Automatic generation of consistency constraints for an OWL representation of the FMA

Olivier Dameron, Julie Chabalier

UPRES-EA 3888, Université de Rennes1, France
olivier.dameron@univ-rennes1.fr

1   Introduction

Ontologies are explicit and machine-processable representations of some domain knowledge. As anatomy is a fundamental discipline to the medical domain [1], an ontology of anatomy would play a pivotal role. Nowadays, the Fundamental Model of Anatomy (FMA) is the reference ontology for canonical human anatomy. It contains over 75.000 classes and 2.5 millions relations [2]. From an application perspective, ontologies contribute both to semantic interoperability of data sources [2] and to reasoning on the data [3]. In order to support these interoperability and reasoning requirements, the OWL family of languages seems to be currently the best solution. However, the FMA is represented using the frames formalism.

For the most part, generating an OWL version of the FMA is straightforward. However, previous works showed that some features are implicit or cannot be represented in frames but are crucial for leveraging the specificites of OWL [4, 5, 6, 7]. It would be possible to address this point by manual processing, but the task is likely to be cumbersome, error-prone, and would increase the workload of maintaining the original FMA. Moreover, the organization of the FMA follows a strict and principled approach that could be exploited.

This article focuses on the automatic generation of OWL-specific constraints for the conversion of the FMA into OWL. Particularly, we report on the appropriate generation of subsumption relations, of disjunction constraints among sibling classes, of necessary and sufficient definitions for coverage and of existential and closure constraints for partonomy relations. We restrict our study to the subclasses of `Physical anatomical entity` and to the relations of subsumption and composition. We show that within this domain, (1) the automatic generation of the constraints is possible, and (2) that the process can identify some situations where the ontology would be logically consistent but in contradiction with reality.

2   Methods

We first generated one OWL class for each original FMA class. The FMAID of the class is used as the `rdf:ID`, prefixed by "fma". The name of the class is added as an `rdfs:label`.

We then identified in the original FMA a set of patterns reflecting situations with an underlying modeling principle that could not be represented in frames, and is therefore missing.

We eventually wrote a set of Python scripts for detecting these patterns among the classes of the original FMA and generating the corresponding OWL constraints.

3   Taxonomy

The FMA taxonomy follows a very principled approach. We duplicated this taxonomy in OWL. However, because the original FMA only uses single inheritance and because the distinction between single and multiple inheritance is not relevant in OWL, we generated some additional taxonomic relationships.

3.1   Representing the original FMA taxonomy

The representation of the original FMA taxonomy in OWL was straightforward. The subclasses—superclass relation between frames was represented by the `rdfs:subClassOf` relation. The resulting hierarchy is homologous to the original FMA one.

3.2 Representing multiple inheritance

Due to the single inheritance constraint, the following situations incompletely account for the subclass—superclass structure:

- left/right and male/female: 3 classes (`Breast`, `Areola`, `Nipple`)

- left/right and enumeration: 65 classes (e.g. `Left first cervical nerve`)

- upper/lower and enumeration: 5 classes (e.g. `Upper first molar socket`)

For example, `Breast` has four direct subclasses: `Left breast`, `Right breast`, `Male breast` and `Female breast`. Each of `Male breast` and `Female breast` have one left and one right direct subclass. Consequently, `Left male breast` is a subclass of `Male breast`, but not of `Right breast`. Moreover, classes such as `Intercostal lymph node` combine the last two patterns.

4 Disjunction

In the original FMA, by default, all the sibling classes are disjoint. For example, the direct subclasses of `Cell` are `Nucleated cell` and `Non-nucleated cell`, and it is clear that a cell cannot be at the same time nucleated and non-nucleated. This feature can be made explicit in OWL with disjunction constraints.

However, systematically making all the sibling classes mutually disjoints requires to rule out situations where all the direct subclasses of a class are not mutually exclusive. In the previous example, `Left breast` and `Right breast`, as well as `Male breast` and `Female breast` are disjoint, but `Left breast` and `Male breast` are not.

Rather than risking some inconsistency or some trivial satisfiability of the ontology, we chose the conservative approach of only generating the disjointness constraints we are certain of. This lead us to the identification of some other patterns among sibling classes:

- `Left X`/`Right X`: 3736 classes (e.g. `Left lung`)

- `X left Y`/`X right Y`: 13989 classes (e.g. `Skin of right breast`)

- `Male X`/`Female X`: 25 classes (e.g. `Male breast`)

- `X male Y`/`X female Y`: 75 classes (e.g. `Right side of male chest`

- enumeration: XX classes (e.g. `First cervical nerve`)

- upper/(middle)/lower: YY classes (e.g. `Upper lobe of lung`)

5 Necessary and sufficient definition

5.1 Coverage

The FMA aims at completeness. It is assumed that for each class, its subclasses provide a complete decomposition (i.e. there are no `X-Other` nor `X-Unspecified` subclasses of the `X` class). For example, the two direct subclasses of `Organ` are `Solid organ` and `Cavitated organ`. The intended meaning is that each organ is either solid or cavitated, and that there is no third possibility. In OWL, this can be made explicit by a coverage definition. In the previous example, `Organ` would be defined as the union of `Solid organ` and `Cavitated organ`.

