Abstracting and Generalising the Foundational Model Anatomy (FMA) Ontology

Eleni Mikroyannidi, Alan Rector and Robert Stevens

School of Computer Science, University of Manchester, UK

ABSTRACT

Motivation: The Foundational Model of Anatomy (FMA) is arguably the largest and most complete ontology of anatomy available. However, its very size makes it difficult to use [1, 2]. Much of its size is due to repetition of identical or very similar features for symmetrical parts, *e.g.* left and right, proximal, middle, and distal, etc. For many purposes only the parent abstraction is needed. A methodology is presented to abstract the common parent from sets of self-similar parts with the goal of making the FMA smaller and more accessible without eliminating information.

Results: An ontology design pattern (ODP) called the selector pattern was applied to an extract of the FMA covering the skeleton and related notions applied to symmetrical portions of anatomy. Application of the pattern shrinks the extract of the FMA by a maximum of 57%. The abstracted ontology was expanded to test if all of the information from the original could be regenerated. Only a few discrepancies were discovered. Some revealed omissions in the FMA while others pointed to deficiencies in our algorithm to be addressed in future versions.

Conclusions: A smaller, more compact and manageable FMA ontology can be generated by application of the selector pattern without loss of information. The ontology can be expanded on demand. The approach can highlight omissions in symmetrical entities in the ontology and can make areas that lack symmetry more prominent.

Availability: All resources can be found via http://www.cs.man.ac.uk/~stevensr/menupages/FMASupplementary.php

1 INTRODUCTION

The Foundational Model of Anatomy (FMA) is a highly detailed reference ontology representing the human anatomy from a developmental perspective [3]. It has a potentially important role in the description of biological phenomena in biology. The FMA is not, however, regarded as an effective solution for bioinformatics data manipulation in part because of its size and detail, consisting of approximately 70,000 concepts. One factor that contributes to the size of the model is that symmetrical parts such as the Right and the Left hand and the Proximal, Middle and Distal Phalanx of each finger are represented individually, whereas most users require only the pattern for "hand" and for "phalanx of finger". In most cases the features of these symmetrical concepts differ only in the indicator of symmetry – right or left, proximal, middle or distal, etc.

This structure was examined in order to specify a more efficient and compact representation that is expected to be easier to use [4, 5]. In addition, an examination of the symmetrical entities in the FMA enabled a check on the consistency and completeness of the representation of the analogous characteristics of symmetrical entities.

The methodology aims first to detect symmetries and then to represent them as a single abstraction plus a set of "selectors". For example, the methodology aims to produce a single hand concept with a mechanism for generating the left- and right-hand classes on-demand.

A further goal of the abstraction methodology is to detect and report omissions in the FMA. Finally, in order to test the abstraction process an expansion algorithm was created. The purpose of developing this algorithm is to check whether the useful information is preserved during the abstraction process.

2 MATERIALS AND METHODS

2.1 FMA version

As a starting point we used the version of FMA in the Web Ontology Language (OWL) produced by [6], available in the supplementary material, which was further transformed to make it more convenient to work with, by using conventional OWL tools. The OWL version was chosen for convenience because of the availability of standard tools and the expectation that the result might be integrated with other OWL ontologies [7]. The methodology is, however, in no way specific to OWL. From this version the segment relating to the skeleton plus a few related notions such as the extremities were extracted using the methods developed by Seidenberg [8]. The subset consists of 6483 classes, 116 Object properties and 18477 SubClass axioms.

* To whom correspondence should be addressed.

2.2 Abstraction Algorithm

The algorithm combines lexical matching on the systematic naming conventions of the FMA and semantic comparison of the characteristics ("restrictions" in OWL parlance) on its classes. Each potential symmetry must be identified manually. Then a *Refining_entity* is created with a *Selector* subtree corresponding to the symmetry – *e.g.* Laterality with subclasses left and right (Fig.1a. A corresponding property is created, *e.g.* has_laterality (Fig1b). The *has_laterality* object property is defined as functional as an individual entity of the human body can have only one kind of laterality.

