Artificial Intelligence in Medicine 69 (2016) 12–21

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

From frames to OWL2: Converting the Foundational Model of Anatomy

Landon T. Detwiler[a,*], Jose L.V. Mejino[a], James F. Brinkley[a,b,c]

[a] Biological Structure, University of Washington Structural Informatics Group, 1959 NE Pacific Street, Box 357420, Seattle, Washington 98195, USA
[b] Computer Science and Engineering, University of Washington, Seattle, Washington 98195, USA
[c] Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington 98195, USA

ARTICLE INFO

Article history:
Received 26 June 2015
Received in revised form 22 April 2016
Accepted 23 April 2016

Keywords:
OWL
Frames
Ontology
Foundational Model of Anatomy
Ontology of Craniofacial Development and Malformation

ABSTRACT

Objective: The Foundational Model of Anatomy (FMA) [Rosse C, Mejino JLV. A reference ontology for bioinformatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* 2003;36:478–500] is an ontology that represents canonical anatomy at levels ranging from the entire body to biological macromolecules, and has rapidly become the primary reference ontology for human anatomy, and a template for model organisms. Prior to this work, the FMA was developed in a knowledge modeling language known as Protégé Frames. Frames is an intuitive representational language, but is no longer the industry standard. Recognizing the need for an official version of the FMA in the more modern semantic web language OWL2 (hereafter referred to as OWL), the objective of this work was to create a generalizable Frames-to-OWL conversion tool, to use the tool to convert the FMA to OWL, to "clean up" the converted FMA so that it classifies under an EL reasoner, and then to do all further development in OWL.
Methods: The conversion tool is a Java application that uses the Protégé knowledge representation API for interacting with the initial Frames ontology, and uses the OWL-API for producing new statements (axioms, etc.) in OWL. The converter is relation centric. The conversion is configurable, on a property-by-property basis, via user-specifiable XML configuration files. The best conversion, for each property, was determined in conjunction with the FMA knowledge author. The convertor is potentially generalizable, which we partially demonstrate by using it to convert our Ontology of Craniofacial Development and Malformation as well as the FMA. Post-conversion cleanup involved using the Explain feature of Protégé to trace classification errors under the ELK reasoner in Protégé, fixing the errors, then re-running the reasoner.
Results: We are currently doing all our development in the converted and cleaned-up version of the FMA. The FMA (updated every 3 months) is available via our FMA web page http://si.washington.edu/projects/fma, which also provides access to mailing lists, an issue tracker, a SPARQL endpoint (updated every week), and an online browser. The converted OCDM is available at http://www.si.washington.edu/projects/ocdm. The conversion code is open source, and available at http://purl.org/sig/software/frames2owl. Prior to the post-conversion cleanup 73% of the more than 100,000 classes were unsatisfiable. After correction of six types of errors no classes remained unsatisfiable.
Conclusion: Because our FMA conversion captures all or most of the information in the Frames version, is the only complete OWL version that classifies under an EL reasoner, and is maintained by the FMA authors themselves, we propose that this version should be the only official release version of the FMA in OWL, supplanting all other versions. Although several issues remain to be resolved post-conversion, release of a single, standardized version of the FMA in OWL will greatly facilitate its use in informatics research and in the development of a global knowledge base within the semantic web. Because of the fundamental nature of anatomy in both understanding and organizing biomedical information, and because of the importance of the FMA in particular in representing human anatomy, the FMA in OWL should greatly accelerate the development of an anatomically based structural information framework for organizing and linking a large amount of biomedical information.

* Corresponding author.
E-mail address: det@uw.edu (L.T. Detwiler).

http://dx.doi.org/10.1016/j.artmed.2016.04.003

1. Introduction

The Foundational Model of Anatomy (FMA) [1,2] is an ontology that represents canonical anatomy at levels ranging from the entire body to biological macromolecules. With over a hundred thousand concepts and over a million relations between concepts the FMA is one of the largest and most complex biomedical ontologies in existence. Because of its extensive coverage of anatomy and principled construction the FMA is rapidly becoming the human anatomy reference standard, and a template for model organisms, with most of the existing biomedical ontologies and terminologies incorporating or aligning to the FMA for their own anatomy axes.

Prior to the work described in this report the FMA was represented in a modeling formalism referred to as Protégé Frames [3] which derives from the Open Knowledge Base Connectivity (OKBC) specification [4]. Although Frames was the representation of choice when we began the FMA project, the current preferred representation is OWL2, hereafter referred to as OWL (Web Ontology Language) [5]. OWL has gained a wide following in recent years, in part because it is the representation of choice for the semantic web [6] and as such facilitates the creation of a worldwide interconnected web of knowledge. There has therefore been an increasing demand, both by us and by other FMA users, to convert the FMA to OWL so that it may more easily contribute to the global knowledge base.

