Practical use of medical terminology in curriculum mapping

Martin Komenda [a], Daniel Schwarz [a,*], Jan Švancara [a], Christos Vaitsis [b], Nabil Zary [b],
Ladislav Dušek [a]

[a] Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Czech Republic
[b] Department of Learning, Informatics Management and Ethics, Karolinska Institutet, Sweden

ARTICLE INFO

ABSTRACT

Background: Various information systems for medical curriculum mapping and harmonization have been developed and successfully applied to date. However, the methods for exploiting the datasets captured inside the systems are rather lacking.
Method: We reviewed the existing medical terminologies, nomenclatures, coding and classification systems in order to select the most suitable one and apply it in delivering visual analytic tools and reports for the benefit of medical curriculum designers and innovators.
Results: A formal description of a particular curriculum of general medicine is based on 1347 learning units covering 7075 learning outcomes. Two data-analytical reports have been developed and discussed, showing how the curriculum is consistent with the MeSH thesaurus and how the MeSH thesaurus can be used to demonstrate interconnectivity of the curriculum through association analysis.
Conclusion: Although the MeSH thesaurus is designed mainly to index medical literature and support searching through bibliographic databases, we have proved its use in medical curriculum mapping as being beneficial for curriculum designers and innovators. The presented approach can be followed wherever needed to identify all the mandatory components used for transparent and comprehensive overview of medical curriculum data.

1. Introduction

Medical coding and classification systems are used in a variety of applications in medicine, public health, education and medical informatics, including their use in statistical analyses, data mining, and expert and decision support systems or knowledge engineering. Creation and maintenance of the coding systems, nomenclatures and standardized vocabularies involves the processes of transforming descriptions of diagnoses and procedures into a universal medical standardized scheme [1]. When talking about standardization, it is necessary to mention also ontologies, which formalize the meaning of terms used in practice, expected to play a major role in the automated integration of medical data with relevant information to support basic discovery and clinical research, drug formulation, and drug evaluation through clinical trials [2]. If we want information systems in health care to process data automatically, to sort them and produce correct statistics and if the principal data should be understood globally, it is necessary to use international coding systems and nomenclatures in the process of data acquisition [3]. From the perspective of medical informatics, the topic covering the broad use and global acceptance of medical coding standards still involves two main unresolved challenges: (i) the number of coding standards developed for medicine and health care is high, whereas their widespread adoption has been rather slow; (ii) the developed standards and vocabularies vary in their coverage, often being contradictory and competing. The basic principles of medical informatics and discussions on medical coding can be found in books and guides, such as [4] or [5]. One working group inside the European Federation for Medical Informatics (EFMI) is focusing on codes, classifications, terminologies and nomenclatures in their overview article [6]. They provide definitions and history of the International Classification of Primary Care (ICPC), and of the Read code and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). The review [7] analyses the use of SNOMED CT over time from 1966 until June 2006. A more recent review [8] then reports on the increasing use of SNOMED CT in literature, showing that the adoption of the nomenclature needs additional efforts toward harmonization with other standardized terminologies.

Medical education is an active area of education research, the importance and value of which has been demonstrated by Harden [9,10] and later by Wartman [11] already in the late eighties and the early nineties of the previous century. Medical curriculum development, mapping and harmonization are the natural components of

* Corresponding author. Tel.: +420 549492854.
E-mail address: schwarz@iba.muni.cz (D. Schwarz).

medical education, regardless of whether the underlying model is systems-based [12], outcome-based [13] or different (competency-based, socially accountable, distributed, etc.). Increasing variety in the form and resulting proficiency of different curriculum models has recently prompted the medical education community to introduce curriculum mapping guides [14] and interoperability standards [15]. The MedBiquitous Curriculum Inventory Standard [16], approved by the American National Standards Institute (ANSI), provides the technical syntax through which a wide range of different curricula can be expressed and subsequently analysed.

