

KEMM: A Knowledge Engineering Methodology in the Medical Domain

Pinar WENNERBERG^a Sonja ZILLNER^a Manuel MÖLLER^b
Paul BUITELAAR^c Michael SINTEK^b

^a *Siemens AG, Corporate Technology, Munich, Germany*

^b *DFKI GmbH, Knowledge Management Dept., Kaiserslautern, Germany*

^c *DFKI GmbH, Language Technology Lab, Saarbrücken, Germany*

Abstract. Medical research and clinical practice deal with complex and heterogeneous data. This requires a systematic approach for semantic integration of information to support clinicians in their daily tasks.

As the clinicians speak and think in a very different language than that of the computer scientists, existing knowledge engineering approaches based on classical expert interviews fall short. Moreover, as human health is a very sensitive subject, the reuse of standardized hence reliable ontologies as medical knowledge resources becomes a key requirement.

In this paper, we first discuss the specific medical knowledge engineering requirements, we identified along a semantic medical image and text retrieval use case. Then we report on ongoing work towards establishing a corresponding methodology based on ontology reuse that is derived from the requirements. The methodology, which will be discussed in detail, relies on a novel technique for semi-automatically generating a set of potential user queries to support the knowledge elicitation process.

Keywords. knowledge engineering methodology, biomedical ontologies, reasoning, image retrieval

Introduction

Clinical care and research deal with large volumes of complex information that originates from different sources, with different structures and different semantics. By establishing an explicit formal specification of the concepts and their interrelations for a particular domain, such as medicine, ontologies (following the definition of the term “ontology” by [1]) facilitate integration and reuse of valuable knowledge across applications.

To incorporate the external medical knowledge in the ontologies and hence to semantically enhance clinical data, one has to identify the query strategies that the clinicians are interested in. As medical knowledge concerns a very sensitive context, *i. e.*, the human health, reusing standardized thus reliable medical ontologies instead of developing them from scratch is an important requirement. Consequently, it becomes necessary to decide for the appropriate, application related ontologies and their fragments (or modules).

The focus of this paper is to discuss the challenges, requirements and best practice solutions for ontology-based knowledge engineering in the medical domain. In particular, work towards establishing the KEMM (Knowledge Engineering Methodology in the Medical Domain) methodology consisting of several knowledge engineering sub-processes is introduced, whereby the focal point here is the linguistic-driven ontology engineering. The KEMM methodology is derived from the experiences gained during the realization of a clinical use case within the context of the THESEUS MEDICO project, which ¹ aims for advanced search and analysis technologies that exploit image semantics. For a more detailed description the reader is referred to [2].

The main research contributions of this paper are twofold. Firstly, the domain specific challenges and requirements for medical knowledge engineering are identified and discussed. Secondly, an initial knowledge engineering methodology is introduced that is being designed to address these requirements.

The envisioned methodology has two main objectives. Firstly, it will support the communication between the knowledge engineer and the clinical expert so that the expert's knowledge can be acquired most efficiently. Secondly, the knowledge acquired in this way will be transferred to the target software application effectively.

The KEMM methodology defines six tasks to achieve these objectives: *Query Pattern Derivation*, *Ontology Identification*, *Ontology Modularization and Pruning*, *Ontology Customization*, *Ontology Alignment*, *Reasoning-Based Ontology Enhancement* and *Ontology Testing and Deployment*. We will discuss all six steps in detail, and will demonstrate how our methodology is applied in a MEDICO specific use case, where the goal is to find all related information (images, text) about patients suffering from head and neck lymphoma.

The remainder of this paper is organized as follows. Section 1 discusses the challenges of the medical knowledge engineering process, whereas Sect. 2 presents the requirements derived to address these challenges. Section 3 introduces the actual KEMM methodology, describing all the tasks that comprise the methodology in detail. In Section 4 we compare our work with related approaches and Section 5 concludes the paper with open issues and further research.

1. Challenges of Knowledge Engineering in Medical Applications

In this paper we refer to knowledge engineering as a collection of systematic activities conducted to build knowledge-based systems as defined in [3]. Additionally, we require that the knowledge-based application incorporates ontologies, which provide the domain model, *i. e.*, they supply the background knowledge necessary to build the application.

