An Architecture for Automatic Generation of Computer Interpretable Guidelines

Daniel R. Schlegel, PhD
Department of Computer Science, SUNY Oswego, Oswego NY, USA

Introduction

Compliance with clinical practice guidelines (CPGs) improves patient care, but is lacking. Compliance improves greatly with the introduction of clinical decision support systems (CDSS) which implement guideline recommendations and are integrated into EHR systems. Unfortunately CPGs, as commonly distributed, are not easily consumable by computers for integration with CDSS.

Systems which assist in the creation of CIGs use top-down information extraction techniques, examining linguistic patterns[1,2] or co-occurances[3] to extract text which may be placed in CIG templates. Each operates best (or only) on simple recommendations, whether they be semi-structured[2] or only containing single disorders and prescriptions/procedures[3]. We propose a system architecture (see Figure 1) based on the application of bottom-up natural language understanding (NLU) techniques which would be capable of handling more sophisticated guideline recommendations, grounding them in the most specific terms available so that they may easily integrate with CDSS, and producing a CIG.

Natural Language Understanding with Clinical Tractor

Natural language understanding techniques have been used in other domains, such as counter-insurgency, with some success. The Tractor system converted short intelligence messages to a knowledge base (KB) containing over 92% semantic relations using a rule-based system wherein the rules fired correctly nealy 98% of the time[4]. In a paper currently in preparation it was found that the transformation is on par with that of humans. We are porting Tractor to the clinical domain – a new system dubbed *Clinical Tractor*.

The proposed architecture (see Figure 2) consists of four main components. Text processing using various "processing resources" (PRs) operating within the open-source General Architecture for Text Engineering (GATE) produces a set of annotations. These annotations are converted to a KB consisting of propositions in the form of the CSNePS knowledge representation and reasoning system[5]. These are mostly syntactic relations. The terms in the KB are aligned with background knowledge in resources such as WordNet, and relevant knowledge is imported (an important step for identifying, *e.g.*, which terms used in the text refer to Actions). The syntactic relations in the KB are simplified using rules which map, *e.g.*, passive voice to active voice, then the syntactic relations are mapped to semantic relations (see[4] for worked examples in the Tractor system). These rules, executed in CSNePS, are designed to be general, operating on entire classes of nouns, verbs, and phrases through the use of ontology.

Concretization and CIG Generation

Concretization maps terms to their definitions/elaborations, with the goal of reaching a level of specificity which eases interop with existing EHR data. Definitions/elaborations appear in guidelines in (at least) four ways: (1) appositional phrases, often in parentheses after a term; (2) explicit definitions, *e.g.,* "*X is defined as Y*"; (3) discourse elaboration in which a later sentence elaborates on the meaning of terms in an earlier one; (4) references to tables and figures. From the semantic KB, this module attempts to concretize terms by associating them with definitions/elaborations within

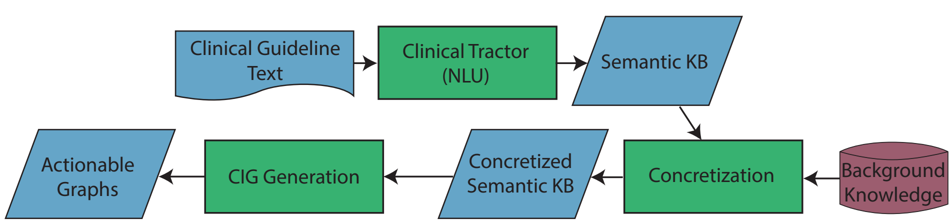

Figure 1: System architecture of the proposed CPG understanding and formalization system.

