Generating a Disease Ontology using specialization and combinatory restriction rules

Ihssen Belhadj[a,b], Christian Jacquelinet[a,b]

[a] *Laboratoire d'informatique médicale et de bioinformatique, Faculté de médecine de l'université Paris 13, Bobigny, France*
[b] *Agence de la biomédecine, Saint-Denis La Plaine, France*

Abstract

This paper reports on the design of a method to generate a wide coverage disease ontology starting from a root concept: here Disease and using its related semantic structure we refer to as its description knowledge model as a seed to generate Disease subtypes by a stepwise specialization and controlled combination of the knowledge model's constituents organized within a primitive pre-existing domain ontology. The core algorithm was developed in Java using Eclipse. The generating module was related to Protégé as the support for the initial and generated ontologies and a connection to the DL toolbox. The generating module was also linked to ConExp used for the examination and validation of the generated ontology with the FCA toolbox. Experimental results showed us that the explicit statement of Principles of opposition of siblings in the primitive ontology provides an efficient mean to restrict the combinatory explosion and to enhance the human readability of generated hierarchy.

Keywords:

Ontology generation, Integration of terminologies, Semantic interoperability

Introduction

The aim of this article is to examine the feasibility of an automatic disease ontology generation by systematic specialization of concepts from a given knowledge model and according to the hierarchical structure of the attributes.

Materials and Methods

To generate concepts from a starting root concept, we use: (1) the description knowledge model (DM) related to the root concept defined with Conceptual graphs formalism, (2) a primitive pre-existing domain ontology that supports the constituents of the seed DM and all their subtypes which was defined with Protégé, (3) a set of specialization, differentiation and combining principles guiding the generation process. We implemented the combination algorithm using an eclipse Java platform.

We experimented our algorithm with 3 different combinatory rules: Combinatory rule 1 assumes that all the subtypes of C1 can be combined with the subtypes of C2 with no restriction. Combinatory rule 2 restricts the generation of concept to those combining terminal specializations of DM attributes. More interestingly, we introduced the possibility to add an exclusivity property to certain subtypes of the attributes related to the DM. The combinatory restriction rule 3 imposes that a specialization of the DM defining a generated concept comprises a maximum of 1 exclusive differentiating attribute, all other combined attributes be non exclusive specializations.

Results

A first test permitted to generate a total of 157 concepts starting from a primitive conceptual hierarchy of 21 concepts and a DDM including three attributes. A second test aimed to observe the combinatory explosion while increasing each time a primitive more detailed conceptual hierarchy. The last one comprises 110 concepts and generated ontology of 12000 concepts according to a DM including four attributes.

A third test consisted of repeating the previous test after adding clauses referring to restriction rules defined in the method above. We can say, then, that combinatory restrictions rules reduced considerably the combinatory explosion using three Primitive models and permits to include siblings opposition principles within the concept hierarchy.

Discussion – Conclusion

We implemented a method of automatic generation of diseases definitions ontology based on a systematic combination of attributes and exclusively by a top-down approach. The implementation of this method showed that it is possible to create a multi-hierarchical concept ontology according different views from a starting domain concepts model and an ontology of primitive domain concepts.