Ontology Patterns for Clinical Information Modelling

Catalina Martínez-Costa¹, Daniel Karlsson², Stefan Schulz¹

¹Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
²Department of Biomedical Engineering, Linköping University, Sweden

{catalina.martinez, stefan.schulz}@medunigraz.at, Daniel.Karlsson@liu.se

Abstract. Motivated by our experiences of representing clinical information using OWL DL, which often resulted in highly complex expressions, we propose the use of ontology content patterns to facilitate this task. They are based on a set of formal ontologies, constrained by the concepts and relations of a top-level one, which reduces arbitrariness in ontology design. We propose their application to information encoded by electronic health records specifications and ontology-based terminologies, in order to provide semantic interoperability across heterogeneously represented data, and to guide the creation of clinical models and detect semantic inconsistencies across them. We provide examples of their application to achieve the above mentioned tasks and discuss the limitations and further research issues.

Keywords: ontology content patterns, electronic health standards, SNOMED CT

1 Introduction

Despite a wide-spread use of computers in clinical documentation, the semantic interoperability of information kept in electronic health record (EHR) systems is insufficient [1]. A plurality of EHR representations together with medical terminologies like SNOMED CT [2], have been proposed in recent years to structure clinical information and to provide standardized codes for frequently used medical terms, respectively.

Existing EHR standards and medical terminologies were developed in isolation and major problems exist when they are combined. Projects such as the HL7 TermInfo [3] or more recently the Clinical Information Modeling Initiative (CIMI) [4] and the European network SemanticHealthNet [5], have attempted to provide solutions by addressing the lack of division between ontology-based medical terminologies and information models (provided by EHR representations). This is commonly known as the boundary problem [6].

TermInfo provides a set of rules for the combined use of the HL7 information model and SNOMED CT; CIMI proposes a set of modelling patterns, defined as clin-
ical models that are intended to act as guide for the creation of new ones. Clinical models constrain information model structures to represent particular data capture and communication use cases. In medicine it is often not possible to impose one universal data form, such as for recording diagnostic information. Thus, CIMI associates each clinical model with a set of iso-semantic models (models heterogeneously structured but with the same meaning), from which one is selected as the preferred one and mappings are established across them.

CIMI or HL7 based models that implement the TermInfo specification might work well in isolation, but semantic interoperability issues arise when interacting with others, which are not necessarily compatible, whilst the anticipation of all possible iso-semantic representations will lead to an explosion of models. The European network SemanticHealthNet addresses this problem by providing clinical model information structures with a set of expressions, based on a shared ontological framework. This framework allows representing both (ontology-based) medical terminologies and information models, and implements the classical distinction between ontology [7] (what exists – independently of being known or observed) and epistemology [8] (what is known, suspected, planned, etc.).

The inherent complexity of this representation is addressed by using semantic patterns as intermediate representations, which is the focus of this paper.

2 Background

2.1 EHR Structured Clinical Models

Several EHR standards and specifications propose representing clinical information by using clinical models based on a reference information model (RM). Clinical models, also known as archetypes (e.g. openEHR/ISO 13606 archetypes) [9,10] or HL7 CDA documents [11], constrain a set of standardized information structures provided by some reference model (RM), to represent EHR data. They are used for modeling particular use cases for clinical data capture and communication. As an example, the ISO 13606 archetype of Fig. 1 constrains information structures (e.g. CLUSTER, ELEMENT, etc.) to represent a medical questionnaire consisting of questions groups. The use of terminologies and ontologies within clinical models is known as terminology binding. Fig. 1 shows how the information structure ELEMENT[at0003] is bound to the SNOMED CT concept Past history of diabetes mellitus. Interpreted within the context of the clinical model, it is a question, and its allowed answers are yes / no.

