9 July 2011

A Methodology For Change Propagation In Health Ontology

Anny Kartika Sari
La Trobe University, aksari@students.latrobe.edu.au

Wenny Rahayu
La Trobe University, W.Rahayu@latrobe.edu.au

ISBN: [978-1-86435-644-1]; Full paper

A METHODOLOGY FOR CHANGE PROPAGATION IN HEALTH ONTOLOGY

Anny Kartika Sari, Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia, aksari@students.latrobe.edu.au; Department of Computer Science and Electronics, Gadjahmada University, Yogyakarta, Indonesia

Wenny Rahayu, Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, Australia, W.Rahayu@ latrobe.edu.au

Abstract

Health ontologies are commonly used as the standardization for interoperability between different health institutions. Since health knowledge changes rapidly, health ontology also evolves frequently. A mechanism of change propagation is needed to maintain the consistency between the ontology with the dependent artefacts.

In this paper, we present the classification of health ontology changes based on their semantic meaning. Based on the classification, we define the basic operations used to perform the changes. Using those basic operations, we develop a method to propagate the ontology changes to one type of the dependent artefacts, i.e. the sub-ontologies which are extracted from the health ontology.

We have applied the method to a case study which is based on the SNOMED CT ontology - one of the ontologies mostly used in health information systems. Our propagation mechanism will minimize the need to extract and re-create sub-ontologies from the base ontology every time the base ontology evolves. We also demonstrate that our approach produces a consistent sub-ontology state with the base ontology at any given time by comparing the propagated sub-ontology with the evolved SNOMED CT base ontology.

Keywords: health, ontology, SNOMED CT, evolution, change propagation.

1 INTRODUCTION

Healthcare is one of the domains with very complex knowledge. Health information systems cover many types of data, from the data of patient and treatment to the data of health organisational management. The complexity leads to the diversity in the vocabularies as well as in the semantic meaning of the data. Interoperability problems will arise between different health institutions which need to exchange their data. Different health workers may have different understanding on the same term used to represent the data.

Ontology is one of the ways to overcome the interoperability problem. Many ontologies in the health domain have been established. Among them are SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), LOINC (Logical Observation Identifiers Names and Codes), and GALEN. Using these ontologies, semantic meaning of the terms can be made uniform. Since the ontologies are aimed for standardization, they should contain complete knowledge of the domain. Because of this reason, the size of an ontology can be very large even though in most cases an application will need only a subset of the ontology. For such an application, the use of the whole ontology is a waste in terms of the storage and the processing time. It is more efficient to use only the relevant part of the ontology, which is called *sub-ontology*. For example, to represent the semantic content of an *archetype*, which describes a complete clinical knowledge concept such as 'diagnosis' or 'test result' (Leslie 2006), sub-ontology is more suitable to be utilized rather than the whole ontology. In fact, Yu et al. (2010) has proposed a kind of sub-ontology referred to as *Terminological Shadow*, which is derived from SNOMED CT, to represent the semantic content of an archetype. Some other works such as in Qamar & Rector (2006) and Sari et al. (2010) also presented the binding mechanism between archetype terms with SNOMED CT concepts. The binding can be considered as the initial step of the development of the sub-ontology.

As health knowledge is not static, which means that it evolves from time to time, health ontology also evolves. For instance, the content of SNOMED CT evolves with each release (IHTSDO 2010). There are on average two releases each year, which means that changes of SNOMED CT are made twice in a year. From the Component History Table of the 20100731 SNOMED CT International Release version, we calculated the number of changes in each year since the 20020131 first release version, and it is presented in Table 1. The average number of changes for each release is more than 50,000 which include addition (45.45 %), status change (30.87 %), and minor change (23.68 %). This shows that the changes of SNOMED CT happen in high frequency.

Year	Release Version	Number of Changes
2002	20020731	97,625
2003	20030131 and 20030731	134,936
2004	20040131 and 20040731	90,861
2005	20050131 and 20050731	36,722
2006	20060131 and 20060731	96,396
2007	20070131 and 20070731	73,636
2008	20080131 and 20080731	242,477
2009	20090131 and 20090731	80,270
2010	20100131 and 20100731	35,525
Total		888,488

Table 1. The number of changes of SNOMED CT in each year after the first release date (31 January 2002) of SNOMED CT International Release

Figure 1 illustrates some sub-ontologies which are derived from SNOMED CT ontology and used to represent the semantic content of different archetypes. Archetypes are used as a guideline for medical practitioners to do specific treatment to the patients via computer-based user interfaces. When SNOMED CT evolves due to the changes in health knowledge, the sub-ontologies must evolve as well so that they keep their consistency with the base ontology, i.e. SNOMED CT, and at the same time can trigger the adjustment on the archetypes to keep them synchronised with the new knowledge. In health domain, it is very important to keep the knowledge up-to-date as it might be critical to human life. For instance, a sub-ontology representing the content of the *lab_test-liver_function*[1] archetype is derived from the previous version of health ontology. New knowledge finds out that a new type of examination should be conducted to know the functionality of a liver. Since the sub-ontology has not been updated, the existing archetype has not captured the new knowledge. As a result, the new type of liver examination is not conducted to the patients while it might be the key result in determining the treatment to the patients.

