

Development of a Customized SNOMED-CT Toolkit for Efficient Clinical Term Searches and Mapping Biobank Resources in Korea

Hyun Sang Park¹, Hwa Sun Kim², Hune Cho^{1□}

¹Department of Medical Informatics, Kyungpook National University,
Daegu 700-842, South Korea

²Department of Medical Information Technology, Daegu Haany University,
Gyongsan 700-715, South Korea

Abstract. This study developed a customized SNOMED-CT toolkit that enables a user to search and map clinical terms for efficient utilization of biobank resources. The developed toolkit used the Java programming language, JDBC, and SWING, and provides the user SNOMED-CT concept searching, pre-coordinated mapping, post-coordinated mapping, and clinical item searching in conjunction with a previously developed research database. In addition, using mapping tables of other terminology provided not only the detailed SNOMED-CT information for the searched concept, but also reference information.

Keywords: Biological Specimen Bank, SNOMED-CT, Customized Toolkit

1 Introduction

Biobanks worldwide have collaborated to construct a Biobank Information Management System (BIMS) for the efficient management of collected biological data, to enable the development of customized methods for treating and preventing disease. Each BIMS was developed with the aim of providing high-quality information to a large number of researchers, regardless of affiliation. However, the BIMS experienced difficulties with information exchange because incompatibility problems between heterogeneous systems, such as the Electronic Medical Record System (EMRS) [1].

National projects in the United States and Europe have constructed network environments that facilitate the exchange of biological data between biobanks, such as the Cancer Biomedical Informatics Grid (caBIG) [2] and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) [3]. Therefore, we studied database development in order to integrate biological data managed by 15 university hospital biobanks in the Republic of Korea, together with the National Biobank of

□ Corresponding Author: Hune Cho (Kyungpook National University.) Tel: +82-53-420-4896,
[email: medinfo@knu.ac.kr](mailto:medinfo@knu.ac.kr)

Korea (NBK). To ensure semantic interoperability of the developed database, clinical items were mapped to the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) and specimen items to the Logical Observation Identifiers Names and Codes (LOINC). In total, 2,504 (70.4%) items were mapped, while the remainder was withheld because of uncertainties in the collected data. The most efficient means of maximizing the semantic interoperability of the database would be to map the biobank data that could enhance the understanding of the significance of an item to the international terminology directly, and it is necessary to develop a tool that can be easily maintained. This study developed a customized toolkit that provides clinical item searching and SNOMED-CT mapping, editing, and viewing upon accessing the database for all users who wish to take advantage of the biological data stored in the biobank. Specifically, experts designed the function and user interface of the toolkit by means of a requirement analysis, and implemented a linkage output function with other terminology.

2 Requirement Analysis and Design

First, the requirements were analyzed by three SNOMED-CT experts so that they could develop a toolkit for efficient utilization of the biological data in the biobank. The toolkit included various functions, such as searching clinical items and mapping of clinical items with concept identifiers, viewing concept information, and editing mapped information in connection with the database (Figure 1).

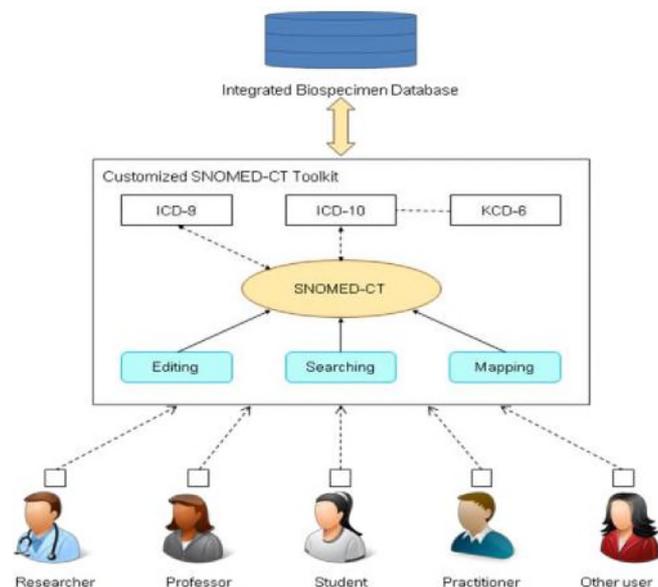


Fig. 1. Architecture of customized SNOMED-CT toolkit based on expert's requirement analysis.