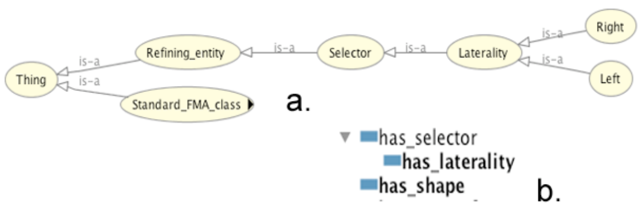

Fig. 1. The hierarchies created by the Abstraction algorithm. In a. the Refining_entity subtree including the Laterality Selector is shown. In b. the object property hierarchy is shown. The has_laterality property is defined as Functional with the class Laterality as Range.

The next step is the detection and abstraction of the symmetrical entities. For example, the Hand class has the *Right_hand* and *Left_hand* as subclasses, which have completely identical characteristics (OWL restrictions) up to the specification of "right" or "left" in their names – the same strings that distinguish the two subclasses. The abstraction algorithm therefore creates an additional restriction on the Hand class indicating that the Hand has right or left laterality ("*Hand has_laterality some Laterality*"). The common restrictions on Left_hand and Right_hand are moved to the parent concept "Hand". Finally, the subclasses *Left_hand* and *Right_hand* are removed. Figure 2 shows the structure of the Hand class before and after the abstraction.

The algorithm will abstract only the concepts referring to the symmetries defined in the Selector hierarchy. The sibling symmetrical classes are detected, checked if they conform to a number of preconditions and then substituted with existential restrictions. The first precondition that is checked is whether both symmetrical classes are present. The second, semantic, precondition checks that the number of existential restrictions on the symmetrical classes are the same. If the number is different they are not included in the abstraction. Finally, the labels of the sibling symmetrical classes are compared lexically with the label of the parent class. If the labels differ by more than one word, they are not included in the abstraction as the lexical algorithm for regenerating the original entities cannot cope with complex phrases. Rejected candidates for the pattern are reported in log files.

Fig. 2. The structure of the *Hand* before and after the abstraction. In a. the hierarchy of the class *Hand* with subclasses *Left_hand* and *Right_hand* before the abstraction is presented. In b. the hierarchy of the *Hand* after the abstraction is shown. The *Left_hand* and *Right_hand* have been removed. In c. the restrictions of the class Hand after the abstraction are presented. The underlined restriction "*has_laterality some Laterality*" has been added by the abstraction algorithm.

The algorithm as sketched above is generalized to consider more complex symmetries, e.g. Posterior and Anterior parts, Medial and Lateral parts etc. The symmetries are to be considered are defined in the input Symmetry hierarchy by the user. Figure 3 shows the Selector subtree and the corresponding property hierarchy that is automatically created by the algorithm for different kinds of symmetries.

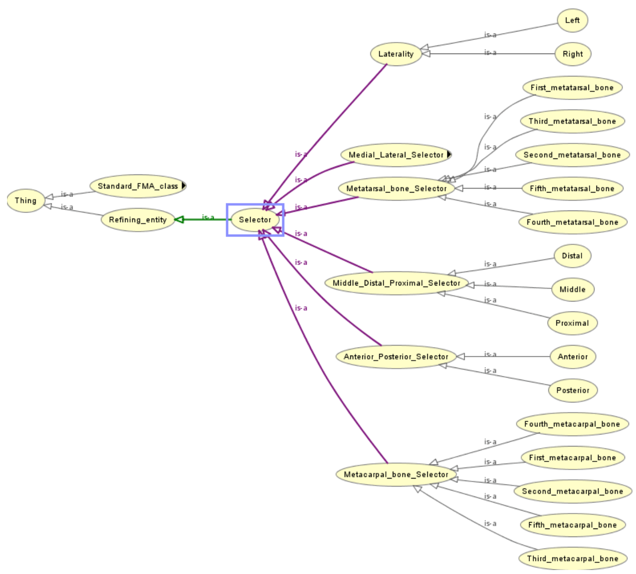

Fig. 3. The Selector subtree that is created by the abstraction algorithm for the definition of different symmetries is shown. All the symmetries are defined as subclasses in the Selector entity.

