Matching Clinical Observations in Published Cardiology Case Reports to LOINC

Charles Sneiderman MD PhD,[1] Graciela Rosemblat PhD,[1] Bruce E. Bray MD,[2] Marcelo Fiszman MD PhD,[1] Thomas C. Rindflesch PhD[1]
[1]National Library of Medicine, Bethesda, MD
[2]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT

Observations in published clinical case reports were matched by hand to entries in the Logical Observations Identifiers Names and Codes (LOINC) in a preliminary analysis of the feasibility of automated natural language processing. Coverage of history, physical, laboratory, and imaging findings was nearly complete, and categorization of terms assigned suggests a strategy for automatic coding.

INTRODUCTION

A significant amount of textual information is stored in electronic medical records and is inaccessible to computer applications. The Logical Observations Identifiers Names and Codes (LOINC) [1] is a terminology that provides a framework to represent clinical observations. We describe our preliminary work in mapping text in cardiology case reports to the LOINC terminology. The case reports function as surrogates for medical reports in an electronic medical record.

METHODS

MEDLINE was searched with the Medical Subject Heading term "Heart Diseases" and the 10 most recent full-text articles were selected from PubMed Central for annotation. Paragraphs describing treatment or disease processes were removed. LOINC (Version 2.25) terms and codes were assigned by a family physician (CS) and reviewed by a cardiologist (BB). Text expressions were matched to LOINC attributes (e.g. Physical findings) and not values assigned to them (e.g. lethargy). After all the codes were assigned, we reviewed the level of complexity that would be required to map from the text to the LOINC terms. It was determined that there were three levels of complexity: (1) An exact match, in which there was a straightforward lexical match between the text and LOINC component name or short common name, (2) A computable match, in which terms could be matched to text using rules, and (3) A domain knowledge (or context) match, which required ontological knowledge to be completed. We then classified all the terms into these three categories.

RESULTS

182 sentences were examined and 230 findings identified. Only two findings had no LOINC match: *12 hour troponin T* and *NYHA Class 2*. 132 LOINC terms were used: 33 laboratory and 99 clinical observations. The distribution of the three categories was: 50 exact match, 160 computable matches, and 20 requiring domain knowledge. *Alanine amino transferase*, for example, matched to a LOINC component, while heart rate mapped to a short common name. In a computable match, a rule is needed to map the text expression to the LOINC terms. Thus, a rule based on UMLS semantic types could link *Kussmaul sign* to the LOINC attribute "Physical findings." Finally, Metathesaurus relations could be exploited to determine that *ventricular contractility* is to be coded with LOINC component "Ejection fraction."

CONCLUSION

Coverage of history, physical exam, laboratory and imaging findings was essentially complete [cf. 2], and generalizations noted during annotation suggest the use of domain-based rules and hierarchies to augment natural language processing techniques based on string matching for automatically identifying concepts in clinical text. Future work includes first implementing the generalizations observed during the hand annotation and then extending the process to automatically coding values as well as attributes.

Reference

1. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem. 2003 Apr;49(4):624-33.
2. Bakken S, Cimino JJ, Haskell R, Kukafka R, Matsumoto C, Chan GK, Huff SM. Evaluation of the clinical LOINC (Logical Observation Identifiers, Names, and Codes) semantic structure as a terminology model for standardized assessment measures. J Am Med Inform Assoc. 2000 Nov-Dec;7(6):529-38.