

Challenges in Converting Frame-Based Ontology into OWL: the Foundational Model of Anatomy Case-Study

Olivier Dameron, PhD, Daniel L. Rubin, MD, MS and Mark A. Musen, MD, PhD

Stanford Medical Informatics, Stanford University School of Medicine, Stanford CA 94305, USA

Abstract

A description logics representation of the Foundational Model of Anatomy (FMA) in the Web Ontology Language (OWL-DL) would allow developers to combine it with other OWL ontologies, and would provide the benefit of being able to access generic reasoning tools. However, the FMA is currently represented in a frame language. The differences between description logics and frames are not only syntactic, but also semantic. We analyze some theoretical and computational limitations of converting the FMA into OWL-DL. Namely, some of the constructs used in the FMA do not have a direct equivalent in description logics, and a complete conversion of the FMA in description logics is too large to support reasoning. Therefore, an OWL-DL representation of the FMA would have to be optimized for each application. We propose a solution based on OWL-Full, a superlanguage of OWL-DL, that meets the expressiveness requirements and remains application-independent. Specific simplified OWL-DL representations can then be generated from the OWL-Full model by applications. We argue that this solution is easier to implement and closer to the application needs than an integral translation, and that the latter approach would only make the FMA maintenance more difficult.

Introduction

In the domain of knowledge representation, ontologies are shared conceptualizations of a domain, and they possibly include the representations of these conceptualizations [1]. Ontologies are independent from the applications that use them. This leads to easier software and knowledge maintenance, and contributes to the semantic interoperability between applications [2]. In the medical domain, anatomy is a fundamental discipline that underlies most medical fields [3]. The Foundational Model of Anatomy (FMA) is the most complete ontology of canonical human anatomy [4]. It strictly follows a principled modeling approach and currently includes more than 70,000 concepts and 1.5 million relationships.

Among the representation formalisms for ontologies, the Web Ontology Language¹ (OWL) is the widely

accepted standard for representing and sharing knowledge in the Semantic Web context. OWL comes in three versions supporting different compromises between expressiveness and computational tractability: OWL-Lite only supports classification hierarchies and simple constraints, OWL-DL is more expressive but still computationally tractable, and OWL-Full is even more expressive but offers no computational guarantee. Particularly, OWL-Lite and OWL-DL belong to the description logics [5] family, which are decidable fragments of first order logics. Description logics-based languages have a precisely and formally defined semantics referring to the set theory. Some generic reasoning tools have been developed that leverage this semantics. Thus, an application can reason about an ontology represented in description logics without having to implement any inference function.

The FMA is represented in a frame language. However, for some applications it is desirable to use an OWL representation of the FMA, either for reasoning purposes [6] or for integrating it with other OWL ontologies, such as the NCI thesaurus². The problem is that frames' semantics is not as precisely defined as description logics' one. Moreover, although superficially similar, these two approaches rely on fundamentally different modeling assumptions, and there is no direct mapping between them. Protégé³, the ontology editing platform that was used to build the FMA supports both formalisms. The frame-based mode has an "export to OWL" option. However, this option only performs a straightforward translation that ignores all the features that do not have a direct equivalent. Moreover, it does not take advantage of all the OWL-specific features that are the basis of the language strength. For these two reasons, the resulting translation would not be usable for reasoning.

We analyze some theoretical and computational issues of representing the FMA in OWL-DL. To address the expressiveness limitation, we propose to use a more expressive formalism ensuring application independence while meeting the expressiveness requirements. To address the computational limitations, we propose a "Virtual FMA-OWL" architecture

¹ <http://www.w3.org/2004/OWL/>

² <http://ncimeta.nci.nih.gov>

³ <http://protege.stanford.edu>

based on a Web Service that returns the OWL-Full representation of a concept given its identifier. Eventually, we advocate the use of this architecture for continuing to maintain the FMA in the current frame-based form while making it accessible to the Semantic Web. Note that the intention of this article is not to discuss the modeling of the FMA [4], but rather to examine different representation formalisms considering the computational requirements of the applications that use them.

