Chapter 28
Clinical Practice Ontology Automatic Learning from SOAP Reports

David Mendes
Universidade de Évora, Portugal

Irene Pimenta Rodrigues
Universidade de Évora, Portugal

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

ABSTRACT

We show how we implemented an end-to-end process to automatically develop a clinical practice knowledge base acquiring from SOAP notes. With our contribution we intend to overcome the "Knowledge Acquisition Bottleneck" problem by jump-starting the knowledge gathering from the most widely available source of clinical information that are natural language reports. We present the different phases of our process to populate automatically a proposed ontology with clinical assertions extracted from daily routine SOAP notes. The enriched ontology becomes a reasoning able knowledge base that depicts accurately and realistically the clinical practice represented by the source reports. With this knowledge structure in place and novel state-of-the-art reasoning capabilities, based in consequence driven reasoners, a clinical QA system based in controlled natural language is introduced that reveals breakthrough possibilities regarding the applicability of Artificial Intelligence techniques to the medical field.

INTRODUCTION

In the *KR*[1] community of the *AI*[2] science domain recent developments are sprouting that lead to the establishment of expressive ontologies based in *OWL2*[3] as the foundation for distributed reasoning in complex domains in the Semantic Web. This approach has revealed a very serious problem when trying to represent highly specialized and complex domains of science, namely the medical domain, that is the

DOI: 10.4018/978-1-5225-2237-9.ch028

"Knowledge Acquisition Bottleneck" (Wong et al., 2012). The ontology creation and development has traditionally been in charge of multidisciplinary domain experts that have to master both their particular scientific field and the ontology engineering novel techniques. These experts are very few and they are normally engaged in defining the ontological terminological structures of their domain ontologies, the *TBox*, and cannot be worried with the *ABox* to enrich the ontologies into *KB*[4]s that can be automatically reasoned upon.

A pragmatic approach is presented in order to identify the different issues faced when addressing the KAB[5] and for each one of them we discuss the possible and feasible solutions according to the State-of-the-Art in the Semantic Web and Artificial Intelligence science fields. Paramount interest arrived due to the very recent acknowledgment of the clinical practice encoding communities about the possibilities of redirecting efforts to capture the "*meaning of data*" instead of coding directed to a particular purpose like reimbursement or government funding and reporting as introduced in (Cimino, 2011). These efforts raise the level of awareness of the clinical reporting into the semantic possibilities that "understanding the meaning" of the encoded information uncovers like *CSI*[6] or *automated reasoning* around such complex domains.

Our pragmatic approach rests on the evidence that to overcome the *KAB* the tools available to the public in general have to be supportive in nature like:

- being based in natural language interaction,
- responsive in order to be accepted has a working tool and
- scientifically extremely accurate and self evident because not even the slightest chance of error is acceptable in the medical domain and especially in healthcare.

Moreover, we accept as evidence that natural language is the Lingua Franca for communication among physicians and we have to gather information in computable form from text that they traditionally use to pass their information. Thus, we can conform the scope of our work in:

1. Develop an Ontology suitable to model accurately the healthcare processes
2. Develop a process for automatic acquisition of the ontology ABox from text
3. Use this framework as support for natural language *CQA*[7] to be used as an *AI* professional aid.

In the remainder of the paper we develop the presentation of our work showing all the foundational reasons of every choice and discuss the results achieved:

- In section named "Ontology population in healthcare" we introduce the specific questions of automatically acquiring knowledge in the healthcare domain.
- We explain our knowledge engineering choices when presenting our ontology in section titled *OGCP*. We pay particular attention to the automatic process of knowledge acquisition in sub section *Automated OGCP enrichment* because it represents the epitome of our elaborate work.
- For our framework to be highly expressive while computationally efficient we explain in section *"Ontological relations learning and enrichment"* what are some assumptions that can be made profiting from the structure of SOAP to elicit expressive ontological relations and present a small example of what can be achieved with our proposed knowledge framework.
- Finally the conclusions are summarized in the last section "Conclusions".