One challenge in knowledge engineering is the so-called "*knowledge acquisition bottleneck*" meaning that the knowledge is hard to acquire as it is in the heads of the domain experts (*i. e.*, tacit knowledge) or it is distributed in different

¹This research has been supported in part by the THESEUS Program in the MEDICO project, which is funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016. The responsibility for this publication lies with the authors.

resources in different formats (*e. g.*, natural language, visual media, structured, semi-structured sources). Another challenge is the discovery of actual knowledge from the overdose of data.

We distinguish between two different types of knowledge engineering challenges; the first one concerns the communication process between the knowledge engineer and the medical expert, whereas the second is specifically about the engineering of medical ontologies.

The communication related challenges were identified during the intensive interviews with medical experts in the context of the MEDICO project. These interviews revealed the fact that each party has quite a different perception of the medical domain. From the computer scientist's perspective, the knowledge, in the way it is presented by the medical expert (and who are also always in lack of time) is often fuzzy if not contradictory or inconsistent.

The medical ontology engineering specific challenges arise due to multiple reasons. One reason is that knowledge in the medical ontologies is hidden, *i. e.*, it is opaque to the ontology engineer as he is not familiar with the terminology. Another one concerns the sheer size of the medical ontologies such as the Unified Medical Language System (UMLS) [4] containing over one million concepts or the Foundational Model of Anatomy (FMA) [5] with 75,000 concepts and over 2.1 million relationship instances about human anatomy. Thus, accessing the contents of complex ontologies of these sizes and hence assessing their appropriateness for the target application becomes an issue. Finally, modeling medical information is another major challenge as it requires accurate and comprehensive domain knowledge, which can only be acquired through intensive domain expert support².

To address these challenges, some knowledge engineering methodologies [6,7] suggest a process, where the so-called "*competency questions*" shall support the knowledge transfer from the expert to the engineer and shall help structure the engineering process. These questions, for which the future application shall be capable of providing answers, are derived from the business use cases. In the medical domain, however, the knowledge is too specific, too technical and sensitive for a knowledge engineer to be able to define such competency questions even when assisted by the clinician.

Thus, other aiding tools and methodologies become necessary for at least two purposes; (1) for supporting the communication between the expert and the engineer to acquire the expert knowledge most efficiently and (2) for engineering the knowledge acquired so that it can be incorporated into the target application most effectively.

2. Requirements Analysis

To achieve the objectives set for MEDICO in a systematic way we identified the following requirements:

²*i. e.*, only for the FMA Ontology more than 30 person years were needed so far according to the FMA FAQ, <http://sig.biostr.washington.edu/projects/fm/FAQs.html>.

Query Pattern Derivation For improved semantic medical applications, we need to identify the right level of information coverage and detail the clinicians are interested in. This type of information is typically contained in the queries the clinicians would want to ask to a clinical search engine. Yet, it is difficult to acquire it by classical interview techniques as the clinicians rely on very specific information that is difficult to talk about in a general manner. Thus, we require to establish means for semi-automatic query pattern derivation, *i. e.*, an approach for generating a set of hypothetical user queries that are subsequently evaluated by the clinicians.

Ontology Identification As human health is a sensitive matter, the quality and the quantity of the medical knowledge to be used in the target application has to be ensured by reusing the work of acknowledged authors and standardization committees that comes in form of medical ontologies. However, these ontologies are typically very comprehensive and cross-linked, which cannot be easily read, navigated or understood. Therefore, we require some automatic support for identifying the appropriate ontologies that fulfill our requirements set by the query patterns that were derived and expert validated in the previous step.

Ontology Modularization and Pruning For an effective reuse of the large medical ontologies we require (modular) ontology subsets that can be easily navigated by humans and reasoned by machines. These modules need to cover all concepts and relationships for describing the particular scenario, thus the derived set of query patterns will also determine the criteria for pruning and modularizing the ontologies that were identified as relevant.

Ontology Customization Quite often the modules extracted from the ontologies have either redundant or missing knowledge, which therefore will be post-processed, *i. e.*, customized to meet the requirements w.r.t. the application.

