Ontologies in Medical Knowledge Representation

Alan Jovic, Marin Prcela, Dragan Gamberger
Rudjer Boskovic Institute, Laboratory of Informational Systems, e-mail: ajovic@irb.hr, mprcela@irb.hr, gamber@irb.hr

Abstract. *In this work structure of medical ontologies and their construction process are presented. For each medical domain one has to specify the scope of the ontology, acquire medical knowledge, select a tool and an ontology language, design the ontology, and present it in an appropriate way. Special attention is devoted to the problem of representing relevant medical knowledge in the form of ontology. Connection of an ontology with a rule base as a part of a decision support system is established. The ontology of the heart failure disorder, designed for the use in Heartfaid platform, is used to explain the ontology construction process.*

Keywords. medical ontology, medical knowledge, ontology design, decision support system, knowledge base

1. Introduction

Knowledge representation presents an important problem in today's science. Especially if this knowledge has to be effectively used for reasoning as a part of the decision support systems (DSS). Medical domain is characterized by the abundance of existing expert knowledge and practically each of its specializations has a constantly growing and interacting number of relevant guidelines. A long-term goal is representation of this knowledge in a form that can be used by systems supporting medical decision making. An approach is necessary that will enable systematic representation of different types of medical knowledge that could be used for various types of reasoning, ranging from off-line and on-line warning systems to planning healthcare activities.

The research presented in this work is stimulated by a project aimed at realization of a knowledge-based platform that should assist in management of heart failure patients. The platform will have to intelligently assist in a set of very different tasks ranging from monitoring patients in their home environment to decision support in specialized hospitals. In this way the platform presents an excellent example of an artificial intelligent system in a real application and presents a challenge that can influence other knowledge based systems in any domain.

The project is still a work in progress and we are not able to report on its final results yet. The first step in the preparation of the knowledge representation has been the construction of the heart failure ontology. This paper explains the technology and experience gained in creating the ontology, that may also be useful for other medical applications. The current version of the constructed heart failure ontology is available at [http://lis.irb.hr/heartfaid/ontology/].

The organization of the paper is as follows: Section 2 presents the ontology concept, structure and languages used in ontology creation as well as rules and reasoning. Section 3 covers medical ontology construction process. Section 4 describes a sample ontology working environment in a medical decision support system.

2. Ontology concepts

Medical ontology is a model of the knowledge from a clinical domain such as the heart failure syndrome. It contains all of the relevant concepts related to the diagnostics, treatment, clinical procedures and patient data. Ontologies are designed in a way that allows knowledge inference and reasoning. They are different from terminologies which are static structures used for knowledge reference. Terminological databases can be categorized based on their basic organization unit from a linguistic point of view. There are two types of terminologies: 1) Headword with its synonyms; 2) A concept with its different wordings. When available, terminologies are an excellent starting point for the ontology construction. An example of a terminology is UMLS (Unified Medical Language System) [UMLS Knowledge Source Server, http://umlsks.nlm.nih.gov]. It contains many clinical terms and integrates about 100 different vocabularies (thesauri). It is currently considered a major reference for medical terms.

In the design of the heart failure ontology we have started from the terms defined in the "Guidelines for diagnosis and treatment of the chronic heart failure" prepared by the European Society of Cardiology [http://www.escardio.org]. In order to connect the ontology concepts with the terms defined in UMLS we have introduced the property that gives the appropriate UMLS reference for every concept in the ontology. In rare cases when the terms used in the guidelines are synonyms of the UMLS terms, we have constructed a special class of UMLS synonyms inside the ontology that is connected with the original guideline term through "UMLS synonym" property.

An important example of a medical ontology is GALEN. GALEN ontology is an axiomatized taxonomy [1] and has been in development since the early 1990's. Today it is available as a commercial product. It is a huge ontology containing several thousands of concepts and procedures obtained from all the fields of medical specialization [8]. While GALEN contains many medical concepts, it is not specialized for the heart failure domain. Some specific disorders and procedures are not included in GALEN. GALEN also includes very few synonyms relevant to heart failure. Therefore, GALEN is not used in the heart failure ontology, but it should be consulted in a case of any larger medical ontology.

