Information Systems 38 (2013) 727–744

Contents lists available at SciVerse ScienceDirect

Information Systems

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

An approach for sub-ontology evolution in a distributed health care enterprise

Anny Kartika Sari [a,b,*], Wenny Rahayu [a], Mehul Bhatt [c]

[a] Department of Computer Science and Computer Engineering, La Trobe University, Victoria 3086, Australia
[b] Department of Computer Science and Electronics, Gadjah Mada University, Yogyakarta 55281, Indonesia
[c] SFB/TR 8 Spatial Cognition, Informatics, University of Bremen, Germany

ARTICLE INFO

Available online 25 April 2012

Keywords:
Ontology evolution
Change propagation
Sub-ontology
Health
Semantics of change

ABSTRACT

In response to the changing nature of health issues, standardized health ontologies such as SNOMED CT and UMLS incline to change more frequently than most other domain ontologies. Yet, semantic interoperability shared among institutions within a distributed health care enterprise relies heavily on the availability of a valid and up-to-date standardized ontology. In this paper, we propose the creation and preservation of sub-ontologies to deal with the frequent changes in health ontologies. Our approach focuses on the nature and characteristics of standard health ontologies, however it can also be applied to other domain ontologies with similar characteristics. Our sub-ontology evolution approach defines ways to create valid sub-ontologies for each specific health application, and to effectively develop a series of propagation mechanism when the main ontology changes. Our approach will (i) isolate the required change propagation to the relevant health applications that utilized the changing concepts only, and (ii) optimize the propagation mechanism to include the minimum number of operations only. Since a sub-ontology should be a valid ontology by itself, the change propagation approach used in this process should contain the rules to assure the validity of the produced sub-ontology while keeping the consistency of the sub-ontology to the evolved base ontology. A change identification process, which considers the nature of the health ontology change logs, is conducted to identify the semantics of the changes. From the evaluation, it is shown that the content of the evolved sub-ontologies produced using our approach is consistent to the evolved base ontology. Moreover, the propagation process can be performed more efficiently because the number of operations required for our change propagation method is lower than the number of operations required for direct re-extraction from the evolved base ontology.

1. Introduction

Health care providers range from hospital to specific care centers such as rehabilitation centers. Patients can choose or be recommended to different health care providers to achieve their health goals. The increased mobility of people often results in them receiving health treatment from caregivers who are geographically separated. In these conditions, interoperability among different health providers is vitally important so that the medical records for each patient can be preserved and exchanged between providers. Brailer in [1] believes that the consumer can suffer from a lack of interoperability and health information exchange because the health care enterprise hopes to gain a comparative advantage by imposing high costs when consumers change health care providers. According to Dogac et al. [2], full 'share-ability' of

* Corresponding author at: Department of Computer Science and Computer Engineering, La Trobe University, Victoria 3086, Australia. Tel.: +61 420852501.
E-mail addresses: aksari@students.latrobe.edu.au, a_kartikasari@ugm.ac.id (A.K. Sari).

0306-4379/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.is.2012.03.006

data and information requires two levels of interoperability: semantic interoperability and functional (syntactic) interoperability. Our focus is on semantic interoperability, which is defined as the ability for information shared by systems to be understood at the level of formally defined domain concepts so that the information is computer processable by the receiving system [3].

It is believed that ontologies are one way to overcome the semantic interoperability problem between different health care providers. Using ontology, the semantic meaning of each health term can be uniformly interpreted. An example of the use of ontologies for health care applications is the binding between the archetype terms and the ontology concepts. Several health ontologies such as SNOMED CT (Systematized Nomenclature of Medicine—Clinical Terms), LOINC (Logical Observation Identifiers Names and Codes) and UMLS (Unified Medical Language System) can be referred to by archetype terms. Archetype has been proposed by the openEHR[1] Foundation as a model of specific domain knowledge. This model has also been adopted by the CEN TC/2512 in its Health informatics—Electronic Health Record Communication (EN 13606) European Standard. The binding between the archetype terms and the ontology concepts is aimed at achieving semantic interoperability between different health care institutions which may use different electronic health record standards. Health terminologies are also used in a similar way in the HL7 (Health Level Seven) standard. Externally defined terms and codes such as SNOMED CT can be utilized in HL7 as an Instance Identifier, which is used to give a unique identity to people, persons, organizations, things and information objects.[2]