The conversion of the FMA into OWL has been attempted several times in the past [7–11]. Our original intension was to extend the implementations from one or more of these earlier efforts. However, these efforts were either incomplete, did not classify under a reasoner, or were simply unavailable to us. We thus decided to do our own conversion, with the added advantage that since we are also the authors of the FMA we not only have domain-specific insight into the ontological formalisms we are trying to represent, but also are able to modify the output of the conversion to be consistent under reasoning.

No other authors of conversion tools had this kind of control and domain expertise, with the result that our converted and post-processed FMA is the only conversion that captures all or most of the information in the original Frames model while also classifying under a reasoner. For these reasons we propose that our OWL version of the FMA should replace any existing OWL conversions, and should become the official release version of the FMA.

The purpose of this paper is to describe our conversion and post-conversion methods. Since so many people use the FMA, either for content development or for informatics research, and since the conversion is a major change from the Frames version, we feel it is important to describe our methods so others can understand the ontological formalisms we chose and the various tradeoffs we made.

In the remainder of this paper we first describe our conversion methods and their potential generalization to other Frame-based ontologies, which we have partially verified by using it not only to convert the FMA, but also to convert our Ontology of Craniofacial Development and Malformation (OCDM) [12]. We then describe our post-conversion work that resulted in an FMA that classifies under an EL reasoner. Next, we describe changes likely to be needed by applications to use the converted FMA, illustrated with changes we needed to make to our own applications. Finally, we speculate about the potential impact that the converted FMA can have on the emerging life sciences semantic web [13] for representing and organizing biomedical knowledge and data. At the end of the paper we provide links for obtaining the converter program, the converted FMA and OCDM, and an online browser for FMA viewing.

2. Conversion procedure

Our overall conversion approach is similar to earlier efforts. In particular it is based on that of Golbreich, et al. [7–9], but was also informed by those of Noy and Rubin [10] and Dameron et al. [11] as well as unpublished investigations by Alan Ruttenberg and by Robert Hoehndorf.

The purpose of the conversion was first and foremost a translation in syntax into a model that is readable, operable, editable by current OWL ontology tools, and amenable to post-processing as we describe in Section 3. We endeavored to capture everything that was said in the Frames model, albeit often with subtle differences. We attempted to do this in a very generic way that would also work across several other ontologies that we are presently working with, in particular the OCDM.

Our approach is configurable, but in broad brush-strokes. We do not support frame-by-frame configuration. Rather, our implementation is relation centric. It allows a user to dictate how relationship types are transformed into OWL constructs, by associating specialized converters. This is a far more manageable way to configure the conversion of the FMA, as there are over 150 types of relationship but around 200,000 frames.

We attempted to capture everything, even if doing so might result in a computationally intractable model (e.g. a model upon which a classifier might not complete in a usable amount of time). We attempted to fix the problems that were easily identifiable in Frames prior to conversion (facet violations). We did not try to correct errors that were not flagged by Protégé, and we did not try to detect logical issues, such as unsatisfiable classes, in the OWL model being constructed. Many of these issues were addressed post-conversion (Section 3).

Very little was added during the translation. As was the case with cleaning the model, much was left as to-do items once we had the FMA in OWL. An important example of this pertains to sufficient conditions. As sufficient conditions are not present in the Frames model, they were not generated in our OWL translation (e.g. only necessary conditions were generated, complete logical class definitions have yet to be constructed).

Our converter is written in Java and configured via an associated XML file. It uses the Protégé Frames Application Programming Interface (API) to interrogate the original FMA in Frames. The OWL API is used to generate the OWL classes, individuals, properties, axioms, etc., and to write the resulting model out to a file. Use of the APIs is an example of one of the changes that we wanted to make to the prior conversion tool most similar to our own, but for which we were unable to obtain the source code [7]. That conversion required that the FMA, originally stored in a Protégé Frames database backend, first be converted into the Protégé flat file format CLIPS. The CLIPS file was then used as input. The conversion of the FMA to CLIPS is time and resource consuming, sometimes introduces errors, and in our view is not necessary. So our converter runs directly over the native FMA database via the Protégé API.

2.1. *Frames and OWL: apples and oranges*

Before we talk further about our specific methods, we mention a few terminological distinctions between Frames and OWL, drawing some rough analogies between the two modeling formalisms. Anatomical concepts are represented as classes, and the notion of a “class” exists in both Frames and in OWL. In Frames a class is a collection of “instances”, whereas in OWL it is a potential collection of “individuals”. In Frames, relationships are referred to as “slots” whereas in OWL they are called “properties”. “Facets” are used in Frames to constrain the allowed values for a slot. In OWL there are similar constructs, referred to as facets or “restrictions”. They