Converting the FMA into OWL-DL

The FMA is currently composed of more than 70,000 anatomical items called concepts, having more than 1.5 million relationships -such as composition, neighborhood or blood supply- between them. The concepts are identified by a unique number called the FMAID, and are associated with one or more designation (e.g. the string “Heart” for the concept 7088 corresponding to the heart), which allows to handle synonyms or multiple languages. The concepts are strictly organized in a principled specialization hierarchy.

Basic concept representation: identifiers and designations

We represented the FMA concepts as OWL classes, and relationships as OWL properties. Classes were identified by their FMAID, relative to the FMA namespace⁴. This allows us to avoid any potential ambiguity with another ontology having a concept with the same identifier, as different ontologies have different namespaces. We used RDF labels to represent the concept designation, explicitly mentioning the language. This is illustrated by the Figure 1, in which the identifier is interpreted against the default namespace, which is declared at the ontology level to be the FMA one):

```
<owl:Class rdf:ID="7088">
  <rdfs:label xml:lang="en">Heart</rdfs:label>
  <rdfs:label xml:lang="fr">Coeur</rdfs:label>
  ...
</owl:Class>
```

Figure 1: Representation of the FMA identifiers and designations in OWL-DL

The previous example illustrates the logical framework. However, it seriously impairs user-friendliness when identifiers are used to display concepts in editing tools. It is possible to configure Protégé to override this behavior and to use a preferred language-specific label⁵.

⁴ <http://sig.biostr.washington.edu/fma#>

⁵ <http://protege.cim3.net/cgi-bin/wiki.pl?HidingIdentifiersWithLabels>

Metaclasses

The FMA features a complex structure of superclasses and subclasses [7]. For example, “Physical anatomical entity” is an instance of “Anatomical entity template”, and a subclass of both “Anatomical entity template” and “Anatomical entity”. OWL-DL does not support metaclasses, so we needed to remove them. We stick to the main taxonomy from the FMA (see Figure 2).

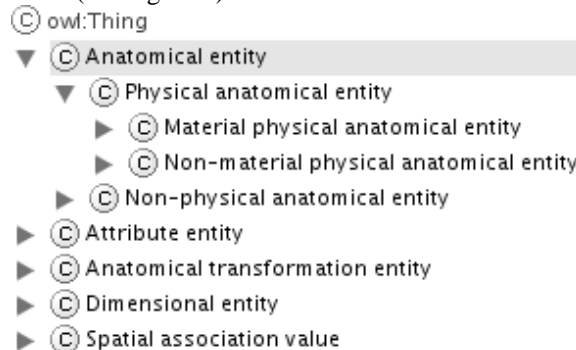


Figure 2: Taxonomy of the OWL-DL representation of the FMA

Disjointness

Description logics’ modeling principles are slightly different from those of the frames. These differences have to be taken into account during conversion. Particularly, the FMA is organized in a hierarchy of mutually-disjoint concepts. However, in description logics (hence in OWL-DL), classes are not disjoint by default (i.e. there can exist an individual that is an instance of both classes). Therefore, in order to respect the FMA modeling principles, we assume that unless specified otherwise by multiple inheritance, all the direct subclasses of a class are mutually disjoint. For example, Esophagus and Stomach are two direct subclasses of “Organ with organ cavity” and they are disjoint (an instance of “Esophagus” cannot be also an instance of “Stomach”). However, “Left breast”, “Right breast”, “Male breast” and “Female breast” should not be specified as disjoint (although they are automatically because “Left female breast” is only described as a subclass of “Female breast”, and not also of “Left breast”).

Note that this knowledge was implicit in the FMA and is made explicit in OWL

Closure

Another difference between frames and description logics is that the latter relies on the “open world assumption” [5] whereas the former assumes a closed world. In a closed world, everything that is not explicitly said is assumed to be false.

