DEVELOPMENT OF FUNDAMENTAL TECHNOLOGIES FOR BETTER UNDERSTANDING OF CLINICAL MEDICAL ONTOLOGIES

Hiroko Kou, Mamoru Ohta, Jun Zhou, Kouji Kozaki, Riichiro Mizoguchi
The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
{Kou, ohta, shuu, kozaki, miz}@ei.sanken.osaka-u.ac.jp

Takeshi Imai, Kazuhiko Ohe
Department of Medical Informatics, Graduate School of Medicine, Tokyo University, 73-1m Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan
{ken, kohe}@hcc.h.u-tokyo.ac.jp

Keywords: Medical ontology, dynamic generation of is-a hierarchy, natural language explanation

Abstract: We have been building a Japanese medical ontology that provides an intelligent infrastructure for systematization and synthetic understanding of medical knowledge on a large scale. The objectives of our research include building a medical ontology and developing application systems based on it. We identified a few common fundamental technologies for understanding the medical ontology and implemented them. The main features of these technologies are summarized as the following two functions: dynamically generating *is-a* hierarchy according to the user's interest, and providing natural language explanations. We built a prototype medical information service system using these fundamental technologies. We conducted an informal evaluation in a workshop and received favorable comments from medical experts.

1 INTRODUCTION

A lot of medical data have been computerized to improve the quality of medical services. However, most of them are stored in different formats depending on their domains and database management systems. In addition, the government has set IT-based structural reform of the healthcare system as the top IT strategic focus to develop more advanced medical information systems. Here we pay special attention to ontology studies, because a medical ontology would be a core technology for various applications, such as electronic medical records, and diagnostic support systems. Medical ontology research is a key to the successful development of various knowledge processing applications that are interoperable with one another. Although some medical ontologies and standard vocabularies such as MeSH (MeSH), ICD-10 (ICD-10), SNOMED-CT (SNOMED-CT), Galen (GALEN), and FMA (FMA, Rosse, C. 2003) have been developed, most of them are based on legacy system terminologies, and some have quite a few ontological problems. For example, Stefan Schulz et al. (Stefan, S. 2007) point out some ontological problems in SNOMED-CT, such as confusion between the *subclass-of* (*is-a*) relation and *instance-of* relation, multiple inheritance without considering the inheritance of an attribute, and others. In addition, these ontologies are not suitable for Japanese medical practice due to many culture-specific differences between Japan and Western countries.

We have developed a medical ontology suitable for Japanese medical practice (Mizoguchi, R. 2009). This ontology is to be translated into English after enough data is gathered, and will be mapped with other standard ontologies for interoperability. This mapping will evaluate the standard terminologies in line with fundamental ontology engineering and uncover culture-specific differences between Japan and Western countries. In addition to building a medical ontology, the objectives of our research include constructing and developing application systems based on it. The underlying philosophy is to "Learn from the ontology" before using it for building another application because ontology itself is a rich source of knowledge about the domain. We identified a few common fundamental technologies for understanding the medical ontology, such as a method for navigation for ontology content exploration and explanation generation, and we implemented them. Also, we built a prototype medical information service system using these fundamental technologies.

This paper is organized into four sections, including this Introduction. In the next section, Section 2, we summarize our major principles for building our medical ontology. In Section 3, we discuss fundamental technologies for understanding clinical medical ontologies and introduce the main functional modules for dynamically generating *is-a* hierarchies. Finally, we present concluding remarks and discuss future work.

2 DEVELOPMENT OF MEDICAL ONTOLOGY

2.1 Outline of our Medical Ontology

Our major principles for building a medical ontology are that we clearly distinguish between context-dependent concepts from context-independent concepts, and that we describe all types of diseases in a common framework. In the framework, diseases are described as combination of abnormal states, and all types of abnormal states of the human body described in the same framework, which is represented in terms of a triple <object, attribute, attribute's value> (Figure 1(a)). For example, "a high level of HbA1c in the blood" is described as <blood, HbA1c, high>. Also, we explicated the specification of causal chains, which are expressed clearly by sets of causal relationships between abnormal states and their causes (Mizoguchi, R. 2009). This ontology is described using Hozo which can describe roles correctly based on the role theory and can export ontology in OWL (Kozaki, K. 2002). For example, diabetes is described as combinations of abnormal states, such as diabetical hyperglycemia, dry mouth, excessive urination, and diabetic glycosuria (Figure 1(b)). These abnormal states play roles such as symptoms and main pathology of the disease. A role is defined as a dependent entity played by another entity in a context. An entity playing a role in a specific context is called a role holder. By a class constraint, we mean a constraint on the class to which an instance playing the role belongs. (Mizoguchi, R. 2007)

