Web Semantics: Science, Services and Agents on the World Wide Web 29 (2014) 53–66

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

# Identifying relevant concept attributes to support mapping maintenance under ontology evolution

Duy Dinh [a,*], Julio Cesar Dos Reis [a,b], Cédric Pruski [a], Marcos Da Silveira [a], Chantal Reynaud-Delaître [b]

[a] Public Research Centre Henri Tudor, 6 avenue des Hauts-fourneaux, L-4362 Esch-sur-Alzette, Luxembourg
[b] LRI, University of Paris-Sud XI, Bât 650, 91405 Orsay Cedex, France

ARTICLE INFO

*Article history:*
Received 1 August 2013
Received in revised form
3 April 2014
Accepted 27 May 2014
Available online 9 June 2014

*Keywords:*
Ontology mappings
Mapping adaptation
Ontology evolution
Semantic similarity
Biomedical ontologies

ABSTRACT

The success of distributed and semantic-enabled systems relies on the use of up-to-date ontologies and mappings between them. However, the size, quantity and dynamics of existing ontologies demand a huge maintenance effort pushing towards the development of automatic tools supporting this laborious task. This article proposes a novel method, investigating different types of similarity measures, to identify concepts' attributes that served to define existing mappings. The obtained experimental results reveal that our proposed method allows to identify the relevant attributes for supporting mapping maintenance, since we found correlations between ontology changes affecting the identified attributes and mapping changes.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

The evolution of semantic technologies has led to the development and publication of a huge amount of ontologies, allowing information systems to better describe data and search for relevant information on the Web. Ontologies offer means to make the semantics of data explicit which, in turn, facilitates its exploitation and management. However, mainly for semantic interoperability issues, we need to establish semantic correspondences between ontologies, named mappings, to allow software applications to explore data annotated using various ontologies. The ever increasing number of large ontologies underlines the major role played by mappings [1].

The dynamic nature of domain knowledge induces continuous changes in existing ontologies like *adding, removing* or *modifying* ontology elements (*e.g.*, classes, properties, *etc.*) [2]. These changes impact dependent artefacts such as mappings, making them invalid. In consequence, domain experts must repair affected mappings taking ontology changes into account. This laborious task consists in identifying changes affecting elements of ontologies, and to adapt mappings impacted by these changes accordingly. This process can be performed manually on small ontologies with a restricted number of mappings, but large and highly dynamic ontologies, like those of the life sciences, require appropriate methods and automatic tools.

Existing tools compute mappings between concepts in an (semi-)automatic way to create semantic correspondences between them [3]. Although concepts are considered in their entirety, a closer empirical analysis of mappings reveals that only partial textual statements characterizing concepts are used to define the semantic correspondences [4,5].

When analysing two consecutive versions of the same ontology, we found cases, for instance, where concepts' attribute values are completely transferred from one concept to its siblings. This had affected the associated mappings since their definition relies on such textual statement. For example, we observed this case with the concept "560.39" of the ICD-9-CM[1] (ICD) biomedical ontology. Such concept contains three attributes and one of them has as value "Fecal impaction" (release 2009). Five mappings are defined with this concept as domain, and one of these mappings has a range called "Fecal impaction (disorder)", from SNOMED CT[2] (SCT).

* Corresponding author. Tel.: +352 425991.
*E-mail addresses:* dinhbaduy@gmail.com, duy.dinh@tudor.lu (D. Dinh), julio.dosreis@tudor.lu (J.C. Dos Reis), cedric.pruski@tudor.lu (C. Pruski), marcos.dasilveira@tudor.lu (M. Da Silveira), chantal.reynaud@lri.fr (C. Reynaud-Delaître).

http://dx.doi.org/10.1016/j.websem.2014.05.002
1570-8268/© 2014 Elsevier B.V. All rights reserved.

[1] http://www.cdc.gov/nchs/icd/icd9-cm.htm.
[2] http://www.ihtsdo.org/snomed-ct.

After evolution (*i.e.*, ICD release 2010), the attribute value "Fecal impaction" is no longer associated with the ICD concept and the previously mentioned mapping has been removed. Moreover, the concept "Fecal impaction" has been newly created in ICD (release 2010) and is remapped to "Fecal impaction (disorder)" of SCT. This illustrates the major role played by concepts' attributes in the definition of mappings [5]. However, when matching systems create a mapping, they fail to keep the ontological entities used to justify such mapping in its definition, preventing thus any future use for maintenance purpose.

In this article, we address this issue by investigating techniques suited to identify textual statements in concepts that might represent the most meaningful attributes for a given correspondence, but lack in its original description. We hypothesize that adequately supporting the mapping adaptation task requires the correct identification of these statements [5,6]. Our identification method relies on the adaptation of various semantic similarity measures targeting: the *lexical* level [7], the *syntactic* level [8] and the *semantic* [9] level. These measures might support our method to identify a sufficient **S**ubset of **C**oncept **A**ttributes (SCA) relevant for interpreting mappings. We conduct a set of experiments to assess the quality of the results yielded by the identification method using two life science ontologies (SCT and ICD) and their associated mappings. In particular, we measure correlations between ontology changes affecting the identified attributes and adaptation of associated mappings. We further study the stability of the underlying similarity measures analysed.

