Opposition principles and Antonyms in Medical Terminological Systems: Structuring Diseases Description with Explicit Existential Quantification

Christian Jacquelinet
Agence de la Biomédecines, 1, avenue du Stade de France, 93 212 Saint-Denis-La Plaine, FRANCE

Abstract

The Agence de la Biomédecine (BMA) maintains a national information system where the transplantation teams record patients data. In order to facilitate data recording and to support inter-operability with local and international registries, the BMA-IS comprises a terminological system structured around a medical diseases-centered ontology. BMA-IS provides tools for mapping multiple terminological standards and ad-hoc specialized thesaurus onto the diseases ontology. It also provides ontology editing functionalities: adding a type within the hierarchy by creating a new subtype from a current concept according (or not) to the current concept schema, creating conceptual relations, defining description schema, refining the schema for a given concept, adding and deleting hyponyms and hyperonyms to a given concept. This paper reports on the treatment of antonymy in diseases terminologies. It focuses on the manner we have systematically built up the diseases ontology in specifying progressively the existence then the type of refining concepts according to a global meta-schema describing diseases.

Keywords:
Medical Ontology,; Decisional Information System,; Antonym.

1. Introduction

The Agence de la Biomédecine, former Etablissement français des Greffes, maintains a national information system (BMA-IS) in which organ procurement organisations register data about donors and the transplantation teams record patient data at the registration on the waiting list, at the time of transplantation if any and during follow-up before or after transplantation [1]. BMA also supports the Renal Epidemiology and Information Network (REIN) devoted to the follow-up of all patients treated by dialysis [2], connecting the cohorts of dialysed and transplanted patients to offer a complete registry for end-stage renal diseases in France. BMA-IS is a decisional information system used to aggregate information, perform evaluation studies and propose strategic decision in the field of organ failure public health policies. It is required to integrate periodically data coming from multiple sources: patient data coming from transplant teams or from dialysis centres, information coming from organ procurement organisations when a potential donor is
identified and geographical information concerning health offer organisation. The fact is that multiple coding systems are in use, resulting in a terminological heterogeneity that must be dealt. In integrative approach based on a semantic mediator, a common domain ontology is used to map entities [4]. In the medical field, a diseases ontology has to support the integration of various medical terminologies by mapping terms to concepts. In such an approach, a term is like a message to relate to the conceptual structure supporting its meaning. Terminological mapping indeed is a first step toward medical information integration. In integrating terminologies, we observed that the negation of relational adjectives and antonyms were frequent in medicine. Description Logics as well as Conceptual Graphs formalism addresses this phenomenon using negation. With such formal approaches, a primary disease is not a secondary disease (antonym of). But when the "positive" concept is defined by a formal conceptual structure, few is said on the consequence of the negation on the definitional conceptual scheme of the antonym. We present in this paper how we address this issue with the introduction of an existential quantifier class concept within the definitional conceptual scheme, and to specify it as positive or not negative.

2. Material and Methods

2.1 The Terminological System

The foundations of our terminological system have been reported in [5]. The terminological system comprises two main components. The first component supports a medical ontology centered on diseases with relevant editing functionalities. The second component regroups terminologies to integrate and relevant mapping functionalities. Another component will only be mentioned here: it deals with the definition of information production contexts and information exploitations contexts. A use-context enables for example a given medical term to be used by heart transplant teams as complication to record patient data, whereas kidney transplant teams use it as an indication or a co-morbidity item only to browse within the hierarchy of terms. The implementation is based on n-tier architecture comprising a web server allowing the terminological system administration and consultation through the BMA portal. Business logics developed to support the functionalities described below rely on Oracle application server and data are stored in a relational Oracle database.