In practice, the division line between ontologies and information models is often crossed both by ontologies (where they represent epistemic and temporal information aspects, such as “known present” or “past history of”) and by RMs and clinical models (where they carry their own ontology without reference to external standards, here the fact that it is a question).
2.2 Ontology-based medical terminologies: SNOMED CT

Ontologies formally describe properties and relations of types of entities. Domain-independent categories, relations and axioms are typically provided by top-level ontologies [12], whereas the types of things that make up a domain are represented by domain ontologies. In the former one we find categories like Process, Material entity, Quality, etc., whereas in a clinical domain ontology we would find Diabetes mellitus type 1, Left index finger, or Aspirin, i.e. the classes of entities corresponding to the terms used in clinical documentation and reporting, and defined by the properties shared by all of their individual members.

Medical terminologies have evolved in the last years to include definitional knowledge about their terms, by using an ontological framework in order to help humans and computers to recognize the intended meaning of their terms, for proper coding of, retrieval of, and inferencing about biomedical data, as well as for maintenance of the terminology itself [13]. An example is SNOMED CT, a clinical terminology covering all aspects of clinical medicine, with about 300,000 representational units (called SNOMED CT concepts) and terms in several languages.

Due to the legacy of its predecessors, SNOMED CT does not only provide codes for clinical terms proper but also for contextual statements, which are often represented in information models. An example of this is the Situation with explicit context concept hierarchy (i.e. context model), in which we find terms such as Suspected deep vein thrombosis or No past history of venous thrombosis. We have largely harmonized the SNOMED CT content with basic top-level classes and relations of BioTopLite upper ontology [14] (e.g. btl:Process, btl:Quality, btl:Condition, btl:Situation, etc), in order to better distinguish clinical from information entities. Based on [15] we interpret SNOMED CT concepts from the clinical finding hierarchy as clinical situations and reinterpreted the SNOMED CT context model [16]. Fig. 2 shows the OWL DL representation of a post-coordinated expression that follows the context model and represents past history of diabetes. Past history is a temporal aspect that specializes the meaning of the finding diabetes mellitus.

---

1 Post-coordination describes the representation of a term using a combination of two or more of them (e.g. past history of clinical finding and diabetes mellitus)
Fig. 2. OWL DL SNOMED CT representation of an expression based on the post-coordination of two terms (past history of clinical finding and diabetes mellitus) linked by the linkage concept associated finding. Terms using italics represent ontology classes, bold face represents ontology object properties.

3 Methods

A shared OWL DL [17] ontological framework is proposed that allows relating EHR information models with medical terminologies [18] in an unambiguous way. It is supported by a the use of semantic patterns in order to provide semantic interoperability across heterogeneously represented data and to guide the creation of clinical models and detect semantic inconsistencies across them.

3.1 Clinical Information Semantic Patterns

The semantic patterns we propose represent recurrent clinical information modelling aspects and can therefore be considered ontology design content patterns applied to clinical information. They are inspired by the experience of modelling clinical information based on ontologies. As ontology patterns they help to reduce the arbitrariness that exists when representing clinical information, by using a set of OWL DL formal ontologies as standard modelling framework [19].

Two ontologies, the SNOMED CT ontology (prefix sct) and an information ontology (prefix shn) are rooted in the biomedical top-level ontology BioTopLite (prefix btl). The use of BioTopLite standardizes the ontology development process, by providing a set of logical axioms which constrain how both ontologies are related. We use SNOMED CT as common reference point for representing the healthcare domain. The information ontology provides a set of classes that represent contextual and temporal information aspects (e.g. diagnostic information, past history, provisional, etc.) and refer to the SNOMED CT concepts.

Each pattern can be considered a small ontology based on the previous framework, to be used as a building block for a particular modelling use case. For that, they can be specialized and composed by following similar principles to object oriented languages [20].

According to [21], content patterns are language-independent and should be encoded in a high order representation language. Nevertheless, their representation in a logic-based language allows the use of DL reasoning [22], which can be used to ensure the consistency of the patterns and to allow inference-related tasks. On the left
side, Fig. 3 shows the graphical representation of a pattern that represents the past history of some patient clinical situation. The right side, shows a concrete instance of that pattern that represents the statement “Past history of diabetes mellitus”. Other examples of patterns are “Family history of clinical situation” or “Plan to perform some clinical process”.

