APPLYING TERMINOLOGICAL METHODS AND DESCRIPTION LOGIC FOR CREATING AND IMPLEMENTING AN ONTOLOGY ON INHIBITION

Sine Zambach
CBIT, Roskilde University, Universitetsvej 1, Roskilde, Denmark

Bodil Nistrup Madsen
ISV, Copenhagen Business School, Dalgas Have 15, Frederiksberg, Denmark

Keywords: Ontology modeling, Formal ontologies, Terminological methods, Description Logic.

Abstract: By applying formal terminological methods to model an ontology within the domain of enzyme inhibition, we aim to clarify concepts and to obtain consistency. Additionally, we propose a procedure for implementing this ontology in OWL with the aim of obtaining a strict structure which can form the basis for reasoning and further processing, and we compare a semi-formal terminological concept modeling approach with a formal Description Logic approach in OWL-DL.

1 INTRODUCTION

Much salient work is put into formalizing biomedical ontologies using Description Logic, usually with the purpose of checking consistency, cf. for example SNOMED CT (Sterns et al., 2001). Description Logic allows for a formal description via a wide range of roles, classes and instances, and it has the possibility of expressing a number of logical descriptions related to each class (Baader et al., 2003). However, this possibility of DL can be inconvenient when a minimization of the number of conditions describing each concept is desired, and we therefore argue that some modeling restrictions could be useful.

Terminological concept modeling uses delimiting characteristics to clarify how subordinate concepts of the same superordinate concept differ from each other (ISO 704, 2000), in this way making it possible to write consistent definitions, consisting of a reference to the superordinate concept, *genus proximum*, followed by one delimiting characteristic.

In 1993, Gruber defined a framework of ontological commitments (Gruber, 1993). Later on, in 1997, the Methontology modeling method was developed which provided a general guidance to ontology construction (Lopez et al., 1997). In our presentation, we will discuss these methods compared to the methods of terminological concept modeling and the methods we propose here.

To construct a formal ontology in the domain of enzyme chemistry, we take as point of departure an ontology in the biochemical subarea "enzyme inhibition" created by means of a semi-formal method (Damhus et al., 2009). The ontology is intended to be used by subject field specialists for the purpose of concept clarification. The long-term goal is to integrate the methodology and the resulting ontologies and descriptions in the standards of IUPAC (International Union of Pure and Applied Chemistry), c.f. (McNaught and Wilkinson, 1997). This ontology, however, is inconsistent and does not adhere to the terminological principles that have been defined by Madsen et al. (2004, 2005).

Therefore we have constructed a new version of the ontology which we implement in Protégé OWL-DL. We apply the terminology of terminological concept modeling when we describe the principles of terminological ontologies and the OWL terminology when we describe the OWL implementation (Horridge et al., 2004). It should be noted that terminological concept modeling differs substantially from Object Oriented Modeling (ORM) that is used for conceptual *data* modeling.

Zambach S. and Nistrup Madsen B. (2009).
APPLYING TERMINOLOGICAL METHODS AND DESCRIPTION LOGIC FOR CREATING AND IMPLEMENTING AN ONTOLOGY ON INHIBITION.
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pages 452-455
DOI: 10.5220/0002304904520455

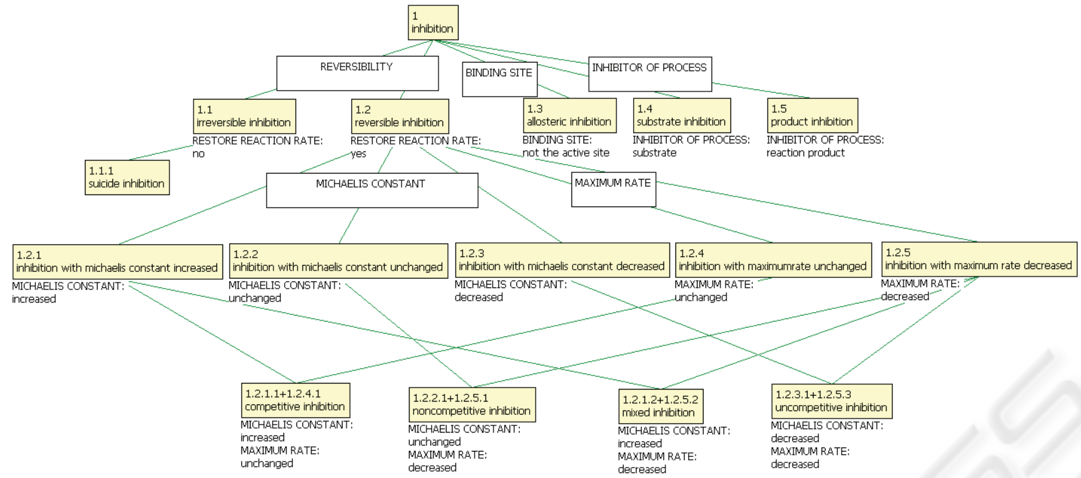

Figure 2: The diagram inhibition with subdivision criteria and an artificial layer of concepts.