Ontology Alignment Each customized ontology module represents a piece of knowledge that is necessary to realize the entire application. These knowledge pieces are not arbitrary but they need to be interrelated within the context of the application. Therefore, the separate ontology modules will be integrated to deliver the whole picture.

Reasoning-Based Ontology Enhancement To discover new knowledge in form of relations and concepts, reasoning processes will need to be incorporated.

Testing and Deployment The results of each step will be tested for validity in an iterative process before the engineered model is deployed.

3. The KEMM Methodology

The methodology being developed is designed to address the two main challenges, *i. e.*, communication and medical ontology engineering. Consequently, it defines an initial task called “*query pattern derivation*” that supports the communication process between the clinical experts and the computer scientists during the interviews. This task also frees the expert from having to imagine arbitrary scenar-

ios that can be potential use cases. As described in the next section, it is based on semi-automatically deriving a set of query patterns that represent potential medical expert queries.

The succeeding tasks of the envisioned methodology are also driven by our query pattern derivation approach. Again, based on these patterns the ontologies to be reused are identified, pruned, and modularized; the relevant modules are customized and finally integrated. Once an integrated model is established in this way, reasoning processes can be applied to infer new knowledge and thus enhance the model. Parallel to this, the system is iteratively tested to guarantee consistency, clarity, coherence and validity before it is deployed. Figure 1 gives an overview of the methodology being developed that comprises the six tasks mentioned above.

3.1. Query Pattern Derivation

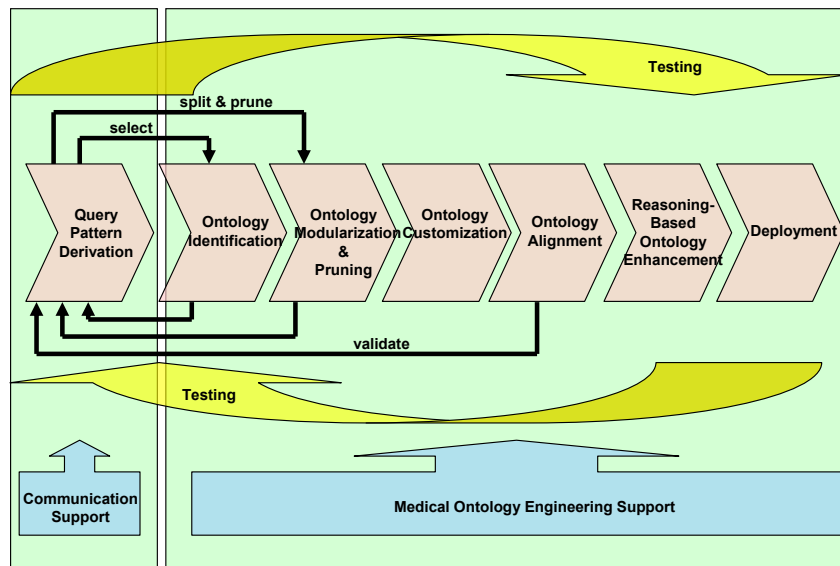


Figure 1. KEMM Workflow

To improve semantic medical text and image search, it is essential to first find out what kind of information the clinician, in our case a lymphoma expert, wants to know. Interview-based approaches such as [8] and [9] to the identification of possible user queries did not provide satisfactory results within the medical domain because of the reasons discussed earlier.

Thus, we concluded that pre-interview preparations were necessary to achieve effective results from the actual interviews on the way to discover the expectations of the clinicians from a semantic medical search engine.

We identified a set of interrelated domain terms, concepts and relations, that we considered as potentially relevant to the clinicians, which therefore can be hypothesized as query patterns, *i. e.*, typical queries that the clinicians would

want to ask. To identify the set of potential query patterns, we referred to the information readily available in ontologies, in patient health records, on the Web *etc.* The initial set of query patterns established in this way, provided the basis for a (more successful) communication with the clinicians at the same level of granularity. Additionally, it released the medical experts from the obligation of having to imagine arbitrary query use cases, thereby guiding them to concentrate on more machine-processable and technologically realizable scenarios.

