Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220483

WikiMeSH: Multi Lingual MeSH Translations via Wikipedia

Mikaël DUSENNE[a], Kévin BILLEY[a], Florent DESGRIPPES[c], Arriel BENIS[d], Stéfan Jacques DARMONI[a,b] and Julien GROSJEAN[a,b,1]
[a] *Department of Digital Health, Rouen University Hospital, France*
[b] *Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), U1142, INSERM, Sorbonne Université, Paris, France*
[c] *LIFEN company, France*
[d] *Faculty of Industrial Engineering and Technology Management, Holon Institute of Technology, Holon, Israel*

Abstract. Objective: The aim of this paper is to propose an extended translation of the MeSH thesaurus based on Wikipedia pages. Methods: A mapping was realized between each MeSH descriptor (preferred terms and synonyms) and corresponding Wikipedia pages. Results: A tool called "WikiMeSH" has been developed. Among the top 20 languages of this study, seven have currently no MeSH translations: Arabic, Catalan, Farsi (Iran), Mandarin Chinese, Korean, Serbian, and Ukrainian. For these seven languages, WikiMeSH is proposing a translation for 47% for Arabic to 34% for Serbian. Conclusion: WikiMeSH is an interesting tool to translate the MeSH thesaurus and other health terminologies and ontologies based on a mapping to Wikipedia pages.

Keywords. MeSH, translating, Wikipedia

1. Introduction

The Medical Subject Headings (MeSH) thesaurus is a controlled and multi-hierarchically organized vocabulary produced by the National Library of Medicine (NLM). It is used for indexing, cataloging, and searching of biomedical and health-related information. MeSH includes the subject headings appearing in MEDLINE/PubMed [1]. Currently, the MeSH is available in 16 languages [2]. In its 2021 version, the MeSH thesaurus contains 29,754 Descriptors. Wikipedia [3] is a free content, multilingual online encyclopedia written and maintained by a community of volunteers through a model of open collaboration, using a wiki-based editing system. The domain Wikipedia.com was created in January 2001. Currently, it contains over 6,400,000 articles in English. All these articles are manually created and maintained. Wikipedia exists in 325 languages in the world. The aim of this paper is to propose an extended translation of the MeSH thesaurus based on Wikipedia pages using a new tool called "WikiMeSH". To our knowledge, no prior work has tested Wikipedia to enhance the multi lingual translation of a reference health terminology: in our example, the MeSH thesaurus. This work has

[1] Corresponding Author, Julien Grosjean; E-mail: julien.grosjean@chu-rouen.fr.

been partially granted by the European project (granted by Horizon 2020) HOspital SMART development based on Artificial Intelligence (AI; HosmartAI).

2. Methods

The goal of this WikiMeSH tool is to find Wikipedia entries in different languages for the descriptors of the MeSH thesaurus. The Wikipedia API allows to find the linguistic links (or mappings) for a given page, allowing for the exploration of a given topic in different languages. For each MeSH descriptor, a search for the corresponding entry on Wikipedia was performed.

3. Results

The WikiMeSH tool has been written in Python 3X and calls the Wikipedia API through the dedicated endpoints depending on the languages. Overall, the WikiMeSH tool was able to map a Wikipedia page for 18,191 MeSH descriptors (61.14%), which means that less than 39% of the MeSH descriptors have no Wikipedia links. 15,268 MeSH descriptors obtained a link based on their preferred term and 2,923 based on a synonym. The average number of detected Wikipedia languages per MeSH descriptor is 17.57 ± 34.80 (min-max: 0 - 300). The average number of added languages thanks to WikiMeSH is 17.61 ± 34.75. Among the top 20 languages of this study, seven are not already officially present in the MeSH thesaurus translations: Arabic, Catalan, Farsi (Iran), Mandarin Chinese, Korean, Serbian, and Ukrainian. For these seven languages, WikiMeSH is proposing a translation for 13,959 out of 29,754 MeSH descriptors (46.91%) for Arabic, 12,039 for Farsi (40.46%), 11,945 for Mandarin (40.16%), 10,839 for Ukrainian (36.43%), 10,770 for Catalan (36.20%), and 10,103 for Serbian (33.96%). The WikiMeSH results are integrated into the crosslingual terminology server HeTOP [4].

4. Conclusion

WikiMeSH is an interesting tool to translate the MeSH thesaurus and other health terminologies and ontologies based on a mapping to Wikipedia pages.

References

[1] Welcome to Medical Subject Headings. Available at https://www.nlm.nih.gov/mesh/meshhome.html (Accessed January, 10 2022).
[2] MSH (MeSH) – Synopsis. Available at: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSH/index.html (Accessed January, 10 2022).
[3] Wikipedia. Available at: https://en.wikipedia.org/wiki/Wikipedia (Accessed January, 11 2022).
[4] Grosjean J, Merabti T, Dahamna B, Kergourlay I, Thirion B, Soualmia LF, Darmoni SJ. Health multi-terminology portal: a semantic added-value for patient safety. Stud Health Technol Inform. 2011;166:129-38. PMID: 21685618.