2 TERMINOLOGICAL MODELING METHOD

A terminological ontology is a domain-specific ontology; cf., for example the categorisation of ontologies by Guarino (1998). We use the term *terminological ontology* as a synonym for the term *concept system*, which is normally used in terminology work, e.g. (ISO 704, 2000). The principles of terminological ontologies are based on principles that have been used for many years in terminology work, cf. e.g. Madsen et al. (2004, 2005).

In terminological ontologies, nodes are referred to as *concepts* which are defined by means of *concept relations* and *characteristics* that denote properties of individual referents belonging to the extension of a concept. In Figure 2 a new version of the terminological ontology, which was constructed in the above-mentioned pilot project on ontologies within the enzyme chemistry domain, is presented. In terminology work, all kinds of concept relations are used: type relations (ISA relations), part-whole-relations and associative relations, such as causal relations. All relations in Figures 1 and 2 are type relations. The full ontology also comprises part-whole-relations and associative relations.

The characteristics of the concepts are presented as *feature specifications* in the form of attribute-value pairs (Carpenter, 1992), e.g. *MICHAELIS CONSTANT: increased*. On the basis of these feature specifications, *subdivision criteria* are introduced which provide a good overview and help the terminologist in writing consistent definitions of coordinate concepts. Subdivision criteria are in Figures 1 and 2 represented by means of boxes with text in capital letters.

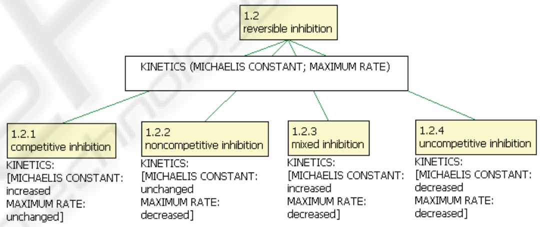

Figure 1: Early version of the diagram Inhibition from the enzyme chemistry project.

According to the terminological principles, two coordinate concepts must not differ with respect to more than one characteristic, except if they belong to a polyhierarchy, where the concepts in question have two or more superordinate concepts belonging to different subdivision criteria. In this case the concept with two or more superordinate concepts is defined by means of a combination of the characteristics of the superordinate concepts.

The first version of the ontology did not adhere to this principle. Figure 1 is a part of this ontology. In the ontology in Figure 2, we have therefore introduced a layer of extra concepts: three concepts that differ with respect to *Michaelis constant* and two concepts that differ with respect to *Maximum rate*. These concepts are "artificial" and not important in concept clarification. However, if we want to adhere to the principle of terminological ontologies for formalizing the ontology with a view to consistency checking, this layer of concepts is important.

3 IMPLEMENTATION IN OWL

The ontology of figure 2 is implemented in OWL-DL using Protégé 3.4 (Horridge, 2004), c.f. figure 3. The OWL-file can be found at the website: ruc.dk/~sz/Inhibition09.owl.

We use OWL DL for its possibility of a fine grained property structure using e.g. the "hasValue" operator for datatype properties and the possibility of more functions in later extensions.

For simplicity, we operate with two kinds of OWL-properties in order to represent concept relations, and feature specifications, as mentioned in section 2.

Type relations and part-whole relations have an obvious formalization in OWL as ISA relations among classes and the so called object properties, respectively.

In addition to these we need to decide which type of property to use for the implementation of the feature specifications. In the present implementation, the features themselves are the data literals "strings of characters" that are inherited throughout the ontology. Therefore we have chosen datatype properties to formalize the feature specifications to avoid introducing all the values of the feature specifications as classes.

As an example, see the string "Substrate" in *SubstrateInhibition* in figure 3: The class *SubstrateInhibition* has the value "Substrate" for the datatype property: *hasInhibitorOfProcess*. This property is inherited through the type relations and every class has exactly one value for each property.

Figure 3: Conditions for the concept substrate Inhibition in OWL-DL.

Any feature specification can be represented as a relation between two concepts, and a concept relation can be represented as a feature specification.

Therefore we could have considered using object properties instead, having the possibilities of creating transitive and symmetric relations. The full ontology does include such relations, namely part-of and has-part, which can be transitive. Data properties are only inherited down in the hierarchy.

The principle of working with only one delimiting feature specification per concept becomes feasible in the formal modeling procedure. Siblings are all separated by characteristics, represented by feature specifications. This supports a consistent ontology with a minimum of logical operators for each predicate since each concept can be described by its inherited characteristics and one "necessary and sufficient" description.

This is in line with the suggestion of Minimal ontological commitment (Gruber, 1993).

4 MODELING PROCEDURE

We suggest that the ontology modeling procedure is implemented as an iterative process.

We present examples of the steps that were used for constructing the Inhibition ontology. If this procedure is followed, the resulting ontology will have a minimum of necessary and sufficient conditions. It will consist of defined classes rather than primitives.