In particular, we implemented an editing function that automatically manages and inputs the mapping results, and a reference terminology (ICD-9, ICD-10, KCD-6) function that utilizes the benefits of SNOMED-CT for differentiating existing toolkits [4-7]. To provide clinical item information and database connection, we added nine new tables to the existing entity relationship diagram (ERD) (Figure 2).

The ERD of the clinical domain consisted of seven tables, including biobank and document information tables (t_bank, t_document), clinical item classification information tables (t_high_class, t_middle_class, and t_low_class), a raw data table (t_clinical_raw_data), and the t_clinical_item table, which contains metadata for 1,796 clinical items. To meet the requirements, key SNOMED-CT tables (t_concepts, t_descriptions, and t_relationships) were added, to create a relationship between the concept field in t_clinical_item and the primary key in t_concepts. Other terminology tables (t_icd9, t_icd10, t_kcd6) were added and mapping tables were placed between the t_concepts. The mapping tables used were from an official source in the Unified Medical Language System (UMLS) that is updated consistently [7].

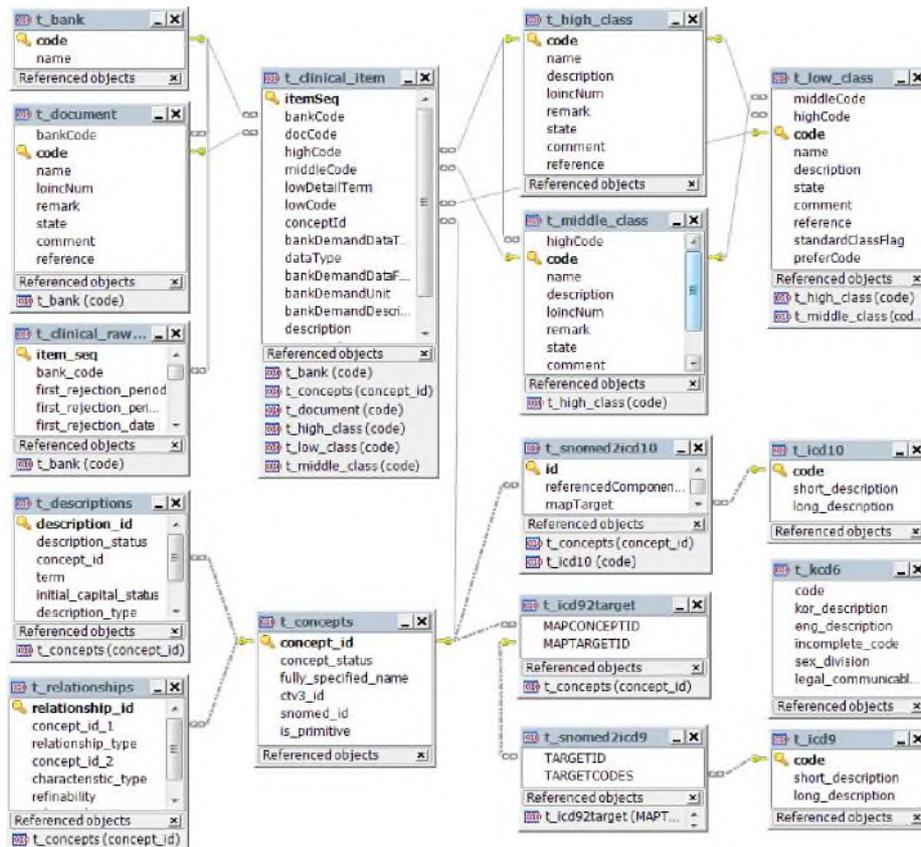


Fig. 2. Entity relationship diagram of customized SNOMED-CT toolkit.

search results table outputs information on the searched term by using the term fields `t_descriptions` and `description_type`.

The hierarchy function outputs parent-child concepts with an “is-a” relationship using the `t_relationships` fields based on the selected concept.

The center panel provides three functions: the top level of the selected concept, detailed information about the concept, and information on other terminology. The top level outputs information about the corresponding concept of the 19 top-level concepts using a recursive call method as used for exploring the parent concept in the hierarchy.

