Enrichment/Population of Customized CPR (Computer-Based Patient Record) Ontology from Free-Text Reports for CSI (Computer Semantic Interoperability)

David Mendes, Universidade de Évora, Évora, Portugal

Irene Pimenta Rodrigues, Universidade de Évora, Évora, Portugal

Carlos Rodriguez-Solano, Universidad de Alcalá, Alcalá de Henares, Spain

Carlos Fernandes Baeta, Unidade Local de Saúde do Norte Alentejano, Portalegre, Portugal

ABSTRACT

CSI (Computer Semantic Interoperability) is a very important issue in healthcare. Ways for heterogeneous computer systems to "understand" important facts from the clinical process for clinical decision support are now beginning to be addressed. The authors present here comprehensive contributions to achieve CSI. EHR (Electronic Health Record) systems provide a way to extract reports of the clinicians activity. In order to formalize an automated acquisition from semi-structured, free-form, natural language texts in Portuguese into a Clinical Practice Ontology an important step is to develop the ability of decoding all the nicknames, acronyms and short-hand forms that each clinician tend to write down in their reports. The authors present the steps to develop clinical vocabularies extracting directly from clinical reports in Portuguese available in the SAM (Sistema de Apoio ao Médico) system. The presented techniques are easily further developed for any other natural language or knowledge representation framework with due adaptations.

Keywords: Clinical Practice, Computer-Based Patient Record (CPR), Computer Semantic Interoperability (CSI), Electronic Health Record (HER), Ontology Enrichment

DOI: 10.4018/jitr.2014010101
INTRODUCTION

We will present in this paper two sections that illustrate our work:

1. The steps that we consider are involved in the complex acquisition procedure of clinical concepts expressed in English from text in Portuguese;
2. A proposal of a software architecture needed for an automated acquisition to articulate the steps presented before.

Our work final objective is to enrich/populate an ontology that shall allow us to devise AI (Artificial Intelligence) tools that reason about clinical practice. Given the reasons explained in (Mendes & Rodrigues, 2011) we chose CPR (Computer-based Patient Record) Ontology as target for population. CPR is a W3C (World Wide Web Consortium) standard for representing clinical practice knowledge. The major problem of personal jargon creation was not properly addressed, however, and so is fully discussed in the present article. We demonstrate the possibility of information extraction from free-text clinical episode reports in an automated manner.

When developing a methodology for automatic population of CPR Ontology we faced the particular problem of clinical concept recognition when dealing with Portuguese natural language text. After consulting with several MD (Medical Doctors) whose activity is the main subject of representing knowledge that way, we found that we can take into our advantage the fact that each one usually develops his/her own way of writing down their daily chores. What we have developed is a way of maintaining an acquired controlled vocabulary. Assuming that this is a task involving NLP (Natural Language Processing) specifically for Portuguese we tried to develop some techniques that can be easily applicable to different languages with the same set of constraints presented ahead.

EXPERIENCES IN THE FIELD

Protocol with ULSNA

The ULSNA, E.P.E. (http://www.ulsna.min-saude.pt/) has as its principal object the provision of primary and secondary health care, rehabilitation, palliative and integrated continued care to the population and the means necessary to exercise the powers of the health authority in the geographic area affected by it. ULSNA is a healthcare providing regional system that includes 2 hospitals (José Maria Grande in Portalegre and Santa Luzia in Elvas) and the primary care centers in all the district counties. Universidade de Évora signed an agreement with ULSNA that enabled the usage of de-identified (according to safe-harbor principles as reviewed in Meystre, Friedlin, South, Shen, and Samore (2010)) clinical data from the SAM system in use both in the Primary Healthcare units and in the Hospitals. Using the clinical data that is available for us we intend to take advantage of the tooling available to reach the objectives mentioned.

Inevitability of English Usage in Bioinformatics

In the Biomedical informatics domain the knowledge representation of choice has been evolving since the initial developments of the Gene Ontology project (http://www.geneontology.org/) in the beginning of the century into ontology based representations. With the advent of research about reasoning capabilities and its applicability in the Semantic Web a lot of services that expose that capabilities have appeared but all of them, that matter, exploring only English based terminology. Recently we have been testifying the appearance of a strong set of Web Services in the area of semantic annotation of texts in the Biomedical domain. These specialized, domain oriented services, are made available mainly by the U.S. NCBO (National Center for Biomedical Ontologies) through
Knowledge Management Systems Acceptance
www.igi-global.com/chapter/knowledge-management-systems-acceptance/14511?camid=4v1a

Multimedia, Information Complexity, and Cognitive Processing
www.igi-global.com/article/multimedia-information-complexity-cognitive-processing/1252?camid=4v1a