To evaluate whether it is possible to describe most clinical observations in clinical practice by our conceptual description framework using abnormal states, we verified a number of physical findings. We used 3465 data of physical findings (which is a standard vocabulary used when symptoms and opinions were described in a care card) that have been opened to the public by The Medical Information System Development Center in Japan (MEDIS-DC). We confirmed that our conceptual description framework is suitable for most clinical observations. Similarly, we verified the adequacy of the conceptual description frame of diseases by describing 50 disease data that were extracted to cover each area from 12,000 disease data of ICD-10 that WHO has made public. After validating the description framework, we created a template for describing knowledge about each disease and developed dedicated input software to record data, and data on 6051 diseases are currently being collected from 12 hospital departments. Our medical ontology imported the collected data, and as the result it consists about 15,000 concepts and 60,000 slots at present. The person in charge of each medical area plans to continue the description and data collection in the future.

2.2 Comparison with Other Medical Ontologies

In SNOMED-CT, the disease is one of the 19 top-level categories. Therefore, there is no super-concept of the disease. SNOMED-CT's 19 top-level categories preserve the legacy of the former SNOMED axes, which do not easily agree with any

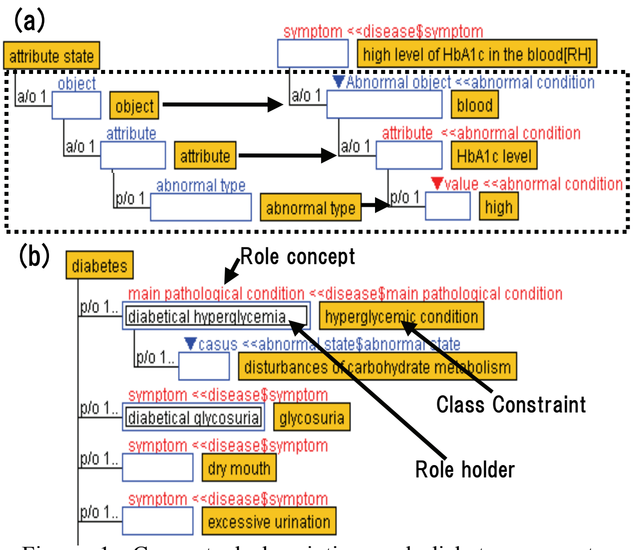

Figure 1: Conceptual description and diabetes concept description as example on ontology

formal upper level ontology. This is in fact acknowledged as one of SNOMED-CT's problems. (SNOMED-CT, Stefan, S. 2007, SNOMED Clinical Terms User Guide). On the other hand, we used top-level ontology of YAMATO (YAMATO), which is compatible with DOLOCE (DOLOCE), and can be viewed as a high-level ontology based on philosophical considerations.

In general, concepts are related to one or more other concepts. In fact, SNOMED-CT allows multiple "*is-a*" relationships for describing that complicate classification without providing any information regarding a viewpoint for systematizing hierarchical classification. We adopt the approach of accepting an essential hierarchical classification and switching between the hierarchical classifications. It is then necessary to develop fundamental technologies that gather related information from the ontology and dynamically generate an *is-a* hierarchy using the information. This function of dynamic generation of *is-a* hierarchy has been realized first in the world.

