OWL-based Reasoning Methods for Validating Archetyped Clinical Knowledge *

Marcos Menárguez-Tortosa
Departamento de Informática y Sistemas
Facultad de Informática
Universidad de Murcia, Spain
marcos@um.es

Jesualdo Tomás Fernández-Breis
Departamento de Informática y Sistemas
Facultad de Informática
Universidad de Murcia, Spain
jfernand@um.es

ABSTRACT

Most advanced Electronic Healthcare Records architectures represent clinical knowledge by means of archetypes. Consequently, guaranteeing the correctness and consistency of such archetypes becomes crucial for the success of those architectures. In this work, we present a method that uses OWL and reasoners for evaluating the consistency of the archetypes. This method has been validated through its application to the openEHR archetype repository, which is the largest available one nowadays. The results of this validation are also reported in this study.

1. INTRODUCTION

The lifelong clinical information of any person supported by electronic means configures his Electronic Health Record (EHR). Most advanced EHR architectures and standards are based on the dual model-based architecture [1] (OpenEHR, ISO EN 13606), which defines two conceptual levels: (1) reference model; and (2) archetype model. The reference model defines the set of entities that form the generic building blocks of the EHR. On the other hand, archetypes define how to represent clinical concepts in the form of structured and constrained combinations of the entities contained in the reference model, so knowledge in the EHR domain is defined at this level.

Archetype-based EHR domain knowledge plays a fundamental role for the achievement of semantic interoperability in healthcare [2]. In addition to this, the requirement of formal methods for validating the design and content of archetypes has been identified in [3]. So far, very few archetypes authoring tools implement techniques for assuring the quality of archetypes and none includes a knowledge-based representation of archetypes in order to perform semantic activities. In previous work, we addressed the representation of archetypes based on semantic web technologies in order to perform such semantic activities [5]. However, that representation requires much implementation work for validating the knowledge of archetypes and dealing with EHR extracts. In [4] a representation that allows for processing EHR extracts is proposed, but quality assurance and validation methods were not provided.

In this work, an OWL-based representation of archetypes that makes it possible to accomplish validation and quality assurance requirements is presented. The methods for checking the consistency of archetypes will be supported by reasoners, which will be in charge of helping to identify the wrong definitions in the archetypes.

2. ARCHETYPES

Archetypes are detailed and domain-specific definitions of clinical concepts in the form of structured and constrained combinations of the entities of the reference model. The ISO EN 13606 and openEHR communities specify them using the Archetype Definition Language (ADL). Next, an excerpt of the definition of an openEHR archetype for working with check lists in healthcare is shown.

```
-- A health oriented check list
EVALUATION[at0000] matches {
  data matches {
    -- Tree
    ITEM_TREE[at0001] matches {
      items cardinality matches {0..*; unordered} matches {
        -- Question group
        CLUSTER[at0002] occurrences matches {1..*} matches {
          items cardinality matches {0..*; unordered} matches {
            -- Question
            CLUSTER[at0003] occurrences matches {1..*} matches {
              items cardinality matches {1..2; ordered} matches {
                ELEMENT[at0004] occurrences matches {1..*} matches {*}
                -- A comment on the answer
                ELEMENT[at0005] occurrences matches {0..1} matches {
                  value matches {
                    DV_TEXT matches {*}
                  }
                }
                ...
}
```

Concepts in archetypes are identified by an id (e.g., `at0003`) and characterized by their *occurrences*, that is, the number of instances that can be part of the association to which they belong. Multivalued attributes may be restricted in different ways: `cardinality`, `order` and uniqueness (`unique`) of the instances. Finally, an archetype can be defined as the *specialization* of another one.

*This work has been possible thanks to the Spanish Ministry of Science and Innovation through grant TSI2007-66575-C02-02.

3. OWL REPRESENTATION

In order to check the correctness of the clinical archetypes based on OWL reasoning, we should be capable of representing archetypes in OWL. As it has been mentioned, archetypes are built by defining restrictions over the entities of the EHR reference model, usually expressed in UML. Therefore, an OWL-based representation of the reference model is needed. Otherwise, the validation of the archetypes would not be possible. This is not a difficult step since this representation can be obtained by applying the following simple rules: (1) UML classes are transformed into OWL classes; (2) attributes are represented as OWL properties and a *SubClassOf* axiom is included in the class in which it has been defined; (3) inheritance relations are transformed into *SubClassOf* axioms and sibling classes are defined disjoint; and (4) multivalued attributes are represented by means of a class which makes use of properties for defining the *order*, *repetitions*, *cardinality* and *range* of the values of the instances.