2.2 The ontological component

The ontology is supported by a concept type lattice with multiple inheritance. It comprises concepts related to biological and chemical substances, anatomy, pathology, physiology, microbiology, medical procedures and of course diseases. The central concepts are heart, lung, kidney and liver diseases. A concept can be related to a Description Model. A Description Model is a set of conceptual relations linking the concept to refining concepts. It can be viewed as a CG based frame-like structure. It is to be shared by many concepts as a template to guide their formal description by a conceptual structure. A Description Model behaves like a meta-model comprising all potential slots describing a class and its subclasses. Some "slots" defining the class are inheritable by hyponyms, some are local and not inheritable. A new concept is always generated from a current concept as a brother or as a hyponym of the current concept. At the generation of the new concept, the instantiation of the Description Model give birth to a Description Graph. Slots inherited of the current concept can be over-specified, some new slots will be instantiated by specifying a refining concept, some won't be instantiated. Typical local slots are the links mapping a disease concept to its related terms in various terminologies. A Description Graph associates
together for a given concept its type definition (class `slots`) and its schema (local `slots`). The description of each concept by a conceptual structure (its Description Graph) makes the ontology formal: the is_a relation is derived from the definitional conceptual structure. It is not based on type labels and their lexical meanings. The ontological component provides ontology editing functionalities (i) to generate, modify or delete a Conceptual Relation and its signatures: a given CR can have multiple signatures; (ii) to generate, modify or delete a type within the hierarchy by creating a new subtype from a current concept according (or not) to the current concept definition; (iii) to generate, modify or delete a description model by insertion of its constitutive CR and the specialization of their signatures if relevant; (iv) refining the description structure for a given concept, adding and deleting hyponyms and hyperonyms to a given concept.

### 2.3 The Terminological Component

The Terminological Component stores terms to integrate with information on the terminologies they are coming from (ICD10, ISHLT, ERA-EDTA and ELTR thesauri, the French specialized terminology for renal diseases). Our ancient thesaurus is also a terminological source to map onto the ontology: indeed, terminology mapping is a solution to the diachronic evolution of a thesaurus. Tools are provided to map terms onto the diseases ontology. A term can be labeled as ambiguous (thus related to more than one concept), implicit (its signification is more precise that its label, due to its situation in the terminology it is coming from) or pertinent (whether it relevant or not for the domain).

### 2.4 The Initial Diseases Description Model

Previous works dealing with the semantic structure of medical terms, led us to build an initial Disease Description Model comprising six main slots relating a given disease to its discriminating characteristics, thus beginning with a: has_for_discriminating completed with: _Location, _Evolution, _Semiology, _Etiological-Process, _Lesion, and _Causative_Disease. The has_for_discriminating_Causative_Disease relation enables to represent diseases that occur during pre-existing diseases, e.g., terminal renal insufficiency in chronic glomerulo-nephritis. Therefore the has_for_discriminating_Causative_Disease slot is valued by concepts that are themselves Diseases. The other slots are valued by concepts from specialized hierarchies representing anatomy, medical signs, physio-pathological process, or pathological lesion. In all the example below, local slots are in italic:

**Description Model for Disease** is

[Disease]-

(has for label) → ['Disease']
(has for discriminating location) → [Body_Component]
(has for discriminating etiological process) → [Etiological Process]
(has for discriminating evolution) → [Evolution]
(has for discriminating sign) → [Semiology]
(occurs during or after a discriminating disease) → [Disease]
(has for discriminating lesion) → [Lesion]

With such a model, it is possible to relate a "positive" disease concept to its antonym through a local (has for discriminating antonym) relation, as in wordnet for example. With the CG model, the formal type definitions CG1 of a Specified Disease can be specified by its differentiae as related to another concept. We can use a co-reference link (- - - -) and the formal definition of a negative context to define the antonym of the Specified Disease as in CG2. This definition says that the Antonym of Specified Disease is a Disease that refers to
something else that the Specified Disease.

Type definition of Specified Disease (x) is CG1

Type definition of Antonym of Specified Disease (z) is CG2

GG1: \[ \text{Disease}: *x] \rightarrow (\text{REL}) \rightarrow \{ \text{Concept} \}

GG2: \[ \text{Disease}: *z] - - - ¬\{\text{Specified Disease}\}

3. Results

3.1 Antonyms and Opposition Principles in Medical Terminologies:

The use of this model to integrate medical terms showed us that it required to be more sophisticated to explicitly express the quantification. Any disease has intrinsically an existing location, an existing evolution profile or an existing etiological process. Indeed, there is no disease without an involved body component, evolution or etiological process, even if it is unknown (cryptogenic disease) or unspecified. In contrast, the existence of a discriminating sign, causative disease or lesion is to specify: a disease can be discriminated according to its semiology, a disease can be primitive or secondary. Negation is a very common phenomenon in medical terminologies. Some terms are defined by opposition to the others (e.g.: non otherwise specified, non Hodgkin's Lymphoma, ...), some comprise the negation of relational adjectives that can be described using negative slots (e.g: non genetic hemochromatosis, non A non B hepatitis, non glomerular lesion,...). We found that the type definition of an antonym with a negative context like CG2 is equivalent to CG3, where \( \neg \) is the logical operator not and * the existential quantifier: CG3: \[ \text{Disease}\rightarrow (\text{REL})\rightarrow \{ \text{Concept} \rightarrow \neg * \}