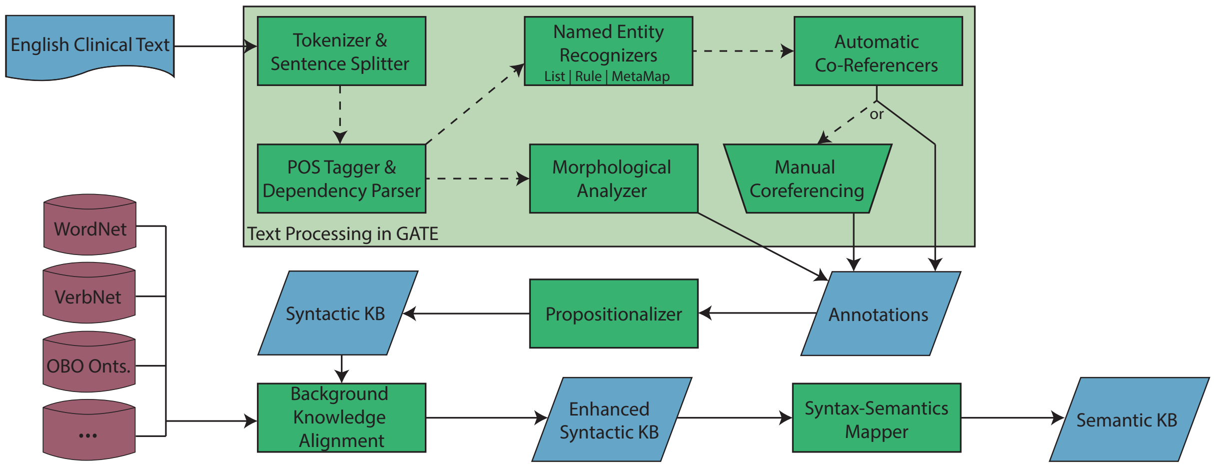

Figure 2: Clinical Tractor system architecture.

the guideline itself, then looking to background knowledge sources where that fails.

The target CIG formalism being considered is Actionable Graphs (AG), as it is used in a multiple-CIG mitigation strategy[6]. An AG has a context node at the root with action and decision nodes below, connected by arcs indicating transitions. Action and decision nodes indicate actions to be taken and decisions which need to be made, respectively. Generating an AG requires understanding conditions and actions, both explicit and implicit, in recommendations.

Preconditions often require background knowledge; consider the ACC recommendation "Every tobacco user should be advised at every visit to quit." The precondition is that *the patient* is a tobacco user, something unstated in the recommendation itself. Clinical Tractor contains a list of roles which may be filled by a person, *e.g.,* "user." A mapping rule recognizes that "tobacco" modifies "user", and therefore "tobacco user" is the role. Since roles are filled by persons, the system produces a KB representing that every person who has the role of being a tobacco user should be advised at every visit to quit. Moreover, we know that the person being discussed is in fact the patient.

Often there are several potential clinical actions suggested by a recommendation, requiring the generation of decision nodes. Representing choices of this type is a currently unrealized goal for Clinical Tractor. These choices will then be directly translatable into decision nodes. Temporal ordering which uses prepositions such as "after" are handled by Clinical Tractor. The written order of the recommendations imply their temporal order within a group with the same set of preconditions. These kinds of temporal ordering will be handled using simple rules over the semantic KB.

References

1. R Serban, A ten Teije, et al. Exploiting thesauri knowledge in medical guideline formalization. *Methods of Information in Medicine*, 48(5):468–474, 2009.

2. Katharina Kaiser and Silvia Miksch. Versioning computer-interpretable guidelines: semi-automatic modeling of 'living guidelines' using an information extraction method. *Artificial intelligence in medicine*, 46(1):55–66, 2009.

3. Wessam Gad El-Rab, Osmar R Zaïane, and Mohammad El-Hajj. Formalizing clinical practice guideline for clinical decision support systems. *Health informatics journal*, 23(2):146–156, 2017.

4. Stuart C. Shapiro and Daniel R. Schlegel. Natural language understanding for soft information fusion. In *Proceedings of Fusion 2013*. IFIP, July 2013. 9 pages, unpaginated.

5. Daniel R Schlegel and Stuart C Shapiro. Inference graphs: Combining natural deduction and subsumption inference in a concurrent reasoner. In *Proceedings of AAAI-15*, 2015.

6. Szymon Wilk, Martin Michalowski, et al. Comprehensive mitigation framework for concurrent application of multiple clinical practice guidelines. *JBI*, 66:52–71, 2017.