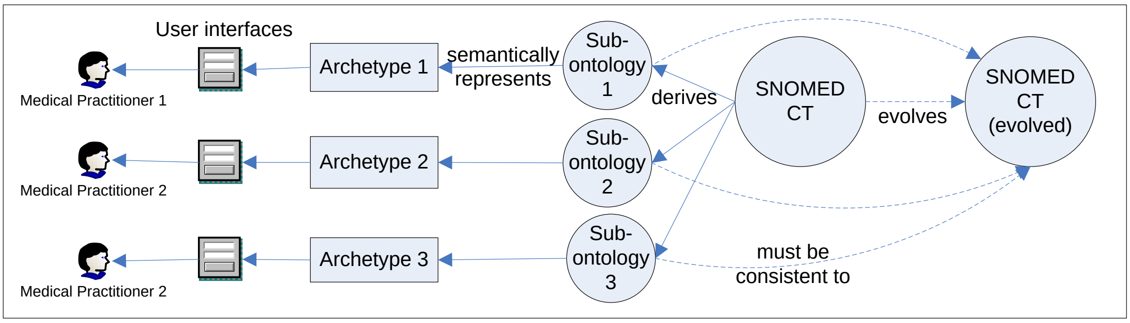

Figure 1. *Sub-ontologies derived from SNOMED CT to represent the semantic content of archetypes must be consistent with the evolved base ontology (SNOMED CT).*

To keep the applications in health domain consistent with the underlying knowledge, the changes of the health ontology should be transmitted to all dependent artefacts, including the sub-ontologies which were derived from it. This process is referred to as the *change propagation*. To the best of our knowledge, currently there is no mechanism for change propagation in the case where the base ontology is very dynamic such as the health ontology. The simplest method to maintain the consistency of the sub-ontology is to re-extract the sub-ontology from the evolved base ontology. However, this approach is not practical as the number of sub-ontologies might be very high and the changes in health ontology such as SNOMED CT are as frequent as twice a year.

In this work, a method in propagating the changes to the sub-ontologies is proposed. The approach makes use of the history mechanism provided by the ontology. To build the framework, we first present the evolution mechanism and a classification of ontology changes. Our classification is based on the nature of the changes in health domain's ontologies. A case study is also presented to evaluate the feasibility of the proposed approach as well as to show that the method can keep the consistency of the sub-ontologies with the evolved base ontology.

The rest of this paper is organized as follows. Related work is discussed in Section 2. It is followed by the description of the sub-ontology and the evolution mechanism of SNOMED CT in Section 3. The discussion of health ontology evolution, which contains the classification of changes and a proposed method to propagate the changes, is presented in Section 4. Section 5 presents a case study and discussion. Section 6 concludes our work.

2 RELATED WORK

Stejanovic et al. (2002a) define 6 phases of ontology evolution: change capturing, change representation, semantics of change, change implementation, change propagation, and change validation. Our work is focusing on the change propagation phase. The task of this phase is to bring all dependent artefacts in a consistent state, i.e. the state after the ontology change has been performed, by propagating changes to these depending artefacts (Plassers & Troyer, 2005). In this

[1] This archetype is displayed in the openEHR Clinical Knowledge Manager at: http://www.openehr.org/knowledge/

work, the depending artefacts are the sub-ontologies extracted from SNOMED CT ontology. We chose SNOMED CT as the model of our health ontology because according to Cornet & de Keizer (2008), there is an increasing awareness of SNOMED CT and its development and implementation. The usage of SNOMED CT as a standard for the electronic health record, which is related to the interoperability problem experienced in health domain as we mentioned previously in this paper, is an emerging subject of study. Furthermore, some characteristics of SNOMED CT make it unique compared to other ontologies. For example, a relationship is represented as an OAV (Object Attribute Value) triplet, in which all three elements are concepts. The availability of relationship groups, which declare associations between sets of OAV triplets, is another specific characteristic of SNOMED CT which is discussed in detail in Schulz et al. (2006) and Cornet & Schulz (2009).

While there have been several frameworks of ontology evolution supporting change propagation mechanism, such as CHAO (Noy et al. 2006), KAON (Gabel et al. 2004), and Evolva (Zablith 2009), the papers which specifically discuss the propagation approaches are not excessively available. One of them is the work by Stojanovic et al. (2002b). The work proposes a method for change propagation, focusing on data on the web. It contains 3 steps: metadata capturing, metadata analysis, and generation of a proposal for modifications. In (Maedche et al. 2003), change propagation is performed among multiple distributed ontologies which are dependent to each other. The phase is divided into 3 components: Ontology Propagation Order, Change Filtering, and Change Ordering. Our work differs from the works in (Stojanovic et al. 2002b) and (Maedche et al. 2003), as well as from CHAO, KAON, and Evolva, in at least 3 things. Firstly, we work with health ontology, more specifically SNOMED CT, which has some specific characteristics in the change process. For instance, while in other ontologies, the deletion of a concept always causes the deletion of all its sub concepts (Maedche 2002), it is not the case with SNOMED CT. Secondly, in other approaches, the propagation directly follows the previous phases of ontology evolution, and hence, the knowledge on the semantic of the changes is known. Our change propagation is based on the available logs which only list the changes in terms of basic operations, not the semantic meaning of the operations. We do not have knowledge on the evolution process occurred before the change propagation phase. Thirdly, our approach is focusing on the change propagation to sub-ontologies derived from the evolved ontology.

In terms of the use of change log, our approach has similarity with the work of Plessers et al. (2007). The work has proposed the use of version log for the purpose of change detection phase in ontology evolution. The term ‘change detection’ is used to represent the phase in which it is checked whether other (composite) changes (besides the ones specified in the change request) or meta-changes occur as a consequence of the ontology modification. While the version log in (Plessers et al. 2007) uses CDL (Change Detection Language), which is based on temporal logic, to represent the changes kept in the version log, our version log contains only the delta of changes consisting of the list of the basic operations which have been executed form the evolved ontology.

3 SUB-ONTOLOGY AND THE EVOLUTION OF SNOMED CT

In this section, we discuss the definition of sub-ontology and the history mechanism in health ontology. The definition is based on SNOMED CT, which is the health ontology used in many health information systems.

3.1 Sub-ontology of SNOMED CT

SNOMED CT has provided a *Subset* mechanism which enables an application to use only a little part of the ontology. A Subset refers to a set of Concepts, Descriptions, or Relationships that are appropriate to a particular language, dialect, country, specialty, organization, user or context (IHTSDO 2010). However, the Subset cannot be used independently, as it only contains the IDs of the components participating in the Subset. To see the semantic information of the components, we have to refer back to the base ontology. Instead of using subset, in this paper we use *sub-ontology* to contain a specific part of ontology relevant to an application. A sub-ontology of a base ontology is a certain sub-set of that base ontology that is a valid ontology in its own right (Wouters 2005). The extraction process of the sub-ontology is not discussed here. The sub-ontology can be used independently, such that all semantic information of its components is provided in the sub-ontology itself. Thus, the look-back process to the base ontology is not needed.