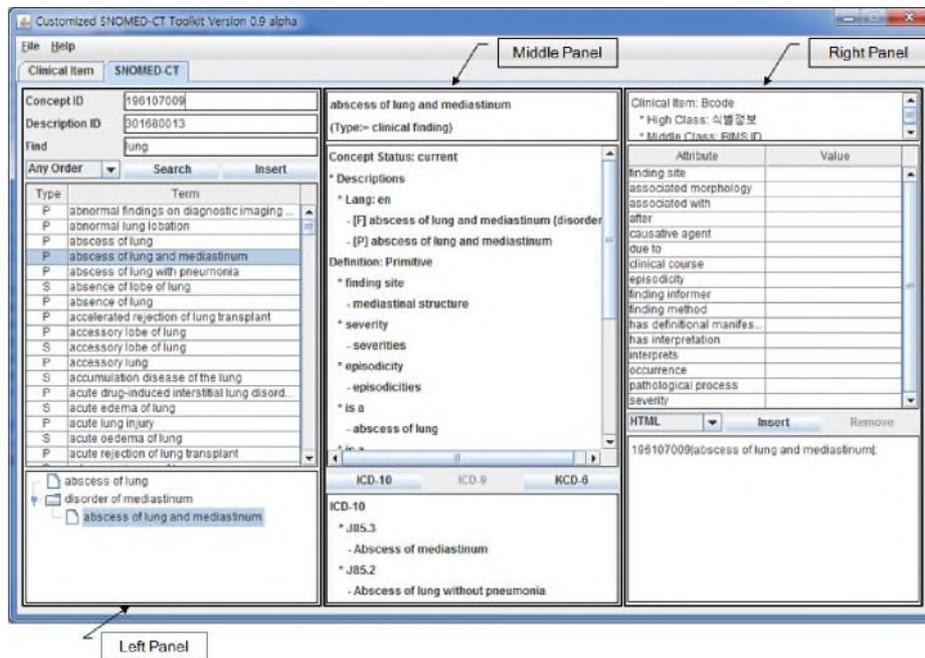


Fig. 4. Screen of SNOMED-CT tab.

Top-level information is output along with the fully specified name (FSN) of the selected concept, which uses the `fully_specified_name` field of the `t_concept` table. Detailed concept information outputs the preferred term, synonym, FSN, parent concept, and relationship type of the selected concept by using all fields in the SNOMED-CT key tables. Other terminology is output to the code and description of the term through the activated button when the identifier of the selected concept is present in the mapping tables. The right panel provides two functions for SNOMED-CT mapping: clinical item metadata output and post-coordinated mapping. Clinical item metadata output is a function for automatically outputting information relating to the items necessary for SNOMED-CT mapping by the recording unit, which utilizes the results of the item selected in the clinical item tab. Post-coordinated mapping is a method for information not represented in pre-coordinated mapping or expressing a

concept in detail. It is expressed by repetition of one or more attribute concepts and the corresponding value concepts related to the selected concept.

4 Discussion

The aim of this research was to maximize the semantic interoperability of a database developed previously that presented a customized SNOMED-CT toolkit for efficient use of biobank resources by users. The requirement analysis by SNOMED-CT experts ensured that the developed toolkit provides optimal functionality to users. Users can easily search clinical item metadata using the developed toolkit, and SNOMED-CT concept searching, pre-coordinated mapping, and post-coordinated mapping can be used for non-mapped items. Unlike existing toolkits, it is possible to confirm the reference terminology information of the selected concept by using mapping tables (ICD-10, KCD-6, ICD-9), and to search the reference terminology, by selecting the appropriate item from a menu. Therefore, users can efficiently search and map international terminology, while maximizing their utilization of biobank data with the developed toolkit, which meets all of the predetermined requirements.

References

1. Hirtzlin, I., Dubreuil, C., Preaubert, N., Duchier, J., Jansen, B., Simon, J., et al.: An empirical survey on biobanking of human genetic material and data in six EU countries. *Eur J Hum Genet.* 11, 475--488 (2003)
2. Oster, S., Langella, S., Hastings, S., Ervin, D., Madduri, R., Phillips, J., et al.: caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical Research. *J Am Med Inform Assoc.* 15, 138--149 (2008)
3. Viertler, C., Zatloukal, K.: Biobanking and Biomolecular Resource Research Infrastructure (BBMRI) Implications for pathology. *Pathologie.* 29, 210--213 (2008)
4. IHTSDO SNOMED CT Browser, <http://browser.ihtsdotools.org/>
5. NPE SNOMED CT Browser, <http://www.snomedbrowser.com/>
6. CliniClue Xplore, <http://www.cliniclue.com/>
7. Unified Medical Language System, <http://www.nlm.nih.gov/research/umls/>