Up until recently, only meta-physical aspect of ontology has been discussed, i.e. the nature of the reality itself. For artificial intelligence applications, the whole ontology concept has been downgraded to a simple model of reality, so called well-behaved reality surrogate [http://ontology.buffalo.edu/ontology(PIC).pdf]. Ontology is used only to model the reality and not to explain it, because it is considered out of its scope. Ontology is based on our knowledge and beliefs and not on the objects in reality themselves [2].

2.1. Ontology structure

An ontology consists of classes, properties or slots, relationships between classes and individuals. An example of a medical ontology class is «Medication». It is the superclass of all the other medication types. Classes represent a specific clinical concept within the model. A class can be more general (upper class) or more specific (subclass), e.g. a specific class of «Medication» is «Anti_thrombotic_medication». An ontology always has a most general class. By convention introduced in [2], this class is called «Thing».

There is no strict and unambiguous way in which medical knowledge must be represented. For instance, class «Diagnosis» can be put on the first level of the heart failure hierarchy; however, in the heart failure domain it has been modeled as a subclass of «Patient_characteristics», on the second level. The difference between placing some concept on a higher or lower level is a question of semantics and is based on our understanding of the domain. For example, does one consider «Diagnosis» as a separate entity (clinical concept) or is it a characteristic of a patient?

Properties of classes and the relationships between classes are closely related. For example, the property of the class «Patients» is «testTaken». It is also a relation from the class of patients to the class of tests. Other examples of implicit relations are *is-a* and *part-of* relations. They always exist between classes and their subclasses. Classes can contain individual objects called instances or individuals. An example is given in Fig 1. «Aspirin» is an individual of the class «Anti-thrombotic_medication». Individuals are specific objects that represent the class and inherit its properties. Ontology classes in a hierarchy also inherit properties from their upper classes. For example, if the class «Medication» has property «indDiagnosis» (meaning «indicatedDiagnosis») and «maxDailyDose», then «Anti_thrombotic_medication» will also have these properties and so will its individual «Aspirin». Also, «Treatment» is obviously the upper class of «Medication». «Treatment» does not have property «maxDailyDose», but it does have property «indDiagnosis». However, «Anti_thrombotic_medication» might also have a property «targetDose» which «Medication» does not have. Individuals such as «Aspirin» cannot have additional properties other than those provided by their class.

2.2. OWL and Frames

One of the most used standard ontology languages today is OWL (Ontology Web Language). It is written in XML format and is considered a semantic upgrade of RDF (Resource Description Framework). RDF is an XML-based framework for describing information on the web. There are three types of

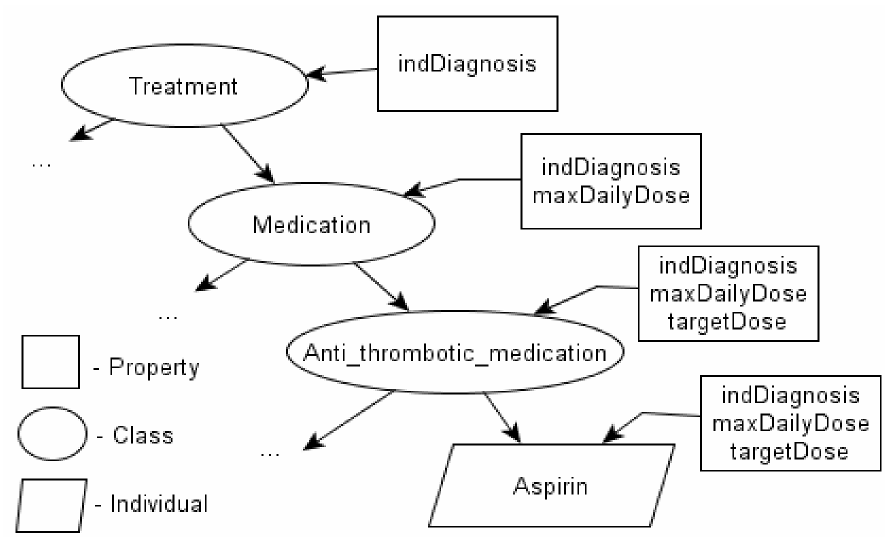

Figure 1. The inherence of properties from classes to subclasses and instances

OWL sublanguages differentiated by their expressiveness. These are called OWL Lite, OWL DL and OWL Full. OWL contains classes, properties, relations and individuals and it allows a reasoner-based inference.

