Original Article

Developing a Reference Terminology Model for Health Care Using an Object-Oriented Approach

Seung-Hee Kim[1], Yoo-Kyung Boo[1], Yoon Kim[1]

R & D Center for Interoperable EHR[1]

Abstract

Objective: A reference terminology is essential to achieve semantic interoperability and enhance the quality of health care. Reference terminologies that have achieved common acceptance contain many concepts that clinicians would not want in healthcare, which preclude their practical use in documentation of patient information. To solve the problems, this document proposes a reference terminology model which contains concepts that physicians can use satisfactorily. **Methods:** We analyzed the structures of the UMLS and SNOMED CT. We also analyzed health care terms which had been collected by the Korea National Health Information Standard Committee. Based on the results of the analysis, we developed an object-oriented reference terminology model. And, we designed database schema with the model. **Results:** Eight components of the UMLS and six components of the SNOMED CT were analyzed. The collected terms had various properties and mapping vocabularies according to the characteristics of their respective domains. A reference terminology model was developed from a three-level view using UML. A database schema was developed using ERD. **Conclusion:** This study mainly focuses on reference terminology modeling. It is hoped that this reference terminology modeling helps the semantic interoperable exchange of clinical documents as the basis of common EMR. ***(Journal of Korean Society of Medical Informatics 13-2, 83-89, 2007)***

Key words: Reference Terminology, Object-oriented Modeling, Unified Medical Language System, Systematized Nomenclature of Medicine, Database schema

Ⅰ. Introduction

The foundation for electronic health records (EHR) is a comprehensive reference terminology (RT) whose structure unambiguously represents concepts by using a knowledge-based approach[1]. Also, an RT is needed to achieve increasingly greater "semantic interoperability", which can be defined as the ability of two applications to share data with no prior negotiations. An RT is the set of canonical concepts, their structure, relationships and, if present, their systematic and formal definitions[2]. It can help standardize information processing among different organizations and thus reduce the overall cost of doing business[3].

Received for review: April 15, 2007; **Accepted for publication:** June 26, 2007.

Corresponding Author: Yoon Kim, Annex Building, Seoul National University, College of Medicine, 199-1 Dongsoong-Dong, Jongno-Gu, Seoul, Korea, 110-810

Tel: +82-2-741-1260, **Fax:** +82-2-741-1240, **E-mail:** yoonkim@snu.ac.kr

*This work was supported by a research grant of the Korea Health 21 R&D project, Ministry of Health and Welfare of Korea (A050909).

In this paper, a well-structured RT model is proposed for developing a comparable, reusable, multipurpose, and maintainable for health care. This backbone will primarily serve to support semantic interoperable clinical documents exchange. Other obvious challenges still need to be overcome, such as natural language processing, indexing medical records, indexing medical literature, and representing medical knowledge.

Though many terminologies are used in EHR systems, there is no proper well-structured RT to use in Korean health care. In fact, no one has even tried to develop an RT to serve as reference for our health care environment. As a pilot study, we propose a well-structured model for a comprehensive clinical reference terminology that differs from the existing RTs.

To achieve this, we first analyzed the structures and contents of existing terminologies (e.g. UMLS, SNOMED CT) and also analyzed collected terms by the Korea National Health Information Standard Committee (KNHISC). Next, we described an object-oriented modeling method to make the RT model using the Unified Modeling Language (UML). Finally, a relational database(DB) schema was designed using the Entity-Relationship diagram (ERD).

Ⅱ. Materials and Methods

To develop the reference terminology backbone, we analyzed the structures of UMLS and SNOMED CT. We also analyzed collected terms by the KNHISC that will be included in the object-oriented RT model. Based on the results of the analysis, we constructed the object-oriented model using UML. Finally, we designed a DB schema to construct RT.

The overall research process is described in Fig. 1.