Since our sub-ontology is extracted from SNOMED CT, it is important to know the characteristic of SNOMED CT. Below is the summary of the important characteristics of SNOMED CT.

- The three core components of SNOMED CT are concepts, descriptions, and relationships. A *concept* is a clinical meaning identified by a unique code in the form of string of digits which is called concept ID. A *description* is a term attached to a concept which explains the meaning of the concept in natural language. Each concept can have one or more descriptions. A description is represented as a unique description ID together with the concept ID it refers to. Concepts are connected to each other using relationship mechanism. The *relationship* mechanism in SNOMED CT is in the form of an OAV triplet (IHTSDO 2010). Object and Value are the two concepts to be connected, while Attribute is the relationship type connecting them. Attribute has to be one of the descendant concepts of the top level concept *Linkage Concept*. The main hierarchy of SNOMED CT is formed by the IS-A (sub-type) relationships.
- There are 18 top level concepts (the concepts that are directly connected to the root concept, i.e. concept *SNOMED CT Concept*) in SNOMED CT. Each concept other than the top level and root concepts is the descendant of at least one of the 18 top level concepts in the IS-A hierarchy.
- If SNOMED CT is represented as a graph, in which nodes represent concepts and edges represent relationships, the graph will be connected. It means that no island is permitted in the ontology graph. The connection between the 18 branches is established via the root concept.

Based on those characteristics, we define a valid sub-ontology of SNOMED CT, which is depicted by Figure 2, by presenting its features as follow:

- A sub-ontology of SNOMED CT consists of a non-empty set of concepts, a set of relationships, and a set of descriptions. The set of concepts is a subset of the set of SNOMED CT concepts, the set of relationships is a subset of the set of SNOMED CT relationships, and the set of descriptions is a subset of the set of SNOMED CT descriptions.
- The ontology graph of the sub-ontology must be connected.
- The sub-ontology must be independent, which means that all information about the concepts, the relationships, and the descriptions is provided in the sub-ontology.

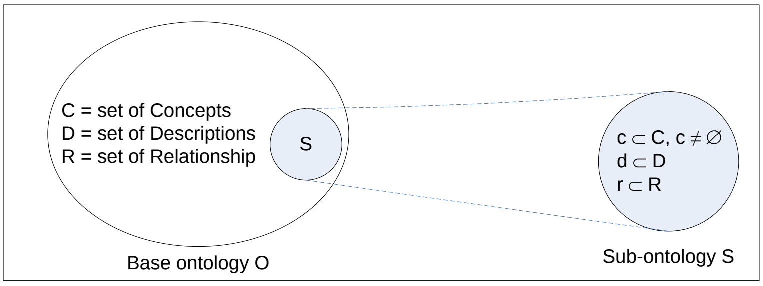

Figure 2. Sub-ontology is a subset of a base ontology which is a valid ontology by itself .

Some consequences arise from the restriction that the sub-ontology must be independent. Firstly, since each concept must be a descendant of one of the top level concepts, it must have at least one subtype hierarchy to the root concept. This is not the case with the concepts which do not have any IS-A relationship in the sub-ontology, because they are included in the sub-ontology only for completing the definition of other concepts. Secondly, the information about the concepts acting as Attribute in the OAV triplet of the relationships can only be obtained if the path of the Attribute concept to the concept linkage concept is included in the sub-ontology. Hence, the sub-ontology will also contain the concept linkage concept and its descendants which are used as the Attributes in the relationship triplets. Thirdly, as the effect of the first and second consequences, the root concept (SNOMED CT Concept) will be automatically included in the sub-ontology because the top level concepts are connected to each other only via the root.

3.2 SNOMED CT Evolution Mechanism

According to SNOMED CT Technical Reference Guide, there are 3 types of change in SNOMED CT. The first one is *addition*, in which a new component (concept, description, or subset) is added to the ontology. The second type is *status change*, in which the status of a component is changed, usually from *active* to *inactive*. The change of a component status from *active* to *inactive* semantically means that the component is deleted from the ontology. However, since in SNOMED CT a concept will never be deleted from the ontology, the term 'deletion' refers to the movement of the concept to a special branch of SNOMED CT which has *special concept* as the root concept. The last type of change is the *minor change* which accommodates a minor change in the presentation such as capitalization, punctuation, etc. This type of change does not alter the meaning of the component. This classification of changes is based on the changes in the presentation of SNOMED CT tables. In section 4.1, we present our proposed classification of health ontology change based on the semantic meaning of the changes.

The evolution of SNOMED CT is managed using the history mechanism which involves 2 tables: Component History Table and References Table. Component History Table is the main table of the SNOMED CT history mechanism and is the one employed in this work. It identifies each change in the status of a component and the release version in which the change was made (IHTSDO 2010). All changes applied to SNOMED CT components since the 19940131 release are archived in this table. The component included in the Component History Table can be a concept, description, or subset. One of the fields contained in the table is *ChangeType*, which can be *Addition* (1), *Status change* (2), or *Minor change* (3). Although relationship can also be changed, the history mechanism does not include a specific table to contain the change of relationships. The information about relationships is available in the Relationship Table. However, there is no status field in the Relationships Table because components other than Concepts and Descriptions are only distributed when they are "current" (IHTSDO 2010). Thus, the only way to know the changes of relationships is by comparing the different releases of Relationship Table.

4 HEALTH ONTOLOGY EVOLUTION

This section discusses the evolution of health ontology. We classify the changes in health ontology in Section 4.1. Based on the basic operations deduced from the classification, in Section 4.2 we propose a method of change propagation to make the sub-ontologies consistent with the evolved ontology.