3 FUNDAMENTAL TECHNOLOG IES FOR UNDERSTANDING ME DICAL ONTOLOGIES

3.1 Applications based on Medical Ontologies

A medical ontology would be a core technology for various applications, such as electronic medical records, diagnostic support systems, medical electronic dictionary and electronic textbook, medical knowledge navigation, portal site of medical information, and so on. The underlying philosophy on application systems in our research is that "Learn from the ontology" before using it for building applications since ontology is a rich knowledge source about the domain. We identified a few common fundamental technologies for understanding the medical ontology, such as a navigation method for ontology content exploration and content management for explanation generation, and we implemented them. The main features of these technologies are summarized as the following two functions: dynamically generating an *is-a* hierarchy according to user's viewpoints, and providing natural language explanation. This paper concentrates on the dynamic *is-a* hierarchy generation function for navigation of ontology content.

3.2 Navigation of ontology content

By navigation of ontology content, we mean to guide users to explore ontology content as a rich source of knowledge about the domain. Navigation technology supports efficient access to concepts defined in medical ontologies. It is very important because locating targeted medical knowledge is otherwise difficult in the vast amount of medical information available. Navigation includes technologies such as providing indexes for access, linking to related information, and retrieval functions. We focus on indexing based on a medical ontology because it provides the basic infrastructure for navigation. For this purpose, ontology content exploration should gain much attention.

Ontology content is usually explored following subclass-of (*is-a*) hierarchy. However, we have obvious problems because many types of hierarchical classification could exist, and disease characterization is susceptible to various interpretations. A disease is interpreted from various viewpoints. Consider diabetes as an example. Clinical technologists may pay attention to the body part that has the abnormality and classify diabetes as an abnormal blood sugar level. On the other hand, a certain specialist may pay attention to the main condition and may classify diabetes as an abnormality in metabolism, and another specialist may classify diabetes as a lifestyle disease. Staff administering the medical care implicitly understands which *is-a* hierarchy should be used for disease interpretation in correlation with their respective interpretations. This suggests that one *is-a* hierarchy of diseases cannot cope with such a diversity of viewpoints, since a single-inheritance hierarchy necessarily represents one viewpoint.

Some researchers would say "this is why we use multiple-inheritance *is-a* hierarchy. Why don't you use it for disease organization?" The answer to such a question is as follows. In ontological theory, *is-a* hierarchy must represent essential property of things and hence it should be single-inheritance, since essential property of things cannot be multiple. Imagine objects, processes, attributes, all of them have their own unique and essential properties. The use of multiple-inheritance for organizing things necessarily blurs what is the essential property of things. This observation is strongly supported by the fact that both of the well-known upper ontologies: DOLCE and BFO use single-inheritance hierarchy. If we add a practical difficulty, we can indicate the instance management issue. Instances of a class must have their own appearance/disappearance policy according to their essential properties. Multiple-inheritance *is-a* hierarchy hides essential property of things and hence you cannot identify what policy to use for their appearance/disappearance.

So, the problem is how to reconcile the conflicting requirements of multiple-views and single-inheritance in a good ontology. Each of these *is-a* hierarchies is significant and we try to keep both requirements. Furthermore, in order to obtain a deep understanding of a disease, it is important to use more than one disease *is-a* hierarchy which represents the essential structure underlying the target world. In order to tackle this important issue, we adopt an approach of dynamically generating *is-a* hierarchies of diseases according to the viewpoint of users from an ontology using single-inheritance.

3.3 Dynamic Generation of *is-a* Hierarchies

To develop a function of dynamic generation of *is-a* hierarchy, we need to classify plausible viewpoints of the hierarchical classifications and relate these viewpoints to the conceptual structures. This function uses the aspect to determine *is-a* hierarchies. This aspect is used to traverse the ontology and collect related information. An *is-a* hierarchy is then generated using this information.

For instance, to generate an *is-a* hierarchy from the viewpoint of the pathological condition, the main pathological condition of metabolic disorders is an abnormal state consisting of a metabolic abnormality, and the subclassification of a metabolic disorder is generated using *is-a* hierarchy information about the metabolic abnormality, which is the disorder's main pathological condition. The subclassification of a metabolic disorder is similar to an *is-a* hierarchy of metabolic abnormalities.

The diabetic main pathological condition is carbohydrate metabolism abnormality, which is a particular type of metabolic abnormality. Therefore, the disease of carbohydrate metabolism is subordinate to metabolic disease, and diabetes is one of the diseases of carbohydrate metabolism.