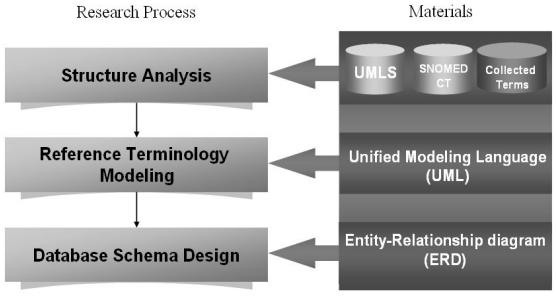

Figure 1. Study Framework

1. Structure Analysis

We analyzed the most prominent terminologies －UMLS, SNOMED CT and also analyzed collected terms used in Korean health care.

(1) UMLS Structure[4)]

The purpose of the UMLS (Unified Medical Language System) is to facilitate the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health. The UMLS combines many well-established authoritative medical informatics terminologies in one knowledge representation system[5)]. It provides a common structure for more than 100 biomedical vocabularies.

The UMLS Metathesaurus(MT) is a very large, multi-purpose, and multi-lingual vocabulary DB that contains information about biomedical and health-related concepts, their various names, and the relationships among them.

It has four components which are attributes, relationships, data about the MT, and the fourth component is concepts, concept names, and their sources. The attributes and data about the MT component are mixed a variety of information. To put pure information into each component, we regrouped into eight components according to their common characteristics.

(2) SNOMED CT Structure[6)]

The SNOMED CT (Systemized Nomenclature of Medicine - Clinical Terms) is the largest, most comprehensive, and precise clinical RT. This terminology enables clinicians, researchers, and patients to share clinical data worldwide, across medical specialties and sites of care. The terminology is comprised of concepts, terms, and relationships that are necessary to represent clinical information across the scope of health care with precision.

The SNOMED CT structure is composed of core tables, history mechanism, subset mechanism, cross-mapping mechanism, and developer toolkit. The core tables have concept and relationship information. To put pure information into each component, we regrouped into six components according to their common characteristics.

(3) The Structure of Korean Health Care Terminology[7)]

The NHISC collected approximately 2.4 hundred thousand terms used in 12 domains－Clinical Findings and Problems, Diagnoses, Operations and Procedures, Nursing, Laboratory, Imaging, Medical Device, Dentistry, Oriental Medicine, Pharmaceuticals, Public Health, and Health Statistics. The reason for collecting these terms was to standardize terminology across the health care domains. The collected terms consisted of simple concept lists and their respective attributes. The NHISC has subsequently been revising these terms to improve accuracy. Until now, 92,063 concepts (about 38% of all terms) in nine domains have been revised. We analyzed these revised terms to determine their attributes and the relation of their attributes to their respective domains.

2. Reference Terminology Modeling

(1) Object-Oriented (OO) Approach[8)]

The OO approach combines data and processes (called methods) into single entities called objects. The main principle in the OO approach is that of abstraction, not of data structures and processes separately but together[9)]. Objects usually correspond to the real referents an information system deals with, such as customers, suppliers, contracts, and rental agreements. The OO model is able to thoroughly represent complex relationships and to represent data and data processing with a consistent notation, which allows an easier blending of analysis and design in an evolutionary process. The goal of the OO approach is to make system elements more reusable, thus improving system quality and the productivity of systems analysis and design[10)].

We used the OO approach method to construct a reference terminology model for comprehensive Korean health care terms in a simplified and homogeneous way. This method was based on the grouping of concepts with the same set of properties as instances of the same object class. The OO model was developed using UML which is a language independent notation system allowing the specification of classes, their data or attributes and methods, inheritance, and other more general relationships between classes[11)].

Our model was constructed from a three-level view based on structure analyses. The three views correspond to the level of abstraction.

