Engineering a Semantic Web for Pathology

Robert Tolksdorf[1], Elena Paslaru Bontas[2]
[1] research@robert-tolksdorf.de, http://www.robert-tolksdorf.de
[2]paslaru@inf.fu-berlin.de

Freie Universität Berlin
Institut für Informatik
AG Netzbasierte Informationssysteme
Takustr. 9, D-14195 Berlin Germany

Abstract. Digital pathology or telepathology intends to extend the usage of electronic images for diagnostical, support or educational purposes in anatomical or clinical pathology. Available approaches have not found wide acceptance in routine pathology, mainly due to the limitations of image retrieval. In this paper we propose a semantic retrieval system for the pathology domain. The system brings both text and image information together and offers advanced content-based retrieval services for diagnosis, differential diagnosis and teaching tasks. The core of the system is a Semantic Web gathering both ontological domain knowledge, and rules describing key tasks and processes in pathology.

1 Introduction

Digital pathology or telepathology intends to extend the usage of electronic images for diagnostical, support or educational purposes in anatomical or clinical pathology. The advantages of these approaches are generally accepted and several applications are already available. Nevertheless, none of the available products has found wide acceptance for diagnostic tasks, mainly due to the huge amount of data resulting from the digitalization process and the limitations of image-based retrieval. In this paper we propose a *semantic* retrieval system for the pathology domain. The system brings both text and image information together and offers advanced content-based retrieval services for diagnosis, differential diagnosis and teaching tasks. The core of the system is a Semantic Web gathering both ontological domain knowledge, and rules describing key tasks and processes in pathology. The usage of *Semantic Web* standards and domain ontologies facilitates the realization of a distributed infrastructure for knowledge share and exchange. The rest of this paper is organised as follows: The remaining introductory sections present the setting of the project, telepathology, and its main ideas and features. Chapter 3 provides an insight into the technical aspects of the retrieval system, by enumerating the technical requirements and the associated system architecture, followed by a detailed description of the system components. At this point we will present our achievements and the challenges we are currently confronted with in the realization of the main components. Chapter 4

delimits our approach from related research efforts in this domain, while Chapter 5 is dedicated to future work.

1.1 Telepathology

Telepathology is a key domain in telemedicine. By using telepathology approaches like virtual microscopy, pathologists analyze high quality digital images on a display screen instead of conventional glass slide at the common light microscope. In a typical digital pathology system, a camera is attached to a microscope and still images are taken. Images (with or without textual annotations) are stored in a database or directly in a patient record. Cases and images can be retrieved from the database or patient record as needed.

Health care information systems, which store and integrate information and coordinate actions among health care professionals, have been realized at various places in the last decades. New developments in telemedicine allow medical personnel to remotely deliver health care to the patient. At the Charité Institute of Pathology in Berlin, the first web-based virtual microscope allows histological information to be evaluated, transfered, and stored in digital format [17, 14]. This technique offers essential advantages compared to the classical approach, by supporting communication and exchange among professionals not sharing the same workplace location and improving quality assurance mechanisms [15]. However, to realize a complete computer-based infrastructure for pathology, one needs not only advanced support in the management of digital images. Necessary is also a more efficient integration of the medical findings, which are produced by pathologists to describe their observations from analyzing the slides at the light/digital microscope.

Common information systems in pathology restrict their retrieval capabilities to automatical picture analysis and ignore corresponding medical findings. Such analysis algorithms have the essential drawback that they operate exclusively on structural – or syntactical – parameters such as color, texture and basic geometrical forms while ignoring the real content and the actual meaning of the pictures. Medical findings, however, contain much more than that since they are textual representations of the pictural represented *content* of the slides. By that they capture the actual semantics of what the picture graphically represent, for example “a tumor” in contrast to “a red blob” or “a colocated set of red pixels”. Therefore, including medical findings in the information retrieval system goes beyond purely syntactic picture retrieval.

In the project described in this paper, we take the semantics aspects a step further: We understand the findings report as semantic metadata for the image prepared by an expert with high quality. We intend to make the semantic content explicit and build a system that takes advantage of the explicitly represented knowledge.