Query pattern derivation is based on a combination of various techniques from natural language processing and text mining as described in detail in [10], [11]. A possible query pattern for the head and neck lymphoma is shown below, where the anatomical structure is instantiated by the head and neck, the radiology image and its modality are instantiated by the head and neck CT scan and the disease, symptom or observation is instantiated by the lymphoma (or lymph node).

```
[ANATOMICAL STRUCTURE] located_in [ANATOMICAL STRUCTURE]
                        AND
[[RADIOLOGY IMAGE] modality] is_about [ANATOMICAL STRUCTURE]
                        AND
[[RADIOLOGY IMAGE] modality] shows_symptom [DISEASE,SYMPTOM,OBSERVATION]
```

3.2. Ontology Identification

Because of the reasons discussed earlier, in MEDICO we follow the rationale of ontology (module) reuse rather than developing them from scratch. Consequently, the available medical ontologies need to be identified to incorporate the knowledge from those of high quality which have been developed over the years as a result of joint efforts of knowledge engineers and health care experts. Reusing existing medical ontologies also requires evaluating their appropriateness w.r.t. the target application, before finally the selecting them (or not).

The identification of the domain and task relevant, high quality ontologies is not a trivial task. On the contrary, it is an active research field on its own with various approaches [12,13,14] offering different solutions such as ontology search engines (*e.g.*, Swoogle³, Watson⁴), ontology libraries (*e.g.*, DAML ontology library⁵) or both (*e.g.*, OntoSelect⁶).

Within the MEDICO context, the interviews with the clinicians and radiologists showed that medical imaging and patient data need to be considered along three different perspectives; (i) the anatomical spatial perspective that addresses body parts and their locations, (ii) the radiology-specific perspective, which describes the relationships between various image modalities and anatomical regions as shown on medical images and finally (iii), the disease perspective that concerns the distinction between the normal and the abnormal imaging features.

We set these three perspectives as our search dimensions in identifying relevant ontologies and thesauri. Consequently, we manually searched in BioPortals,⁷

³<http://swoogle.umbc.edu/>

⁴http://watson.kmi.open.ac.uk/editor_plugins.html

⁵<http://www.daml.org/ontologies/>

⁶<http://olp.dfki.de/ontoselect/>

⁷<http://www.bioontology.org>

or used ontology search engines to determine a first set of potentially relevant ontologies. Upon agreement with the clinicians, from this set we decided for the FMA in anatomy, RadLex in radiology and for ICD-9 CM to represent the disease dimension.

The main criterion we set for the ontology selection process in the MEDICO scenario was the capability of the ontologies to reflect the three joint perspectives (or dimensions) that are the in focus of the radiology images and hence in that of the semantic search. Consequently, we excluded ontologies as (*e. g.*, UMLS, MeSH, GALEN *etc.*) since these were far too generic for our purposes. Further criteria we considered were:

Representation Language The Web Ontology Language OWL⁸ was a pre-defined requirement for the THESEUS-MEDICO use case, therefore we considered only those ontologies that were available in OWL format or were easily transferable.

Comprehensiveness Even though the amount of available knowledge in the medical ontologies is a challenge, we concentrated on those ontologies that are most comprehensive (*e. g.*, FMA) to avoid missing relevant important information.

Popularity Being non-experts, the popularity of a given ontology was a helpful guide for determining its appropriateness for our use case. Popularity, in our case, is defined in terms of how well-known it is, the amount of documentation available and the number of projects using the ontology.

Semantic formalism The logical formalism was also a pre-defined requirement of the THESEUS-MEDICO use case, therefore we considered only those ontologies that were formalized in Description Logics.

While we have already described the overall MEDICO ontology hierarchy (see Fig. 2) in another publication [15], Fig. 3 details the relationships between medical thesauri and ontologies in MEDICO. On the bottom of the diagram we list the different thesauri and on the top the corresponding ontologies. Thesauri are dictionaries of words but they additionally contain synonyms and antonyms. Ontologies contain abstract concepts, which are language-independent and they represent formal specifications of the represented entities. *Terms* in the thesauri are connected to concepts in the ontologies via simple `rdfs:label` or preferably via `LingInfo` relations (that allow more complex linguistic information to be attached to concepts, see [16]), which thus conceptualize these entities by using ontology mapping techniques as reported in [17].

