

A methodology for extracting ontological knowledge from Spanish documents

Rafael Valencia García¹, Dagoberto Castellanos Nieves²,
Jesualdo Tomás Fernández Breis¹, Pedro José Vivancos Vicente¹

¹ Departamento de Informática y Sistemas. Facultad de Informática.
Universidad de Murcia 30071 Espinardo (Murcia). España
Tel: +34 968364613, Fax: +34 968364151
{valencia, jfernand, pedroviv}@um.es

² Facultad de Informática y Matemática,
Universidad de Holguín, Holguín, Cuba
dcn.ncd@gmail.com

Abstract. This paper presents a semi-automatic approach for extracting knowledge from natural language texts in Spanish. The knowledge is acquired and learned through the combination of NLP techniques for analyzing text fragments, the ontological technology for representing knowledge and MCRDR, a case-based reasoning methodology. This approach has been applied in the ontology domain and the results of this application are discussed in this work.

1 INTRODUCTION

Spanish is the official language of a significant amount of countries, and it has millions of speakers world-wide. Hence, there is a huge amount of information and knowledge in Spanish documents. So, extracting knowledge from such texts would be beneficial and of great help for the Spanish speaking community. The recognition of natural language has been traditionally viewed as a linguistic issue and based on grammars. However, grammars have different drawbacks, such as the fact that they are unable of managing ambiguity, imprecision, variability, etc. In order to overcome the drawbacks of grammar approaches, we have developed a methodology for acquiring knowledge from texts in an incremental way based on knowledge engineering and natural language processing techniques. In this paper, we describe such methodology and how it is capable of extracting knowledge from pieces of Spanish free texts. The combination of knowledge engineering technologies with natural language processing techniques provides us the goodnesses of both areas. As far as knowledge engineering technologies are concerned, two have been included in the methodology, namely, ontologies and MCRDR. Let us introduce now both technologies and the reason why they are used in the methodology.