

Ontology-Driven Semantic Search for Brazilian Portuguese Clinical Notes

Sadid A. Hasan^a, Xianshu Zhu^a, Joey Liu^a, Claudia M. Barra^b, Lucas Oliveira^b, Oladimeji Farri^a

^a Philips Research North America, New York, United States

^b Pontifical Catholic University of Paraná, Curitiba, Brazil

Abstract

The emerging penetration of Health IT in Latin America (especially in Brazil) has exacerbated the ever-increasing amount of Electronic Health Record (EHR) clinical free text documents. This imposes a workflow efficiency challenge on clinicians who need to synthesize such documents during the typically time-constrained patient care. We propose an ontology-driven semantic search framework that effectively supports clinicians' information synthesis at the point of care.

Keywords:

Natural Language Processing (NLP); Semantic Search; Ontology Classification; Brazilian Portuguese

Introduction

Traditional semantic search applications mainly focus on the general knowledge domain while understanding the English language. Furthermore, existing semantic search applications based on Brazilian Portuguese are not designed to support routine healthcare tasks and clinical workflow. We propose an ontology-driven semantic search framework to provide clinicians better support in their decision-making process.



Figure 1 - Screenshot of semantic search user interface

Methods

Our semantic search framework uses open source NLP toolkits to analyze the Brazilian Portuguese grammar. The resulting grammatical units are processed for clinical concept extraction and mapping to Concept Unique Identifiers (CUI) in the Unified Medical Language System (UMLS) Metathesaurus. We generate relevant hierarchical representations of clinical concepts in the Portuguese version of the International Classification of Diseases (CID-10) based on publicly available language translations, and mapped these concepts to their English equivalents within the UMLS, including the Portuguese clinical ontologies: MSHPOR (Medical Subject Headings in Portuguese) and MDRPOR

(Medical Dictionary for Regulatory activities in Portuguese). CID-10 concepts are then mapped to the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). Our ability to achieve Portuguese-English mapping of controlled clinical vocabularies represents a significant contribution to Brazilian Portuguese clinical NLP research. Our framework also offers patient data visualization via an intuitive search engine interface (Figure 1) that allows clinician users to efficiently retrieve relevant clinical information from large data sources.

Results

To evaluate the effectiveness of our semantic search framework, we conducted a pilot study using a subset of 80 cardiology discharge summaries written in Brazilian Portuguese. 4 clinical students (2 Medical and 2 of a Nursing background) were each asked to perform 2 clinical summarization tasks using a traditional EHR-like system (without semantic search functionality) for one task and our semantic search system for the second task. Our study results showed that on average, the students required 39.9% less time to complete task 1 and 10.7% less time to complete task 2 when they used our semantic search application. Further analysis of the results suggested that the differences in time to task completion may be due to each student's decision to read every discharge summary while using the traditional EHR-like system. In addition, qualitative data from the interviews conducted after task completion revealed that the students unanimously preferred having access to the semantic search application for task completion as they could search for keywords and browse through fewer summaries in contrast to manually synthesizing information. Limitations in our study include possible disparities in domain expertise, documentation speed, degree to which the tasks were understood by the students, and differences in task complexity. In our future research, we would perform a larger and more robust evaluation of our application while minimizing the aforementioned limitations.

Conclusion

Preliminary evaluation of our ontology-driven semantic search framework for Brazilian Portuguese clinical documents reveals that it effectively improves clinical task workflow and can contribute to the decision making process of the clinicians through faster, and more accurate information retrieval and visualization at the point of care.