Probably the most well known open, but not standard ontology language is Frames. It was developed by Minsky in 1975 [6]. Although it is an old framework, Frames is still used worldwide for knowledge representation. A Frames-based ontology is usually written using RDF files. Frames representation uses classes, slots and individuals. A slot can be considered very similar to a property. It differs from an OWL property, because a slot cannot be stated as reflexive or transitive and it cannot include logical operations. There are several other differences between OWL and Frames. Frames representation uses Unique Name Assumption (UNA), which means that different names mean different things by default, while OWL can use different names for the same thing. The other difference is the closed world assumption of Frames. This means that if nothing can be inferred about a statement, then that statement is false. In OWL, nothing is assumed if nothing can be inferred (open world assumption) [10].

2.3. Ontologies and rule-based languages

Rules present the form that can be effectively used in order to present actionable medical knowledge. It is a good practice to construct rules from concepts included in the ontology. In this way ontology design is the first and necessary step in the actionable knowledge construction process. The rules can be used for reasoning as a part of a stand-alone expert system or they may be used together with the concepts presented in the ontology. The latter approach has been experimentally designed for the heart failure knowledge based platform. Rule-based engine is a part of a DSS (section 4).

SWRL (Semantic Web Rule Language) is a rule language that expands OWL with the rule-based inference [3]. It presents a simple and natural way to add actionable knowledge to the domain concept knowledge. Another option is RuleML [5]. It is a very expressive language with strict syntax. Its logic is based on Prolog, but it also allows Horn clauses as well as Herbrand terms. Due to its complex syntax, it is not used as often as SWRL, however its expressiveness is its main advantage.

2.4. Reasoning and rules

The authors have been developing the heart failure ontology using the Protégé graphical tool in both Frames and OWL languages. A Frames version of an ontology should be built first because of its simplicity. If reasoning is required then a transformation from Frames to OWL format can be easily achieved using a built-in Protégé tool. Reasoning can only be pursued on OWL ontologies and not on Frames ontologies. Reasoning using known logic reasoners such as Pellet allows inference upon classes and their properties. If a Frames representation is used in a system, then rule-based engines are necessary. In such a case, JESS (Java Expert System Shell) [http://herzberg.ca.sandia.gov/jess/] rules can be written and when they are triggered, they would modify the Frames ontology. The problem with JESS rules is a relative complexity of the syntax. JESS rules were originally developed for expert systems, so they are not specialized in dealing with specific ontology problems. Although JESS can be used by both Frames and OWL languages, OWL can benefit from SWRL which has easier syntax. Frames representation is limited to the use of JESS rules.

A problem arises when reasoning with classes in OWL. Since OWL is used under the open world assumption, the negation of classes can not be inferred if the class itself is not inferred true. Basically, this is the only reason why rules written in JESS should be used instead of the reasoning on OWL classes or the appliance of SWRL rules.

3. Medical ontology creation process

It was mentioned that there exists no single protocol on how to construct a medical ontology or any other type of ontology. An ontology can be constructed manually [4] or (semi)automatically [7]. Manual extraction has been done for the heart failure ontology. In any case, a person who constructs the ontology needs to have some experience in ontology construction and some knowledge of the domain. Usually, domain experts are consulted to explain the meaning of domain-specific concepts. The process of ontology construction can be divided into several steps.

3.1. Scope and sources

Constructing ontologies usually starts with the specification of the desired area of reasoning, especially determining the model boundaries and the level of detail. There is an option to use already existing ontologies or some of their parts in the designing process. Which parts of the existing ontologies are used depends on the domain and application. After the ontology has been finished, it becomes possible to import it into some previously constructed ontology of a higher generalization level as well as to reuse it later in a similar domain. This is the preferred way to achieve cooperation with existing knowledge models.

When one has to construct a higher-level ontology, then one also has to use concepts that are more abstract. In this case, many higher-level classes would have to be only abstract, thus containing no individuals. These classes would create a framework for other, more specific classes to fit in. Discerning relevant from irrelevant concepts should be pursued. This will determine the level of detail that the ontology models.