We structure the remainder of this article as follows: Section 2 presents the related work. Section 3 introduces the preliminaries. Section 4 presents our approach to identifying the relevant attributes. Section 5 presents the experimental evaluation while Section 6 discusses the obtained results. Section 7 wraps up with concluding remarks.

## 2. Related work

Semantic interoperability among heterogeneous systems has increasingly pushed the efforts on semi-automatic matching approaches to aligning ontologies, by finding correspondences between concepts [10]. Ontology mappings play a key role in the biomedical domain [11] where various efforts such as Bioportal [12] and UMLS [13] aim at supporting interoperability.

The *Unified Medical Language System* (UMLS) [13], developed by the *U.S. National Library of Medicine*,[3] aims at facilitating the exchange of clinical data and improve retrieval of health information. UMLS includes over than one hundred of ontologies and mappings between them. For example, UMLS also integrates the mappings between SNOMED-CT and ICD9-CM which are created and maintained in collaboration with the IHTSDO organisation.[4] These mappings support a transition from the use of legacy ICD9-CM procedure codes to SNOMED-CT.[5]

A large number of works have investigated approaches to ontology matching and alignment [14–17]. Various approaches focus on string-based similarity metrics to establish mappings between ontologies. Some surveys have reported on the performance of these metrics in the course of developing ontology alignment systems [18–20].

Indeed, despite advancements over the last years on ontology alignment, many issues remain open and represent a real challenge for the Semantic Web community [1]. Unaddressed aspects such as user's interaction [21] and crowdsourcing [22] have recently gained interest and the mapping maintenance under evolving ontologies remains an open research problem [23]. To the best of our knowledge, only few studies have investigated fully automatic methods to keep ontology mappings semantically valid over time.

We distinguish three main categories of approaches for mapping maintenance. The first one relies on the revision of mappings by identifying and repairing invalid mappings. For example, Meilicke et al. [24] propose an automatic debugging of mappings between expressive ontologies eliminating inconsistencies, caused by erroneous mappings, by means of logical diagnostic reasoning. Similarly, Castano et al. [25] suggest a probabilistic reasoning approach to performing the validation of mappings. These techniques can be applied after ontology evolution to detect invalid mappings. However, they require logically expressive ontologies at a high level of formalization, making this approach unavailable for information systems that rely on semantic resources of low level of formalization such as nomenclatures, thesauri, *etc.*

The second category performs a *full* or a *partial* re-calculation of mappings. While the former fails to consider any information from ontology evolution nor existing mappings, the latter aims at exploiting those information for recreating only mappings that are associated with changed concepts in ontologies. In addition, if large ontologies are frequently released, fully re-calculating mappings becomes less flexible than a partial re-calculation approach because the cost in terms of processing time for re-aligning ontologies still remains too expensive. Khattak et al. [26] propose a partial re-calculation approach, which re-creates only those mappings associated with concepts whose elements have changed. They use matching algorithms to perform a new alignment between changed concepts issued from source ontology and the whole target ontology. However, large ontologies, as in the biomedical domain, so far represent a big challenge for methods relying on mapping calculation [1].

The third category concerns approaches that attempt to adapt semantic mappings in response to ontology evolution. These approaches usually use ontology changes to support mapping adaptation, avoiding to perform calculations for re-aligning ontologies. The first propositions appeared in the context of database schema mappings [27] based on primitive schema changes. Composition of mappings [28] explores mappings between different schema versions for adapting mappings. Concerning ontologies, Tang & Tang [29] propose a method for ontology evolution to find the minimal impact of ontology change propagation. Nevertheless, they assume that only removal of axioms can impact mappings. Martins & Silva [30] propose that the evolution of mappings should behave similarly with the strategies applied for ontology evolution, but their method only adapts mappings when concepts are removed from the ontology. More recently, researches have empirically investigated the evolution of life science ontology mappings to understand the mapping evolution phenomenon [31,5]. On this basis, preliminary investigations have proposed ontology changes-based techniques to adapt mappings [32].

Despite these recent investigations, it still lacks adequate methods to fully perform mapping adaptation in an automatic way according to ontology evolution. Mapping adaptation under evolving ontologies should consider the whole set of possible types of ontology changes rather than only concept removal. We have defined the *DyKOSMap* framework [33,6] for handling the adaptation of mappings based on ontology changes and mapping interpretation. We believe that the similarity between concept's attributes that have been changed constitutes a key factor that we might take into account in automatic mapping adaptation. Beyond the literature, this article proposes a novel technique for identifying relevant attributes that we could exploit for adapting mappings under evolving ontologies.

[3] http://www.nlm.nih.gov.

[4] http://www.ihtsdo.org/snomed-ct.

[5] http://www.nlm.nih.gov/research/umls/mapping_projects/icd9cmv3_to_snomedct.html.

Download English Version:

https://daneshyari.com/en/article/557796

Download Persian Version:

https://daneshyari.com/article/557796

Daneshyari.com