Evaluation of a semi-automated code mapping and management system

Soondong Kim, MCS[1], Aelan Park, RN,CMD, PhD[1], Soo-Yeon Cho, RN, MPH [1], Sukyoung Lee, RN, MPH[1], Hye Jin Kam, PhD[1]
EvidNet, Seongnam-si, Gyeonggi-do, Republic of Korea

Is this the first time you have submitted your work to be displayed at any OHDSI Symposium?

Yes _____ No___V___

Abstract

In CDM (Common Data Model), vocabulary mapping plays a vital role in global-network studies. In order to map Korean vocabularies effectively and efficiently, EvidNet developed an application named 'MASO' (Mapping Assistance System on OMOP), and evaluated the suitability of the vocabulary mapping protocol, efficiency of entire system functions, and usability of candidate target concept providing algorithm. MASO showed good performance in quantity and quality, it provided time efficiency by offering semi-automatic system and increased agreement rate among experts.

Introduction

In CDM, it is indisputable that vocabulary mapping is vital to researches via international networking. EvidNet have chosen a mapping protocol to increase agreement on relationship between Korean and international codes: this mapping process has included initial mapping, cross-check, and the third review[1]. Even though it could draw better agreement, this process leads to some human-error without a management tool. We developed an application named 'MASO' (Mapping Assistance System on OMOP) and evaluated the (1) suitability of the vocabulary mapping protocol, (2) efficiency of the entire system functions, and (3) usability of the algorithm which will be applied on MASO.

Methods

MASO is an application based on Server-Client model to manage mapping process. It was designed with WPF (Windows presentation foundation) and implemented by Node.js. The goals of MASO are to reduce laborious jobs besides vocabulary mapping, and to decrease time spending through automatic recommendation of proper vocabulary candidates. The basic function of MASO is to keep 3 mapping phases of initial mapping, cross-check and review to get better mapping results (Figure 1). We also have implemented essential functions for automation of manual jobs such as source code preprocessing, job assignment, importing, recording, searching, note, discussion, feedback and history management. Also, an algorithm, called RPM (Recommender from Past mappings) was applied on MASO to provide proper candidate concepts from past mapping results. It extracts keywords from the name of source code and searches those from previous mapped source code names. MASO automatically recommends candidates sorted by similarity score with Levenshtein algorithm[2] when a mapping expert tries to map a new source code (Figure 2). Also if the recommendation of candidate seems not proper, the expert can try to put the the other keywords the expert wants on MASO (Figure 3).

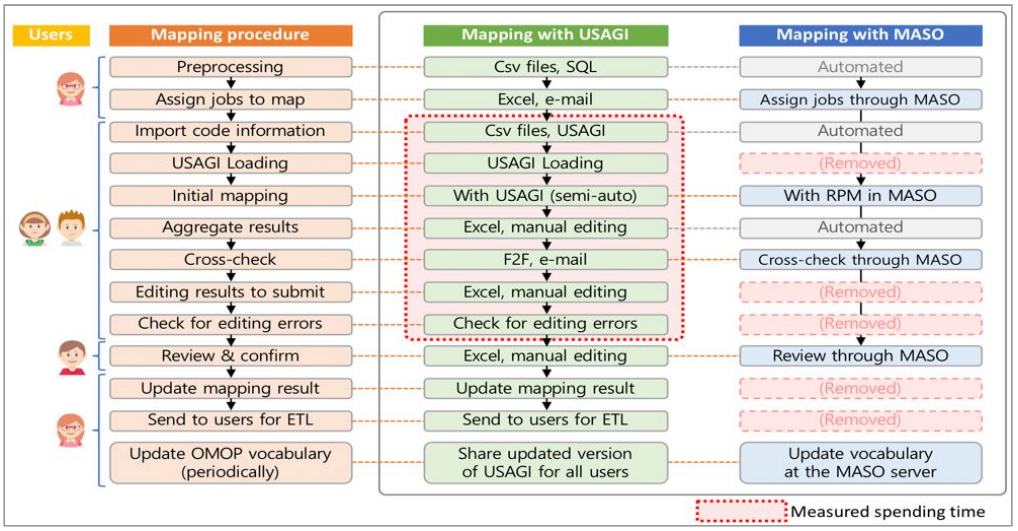

Figure 1. Simplified mapping procedure and activities with MASO.

