Proceedings of the 13th onference on Language Resources and Evaluation (LRE 2022), pages 2850–2855
Marseille, 20-25 June 2022
© European Language Resources ssociation (ELR), licensed under CC-BY-NC-4.0

Whole-Person Function Dictionary for the Mobility, Self-Care and Domestic Life Domains: a Seedset Expansion pproach

yah Zirikly [2] Bart Desmet Julia Porcino
Jonathan Camacho Maldonado * Pei-Shu Ho * Rafael Jimenez Silva * Maryanne Sacco *
[1] Rehabilitation Medicine Dept., Clinical Center, National Institutes of Health
Bethesda, MD
[2] Department of Computer Science, Johns Hopkins University
Baltimore, MD
{ayah.zirikly, bart.desmet, julia.porcino,
jonathan.camacho, pei-shu.ho, rafael.jimenezsilva, maryanne.sacco}@nih.gov

bstract

Whole-person functional limitations in the areas of mobility, self-care and domestic life affect a majority of individuals with disabilities. Detecting, recording and monitoring such limitations would benefit those individuals, as well as research on whole-person function and general public health. Dictionaries of linguistic expressions related to whole-person function would enable automated identification and extraction of relevant information. However, no such resources currently exist, due in part to a lack of standardized coding and their availability mainly in free text clinical notes. In this paper, we introduce dictionaries of whole-person function in the domains of mobility, self-care and domestic life, built and evaluated using a small set of manually annotated clinical notes, which provided a seedset that was expanded using a mix of lexical and deep learning approaches.

Keywords: disability, terminology, terminology expansion, whole-person function, mobility, self-care

1. Introduction

Functional limitations affect a large proportion of the world's population, and it is estimated that over 1 billion people (WHO, 2011) live with some form of disability. In Europe, about 135 million people live with disability [1], and the community of people with disability in the United States is 61 million large (Centers for Disease Control and Prevention, 2020). Social insurance and security organizations, such as the Social Security dministration (SS) in the United States, aim to support this population, for example by providing financial aid in the form of disability benefits. Given the scale of the problem, such organizations need to allocate considerable and rising resources, both financial and in terms of personnel, to review disability applications and provide an eligibility decision. These responsibilities are even more challenging when there are no diagnosis codes associated with an applicant's medical evidence. s an example, a claim for a person who suffers from impairments due to cancer would usually be coded using the International Classification of Diseases (ICD), which eases the eligibility determination process. However, disability benefits can also be granted on the basis of whole-person function limitations which cannot be attributed to a single medical diagnosis. Whole-person functioning is assessed at the activity and participation level and represents a transactive process of interactions between a person, their environment, and the nature and demands of the activities that the person wants to do, needs to do, or is expected to do to fulfill one's roles. Deciding cases at the whole-person level is more challenging, due to the lack of use of standardized codes for whole-person function within the medical community, and the fact that their descriptions are heavily semantically dependent. For instance, a note by social worker may include the following: *the patient cannot bathe himself alone without the help of a family member, if available*. Such mentions imply that this person cannot perform the bathing task independently, and that help is not always available. It is important to note that the World Health Organization (WHO) developed the International Classification of Functioning, Disability, and Health (ICF) to address function-related information coding, but unfortunately its adoption in the clinical community is optional and not widespread, especially in the United States. On the other hand, some European countries, such as the United Kingdom, have seen more initiatives to adopt ICF coding (Gimeno and Lin, 2017).

