User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-806-9-714

Creating an Ontology Driven Rules Base for an Expert System for Medical Diagnosis

Valérie BERTAUD GOUNOT[a,1] Valéry DONFACK[a], Jérémy LASBLEIZ[a], Annabel BOURDE[a], Régis DUVAUFERRIER[a]
[a] *Unité Inserm U936, IFR 140IFR 140, Faculté de Médecine, University of Rennes 1, France*

Abstract. Expert systems of the 1980s have failed on the difficulties of maintaining large rule bases. The current work proposes a method to achieve and maintain rule bases grounded on ontologies (like NCIT). The process described here for an expert system on plasma cell disorder encompasses extraction of a sub-ontology and automatic and comprehensive generation of production rules. The creation of rules is not based directly on classes, but on individuals (instances). Instances can be considered as prototypes of diseases formally defined by "restrictions" in the ontology. Thus, it is possible to use this process to make diagnoses of diseases. The perspectives of this work are considered: the process described with an ontology formalized in OWL1 can be extended by using an ontology in OWL2 and allow reasoning about numerical data in addition to symbolic data.

Keywords. NCI Thesaurus, OWL, SWRL, Expert Systems, Biomedical ontologies, Ontology modularization, data-element, value-set.

1. Introduction

Expert systems of the 1980s have failed to maintain large rule bases[1][2]. The aim of the current work is to show how semantic web tools can help in this area. We propose a method for (1) extracting a sub-ontology of a particular field (plasma cell neoplasms) from a medical ontology (the NCIT in OWL), and (2) automatically translating this ontology into production rules using SWRL formalism.

The goal is to enable easy building of the knowledge base of an expert system. This process based on a formal ontology allows to easily generating a large number of production rules, ensures consistency of the expert system knowledge base and thus makes it easier to understand the reasoning.

2. Material and Methods

The NCIT (v10.07) is an ontology and a terminology in the cancer domain with over 80,000 classes, 187 properties (or relations) and 57,000 restrictions. It is currently

[1] Corresponding Author: Valérie Bertaud Gounot.

available as a free way OWL 1.1 of [3]. OWL reasoners like Pellet [4] allow to check consistency of the ontology based on the formal definitions of classes. They can also classify an instance as being an instance of a specific class if the instance at least all the necessary and sufficient conditions of the class.

2.1. Reorganizing the Relationships in the NCIT

The NCIT has relationships ("Object Properties") which are not common in ontologies, such as "may_have" and "excludes". These "Object properties" link each disease to its manifestations (signs, symptoms). It's a fine representation of signs in diseases, but it doesn't allow classifying instances of diseases according to their signs. Indeed, for a given patient, the signs are "present" or "absent": the patient will or will not have a sign (relationship "has" and not "may-have" or "exclude"). This leads us to propose that the relationship "disease_has_finding" is a special case of the relation "disease_may_have_finding" as did Natalya Noy [5]. If "disease_may_have_finding" subsumes the relation "disease_has_finding", it is in accordance with the description logic and also with the reality of the domain. Moreover it enables the reasoners to classify the instances that have or don't have the sign.

2.2. Extracting a Sub-Ontology

We didn't want to work with the whole NCIT for processing time reasons.
We created a "sub-ontology extractor" able to extract a sub-ontology that is to say a subset of the NCIT (classes and their formal definitions). It takes as input parameters (1) an ontology in OWL format, (2) a list of key concepts from which the extraction shall start, (3) the directions in which the extractor shall search, ie to parents, to children and/or to connected concepts thanks to relationships ("Object Properties") and (4) the list of the relationships to be followed. We ran the extractor with PLASMA_CELL_NEOPLASM as key concept (Figure 1). The extractor retrieved all ancestors up to the top, all children down to the leaves, all target concepts connected by a relationship to PLASMA_CELL_NEOPLASM or its children. Then it retrieved parents of all these target concepts in order to link them to the root.

2.3. Developing Production Rules for Medical Decision Support

2.3.1. Logic Background

Ontology classifiers(Pellet, FACT ...) implement deductive reasoning. However, in the diagnostic process, we must propose diagnostic hypotheses in an abductive reasoning starting from an observation in which information is inherently incomplete[6].
Deductive reasoning: if a→b and if a is true, then b is true.
Abductive reasoning: if a→b and if b is true, then a is possibly true.

2.3.2. Creating the Prototypical Cases

SWRL reasons on instances. Thus it was necessary to generate an ABox (Assertional Box: all information related to specific instances of the domain) from the myeloma TBox (Terminology Box: all classes of the ontology with their formal definitions). As Protégé does not have an assistant to automatically create instances, we used the OWL

API to automatically generate an instance and define the various assertions for each of the 27 prototypical cases of plasma cell disease.