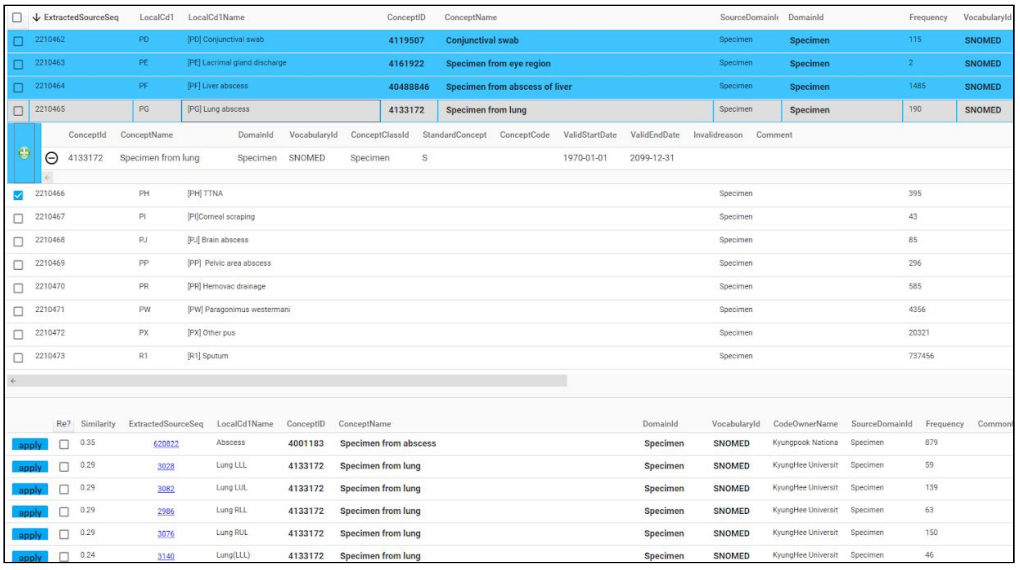

Figure 2. MASO UI for recommending candidate concepts with similarity.

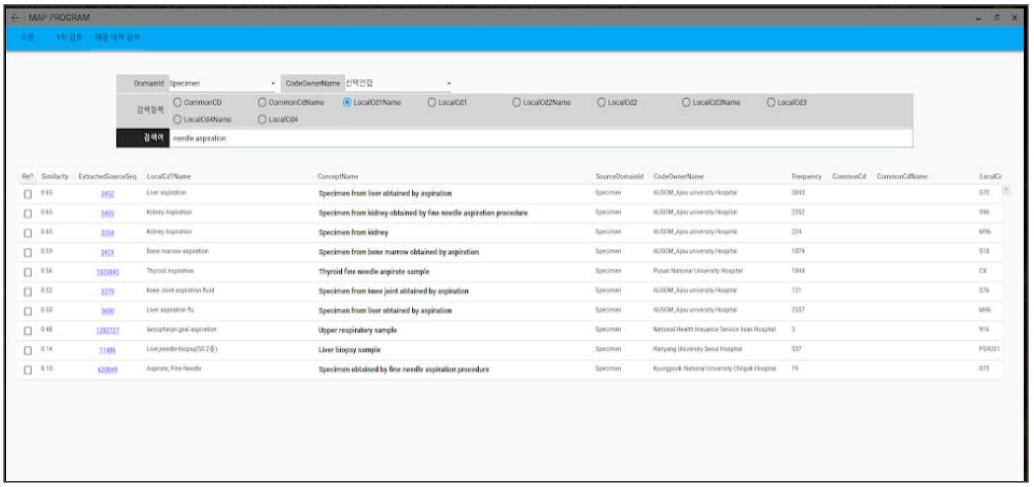

Figure 3. MASO UI for searching keywords and showing past mapping results.

We evaluated our mapping processes with Usagi and MASO in the aspects of 3 factors in the following Table 1.

Table 1. Three factors (Suitability, efficiency, usability) with evaluation criteria and their evaluation methods

Evaluation factor	Evaluation method	source data for evaluation
1. suitability - agreement rate difference	- calculate disagreement rate between expert A vs. B, A vs. C, B vs. C, and cross-checked result (agreed result of A and B) vs. C and compared each rate (initial mapping was done by A and B; C was 3rd reviewer)	- mapped results of Procedure domain (n=3,452)
2. Efficiency - saved time - prevention of errors by hand	- measure average spent time for each action without MASO during mapping procedure from assignment of new job to submission of cross-checked results (e.g. downloading, comparing with other expert's results or organizing the result report)	- hand recorded time for source codes of Specimen domain (n=468) by 2 expert each
	- aggregate survey results from experts for what kinds of errors had they experienced during mapping procedures without MASO	- survey from experts participated in mapping (n=9)
3. Usability - changes of performance	- measure average hitting rate of proper concept, the number of average candidate concepts and expectation rate for proper concep t as completed mapping data increases	- test: mapped results of Specimen domain (n=468) - search source: mapped results of Specimen domain (n=3,635) and OMOP DB (n=1921) of Specimen domain

Results

1 [Suitability] In terms of agreement rate, while comparison of mapping result in expert A vs. C and expert B vs. C was 62.3 - 69.3%, comparison in cross-checked result vs. expert C was 73.6%.

2 [Efficiency] Mapping 468 source code of Specimen took around 12 hours per each expert. It saved 52 mins and 90 mins for expert A and B respectively, with mapping on MASO. According to the survey, all mapping experts have experienced recording errors quite frequently on editing phase of file (n=9). Missing information or file during communication via email or summary files was ranked the second reason of error(n=5).