To represent such entities, we found that an efficient solution was to render more explicit the type and the quantification. The improved Description Model guiding the generation of disease concepts can be summarized as follow:

\[
\text{[Disease]} -
\begin{align*}
(00: \text{has for label}) & \rightarrow \{\text{"disease"}\} \\
(10: \text{has a location quantified as}) & \rightarrow \{ * \} \\
(11: \text{has for location's type}) & \rightarrow \{\text{Body_Apparatus}\} \rightarrow (\text{Part_of}) \rightarrow \{\text{Human_Body}\} \\
(12: \text{has for location's type}) & \rightarrow \{\text{Organ}\} \rightarrow (\text{Part_of}) \rightarrow \{\text{Body_Apparatus}\} \\
(13: \text{has for location's type}) & \rightarrow \{\text{Body_Component}\} \rightarrow (\text{Part_of}) \rightarrow \{\text{Organ}\} \\
(20: \text{has an etiological process quantified as}) & \rightarrow \{ * \} \\
(21: \text{has for discriminating etiological process type}) & \rightarrow \{\text{Etiological Process}\} \\
(22: \text{has for etiological process agent's type}) & \rightarrow \{\text{Etiological Process Agent}\} \\
(30: \text{has an evolution quantified as}) & \rightarrow \{ * \} \\
(31: \text{has for discriminating evolution's type}) & \rightarrow \{\text{Evolution}\} \\
(40: \text{has a discriminating semiology quantified as}) & \rightarrow \{\text{Quantifier}\} \\
(41: \text{has for discriminating sign's type}) & \rightarrow \{\text{Semiology}\} \\
(50: \text{has a causative disease quantified as}) & \rightarrow \{\text{Quantifier}\} \\
(51: \text{has a discriminating causative disease whose type is}) & \rightarrow \{\text{Disease}\} \\
(60: \text{has a lesion quantified as}) & \rightarrow \{\text{Quantifier}\} \\
(61: \text{has for discriminating lesion's type}) & \rightarrow \{\text{Lesion}\} \\
(62: \text{has for discriminating lesion distribution's type}) & \rightarrow \{\text{Lesion Distribution}\}
\end{align*}
\]

4. Discussion-Conclusion

As reported in [5], the use of a knowledge extraction tool such as RIBOSOME is an important help to compare the semantic structures generated for medical terms and to propose a general schema for the description of a disease based on an experimental basis. This approach appeared to us as a starting point toward both a top-down approach because it produces relevant and robust definitional scheme - a prerequisite for a generative ontology, and toward a bottom-up approach where a classifier is used to insert concepts within the type hierarchy. The definition of diseases according to description frame like
structures with a limited and stereotyped set of discriminating nosologic slots is effective and the terminological server will be used for the REIN registry at the end of 2005. Linking medical terms to concepts described using CG based scheme permits to disambiguate and to improve explicitness and consistence of existing terms, to insert them at their right place within a hierarchy. This approach enables the integration of subsets of relevant medical terms coming from various terminologies in a structured poly-hierarchical semantic network devoted to support the domain ontology, the so-called terminology mapping. The use of precise rules to assess hyponymy permits to set up the hierarchy of terms on sound ontological foundations. Especially, we found that the systematic use of a discriminating associated disease to specify the chaining of primary, secondary and ternary diseases is very important to avoid errors in the hyper/hyponyms relations and thus helps a lot in the building of the ontology.

The completion of our initial description model with the explicit introduction of negative and positive quantifiers let us now consider that such an approach permits to systematically generate an ontology of diseases from a skeleton ontology. The notion of generative ontology is inspired by the generative grammar paradigm and provides semantic domains for a compositional ontological semantics, as reported in [7-9]. Constraints due to the disease's schema in combination with the anatomical, etiological process, semiology and lesions type hierarchy precludes artificial and excessive disease's concept generation. This procedure appears as a first step toward a generative ontology for diseases, in contrast with classifying approaches where all concepts are to be described before their automated classification. This kind of systematized diseases "periodic table" should be helpful non lonely for terminology mapping, but also to deal with the coming huge impact of genomics on medical knowledge.

5. References


Address for correspondence

Christian Jacquelinet, MD, PhD
Agence de la biomédecine,
1, avenue du Stade de France, 93 212 Saint-Denis-La Plaine, FRANCE
christian.jacquelinet@biomedecine1.fr