A naive approach would consist in generating a coverage definition for each class using the union of its direct subclasses. This would successfully generate the definition that a lobe of lung is either an upper lobe of lung or a middle lobe of lung or a lower lobe of lung. The result would always be correct from a logical point of view. However, in some situations, it can still be refined. If we return to the class `Breast`, we could further specify that a breast is either a left breast or a right one, and also that it is either a male breast or a female one.

In order to generate coverage definitions as specific as possible, we reused the patterns identified for generating multiple inheritance (Section 3.2) and disjunction (Section 4). For each of the matching pattern, we generated the corresponding coverage (this is what allows us to generate two definitions for `Breast`). We also generated a coverage definition with the classes matching none of the patterns (e.g. for `Lobe of lung`).

5.2 Auxiliary classes

The FMA is composed of numerous classes for which we can provide a definition considering the name and their place in the taxonomy. Typically, a class `Region of X` that is a sibling of the class `X` can be defined as the anatomical structures being direct regional parts of `X`. A similar approach can be applied to variants such as `Part of X`, `X part` or `Zone of X`.

6 Closure

The FMA describes canonical anatomy. Representing the range of possible anatomical variations is beyond its scope. Therefore, relations between anatomical structures can be represented in OWL using existential constraints, which would require that the relation always hold between every instance of the domain of the relation and at least one instance of its filler (and possibly instances of other classes as long as they are consistant with the range of the relation).

Closure constraints allow to further specify a class by enforcing that some relation always lead to some class. Such constraints are important for preventing a foot to be a part of a lung for example. The automatic generation of such closure constraints, if possible, is currently under investigation.

7 Discussion

The work by Noy and Rubin [7] focuses on a translation that represents all the information that is present in the original FMA, and only the explicit information. Our approach is complementary and focuses on the information that is implicit in the FMA.

From a logical point of view, our approach is based on a combination of lexical and structural properties. However, it also exploits the underlying implicit semantics bore by the modeling principles. In this sense, it can be qualified of hybrid.

We do not claim to have identified all the possible patterns. Moreover, it might be necessary to add specific additional constraints for some classes (e.g. the definition for `Nucleated cell` as any cell having a nucleus, and conversely the definition for `Non-nucleated cell` as any cell having no nucleus). These can be represented directly in OWL for reusability and be incorporated when necessary.

We observe dependencies between some of the patterns. For example, the additional taxonomic relations have to be generated before the coverage constraints. In the previous examples, it is necessary to have generated the fact that `Left male breast` and `Left female breast` are subclasses of `Left breast` for being able to define the latter as equivalent to the union of the former two classes.

References

[1] C. Rosse, J.L. Mejino, R. Modayur, B.R.and Jakobovits, K.P. Hinshaw, and J.F. Brinkley. Motivation and organizational principles for anatomical knowledge representation: The digital anatomist symbolic knowledge base. *Journal of the American Medical Informatics Association*, 5(1):17–40, Jan/Feb 1998.

[2] C. Rosse and J.L.V Mejino. A reference ontology for bioinformatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, 36:478–500, 2003.

[3] Daniel L. Rubin, Olivier Dameron, Yasser Bashir, David Grossman, Parvati Dev, and Mark A. Musen. Using ontologies linked with geometric models to reason about penetrating injuries. *Artificial Intelligence in Medicine*, 37(3):167–176, 2006.

[4] Olivier Dameron, Daniel L. Rubin, and Mark A. Musen. Challenges in converting frame-based ontology into OWL: the Foundational Model of Anatomy case-study. In *American Medical Informatics Association Conference AMIA05*, pages 181–185, 2005.

[5] Christine Golbreich, Songmao Zhang, and Olivier Bodenreider. The foundational model of anatomy in OWL: Experience and perspectives. *Web Semantics*, 4(3):181–195, 2006.

[6] Songmao Zhang, Olivier Bodenreider, and Christine Golbreich. Experience in reasoning with the foundational model of anatomy in OWL DL. In *Proceedings of the Pacific Symposium on Biocomputing*, number 11, pages 200–211, 2006.

[7] Natalya F. Noy and Daniel L. Rubin. Translating the foundational model of anatomy into owl. Technical report, Stanford Medical Informatics, 2007.