Moreover, one might want to see an *is-a* hierarchy of diseases analogous to the location where their main pathological condition appears. In such a case, the part-of relationships of the human body converted into an *is-a* hierarchy and is used to generate such an *is-a* hierarchy. This *is-a* hierarchy is similar to the part–whole relationship of the human body. For instance, mitral valve disease and tricuspid valve disease are classified as subcategories of cardiac disease, because the tricuspid valve and the mitral valve have a part–

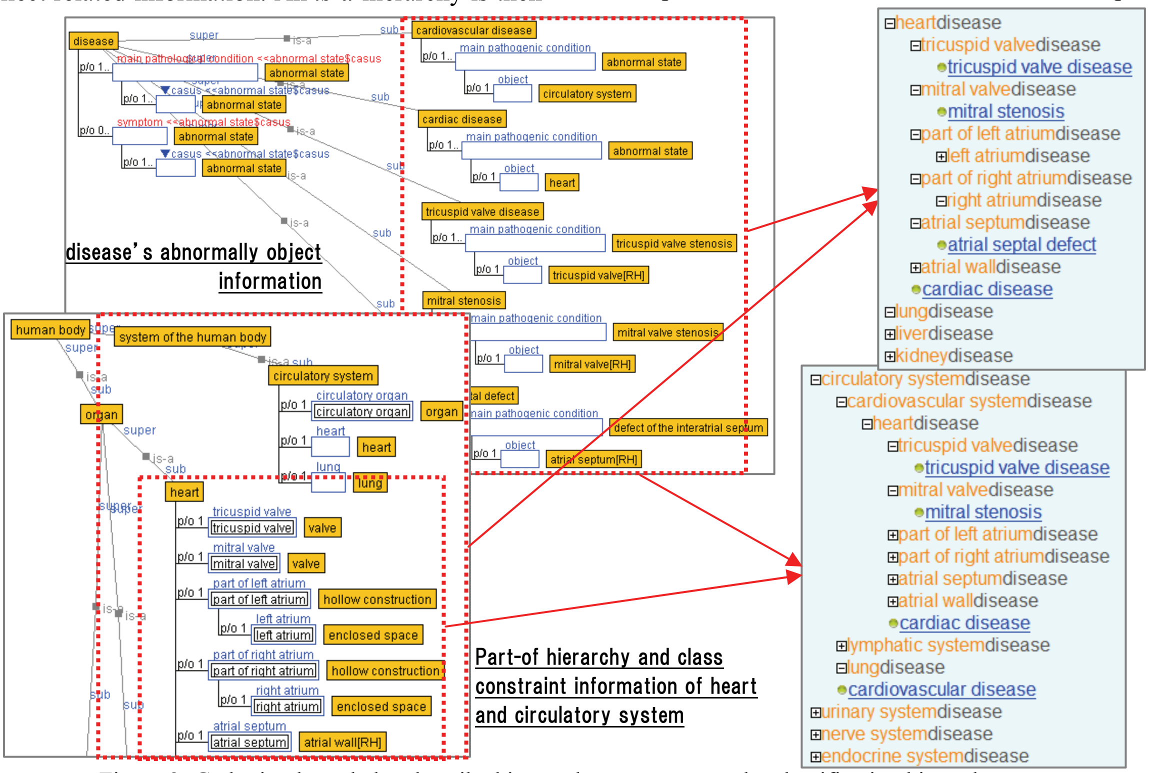

Figure 2: Gathering knowledge described in ontology to generate the classification hierarchy with paying attention to particular thing that caused the abnormal finding

Table 1: Some hierarchical classifications

(a) Viewpoints for classification of diseases		(b) Aspects for dynamic generation of is-a hierarchies of diseases from ontology
finding site (organs)	(m)	organs of the human body (part-of), main pathological condition information (slot)
	(s)	organs of the human body (part-of), symptoms information(slot)
finding site (organ systems)	(m)	organ systems(part-of), main pathological condition information (slot)
	(s)	organ systems, symptoms information(slot)
types of pathogenic abnormality		pathogenic abnormality information (is-a), main pathological condition information (slot)
types of symptoms		symptoms of abnormality hierarchy information(is-a),
first choice for diagnosis and treatment department		diagnosis and treatment department information, symptoms information(slot)
ICPC2		Mapping
ICD10		Mapping