4.1 Classification based on the Semantic of Changes

To develop the appropriate method for change propagation, it is important to understand the semantic of changes in health ontology. We do the semantic classification of changes based on the type of components (concept, relationship, and description). Minor changes such as change in the punctuation and spelling correction are not examined because it does not influence the hierarchy of the ontology. This classification is mainly based to the changes of SNOMED CT, especially on that the ontology consists of concepts, descriptions, and relationships and that the main hierarchy of the ontology is formed by the IS-A (sub-type) relationships.

4.1.1 Concept Changes

We divide the changes into 3: *addition*, *deletion*, and *movement*. Since the hierarchy of the ontology is formed by the IS-A relationships, the changes discussed in this section is concerned with the IS-A relationships. Figure 3 depicts the 3 operations. In the figure, a concept is represented as a circle, a line represents a relationship, and dashed line is used to represent the changes.

There are 2 types of addition: (i) addition of a leaf concept and (ii) insertion of a new concept to an existing relationship. In Figure 3, addition of a new leaf concept can be seen by the addition of concept X to the hierarchy. It must be followed by the addition of the new description(s) for the concept as well as the new relationship (XE) which connects it to other concept (E). Insertion of a new concept is shown by the insertion of concept Y to the hierarchy. This operation also involves the deletion of the existing relationship (relationship NB) to which the new concept is inserted and the addition of new relationships (YB and NY) for the new concept.

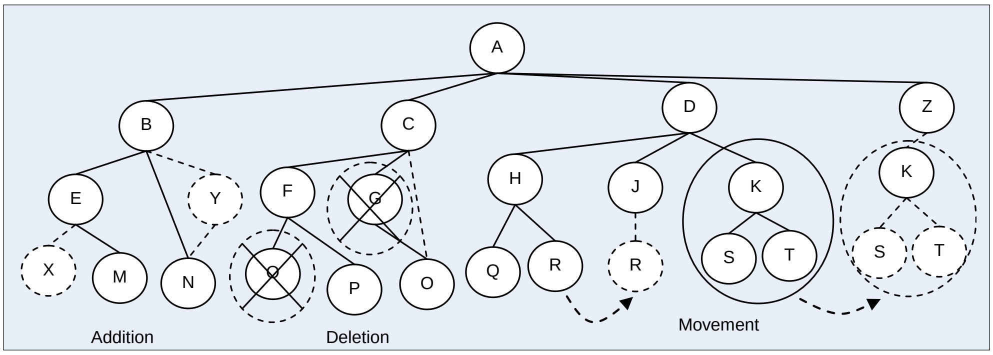

Figure 3. Examples of addition, deletion and movement of concepts in health ontology.

In Figure 3, two types of deletion are presented. The first one is the deletion of a leaf concept, in which concept O is deleted from the ontology. It involves the deletion of relationship (OF) that connects the concept with its parent and other concepts, the deletion of description(s) for the concept, and the deletion of concept itself. The second type of deletion is the deletion of a concept with child. In the example, concept G is deleted. This operation involves the deletion of the former relationships (OG and GC), the deletion of the description and concept G itself, and the addition of a direct relationship (OC) between the concept child and the former parent of the deleted concept.

We classify the movement operation into 2 types. The first one is the movement of a single concept, depicted in Figure 3 by the movement of concept R to another branch. This operation involves the deletion of the relationship (RH) between the concept and its former parent (and child if it has a child), and the addition of the new relationship (RJ) between the concept and the new parent (and child, if the parent has a child). The second type of movement is the movement of a fragment, i.e. a group of connected concepts, usually in a form of a parent concept and all its successors. In Figure 3, concept K, S and T with all relationships between them form a fragment. The fragment is moved to another branch, i.e. under concept Z. In this operation, the relationship (KD) between the fragment root and its parent is deleted. Then, a new relationship (KZ) which connects the fragment root with the new parent concept is added.

4.1.2 Relationship and Description Changes

Changes to relationship can be caused by two reasons: (i) changes of concepts (addition, deletion, movement), and (ii) changes of relationship itself. There are two types of relationship change: addition and deletion. For the deletion of a IS-A relationship, it must also be assured that deletion does not cause a concept to be separated from the rest of the ontology.

Similar to relationship, changes to description are mostly caused by the addition and deletion of concepts. However, it is possible that a description is added or deleted without the influence of a concept addition/deletion. The addition or deletion of a description does not influence the hierarchy of ontology. In case of deletion, it must be assured that a concept must be referred by at least one description. Thus, deletion of a description which happens to be the only description of a concept is not permitted.

4.2 Propagation Mechanism

The changes to ontology can have impacts on sub-ontologies. A mechanism of change propagation from the base ontology to the sub-ontologies is needed. This mechanism will keep the sub-ontology to be up to date with regard to the base ontology. In this section, we discuss the propagation mechanism of the evolution of health ontology. The rules in this method are based on the concept of OAV triplet relationship definition and the IS-A ontology hierarchy, which are commonly used in health ontologies such as SNOMED CT.

Figure 4 depicts the proposed propagation mechanism of the ontology changes. Each of the original sub-ontologies was directly extracted from the base ontology. During ontology evolution, which changes the base ontology to the evolved base ontology, delta of changes is derived. The delta of changes is the guidance in evolving each of the sub-ontology into the evolved sub-ontology. The delta of changes is used in this propagation mechanism because the history of changes is usually recorded as table logs. For example, in SNOMED CT, the table logs used to record the changes are the Component History Table and the References Table. From those tables, we can deduce the delta of changes. Although not all sub-ontologies are affected by the ontology evolution, to see whether they are affected or not, each sub-ontology must be checked using the guidance. Hence, after the checking process we call each sub-ontology as the evolved sub-ontology, whether it is affected by the ontology evolution or not. The evolved sub-ontologies are maintained to be consistent with the evolved base ontology.