3. Database Schema Design[12)]

DB schema is a description of the DB structure[13)]. The DB can be generated from either Entity Relationship Diagram (ERD) or Class Diagram. ERD is a major data modeling tool and helps organize data into entities and define the relationships between them. This process has proved to produce a good DB structure that allows the data to be stored and retrieved in an efficient manner. Using a graphical format enhance communication about the design between the designer and the user as well as between the designer and the people who implement it. For these reasons, we generated a DB schema from the OO model using ERD.

Ⅲ. Results

1. Structure Analysis

(1) UMLS Structure

The four components of the UMLS were regrouped into eight components according to each property. The attributes component was divided into three components which are concept, semantic type, and history. The relationship component was divided into two components which are relationship and history. The data about the MT component was divided into four components which are metadata, concept, history, and source. The fourth component did not only divide into several components. This component was concept component. Table 1 shows how regrouped components are connected with UMLS.

The Metadata component consists of data about the MT itself. The Concept component includes all information pertaining to the concepts in UMLS. The Semantic type component assigns basic semantic types to categorize the concepts. The History component has data that can track changes between the current version and the previous version. The Relationship component defines the relationships that may hold between the semantic types

and concepts. The Mapping component shows the mapping concepts between the two different source vocabularies. The Source component presents a complete list of the source vocabularies in UMLS. The Index component assists system developers in building applications that retrieve all strings or concept names.

(2) SNOMED CT Structure

A reconstruction of the SNOMED CT structure is shown in Table 2. We separated only core tables into the concept and relationship components. The subset component refers to the set that is appropriate to a particular subject. The developer toolkit component is similar to the UMLS Index component. The concept, relationship, mapping, and history components are the same as those in the UMLS.

(3) The Structure of Korean Health Care Terminology

Because the collected terms are organized into a list, they have a concept component and because they are collected for standardization purpose, they have a mapping component as well. But the collected terms have no other components.

Until now, nine domains－Clinical Findings and Problems, Diagnoses, Operations and Procedures, Nursing, Laboratory, Imaging, Dentistry, Oriental Medicine, and Public Health have been revised. The Dentistry domain and Oriental Medicine domain were not included in this paper because so many of their properties are uncommon. As shown in Table 3, the collected terms have various properties and mapping vocabularies according to the characteristics of the domain.

2. Reference Terminology Modeling

The OO model has three views－meta, conceptual, and instance. The meta view is the most abstract and consists of eight elements－concept, semantic class, term, concept relationship, semantic class relationship, term relationship, concept-term relationship, and concept-semantic class relationship. The conceptual view is focused on the properties of the element and is more specific than the meta view. The instance view defines the structure and relationship of elements and is the least abstract level.

From these views, we developed an OO model using UML (Fig. 2). The model contains six classes: Semantic Group, Semantic Class related to itself, Concept related to

Table 1. Regrouped components of the UMLS

Regrouped Component	UMLS tables (.RRF)	Original UMLS component
Metadata	MRFILES, MRCOLS, MRDOC, MRRANK	Data about the MT
Concept	MRCONSO	Concepts, Concept Names, and their sources
	MRSAT, MRDEF	Attributes
	AMBIGLUI, AMBIGSUI	Data about the MT
Semantic type	MRSTY	Attributes
History	MRHIST	Attributes
	CHANGE/MERGEDCUI, CHANGE/DELETEDCUI, CHANGE/DELETEDLUI, CHANGE/DELETEDSUI, MRCUI	Data about the MT
Relationship	MRREL, MRCOC, MRCXT, MRHIER	Relationships
Mapping	MRMAP, MRSMAP	Relationships
Source	MRSAB	Data about the MT
Index	MRXW, MRXNW, MRXNS files	Indexes

Table 2. Regrouped components of SNOMED CT

Regrouped Component	SNOMED CT tables	Original SNOMED CT component
Concept	Concepts, Descriptions	Core Tables
Relationship	Relationships	Core Tables
Subset	Subsets, Subset Members	Subset Mechanism
Mapping	Cross Map Sets, Cross Maps, Cross Map Targets	Cross-Mapping Mechanism
History	Historical Relationships, Component History	History Mechanism
Developer Toolkit	Indexes, Canonical Table, navigation Hierarchies, Subsets, Word Equivalents, Duplicate Terms	Developer Toolkit

itself, Terms related to itself, MapSysConcept which is mapped to concept, and History which has its own concept. Semantic Class is a subclass of Semantic Group and Concept is a subclass of Semantic Class. Term is a part of Concept. A subclass inherits all the members (fields, methods, and nested classes) from its super class.