2 A Semantic Web for Pathology

The project "Semantic Web for Pathology" aims to realize a Semantic Web-based text and picture retrieval system for the pathology domain. For this purpose we concentrate our efforts in three interrelated directions: 1) the construction of a *knowledge base*, 2) the development of *knowledge reuse algorithms* and of a 3) *semantic annotation schema* for medical findings and digital histological images.

The knowledge base contains domain ontologies, generic ontologies and rules. Domain ontologies are used for the machine-processable representation of specific pathology knowledge, while generic ontologies capture common sense knowledge that can be useful in knowledge-intensive tasks. Several very complex libraries of ontologies are already available for this purpose. Rules are intended to formalize the key tasks in everyday pathology. While ontologies model the background knowledge of the pathologists, the rules are used to describe the decision processes using this knowledge: diagnostics, microscope analysis, observations etc. The acquisition of such rules, which play a crucial role for the retrieval, will be accomplished during an intensive collaboration with domain experts.

Further on, we analyze the textual data with text processing algorithms and annotate it with concepts from the knowledge base in order to improve precision and recall in retrieval operations. The annotation scheme is harmonized with the pathology knowledge base by using the corresponding medical ontologies as controlled vocabulary for the annotations. Text analysis is also used to extract implicit factual knowledge, which is subsequently integrated in the knowledge base.

2.1 Main features

We foresee several valuable uses of the planned system in routine pathology. First, it may be used as an assistant tool for diagnosis tasks. Since knowledge is made explicit, it supports new query capabilities for diagnosis tasks: similarity or identity of cases based on semantic rules and medical ontologies, differential diagnosis, semantically precise statistical information about occurrences of certain distinguishing criteria in a diagnosis case. The provided information will be very valuable in diagnosis work especially for the underdiagnosed cases, since such situations require deeper investigations of the problem domain and a very strict control mechanism of the diagnosis quality ([5]).

Second, advanced retrieval capabilities may be used for educational purposes by teaching personnel and students. Currently, enormous amounts of knowledge are lost by being stored in data bases, which are behaving as real data sinks. They can and should be used for teaching, e.g. for case-based medical education.

Third, quality assurance and checking of diagnosis decisions can be effectuated more efficiently because the system uses axioms and rules to automatically check consistency and validity.

Finally, explicit knowledge can be exchanged with external parties like other hospitals. The representation within the system is already the transfer format for information. Semantic Web technologies are by design open for the integration of

knowledge that is relative to different ontologies and rules. Therefore we intend to use mainly such technologies for the realization of the retrieval system.

2.2 Use cases and technical requirements

The technical analysis and design of the pathology retrieval system is closely related to typical usage scenarios, which are not necessarily related to routine pathology. Most probable, the system will be used for underdiagnosed cases, where a second or third opinion is to be consulted or the specialist usually reverts to certified controll sources, like Internet or printed material. Such information sources have an essential drawback: they offer limited capabilities for a thematically focused search. Both manual search within printed materials and Internet search, based on common or medicine-related search engines, is time-consuming and not specific enough to be integrated in everyday pathology. Instead, our system will offer the possibility to search the archiv of medical findings for similar cases or differential diagnosis. It is improbable that the system will be consulted for routine cases, covering approximately 80 percent of the total amount, which are on the fly analyzed by the pathologists without the need for additional information sources.

The acceptance of the system is strictly related to its minimal invasive character: it should not imply any change of the current work flows and should achieve good precision results. Recall is also important, but since the two parameters are usually contradictory, we favor precision, mainly because of the predominant usage of the system for underdiagnosed cases, within which every detail may play an important role for the final results. The minimal invasive feature will be reflected in a careful design of the user interfaces and a intuitive query language.

Another important setting is teaching: therefore, the system should be able to generate different reference materials and to retrieve information about typical pathology cases and their diagnosis. The key feature for the second scenario is the flexibility to generate and present domain information.

The network aspect is important for both settings. Pathologists use the system for cases where they need the remote collaboration of other specialists. The teaching scenario asumes also a distributed infrastructure, so that the resources can be accessed anytime, anywhere. The usage of Semantic Web technologies on one side, and of standards like XML/OWL and the medical HL7/DICOM is a condition for the realization of this requirement.