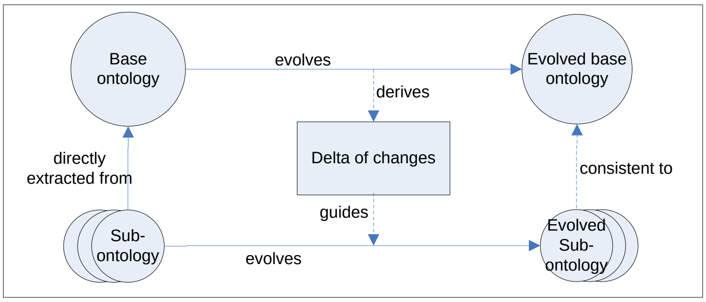

Figure 4. The changes of the base ontology are propagated to the sub-ontology using the delta of changes derived from the evolution.

The propagation mechanism, which utilizes the delta of changes, is depicted in Figure 5. The delta of changes consists of some lists of basic operations. Based on Section 4.1, it can be inferred that each change comprise one or more of some basic operations. For instance, addition of a leaf concept consists of an addition of concept, an addition of description(s), and an addition of relationship(s). The basic operations used to perform the changes are addition of concept, addition of description, addition of relationship, deletion of concept, deletion of description, and deletion of relationship. Each list in the delta of changes contains all operations of the corresponding type which are performed to evolve the ontology. Hence, there will be 6 lists: list of additions of concepts, list of deletions of concepts, list of additions of relationships, list of deletions of relationships, list of additions of descriptions, and list of deletions of descriptions.

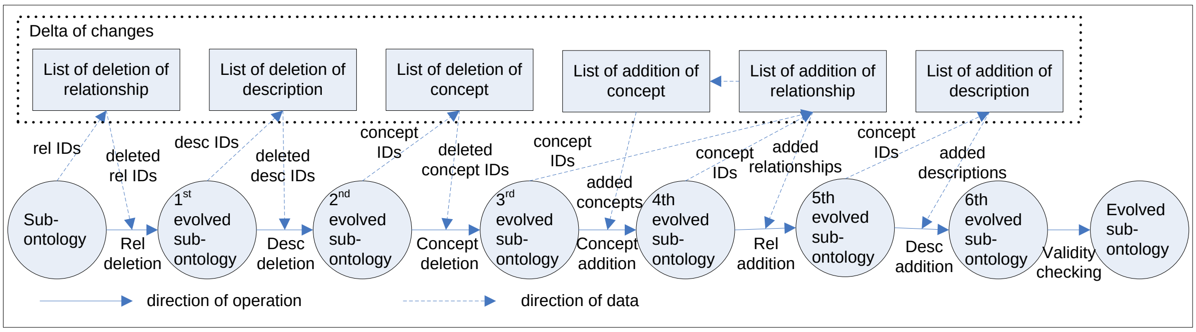

Figure 5. The propagation mechanism of ontology changes by utilizing the delta of changes.

The delta of changes becomes the guidance in evolving the sub-ontologies in order that the sub-ontologies consistent with the evolved ontology. The ontology changes can influence the size as well as the semantic content of the affected sub-ontologies. To prevent the growing of sub-ontology and to maintain the semantic content of sub-ontology, we use the following assumptions:

- The deep of the IS-A hierarchy of the sub-ontology can only be added by 1 level. It means that there can only be maximum 1 level addition of leaf concept.
- Each concept in the sub-ontology can only have maximum 1 level deep of relationship other than IS-A. This means that a relationship other than IS-A can be added to the sub-ontology only if the concept acting as Object (O) in the OAV triplet is included in the sub-ontology as well.
- A fragment cannot be inserted to the sub-ontology. This is to maintain the level of detail intended for the sub-ontology, i.e. preventing the sub-ontology from becoming too detail.

Since the deletion operations will reduce the number of concepts, relationships, and descriptions in the sub-ontology, which will subsequently reduce the number of IDs to be checked for the evolution, the propagation of changes for each sub-ontology starts with the deletion operations. Below is the sequence of operations, which consists of 7 operations, performed to obtain the evolved sub-ontology.

- **Relationship deletion**
  Since a relationship contains the OAV triplet and each component in the triplet refers to a concept, the operation to delete relationship should be executed before the execution of the operation to delete concept. The key of this operation is the relationship IDs. When a relationship ID contained in the sub-ontology also exists in the list of deletions of relationships, the relationship is deleted from the sub-ontology. The final result of this operation is the 1st evolved sub-ontology.
- **Description deletion**
  A description also refers to a concept, thus, the operation to delete descriptions must be performed before the execution of the operation to delete concepts. The key of this operation is the description IDs. Similar to that in relationship deletion, when a description ID contained in the sub-ontology also exists in the list of deletions of descriptions, the description is deleted from the sub-ontology. This operation produces the 2nd evolved sub-ontology.
- **Concept deletion**
  The key of this operation is the concept IDs. Every time it is found that a concept ID included in the sub-ontology exists in the list of deletions of concepts, the concept is deleted from the sub-ontology. The final result of this operation is the 3rd evolved sub-ontology.
- **Concept addition**
  This type of operation must be executed before the execution of the additions of relationships as well as the additions of descriptions because the definition of a relationship and a description contains concept(s). To check whether a concept added to the base ontology must be included in the sub-ontology, we use the concept IDs as the key and the list of additions of relationships as the reference. In this case, there are 2 scenarios in which the concept should be included in the sub-ontology. Firstly, when there is a new IS-A relationship in which the Value of the triplet is a concept included in the sub-ontology, and the concept acting as Object in the triplet is a new concept, then the new concept is included in the sub-ontology. Secondly, when there is a new non IS-A relationship in which the Object of the triplet is a concept which has an IS-A relationship in the sub-ontology, and the concept acting as Value in the triplet is a new concept, then the new concept is included in the sub-ontology. From this operation, the 4th evolved sub-ontology is produced.
- **Relationship addition**
  The key of this operation is the relationship IDs, and the reference to base the decision of the inclusion of a relationship is the list of additions of relationships. There are also 2 scenarios when a new relationship should be added to the sub-ontology. The first one is when the new relationship connects two concepts included in the sub-ontology, and the second one is when the new relationship connects a new concept acting as Object in the sub-ontology with a concept not in the sub-ontology. The final result of this step is the 5th evolved sub-ontology.
- **Description addition**
  The key of this operation is the ID of the concepts included in the sub-ontology. Whenever a description in the list of additions of descriptions refers to a concept in the sub-ontology, the description must be included in the sub-ontology. This operation produces the 6th evolved sub-ontology.

