

An Advanced Clinical Ontology

Riichiro MIZOGUCHI¹, Hiroko KOU¹, Jun ZHOU¹, Kouji KOZAKI¹,
Ken IMAI², Kazuhiko OHE²

¹The Institute of Scientific and Industrial Research (ISIR), Osaka University, Japan; ²Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo, Japan

Abstract

This article discusses a fundamental issues of medical ontology based on ontological theory. We focus on "anatomical structure of organs" and "abnormal states in the human body". On the basis of the investigation, we distinguish organ-specific types from those independent of any organ to maximize the explicitness of ontology. The next feature of our ontology is to allow on-demand reorganization of is-a hierarchy of diseases instead of one fixed hierarchy to cope with various viewpoints which physician might have. We also take care of the notorious issue related to conflict of is-a and part-of relations.

Introduction

Ontology is one of the most promising techniques for enabling semantic interoperability of medical information among various data across domains/tasks. This is why there have been developed some ontologies such as SNOMED-CT¹, ICD-10², Galen³, etc. In this situation, there has been strong need of a sophisticated medical ontology in Japanese which is highly expected to be compatible with those existing ontologies. The authors believe that the ontology cannot be a simple translation of the existing ontologies because that would hide some possible concepts specific to Japanese clinical practice. We should first establish our own ontology to reflect Japanese clinical practice and then investigate alignment between the Japanese one and existing ones to make them interoperable with each other. Another reason for this policy is that those existing ontologies suffer from so-called "legacy problem", that is, some of them are incomplete in terms of ontological theories since they had started their project when ontological engineering was not matured enough. As a late comer, we aim at building a medical ontology which is ontologically sound.

In this background, the Japanese ministry of health, labour and welfare has launched a three-year project on Foundation of Database for Clinical Knowledge in 2008. The expecting deliverable is a clinical ontology composed of roughly 30,000 concepts or more covering a couple of thousands of diseases in typical clinical and anatomical domains. This paper is

an intermediate report on the ontology development conducted in the project and is structured as follows. The next section discusses the underlying policy in the ontology development. Human body structure with the focus on organs is discussed in Section 3. Diseases are discussed in Section 4. Section 5 presents related work to locate our project in the right context followed by concluding remarks.

Underlying policy

Our ontology is being developed having the following issues in our mind. These issues lead us to introduce several new theories and ideas as explained below.

a) Commonality vs. specificity: In order to make it more articulate, common characteristics and target-thing-specificities should be clearly captured and differentiated. We introduced *generic structural /disorder* components each of which represents common characteristics of structural and disorder components as much as possible.

b) *is-a vs. part-of* issue 1: For example, the two relations <disease of a pulmonary valve *is-a* disease of heart> and <pulmonary valve *part-of* heart> cause a problem, since both "disease of a pulmonary valve" and "disease of a heart" have a slot of site of the disease and the filler of the former must be a subclass of the latter from the theory of inheritance, in reality, however, the former must be a part of the latter. To solve this problem, on the basis of our latest theory of roles⁴, we introduced "*p*-" operator in our ontology building tool Hozo^{4,5} which automatically generates a generic concept representing all the parts of the thing the operator is attached.

c) *is-a vs. part-of* issue 2: The atrium is composed of left and right atriums. At the same time, however, both left and right atriums are subclasses of the atrium. Fortunately, the "*p*-" operator can solve this issue at the same time.

d) No single hierarchy of diseases does not work well for all the stake holders such as pathologists, clinicians and surgeons, etc. To cope with these various viewpoints, we introduced an innovative technique to realize **on-demand** reorganization of *is-a* hierarchy according to the specified viewpoints.

