

A new proposal of Word Sense Disambiguation for nouns on a Question Answering System^{*}

S. Ferrández¹, S. Roger^{1,2}, A. Ferrández¹, A. Aguilar¹ and P. López-Moreno¹

¹ Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante, Spain

² Natural Language Processing Group
Department of Computing Sciences
University of Comahue, Argentina
{sferrandez,sroger,antonio}@dlsi.ua.es,
{P.Lopez}@ua.es

Abstract. This paper describes the impact of the application of a Word Sense Disambiguation (WSD) algorithm for nouns on AliQAn [16], the Question Answering system with which we have participated in the CLEF-2005. Applying the traditional WSD decreases the performance in 4.7% on the Mean Reciprocal Rank (MRR). To solve this problem, we propose two different uses of WSD: (1) to choose a set of synsets instead of the traditional use of WSD, in which only one synset is chosen; (2) to disambiguate the words not present in EuroWordNet (EWN). Using our proposal of WSD the MRR increases a 6.3% with regard to the baseline without WSD. Furthermore, our proposal of WSD increases the MRR with regard to the traditional use of WSD in an 11%. Finally, the implementation of our approach of WSD is computationally efficient by means of a preprocessing of EWN.

1 Introduction

In this paper we analyze the benefits of a Word Sense Disambiguation (WSD) algorithm for nouns in AliQAn [16], a Spanish Question Answering (QA) system, with which we have participated in the CLEF-2005¹ competition.

QA objective consists of identifying the answer of questions in large collections of documents. QA is not a simple task of Information Retrieval (IR), QA tries to go beyond and returns a concrete answer in response to an arbitrary query. For the users, it is very interesting to find accurate information, thanks to the increment of available information. The QA systems are capable to answer questions formulated by the users in natural language.

^{*} This research has been partially funded by the Spanish Government under project CICYT number TIC2003-07158-C04-01 and by the Valencia Government under project number GV04B-268.

¹ <http://www.clef-campaign.org/>