Consequently, when the FMA describes the parts of an anatomical structure such as the hand, the fact that the structures other than the palm or the finger are not said to be parts of the hand is interpreted as they are not part of the hand. However, in description logics, providing a list of the possible parts of the hand does not prevent other structures to be also parts of the hand. Therefore, we have to add an extra constraint saying that the structures in the list are the only possible parts of the hand. This is called introducing a closure axiom [8], and it has to be done for all the relationships (for an example see [6]).

However, generating closures is much more complicated than it may seem at first sight. For example, the possible parts of the hand are the palm and the five fingers. Now, we have to take overloading by subclasses into account so that the possible parts of the “Left hand” are “Left palm”, “Left thumb”, ..., “Left little finger” (note that the closure does not mention non-lateralized concepts such as “Thumb” anymore). However, the same approach cannot be applied to the lungs: a lung has an upper lobe and a lower lobe as parts. Its subclass “Right lung” not only has parts “Upper lobe of right lung” and “Lower lobe of right lung” (same approach as for the hand), but also a middle lobe that doesn’t exist for the left lung. As a consequence, it would be incorrect to generate a closure for the lung based on its parts, whereas it should be done for the hand. The first situation occurs when the child overloads its parent. The second one occurs when subclasses introduce new properties. Unfortunately, real world situations can mix these two situations.

In order to automate the systematic generation of closures, we have to check if all the classes that define the range of a relation for a concept are subclasses of the range of this relation for the superclasses of the concept.

N-ary and attributed relationships

N-ary relationships associate more than two entities. Particularly, this is extensively used in the FMA to qualify a relation between two entities. Those attributed relationships are used to qualify part or continuity relationships for example (the lung is continuous *medially* to the pulmonary veinous tree).

The modeling in description logics of such relationships has been studied by the W3C Semantic Web Best Practice working group, and we followed their recommendation [9].

Computational issues

The expressive OWL-DL ontology resulting from the extensive conversion of the FMA cannot be handled

by current classifiers such as Racer or FaCT++. They require that the entire model be in memory, which is not possible with the 70,000 concepts and 1.5 million relationships of the FMA, and the processing time would be prohibitive anyway. Therefore, it is impossible to check the consistency of the whole model. Consequently, the OWL-DL representation of the FMA cannot be combined with ontologies of other domains such as the NCI ontology of tumors as it would further increase the size of the ontology.

Our solution is to assume that depending on their scope, many applications do not need the whole FMA but a fraction of its concepts (e.g. those involved in the description of the neighborhood of the heart for the Virtual Soldier project) and a fraction of its relationships (mainly paronomies and blood supply in the previous example). Such considerations about large ontologies resulted in the notion of a view as an application-dependent part of an ontology [10].

Synthesis

The conversion from frames to OWL-DL required us to forgo representing some features of the FMA such as metaclasses. On the other hand, it allowed us to represent explicitly other features such as disjointness or closure.

However, one must select the features to give up and choose how to represent additional features. Both of which are application dependent.

Moreover, there is clearly a scalability issue as an OWL-DL representation of the complete FMA cannot be managed by the current classifiers (or it requires a very crude representation with hardly any concepts), thus failing to meet the initial goal of using OWL-DL.

These two points suggest that OWL-DL is not the optimal representation formalism for the FMA. In addition, such a representation of the whole model would not be usable with regards to the initial goal of reasoning.

Addressing expressiveness and application-independence: OWL-Full

From the previous section, we have seen that some of the FMA features are simply out of the scope of OWL-DL. We propose a two-layered approach. The first layer consists of a generic conversion tool that generates a representation of the FMA in OWL-Full. The second layer consists of several application-specific optimization tools that simplify the OWL-Full representation of concepts into OWL-DL ones by removing all the features unnecessary according to the application context.