2.3.3. Creating SWRL Rules for Diagnostic Reasoning

We then created the SWRL rules used for abductive reasoning.

3. Results

3.1. Extracting a Sub-Ontology

From the NCIT ontology encompassing over 60,000 classes, 187 relations (object properties) and 57,000 limitations, we extracted automatically a sub-ontology : 281 class, 17 relations and 25 restrictions. The resulting sub-ontology completely defines the concepts of the taxonomy of the PLASMA_CELL_NEOPLASM.

3.2. Creating Production Rules for Medical Decision Support

Formulating these production rules required the use of four new data properties:

"Finding_Has_Diagnosis ", "Finding_ Excludes_Diagnosis"
"Finding_Absence_Has_Diagnosis", "Finding_Absence_Excludes_Diagnosis.

For diagnostic reasoning, four generic rules were defined. They are meant to be used to reason on the instances (prototypical cases).

(1) If a patient has a sign f and if this sign may be a manifestation of the disease d, then this disease d is a possible diagnostic:

Disease_Has_Finding (?d, ?f) ^ Finding (?f)
-> Finding_May_Have_Diagnosis (?f, ?d)

(2) If a patient has a sign f and if a disease d excludes this sign, then the disease d is not a possible diagnosis:

Disease_Excludes_Finding (?d, ?f) ^ Finding (?f)
-> Finding_Excludes_Diagnosis (?p, ?d)

(3) If a sign f is absent in the patient and if this sign f is required for the disease d, then the absence of sign f excludes the diagnosis:

Disease_Has_Finding (?d, ?f) ^ Finding (?f)
-> Finding_Absence_Excludes_Diagnosis (?p, ?d)

(4) If a sign f is absent in the patient and if a disease d excludes this sign, then the absence of sign f makes the diagnosis possible:

Disease_Excludes_Finding (?d, f) ^ Finding (?f)
-> Finding_Absence_May_Have_Diagnosis (?p, ?d)

Given the fact that we consider that "Disease_Has_Associated_Disease", "Disease_Has_Abnormal_Cell", "Disease_Has_Cytogenetic_Abnormality" (...) also express Disease-Sign relationships in our sub-ontology, 9 generic SWRL rules are needed in order to be able to drive abductive reasoning on patients and to expoit “excludes” and “may have” relationships. For the whole NCIT, we would have to write the 4 rules for 5 different types of relationships, thus 20 rules would be needed.

These production rules define the semiological relationships (relationships between diseases and their manifestations) that will allow to suggest or eliminate a diagnosis depending on the presence or absence of a sign. They follow a first order logic (with variables: “?d” and “?f”). Variables can also be automatically instantiated with all

instances (individuals) of the ontology resulting in three hundred and five production rules (0 order logic) for our subontology (Fig.1).

The system was evaluated with 10 real patient records and exit letters. An input form (http://www.med.univ-rennes1.fr/OntoDiag/) gathering all very bottom medical findings from the ontology was filled for each patient. The production rules were used to provide 2 lists of possible and excluded diagnoses. All diagnoses made by the doctors (domain experts) were in the list of the possible diagnoses. For each possible diagnosis, the number of signs present or absent compared to the number of signs described for the disease in the ontology are also displayed allowing to classify the possible diagnoses by relevance.

```
Finding_Absence_Excludes_Diagnosis (Extraosseous_Lesion, Extramedullary_Plasmacytoma)
Finding_Absence_Excludes_Diagnosis (Localized_Lesion, Extramedullary_Plasmacytoma)
Finding_Absence_May_Have_Diagnosis (Neoplastic_Plasma_Cells_Present_in_Bone_Marrow,
Extramedullary_Plasmacytoma)
Finding_Excludes_Diagnosis (Neoplastic_Plasma_Cells_Present_in_Bone_Marrow,
Extramedullary_Plasmacytoma)
Finding_May_Have_Diagnosis (Arthritis, Heavy_Chain_Deposition_Disease)
Finding_May_Have_Diagnosis (Coagulation_Disorder, Heavy_Chain_Deposition_Disease)
```

Figure 1: Example of reified production rules: the SWLR rules were instanciated with the diseases and findings defined in the ontology. For example, the first rule means "if extra-osseous lesion is absent, then extramedullary plasmocytoma diagnosis is excluded".

4. Discussion

Previous work has already demonstrated it was possible to use semi-formal knowledge bases to build expert systems [7]. Our study shows how it is possible to generate the inference rules of an expert system from an ontology written in OWL.

