



HAL
open science

Elaboration of a medical knowledge organization system supported by automatic tools: towards ecological, intelligent and ethical action

Marcin Trzmielewski

► To cite this version:

Marcin Trzmielewski. Elaboration of a medical knowledge organization system supported by automatic tools: towards ecological, intelligent and ethical action. 12' ISKO Italy Meeting "Knowledge Organization and Representation: Convergences and New Opportunities", University of Calabria, Apr 2025, Rende (CS), Italy. <hal-05031555>

HAL Id: hal-05031555

<https://hal.science/hal-05031555v1>

Submitted on 11 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Elaboration of a medical knowledge organization system supported by automatic tools: towards ecological, intelligent and ethical action

Marcin Trzmielewski

Laboratoire d'Etudes et de Recherches Appliquées en Sciences Sociales (LERASS)

Université de Montpellier Paul-Valéry

Marcin.trzmielewski@univ-montp3.fr

While intellectual tools (*e.g.* interview, thematic and facet analyses) are widely used to gather and analyze data for elaboration of medical knowledge organization systems (MKOSs) (Iyer and Raghavan 2018, Raghavan and Sajana 2010), the abundance and the heterogeneity of medical knowledge to be represented and organized lead to consider automatic techniques to minimize tedious, repetitive, time-consuming, costly tasks (Hassan 2017, Toro *et al.* 2024). For these reasons, natural language processing, text mining and machine learning have been used to extract and organize medical terminology and to match terms with corresponding definitions from large textual corpora (Choi 2016, Muresan and Klavans 2013, Tapi Nzali *et al.* 2019). Recently, large language models and retrieval augmented generation have started to be integrated to ontology editing tasks such as prediction of relationships, term definitions, and logical definitions (Liu *et al.* 2024, Toro *et al.* 2024). The use of automatic tools, often connected to ontological approach leads to elaboration of MKOSs in laboratory methods and conditions, mostly on limited samples that often are not part of collections of documents that actually will be organized and sought by potential users, as well as to test only the internal structure of elaborated systems (Author 2024). Such methods pay less attention to the human and socio-organizational aspects, linked to the use of MKOSs in concrete situations (Author 2024). Designers provide some *a priori*, content-related quantitative criteria of assessment, by

calculating the relevance of extracted terms or the performance of MKOSs to accomplish document processing tasks (Lopes *et al.* 2009, Tang *et al.* 2017, Tapi Nzali *et al.* 2019). These methods seem to be reductive, and the use of automatic tools should be connected to epistemological approaches in which designers search for criteria that are “external to the corpus” (Clavier 2010, Tennis 2012). However, such elements are not provided in advance. The aim of our paper is to assess how and where to find them to perform the “right action” (Tennis 2012) while using automatic techniques for the elaboration of MKOSs.

We show some examples from the elaboration of an allergy ontology for the Allergy Unit of the Montpellier University Hospital in France. We have performed a semi-automatic term extraction supported by terminology extractor TermoStat and an automatic text annotation with the Annotator tool to collect useful data for the ontology construction and assessment. These processes were preceded by the study of the context of use of knowledge in the unit, obtained through participant observation and semi-structured interviews.

The inputs from our analyses allowed the identification of a set of useful human, informational and socio-organizational elements, that guided the automatic extraction and annotation tasks. First elements allowed to perform an *ecological action*: involve actors, practices and information resources that are part of the allergy unit environment. We identified and collected documents daily used by professionals, and these documents composed a corpus analyzed with automatic tools. Moreover, the utility of extracted terms to professionals’ practices was assessed by an allergist working in the unit. The utility of the constructed ontology to perform automatic annotation of documents was assessed by allergy professionals as well, because they will be using the MKOS under development to process documents in their daily activities.

Other elements identified by the study of the context use led us to an *intelligent action*. The specialized vocabulary used by professionals, extracted from the content of the observation reports and interview transcripts, as well as user-oriented facets representing informational

targets of the employees, guided the process of automatic extraction of useful candidate terms, discrimination of irrelevant terms, and assessment the performance of the extraction software. During the assessment of the ontology to perform automatic annotations, professionals assessed *utility* dimensions related to the content, intentionality and operationality of this MKOS, defined by allergy-domain actors themselves.