![Graphical representation of history-situation pattern](image)

**Fig. 3.** (Left) Graphical representation of the history-situation pattern; (Right) Instance of the history-situation pattern; Squares represent ontology classes and unidirectional arrows predicates enhanced by cardinality constraints.

Within SemanticHealthNet, we have elaborated two representations of semantic patterns: an OWL 2 DL and a RDF [23] representation. The OWL-based representation describes a pattern as a set of logical axioms. **Table 1** shows the OWL rendering of the history-situation pattern as pieces of information (shn:InformationItem) that are acquired by performing some clinical process (shn:ClinicalProcess) and that refer to clinical situations (shn:ClinicalSituation) of a given type (if any), which happened in the past (sct:InThePast). Additionally, it allows expressing epistemic information aspects (shn:InformationAttribute) that indirectly refer to the situation (e.g. severe, present, etc.).

<table>
<thead>
<tr>
<th>shn:InformationItem</th>
</tr>
</thead>
<tbody>
<tr>
<td>and shn:isAboutSituation only shn:ClinicalSituation</td>
</tr>
<tr>
<td>and btl:isOutcomeOf some shn:ClinicalProcess</td>
</tr>
<tr>
<td>and shn:hasInformationAttribute some shn:InformationAttribute</td>
</tr>
<tr>
<td>and shn:hasInformationAttribute some sct:InThePast</td>
</tr>
<tr>
<td>and shn:hasInformationAttribute some sct:FindingContextValue</td>
</tr>
</tbody>
</table>

**Table 1.** OWL DL representation of history-situation pattern; Terms using *italics* represent ontology classes, *bold face* represents ontology object properties.

**Table 2** shows the RDF representation, which consists of a set of Subject-Predicate-Object (SPO) triples. Both representations are connected as follows: The subject and object parts of a triple correspond to ontology classes, and the predicates to ontology expressions. **Table 3** provides the OWL DL translation of the RDF predicates. This allows the implementation of automatic translations from a ‘closer to user’ RDF representation into a representation in OWL DL, which would require a more in-depth understanding of DL syntax and semantics. In the following we will describe the use.
of semantic patterns regarding EHR clinical models and ontology-based terminologies as SNOMED CT.

<table>
<thead>
<tr>
<th>Predicate</th>
<th>OWL DL expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘describes situation’</td>
<td>SUBJ subClassOf shn:isAboutSituation only OBJ</td>
</tr>
<tr>
<td>‘results from process’</td>
<td>SUBJ subClassOf btl:isOutcomeOf some OBJ</td>
</tr>
<tr>
<td>‘has attribute’</td>
<td>SUBJ subClassOf shn:hasInformationAttribute some OBJ</td>
</tr>
<tr>
<td>‘has temporal context’</td>
<td>SUBJ subClassOf shn:hasInformationAttribute some OBJ</td>
</tr>
<tr>
<td>‘has situation context’</td>
<td>SUBJ subClassOf shn:hasInformationAttribute some OBJ</td>
</tr>
</tbody>
</table>

Table 2. Subject-Predicate-Object (SPO) triple representation; Italic terms represent ontology classes and terms in quotes represent predicates. Note that predicates are not equivalent to OWL object properties.

Table 3. Translations of RDF predicates into OWL DL axioms within the shared ontology framework

3.2 The role of semantic patterns regarding EHR clinical models and medical domain ontology-based terminologies

Assuming that a limited set of top-level semantic patterns that can be specialized and composed is sufficient to represent a great variety of clinical information, we propose the use of semantic patterns as proxy to the semantic representation of clinical information encoded by EHR structured clinical models and ontology-based medical terminologies. They act as a template, with fix and variable parts, and guide the mapping process in which the correspondences between information model structures and their values are defined with regards to the ontology. Dashed arrows in Fig. 4 indicate the correspondences between the clinical model from Fig. 1 and the history-situation pattern.