- **Validity checking**
  After the application of those changes, the produced sub-ontology might be not valid anymore. This step is aimed at ensuring the validity of the evolved sub-ontology. The following rules are used to maintain the validity of the sub-ontology.
  - If there is *island* in the ontology graph of the sub-ontology, the island, which contains concepts and relationships, must be deleted from the sub-ontology. The descriptions of the concepts deleted in this rule must also be deleted from the sub-ontology.
  - All concepts acting as Attribute in the added relationships must be included in the sub-ontology together with its relationship path to the concept *linkage concept*.
  - All concepts not included in the sub-ontology which acts as Value in any relationship in the sub-ontology must be included in the sub-ontology together with its description(s).

  The final result of this step is the evolved sub-ontology.

5 CASE STUDY AND DISCUSSION

In this section, we present a case study of the application of change propagation to a sub-ontology. It is followed by the discussion on the result.

5.1 Case Study

Since the main goal of the change propagation mechanism is to maintain the consistency of the dependent artefacts, in this case the sub-ontologies, with the evolved base ontology, we consider that a case study is appropriate to evaluate our proposed method in terms of its feasibility and the consistency of the produced evolved sub-ontology with the evolved base ontology. For the purpose, we first created a sub-ontology extracted from the version 20100131 of SNOMED CT. The list of the concepts used in the sub-ontology is shown by Table 2. Note that even though the sub-ontology does not represent a specific domain of interest in healthcare, it is formed such that it covers most of the change operations discussed in Section 4.1. The sub-ontology will be changed according to the changes happen in the base ontology based on the delta of changes. The delta of changes contains the list of additions of concepts, the list of deletions of concepts, the list of additions of descriptions, and the list of deletions of descriptions which are derived from the version 20100731 of Component History Table, and the list of additions of relationships and the list of deletions of relationships derived from the version 20100131 and 20100731 of SNOMED CT Relationship Tables. We use a graph representation for the sub-ontology in order that the propagation process can be easily understood. In the graph, a circle represents a concept in the sub-ontology. A relationship is shown by the edge connecting two vertices. To simplify the representation, we do not include some parts of the sub-ontology such as the concept *linkage concept* and all its descendants which are used for relationship type (Attribute) of the relationships in the sub-ontology.

Figure 6 shows the initial state of the sub-ontology. In the figure, an IS-A relationship is represented as a solid line, while a non-IS-A relationship is represented as a dashed line. The dots connecting concept L and M mean that M is not the direct descendant of L. To this sub-ontology, we apply the propagation mechanism discussed in Section 5, based on the operations in the delta of changes. Five operations are carried out, i.e. the relationship deletion, the concept deletion, the concept addition, the relationship addition and the validity checking. The description deletion and the description addition are not discussed in this case study because they do not influence the structure of the sub-ontology.

The first operation in the propagation mechanism is the relationship deletion, in which relationships ON, PO, QO, RP, YX, and ZX are deleted from the base ontology, and thus, they should be deleted from the sub-ontology as well. Then, the process concept deletion is performed which results in the deletion of concept Y. In the execution of concept addition, it is found out that the new concepts J and T should be included in the sub-ontology. The state of the sub-ontology after the execution of those operations, which in Section 5 is named the 4^{th} evolved sub-ontology, is shown by Figure 7a.

Letter representation	Concept ID	Concept Name (Fully Specified Name)
A	138875005	SNOMED CT Concept
B	404684003	clinical finding
C	64572001	disease
D	414029004	disorder of immune function
E	234646005	graft versus host disease
F	402355000	acute graft-versus-host disease
G	402360001	materno-fetal graft-versus-host disease
H	47650006	graft versus host reaction
I	119561005	grafting procedure
J	444547006	cutaneous graft-versus-host disease
K	39937001	skin structure
L	128139000	inflammatory disorder
M	64226004	colitis
N	54597004	chronic colitis
O	64766004	ulcerative colitis
P	442159003	chronic ulcerative pancolitis
Q	13470001	chronic ulcerative ileocolitis
R	90734009	chronic
S	62814004	chronic ulcerative inflammation
T	444546002	chronic ulcerative colitis
U	71854001	colon structure
V	267038008	edema
W	424372002	edema of extremity
X	102558002	oedema of the upper extremity
Y	102559005	oedema of arm
Z	102562008	oedema of elbow

Table 2. List of concepts used in the example of sub-ontology.

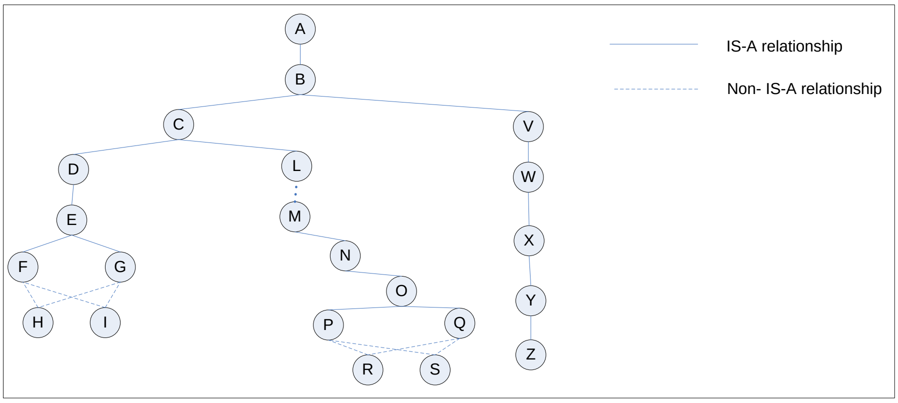

Figure 6. The sub-ontology which is extracted from the version 20100131 of SNOMED CT.

