

# Experiences Using Semantic Web Technologies to Realize an Information Retrieval System for Pathology

Elena Paslaru Bontas<sup>1</sup>, Sebastian Tietz<sup>1</sup>, Thomas Schrader<sup>2</sup>

<sup>1</sup>Freie Universität Berlin

Institut für Informatik

AG Netzbasierte Informationssysteme

Takustr. 9

D-14195 Berlin, Germany

*paslaru@inf.fu-berlin.de, tietz@inf.fu-berlin.de*

<sup>2</sup>Institute of Pathology Charité

Rudolf-Virchow-Haus

Schumannstr. 20-21

D-10117 Berlin, Germany

*thomas.schrader@charite.de*

**Abstract:** Building the Semantic Web requires the use of powerful tools to create, manage and extend domain ontologies represented with Semantic Web languages. Though many tools have been already developed, unfortunately there is a lack of practice using them in real-world applications. The suitability of these tools and the underlying technologies, especially w.r.t. performance and scalability parameters, is still an open issue. This paper focuses on problems and experiences gained in the project "A Semantic Web for Pathology", a research project using domain ontologies and ontology-driven natural language processing to support a content-based storage and retrieval for text and image-based medical information.

## 1 Introduction

Ontologies are a key technology for realizing the Semantic Web. They have already shown their usability and potential in different application domains [Fe01]. In medicine large-scale ontology domains have been implemented: UMLS [UML02], GALEN [Gal01], NCI [NCI]. Though containing huge amounts of useful domain knowledge, most available medical ontologies have been designed under different design principles than those required for Semantic Web applications and therefore can not be directly integrated in such applications. On one hand most of the available ontologies are not formalized in an appropriate representation language to be shared and reused. On the other hand they have been realized for very concrete tasks and their content is modelled in an ambiguous way.

In order for these knowledge sources to be easily integrated and reused by real-world applications one needs powerful tools, which should assist the ontology engineer in managing and evaluating available ontologies, and programming interfaces to support the efficient generation, query and storage of Semantic Web-based knowledge bases.

In this paper we describe our experiences using available tools and medical knowledge sources for the development of a Semantic Web-based retrieval system for the domain of "lung pathology". As an initial input for the domain ontology we used UMLS, since it integrates most of the significant medical libraries so far, and adapted it to the requirements of our concrete application setting. This domain ontology is part of a knowledge base also embedding rules describing decision processes in typical pathology work flows, and concept instances extracted from real pathology reports.

The rest of the paper is organized as follows: Section 2 gives a brief overview of the application scenario and the corresponding requirements w.r.t. the domain ontology and its usage. Section 3 presents the core ontology used by the retrieval system, introduces requirements for assistant tools needed during this process and presents our experiences in using the available tools. Section 4 outlines our conclusions as a result of the ontology development process.

## **2 A Semantic Web for Pathology: Architecture and Main Components**

The aim of the project "A Semantic Web for Pathology"<sup>1</sup> is the realization of a content-based retrieval system for pathology-specific data. In order to improve the efficiency of the retrieval mechanisms for the available data the system uses an explicit semantic representation of textual pathology reports and links them to the digital images they describe [TPB04]. The semantic representation of the data consists of a set of instances of ontology concepts belonging to a pre-defined knowledge base. The medical knowledge base is supposed on one hand to control the annotation vocabulary and to assist this transformation process. On the other hand it is used to refine the retrieval capabilities of the system. By using the background knowledge including both application-relevant (i.e. pathology) and generic facts the system will be able to answer content-related queries like "images containing a tumor with certain properties" or "images where the tissue sample presents certain morphological characteristics", to compare case reports or support differential diagnostics tasks (i.e. retrieve reports with a similar appearance and alternative diagnosis).

The most important design criterion to obtain a high user system acceptance is its integration into the work flow of a pathologist. This minimal invasive usage will be the result of a careful design of the user interface, query language and content presentation methods. Furthermore, the system offers the ability to write a medical report while examining the images with the conventional microscope or related software (for a detailed description of the application setting w.r.t. "Digital Pathology" see [TPB04, PBTS04]). The new report is annotated with ontology concepts by a text processing component [SSPB] and the resulting semantic representation is forwarded to a quality assurance module to check its validity w.r.t. the content of the knowledge base. The annotated report is stored in a database containing all available reports, while the instantiated concepts and relations found in the newly created report (i.e. its semantic representation) become part of the knowledge base. A retrieval component performs queries and retrieves images corresponding to the pre-defined user criteria. Data (e.g. patient records and case reports, as