An archetype defines a clinical concept by constraining an entity of the reference model. Such concept represents a specialization of that reference model's entity. Those restrictions are applied to the attributes defined for each entity: range, cardinality, and so on. In this way, each restricted entity is defined by means of an OWL class which contains the definition of the corresponding constraints. Most constraints are applied to multivalued attributes. These constraints are defined over the *collection* class that represents the range of the property. This is the case of the attribute `items` of `CLUSTER[at0003]` that is shown next.

```
Class: CLUSTER_at0003
  EquivalentTo: ARCHETYPED_CLASS
    and id value "at0003"
    and CLUSTER
    and op_items only COLLECTION_CLUSTER_at0003_items

Class: COLLECTION_CLUSTER_at0003_items
  EquivalentTo: COLLECTION
    and id value "COLLECTION_CLUSTER_at0003_items"
    and elements min 1 ITEM
    and elements max 2 ITEM
    and elements min 1 ELEMENT_at0004
    and elements max 1 ELEMENT_at0005
    and ordered value true
    and unique value false
```

In case of defining an archetype that specializes another one, its definition includes an additional annotation in each class. That annotation indicates the name of the class in the parent archetype that is being specialized, if any. This annotation will then be used in the consistency checking process.

4. CONSISTENCY CHECKING

An archetype is consistent if its set of constraints defined over both the reference model and the parent archetype are valid. The following types of errors can be identified: incorrect restrictions and incorrect specializations. In this work, such inconsistencies are detected by using a strategy based on OWL reasoners. In the first case, if a class is not consistent with respect to the reference model, the reasoner will find that class unsatisfiable. Specialization errors requires the processing of the reasoner results. In this way, a specialization is wrong if the reasoner cannot infer a subsumption relation.

In order to provide detailed information about the causes of the inconsistency, support classes, that allow the isolation of each archetype constraint, are defined. For instance, a class is generated for defining the constraint of maximum cardinality for the attribute `items` of `CLUSTER_at0003`. The definition of these classes for the specialized archetype would be similar, except for the addition of the corresponding annotation to define its parent archetype.

```
Class: MAX_CARDINALITY_COLLECTION_CLUSTER_at0003_items
  EquivalentTo: MAX_CARDINALITY
    and id value "COLLECTION_CLUSTER_at0003_items"
    and max_cardinality_value only xsd:int[ <= 2 ]
```

5. VALIDATION AND TOOLING

The OWL-based method for checking the consistency of archetypes has been implemented in the tool *Archeck* that is available at http://miuras.inf.um.es/archeck. The tool has been implemented in Java and makes use of the openEHR Java tools. Ontologies are processed with the OWL API and we have used Pellet and Fact++ for the reasoning.

Our experimental validation has used the archetypes available in the openEHR repository. The validation of this repository has reported that 14 over 931 archetypes are inconsistent. In terms of reasoning time, the results obtained with Fact++ are better than with Pellet. The complete results of this experiment are available at the referred website.

6. CONCLUSIONS

This work proposes a representation of ADL clinical archetypes as OWL classes in order to check the correctness of their definitions by using OWL reasoners. The approach has been implemented in the *Archeck* tool and has been validated using the openEHR archetype repository. The tool has proved to be useful since a number of archetypes have been found inconsistent in that repository and the validation errors have been reported. *Archeck* will also be deployed as a web service in order to integrate the validation process in other tools.

7. REFERENCES

[1] T. Beale. Archetypes: Constraint-based Domain Models for Futureproof Information Systems. In *Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer*, 2002.

[2] European Commission. Semantic interoperability for better health and safer healthcare. deployment and research roadmap for europe. ISBN-13 : 978-92-79-11139-6, 2009.

[3] D. Kalra. EHR Archetypes in practice: getting feedback from clinicians and the role of EuroRec. In *eHealth Planning and Management Symposium*, 2007.

[4] O. Kilic, V. Bicer, and A. Dogac. Mapping Archetypes to OWL. Technical report, Middle East Technical University, Turkey, 2005.

[5] C. Martínez-Costa, M. Menárguez-Tortosa, J. T. Fernández-Breis, and J. A. Maldonado. A model-driven approach for representing clinical archetypes for Semantic Web environments. *Journal of Biomedical Informatics*, 42(1):150–164, 2009.