4.1 Terminology Modeling Overview

Below we describe the methods used to construct formal terminological ontologies.

a. Find sibling concepts related to one superordinate concept.
b. Identify the characteristics of the concepts.
c. Can the sibling concepts be separated by one characteristic? If yes, introduce an attribute-value pair on each concept.
d. Group the siblings by means of one or more subdivision criteria.
e. If step *c-d* are not possible and there is a need for more delimiting characteristics on each concept, introduce an extra layer of concepts so that the sibling concepts form part of a polyhierarchy, i.e. inherit characteristics from two (or more) superordinate concepts belonging to two (or more) different subdivision criteria.
f. Define the concepts as classes in e.g. OWL-DL.
g. Define the delimiting features of the sibling concepts by means of the logical equivalence operator. If a polyhierarchy is present, the super classes are added as equivalents.

5 DISCUSSION

The resulting ontology of our modeling procedure as described in section 4, will, as already mentioned, be in line with the design criteria related to the *ontological commitments* suggested by Gruber (1993).

Minimal ontological commitment corresponds to the results of our procedure that lead to the use of only one operator for each necessary and sufficient condition.

Clarity is achieved by formulating statements in a logical axiomatic form. On the other hand, we loose some readability using Protégé since words and relations are not formulated in natural language. A more appropriate and "clear" way of designating concepts and relations is used in terminological concept modeling. Also the visualization of characteristics and subdivision criteria is very clear and user friendly in terminological ontologies like the one in figure 2.

Coherence is achieved by using the reasoning function in Protegé and this application also facilitates extendability, since other specialists are able to extend the ontology by using the same software and the same method to add new concepts.

It may be argued that the encoding in OWL to some degree suffers from encoding bias. Although the software generally supports the functionality we require, a possibility of translating the descriptions to something more natural language-like would be appropriate for non-experts.

We propose that the modeling procedure as described in section 4 should be studied and tested in development of ontology modeling methodologies such as Methontology (Lopez, 1997). The terminological modeling method, described here, may fine grain the methods of Methontology, especially in the process of conceptualization, formalization and implementation.

6 CONCLUSIONS

In this paper, we have presented some central principles of terminological concept modeling, applied to an ontology within the subject area of enzyme chemistry. We have implemented this ontology in OWL-DL by means of Protégé, and may conclude that it is possible to implement the basic features of terminological ontology modeling (characteristics and concept relations) in OWL-DL, and in this way it will be possible to check consistency by using a reasoning function in Protégé.

Also we may conclude that the visualization functionality of Protégé does not yet support the presentation of characteristics and subdivision criteria in the same way as they are used in terminological ontologies.

REFERENCES

Baader, F., Calvanese, D., McGuiness, D.L., Nardi, D. And Patel-Schneider, P.F. editors, 2003. *The Description Logic Handbook. Theory, Implementation and Applications*. Cambridge, UK.: Cambridge University press.

Carpenter, B., 1992. *The Logic of Typed Feature Structures*. Cambridge, Mass.: Cambridge University Press

Gruber, T.R., 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *In Formal Ontology in Conceptual Analysis and Knowledge,* Kluwer Academic Publishers.

Damhus, T., Olesen Larsen, P. Madsen, B.N. and Zambach, S., 2009. How to work systematically towards a consistent and codified chemical terminology – a pilot study. To be published in *Chemistry International*, July 2009.

Guarino, N., 1998. Formal Ontology and Information Systems. In: *Formal Ontology in Information Systems, Proceedings of the First International Conference ,* June 6-8, Trento, Italy, 3-15. Amsterdam: IOS Press.

Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C., 2004. *A Practical Guide To Building OWL Ontologies Using The Protegé-OWL Plugin and CO-ODE Tools, Edition 1.0*, August 27, University of Manchester, pp. 1-99.

ISO 704, 2000. *Terminology work — Principles and methods*. Genève: ISO.

Lopez, M.F., Gomez-Perez, A., Juristo, N., 1997. In *Proc. AAAI Spring Symp. Series*, AAAI Press, Menlo Park, pp. 33-40.

Madsen, B.N., Thomsen, H.E., and Vikner, C., 2004. Principles of a system for terminological concept modelling. In*: Proceedings of the 4th International Conference on Language Resources and Evaluation*, Vol. I. Lisbon: 15-18.

Madsen, B.N., Thomsen, H.E., and Vikner, C., 2005. Multidimensionality in terminological concept modelling. In: *Terminology and Content Development, TKE 2005, 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen: 161-173.

McNaught, A.D. and Wilkinson, A., 1997. IUPAC. *Compendium of Chemical Terminology, 2nd ed.* Blackwell Scientific Publications, Oxford. XML on-line corrected version: http://goldbook.iupac.org (2006-)

Stearns, M.Q., Price, C., K.A. Spackman, and A.Y. Wang, 2001. SNOMED clinical terms: overview of the development process and project status, in *Proc AMIA Symp*, pp. 662–666.