---

<sup>1</sup><http://www.inf.fu-berlin.de/inst/ag-nbi/research/swpatho/>

well as the ontology knowledge) should be shared or exchanged among domain experts and health-care organizations. Therefore the system makes use of standard, platform independent technologies which support these requirements, like Semantic Web technologies and Web Services. The system has a Web-based distributed architecture to simplify the exchange among health-care organizations and insurance companies and the possible involvement of various hospitals in the deployment of the knowledge base, as well as the usage of the medical data for teaching purposes. The derived system architecture is illustrated in Figure 1.

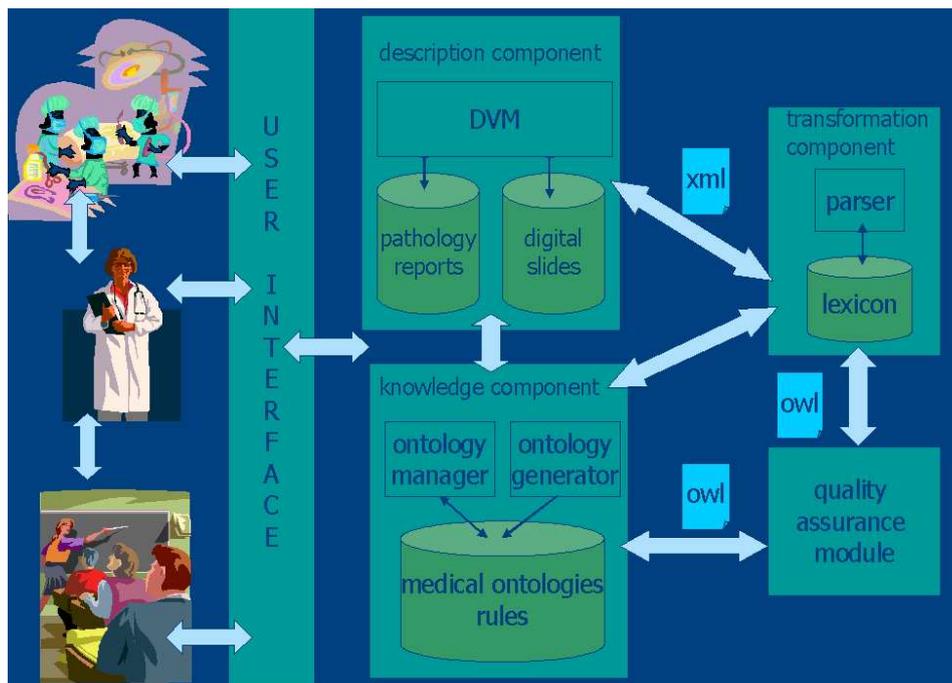


Figure 1: System architecture "Semantic Web for Pathology"

The system components and their main features are described in [TPB04]. By now we developed a first draft of the medical knowledge base (see Section 3), a prototypical client application including a component to generate and edit medical reports and annotate digital images and a prototypical ontology-based retrieval component. The client communicates with a number of Web services, which analyze, annotate and store the given information and query the knowledge base. While developing a prototype of the retrieval system, a variety of tools and applications have been used and developed to create, store, query, inference and extend the domain ontology. This paper concentrates on Semantic Web technologies and associated tools being used during the development of the system.

## 3 Implementing the Knowledge Base

### 3.1 The Content of the Knowledge Base and its Information Sources

The core of the knowledge base is based on the Unified Medical Language System (UMLS) [UML02]. UMLS integrates more than 100 medical libraries, containing about 1.8 million concepts with more than 11 Millions of relationships. It unifies the most important medical libraries like SNOMED [SNO98], MeSH [MeS03] or the Gene Ontology [Th00]. The UMLS consists of two parts: The Semantic Network [UMLb] and the Metathesaurus [UMLa]. The Semantic Network contains generic medical and non-medical categories and relationships (approximately 150 concepts and 50 relations, e.g. "body part", "disease", "physical concept", "organization"). It is used as a meta level for the information in the Metathesaurus, which brings together the particular UMLS libraries. The Metathesaurus consists of a list of uniquely identified concepts and several generic relations. Every concept in the Metathesaurus references at least one semantic type of the Semantic Network. The relations between concepts are usually typed by means of the semantic relations from the Semantic Network. Though containing a huge amount of domain information, the ambiguous and error-prone integration schema and the heterogeneity of the libraries made the reuse of UMLS as information knowledge source in applications as retrieval, annotation and text processing difficult [HMS96, GPS99, SH01, HS02].