The topic ontology and decision support has led us to consider semiotic ontologies in which the entities were not diseases but diagnoses [8][9]. It is clear that this approach is a minority. In a classical medical ontology, diseases are entities that have manifestations (Disease_Has_Finding). The concept of diagnosis is not mentioned. We had to add the "Finding_Has_Diagnosis" relationship. It is not the inverse of the previous one, but a really new relationship. Indeed, the diagnosis is not a disease but a hypothesis of the disease.

Computer Assisted Decision Systems based on classical ontologies have rarely been proposed, however, we can highlight the work Jovic [10] who described an architecture similar to ours. One of the special features of our work is the use of abductive reasoning on an ontology. Abductive logic and ontology were already mentioned in several publications[11]. Querying an ontology is classically based on Description Logic and deductive reasonning. This deduction may either be made on the Terminological Box (TBox) or on the Assertional Box (ABox). In accordance with Description Logic, it doesn't allow to classify a class or an instance if it doesn't have every necessary and sufficient conditions of at least one class in the ontology. Production rules linking signs to diagnoses could be considered as a HBox (Hypothesis Box), which allows us to use the ontology for decision support in abductive reasoning. The abductive reasoning allows getting results (hypotheses of diseases) even if all the necessary and sufficient findings are known.

In the NCIT, a relationship is unusual in ontologies: the "Excludes" relationship. It is useful for medical diagnosis which can be based on negative signs (absent signs).

The production rules make it possible reasoning on findings known to be absent for a given patient: the fact that the sign is known to be absent either (1) has no influence on the diagnosis if the sign is not mandatory; or (2) excludes the diagnosis if the sign is mandatory in a disease, or (3) strengthens the possibility of the diagnosis if the sign is excluded for the disease. For example, it could be useful to formally define eligibility criteria of clinical trials. Some "may-have" relationships defined in the class could be transformed at the level of the instance into "has" or "exclude" relationships according the definition of the disease chosen in the clinical trial. This approach is an adaptation of a case-based reasoning (CBR) system in which the source case is modified to be in accordance with a new situation: the target case which is expressed in description logic as proposed by Cojan Lieber [12].

5. Conclusion

A major problem of expert systems was the creation and maintenance of rule bases. Driving this creation by an ontology can greatly facilitate this process. In this context, the generated rule base could be considered as the HBox of the ontology, that is to say the diagnostic hypotheses box as the Tbox is the terminology box and the ABox is the descriptions of diseases prototypes.

References

[1] Liao S. Expert system methodologies and applications—a decade review from 1995 to 2004, *Expert Systems with Applications* (2004), 1–11.
[2] Hayes-Roth B. A blackboard architecture for contro, *Artificial Intelligence* 26 (1985), 251-321.
[3] Fragoso G, de Coronado S, Haber M, Hartel F, Wright L. Overview and utilization of the nci thesaurus, *Comp Func Genomics* 5(8) (2004), 648–54.
[4] Sirin E, Parsia B, Grau B, Kalyanpur A, Katz Y. Pellet: a practical owl-dl reasoner, Web Semantics: Science, *Services and Agents on the World Wide Web* 5 (2007), 51–53.
[5] Noy N, de Coronado S, Solbrig H, Fragoso G, Hartel F, Musen M. Representing the nci thesaurus in owl dl: modeling tools help modeling languages, *Appl Ontol* 3(3) (2008 Jan 1), 173–190.
[6] Pottier P, Planchon B. *Description of the mental processes occurring during clinical reasoning*, Doi : 10.1016/j.revmed (2010), .
[7] Achour S, Dojat M, Rieux C, Bierling P, Lepage E. A umls-based knowledge acquisition tool for rule-based clinical decision support system development, *J Am Med Inform Assoc* 8(4) (2001 Jul-Aug), 351-360.
[8] Bertaud-Gounot V, Lasbleiz J, Mougin F, Marin F, Burgun A, Duvauferrier R. A unified representation of findings in clinical radiology using the umls and dicom, *Int J Med Inform* 77(9) (2008 Sept), 621-629.
[9] Bertaud-Gounot V, Belhadj I, Dameron O, et al. Computerizing the radiological sign, *J Radiol* 88 (2007 Jan), 27-37.
[10] Jovic A, Prcela M, Gamberger D. Ontologies in medical knowledge representation, Proceedings of the ITI 2007 29th *Int. Conf. on Information Technology Interfaces* June 25-28 (2007), .
[11] Elsenbroich C, Kutz O, Sattler U. A case for abductive reasoning over ontologies, *OWLED* Vol. 216CEUR-WS.org ((2006)).
[12] Cojan J, Lieber J. An algorithm for adapting cases represented in an expressive description logic, *ICCBR* (2010), 51-65.