As observed, the pattern is applied to both, the SNOMED CT term used as binding and the clinical model information structures. Three correspondences have been provided. Two between the CLUSTER[at0002] binding and the pattern triples that represent the situation and its temporal context. Diabetes mellitus is placed as subclass of shn:ClinicalSituation. One between the value of ELEMENT [at0003] and the pattern triple that represents if the situation is present (sct:KnownPresent) or absent (sct:KnownAbsent). Both are represented as subclasses of sct:FindingContextValue, and will be selected depending of the value of the model instance (True or False).
ENTRY[at0000] matches { -- Question group
  items matches {
    CLUSTER[at0001] matches { -- Question group
      items matches {
        CLUSTER[at0002] matches { -- Question
          items matches {
            ELEMENT[at0003] matches { -- Answer
              value matches {
                BL matches (True, False)
              }
            }
          }
        }
      }
    }
  }
}

Fig. 4. (Left) ISO 13606 archetype and SNOMED CT binding to record the question "past history of diabetes" (Y/N); (Right) Graphical representation of the history-situation pattern

4 Results

In the following we will describe the potential of semantic patterns for each of the tasks introduced in the Methods section. We will use the history-situation pattern as example.

4.1 Semantic patterns provide interoperability across heterogeneously represented data

We will use the history-situation pattern to provide semantic interoperability across two past history data instances captured by two heterogeneous fictitious applications used at a GP consultation and at a hospital. Fig. 5 shows their interfaces. They have been designed attending to different requirements and therefore record the information at different levels of detail. At the hospital (right), the specialist records additional information about the patient past situation (i.e. cause and severity). However, the GP only records the situation itself (left).

Fig. 5. (Left) Past history recording at the GP; (Right) Past history recording at the specialist.

Each of the above applications is based on a different ISO 13606 clinical model. The GP application is based on the questionnaire model introduced in Section 2.1. The left part of Fig. 6 shows the model used by the hospital application. Both are different in terms of structure but not syntax, since both implement the same standard.
In order to access information recorded by both applications, independently of their source representation, the correspondences between each clinical model and the history-situation pattern are defined. Fig. 4 depicted the correspondences between the questionnaire model and the pattern. Following, dashed arrows in Fig. 6 show the correspondences for the hospital model. This model allows recording the severity of the past disease and its cause, requiring the use of the situation pattern, by composition. The situation pattern, allows providing more detail information such as when it occurs, where, associated situations, etc.

Once the correspondences between the models and the patterns are established, when the former ones are instantiated with patient data, the instances of the patterns are also created, in a similar way to the one shown in Fig. 3. If OWL DL instances are created, it is possible to perform homogeneous queries on instances from both applications and retrieve their results [24].

Besides, the use of the ontology framework and DL reasoning allows performing queries at different granularity level: E.g. “Information about all patients with past history of some endocrine disease”, without specifying whether diabetes or a different one.

4.2 Semantic patterns guide the creation of clinical models and detect semantic inconsistencies

Semantic patterns can guide the development of new clinical models if the latter are created by following the constraints dictated by a set of limited top-level patterns.

Top-level patterns are based on a set of generic ontology classes and predicates that can be specialized and composed by following the ontology constraints. These constraints can be used to determine which elements include in a clinical model or in a terminology binding.
As a difference with clinical models, where their elements are only structurally related (e.g. list, tree, etc.), within patterns they are connected by semantic relationships (e.g. shn:isAboutSituation, btl:isOutcomeOf, etc.). These relationships can be used to guide the decision of the elements to include in a model, reducing the existing arbitrariness. Now this is mainly a non-constrained modeller decision that might lead to the creation of non-interoperable models even for the same use case.

If semantic patterns are not applied at clinical models design time, they still can be used to detect semantic inconsistencies across them. As an example, Fig. 7 shows an excerpt of a CIMI model that records observation results. It records: (i) what is observed, ELEMENT[at0001] (e.g. color of the eye); (ii) the reason to perform the observation, ITEM[at0002] (e.g. problem wearing contact lens); (iii) the method used to observe, ITEM[at0003] (e.g. eye examination); (iv) the status of the observation, ELEMENT[at0004] (e.g. performed, planned); and (v) the priority to perform the observation, ELEMENT[at0005] (e.g. high, normal).

Fig. 7. Excerpt of the CIMI model (CIMI-CORE-CLUSTER.observable.v1.0.0) to record observation results

Fig. 8 shows another CIMI model that records observation requests and references the above model by composition (keyword “use_archetype”). Besides, it also references a model to record observation actions. Within this last model we have found a content overlapping with the observation result one, since it also provides elements for recording the reason, method, status and priority of the observation.

Fig. 8. Excerpt of the CIMI model (CIMI-CORE-ENTRY.observation.v4.0.0) to record an observation request and its result
Semantic patterns could avoid such an overlapping situation, by providing formal modelling guidelines, based on the ontological framework, to distinguish across what is observed, the observation procedure and the result of the observation.

Additionally, as already mentioned, they can help to guide or detect inconsistencies regarding terminology bindings. For instance, the pattern logic axiom (shn:InformationItem and shn:isAboutSituation only shn:ClinicalSituation), relates an information entity (i.e. shn:InformationItem) with a clinical entity (shn:ClinicalSituation) and the latter is equivalent to SNOMED CT clinical findings. Therefore, if a model information structure is mapped to that axiom, its value is only valid if it is of the type clinical finding.

When clinical models are instantiated with patient data, semantic patterns can also be used to check that the data entered complies with the constraints defined at the model level.

5 Discussion and conclusions

In this work we have proposed semantic patterns as ontology design content patterns applied to the representation of clinical information. They were motivated by our experiences of representing clinical information using OWL DL, which often resulted in highly complex expressions.

The EHR standards community has put a lot of effort in providing standardized means to represent the EHR. However, the complexity of the medical domain and their heterogeneous data capture and re-use needs does not make it easy. One of the reasons might be the high degree of freedom provided when modelling clinical information, which is mainly formally constrained in terms of structure but without considering the meaning of what is being represented.

Aware of this gap, and concerned about the need of providing standardized modelling means, we propose an ontological framework, in order to represent both information and medical entities, constrained by a top-level ontology which reduces arbitrariness in ontology design. Semantic patterns are based on this framework and therefore constrained by their concepts and relations. In [25], the advantages of using a top-level ontology for creating ontology design content patterns were described, stating that it provides it with an existing backbone structure and well-defined relations.

Semantic patterns provide a more intuitive representation and standardize their development process, yet allowing flexibility through specialization and composition. We have proposed their representation in OWL DL and in RDF. The former one allows logical reasoning and therefore more advanced exploitation of information, although it might be more difficult to implement in a real system, due to performance issues. In the latter case, the RDF representation although less expressive and therefore more limited in terms of information exploitation, might be more adequate. Correspondences between both representations exist, what might allow using the most suitable one for each use case.

In this work we have demonstrated how semantic patterns can be applied to EHR clinical models and ontology-based terminologies (1) to provide semantic interopera-
bility across heterogeneously represented data and (2) commented their potential use to guide the creation of clinical models and detect semantic inconsistencies across them.

By looking at the content patterns available at the NeOn repository [26], we did not find specific patterns for the modelling of clinical information. However, patterns such as the agent-role or the action ones can be applied.

There are numerous new issues that arise from the use of semantic patterns for EHR modelling that still have to be investigated. These include the selection of the right set of patterns to be used for modelling specific pieces of clinical information, who would create and maintain the patterns and who would manage and validate them.

Other issues must be further investigated, such as providing evidence that a set of top-level semantic patterns for modelling clinical information can be rather small, with increasing complexity and expressiveness coming from specialization and composition. So far we have only worked with limited modelling examples and we need more evidence of the real benefit of using patterns; what is hard to obtain without appropriate tools that implement them.

Further research should include the potential of semantic patterns for detecting semantic inconsistencies across existing clinical models, considering their specialization, composition and cardinality constraints. Languages such as SPIN [27] or RDF shapes [28] could be helpful for their representation and are subject of our research.

Acknowledgements. This work has been funded by the SemanticHealthNet Network of Excellence within the EU 7th Framework Program, Call:FP7-ICT- 2011-7, agreement 288408. http://www.semantichealthnet.eu/

References