Finally, considering the use of sensible and personal data in our corpus, we must respect French and European legislation to perform an *ethical action*. Our study had to be officially accepted by the hospital Ethics Committee and registered on ClinicalTrials.gov. All personal data that appeared in the analyzed corpus was manually anonymized and concerned persons were informed about data collecting.

Our results revealed that the use of automatic tools is helpful to elaborate MKOSs, but the use of these tools should be guided by elements that are external to the analyzed corpora. We interpreted these elements from the data on context of use and contextual interactions with actors considered as experts and potential users. Such an approach allowed ecological and intelligent actions, as well as the respect of ethical and legal obligations. However, it revealed a paradox: the use of automatic tools supported time-consuming tasks, but qualitative methodology required a considerable amount of time to collect and analyze meaningful data.

Bibliography

Choi, Yunseon. 2016. "Supporting better treatments for meeting health consumers' needs: extracting semantics in social data for representing a consumer health ontology". *Information Research* 21, no. 4.

Clavier, Viviane. 2014. "L'organisation des connaissances au prisme du langage, du texte et du discours. Un parcours en recherche d'information". Tenure diss., Université Stendhal-Grenoble 3.

Hassan, Mohsen. 2017. "Knowledge Discovery Considering Domain Literature and Ontologies: Application to rare Diseases". PhD diss., Université de Lorraine.

Liu H., S. Zhou, Z. Chen, Y. Perl and J. Wang, "Using Generative Large Language Models for Hierarchical Relationship Prediction in Medical Ontologies," *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, Orlando, FL, USA, 2024, pp. 248-256, doi: 10.1109/ICHI61247.2024.00040.

Iyer, Hemalata and K. S. Raghavan. 2018. "Medical ontology: Siddha System of Medicine". In *Challenges and Opportunities for Knowledge Organization in the Digital Age: Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal*, edited by: Fernanda Ribeiro and Maria Elisa Cerveira. *Advances in Knowledge Organization* 16. Baden-Baden: Ergon, 347-355.

Lopes, Lucelene, Renata Vieira, Maria José Finatto, Daniel Martins, Adriano Zanette, Luiz Carlos Riberio Jr. 2009. "Automatic extraction of composite terms for construction of ontologies: an experiment in health care area". *Electronic Journal of Communication, Information & Innovation in Health* 3, no. 1: 72-84.

Muresan, Smaranda and Judith L. Klavans. 2013. "Inducing terminologies from text: A case study for the consumer health domain". *Journal of the American Society for Information Science and Technology* 64, no. 4: 727-744.

Raghavan, K. S. and C. Sajana. 2010. "NeurOn. Modeling ontology for neurosurgery". In *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference 23-26 February 2010 Rome, Italy*, edited by Claudio Gnoli and Fulvio Mazzocchi. *Advances in Knowledge Organization* 12. Würzburg: Ergon, 208-215.

Tang, Xiao-Bo, Wei Wei, Guang-Chao Liu, Juan Zhu. 2017. "An Inference Model of Medical Insurance Fraud Detection: Based on Ontology and SWRL". *Knowledge Organization* 44, no. 2: 84-96.

Tapi Nzali, M. D., J. Aze, S. Bringay, C. Lavergne, C. Mollevi, T. Optiz. 2019. "Reconciliation of patient/doctor vocabulary in a structured resource". *Health Informatics Journal* 25, no. 4: 1219-1231.

Tennis, Joseph T. 2012. "The Weight of Language and Action: Epistemic Stance, Methodological Action, and Theoretical Perspective in Knowledge Organization". *Études de communication*, no. 39: 15-40.

Toro, S., A.V. Anagnostopoulos, S.M. Bello *et al.* 2024. "Dynamic Retrieval Augmented Generation of Ontologies using Artificial Intelligence (DRAGON-AI)". *Journal of Biomedical Semantics* 15, no. 19.