From left to right we differentiate between different aspects—throughout this document we will also refer to them as *dimensions*—of the conceptualization of the medical domain.

The Information Element Ontology contains the information elements that we want to annotate (images, text documents, videos, ...). [16] have shown how MPEG⁹ can be used as a generalized formalism for segmentation of arbitrary document formats and annotation of segments. We will apply this generalized

⁸www.w3.org/TR/owl-features/

⁹MPEG Homepage: <http://www.chiariglione.org/mpeg/>

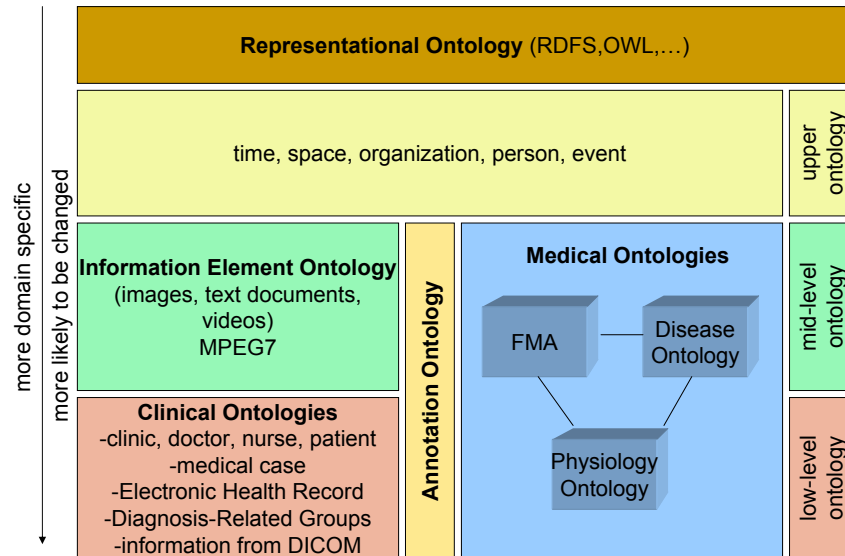


Figure 2. MEDICO Ontology Hierarchy

segmentation to treat parts of images and documents in the same way as complete documents.

The *Image Parameters* ontology contains abstract descriptions of imaging procedures (*e. g.*, the peculiarities of CT scans, of ultrasound examinations), about the image quality *etc.*

The *Annotation Ontology* includes concepts, which are used to annotate information elements with concepts from the Medical Ontologies that were detected during object recognition. Using an ontology (instead of just a simple relation) allows us to express that, *e. g.*, an image *partially* deals with a specific concept from the anatomical ontology, because only parts of it are on the picture. At least in some cases we want to annotate the relations with a probability. The Annotation Ontology allows us to express such qualifications as properties of attributes.

On the right side of the *Annotation Ontology* are the different medical ontologies and thesauri, which cover different dimensions of image annotation. Here we differentiate between *Anatomy*, *Disease* and other aspects like medical treatment and applied substances. RadLex is a thesaurus, which covers different aspects: anatomical, pathological as well as aspects about imaging parameters, quality and even applied treatments. Thus, the according parts of RadLex appear in each of these dimensions. For the disease dimension we additionally leverage on ICD9 CM. Together with the lymphoma part of NCI Thesaurus they are conceptualized to the *Disease Ontology*.

3.3. Ontology Modularization and Pruning

Ontology modularization can be addressed automatically or user-driven, but in both cases the segmentation of the ontology is a difficult task. For instance, ontology modularization approaches guaranteeing logical consistency, such as [18],

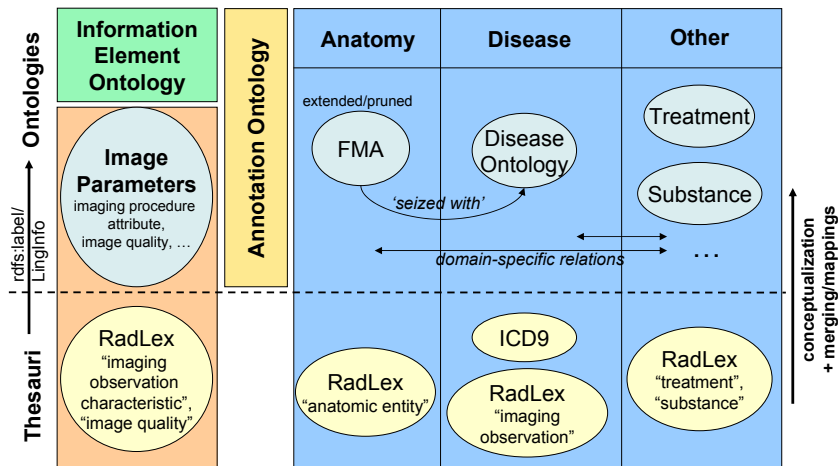


Figure 3. Relationships Between Medical Thesauri and Ontologies in MEDICO

occasionally result in too large fragments and a slow algorithm. On the other hand, graph-based approaches such as [19] perform well in terms of speed and of fragment size but do not guarantee the logical completeness. User-driven, semi-automatically created fragments established in the way as described in [20], have the advantage of covering the necessary level of detail, but they require additional user interaction and might miss relevant relational knowledge.

Within the KEMM methodology, ontology pruning can be realized based on the query patterns, as these patterns reflect the user relevant level of detail and coverage that the ontology is expected to contain. This implies that those parts of the selected ontologies that are not relevant to the initial set of query patterns need not be considered. Thus, the query patterns derived as the initial task of our methodology act as a filter in a way for determining not only the relevant ontologies but also their relevant parts which allows them to be pruned to (hopefully) useful sizes. In case query patterns are clustered (*e. g.*, based on query situations), pruning the ontologies based on these clusters will derive ontology modules (one module per cluster). These pruned (and modularized) ontologies are then represented to the clinical experts to confirm relevance and validity.

We followed this approach when determining the lymphoma use case relevant parts of the NCI thesaurus, the FMA ontology and the RadLex annotation schema. Consequently, we semi-automatically extracted the lymphoma related parts of these resources based on simple string-similarity. Assuming that anything containing the base word form *lymph* would in some way be related to lymphoma, we recorded all terms/concepts/entities that matched *lymph**. Thus, we obtained a return set with *lymphoma*, *lymph node*, *Hodgkin's lymphoma* *etc.*, which was then statistically profiled yielding a ranked term list as described in [10] and was discussed with the clinical expert.

3.4. Ontology Customization

Often the ontologies that were modularized/pruned in this way did not qualify for an immediate deployment because of several reasons. Sometimes they did not

cover the information required so that additional attributes, relations or concepts were needed. In contrast, there were also cases with redundancies or some knowledge was irrelevant. Finally, in some cases the initial ontology syntax (*i. e.*, the language) did not fit.

During our work, we came across too all three cases. As we were interested in modeling the three dimensions, *i. e.*, the joint perspectives of anatomy, disease and radiology it was important for us to establish the correspondences between the three semantic resources. As cross-references did not exist between them, we semi-automatically created relationships, where the concepts (or terms for that matter) from these resources can be referenced to each other within the lymphoma context. For example, we defined a relationship called *has_nci_code* that relates the concepts in the lymphoma module to the entities in the NCI thesaurus. Similarly, the associated RadLex terms and the FMA concepts are related to each other along *the relates_to* relationship. Finally, as the the lymphoma extract from the NCI thesaurus was in flat text format, it was semi-automatically converted to the OWL syntax.

3.5. *Ontology Alignment*

We use the term ontology alignment to imply ontology mapping and integration. Once again, since our goal was to obtain the joint view of anatomy, disease and radiology as observed on the images, separate knowledge pieces needed to be integrated to deliver the whole picture. This ontology alignment process, however, happens in our case at the ontology module level. In other words, the KEMM methodology targets integrating the customized ontology fragments or modules rather than the actual ontologies themselves because of the reasons discussed earlier.

Ontology alignment is studied under various subtopics such as *ontology merging*, *ontology mapping*, and *ontology integration* and is tackled by different linguistic, logic or graph based approaches [21].

When applying the KEMM methodology within the MEDICO context we concentrated on the linguistic approach and consequently performed ontology mapping based on string-similarity. Hence, cross-references were made between the similar concepts/terms/entities of the corresponding ontology modules by using the relationships defined in the previous ontology customization task.

3.6. *Reasoning-Based Ontology Enhancement*

The MEDICO use case is characterized by reuse and integration of distributed ontological knowledge that may introduce inconsistencies. Inconsistencies can be avoided or detected early more easily by concentrating the reasoning process on the smaller ontology modules instead of the large ontologies themselves.

With the KEMM methodology we concentrate on two specific reasoning services. Thus, in our lymphoma use case one objective is to be able to deduce the relevant image modalities (MR, CT scan *etc.*) given the symptoms of head and neck lymphoma.

Via deductive reasoning we also target the discovery of valid relationships—spatial as well as pathological and physiological— between anatomical structures.

3.7. Testing and Deployment

To avoid the propagation of inconsistencies and modeling mistakes, each and every task shall be tested for validity, completeness and coherence. The query patterns additionally need to be verified iteratively by the clinical experts.

During deployment one MEDICO specific reasoning scenario is the identification of possible diseases or symptoms given an anatomical structure and an image modality. Furthermore we can facilitate inductive reasoning to discover relations based on existing facts *e. g.*, between disease symptoms and body regions.

4. Related Work

There has been a rapid increase in ontology building, especially within the context of the Semantic Web activities initiated by various groups from industry and academia. It has been suggested that these separate activities should be executed in a systematic manner [22,23], as a result of which several ontology engineering methodologies have been proposed [7,24,6].

Some of these methodologies are the outcome of experiences collected during the ontology development process such as the Enterprise Methodology [7] or the TOVE (Toronto Virtual Enterprise) Methodology [24]. The TOVE methodology is characterized by its definition of the so-called “competency questions” that determine the scope of the ontology to be modeled and for which the future ontology shall be capable of providing answers.

Other communities concentrated on developing stand-alone ontology engineering methodologies that shall be applicable across domains and tasks. Examples of these are METHONTOLOGY [6], ON-TO-KNOWLEDGE [9] and COMMON-KADS [8]. METHONTOLOGY proposes to align ontology development with software development activities and consequently defines an ontology development life cycle process (similar to software development) that consists of several phases.

ON-TO-KNOWLEDGE, on the other hand, is a process oriented methodology that concentrates on ontology based knowledge management and maintenance in distributed enterprises.

The common view of the stand-alone methodologies is that they consider ontology development processes as equivalent to business processes and consequently they introduce higher level activities such as requirements analysis, development, evaluation, maintenance and project management. The methodology described in this paper also includes some of these activities; however, it has other components such as the query pattern derivation and ontology modularization in order to meet the specific requirements of the medical domain. More concretely, the special focus of our methodology is not on the business processes, but more on the specific medical processes such as the clinicians’ analysis and diagnosis perspectives.

The TOVE alike methodologies also share a common point with our methodology in that KEMM is also based on the experiences collected along the MEDICO project. However, it differs in the sense that it facilitates ontology reuse by emphasizing the ontology identification, modularization and customization processes.

5. Conclusions and Future Work

In this paper we reported on ongoing work about KEMM, a knowledge engineering methodology for the medical domain, which is derived from the experiences collected along the MEDICO project. KEMM addresses the two domain specific challenges, *i. e.*, communication and medical ontology engineering by defining six separate tasks, whereby the initial task of *Query Pattern Derivation* influences the other tasks as shown in the KEMM workflow.

Next, we will concentrate on deploying the reasoning mechanism over the joint disease, radiology and anatomy perspectives to discover additional knowledge, which might not have caught our attention or is simply unknown to us. In the long run, we target defining new use cases to transfer the KEMM methodology to domains other than medicine, such as engineering or law.