OWL-Full does not suffer from the expressiveness limitations of OWL-DL. For example, it supports metaclasses. Figure 3 shows that the “Heart” (FMAID: 7088) can be represented in OWL-Full both as a subclass and as an instance of “Organ with cavitated organ parts” (FMAID: 55673), which complies with the original FMA structure.

```
<owl:Class rdf:ID="7088">
  <rdfs:label xml:lang="en">Heart</rdfs:label>
  <rdfs:label xml:lang="fr">Coeur</rdfs:label>
  <rdfs:subClassOf rdf:resource=
    "http://sig.biostr.washington.edu/fma#55673"/>
  <rdfs:type rdf:resource=
    "http://sig.biostr.washington.edu/fma#55673"/>
  ...
</owl:Class>
```

Figure 3: Representing the original FMA metaclasses and subclasses structure in OWL-Full (concept #7088 is the heart; concept #55673 is “Organ with cavitated organ part”)

Addressing the computational limitation: the “Virtual FMA-OWL”

We have seen that the excessive number of concepts and relationships of a complete representation of the FMA in OWL-DL was a serious computational limitation. Moreover, in order to address the expressiveness requirements and the desire to provide an OWL representation of the FMA that would be application-independent, we proposed to replace OWL-DL by OWL-Full, which offers no computational guarantee. In order to meet the primary goal and to provide an OWL version of the FMA actually usable for reasoning purposes by applications, we propose a solution that restricts the model to only those concepts relevant for an application, which allows to simplify the OWL-Full representation of these concepts into an optimized OWL-DL form.

Granular access (Virtual FMA-OWL)

The same reasoning requirement that triggered the consideration of converting the FMA into OWL makes it both difficult and useless to have a complete standalone OWL representation. Therefore, we propose to implement a more granular access by a Web Service that returns the generic OWL-Full description of a single concept given its identifier. Such descriptions are similar to the code snippet presented at Figure 3. This can be achieved because the OWL syntax relies on URL that can be resolved at run time.

By following the relations that are relevant to its context, an application can build an OWL-Full representation on a small fraction of the FMA. For example, the view of the FMA composed of the heart, its parts and their respective superclasses contains

3294 concepts, while the same view about the thorax gathers 45,085 concepts.

This architecture has been deployed on the intranet using the Apache Tomcat server and the axis library⁶. For a simpler usage, we added another Web Service that given a name returns the matching concept identifier (or a list thereof).

Simplifying OWL-Full into OWL-DL

The OWL-Full representation of a concept can then be simplified by deleting all the constructs (typically metaclasses and some relationships) that are not used in the application context.

These simplifications are completely specific to the application context.

Discussion

The first section demonstrated that the conversion of the original FMA into OWL-DL was hindered (1) by the impossibility to represent some of the frame-based constructs in an application-independent way, and (2) by the inability of the reasoners to handle the FMA as a whole, which was the primary reason for considering the conversion. We proposed to use OWL-Full, which is more expressive than OWL-DL to alleviate the first point. We also implemented a granular access to this OWL-Full representation on a concept-by-concept basis.

OWL-Full vs OWL-DL

OWL-Full allows us to generate a layer that has all the expressiveness we may need and that is application-independent. The simplification of this layer back into an application-specific optimized OWL-DL representation relies on generic functions. We argue that it is simpler than generating a specific OWL-DL converter of the original FMA for each application.

Moreover, this approach promotes interoperability: any application that requests the concept 7088 gets the same description in OWL-Full. The application is then free to modify this description internally according to its specific needs (namely simplify it to meet its computational requirements), but at least, the communication between applications refers to a shared representation.

Granular access to the FMA-OWL

Providing access to an OWL representation of the FMA on a concept-to-concept basis relies on the assumption that nobody actually needs a whole dump of the FMA in OWL.

Moreover, this approach can be applied to both OWL-Full and OWL-DL.

⁶<http://ws.apache.org/axis/>

Eventually, it can support some precomputing or caching. The representation of each concept does not have to be computed on the fly.

Switching vs. converting the FMA into OWL

The last point raises another question: should the current representation of the FMA in frame be abandoned in favor of an OWL version (preferably in OWL-Full)?

