Extending a Medical Language Processing System to the Functional Status Domain

Michael Bales, MPH[a], Rita Kukafka, DrPH, MA[a,b], Ann Burkhardt[c], OTD, OTR/L, BCN, FAOTA, Carol Friedman, PhD[a]

[a]Department of Biomedical Informatics, Columbia University in New York, NY, USA
[b]Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University in New York, NY, USA
[c]Department of Rehab Medicine, New York Presbyterian Hospital/Columbia, New York, NY, USA

Abstract

The World Health Organization's International Classification of Functioning, Disability, and Health (ICF) provides a common framework for describing functional status information (FSI) in health records[1]. Given the expense of manual coding, we are investigating the use of natural language processing (NLP) for automated FSI coding. We used an existing NLP system that was originally designed to encode clinical information. The system's lexicon and coding table were modified and preprocessing and postprocessing programs were created, allowing for automated assignment of selected ICF codes.

Introduction

Diagnostic information in medical records often lacks detail about the effect of a patient's health status on the individual's daily activities. FSI, which addresses emotional, environmental, and other factors that influence health status, is useful in assessing the use of resources and the need for services. ICF codes differ from traditional codes for clinical information. For example, ICF code $d5400.10$[2] refers to *putting on clothes, minimal assistance*. The last two digits are numeric qualifiers ranging from 0–4, which refer to the level of facilitation or impairment conferred by a condition.

Little is known about how FSI is being expressed in medical records, but it is known that manual coding is costly. In medical language processing, a natural language text file is input into a software program that outputs data in coded format. These data can then be used for a number of purposes, such as for assessing caregiver burden. The Medical Language Extraction and Encoding system[3] (MedLEE), has been in routine use at the Columbia University Medical Center (New York, NY) since 1995, for encoding information in clinical reports. Its components include a lexicon of terms, a parser, and a coding table that maps natural language terms to controlled terms.

This work is one aspect of a two-year pilot research project in which a subset of ICF codes was used in an analysis of human and automated ICF coding.

Methods

A domain expert selected five ICF main codes pertaining to a person's ability to function in society. Appropriate functional status language was identified in patient records, and specific terms were assigned more general target forms and added to MedLEE's lexicon. For example, the phrases "clothing item" and "tying shoes" were assigned the target form "putting on clothes". The phrases "not always independent" and "a little trouble" were assigned the target form "minimal assistance". Entries were also added to the system's coding table. The two target forms above were entered alongside the ICF code $d5400.10$.

Certain terms, such as "transfers" and "bladder", were already in the MedLEE lexicon, but had different senses in the FSI domain. A preprocessing step was performed on all discharge summaries to mark these terms so that they would be recognized as FSI-related terms. Postprocessing was performed to identify all ICF main codes and related qualifiers, and to perform final code assignments.

Discussion

A total of 549 entries were added to the lexicon, and 181 entries were added to the coding table. We are currently evaluating the performance of the system in assigning ICF codes.

Developing medical language processing software for the FSI domain is challenging. Information describing functional status is very descriptive. As described above, rehabilitation experts use a variety of similar phrases to convey the same idea, and these variations must be added to the lexicon. Additionally, the ICF coding process can involve complex medical inferencing, which is difficult to formalize. For example, two patients who have had a stroke might have different levels of functional status, depending on their age and the presence of comorbidities such as diabetes and hypertension.

References

1. Ustun TB, Chatterji S, Kostansjek N, Bickenbach J. WHO's ICF and functional status information in health records. Health Care Financ R 2003;24(3):77-88.
2. World Health Organization. ICF classification hypertext version. Available from: URL: http://www3.who.int/icf/onlinebrowser/icf.cfm
3. Friedman C, Alderson PO, Austin J, Cimino JJ, Johnson SB. A general natural language text processor for clinical radiology. J Med Inform Assoc. 1994;1(2):161-174.