Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200429

Manual Evaluation of the Automatic Mapping of International Classification of Diseases (ICD)-11 (in French)

Julien GROSJEAN[a,b], Kévin BILLEY[a,c], Jean CHARLET[b,d], Stefan J. DARMONI[a,b]
[a]*Department of Biomedical Informatics, Rouen University Hospital, Normandy, France*
[b]*LIMICS U1142 INSERM, Sorbonne Université, Paris & Rouen University, France*
[c]*LITIS EA4108, Rouen University, France*
[d]*Assistance Publique-Hôpitaux de Paris, DRCI, Paris, France*

Abstract. A lexical method was used to map ICD-11 to the terminologies included in the HeTOP server. About half of ICD-11 codes (47.76%) were mapped to at least one concept. The developed tool reached a global precision of 0.98 and a recall of 0.66. Lexical methods are powerful methods to map health terminologies. Supervised and manual mapping is still necessary to complete the mapping.

Keywords. Terminology mapping; Semantics; Terminology as topic; Unified Medical Language System; Vocabulary, controlled

1. Introduction

The International Classification of Diseases (ICD) version 11[1] is a major revision of the ICD-10 which is widely used to code diagnoses in healthcare systems. Furthermore, the purpose of the Unified Medical Language System® (UMLS) [1] is to facilitate the development of computer systems that behave as if they "understand" the meaning of the health language. The process of terminology mapping consists of identifying identical (or approximately identical) concepts or relationships between terminologies [2]. Since 2007, the Rouen University Hospital (RUH) Department of Biomedical Informatics (DBI) has developed the HeTOP cross-lingual health termino-ontology server [3], which includes 75 terminologies (or ontologies) in 32 languages[2].

The objective of this work is to propose a mapping of ICD-11 to the UMLS Metathesaurus and to French terminologies which are not (yet) included in the UMLS.

2. Methods

French natural language processing tools and mapping algorithms were developed by the RUH DBI team to map health terminologies in French. These tools were used in previous works [3][5] and extended to link terms in multiple French health terminologies. This approach allows from a given term, to find a HeTOP concept with French (or English) terms that are most lexically similar to it.

[1]WHO International Classification of Diseases, 11th Revision (ICD-11); https://www.who.int/classifications/icd/en/ (Accessed September 5, 2019).
[2] https://www.hetop.eu/

3. Results

The RUH DBI automatic mapping tool has generated 144,823 ICD-11 mappings to the 75 terminologies integrated into HeTOP. Concerning mappings between ICD-11 and ICD-10, WHO has manually provided 12,767, whereas RUH DBI has provided 6,925. The intersection between WHO and RUH DBI is 3,492 mappings (respectively 27.35% and 50.43%). This French automatic mapping tool has provided at least one mapping for 28,521 ICD-11 distinct codes (using their preferred labels and its synonyms and acronyms) among 59,709 ICD-11 codes (47.76%). The Top 6 terminologies with the most mappings to ICD-11 are SNOMED CT, MeSH and NCIT, which is coherent with the size of these terminologies and their translations as well as their main objectives, respectively 326,927 (201,659 in French), 378,024 (including the MeSH concepts; 105,933 in French) and 148,215 (79,409 in French). Among the 144,823 automatic mappings from ICD-11 to the 75 terminologies integrated into HeTOP, 2,371 were manually supervised by one expert (SJD), using six relations types: EM (exact match), BTNT (broader than), NTBT (narrower than), RT (related), unknown and false. The total number of distinct ICD-11 codes manually curated was 464 (mean mappings number of 5.1 per code). 82% of these automatic mappings were validated in exact match. 15% were evaluated as "close" mappings (i.e. BTNT, NTBT, RT). Only 46 mappings (2%) were evaluated as false. Furthermore, among the 464 ICD-11 curated codes, 183 have no mappings to any HeTOP concept. On the other hand, the expert manually added new mappings to 26 distinct ICD-11 codes without any automatic mappings. Thus the false negative rate is estimated to 26/(183+26)=12.4%. 1,226 new mappings from transitive closure were also added whose 154 automatically validated as Exact Match. Those figures allow an estimation of the recall: 2,371/(2,371+1,226)=65.92%.

4. Discussion and conclusion

To the best of our knowledge, this is the first study to map the entire ICD-11 classification to terminologies in French. The quality of these mappings can be considered as very good, as 82% were evaluated as an "exact match", and only 2% are false. However, the cover of such tool is relatively low (47.76% of all ICD-11 codes) as many labels are too precise or not based on a state-of-the-art terminological method.

References

[1] D.A.B. Lindberg, B.L. Humphreys, A.T. McCray. The unified medical language system. *Methods Inf Med* **32** (1993), 281–91.

[2] A.R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proc AMIA Symp* (2007), 17–21.

[3] J. Grosjean, T. Merabti, B. Dahamna, I. Kergourlay, B. Thirion, L.F. Soualmia, S.J. Darmoni. Health multi-terminology portal: a semantic added-value for patient safety. Stud Health Technol Inform. **166** (2011), 129-38.