Switching to an OWL-DL representation of the FMA seems to be the wrong approach because it requires to adopt implementation choices based on application-specific considerations. Moreover, it means being limited by the expressiveness of OWL-DL. This point is in contradiction with analyses that suggest it is best to use a formalism that is as expressive as necessary for modeling knowledge, and to use a different formalism for reasoning [11]. Eventually, this approach ignores the legacy of the applications that currently use the FMA without resorting to description logic classification for reasoning.

Switching to an OWL-Full representation of the FMA would address the expressivity problem, but not the legacy one. Moreover, it introduces an additional overhead by making the maintenance of the FMA much more difficult than it currently is (e.g. all the closures have to be re-examined every time you add a relationship between two concepts). This is a daunting prospect, as curating the frame representation of the FMA is highly error-prone.

Eventually, maintaining the FMA in its current form and developing a general function that provides an OWL-Full representation for any concept appears to be the best solution.

Conclusion

We have demonstrated that the direct generation of an OWL-DL representation of the FMA is possible, but requires us to give up some of the original FMA features, and to use some application-dependent shortcomings. However, this process also makes explicit some features that were implicit in the original FMA. Eventually, the result cannot be managed by the current reasoners, which was the original goal. Therefore, we proposed to generate an intermediate OWL-Full representation of the FMA, that would address the expressiveness requirements, and would be application independent. We have shown that this representation can be computed on a concept by concept basis, thus alleviating the burden of an integral conversion. Eventually, the OWL-Full representation of a concept can be simplified according to the application's computational requirements. We believe that this architecture will be beneficial by making

the FMA accessible to the Semantic Web while retaining the legacy.

Acknowledgments

This work was supported by a contract from DARPA, executed by the U.S. Army Medical Research and Materiel Command/TATRC Cooperative Agreement, Contract W81XWH-04-2-0012. This work was also supported by the Protégé resource, under grant LM007885 from the U.S. National Library of Medicine.

The authors are grateful to the FMA team for their advice and their support.

References

- [1] Gruber R. Formal Ontology in Conceptual Analysis and Knowledge Representation. Chapter "Towards principles for the design of ontologies used for knowledge sharing" in *Conceptual Analysis and Knowledge Representation*. 1993.
- [2] Chandrasekaran B, Josephson JR and Benjamins VR. What are ontologies and why do we need them? *IEEE Intelligent Systems*. 1999;14(1):20-26.
- [3] Rosse C, Mejino JL, Modayur BR, Jakobovitz R, Hinshaw KP and Brinkley JF. Motivation and Organizational Principles for Anatomical Knowledge Representation. *JAMIA*. Jan-Feb 1998;5(1):17-40.
- [4] Rosse C, Mejino JL, Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003;36(6):478-500.
- [5] Baader F, Calvanese D, McGuinness D, Nardi D and Patel-Schneider P. *The Description Logic Handbook*. Cambridge University Press. 2003.
- [6] Rubin DL, Dameron O and Musen MA. Use of Description Logic Classification to Reason about Consequences of Penetrating Injuries. *AMIA* 2005.
- [7] Noy N, Musen MA, Mejino JL and Rosse C. Pushing the Envelope: Challenges in a Frame-Based Representation of Human Anatomy. *Data and Knowledge Engineering Journal*; 48(3):335-359. 2002.
- [8] Rector A, Drummond N, Horridge M, Rogers J, Knublauch H, Stevens R, Wang H, Wroe C. OWL Pizzas: Practical Experience in Teaching OWL-DL: Common Errors and Common Patterns. *European Conference on Knowledge Acquisition (EKAW-2004)*. 63-81. 2004.
- [9] Noy N and Rector A. Defining N-ary Relations on the Semantic Web: Use with Individuals. W3C Techn. Report. <http://www.w3.org/TR/swbp-n-aryRelations/>
- [10] Noy N and Musen MA. Specifying Ontology Views by Traversal. *International Conference on the Semantic Web (ISWC-2004)*. 2004.
- [11] Doyle J and Patil RS. Two Theses of Knowledge Representation. *Artificial Intelligence*;48(3):261-297. 1991