The relationship addition operation adds the relationships JE, HJ, IJ, OM, TO, PT, QT, QN, TU, and ZX to the sub-ontology. However, relationships JK and TU involve the concepts not included in the sub-ontology, i.e. concepts K and U. The inclusion of the relationships which involve concepts not in the sub-ontology will violate the validity of the sub-ontology. Thus, in the validity checking process, these two concepts are included in the sub-ontology so that the validity is maintained. Figure 7b shows the final result of the sub-ontology.

5.2 Discussion

We have compared the evolved sub-ontology produced using the propagation mechanism to the base ontology, i.e. the version 20100731 of SNOMED CT. Table 3 shows the screen shots of the subtype hierarchy of SNOMED CT captured by the CliniClue Xplore tool. Only the concepts included in the sub-ontology are presented here. Using the tool, it is not possible to view our sub-ontology as an ontology graph, and thus, we have to divide the sub-ontology into fragments. The subtype hierarchy is presented from the leaf concept to the root. Due to the space limitation, we only present the subtype hierarchy of three concepts: *cutaneous graft-versus-hot-disease* (in Table 2 represented as J), *chronic ulcerative ileocolitis* (Q), and *oedema of elbow* (Z).

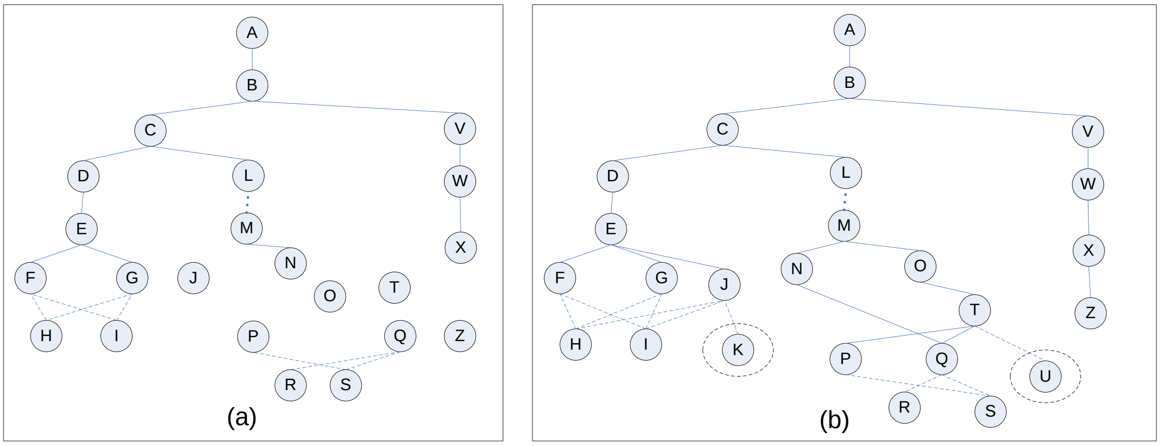

Figure 7. *(a). The 4th evolved sub-ontology after the execution of step 4, and (b) the evolved sub-ontology after the validity checking process.*

Subtype (IS-A) hierarchy	Non-IS-A relationship
Hierarchy: Subtype hierarchy clinical finding disease disorder of immune function graft versus host disease cutaneous graft-versus-host disease	finding site skin structure has definitional manifestation graft versus host reaction associated with grafting procedure
Hierarchy: Subtype hierarchy disease inflammatory disorder inflammation of specific body structures or t inflammation of specific body organs inflammation of large intestine colitis chronic colitis chronic ulcerative ileocolitis Hierarchy: Subtype hierarchy disease inflammatory disorder inflammation of specific body structures or tissu inflammation of specific body organs inflammation of large intestine colitis ulcerative colitis chronic ulcerative colitis chronic ulcerative ileocolitis	clinical course chronic Group associated morphology chronic ulcerative inflammation
Hierarchy: Subtype hierarchy clinical finding edema edema of extremity oedema of the upper extremity oedema of elbow	

Table 3. *Subtype hierarchy and the non-IS-A relationships of concept* cutaneous graft-versus-hot-disease, chronic ulcerative ileocolitis, *and* oedema of elbow.

From the first row of the table, it is shown that the subtype hierarchy of the leaf concept *cutaneous graft-versus-hot-disease* to the top level concept contains the concepts *cutaneous graft-versus-hot-disease* (J), *graft versus host disease* (E), *disorder of immune function* (D), *disease* (C), and *clinical finding* (B). There are 3 non-IS-A relationships which involve concept and connect it with concept *grafting procedure* (H), *graft versus host reaction* (I), and *skin structure* (K). This condition is the same as the evolved sub-ontology produced using our approach, shown by in the left fragment of the sub-ontology in Figure 7b. For the second concept, i.e. *chronic ulcerative ileocolitis*, there are two hierarchies available. The first one contains concept *chronic ulcerative ileocolitis* (Q), *chronic colitis* (N), *colitis* (M), ..., and *inflammatory disorder* (L), while the second one contains concept *chronic ulcerative ileocolitis* (Q), *chronic ulcerative colitis* (T), *ulcerative colitis* (O), *colitis* (M), ..., and *inflammatory disorder* (L). In both hierarchies, concept *clinical finding* is not shown. The non-IS-A relationships connect the concept with concept *chronic* (R), and *chronic ulcerative inflammation* (S), respectively. This condition is also the same as the evolved sub-ontology produced using our approach, shown by middle fragment of the sub-ontology in Figure 7b. For the last concept, i.e. *oedema of elbow,* only the IS-A hierarchy is shown in the table because our sub-ontology does not consider the non-IS-A relationships of the concept. The hierarchy contains concepts *oedema of elbow* (Z), *oedema of the upper extremity* (X), *edema of extremity* (W), *edema* (V), and *clinical finding* (B). This also conforms to our evolved sub-ontology, shown by the right fragment of the sub-ontology in Figure 7b. From the comparison, we can conclude that the concepts and relationships contained in our evolved sub-ontology are consistent with the corresponding concepts and relationships of SNOMED CT. Although we only work with a small part of SNOMED CT, it can be expected that our propagation mechanism is a promising method in term of consistency of the sub-ontology with the base ontology.

