

README:: Sense Inventory of Clinical Abbreviations and Acronyms
Version 1 (October 30, 2012)

Contact information:

Natural Language Processing/Information Extraction (NLP/IE) Program
Email: nlp-ie@umn.edu

This clinical sense inventory is provided in two versions. It contains 440 common abbreviations and acronyms from a corpus of 352,267 dictated clinical notes (2004-7) containing discharge summaries, consultation notes, and admission notes. In addition, senses for each of the abbreviations are compared with senses in the Unified Medical Language System (UMLS) (Version 2012AB), Another Database of Abbreviations in Medline (ADAM) (1), and Stedman's Dictionary, Medical Abbreviations, Acronyms & Symbols, 4th edition.

For each of the 440 clinical abbreviations and acronyms (short forms), the corresponding 949 senses (long forms) and their prevalence are included from a random sample of 500 random manually annotated samples.

After eliminating names, general English, unsure senses, error and misused senses, 752 senses were compared and mapped to 17,359 UMLS, 5,233 ADAM, and 4,879 Stedman's long forms.

In addition to direct mappings using UMLS source files (MRCONSO and LRABR), semantic mappings with MetaMap (Version 2011) was performed and Concept Unique Identifiers (CUIs) extracted.

Clinical sense inventory I is a '|' delimited file (MasterFile) containing exact mappings of lexical forms for each of the long forms from given a given resources.

Clinical sense inventory II is a '|' delimited file (RefinedMasterFile) containing mappings of forms for each resource after merging forms using Lexical Variant Generation (LVG) (2) normalization and then performing semantic mappings.

Column descriptions of Clinical Sense Inventory I and II:

Column 1: Short form (SF, abbreviation or acronym)
Column 2: Long form (LF, sense/meaning)
Column 3: CUIs produced by running MetaMap with term processing option (-z) and Long Form (Column 2)
Column 4: Existence of long form in the Clinical Sense Inventory ('1' indicates existence)
Column 5: Ratio in the Clinical Sense Inventory
Column 6: UMLS CUI(s)
Column 7: Source information in the UMLS (UMLS SOURCE)
Column 8: Existence of long form in ADAM ('1' indicates existence)
Column 9: Ratio in ADAM
Column 10: Existence of long form in Stedman's Long Form ('1' indicates existence)

Reference:

1. Zhou W, Torvik VI, Smalheiser NR. ADAM: another database of abbreviations in MEDLINE. *Bioinformatics*. 2006;22(22):2813-2818.
2. McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A, Srinivasan S. UMLS knowledge for biomedical language processing. *Bull Med Libr Assoc*. Apr 1993;81(2):184-194.

Addendum (January 31, 2013)

a. We removed 'comprehensive' from title of README

- b. We corrected the number of entire notes
- c. We removed 'operative notes' and '2008' from README