* (m): main pathological condition occurred, (s): symptoms

whole relationship with the heart. Mitral stenosis and mitral prolapse are classified as diseases of the mitral valve, and tricuspid stenosis is classified as a disease of the tricuspid valve. Furthermore, many different classifications are possible for human organs. For example, the heart might be classified occasionally as just a human organ, and on other occasions as a circulatory organ. Converting between these *is-a* hierarchies can be achieved by viewpoint switching (Figure 2).

Thus, the dynamic generation of an *is-a* hierarchy enables us to switch some *is-a* hierarchies using collected information. To collect information, we should trace not only *is-a* and *part-of* relationships but also the relationships based on the role played by a particular concept. In addition, as in ICD-10 and ICPC2 (ICPC2), there are *is-a* hierarchies that were established as useful to those particular purposes. We listed generally needed *is-a* hierarchies as determined by a general medical textbook, some medical information web services, and the opinion of clinicians. For comparison, the matrix below lists generally needed *is-a* hierarchies, the kind of information in the ontology used to generate *is-a* hierarchy (see. Table 1).

There are many advantages in choosing not to realize all the classifications within an ontology, but to dynamically generate *is-a* hierarchy from the ontology. For instance, the inconvenience of multiple inheritance might be averted as discussed above. A longstanding problem is the incompatibility of theory and practice in the use of *is-a* relation: ontological theory tells us that *is-a* relation must be used only between concepts which the lower concepts are genuine subconcepts inheriting essential property from their upper concept, and hence *is-a* hierarchy is unique, while concepts are often classified into multiple classes in practice. Our method proposed in this paper is a good compromise of this conflicting situation. It

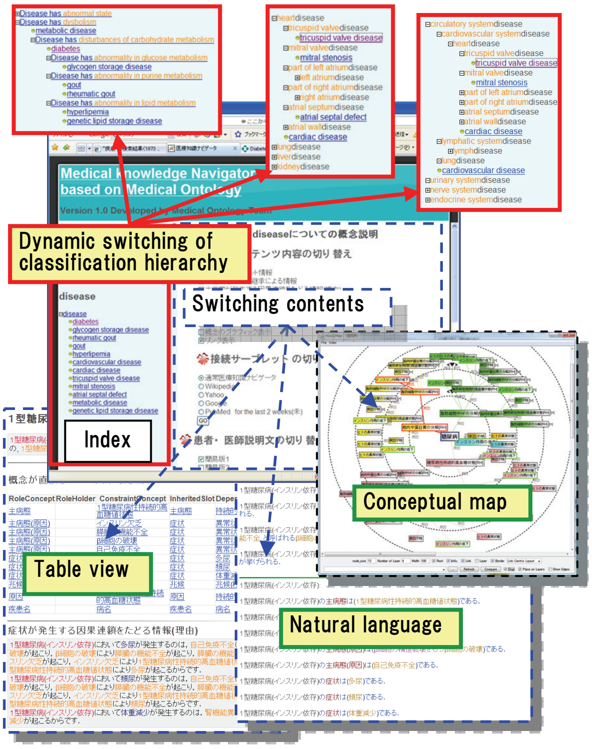

Figure 3: Switching the index using the dynamic generation of *is-a* hierarchy function, and switching the content with the natural language explanation.

allows people to consider their particular viewpoint is "the" essential aspect to them without any harm to others. This function could readily apprehend systematic knowledge, for example, about diseases that have many facets, from various perspectives. This function would also support medical students in systematically understanding that knowledge effectively. Furthermore, this function is important on a clinical site in which many specialists with varied backgrounds cooperate.