2.3 Expansion Algorithm

The expansion algorithm consists of two steps of execution.

In the first step all the sibling classes referring to the symmetries are created and in the second all their existential restrictions are created. For example, in order to expand the abstracted entities with Right and Left laterality, the algorithm detects the classes, which have the existential restrictions referring to the Selector type (e.g. *"has_laterality some laterality"*). Then two symmetrical sibling classes are created as subclasses of the class with the restriction. The labels of these classes are created by lexically modifying the label of their superclass. The lexical modification is *"Symmetry type + class name"*.

The second step is the creation of the existential restrictions in the new classes. The creation of the restrictions is based on the restrictions of the sibling classes' parent. The restrictions are lexically modified and adapted to the specialised class. For example, in order to recreate the restrictions of the Left_hand, the restrictions of the Hand class are copied and modified. Thus, the *"regional_part some Thumb"* restriction of the Hand class will be specialised into *"regional_part some Left_thumb"*. The outcome of the expansion algorithm cannot be considered as an exact copy of the input FMA subset as there are modifications, mainly in the labels of the recreated classes.

3 RESULTS

In Table 1 the ontology metrics are presented for the initial, abstracted and expanded FMA subsets. The number of the object properties in the abstracted ontology (column B) is greater than the number of object properties in the initial and the expanded ontology (column A and C), because in the abstracted ontology the property hierarchy referring to the selector has been added. As shown in Table 1B. the numbers of classes and the subclass axioms of the abstracted ontology have decreased. In addition, the ontology metrics of the expanded ontology indicate that the useful information, such as the number of classes and subclass axioms, is preserved.

Table 1. The FMA ontology metrics of (A.) the initial subset, (B.) after the application of the expansion algorithm and (C.) after the application of the expansion algorithm.

	A. Metrics of the initial FMA subset	B. Metrics of the FMA subset after the abstraction	C. Metrics of the FMA subset after the expansion
Class Count	**6483**	**3149**	**6483**
Object property count	116	118	116
Data property count	55	55	55
Subclass axioms count	**18477**	**11076**	**19453**
Inverse object properties axioms count	27	27	27
Functional object property axioms count	24	25	24

The abstraction and expansion algorithm are tested incrementally starting from the Laterality selector and by adding in each test a different kind of selector in the Symmetry hierarchy file. The metrics that are checked in each test are the number of Classes of the ontology and the number of SubClass axioms. The number of the existential restrictions is included in the SubClass axioms metric. The metrics of the abstracted and expanded ontologies for different symmetries are shown in Table 2 and Table 3.

Table 2. The number of classes of the (B.) abstracted and (C.) expanded FMA for each kind of symmetry (A.) that is added in the Symmetry hierarchy.

A. Type of symmetry that is added in each test	B. Class count of the abstracted FMA	C. Class count of the expanted FMA
Laterality selector	3149	6483
Metacarpal_bone_selector & Metatarsal_bone_selector	3032	6486
Anterior_Posterior_Selector	3029	6484
Middle_Distal_Proximal	2787	6482
Medial_Lateral_Selector	2782	6482

The number of classes of the initial FMA subset is 6483.

Table 3. The number of subclass axioms of the (B.) abstracted and (C.) expanded FMA for each kind of symmetry (A.) that is added in the Symmetry hierarchy.

A. Type of symmetry that is added in each test	B. Subclass axioms count of the abstracted FMA	C. Subclass axioms count of the expanted FMA
Laterality selector	11076	19453
Metacarpal_bone_selector & Metatarsal_bone_selector	10628	18892
Anterior_Posterior_Selector	10617	18868
Middle_Distal_Proximal	9944	18430
Medial_Lateral_Selector	9913	18374

The number of subclass axioms of the initial FMA subset is 18477.

