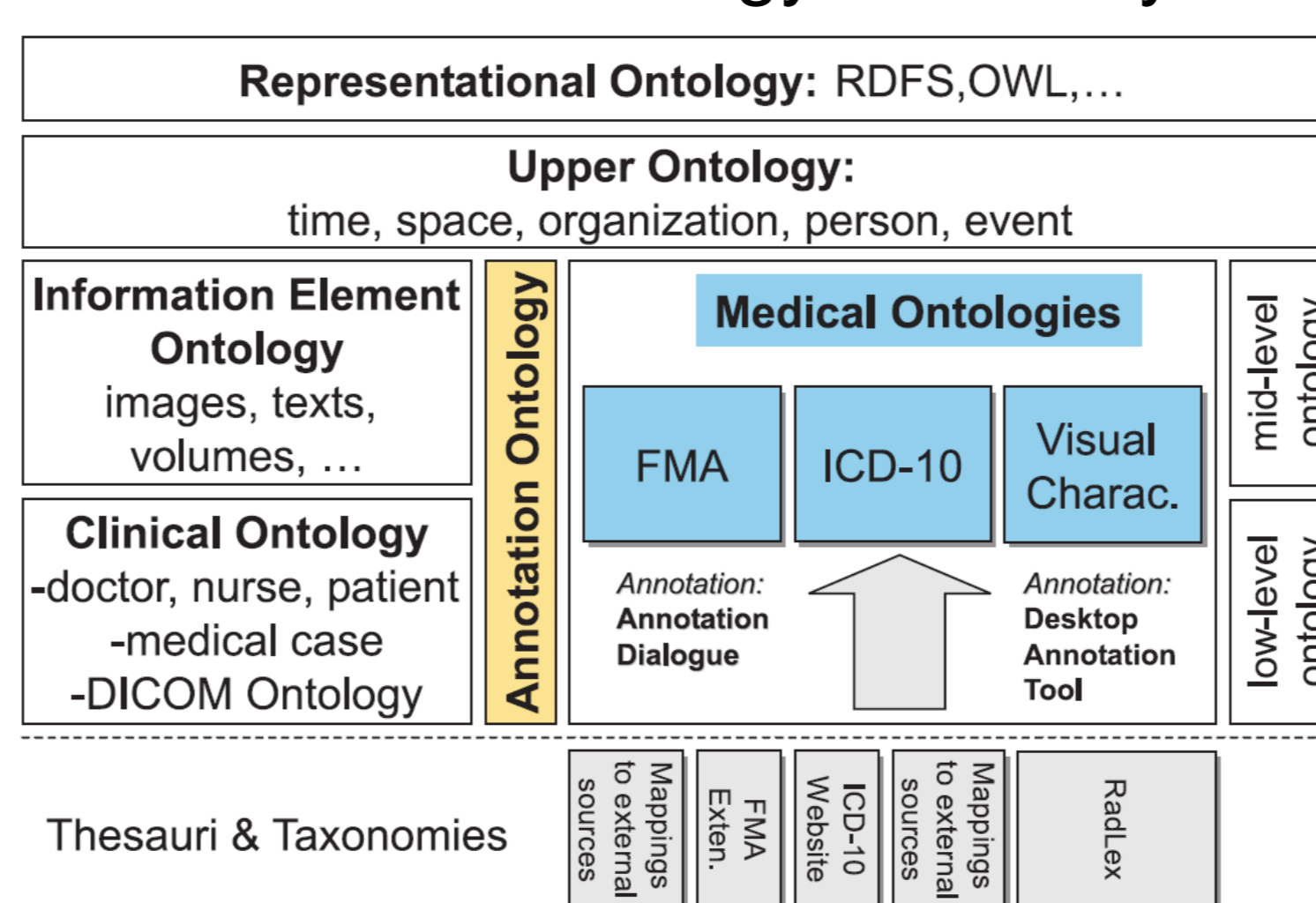
Desktop Application



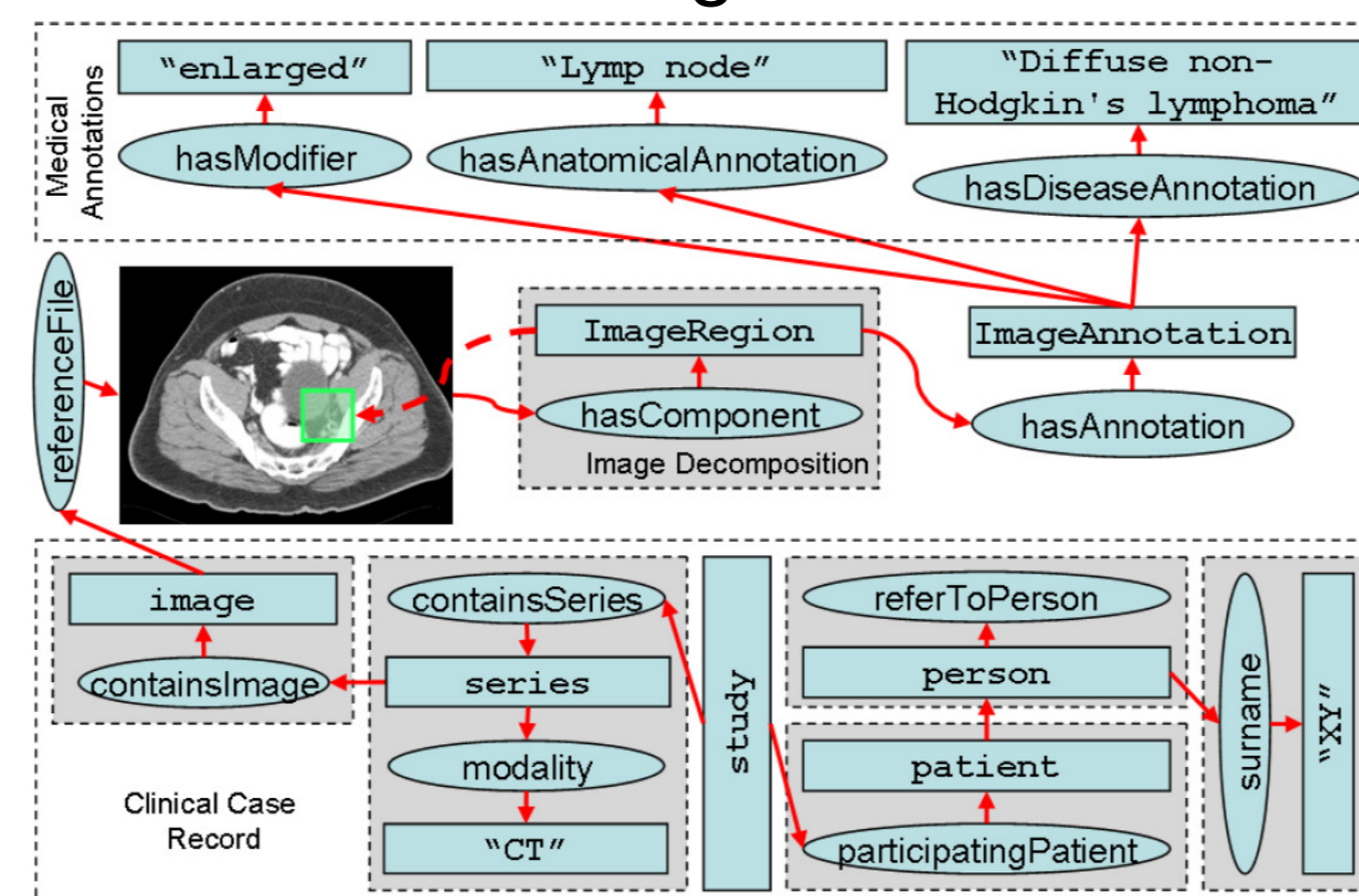
RadSem implements a method to annotate images and upload / maintain a remote RDF repository of the images and semantics.

Enabling Technologies

Medical Ontology Hierarchy



Semantic Image Annotations



Remote RDF Repository

The semantic image repository, a triple store setup at the remote RDF repository site, is based on two VMWare instances which differentiate between development and production environment. A direct access to the RDF statements is possible while using the query language SPARQL.

Multimodal Dialogue

1. U: “Show me the CTs, last examination, patient XY.” (retrieval stage)
2. S: Shows corresponding patient CT study picture series.
3. U: “Show me the internal organs: lungs, liver, then spleen.”
4. S: Shows patient images according to referral record.
5. U: “Annotate with lymph node enhancement (+ pointing gesture on region)”; so lymphoblastic (expert finding).”
6. S: “Region has been annotated.”
7. U: “And replace the characteristic of the other by RadLex: shrunken.”
8. S: “Region characteristic has been updated.”

The radiologist switches to another patient with a broken finger and asks for a summary (retrieval stage).

1. S: “This is a summary of the fracture: ...”
2. S: “Five corresponding CTs will be displayed.”

The radiologist switches to the differential diagnosis of the suspicious case (first patient), before the next organ (liver) is examined and the image annotations can be completed.