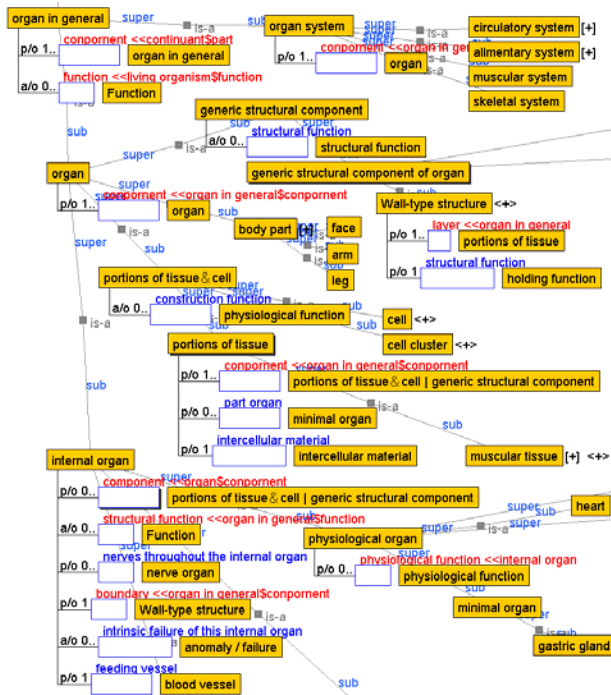


Fig. 1 Top-level structure of human body.

Human body structure

Upper-level types

Fig. 1 shows the top-level types of the structural part of human body. *Organ in general* consists of *organ* and *organ system*. *Organ* consists of *internal organ*, *body part*, *portion of tissue & cells* and *generic structure* component. *Internal organ* represents ordinary organs such as heart, *portion of tissue & cells* includes finer-grained organs such as *gastric gland* as well as tissue and cells, *body part* includes structural parts such as face, arms, legs, et al. The design rationale of the top-level structure is that to represent it in a compact recursive structure reflecting the essential properties of several important types of organs. In fact, the nested structure of *cells*, *tissues*, *minimal-organs*, *organs* and *organ system* are nicely represented in the recursive structure. The details of *generic structure* are explained below.

Generic structural components

Many of the organs which consist of several sub-organs share common structural components. For example, we can identify *hollow structure* component in a stomach and atrium. *Tubular structure* components are found in *blood vessel* and *esophagus*. Although the materials are different, *gastric wall* and *cardiac wall* share three-layer structure of tissue which we call *wall-type structure* component. Those common structural components enable us compact representation of those organs

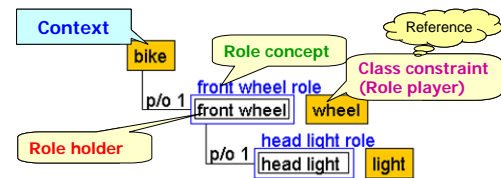


Fig. 2 Legend of concept and role definition in Hozo.

because common properties can be defined once at those components. Typical examples of the utility of such representation include that *tabular structure* has a potential malfunction of *arctation* caused by narrowed cross section area. When it occurs at *blood vessel*, it is called *angiostenosis* which would occur at *cerebral vasculature*, *coronary artery*, etc. as well. All share similar properties. *Arctation* occurs at esophagus and it blocks the flow of pieces of foods down to the stomach, which is analog to the *angiostenosis* case. All of them share widening operations as a possible treatment for its remedy, though how to implement it would be different from each other. Properties specific to each organ can be defined additionally or by specializing the properties inherited from the common structural component. For example, although vein and esophagus share the *tubular structure* component, vein has a valve of vein but esophagus does not. While both vein and esophagus are composed of three-layer wall, that of esophagus has two-layer *muscle fiber* structure to perform *peristaltic action*.

In order to represent such specificity, we introduce the concept of Roles supported by Hozo which is a tool for building ontology developed by us⁵. Fig. 2 shows the legend of type definition in Hozo as well as role definition in which “bike” is defined by specifying its part. “p/o” stands for “part-of” link. At the same time, a role named “front wheel role” is defined by referring to “wheel” defined elsewhere. In Hozo, an entity playing a role is called “role-holder”. In the case of Fig. 2, a wheel which is incorporated as a part of a bike and playing the role of “front wheel role” is thereby called a “front wheel”. Hozo, thus, realizes representation the mutual dependency between the whole and its parts⁴.

Advanced issue of is-a and part-of relation

Contrary to the superficially simple characteristics of *is-a* and *part-of* relations, there have been two serious issues to solve such as b) and c) shown in section 2. Fig. 3 shows the first difficulty b). Following the property inheritance, pulmonary valve must be a specialization of heart, but it is not. To represent the *is-a* relation between disease of heart and disease of pulmonary valve, we need to invent to inherit parts of heart rather than its subclass from heart.