3. Database Schema Design

Based on the OO model, we constructed a DB schema using ERD (Fig 3.). The schema contained Semantic Tables, Concept Tables, Map Tables, and Meta Tables. The relationships in the Semantic Tables form the relationship between the semantic classes. Semantic Class is included

Table 3. Properties and vocabularies of collected terms

Domain	Property	Mapping vocabulary
Clinical Findings and Problems	Concept identifier, Concept name, Creative date, Expire date, Status, Synonym, Source	UMLS
Diagnoses	Concept identifier, Concept name, Synonym, uncommon properties	KCD4, UMLS
Operations and Procedures	Concept identifier, Concept name, Source of data, Format type, Length, Creative date, Expire date, Status	EDI, UMLS
Nursing	Concept identifier, Concept name	NPG, PNDS, NANDA, NIC, Nurse's Pocket Guide, UMLS
Laboratory	Concept Identifier, Concept Name, Synonym, Component, Property, Time_Aspct, System, Scale_Typ, Method_typ, Class, Insurance code	LOINC, EDI, UMLS
Imaging	Concept identifier, Concept name, Classified name	EDI, UMLS
Public Health	Service, Concept Identifier, Concept name, Semantic Type, Definition, Source of data, Synonym	UMLS

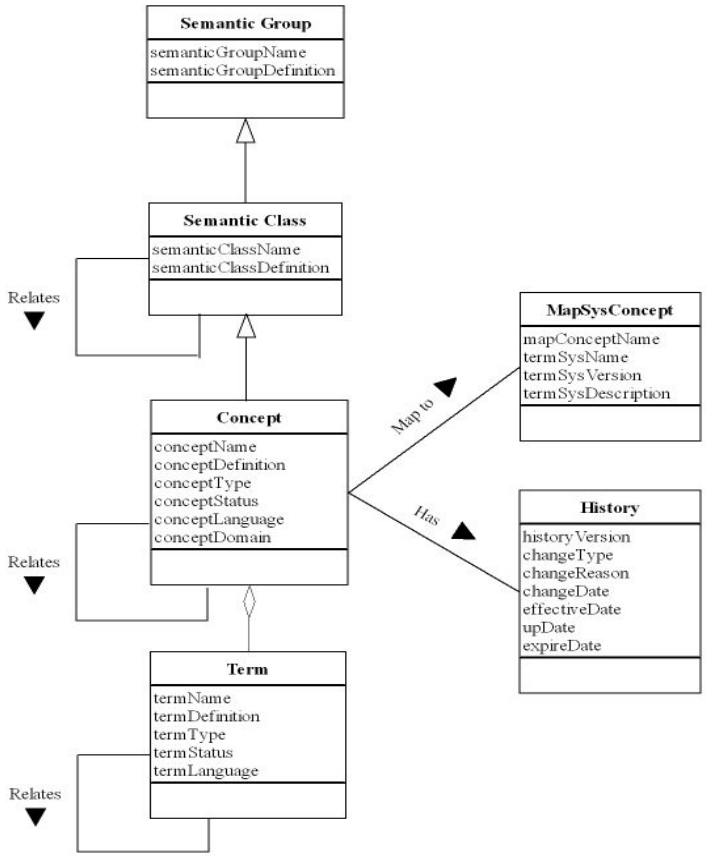

Figure 2. Reference terminology model for health care

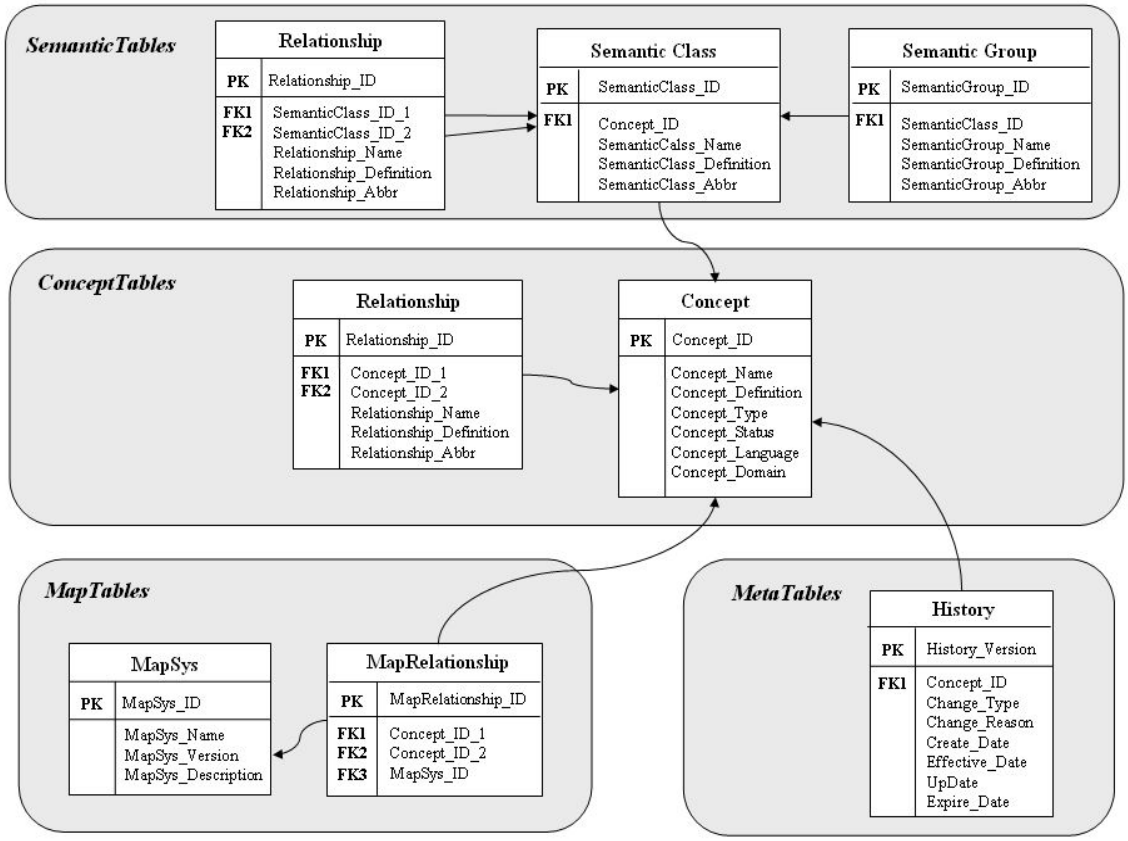

Figure 3. DB schema for reference terminology in health care

in Semantic Group and includes Concept in Concept Tables. The relationships between the concepts are expressed through Relationship in Concept Tables. Also, their relationships with different vocabularies are expressed through MapRelationship in Map Tables. MapSys contains other information related to the vocabularies, for example, the vocabulary version, name, etc. History in Meta Tables contains concept information related to past events.

Ⅳ. Discussion

To provide a well-structured backbone for RT in Korean health care, we constructed an OO model based on structure analyses of UMLS, SNOMED CT, and collected terms. We also developed a DB schema to support the OO model.

The UMLS is a complex collection of medical concepts, terms, and relationships issued from standard classifications[14]. The enormous size of the UMLS means we have to customize the UMLS to make it suitable for our environment. Because the entire UMLS is unnecessary and its default preferred name does not best suit our healthcare[15]. But customizing the UMLS alone is not a satisfactory solution as the UMLS does not have Hangul yet and the mapping rate between the concepts in MT and the collected terms is not high[16].