From the results in Table 2 and Table 3 it can be concluded that the smaller the hierarchy is in the Symmetry hierarchy file, the easier it is for the expansion algorithm to reinstate the classes and the SubClass axioms of the initial ontology.

When more symmetries are considered, the current abstracted ontology does not preserve all the information of the initial FMA subset. In particular, the results of Table 3 have shown that not all the existential restrictions could be recreated. The reason is that both the abstraction and expansion algorithms need to be extended to track all the restrictions reliably, especially when an entity refers to many symmetries. In addition, the expansion algorithm makes the assumption that the FMA has no omissions and searches for classes with different labels compared to the initial FMA subset, in order to create an existential restriction. If, however, the searched class has not been abstracted because of some kind of omission, then it is not tracked by the expansion algorithm.

The expanded ontology sometimes gave better results in comparison with the initial subset, especially in the case of recreating the restrictions of the entities with Right and Left laterality. The initial FMA subset has a number of apparent omissions such as missing restrictions and missing classes. An example of apparent omission is the different number of restrictions appearing in *Left_shoulder* and *Right_shoulder*. Although these two sibling classes are considered as symmetrical and should have the same number and type of restrictions, they do not. If the algorithm considers the symmetrical classes with different number of restrictions, the *Left_shoulder* and *Right_shoulder* will be abstracted as well. Then, the number of existential restrictions is corrected in the expansion procedure and both classes have the same number.

4 CONCLUSIONS

In the abstracted version of our FMA subset the same information is provided with fewer classes and axioms. A more concise and compact ontology resulted from the application of the selector ODP. No semantic errors were traced during the manual checks in both the abstracted and the expanded ontology. The abstraction algorithm has the additional feature of detecting and reporting in log files all the symmetrical classes that were not abstracted because of omissions.

The abstraction methodology was generalised to cover different kinds of symmetry, although further development is needed for the most complex that involve interactions of multiple symmetries. Both the abstraction and expansion algorithms are generic and reusable, as they depend only on the input ontology and the input symmetry hierarchy defined by the user.

In summary, the application of the selector pattern can be used to abstract a more compact representation of the FMA. Considering only the abstraction for the symmetry in Left and Right parts, the number of classes is less than half of the initial class number. By considering other kinds of symmetries for abstraction, the ontology's size could be decreased even more. The "expanded" classes can be provided on demand. The process also affords opportunities for quality assurance. Subjectively, the smaller abstracted version appears easier to use, but this needs to be confirmed by empirical studies.

REFERENCES

1. Mork P, R. Pottinger, and P. Bernstein: Challenges in precisely aligning models of human anatomy. in Proceedings of Medinfo 2004. *San Francisco, CA: IMIA* 2004.
2. Mork PaPB: Adapting a generic match algorithm to align ontologies of human anatomy. *20th International Conference on Data Engineering* 2004, Boston:IEEE.
3. Cornelius Rosse JLVMJ: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Biomedical Informatics 36 (2003)* 2003: 478 - 500.
4. Alan Rector JR: Ontological and Practical Issues in Using a Description Logic to Represent Medical Concept Systems: Experience from GALEN. *Lecture Notes in Computer Science, Reasoning Web* 2006, Volume 4126/2006:197-231.
5. Aranguren ME, Antezana E, Kuiper M, Stevens R: Ontology Design Patterns for bio-ontologies: a case study on the Cell Cycle Ontology. *BMC Bioinformatics* 2008, 9(Suppl 5 doi:10.1186/1471-2105-9-S5-S1):S1.
6. Natalya F. Noy DLR: Translating the Foundational Model of Anatomy into OWL. *Preprint submitted to Elsevier* 2007.
7. Natalya F. Noy MAM, Jose L.V. Mejino, Jr., Cornelius Rosse: Pushing the Envelope: Challenges in a Frame-Based Representation of Human Anatomy. *Data and Knowledge Engineering, 48(3):335–359,* 2004.
8. Julian Seidenberg AR: Web Ontology Segmentation: Analysis, Classification and Use. *ACM 1-59593-323-9/06/0005* 2006.