While there are a plethora of references, surveys and books on medical coding and classification for healthcare informatics applications, much less can be found on standardized vocabularies employed in medical education. The existing cataloguing schemas and thesauri relating, at least partly, to medical education are reviewed in [17] by the Medical Education Taxonomy Research Organization (METRO). Following the results of their research among stakeholders, METRO uses two existing thesauri: (i) the Medical Subject Headings (MeSH) thesaurus, being generic for medicine, and the British Education Thesaurus (BET), being generic for education. The thesaurus of descriptors for topics in medical education resulting from METRO's efforts is not available today, but their set of descriptors for assessment in medical education [18] using as much as possible from the MeSH and BET thesauri is still a living project with results available on-line. The need for identifying a suitable ontology or taxonomy for annotating the content of medical educational resources is emphasized in the systematic review [19], and especially from the perspective of semantic interoperability of information systems used in medical education and the potential of the semantic web (Web 3.0) to support medical curriculum mapping. The review provides a tabular overview of all existing standardized vocabularies, including their features, such as language, scope and the underlying structure, dividing them according to their context into biomedical, educational and administrative. Education resources shared through social media platforms such as YouTube have undergone exponential growth in recent years. Annotating education videos from YouTube and enriching them with other resources through Linked Data approach [20,21] is the main focus of research presented in [22]. SNOMED CT nomenclature is used to bridge YouTube tagging data and the metadata of enriching education resources. The resulting web-based architecture includes also the application interface (API) of BioPortal providing services from the National Center for Biomedical Ontology [23].

In this research, we focus on exploiting the well-known polyhierarchical MeSH thesaurus while helping the medical curriculum designers and innovators to navigate through multidimensional and extensive datasets embedded in the curriculum mapping systems. We are building on our previous experience with various nomenclatures and standards used in the development of the education web portal platform [24] and its extensions [25] adopted by all medical faculties in the Czech Republic and Slovakia under the umbrella of the MEFANET network [26]. In terms of methods and tools for curriculum mapping and reengineering, we are using our original web-based OPTIMED platform [27] and the recently adopted visual analytics tools [28–30] with the aim to develop new insights into curricular datasets and to demonstrate that the use of standardized thesaurus can facilitate curriculum reengineering processes as well as inter-institutional curriculum comparisons. In short, this paper shows an innovative way how to easily identify all essential components of a medical curriculum by adopting the MeSH biomedical nomenclature.

The remainder of this paper is as follows. In the methods section, detailed information is provided regarding: (i) existing and widely-accepted medical terminologies; (ii) our implementation of a web-based system for medical curriculum management; (iii) background and statistical methodology used to develop analytical reports over curricular datasets. In the following section, two data-analytical reports are shown, depicting the results obtained from frequency and association analyses of the particular medical curriculum while employing the MeSH thesaurus. The importance of these achievements together with their current limitations and future prospects are discussed and concluded in the final sections.

2. Materials and methods

The main contribution of this paper lies in the use of a standardized medical thesaurus in the analyses of a dataset extracted from a curriculum mapping system. We believe that our approach can be generalized, although we present results obtained from data stored in a particular curriculum mapping system, OPTIMED, and a specific thesaurus, MeSH. The selection of this specific thesaurus is the result of our scoping review carried out according to the principles laid out in the framework for scoping studies by Arksey and O'Malley [31] and recently extended and clarified elsewhere [32,33].

2.1. Overview of the existing medical terminologies

In this subsection, we refer to the summary overview relating to the up-to-date use of various specialized terminologies in information sciences and education. Our close attention is primarily focused on medically oriented fields where these taxonomies promote an efficient way of organizing and understanding data. It is clear that specialized vocabularies matter, not only in the traditional library and information sciences world and but also for many different digital information stakeholders [21]. For instance, the use of the vocabulary of a particular domain is an important initial step in creating formalized knowledge representations as an essential part of the education process. These vocabularies follow the ratchet principle: it moves from basic understanding to thorough understanding, from simple to complex education [34]. When a virtual learning environment (VLE) turns to the task of consolidated education data collection (for instance content management, curriculum mapping and planning, student engagement and administration, communication and collaboration domain), these vocabularies will present considerable challenges to standardizing medical education. A prerequisite to more comprehensive categorization of the education content is the implementation of standardized terminology directly in the VLE systems. One of the primary aims is to overcome two significant barriers to effective retrieval of machine-readable and processable information: (i) the variety of names used to express the same concept, (ii) the absence of a standard format for distributing terminologies [35]. The purpose of medical vocabularies is to embody what has been known in the past about every phase of medicine [36]. These vocabularies continue to increase and grow, not only in its technological aspects, but also from the perspective of medical education quality, which is logically reflected at the global level of health care. Below, the most widespread, suitable and commonly used standardized methodologies are introduced.

2.1.1. UMLS

Unified Medical Language System (UMLS) brings together many health and biomedical vocabularies, ontologies and standards to enable interoperability between computer systems. It was developed by the National Library of Medicine and covers the entire terminology domain by integrating more than 60 families of