Scalability and performance are critical factors for the acceptance of retrieval system. In our application, the amount of image data is impressive. Every particular case contains up to 10 medical findings. Each of these are based on up to 50 digital histological images, which usually have a size of 4-5 GB each. Our first prototypical implementation of the system will deal with approximately 400 findings and a part of the corresponding digitised slides.

The storage of images will still be subject to the use of specialized image databases. Our approach of resorting to the description of images contained in the findings and their processing in the system makes the requirements on

scalability with the number and complexity of cases independent on the size of the image data. There is no image processing foreseen, instead we use the result of the image analysis performed by human experts, the pathologists.

Remaining scalability and performance issues are affected by the quality of the underlying Semantic Web components and the complexity of models used and inferences drawn therein. Currently, there are strong effort to produce industrial strength Semantic Web components, such as inference engines that go beyond the poor performance of early research prototypes. Our system will benefit from this performance gain in the infrastructure.

The complexity of models, rules and queries triggering inferences remains a critical issue. While we have a substantial basis of models with existing standards it it not clear yet, what heuristics should guide the selection of the granularity of models eventually used and of the details of rules applied when finding "similar" cases. We will restrict ourselves to small models and rulesets that generate a suffient precise answers by the system with minimal inferencing effort. The precise methodology for doing so is subject of our current studies.

3 Engineering the System

Technically the system resorts to Semantic Web technologies. The Semantic Web ([1]) aims to provide automated information access based on machine-processable semantics of data. The final vision is to develop a technological framework that will transform the Web in an huge network of both human- and machine-understandable knowledge with various specialized reasoning services. The first steps in this direction have been made through the realization of appropriate representation languages for Web knowledge sources like RDF(S) and OWL and the increasing dissemination of ontologies, that provide a common basis for annotation and support automatic inferencing for the generation of knowledge.

Our approach makes use of these Semantic Web technologies in order to represent pathology knowledge explicitly and, consequently refine the retrieval algorithms on a semantic level: medical and generic ontologies are integrated into a pathology knowledge base, which serve also as annotation vocabulary for medical findings and histological images. We use OWL and RDF(S) both for the representation of the knowledge base and for the annotation of the information items and XML-based medical standards like HL7/CDA ([10, 9]) for the medical findings.

In medicine and biology exhaustive domain ontologies have been developed and are constantly incorporating new pieces of knowledge. Ontologies like UMLS ([16]), GALEN ([6]), Gene Ontology ([4]) provide a good basis for the development of Semantic Web applications for medicine purposes. These ontologies are therefore used as the initial knowledge base of the semantical retrieval system for pathology. In addition, to put our goals into practice we still need to integrate the individual domain knowledge sources and to adapt them to the requirements of the Semantic Web, which means in the first place to formalize them in a Se-

mantic Web representation language. Our analysis in the application domain has revealed the necessity of a powerful representation language, which can capture most of the semantical features of the medical knowledge. For this purpose we will use mostly OWL instead of RDF(S), mainly because of its expressiveness and inferencing capabilities. The main issues we address w.r.t. the available medical ontologies will be explained in detail in Section 3.2

3.1 System architecture

We propose the following system architecture, which has arisen from the use cases and the corresponding technical requirements (Figure 1):

- **description component**
- **knowledge component**
- **transformation component**
- **application components**

In the following we briefly explain the role of each component and their interaction, a detailed description of the features and related research issues is presented in Section 3.2.