Finding and coding whole-person function information in clinical text is a nascent field in the Natural Language Processing (NLP) community. Technological systems that could identify relevant information could improve capture by the clinician, and unlock it for research and disability determination purposes. This paper joins previous work such as (Desmet et al., 2020) that calls for more work in this area to build NLP so-

* Equal contribution

[1] https://www.euro.who.int/en/health-topics/Life-stages/disability-and-rehabilitation/data-and-statistics/facts-on-disability

lutions that can perform information extraction using terminologies or machine learning models trained on annotated data. However, very few systems have been developed to address function and disability information extraction (Thieu et al., 2021; Newman-Griffis et al., 2019b; Newman-Griffis and Zirikly, 2018). In this work, we focus on building dictionaries of linguistic expressions such as terms, collocations and short phrases, which we will collectively refer to as terms, for the following whole-person function types: mobility, self-care, and domestic life, where these types account for 13.7% (the top disability limitation), 3.6%, and 6.3% of total disabilities [2]. We start with a seedset extracted from datasets that were manually annotated by domain experts, consisting of clinical notes from two sources: the National Institutes of Health (NIH) and SS . This seedset contains mentions for each of the domains of interest, in addition to their associated ICF codes that are also captured during annotation. We expand the seedset by clustering the expanded terms using a combination of lexical and deep learning techniques, where we provide both the mention and the suggested ICF code. The final set of terminologies is evaluated against manually annotated datasets for recall and is also filtered by our domain experts to contain only relevant terms with the correct ICF code. This paper also makes these terminologies available to the research community. We summarize our contributions as follows:

- To our knowledge, these are the first publicly available dictionaries of terms related to mobility, self-care, and domestic life, to help in extracting information relevant to disability.
- Our terminologies provide the associated ICF codes, where both term and ICF code have been verified by domain experts.
- We push for sharing and developing a common understanding and background about the studied domain in the medical-related communities.

2. Related Work

Whole-person function research in the NLP and bioinformatics communities is a recent field that has been attracting momentum in recent years. For instance, Newman-Griffis et al. (2019a) emphasize the importance and the significant impact of the problem on public health in their call for action for more research in the field. Many reasons exist for the slower advances in this field compared to the general clinical one, including the lack of standardized terms, data access, and the challenges associated with the language to describe such terms (Desmet et al., 2020).

There has been some work that introduced datasets annotated for whole-person function, along with fine-grained descriptions of the annotation scheme, given the recency of the problem and the need for many domain experts to collaborate on establishing a scheme (Thieu et al., 2021). Other studies focused on the information extraction aspect of the problem and extracted whole-person function phrases and mentions in a similar way to other Named Entity Recognition (NER) tasks (garonnik et al., 2020; Newman-Griffis and Zirikly, 2018; Newman-Griffis et al., 2019b). Terminologies in the clinical NLP field have proven to be effective in many extraction tasks such as drug names (Lerner et al., 2020), disease (Chun et al., 2006), and symptoms (Silverman et al., 2021). Creating and expanding terminologies is an active area of NLP research as well. Fan et al. (2019), for example, discuss terminology augmentation of dietary supplements using word embeddings. In recent work by Newman-Griffis et al. (2021) in the area of ICF coding, the authors aim at linking mobility terms to their corresponding ICF codes using classification methods. However, they emphasize the importance of having domain-specific terminologies, since they could alleviate some of the challenges they faced and reduce the need for expensive manual annotation of documents by domain experts. Thus, building publicly available terminologies that can be used by the NLP community and other entities that handle data related to function, such as SS 's disability determination program, is timely and necessary.

3. pproach

Our work to build terminologies for whole-person function-related domains aims to foster more interest and work in this field by the NLP community, and is also motivated by the lack of adoption of standardized ICF codes in clinical notes. We focus on two domains: i) mobility (MOB); and ii) self-care and domestic life (SCDL). We start with an example of how this information is reported in clinical notes.

> Patient is **able to walk with the help of cane or another assistive device**. He reports that he **can take a bath and clean himself**.

The mention "able to walk with the help of cane or another assitive device" represents the patient's mobility state, where *walk* is the action of the mobility mention. Similarly, the mention "can take a bath and clean himself" represents information related to SCDL, where *take a bath* and *clean himself* are the action terms for SCDL. In our work we focus on the action terms as opposed to the full mentions, since the former are the key candidates for building terminologies.

[2] https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html