3 [Usability] As completed mapping results with new institutions increase in numbers, the number of hits and probabilities of source code to hit the proper target concept were increased. The percentage of a concept that fits to perfectly proper target concept per single candidate from the algorithm (hit expectation rate per candidate) was 8.6 to 12.5%: It was 5-7 times higher than that of recommended concepts in OMOP concepts (1.8%) (Figure4).

Table 2. target concept searching performances with Atlas and the newly proposed algorithm

searching source		**number of record (n)**	**number of hits (n)**	**hit rate(%)**	**avg. number of candidates (n)**	**max. number of candidates (n)**	**hit expectation rate per candidate (%)**
DB	**number of site (n)**						
OMOP	-	1,921	201	42.9	24.2	113	1.8
RPM utilizing past mapping result	1	503	196	41.9	3.35	19	2.5
	2	973	217	46.4	4.88	20	9.5
	3	1,380	217	46.4	4.93	20	9.4
	4	1,626	217	46.4	5.19	23	8.9
	5	1,989	217	46.4	5.21	23	8.9
	6	2,515	237	50.6	5.44	23	9.3
	7	3,148	241	51.5	5.99	26	8.6
	8	3,635	252	53.8	5.88	26	9.2

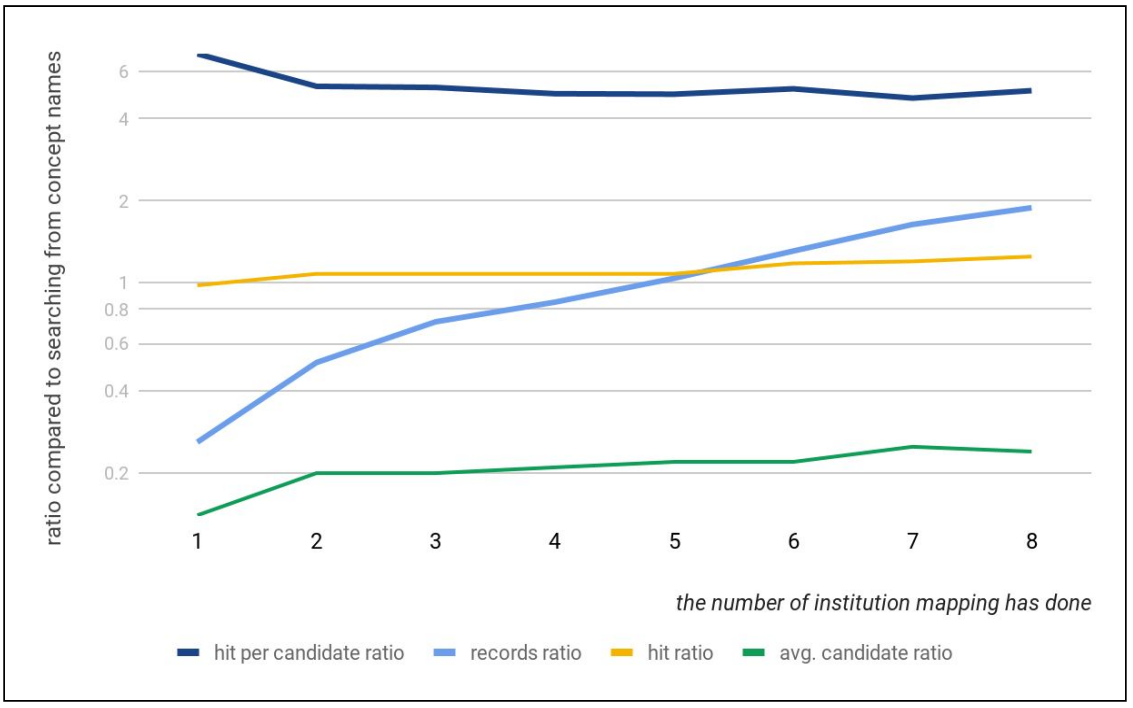

Figure 4. Target concept searching performance ratio of the new algorithm compared to the search from concept names for local source codes.

Discussion

MASO is based on 3 mapping phase with multiple experts to draw the agreement through discussion. With MASO, the agreement rate of vocabulary mapping increased because several experts were involved in mapping process through online discussion without interval. And, it also helped to reduce human-errors as shown by the user survey. Through mapping process on MASO, we expect to save time and raise the quality of vocabulary mapping.

Although Usagi had developed to help mapping process, not only it looks more helpful for codes in English rather than codes in other languages, but it is offline program and saves mapping data in file-form instead of databases exposed to various human errors. For these reasons, we need an algorithm for codes having Korean vocabularies that utilized experience and knowledge from the previous mapping results. We could see the effects of the former mapping knowledge to new mappings as shown in the result section, and when the development of an algorithm is completed to apply to MASO, we expect to obtain higher quality of vocabulary mapping outcome with better efficiency.

Acknowledgement

This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea)

References

1. G Song, HK Woo, DH Shin, SY Cho, HJ Kam. Implementing the OMOP CDM and OMOP Ecosystem in Korea. 2019 OHDSI Symposium poster abstract: 29 MAR 2019
2. https://www.npmjs.com/package/levenshtein