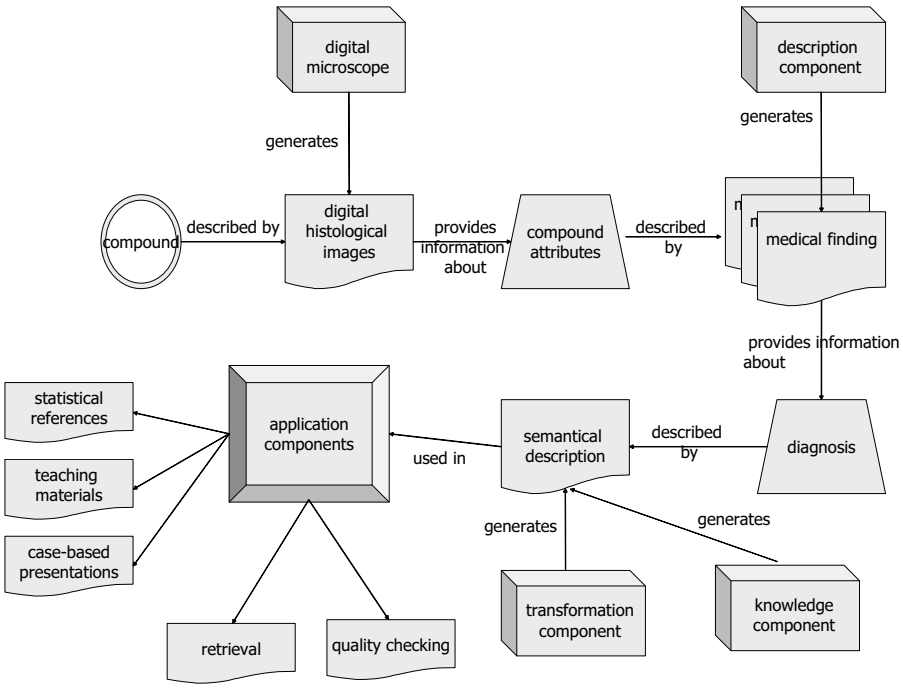

Fig. 1. Systemarchitecture "Semantic Web for Pathology"

The core of the system architecture is the knowledge component (Figure 3), which consists of domain and generic ontologies, as well as a rule engine. The knowledge component influences every process of the remaining components. The medical findings and histological images are analyzed semantically and linguistically within the description component. The explicitely represented knowledge is used to check the consistency of medical findings and picture annotations during their generation. The description component also allows the XML encoding of the textual and pictural data. Both the available pathology data base at the Charité hospital and data to be generated are described in XML in this manner. The transformation component takes the XML-structured data set and integrates it within the semantic network underlying the knowledge component. Due to the application-oriented caracter of the system, special attention in the architecture is paid to the application components, which implement the functionality of the system as presented in Section 2. The search component is used both by pathologists in order to retrieve information concerning diagnosis tasks or by teaching personnel and students. We plan also a component for the generation of statistical evaluations (e.g. related to the most frequent desease symptoms, relationships between patient data and desease evolution etc.) and for the generation of case-oriented teaching materials and presentations (see Figure 1). The quality checking service is intended to evaluate the consistency of medical findings.

3.2 Main components

The Description Component The description component is concerned with the basic formalization of medical findings and digital histological images. For this purpose it deals with two principal data sources: data, which is already available at the Institute of Pathology at the Charité hospital and future data. The goal of this process is to offer a homogeneous encoding of medical findings, on one side and picture annotations on the other side, both for existent and future material. The data should be first encoded in XML and subsequently analyzed using ontology-enhanced text analysis algorithms in order to be annotated with ontology concepts. For the generation of new XML-based information we developed an editor tool, which can be integrated in the actual version of the Digital Virtual Microscope ([17, 14]). By means of this tool pathologists can analyze digitised histological images and simultaneously enter or update the corresponding medical finding, which is subsequently stored in a XML data base. The second source of raw data was naturally the medical findings archive at the Charité. The medical findings of this type have been extracted from their primary text-oriented storage and transformed in XML.

We developed a HL7/CDA compatible XML-scheme for the medical findings, which reflect the logical structure of the data. Such medical data is organized more or less consequenlty in four major parts:

- **macroscopy** describing physical properties and the appereance of the original compound.

- **microscopy** concerned with the detailed description of the slides analyzed at the microscope.
- **diagnosis** sumarizing the conclusions and the diagnosis
- **comments** usually presenting additional facts playing a role in the diagnosis argumentation (patient data, patient history etc.) or an alternative diagnosis for ambiguous cases.

Besides, such a medical finding contains also information from the patient record and references to digital images. The connection to the digital images is fundamental for an efficient retrieval, which should contain apart from the relevant textual information the corresponding image region the pathologist refers to in a certain portion of text. Since the size of such images is 4-5 GB, it is not sufficient to retrieve complete images to a certain user query, but the concrete image sector. For this purpose we use the functionality of the Digital Virtual Microscope, which allows digital slides to be annotated with so-called "observation paths" on one side, and registry an additional "dictation path". The observation path contains image coordinates, image resolution and time stamps registered while the pathologist was analyzing a specific digital image. The dictation path sums up the same data, this time registered while the pathologist was typing the medical finding. The complete path-related information flows in the "diagnosis path", which mirrors the way the diagnosis decision was accomplished.

The proposed XML-Scheme reconstructs the structure of the real medical findings and is HL7-compatible. Though the compatibility restricts the format of the XML findings (the information must be encoded within "section", "paragraphs" and "coded_entry" tags, which is not necessarily the most straight forward manner of formalizing it), it is an important issue, especially for the distributed setting, for the exchange and reuse of information.

The Knowledge Component The knowledge component includes the medical knowledge base and the algorithms for the realization of the applications. As mentioned in Section 3.1 it is build of a library of domain and generic ontologies, a rule engine and the annotated pathology data (Figure 3). We use available medical ontologies as a foundation of the knowledge base, starting with UMLS ([16]) and Gene Ontology ([4]).

The most important issue we have to address when building the pathology knowledge base is the integration and the enrichment of the available medicine standards. Medicine ontologies though containing a huge amount of concepts or termini have seldom been developed for machine processing, but rather as controlled vocabularies and taxonomies for specific tasks in medicine ([13]).

From a strict Semantic Web point of view they proved to be deficiently designed and incomplete. Apart from the absence of an at least Semantic Web compatible representation language, UMLS and Gene Ontology adopt an error-prone modeling style, which is characterized by few semantic relations among concepts and an ambiguous way to interpret such relations (e.g. concepts of the UMLS Metathesaurus are connected through relations like "related", "broader",

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no" ?>
<levelone xmlns="urn::hl7-org/cda"
xmlns:sciphox="urn::sciphox-org/sciphox"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn::hl7-org/cda sciphox-cda.xsd"
xmlns:swpatho="urn::swpatho-org">
<clinical_document_header>...
<local_header ignore="all" descriptor="swpatho">
<swpatho:swpatho-ssu type="Kostentraeger" country="de" version="v1">
<swpatho:Kostentraegerbezeichnung V="CHA" /></swpatho:swpatho-ssu>
<swpatho:swpatho-ssu type="Schreibkraft" country="de" version="v1">
<swpatho:Schreibkraftkuerzel V="SKFX" /></swpatho:swpatho-ssu>
<swpatho:swpatho-ssu type="E-Nummer" country="de" version="v1">
<swpatho:E-Nummer V="E01152-01" /></swpatho:swpatho-ssu>
</local_header>...</clinical_document_header>
<body>
<section><caption>Befund</caption>
<section><caption>Makroskopie</caption>
 <paragraph><content>Zwei Gewebszylinder von 15 und 4 mm Länge.
 </content></paragraph>
</section>
<section><caption>Mikroskopie</caption>
 <coded_entry><coded_entry.value V="5" S="UID" /></coded_entry>
 <coded_entry><coded_entry.value V="6" S="PID" /></coded_entry>
 <coded_entry><coded_entry.value V="Mon Feb 09 13:53:16 CET 2004" S="Start"/>
 </coded_entry>
 <coded_entry><coded_entry.value V="Mon Feb 09 13:53:18 CET 2004" S="Stop"/>
 </coded_entry>
 <paragraph>
 <content>Stanzbiopsate aus Lungengewebe mit deutlicher Störung
 der alveolären  Textur, soweit noch nachweisbar deutlich
 Verbreiterung der Alveolarsepten, stellenweise Nachweis von
 Bronchialepithelregeneraten. Restliche Alveolarlumina z.T. durch
 Fibroblastenproliferate verlegt. Im Interstitium ein gemischt
 entzündliches Infiltrat, bestehend aus Plasmazellen und
 Lymphozyten. ... </content>
 </paragraph>
</section>
<section><caption>Kritischer_Bericht</caption>
<paragraph><content>Stanzbiopsate aus der Lunge mit Zeichen der
organisierenden Pneumonie (klin. Mittellappen).</content></paragraph>
</section>
<section><caption>Kommentar</caption>
 <paragraph><content>Nach klinischer Angabe vordiagnostiziertes
 kutanes T-Zell-Lymphom, jetzt 2 bis 3 cm große
 pleuraständige Raumforderung im Mittellappen. Im vorliegenden
 Material kein Anhalt für eine Lymphom-Manifestation. Kein
 Karzinom. </content></paragraph>
</section>
</section>
</body>
</levelone>
```

Fig. 2. Fragment of an XML-encoded medical finding

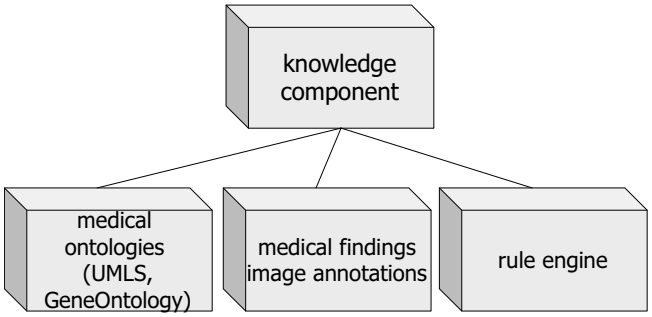

Fig. 3. The knowledge component

"narrower"). A typical example is the usage of the relation "is-a" for both instantiation and specialization/generalization, the usage of a unique "part-of" relation with different meanings ("functional part", "content", "component", "substance") or the usage of one of these relations instead of the other. Mathematical properties of the same semantical relation (e.g. transitivity) are not fulfilled for each pair of concepts connected by the relation and the "is-a" relation between two concepts does not always guarantee the inheritance of the properties of the parent concept to its children (so-called "blocked" relations in UMLS). Besides relations, both UMLS and GeneOntology contain a huge set of conceptual entities, organized in several taxonomies. The classification criteria for concepts are inconsistent and incomplete. Different, unspecified granularities are used within a hierarchy and properties may not be inherited along inheritance paths.

The issue of the restricted representation language is addressed in the several projects, which usually develope their own representation languages, adapted to specific requirements of the medical domain w.r.t expressiveness and inferencing capabilities. Such an ontology, though with increased inference capabilities compared to UMLS or Gene Ontology can not be embedded offhand in a Semantic Web application, shared or completed in a Semantic Web setting. Even more, various ontologies have been developed for particular purposes and and can not be integrated automatically. Besides integration and completion such ontologies do seldom contain axiomatic knowledge which is essential for diagnostics or therapy settings.

Therefore we need to adopt a Semantic Web representation scheme for the available ontological knowledge, complete it with additional axioms and definitions on one hand and on the other hand encode therapy, diagnostic and task knowledge in a supplementary module as rules. For this purpose we will use standards like RuleML. In order to design an appropriate representation based on Semantic Web we will first identify in collaboration with domain experts the fragments of UMLS/Gene Ontology, which are relevant in the pathology. Secondly we will analyze the deficiencies of the available medical standards by

transforming their content in OWL and automatically discovering inconsistencies. The next step will be the manual adaptation of the OWL ontology according to the results of the previous procedure. Currently we are implementing an algorithm for the OWL transformation of UMLS knoweldge sources. The underlying modelling primitives are illustrated in Figure 4.

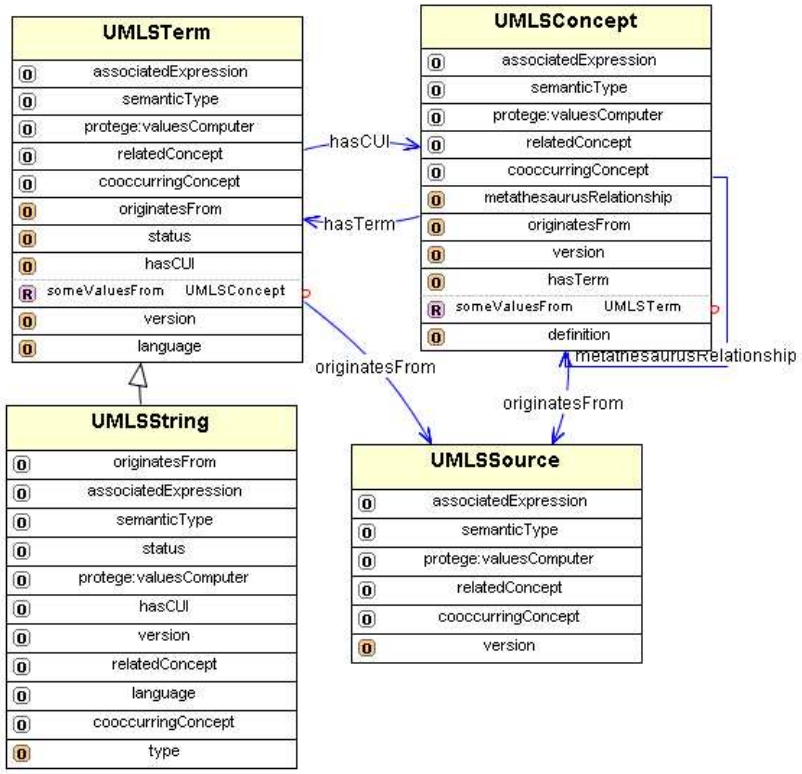

Fig. 4. UMLS modelling primitives in OWL

Besides an ontology-based background the knowledge base also contains the complete set of medical findings and image descriptions, both represented in XML by means of the description component. However, in order for this additional information to be involved in retrieval and knowledge discovery, the XML basic scheme needs to be enriched with annotations referencing ontology concepts and relations. For this purpose we intend to use text processing algorithms for an initial automatic annotation phase and to implement an annotation tool for a subsequent manual annotation phase, which completes the automatic process.

The Transformation Component The tranformation component implements features required for the text-based processing of the medical findings and image descriptions. For this purpose we are currently developing a noun phrasing module, which identifies domain-specific phrases from medical findings. The modules incorporates a tokenizer, a tagger and a ontology-based phrase generator. The phrase generation process interracts with the knowledge base, since it uses medical ontologies to identify relevant (multi-word) phrases and in the same time puts together a lexicon, tailored for the particular application setting: the domain of lung pathology and the language used in the medical findings, which is German. The lexicon provides us indications about the usage limitations of an essentially English-oriented thesaurus like UMLS in our concrete setting. As a result of the phrasing module, the XML-encoded medical findings contain semantic relevant phrases, which can be referenced to concepts of the knowledge base. This task will be realized by the annotation component.

Application Components The Semantic Web for Pathology will assist the following application components:

- **search component** will be used primarily for diagnosis tasks. It will allow not only the basic retrieval of text/image information items, but also support differential diagnosis tasks. The semantic retrieval is oriented towards several typical categories of queries:
  - **statistical queries** e.g. the probability/frequency of a particular carcinoma in a certain age group.
  - **matching queries** e.g. comparison of cases with common characteristics, text and image information to similar cases.
  - **image queries** e.g. cases containing images with certain content- or image-specific constraints.

  Besides, the retrieval should be adapted to the characteristics of the pathology domain and involve issues like the diagnosis path. (Section 3.2).
- **quality checking component** will be used in quality assurence and management of diagnosis processes. Quality criteria, diagnosis standards and their verification are expressed by means of rules.
- **statistical component** will generate statistical material related to the relative frequency or demographic distribution of diseases and their complications.
- **teaching component** will generate teaching materials, using features of the previous components (statistical studies, reference cases)

4 Related Work

Medicine is one of the best examples of application domains where ontologies have already been deployed at large scale and have already demonstrated their utility. Most of these domain ontologies (UMLS inclusively)underlie different design requirements as computer supported and even more specific Semantic

Web applications. They are actually huge collections of medical terms, organized in hierarchies and cannot be used directly in Semantic Web applications. This issue has been addressed in project GALEN ([6]), where the authors developed a special description logic representation, tailored for the particularities of the (English) medical vocabulary. However, the usage of a proprietary representation makes the ontological knowledge difficult to be extended by third parties or exchanged in a Semantic Web.

The usage of ontologies for building knowledge bases for medicine has already been subject of several research projects ([2, 12, 7, 3, 8]). The most important representatives are the ONIONS ([7]) and MEDSYNDIKATE ([12]) projects. In ONIONS the authors aim to develop a generic framework for ontology merging and use UMLS as an example to apply their methodology. Therefore they need a detailed analysis of the ontological properties of UMLS, using a Loom formalization. MEDSYNDIKATE is also confronted with the ontological commitment beyond UMLS in order to use it in text processing algortihms for knowledge discovery. UMLS serves in this case as an annotation vocabulary for medical texts. Both projects offer valuable experiences and facts concerning UMLS and medical ontologies generally, but they do not use Semantic Web technologies to facilitate knowledge share and reuse, which is the crucial feature of ontologies. An interesting approach can also be found in [2], where the authors compare UMLS with other ontologies (e.g. WordNet ([11], GeneOntology) to establish its appropriateness as terminology for biomedical applications.

5 Conclusions and Future Work

In this paper we presented our work towards a Semantic Web based retrieval system for pathology. The system is based on a comprehensive knowledge base, which formalizes pathology-relevant knowledge explicitly by integrating available medicine ontologies like UMLS and rules describing diagnostic guidelines. It is intended to provide both retrieval and knowledge management functionalities. In order to achieve these goals we designed by now the system architecture, adopted XML-based schemes for the uniform representation of medical findings and digital images and developed a methodology for the construction of the pathology knowledge base. Current work includes the specification and implementation of an algorithm for the OWL formalization of medical ontologies and their integration in the knowledge base.

Acknowledgement The project "Semantic Web in the Pathology" is funded by the Deutsche Forschungsgmeinschaft, as a cooperation among the Charité Institute of Pathology, the Institute for Computer Science at the FU Berlin and the Department of Linguistics at the University of Potsdam, Germany.

References

1. T. Berners-Lee, J. Hendler, and O. Lassila. "The Semantic Web". *Scientific American*, 284(5):34–43, 5 2001.
2. A. Burgun and O.Bodenreider. Mapping the UMLS Semantic Network into General Ontologies. In *Proc. of the AMIA Symposium*, 2001.
3. G. Carenini and J. Moore. "Using the UMLS Semantic Network as a Basis for Constructing a Terminological Knowledge Base: A Preliminary Report". In *Proceedings of 17th Symposium on Computer Applications in Medical Care (SCAMC '93)*, 1993.
4. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–30, 2000.
5. F. Demichellis, V. Della Mea, S. Forti, P. Dalla Palma, and C.A. Beltrami. "Digital storage of glass slide for quality assurance in histopathology and cytopathology". *Telemed Telecare*, 8(3):138–42, 2002.
6. Ontology GALEN. http://www.opengalen.org, 2001.
7. A. Gangemi, D. M. Pisanelli, and G. Steve. "An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies". *Data Knowledge Engineering*, 31(2):183–220, 1999.
8. H. Gu, Y. Perl, J. Geller, M. Halper, L. Liu, and J. Cimino. "Representing the UMLS as an OODB: Modeling issues and advantages", 2000.
9. HL7 Standard. http://puck.informatik.med.uni-giessen.de/people/messaritakis/-hl7xml/hl7stand.htm, 2000.
10. The HL7/CDA Standard. http://www.hl7.org, 2000.
11. G. A. Miller. "WordNet: a lexical database for English". *Communications of the ACM*, 38(11):39 – 41, 1995.
12. S. Schulz and U. Hahn. "Medical knowledge reegineering - converting major portions of the UMLS into a terminological knowledge base". *International Journal of Medical Informatics*, 2001.
13. S. Schulz, M. Romacker, and U. Hahn. "Knowledge engineering the UMLS". *Stud Health Technol Inform*, 77:701–5, 2000.
14. Patentanmeldung: Slide Scanner – Vorrichtung und Verfahren, 2002. Aktenzeichen 102 36 417.6 des DPMA vom 5.8.2002.
15. J. Slodkowska, K. Kayser, and P Hasleton. "Teleconsultation in the Chest Disorders". *Eur J Med Res*, 7 (Suppl I):80, 2002.
16. Unified Medical Language System. http://www.nlm.nih.gov/research/umls, 2002.
17. Patentanmeldung: Virtuelles Mikroskop – Vorrichtung und Verfahren, 2002. Aktenzeichen 102 25 174.6 des DPMA vom 31.05.2002.