Since ontologies aim for standardization, their size is usually very large. Many ontologies in the health domain such as SNOMED CT, LOINC and especially UMLS have hundreds of thousands and even millions of concepts. For a specific health application, the use of the whole ontology is not appropriate since actually, only a small part of the ontology is relevant to the application. For example, the health information system of a pharmacy, which uses SNOMED CT as the base ontology, requires only terms related to drugs, while there are many more terms in SNOMED CT, such as terms related to the examination of patients and medical procedures, which obviously are not relevant to the application. For an application with such a specific focus, sub-ontologies can be utilized instead of the whole ontology.

A sub-ontology is a subset of an ontology derived from that base ontology using a specific extraction process. A characteristic which differentiates a sub-ontology from a subset is that a sub-ontology should be a valid ontology in its own right [4]. A sub-ontology refers to a particular part of a base ontology which is appropriate to a specific context, user, specialty, etc. An example of the use of a specific context of knowledge in the clinical domain is the concept of archetype, which has been mentioned previously. An archetype describes a complete clinical knowledge concept such as ‘diagnosis’ or ‘test result’ [5]. An archetype may contain clinical terms which refer to the terms in health ontologies such as SNOMED CT and LOINC. Sari et al. propose the use of *archetype sub-ontology* to represent the semantic content of an archetype [6]. The archetype sub-ontology is extracted from the health ontology. Similarly, Yu et al. [7] has proposed a kind of sub-ontology referred to as the *Terminological Shadow*, which is derived from SNOMED CT, to represent the semantic content of an archetype. These studies show the applicability of the use of sub-ontologies in the health domain.

Ontology and sub-ontology should represent the current knowledge in the domain. When the knowledge changes, they should be adjusted. This process is known as ontology evolution and is one of the prominent issues in the use of ontologies in the health domain as it is one of the domains in which knowledge changes frequently. This is shown by the high frequency of health ontology changes. As an example, in each version of SNOMED CT, which is released twice a year, the average number of changes is more than 50,000 which consist of additions (45.45%), status changes (30.87%), and minor changes (23.68%) [8].

When an ontology evolves, the sub-ontologies derived from it should be adjusted as well so that they are consistent with the base ontology. Re-extraction based on the evolved ontology can be the simplest method to maintain the consistency of the sub-ontologies. However, this approach is not practical when the number of sub-ontologies is high and the changes in health ontologies occur frequently. In this case, it is more appropriate to change the sub-ontologies according to the changes which have taken place in the base ontology. In the notion of ontology evolution, this process is known as *change propagation*.

In addition to ensuring consistency to the base ontology, another important reason for the change propagation process to occur in sub-ontologies is that sub-ontologies should be kept valid. In other words, it should be assured that the evolved sub-ontologies produced from the change propagation process are the same as the ones extracted directly from the evolved base ontology. Rules are needed to determine which changes should be propagated to the sub-ontologies to avoid the sub-ontologies from becoming too big or too small but, at the same time, keep its semantic content. Most of the existing change propagation approaches [9–11] do not consider this requirement because they have not been applied to sub-ontologies. Moreover, the approaches are usually based on the assumption that the semantics of the changes are already known from the version log which contains the list of changes which occur from a previous version of the ontology to the next one. This is not appropriate for health ontologies as most of them do not provide version logs containing a list of changes which are semantically meaningful. For example, SNOMED CT provides a list of basic change operations such as additions and deletions of concepts and descriptions, while they actually permit more complex change operations such as the movement of concept. To capture such types of changes, an approach to identify the semantics of change is needed which should be based on the nature of health ontologies.

In this paper, we propose the use of sub-ontologies as a way to simplify ontology evolution management in a distributed health enterprise. The main issue addressed in this work is the change propagation mechanism from the base

[1] See: http://www.openehr.org/.

[2] See: http://www.hl7.org.au/HL7-V3-Resources.htm.