In addition to scope, it is important to determine the sources of medical information. The most common case in building an ontology is to base the ontology vocabulary on related medical guidelines. This means that all the relevant data from the guidelines has to be represented in a systematic way using a hierarchy of concepts and relations. Other sources of medical knowledge include medical articles, other medical ontologies or terminologies and most importantly, experts' knowledge. The manual extraction of facts and terms by human reading from sources of medical knowledge is a reliable method when one has to construct ontologies for decision support tasks.

3.2 Tools and languages

After determining the knowledge sources, the next step is to decide which tool and language will be used in order to design the ontology. The choice of the language is usually between Frames and OWL, although other open ontology languages like DAML+OIL can be used. If reasoning and web presentation should be supported and the open-world is assumed, then OWL is the best choice. If the purpose is only knowledge sharing and terminology/taxonomy, while the closed world assumption is required, then Frames ontology is both sufficient and adequate [10].

The choice between ontology representation tools is another matter. There is always an option of constructing the ontology by directly writing an OWL/RDF file. However, this approach is not practical and requires in-depth understanding of both OWL and RDF syntax and semantics. Graphical tools for the ontology development such as Protégé, SWOOP and many others are freely available. It is the opinion of the authors that Protégé is one of the best choices for a free software ontology development platform. SWOOP is practical when one wants to consult the existing ontologies on the web and compare them or use them as a reference.

3.3. Ontology design

After a language and a tool have been selected, the process of designing the ontology begins. Essentially, there are two standard approaches to the ontology design. First one is that smaller parts of the ontology are constructed first and then later integrated to form the ontology using higher-level abstract classes. This is the bottom-up approach that is not used often in medical applications, but can be used in, for example, chemical engineering [9]. The other way is to principally design the upper classes (i.e. the skeleton of an ontology) and then develop small pieces of the hierarchy, so called top-down approach. This is used for large medical ontologies as well as terminologies [8]. Though, probably the best way of creating an ontology is to combine both approaches in an iterative way. It is recommendable to begin the process by creating classes first, then add properties or slots and finally conclude with individuals.

It is noteworthy to mention that there exist some regularities concerning the ontology classes, which the ontology creator should bear in mind. First, the concept from which a class is named should be known and already described in some terminology. This is particularly true for the smaller scale, lower-level classes. For instance, «Hypertension» is a class that exists in most of the medical terminologies and signifies a disorder of high blood pressure. It can be further divided into two classes or individuals called «Systolic hypertension» and «Diastolic hypertension». It is prudent to give a class the most recognized name for that concept. «Hypertension» could also be named «High blood pressure», but it should not be named «An elevation of the vein pressure». Second, there should be at least one reference per class to a known medical terminology like UMLS or ICD. If the class has no references in any medical terminology, then there should at least exist a reference to a guideline page, or an article from which this concept was taken. It is possible, though, to have higher-level classes with no references, since they represent more general concepts that sometimes do not exist in the medical terminologies, like «Classification» or «Feature». This should be avoided for lower-level classes and especially for individuals.

The number of properties that a class possesses should always be kept as minimal as possible. In larger ontologies, it is usual that two or more classes use the same property. However, the semantics of this property can differ. For instance, the property «Weight» is a general property that describes a physical property of an object. When this property is used for the class «Patient» and for the class «Aldosterone_receptor_blocker» (which is a medication group), the meaning is quite different. Patient's weight is presented in kilograms. It also varies frequently in time. An Aldosterone receptor blocker's weight is the weight of a pill, given in miligrams and usually a constant value. The solution is to reorganize the property «Weight» into two properties, «PatientWeight» and «PillWeight».