3.4 Implementation of Medical Information Service System

We implemented fundamental technology discussed above, and built a prototype of medical information service system using them. They are developed using HozoCore, which is java API for ontology building by Hozo, and Java Servlet. Figure 3 shows the prototype of medical information service system. The system generates *is-a* hierarchies according to the specified user's intentions. When a concept in the index is clicked, detailed explanation is displayed. It changes the medium used for expressing content, such as a table or natural language. For the natural language explanations, the users can choose out of the three types according to their understanding about diseases. When the system displays a network of complex relationships such as causal chains, the user can use the conceptual map generation tool developed in our laboratory. It can generate conceptual maps based on any viewpoint and help users understand the knowledge extracted from ontologies (Kozaki, K. 2008).

We performed an informal evaluation of the implemented system in a workshop and received favorable comments from medical experts. They especially liked the dynamic *is-a* hierarchy reorganization, which is the first solution to the multi-perspective issues of medical knowledge in the world. We also conducted a feasibility study of the fundamental technologies for building application systems using medical ontologies and found that the technologies are independent of the size of the ontology and are a common foundation for the development of various systems.

4 CONCLUSIONS

After the conceptual framework of human anatomy and disease were validated, we developed data input software and templates to scale up the content of the ontology. Our next tasks are to check data consistency, provide feedback regarding the checked data to the experts who input the data, and improve and adjust the ontology. Some problems still need to be resolved. We need to deepen investigation of fundamental technology regarding each of the three main parts (navigation method, content processing, and media selection). Also, we should identify functional enhancements of this fundamental technology based on users' needs and implement an enhanced version. For the navigation method, the added functionality of presenting ancillary information of linked desired content pages would help users decide whether to obtain information they want, and this function might respond effectively to user needs. So we have to consider additional functions for better navigation based on users' needs.

Also, the search system needs improvement of information retrieval. Determining what kind of searches are required in medicine, providing search functions suitable to medicine, and functional extension. Especially, in medicine, it is important not only to search for diseases based on simple symptoms, but also to find all diseases that may cause the patient's symptoms. In media selection, it is necessary that the optimal media be chosen according to timing and case. This selection method is also a subject of ease of using.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Health, Labour and Welfare, Japan, as Development business and research of "Medical-knowledge-based database for medical informatics system."

REFERENCES

DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering, http://www.loa-cnr.it/DOLCE.html

FMA: The Foundational Model of Anatomy ontology,UW Medicine,http://sig.biostr.washington.edu/projects/fm/

GALEN: OpenGALEN, http://www.opengalen.org/

ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision, http://www.who.int/classifications/icd/en/

ICPC2: International Classification of Primary Care, Seco nd edition, http://www.who.int/classifications/icd/adap tations/icpc2/en/

Kozaki, K., et.al, 2002. Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of "Role" and "Relationship". In *Proc. of EKAW2002, pp.213-218.*

Kozaki, K., Hirota, T., Mizoguchi, R., 2008. Development of a Conceptual Map Generation Tool for Exploring Ontologies, In *Poster Noets of ESWC 2008*.

MEDIS-DC: Medical Information System Development Center, http://www.medis.or.jp/

MeSH: Medical Subject Headings, http://www.nlm.nih.gov/mesh/

Mizoguchi, R., et.al, 2007. A Model of Roles within an Ontology Development Tool: Hozo, In *Journal of Applied Ontology, 2, pp.159-179.*

Mizoguchi, R., et.al., 2009. An Advanced Clinical Ontology, In *Proc. of ICBO, pp.119-122*

Rector, A., 2002. Analysis of propagation along transitive roles: Formalisation of the GALEN experience with Medical Ontologies, In *International Workshop on Description Logics, CEUR-Proceedings 53.*

Rosse, C., et.al, 2003. A reference ontology for biomedical informatics: the Foundational Model of Anatomy, In *Journal of Biomedical Informatics, 36, pp.478-500.*

SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms, http://www.ihtsdo.org/snomed-ct/

SNOMED Clinical Terms User Guide, The International Health Terminology Standards Development Organisation, 2008

Stefan, S., et.al, 2007. SNOMED CT's Problem List: Ontologists' and Logicians' Therapy Suggestions. In *Proc. of the Medinfo 2007 Congress, Studies in Health Technology and Informatics (SHTI-series).*

YAMATO: Yet Another More Advanced Top-level Ontology, http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm