Analyzing the Evolution of Semantic Correspondences between SNOMED CT and ICT-9-CM

Julio Cesar Dos Reis, Cédric Pruski, Marcos Da Silveira, Chantal Reynaud-Delaître

Analyzing the Evolution of Semantic Correspondences between SNOMED CT and ICD-9-CM

J.C. Dos Reis[1,2], C. Pruski[1], M. Da Silveira[1], C. Reynaud-Delaître[2]
[1]CR SANTEC, Public Research Centre Henri Tudor julio.dosreis@tudor.lu, cedric.pruski@tudor.lu, marcos.dasilveira@tudor.lu
6, Av. des hauts fourneaux, L-4362, Belval - Luxembourg
[2] LRI, University of Paris-Sud XI
chantal.reynaud@lri.fr
PCRI, bat 650, Université Paris-Sud rue Noetzlin 91190 Gif-sur-Yvette, France

Abstract: **The combined use of Knowledge Organizations Systems (KOS) including ontologies, terminologies or codification schemas has widespread in e-health systems over the past decades due to semantic interoperability reasons. However, the dynamic aspect of KOS forces knowledge engineers to maintain KOS elements, as well as semantic correspondences between KOS up-to-date. This is crucial to keep the underlying systems exploiting these KOS consistent over time. In this paper we provide a pragmatic analysis of the evolution of mappings between SNOMED CT and ICD-9-CM affected by the evolution of these two KOS.**

Introduction

The growing quantity of the produced medical data requires a new generation of tools to exploit this data to reduce costs and optimize the quality of care. To this end, automatic interpretation of data, information retrieval and sharing are of utmost importance. They implement Knowledge Organization Systems (KOS) like ontologies, thesaurus or classification schemas to automatize the treatment of digital information. However, the size and characteristics of the domain make impossible the definition of a single KOS able to represent the entire medical knowledge. This is why managers of information systems are forced to use a combination of KOS in order to optimize the coverage (*e.g.*, use ICD-9-CM (ICD) for classifying diseases or SNOMED-CT (SCT) for clinical knowledge). This is done through the definition of semantic correspondences (or mappings) between KOS [1]. It consists in defining the semantic relation (*e.g.*, equivalence, subsumption) that exists between elements (*e.g.*, concepts) belonging to different KOS. But, due to the dynamic aspect of the medical knowledge [2], KOS have to follow this evolution and are likely to be modified over time, which can potentially invalidate previously created mappings. In consequence, modifications occurring in KOS must be propagated to mappings in order to keep the underlying information systems consistent over time [3].

We focus on the mapping maintenance problem (*i.e.*, the adaptation of existing mappings according to the changes affecting underlying KOS elements). Our final objective is to propose a formal framework able to propose new semantic correspondences between concepts of medical KOS when those evolve [4]. The goal is to automatize as much as possible the maintenance process reducing the validation time as well as the quantity of errors. We have empirically analyzed the evolution of the mappings between several KOS in order to identify correlation between the way KOS evolve and the impact of this evolution on the behavior of mappings. In this paper we report on this aspect. We have investigated 4 successive versions of SCT, 2 versions of ICD and 4 versions of their associate set of official mappings. Our study highlights such correlation between the way KOS evolve, and especially the type of change affecting elements (*e.g.*, split or merge of concepts), and the way mappings behave.

Understanding Mappings Maintenance

Our first set of experiments consisted in analyzing successive releases of ICD in order to identify the various types of changes that can affect its elements with a particular attention paid to concepts. The analysis of the obtained results show that two kinds of changes can occur:

- *Atomic changes* like the addition or removal of concepts [5].
- *Complex changes* corresponding to a combination of atomic changes.

In our work, we focus on complex changes. We observed that most of complex changes can be characterized as merge or split of concepts and can have various forms (see figure 1 below). These experiments revealed 9 types of major changes (7 complex and 2 atomic modifications) leading to a release of a new version of ICD. In the figure 1, *S* and *T* represent concepts; *Sim* indicates the existence of a similarity between the two involved concepts while *code* denotes the concept ID in its respective KOS.