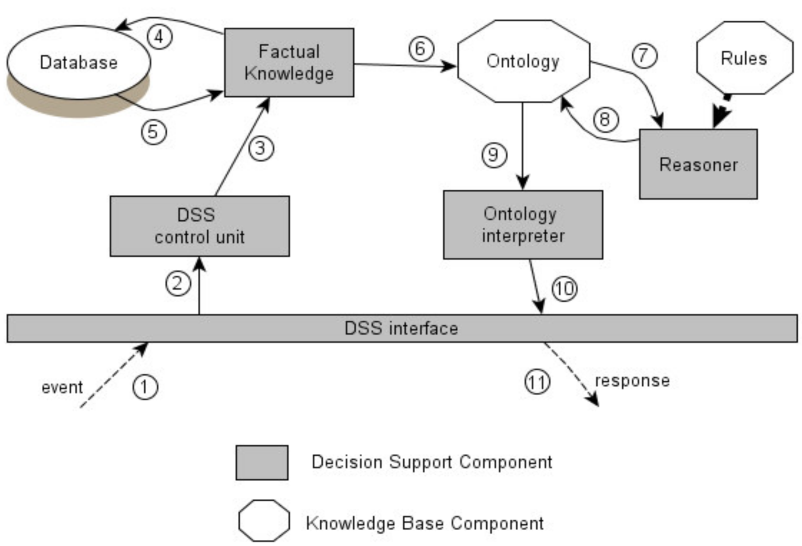

Figure 2. Ontology as a part of a decision support system

4. Ontology as a part of the reasoning process

It is important to point out that any ontology is only a knowledge base. If one wants to reason using the ontology, one has to design and implement a decision support system. An example is given in Fig. 2. This figure illustrates an example scenario in the experimental decision support system in which the ontology has a central position. The event (1) that occurred in a system is served through the DSS interface to the DSS control unit (2). The control unit initiates (3) the extraction of factual knowledge from the database (4, 5). Relevant patient data is then transformed to the ontology format (6) and prepared for reasoning (7) as a set of facts. The reasoning process is performed and conclusions reached are loaded back into the ontology (8), which is than analyzed by the ontology interpreter (9). The information acquired by the analysis is served through the DSS interface (10) back to the system user (11).

5. Discussion and conclusion

The advantages of using ontology for knowledge representation are: standardization of medical terms, knowledge sharing, and support for automatic reasoning. The contribution of the work is presentation of the construction process for medical ontologies. The lesson learned from the presented work is that OWL+SWRL is an interesting combination for reasoning in complex medical systems. The problem is the open world assumption of the OWL approach. It disables representation of actionable knowledge that requires negation-as-failure. An open research task is development of OWL interpreters that will be able to mimic closed world reasoning enabling inclusion of complete actionable knowledge into ontological form. In this way ontologies will be able to integrate descriptive, actionable, and factual knowledge.

6. Acknowledgements

This work is supported by the European Community, under the Sixth Framework Programme, Information Society Technology – ICT for Health, within the STREP project "HEARTFAID: a Knowledge based Platform of Services for supporting Medical-Clinical Management of the Heart Failure within the Elderly Population" and by Croatian Ministry of Science, Education and Sport project "Machine Learning Algorithms and Applications".

7. References

[1] Gangemi, A., Pisanelli, D., Steve, D., "Ontology Integration: Experiences with Medical Terminologies", *Proceedings of Formal Ontology in Medical Systems, (FOIS),* 1998

[2] Gruber, T. R., "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition, 5(2), 199-220,* 1993

[3] Horrocks, I., Patel-Schneider, P.F., Boley. H., Tabet, S., Grosof, B., Dean, M., "SWRL: A semantic Web Rule Language Combining OWL and RuleML", W3C Member Submission, 2004

[4] Hyvönen, E., Saarela, S.,Viljanen, K., "Application of Ontology Techniques to View-Based Semantic Search and Browsing", *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium (ESWS),* 2004

[5] Lee, J.K., Sohn, M.M., "The eXtensible Rule Markup Language", *Communications of the ACM, 46:5, 59-64*, 2003

[6] Minsky, M., "A framework for representing knowledge", *Psychology of Computer Vision, editors P. Winston, pp. 211-281*, McGraw-Hill, New York, USA, 1975

[7] Ohno-Machado, L., Gennari, J.H., Murphy, S.N., Jain, N.L. et al., "The GuideLine Interchange Format", *Journal of the American Medical Informatics Association, 5:357-372*, 1998

[8] Rector, A., Rogers, J., Pole P., "The GALEN high level ontology", *Fourteenth International Congress of the European Federation for Medical Informatics, MIE-96*, Copenhagen, Denmark, 1996

[9] Van der Vet, P.E., Mars, N.J.I., "Bottom-up construction of ontologies: the case of an ontology of pure substances", *Memoranda Informatica 95-35*, 1995

[10] Wang, H., Rector, A., Drummond, N., Horridge M. et al., "Frames and OWL Side by Side", *9th International Protégé Conference*, Stanford, California, USA, 2006