The domain ontology has been constructed using a combination of a top-down and a bottom-up approach: In the first place, a certain amount of domain knowledge has been extracted out of UMLS (the top-down case), afterwards the domain was extended manually by inserting missing concepts, which occurred frequently in concrete medical reports, and alternative names, especially German translations of the English UMLS concepts<sup>2</sup>. The extraction of the "lung pathology"-relevant concepts is not a trivial task, because of the overall dimensions of UMLS, the heterogeneity of specific UMLS libraries and the lack of detailed information concerning their content. A tool called MetamorphoSys developed as part of the UMLS Knowledge Server [UML03] eliminates complete libraries and vocabularies in a specific language<sup>3</sup>. Domain experts identified the 10 most important knowledge sources in order to cut down the remaining ontology to the potentially relevant content. Unfortunately most of the UMLS knowledge is concentrated in a few large source libraries, therefore the selection resulted in more than 500.000 potentially important concepts. In the bottom-up phase pathologists selected four central concepts for lung pathology ("lung", "pleura", "trachea", "bronchia") and the concepts being connected to them by a semantic relationship. We obtained a list of about 1000 concepts having 10.000 relations which describe the anatomy of the lung and lung diseases and used it as initial input for the domain ontology (Figure 2). After selecting the relevant concepts we translated the UMLS data model to the OWL model and transformed the relevant data from one format to the other. We choose OWL [OWL0] as a representation language since it is one of the standard Semantic Web representation languages and because of its expressiveness.

<sup>2</sup>After a first selection of useful libraries, among 500.000 concepts only 12.000 had a representation in the German language.

<sup>3</sup>Note also that the non-English terms are obtained after a translation of a corresponding English-based library and there is no autonomous library in a different language other than English.

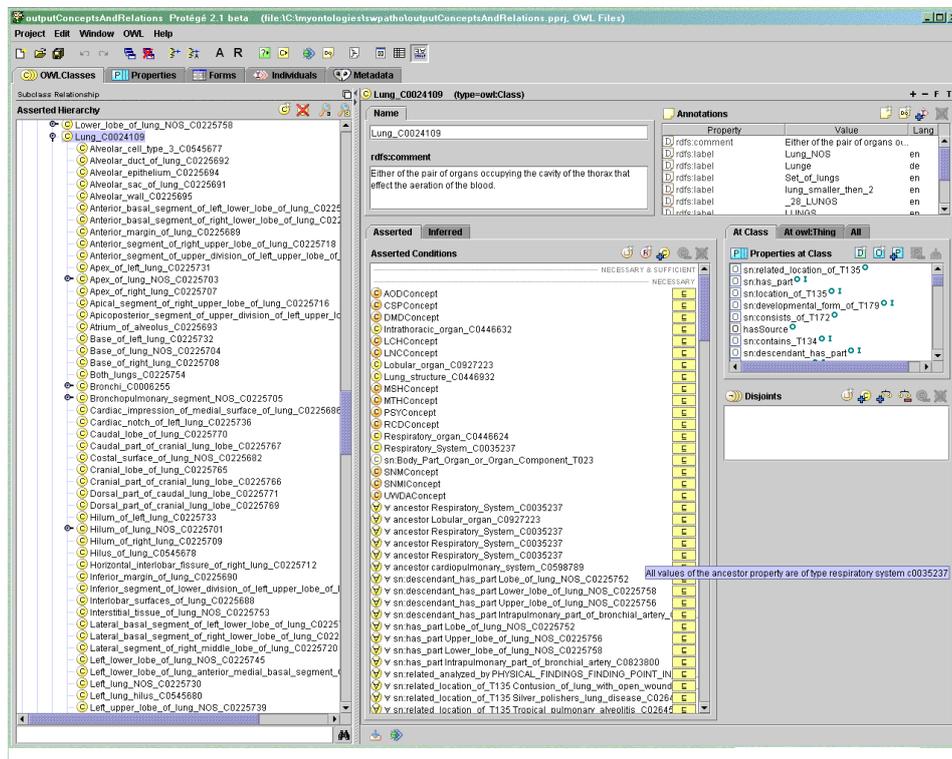


Figure 2: A Fragment of the core domain ontology using the Protégé environment

The linguistic analysis of the patient report corpus evidenced the content-related limitations of UMLS w.r.t. the concrete vocabulary of the report archive. We modelled additional pathology-specific concepts, like the components and typical content of a medical report, and integrate them in the available ontology library. Besides content-related adaptation needs, the analysis of the generated ontology outlined several “syntactical” issues for further adaptations:

- concept names in UMLS: concepts like “ARF-smaller-than-2”, “RESECTION OF LUNG WITH RECONSTRUCTION OF CHEST WALL WITHOUT PROSTHE-SIS”, “Unspecified injury of lung with open wound into thorax” are unlikely to be relevant for the retrieval of pathology reports. Often, the only mapping between natural language (i.e. search terms) and concepts in an ontology is the “name” of the concept. In order to detect all occurrences of a search term in a text (in our case, a medical report) one has to match the query string against the names of the concepts mentioned in the documents. Besides, the way certain concepts are named in UMLS evidences that such concepts should be modelled as such in the ontology (i.e. as a single concept), but rather using corresponding properties (e.g. the concept “ARF”

- acute renal disease - should have a property called "smaller-then").

- the absence of concept names in German language: due to the predominance of English in denominating UMLS concepts and the predominance of German terms in the pathology report archive in our application setting one needs to translate the English terms in order to achieve an efficient retrieval.

The comparison of the vocabulary of the medical reports archive with the generated ontology also emphasized the need to extend the knowledge base with non-medical content. Especially part-whole and spatial relationships are often encountered in medical findings and are therefore included to the ontology library. Medical reports frequently contain ambiguous terms to describe the results of the examinations (e.g. terms like "high-grade", "slightly", "mainly" ), which play an important role for the overall interpretation of the reports. The representation of such terms is still subject of future work.

### 3.2 OWL Representation of the Medical Knowledge

In order to formalize the application-relevant knowledge in OWL we implemented a mapping from the UMLS data model to OWL. As described in the previous section, every UMLS concept references a semantic type and is connected by several relation types to other UMLS concepts. UMLS defines two categories of relationships:

- Semantic Network relationships: the Semantic Network contains a taxonomy of approximately 50 generic relationships, including "is-a" and "part-of".
- Metathesaurus relationships: the Metathesaurus includes a couple of relationships which are not given a specific meaning (e.g. "broader" or "narrower"). Their meaning is given by an associated Semantic Network relationship. The main goal of the Metathesaurus relationships is rather to describe the the kind of semantic link between the associated concepts (e.g. a direct link, descendant or ancestor concepts). The absence of a Semantic Network reference reduces the Metathesaurus relations to their original meaning, e.g. a relation "child" with no attribute is interpreted as "subClassOf".

Every UMLS concept is represented by an OWL class. The Metathesaurus relations "parent" and "child" are formalized as OWL "subClassOf" constraints. The "narrower" and "broader" relations , which define indirect subsumption relations, are formalized as "ancestor" and "descendant" in the OWL ontology. These relations could also be ignored, since their meaning can be inferred from the ontology using a reasoner. Due to the fuzzy definition of the rest of the Metathesaurus relations, we merged them to a single "related\_to" property. "Broken links" (i.e. relationships which are not inherited by subclasses) are not represented in the ontology.

After translating the UMLS data to OWL the ontologies were checked for consistency and the inferred classification hierarchy, which pointed out few differences compared to the original UMLS hierarchy, was analyzed. The UMLS contains several problematic modelling decisions which have been often described in research projects aiming to integrate UMLS in knowledge-based applications. Still, a comprehensive analysis of the quality of

UMLS in such a setting or especially for Semantic Web applications has not delivered an optimal solution to cope with this problem. A possible start point could be the Semantic Network, since every Metathesaurus concept is related to it. Besides, the Semantic Network is supposed to be independent of a particular area in medicine. [SKSK04] and [BO01] describe some of the deficiencies of the Semantic Network at an ontological level. [PGS98] analyzes the same issue for the UMLS Metathesaurus. At this point it is not clear how important such issues are for the quality of our retrieval system, but we intend to extend the Semantic Network with a more detailed and coherent upper level ontology.

