PTXT Finder—an Application for Finding Appropriate EHR Data Elements for Data Analysis Using Cross Referencing Information Sources

Jau–Huei Lin, MD1,2, Peter J. Haug, MD1,2

1Department of Biomedical Informatics, University of Utah, Salt Lake City, UT
2Intermountain Healthcare, Salt Lake City, UT

Abstract

PTXT Finder was developed to reduce the manual efforts necessary to map from clinical variables in a decision support system to data elements in an EHR. The descriptions in a data dictionary may be inadequate for pinpointing data elements that represent a clinical variable. Semantics implied in taxonomy and real usage of the element are two important supporting information sources. PTXT Finder provides description, hierarchy, and statistics displays so users can cross-reference among these views.

Introduction

Even in systems with an established data dictionary, it is not a trivial task to identify appropriate EHR data elements that represent a specific clinical variable. Ambiguity may exist in the descriptions of the data elements. To ascertain the selected element, one may need to look up other information sources. This poster describes the design of an application that can help researchers, intent on building a decision support system, in this process.

Background and Motivation

PTXT Finder is a GUI application that can help researchers reach appropriate data elements in an EHR’s data dictionary, the “PTXT code” dictionary1, used in HELP hospital information system at Intermountain Healthcare, Salt Lake City, UT. The PTXT code is a hierarchical coding scheme. Each code comprises eight numeric sub–codes that indicate a taxonomic structure. To find an appropriate data element, one must refer to the dictionary’s definitions of codes and then explore real data to see the usage for confirmation. However, most researchers in our institute do not master the skills to access the required information.

In addition, the dictionary description does not completely represent the semantics of a data element. Part of the meaning is implied in the taxonomy. For example, the code ’13.1.76.1.1.0.0.0’ is for the fasting sample value of Glucose Tolerance Test. However, the description of this specific code is simply “fasting”. To acquire the complete meaning, one must also refer to the description of its ancestor, ‘13.1.76.0.0.0.0.0’—“Glucose Tolerance Test”. Therefore, the location in the taxonomy is important for the exact meaning.

Design

The PTXT Finder application provides three views of data elements—description, hierarchy, and data statistics displays—and two types of constraints—constraints on descriptions and on data statistics2 (Figure 1). Users can restrict the scope of data elements using constraints, show them in three displays, and make a final decision on the element.

Conclusion

For the PTXT code, this application can help users identify appropriate data elements. This is useful for those who do not possess knowledge of the coding scheme and skill in data exploration. With proper modifications, a similar design can be used in other coding schemes that require views of a taxonomy and data statistics to identify needed data elements.

References