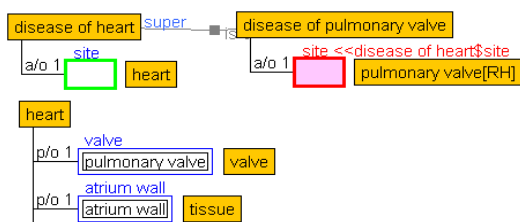


Fig. 3 A problematic situation related to the property inheritance from super to sub concepts.

The second difficulty seems to be more serious than b), since it is related to fundamental conceptualization of what is a whole of collectives. This difficulty has been discussed by Udo Hahn and his group⁶. In order to solve this difficulty, they introduced SEP-triple which consists of three concepts: the original entity called E-node as a whole together with two concepts derived from the original one: One called S-node and the other called P-node. S-node is a super class of both E-node and P-node. The key idea is the introduction of a generic concept representing all the parts of the original entity under consideration.

We first tackled the issue b) and came up with a new operator named “*p*–” operator explained above. The operator enables parts to be inherited by ordinary property inheritance mechanism. In the case of Fig. 3, for example, we write “*p-heart*” instead of “heart”, then the slot of its subclass inherits not subclass of “heart” but its parts. Although this method would suggest we need complicated hidden processes in Hozo, it is not the case. When *p-X* is used, Hozo automatically generates a generic concept representing all defined parts of X including all parts which have X as their ancestor. This is valid because each part *is-a* subclass of “X’s parts class” which coincides with *p-X*. According to mereology, the theory of parts, *p-X* includes itself which is not the very X as an entity but X as its part. This is why “*p*–” operator can solve the issue c).

Definition of disease

Disease as an abnormal state

It is apparent that capturing diseases is a tough problem from the beginning. In fact, there are many ways of categorization of diseases. Patients use common sense names of diseases. What doctors of primary care deal with and what the government deals with to calculate statistics of the cause of death are very different. In addition, pathologists, clinicians and surgeons see the same disease different points of view. Those differences result in multiple taxonomies of diseases. When ontology developers build an ontology, they tend to present one *is-a* hierarchy believing it is the essential structure of the world

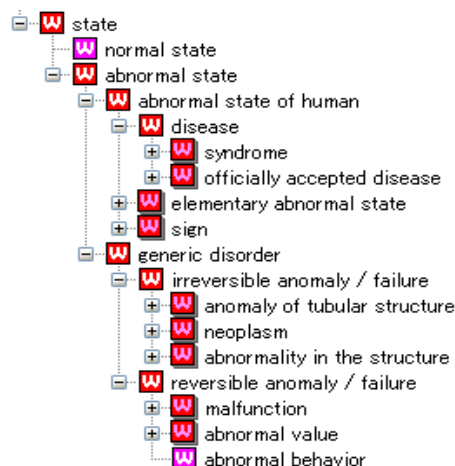


Fig. 4 Top-level categories of state.

under consideration. Although it is often true, it is not the case in medical ontology in which disease classification is essentially perspective-dependent as we see above. To cope with this well-known difficulty, we adopted the strategy as follows: (1) building the most fundamental *is-a* hierarchy of diseases based on “state” and (2) on-demand generation of *is-a* hierarchy according to the viewpoint specified. Ontologically, pathological state, disease, symptom, syndrome, disorder, dysfunction, failure, cause, etc. are kinds of disorder of human body and can be represented as “states”. On the basis of this fact, the top-level categories of disease are developed as shown in Fig. 4.

Upper level categories of disease quality

State, the top-level category, has *normal state* and *abnormal state*. The latter has two subclasses such as *abnormal state of human* and *generic disorder* which is the type corresponding to the *generic structure* component. *Officially accepted disease* is the central type of disease and is defined by referring to other states. *Elementary abnormal state* is disorder which cannot be disease by itself and is mainly used for characterizing *officially accepted disease*. The main component of disease is *pathological state* which is a role played by *abnormal state* in the context of diseases which clinical experts accept as disease. Basically, each state is defined in terms of <Entity, Attribute, Value>, EAV-triple. We have investigated the survey of the current clinical practice in Japan and found EAV-triple works quite successfully. We also analyzed quality descriptions in ICD-10 and Galen and found that most of them are covered by the ontology of quality and quantity defined in YATO⁷ and we can convert them into the form of EAV-triple.

