The Development and Application of a Korean Clinical Data Dictionary

Hyeong-Yun Choi[a], Mi-Hyun Kim[a], Jae-Il Lee[a,b], Yoon Kim[a,c], Hong-Ki Kim[d]

[a] R&D Center for Interoperable EHR, Seoul, Republic of Korea
[b] Department of Oral Pathology and Dental Research Institute, College of Dentistry, Seoul National University
[c] Seoul National University College of Medicine, Seoul, Republic of Korea
[d]Biomedical Knowledge Engineering Laboratory, Seoul National University, Seoul, Republic of Korea

Abstract and Objective

Several English-speaking countries have already been developing concept-based clinical data dictionaries for effective data integration and management that guarantees interoperability in EMR environment, but such dictionaries are hardly usable in non-English-speaking countries, and this cannot be resolved simply through translation. In mixture environment Korean and English, we are in need of a clinical data dictionary for overcoming the problem. For that reason, we developed a Korean-type clinical data dictionary and proved its usefulness with EMR. As a result, we confirmed that our Clinical Data Dictionary has structure and contents that can effectively support the integration of data in different clinical environments.

Keywords:

Medical data dictionary, Data dictionary for EMR.

Methods

Stage 1, referring to development methodologies in advanced cases [1], we designed the structure of the Clinical Data Dictionary (CiDD) so that all concepts have their synonyms and association becomes clear not only with the Korean terminology system but also with English-based standard terminology systems. Stage 2, we selected the Korean standard terminology system as seed data and formed contents with 160,828 concepts and 425,219 terms including English and Korean ones. Stage 3, it applied to two hospitals having more than 500 beds, one of which collected mainly terms related to diagnosis, surgery and treatment, and the other gathered all terms used in medical records. Then, the terms were mapped to the contents of the CiDD according to agreed conditions

Results

According to the results of the mapping analysis, The lexical mapping rate for the hospital from which all terms in medical records were collected was not high as around 50%, but the coverage increased to 70% if semantic mapping was included. The hospital from which terms related to diagnosis, surgery and treatment were collected showed a lexical mapping rate of 98%. This result confirmed the fact that our Clinical Data Dictionary can support almost every term related to diagnosis, surgery and treatment, and can cover around 70% of all terms used in clinical services. To supplement the remaining 30%, we analyzed contents obtained from the application of the dictionary in hospitals and established rules for reflecting the contents in it, and expanded its contents according to the rules.

Conclusion

Our Clinical Data Dictionary proven from this study is expected not only to improve the quality of domestic medical services but also to be a good model for the development of clinical data dictionaries in non-English-speaking countries.

Acknowledgement

This research was funded by a grant (#A050909) from the Ministry for Health, Welfare and Family Affairs in Korea.

References

[1] Rocha RA, Huff SM, Haug PJ, Warner HR. Designing a Controlled Medical Vocabulary Server: the Voser Project. Comput. Biomed. Res. 1994;27:472-507