From the example of our sub-ontology, we can see the various complexities of the semantic of changes. Sometimes the change is simple, such as the addition of concept J and the deletion of concept Y. However, the semantic of the changes in the centre fragment (the predecessors of concept L) is complex. It is started from the deletion of concept N from the path, and followed by the insertion of concept O to the path and some additions of relationship. In such complex change, the order of changes is important to understand the semantic of changes and to find the final state of the components so that the sub-ontology will be consistent with the base ontology. If we know the order of the changes, we can propagate them easily to the sub-ontology. However, in case of SNOMED CT, we can only see the evolved ontology together with the list of additions and deletions of some of the components (concepts and descriptions). It is difficult to deduce the order of changes from the list. Using our approach, the problem of determining the order of changes is avoided because the delta of changes is represented as some lists of basic operations. The semantic of changes has been implied in the lists, and thus, we do not need to think about it. Furthermore, the validity of sub-ontology in our method is also assured by the validity checking process performed in the final step.

By using the propagation mechanism, sub-ontologies do not need to be re-extracted every time the ontology changes. The re-extraction is a high cost approach in dealing with ontology evolution since the number of sub-ontologies as well as the number of components in the sub-ontologies can be very large. Moreover, the high frequency of ontology changes will cause the high frequency of re-extraction, and this increase the resource needed for managing ontology changes. Our approach needs minimum resource. The execution of most of the steps has the complexity of *n*, in which *n* is the number of the processed component in the step which is included in the sub-ontology. The step which requires most resource is the validity checking, especially the checking process of the existence of any island.

6 CONCLUSION

In this paper, the evolution of SNOMED CT health ontology has been presented with the discussion of the history mechanism and the classification of changes. To propagate the changes to the sub-ontologies extracted from the ontology, a specific mechanism has been proposed which utilizes the delta of changes derived from the tables included in the history mechanism. A case study is used to evaluate the feasibility of the method. From the case study, it is shown that the evolved sub-ontology produced from the approach is consistent with the changes of the base ontology. Moreover, we believe that the use of the propagation mechanism will save resources in keeping the consistency of the sub-ontologies because the re-extraction of the sub-ontologies is not needed when the base ontology changes.

Our future work will focus on the generalization of the method such that it can be applied to different health ontologies such as LOINC and GALEN with some adjustment. For that purpose, we need to conduct a thorough evaluation to assess the application of the method to those health ontologies.

Acknowledgement

This work is partially supported by the Ministry of National Education of the Republic of Indonesia through the scholarship granted to the first author.

References

Cornet, R. and de Keizer, N. (2008). Forty years of SNOMED: a literature review. BMC Medical Informatics and Medical Decision Making, 8(Suppl 1):S2.

Cornet, R. and Schulz, S. (2009). Relationship Groups in SNOMED CT. Studies in Health Technology and Informatics, 150, 223-227.

Gabel, T., Sure, Y., and Voelker, J. (2004). KAON – Ontology management infrastructure. D3.1.1.a, SEKT Project: Semantically Enabled Knowledge Technologies.

IHTSDO. (2010). SNOMED Clinical Terms Technical Reference Guide. International Health Terminology Standards Development Organisation.

Leslie, H. and Heard, S. (2006). Archetypes 101. In Proceedings of the Health Information Conference (HIC) 2006 (Westbrook, J. and Callen, J. Eds.), Sydney.

Maedche, A., Motik, B., Stojanovic, L. Studer, R. and Volz, R. (2002). Managing multiple ontologies and ontology evolution in Ontologging. In Proceedings of the IFIP 17th World Computer Congress. Deventer.

Maedche, A., Motik, B., and Stojanovic, L. (2003). Managing multiple and distributed ontologies on the Semantic Web. VLDB Journal, 2003 (12), 286 – 302.

Noy, N.F., Chugh, A., Liu, W., Musen, M.A. (2006). A framework for ontology evolution in collaborative environments. In Proceedings of the International Semantic Web Conference 2006, p. 544 – 558.

Plassers, P., and Troyer, O.D. (2005). Ontology Change Detection Using a Version Log. In Proceedings of the 4th International Semantic Web Conference, p. 578 – 592. LNCS 3729, Springer-Verlag, Berlin.

Qamar, R. and Rector, A. (2007). Semantic mapping of clinical model data to biomedical terminologies to facilitate data interoperability. In Proceedings of the Healthcare Computing 2007 Conference, Brisbane.

Sari, A.K., Rahayu, W., and Wollersheim, D. (2010). Utilization of ontology in health for archetypes constraint enforcement. Lecture Notes in Computer Science (Taniar, D., Gervasi, O, Murgante, B., Pardede, E. and Apduhan, B. O.Eds.), 6018: 380-394, Springer-Verlag, Berlin.

Schulz, S., Hanser, S., Hahn, U. And Rogers, J. (2006). The Semantics of Procedures and Diseases in SNOMED CT. Methods of Information in Medicine, 45(4): :354-358.

Stojanovic, L., Madche, A., Motik, B. and Stojanovic, N. (2002a). User-driven ontology evolution management. In Proceedings of the European Conference Knowledge Engineering and Management, p. 285–300. EKAW 2002, Springer-Verlag, Berlin.

Stojanovic, L., Stojanovic, N., and Handschuh, S. (2002b). Evolution of the Metadata in the Ontology-based Knowledge Management Systems. In Proceedings of Experience Management. Berlin.

Wouters, C. (2005). A Formalization and Application of Ontology Extraction. PhD Thesis, Department of Computer Science and Computer Engineering, La Trobe University, Melbourne.

Yu, S., Berry, D. and Bisbalz, J. (2010). An Investigation of Semantic Links to Archetypes in an External Clinical Terminology through the Construction of Terminological “Shadows". In TeaPOT: People Oriented Technology Conference Papers, Dublin Institute of Technology.

Zablith, F. (2009). Ontology evolution: a practical approach. In Proceedings of Workshop on Matching and Meaning at Artificial Intelligence and Simulation of Behaviour. Edinburgh.