Officially accepted disease

Fig. 5 shows the framework of *officially accepted disease*. It is defined by specifying typical *disorder*

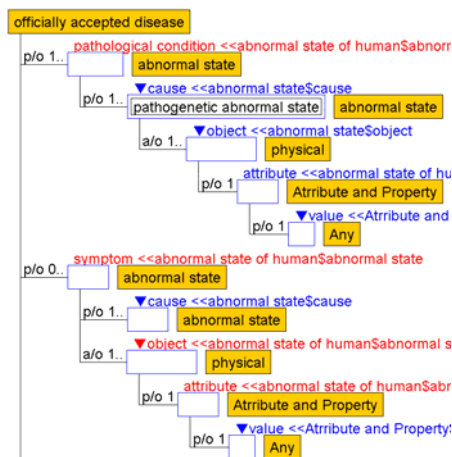


Fig. 5 The framework of *officially accepted disease*.

roles played by *abnormal state* which is based on EAV-triple. Depending on expert's decision, some abnormal states become *pathological state*, some become *symptom* derived by the *pathological state*. This is our way of defining diseases based on *states* with the help of role-defining function of Hozo. Exploiting the fundamental characteristics of states and the expressive power of role representation of Hozo, definition of disease can be easily adjusted to the current understanding of the disease under consideration, which makes the ontology both solid and flexible at the same time. In fact, the same abnormal state can be *pathological state* or *symptom* according to the context of disease of interest. Furthermore, the fact that boundaries between those states are intrinsically vague prevents us from defining them as established types which are hard to change. Roles which intrinsically change according to the context best fits to definition of those states.

On-demand reproduction of is-a hierarchy

As discussed above, on-demand function is critically important to make clinically acceptable for experts in many divisions in medicine. The state-based disease definition with Hozo tool allows us to achieve this demanding goal. Diseases are defined as subclasses of *officially accepted disease* shown in Fig. 5 and they have several slots with classes for constraining slot values. In Fig. 5, examples are *abnormal state* and *physical*. If users want to see *is-a* hierarchy of diseases in terms of the partonomy of organs, then they just specify *physical* which is where diseases locate. Of course, the partonomy is converted into *is-a* hierarchy by Hozo using "p-" operator technology. If they want to see the hierarchy in terms of *pathological state*, then they specify it. We have already built a

prototype system for this functionality and confirmed its feasibility. We have also manually mapped the ICD-10 concepts to ours in the prototype system of ontology navigation to enable users who want to browse our ontology from the ICD-10 contents. The demonstration of the functionality is available at the following URL: <http://www.ei.sanken.osaka-u.ac.jp/MedOnto/>.

Concluding remarks

We did preliminary comparison between our ontology with existing ontologies such as SNOMED-CT, FMA, CARO⁸, ICD-10 and GALEN and confirmed the ontological soundness of our ontology which is compliant with YATO which is comparable to BFO⁹ and DOLCE¹⁰. In addition to this, it has major advantages over them with respect to the following three perspectives: 1) explicit representation of commonality and specificity of structure and diseases, 2) resolution of the notorious problem of inter-dependence between *is-a* and *part-of* relations and 3) on-demand reorganization of *is-a* hierarchy of diseases. In the preliminary comparison, we investigated FMA in terms of the difficulty 2) and found that FMA tries to solve it by introducing a lot of redundant virtual classes and ends up with partial solution of the problem in the sense that it fails to solve the issue of c) in Section 2. We are currently in the phase of increasing diseases of several clinical divisions by tight collaboration with clinical doctors using a description support system we developed to help them input data.

Acknowledgement

This research is supported by the Ministry of Health, Labour and Welfare, Japan.

References

1. SNMED-CT, http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html
2. ICD10, <http://www.who.int/classifications/icd/en/>
3. OpenGALEN, <http://www.opengalen.org/>
4. Mizoguchi R, et.al., A Model of Roles within an Ontology Development Tool: Hozo, *J of Applied Ontology*, 2007:2:159-179
5. HOZO, <http://www.hozo.jp/>
6. Hahn U, et al. Turning Lead into Gold? *EKAW2002, LNCS*; vol. 2473:pp182-196.2002
7. Mizoguchi R, Yet Another Top-level Ontology: YATO, *Proc. of InterOntology09*: 91-101
8. CARO, http://www.bioontology.org/wiki/index.php/CARO:Main_Page
9. BFO, <http://www.ifomis.org/bfo>
10. DOLCE, <http://www.loa-cnr.it/DOLCE.html>