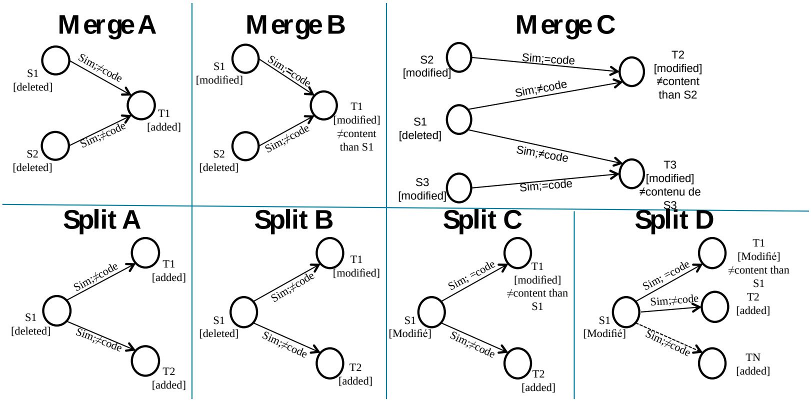

Figure 1: Types of complex changes affecting ICD concepts

Through the second set of experiments we have conducted, we crossed the results obtained during the first phase, and the analysis of the mappings evolution in order to understand the impact of KOS evolution on mappings. We, therefore, analyzed the impact of complex changes on the way mappings behave over time. To do so, we have isolated the concepts of ICD involved in one of the seven complex changes, which contained a mapping with an element of SCT. Then, we regrouped the elements of ICD that have evolved according to the same type of complex change. Afterwards, we analyzed how their associated mappings have evolved. As results, we have identified three types of changes for mappings (see table 1 below in which *(s, t, r)* represents a mapping, *s* denotes the source concept, *t*, the target concept and *r* the semantic relation between *s* and *t*).

Table 1: Behavior of mappings

Type of behaviour	**Formalization**	**Link with types of changes in KOS**
Full transfer	Before evolution *(s, t, r)* After evolution *(s', t, r)*	100% (2/2) Split C 27% (3/11) Split D 28% (16/57) Addition of concept
Partial transfer	Before evolution *(s, t, r)* After evolution *(s', t, r')*	55% (6/11) Split D 14% (8/57) Addition of concept
Full duplication	Before evolution *(s, t, r)* After evolution *(s, t, r)* ∧ *(s', t, r)*	18% (2/11) Split D 58% (33/57) Addition of concept

Results (Table 1) show that the strength of the link between the type of changes affecting KOS concepts and the behavior of mappings varies. In consequence, a split of type C gives always place, according to our observations, to a full transfer of mappings. In most of the cases, a split of type D gives place to a partial transfer, while an addition of concepts causes a full duplication of mappings. Additional investigation on the formal definition of the type of changes affecting KOS concepts is still needed in order to fully define the previously evoked link, as well as the way an automatic process for maintaining mappings can exploit it. Ongoing work on similarity measures between two successive versions of a concept would bring more information in order to better characterize this phenomenon.

Conclusion

In this paper we have highlighted, through a set of experiments, the role played by the type of evolution affecting KOS in the maintenance of mappings. Our future work will focus on the formal definition of change patterns for a better characterization of KOS evolution, as well as a set of heuristics for driving the maintenance of mappings. We also plan to integrate other KOS like NCI thesaurus or MedDRA to understand the importance of the KOS model in the mapping maintenance problem.

Acknowledgment

The research presented in this paper has been entirely funded by the National Research Funds of Luxembourg.

References

[1] J. Euzénat, and P. Shvaiko, “Ontology Matching”, 2007.

[2] A. Baneyx and J. Charlet, “Évaluation, évolution et maintenance d’une ontologie en médecine : état des lieux et propositions méthodologiques” Revue Information - Interaction – Intelligence, 2006 (in French).

[3] L. Stojanovic “Methods and Tools for Ontology Evolution,” PhD thesis Universitat Karlsruhe (TH), Institut AIFB, D-76128 Karlsruhe, 2004.

[4] J.C. Dos Reis, C. Pruski, M. Da Silveira, and C. Reynaud-Delaître, “Analyzing and Supporting the Mapping Maintenance Problem in Biomedical Knowledge Organization Systems,” in the proceedings of the Semantic Interoperability in Medical Informatics (SIMI) workshop, 2012.

[5] M. Klein and N. Noy, “A Component-Based Framework for Ontology Evolution,” in the proceedings of Workshop on Ontologies and Distributed Systems at IJCAI-03, 2003.