Further, the SNOMED CT is not freely available. Like UMLS, the SNOMED CT does not have Hangul either and does not have a high mapping rate between its concepts and those in the discharge summary notes[17].

Our model was made to include appropriate Korean health care terms. Not only were properties in all twelve domains included but common properties were as well to fit the model for reference terminology. At this point, however, we have not put the collected terms into the OO model yet. This work will be carried out later as our model could yet change.

In this study, a well-structured backbone for terms in Korean health care was developed. To do this, two vocabularies were analyzed along with our own terms. The analysis mainly focused on modeling reference terminology for Korean health care terms. Through this study, it was found that using the UMLS and SNOMED CT in Korean health care poses various inconveniences. With the results of the modeling, a DB schema was designed to construct RT terminology. It is hoped that this RT will help facilitate the semantic interoperable exchange of clinical documents as the basis of a common EMR.

Reference

1. ISO/TC215 WG1. Health informatics, requirements for an electronic health record reference architecture. Published Standard ISO/TS 18308:2004. January 22, 2004.
2. ISO/TC215 WG 3. Standard specification for quality indicators for controlled health vocabularies. ASTM Committee E31:2000. July 2000. Report No.: E2087.
3. Liu L, Halper M, Gu H, Geller J. Modeling a vocabulary in an object-oriented database. Proc 5th Int Conf Inf Knowledge Manage (1996) 179-188.
4. Unified Medical Language System. U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, National Institutes of Health. Available at: http://www.nlm.nih.gov/research/UMLS/. Accessed March 7, 2007.
5. Gu H, Peal Y, Geller J, Halper M, Liu LM, Cimino JJ. Representing the umls as an object-oriented database: modeling issues and advantages. J Am Med Inform. Assoc 2000; 7(1): 66-80.
6. SNOMED International Web site. Available at: www.snomed.org. Accessed March 7, 2007.
7. The National Health Information Standard Committee Web site. Available at: http://www.medistds.or.kr/. Accessed March 7, 2007.
8. Wang J. Object-Oriented Analysis Methodology. Available at: http://www.umsl.edu/~sauter/analysis/488_f01_papers/wang.htm. Accessed April 9, 2007.
9. The Object Oriented Approach. Available at: http://www.ecs.soton.ac.uk/~phl/ctit/oo/node3.html. Accessed April 9, 2007.
10. Hoffer JA, George J, Valacich J. Modern Systems Analysis and Design. 3rd ed. Upper Saddle River, N.J.: Prentice Hall, 2002.
11. Collins-Cope M. Object Oriented Analysis and Design Using UML. A whitepaper of Ratio Group.
12. Developing Entity Relationship Diagrams (ERDs). Available at: http://webfuse.cqu.edu.au/Courses/2006/T2/COIS20025/Assessment/Item_2/Part_A_Resources/erd.pdf. Accessed April 13, 2007.
13. WIKIPEDIA. Available at: http://en.wikipedia.org/wiki/Schema. Accessed April 13, 2007.
14. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med.1993; 32: 281-91.
15. Hole WT, Humphreys BL, Srinivasan S, Roth L. Customizing the UMLS Metathesaurus for Your Applications. AMIA 2000 Fall Symposium, Tutorial T13. November 4, 2000.
16. Boo YK, Kim SH, Kwak MS, Yoon SY, Kim Y. Mapping of Korean Healthcare Vocabulary to the UMLS. Poster of the Medinfo 2007 Congress; 2007 Aug 20-24; Brisbane, Australia.
17. Kim SH, Han SB, Choi J. The Expressive Power of SNOMED-CT Compared with the Discharge Summaries. J Korean Soc Med Inform. 2005 Sep; 11(3): 265-272.