Representing medical knowledge using Description Logics is not a trivial task [CSF03]. Although translating the UMLS data format to OWL was a straight-forward procedure, the expressivity limitations of the language become clear after a detailed analysis of the semantics of the medical knowledge. Reasoning beyond subsumption hierarchies and an extended support for concrete domains are very important for an efficient semantic retrieval system. By now we realized a first implementation of the system and are testing the appropriateness of the generated ontology. The results of the testing process will be used as input for the evolution of the ontology, both w.r.t. the representation language (do we need additional knowledge or reasoning capabilities which go beyond OWL DL) and w.r.t. UMLS and its integration in a Semantic Web application.

### **3.3 Technical Requirements for Tools**

The tools needed for large-scale real world applications require features that are currently not available or still under development. Ontology editors have to be used to generate and maintain the knowledge base. They should facilitate a intuitive and comfortable visualization of the structure and content of the ontology and allow manual or automatic insertions, deletions or updates within an ontology. For extracting knowledge from medical sources like UMLS one needs an API is needed for the automatic generation of the OWL data. Additionally the large amounts of domain-specific information (a constantly growing number of pathology reports and the complex medical ontology) should be stored persistently in a database, which not only allows a user-friendly retrieval of the semantic data, but also supports OWL reasoning. Three basic requirements for every category of tools are efficiency and scalability when dealing with large data sets, and a detailed documentation. The APIs and ontology editors should be able to process large ontologies both in main memory and persistently in a database or at least apply caching functionality. Furthermore, an API should offer comprehensive possibilities to manipulate, store and reason over ontologies by means of alternative reasoning engines. The persistent storage of the Semantic Web data should allow integrated query and reasoning capabilities for both domain ontologies and their instances.

## **4 Experiences and State of the Art**

During the development of the system we tested several ontology engineering tools with a variety of features. Though most of them are subject of research or under development, they already offer a basic environment to develop or generate ontologies for Semantic Web applications.

#### 4.1 Ontology Editors and Knowledge Acquisition

For the manual generation or modelling of ontologies the best choice is an ontology editor like Protégé [Pro], Oiled [Oil] or OntoEdit [Ont]. While OntoEdit is a commercial product, the competitors offer already open source solutions and sophisticated tools with similar features. Protégé provides an extensible plug-in architecture and supports a significant part of the OWL language. Its main advantages compared to Oiled are the flexible extendable architecture, the more comfortable user interfaces, robustness w.r.t. the size of the ontology and a persistent storage back end. Furthermore a diversity of plug-ins is available for basic data visualization (e.g. ezOWL[ezO] or OWLviz[OWLc]) and for the integration of additional rule-based knowledge. For these reasons Protégé was the ontology editor of our choice. There are still important features Protégé does not support yet, e.g. the missing support for a concurrent editing of different namespaces. We had to find workarounds to the missing features and correct the produced errors manually.

A programming interface is also essential for the automatic generation of domain ontologies, as well as for the development of ontology-based applications, which query, modify and extend the ontology. Two programming interfaces with OWL support, Jena2[Jen] and OWL API[OWLb], are currently available. Jena2 is an open source Java Framework for the development of Semantic Web applications, offering a variety of tools to engineer RDF/OWL ontologies. It covers a RDF(S)- and an OWL-API for creating and editing ontologies, a reasoning subsystem supporting RDFS and OWL-Lite, offers persistent storage in a relational data base management system and provides query capabilities using the RDQL[RDQ] query language. Jena2 is the continuation of the Jena1 project, which supported the DAML+OIL[DAM] and RDF(S) languages, though covering numerous features of OWL and generic ontology engineering. The OWL API is in an early stage of development. It offers similar basic features like Jena2, a solid integration with the RACER reasoner and an efficient internal model management for OWL ontologies. Since both implementations are still subject of further research they do not cover all features required for the realization of a Semantic Web application. Especially the lack of tests and best practices as well as a detailed documentation and tutorials makes this process inefficient. We used Jena2 because of the more advanced support of the concrete system requirements.

#### 4.2 Persistent Storage of Semantic Web Data

Depending on the size of the involved ontological knowledge the realization of a real-world Semantic Web applications is unimaginable without powerful and flexible storage back ends. In our application scenario, where the complete knowledge base is too large to be processed in the main memory of the ontology server, we need to store the data persistently in a database and retrieve only relevant fragments at execution time.

There are two kinds of knowledge within a DL-ontology: the T-Box containing the concepts and relations between them and the A-Box containing the instances/individuals of the classes. Querying and inferencing over T-Box knowledge is a different task than over A-Box knowledge, therefore both components of the ontology are stored independently. Different products are available to store and perform queries over the knowledge base. At this moment the most advanced storage systems with embedded inferencing capabilities

over T-Box Semantic Web-based knowledge are Jena2, Sesame and Protégé.

Sesame is an open source RDF-database with support for RDFS inferencing and querying. It is a system consisting of a repository, a query engine and an administration module the management of RDF(S) information, supporting expressive querying using the RQL[RQL] query language. The data is stored in a relational data base (e.g. MySQL, PostgreSQL). Comparing Jena2 and Sesame, Sesame seems to be a more consistent, solidly designed product with fewer, but more stable tools. However Jena2 supports OWL explicitly, while Sesame focuses on RDF(S). Jena2 comes with several built-in reasoners for RDF(S), which could be used in connection with the storage backend, but the OWL coverage is still under development. On the other hand it uses the DIG interface [DIG] making possible to plug in external reasoners like RACER[RAC] or FaCT[FaC], which already cover most parts of OWL DL. Currently there is a lot of research on reasoning with OWL DL (e.g. A-Box reasoning within RACER), but no available program fulfills the requirements completely. A similar situation is encountered for Protégé. Protégé is primarily an ontology editor, but it offers a comprehensive environment for the construction and management of knowledge bases. The storage back end focuses on RDF triples, which are stored directly into a relational data base. Reasoning capabilities are provided by activating the RACER reasoning engine.

Momentarily we use Sesame for ontology management and testing purposes because of performance and stability reasons. However, storing and querying OWL ontologies, especially when considering instance knowledge remains a different problem for which the appropriateness of available approaches (e.g. Instance Store [IS] which supports DL reasoning over a large numbers of individuals) remains subject of future work. A second alternative to be analyzed is the usage of deductive databases [XSB, FDES98], which offer efficient query capabilities over large amounts of individuals. The first step in this direction is the representation of the domain knowledge in a fragment of OWL DL which is compatible with a subsequent Datalog storage [GHVD03]. However this approach would also diminish the amount of knowledge representable in the knowledge base, for a domain which usually requires representation formalisms which go beyond Description Logics [CSF03, HS02].

## 5 Conclusions

Most of the OWL ontologies available online do not exceed 50 to 100 concepts. Unfortunately most of the tools have not been tested for real-world applications and are not completely compatible. They have problems when dealing with large amounts of data. Currently RACER has a limitation of about 100.000 concepts. Jena2 produces Out-Of-Memory-Errors at about 10.000 concepts with great performance loss for much fewer concepts. To work around this problem, we had to select our domain knowledge more precisely in order to keep the ontology as small as possible. This solution is not acceptable for more general solutions, e.g. if the domain would be extended to "medicine" or "pathology" instead of being limited to "lung pathology". Currently, there are intensive efforts to produce Semantic Web components and assistance tools with industrial strength,

such as inference engines that go beyond the performance of early research prototypes or methodologies for pragmatical ontology engineering. Our system will benefit from these progresses, as more powerful open source projects are available, but for the time being, the complexity of models, rules and queries triggering inferences is a critical issue.

The choice of an efficient storage infrastructure for the complete pathology knowledge base is subject of current work. Jena2 is one example offering topmost functionality, but still integration of a powerful reasoner working seamlessly together with the storage- and the query-subsystem is expected. In the same time we intend to examine the possibility of using deductive databases for this purpose while taking into account the usage of a less expressive representation language.

Nevertheless we see our project as a challenge and an experiment to realize a Semantic Web knowledge-intensive system in an area, which has been often cited as use case of Semantic Web technologies.

**Acknowledgment** The project “Semantic Web in the Pathology” is funded by the Deutsche Forschungsgemeinschaft, 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

- [BO01] Burgun, A. and O.Bodenreider: Mapping the UMLS Semantic Network into General Ontologies. In: *Proc. of the AMIA Symposium*. 2001.
- [CSF03] Ceusters, W., Smith, B., and Flanagan, J.: Ontology and medical terminology: Why description logics are not enough. In: *Proc. Towards An Electronic Patient Record, TEPR2003*. 2003.
- [DAM] DAML+OIL Reference. <http://www.daml.org/2001/03/daml+oil-index.html>.
- [DIG] DIG Interface. <http://dl-web.man.ac.uk/dig/>.
- [ezO] ezOWL Plugin for Protege. <http://iweb.etri.re.kr/ezowl/index.html>.
- [FaC] FaCT Reasoner. <http://www.cs.man.ac.uk/horrocks/FaCT/>.
- [FDES98] Fensel, D., Decker, S., Erdmann, M., and Studer, R.: Ontobroker in a nutshell. In: *European Conference on Digital Libraries*. S. 663–664. 1998.
- [Fe01] Fensel, D.: *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer Verlag. 2001.
- [Gal01] GALEN Ontology. <http://www.opengalen.org>. 2001.
- [GHVD03] Grosz, B. N., Horrocks, I., Volz, R., and Decker, S.: Description logic programs: combining logic programs with description logic. In: *Proceedings of the twelfth international conference on World Wide Web*. S. 48–57. ACM Press. 2003.
- [GPS99] Gangemi, A., Pisanelli, D. M., and Steve, G.: An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies. *Data Knowledge Engineering*. 31(2):183–220. 1999.
- [HMS96] Hahn, U., M.Romacker, and Schnattinger, K.: Automatic knowledge acquisition from medical text. In: *Proc. of the 1996 AMIA Annual Symposium*. S. 383 – 387. 1996.

- [HS02] Hahn, U. and Schulz, S.: Turning Lead into Gold? Feeding a Formal Knowledge Base with Informal Conceptual Knowledge. In: *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*. S. 182–196. Springer-Verlag. 2002.
- [IS] Instance Store. <http://instancestore.man.ac.uk/>.
- [Jen] Jena2. <http://www.hpl.hp.com/semweb/jena.htm>.
- [MeS03] Medical Subject Headings. <http://www.nlm.nih.gov/mesh/meshhome.html>. 2003.
- [NCI] NCI ontology. <http://www.daml.org/ontologies/318>.
- [Oil] OilEd Ontology Editor. <http://oiled.man.ac.uk/>.
- [Ont] OntoEdit Ontology Editor. <http://www.ontoprise.de/ontoedit.htm>.
- [OWLa] W3C OWL Reference. <http://www.w3.org/TR/owl-ref/>.
- [OWLb] OWL API. <http://owl.man.ac.uk/api.shtml>.
- [OWLc] OWLViz Plugin for Protege. <http://www.co-ode.org/downloads/OWLVizPlugin.html>.
- [PBTS04] Paslaru Bontas, E., Tolksdorf, R., and Schrader, T.: Ontology-based Knowledge Organization in a Semantic Web for Pathology. In: *to appear in J.UCS 2004, online edition*. 2004.
- [PGS98] Pisanelli, D., Gangemi, A., and Steve, G.: Ontological Analysis of the UMLS Metathesaurus. *JAMIA*. 5:810 – 814. 1998.
- [Pro] Protege. <http://protege.stanford.edu/>.
- [RAC] RACER Reasoner. <http://www.cs.concordia.ca/haarslev/racer/>.
- [RDQ] RDQL Description. <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>.
- [RQL] RQL Language Reference. <http://139.91.183.30:9090/RDF/RQL/>.
- [SH01] Schulz, S. and Hahn, U.: Medical knowledge reengineering - converting major portions of the UMLS into a terminological knowledge base. *International Journal of Medical Informatics*. 2001.
- [SKSK04] Schulze-Kremer, S., Smith, B., and Kumar, A.: Revising the UMLS Semantic Network. In: *Proc. Medinfo 2004*. 2004.
- [SNO98] Snomed International. <http://www.snomed.org>. 1998.
- [SSPB] Schlangen, D., Stede, M., and Paslaru Bontas, E.: 7feeding owl: Extracting and representing the content of pathology reports.
- [Th00] The Gene Ontology Consortium: Gene Ontology: tool for the unification of biology. *Nature Genetics*. 25:25–30. 2000.
- [TPB04] Tolksdorf, R. and Paslaru Bontas, E.: Organizing Knowledge in a Semantic Web for Pathology. In: *to appear in Proc. NetObjectDays 2004*. 2004.
- [UMLa] UMLS Metathesaurus. <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html>.
- [UMLb] UMLS Semantic Network. <http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html>.

- [UML02] Unified Medical Language System. <http://www.nlm.nih.gov/research/umls>. 2002.
- [UML03] UMLS Knowledge Source Server. <http://umlsks.nlm.nih.gov>. 2003.
- [XSB] XSB System. <http://xsb.sourceforge.net/>.