

Efficient Population of Structured Data Forms for Medical Records

Using Syntactic Constraints and Intermediate Text

Ronald P. Loui

Department of Computer Science
University of Illinois
Springfield, IL
r.p.loui@gmail.com

Ashley Hollinshead

University of Illinois
Springfield, IL

Abstract—This poster presents an innovative user-support method for structured data entry that resulted from working with a large patient registry reporting group in a leading hospital, tasked with efficient electronic medical record discretization. It is a method of using syntactic constraints on various fields during web-forms coding, to auto-select the fields where values belong; this allows coders to spend more time looking at the clinical narrative documentation and less time attending to the mouse, the browser, and field-selection on the data entry pages. The method is to meta-program a “floating text box” for intermediate, semi-structured notes, with behavior specific to the choices on that page. It bridges the gap between abbreviations, which are easy, and datamining natural language text, which is hard. The ideas can be generalized to auto-filling forms and semi-autonomous forms filling beyond the medical domain.

Keywords—forms, knowledge extraction, syntactic constraint, semantic constraint, automated form filling, field-selection, menu-item selection, electronic medical record, structured data, discrete data, menu shortcut, data entry.

I. INTRODUCTION

Structured and discrete data are important to medical records. The process of transferring data from clinical, written, or hard-copy documents, often from source text written in unconstrained language into discrete and structured form can be done more efficiently by adding an intermediate text.

Observing the practices of the coders in the Cleveland Clinic's SemanticDB project [Pierce12] resulted in the prototyping of a “floating textbox” for unstructured but disciplined typing of notes and abbreviations, from which structured data input elements (checkbox options, radio options, menu selection options, as well as integer, real, date, and time fields) could automatically be populated. The auto-fill is supervised by the human doing data entry, so this is not as autonomous as structured data extraction from natural language text. It is a note-taking with the anticipation of being automatically interpreted. This notation has some of the advantages of natural language narrative, but is close enough to the structured form of the data that extraction efficacy is

extremely high and error is extremely low. The interface is altered, with the addition of a small amount of scripting on each page, to give support for the semi-formal textual input that serves as an intermediary between hard-copy medical notes and electronic medical record.

II. INNOVATION

Our innovation was to relegate the mouse to secondary selection (available as needed, but not mandated), and provide syntax-based automated assistance for selecting and auto-populating input fields. Page-specific, auto-generated client-side scripting (in javascript) permits the coder simply to type a value in the floating text box, and it populates the appropriate field based on syntactic well-formedness constraints.

If the user typed “Male” or even “M” while facing a screen where there is only one field where that could properly be put, no mouse operation would be needed. Similarly, “NY”, “63130”, “808-555-555”, “Dr. Saltpresz”, and even “No AFIB” would be unambiguous entries. Mouse time, mouse error, and keyboard interruption are exchanged for more typing.

Any string matching the regular expression $/[0-9]\{3\}-[0-9]\{2\}-[0-9]\{4\}/$ would populate the SSN field. One can vary and limit the aggressiveness with which imperfect matches are made. Auto-generated shorthands and alphanumeric id's are a trivial form of syntactic constraint. Semantic and statistical constraints can also be added, which provide real-time error checking.

ACKNOWLEDGMENT

Brian Beck helped with implementation, Tanya Ashinhurst with testing. The Clinical Investigations group in the Heart and Vascular Institute at the Cleveland Clinic, and Doug Lenat at Cycorp, gave encouragement.

REFERENCES

- [1] Pierce, Christopher D., David Booth, Chimezie Ogbuji, Chris Deaton, Eugene Blackstone and Doug Lenat, "SemanticDB: A semantic web infrastructure for clinical research and quality reporting," *Current Bioinformatics* 7, 2012.