References

- [1] T. R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.
- [2] Manuel Möller, Michael Sintek, Paul Buitelaar, Saikat Mukherjee, Xiang Sean Zhou, and Jörg Freund. Medical image understanding through the integration of cross-modal object recognition with formal domain knowledge. *Proceedings of Healthinf 2008*, 2008.
- [3] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2003.
- [4] Oliver Bodenreider. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*, 32 (Database Issue):D267–D270, 2004.
- [5] Cornelius Rosse and José L. V. Mejino Jr. A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6):478–500, 2003.
- [6] Mariano Fernández-López, Asunción Gómez-Pérez, J. Pazos, and Alejandro Pazos. Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent Systems and Their Applications*, 14(1):37–46, 1999.
- [7] Mike Uschold. Building ontologies: Towards a unified methodology. In *16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems*, Cambridge, UK, 1996.
- [8] Guus Schreiber, Hans Akkermans, Anjo Anjewierden, Robert Dehoog, Nigel Shadbolt, Walter Vandevelde, and Bob Wielinga. *Knowledge Engineering and Management: The CommonKADS Methodology*. The MIT Press, 1999.
- [9] Y. Sure and R. Studer. To-knowledge methodology - final version, 2002.
- [10] P. Buitelaar, P. O. Wennerberg, and S. Zillner. Statistical term profiling for query pattern mining. In *Proceedings of the ACL BioNLP 2008*, June 2008.
- [11] P. O. Wennerberg, P. Buitelaar, and S. Zillner. Towards a human anatomy data set for query pattern mining based on wikipedia and domain semantic resources. In *Proc. of a workshop on Building and Evaluating Resources for Biomedical Text Mining at LREC 2008*, May 2008.
- [12] P. Buitelaar and T. Eigner. Evaluating ontology search, November 2007.
- [13] Y. Ding and D. Fensel. Ontology library systems: The key to successful ontology reuse, 2001.
- [14] Marta Sabou, Vanessa Lopez, and Enrico Motta. Ontology selection for the real semantic web: How to cover the queen's birthday dinner? In Steffen Staab and Vojtech Svátek, editors, *EKAW*, volume 4248 of *Lecture Notes in Computer Science*, pages 96–111. Springer, 2006.

- [15] Manuel Möller and Michael Sintek. A generic framework for semantic medical image retrieval. *Proceedings of the Knowledge Acquisition from Multimedia Content (KAMC) Workshop, held at 2nd international conference on Semantics And digital Media Technologies (SAMT)*, November 2007.
- [16] Massimo Romanelli, Paul Buitelaar, and Michael Sintek. Modeling linguistic facets of multimedia content for semantic annotation. In Bianca Falcidieno, Michela Spagnuolo, Yannis S. Avrithis, Ioannis Kompatsiaris, and Paul Buitelaar, editors, *SAMT*, volume 4816 of *Lecture Notes in Computer Science*, pages 240–251. Springer, 2007.
- [17] Yannis Kalfoglou and Marco Schorlemmer. Ontology mapping: The state of the art. In Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, and M. Uschold, editors, *Semantic Interoperability and Integration*, number 04391 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2005. <<http://drops.dagstuhl.de/opus/volltexte/2005/40>> [date of citation: 2005-01-01].
- [18] Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov, and Ulrike Sattler. Just the right amount: extracting modules from ontologies. In *WWW'07: Proceedings of the 16th international conference on World Wide Web*, pages 717–726, New York, NY, USA, 2007. ACM.
- [19] E. Jimenez-Ruiz, Rafael Berlanga, V. Nebot, and Ismael Sanz. Ontopath: A language for retrieving ontology fragments. In Robert Meersman and Zahir Tari, editors, *OTM Conferences (1)*, volume 4803 of *Lecture Notes in Computer Science*, pages 897–914. Springer, 2007.
- [20] *Semantic Visualization of Patient Information*, June 2008.
- [21] J. Davies. *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. John Wiley & Sons, July 2006.
- [22] J. BOUAUD, B. BACHIMONT, J. CHARLET, and P. ZWEIGENBAUM. Methodological principles for structuring an ontology, 1995.
- [23] D. Jones, T. Bench-Capon, and P. Visser. Methodologies for ontology development.
- [24] M. Gruninger and M. Fox. The design and evaluation of ontologies for enterprise engineering, 1994.