Formal Top-Level Ontologies Applied to the Integration of Biomedical Terminologies

Mauricio B Almeida* and Jeanne Louize Emygdio

Graduate Program in Management & Knowledge Organization, Federal University of Minas Gerais, Brazil

Abstract

Issues of integration, the so called lack of interoperability between clinical systems, are a world-wide challenge. While the systems have evolved in controlling technical problems, the question of semantic integration is still complex. We here present an ongoing research that demonstrate the resources and the effectiveness required to interoperate large-scale terminologies, such as SNOMED CT and ICD, within the field of obstetrics.

Keywords: Ontology; Interoperability; Data integration; Biomedical terminologies

*Corresponding author: Mauricio B Almeida, Federal University of Minas Gerais, Belo Horizonte, Brazil

Introduction

The healthcare field is characterized by a complex set of heterogeneous resources. One of the main components of information systems in Medicine, the Electronic Patient Record (EHR), is a complex entitu bringing together several specialties in a single document. The adoption of standards seeks to establish common principles so that these elements can cooperate efficiently and productively, favoring interoperability at multiple levels. However, the proliferation and the aftermath overlapping of standards promotes challenges to practices that involve understanding, adopting, integrating and evaluating. This happens because of the distinct purposes of clinical terminologies, which adds a greater burden to the aforementioned practices [1]. The purpose of this article is to present an ongoing research that demonstrates the efficiency required to effectively interoperate large-scale clinical terminologies, such as SNOMED CT and ICD, in the field of obstetrics. We use a convergence of methods, namely, both linguistic methods and realistic philosophical principles for the representation of knowledge in ontologies.

Epistemological and Ontological Aspects in Clinical Terminologies

The existence of epistemic overlapping in clinical terminologies has been discussed in the biomedical ontology literature since the nineties [2,3]. It refers to the presence of additional information, which could be of clinical relevance, that should not participate in the representation of the real entities.

This issue can be observed, for example, in SNOMED CT:

A.   "Natural death with suspected probable cause" allows the communication of a clinical imprecision about a diagnosis of death

B.   "Heart disease excluded" reflects a momentary conviction of the physician and not the nature or severity of a diagnosis

C.   "Operation in the heart, re-scheduled" communicates the intention to change in the situation of a process that has not yet occurred. The presence of this additional information leads to the definition of classes that do not comply with classification

principles, negatively impacting in the terms for alignment, mapping, integration and their evaluation [2].

The use of ontological principles, such as those of the Basic Formal Ontology (BFO), is based on the application of the ontological sextet [4], associated with the rules of construction of quality taxonomies. Considering these aspects, we can increase the chances of accurate discoveries of semantic relations, through both lexical and structural anchors, which characterize shared borders of knowledge between the clinical terminologies. In doing this, we can also open space for the definition of quality alignments, saving time and resources because of the comparisons of ontologically elucidated terms.

Methodology

The methodology takes a qualitative approach, providing quality principles in two levels:

a) construction and distribution of realist ontologies, based on the principles of OBO Foundry and BFO

b) integration method, based on the indirect alignment method [5].

The strategy of combined approaches for alignment and integration between terminologies were implemented in two phases: i) direct alignment between clinical terminologies and formal ontologies, resulting in integration ontologies BFO-SCT(Oi1) and BFO-CID(Oi2) and, among the clinical terminologies themselves, resulting in the integration ontology SCT-CID(Oi3) and ii) indirect alignment between terms of the ontologies BFO-SCT(Oi1) and BFO-CID(Oi2), plus the addition of new classes, axioms and annotations, if necessary, resulting in the integration ontology SCT-CID(Oi4).

The comparison of terms and relations requires four recurring approaches for interoperability: i) ontology matching, ii) ontology mapping, iii) ontology alignment and iv) ontology integration.

Finally, to perform these approaches, four tasks will be necessary:

A. Acquisition of terms

B. Identification of lexical combinations (lexical anchors)

C. Identification of semantic relations

D. Identification of structural anchors (structural similarity).

The analysis of structural similarity allows the discovery of positive evidences for the accomplishment of the alignments and integration, as well as the conflicts between the representations that indicate semantic incompatibility between terminologies. The efficiency of each alignment method to more accurately interoperate clinical terminologies is calculated, at this phase, as a measure of the proportion of positive evidence for alignments over the total number of combinations found: efficiency=(number of positive evidences for alignments ÷ total combinations found)×100.

Preliminary Results

An sample of 2218 terms from the Ontology for the Obstetric and Neonatal Domain (Onto NEO) [6] was established for the execution of the research. Other tree ontologies have been used: Foundational Model of Anatomy (FMA), Information Artifacts Ontology (IAO) and Ontology for General Medical Science (OGMS). The established conventions and rules for alignments and integrations should follow the principles of OBO Foundry and BFO. The production environment will be composed of Protegé ontology editor, and plugins: Bio Portal Import Plugin, HermiT (verification of inconsistencies by reasoning), OWL2 Query (queries in SPARQL), YAM ++ (discoveries of combinations). The Ontofox and PROMPT tools were applied for importing terms and properties and comparing terminologies. There are additional studies underway on the Snow OWL (Snomed CT browser for Protegé) and OWL Diff (ontology comparison) plugins. The production process initially takes human alignment. For the definition of the aforementioned Oi1, Oi2, Oi3 and Oi4, the direction of the alignments departs from BFO to SCT, BFO to CID, and from SCT to CID, respectively.

The validation process of the alignments and integrations is carried out in two stages:

a) Internally, through the verification of inconsistencies and errors of inferences

b) Externally, from the verification of specialists in the healthcare area. We do not have quantitative results yet, but as expected, any task regarding integration and interoperability is hard, both from the computational and human resources.

References

1. Schulz S, Stegwee R, Chronaki C (2019) Standards in healthcare data. In: Kubben P, Dumontier M, Dekker A (Eds.), Fundamentals of Clinical Data Science, Springer International Publishing, pp. 19-36.

2. Bodenreider O, Smith B, Burgun A (2004) The ontology-epistemology divide: A case study in medical terminology. Form Ontol Inf Syst 185-195.

3. Rector AL (1999) Clinical terminology: Why is it so hard? Methods Inf Med 38(4-5): 239-252.

4. Smith B (2005) Against fantology. In: Reicher M, Marek JC (Eds.), Experience and analysis, Vienna: HPT&ÖPV, pp. 153-170.

5. Zhang S, Bodenreider O (2005) Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. AMIA Annu Symp Proc 864-868.

6. Almeida MB, Farinelli F (2017) Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. Journal of the Association for Information Science and Technology 68(11): 2529-2542.

For possible submissions Click below:

Submit Article