SPARTE, a Test Suite for Recognising Textual Entailment in Spanish

Anselmo Peñas, Álvaro Rodrigo, Felisa Verdejo

Dpto. Lenguajes y Sistemas Informáticos, UNED
{anselmo,alvarory,felisa}@lsi.uned.es

Abstract. The aim of Recognising Textual Entailment (RTE) is to determine whether the meaning of a text entails the meaning of another text named hypothesis. RTE systems can be applied to validate the answers of Question Answering (QA) systems. Once the answer to a question is given by the QA system, a hypothesis is built turning the question plus the answer into an affirmative form. If the text (a given document) entails this hypothesis, then the answer is expected to be correct. Thus, a RTE system becomes an Answer Validation system. Within this framework the first problem is to find collections for training and testing RTE systems. We present here the SPARTE corpus aimed at evaluating RTE systems in Spanish. The paper presents the methodology to build SPARTE from the Spanish QA assessments performed at the Cross-Language Evaluation Forum (CLEF) during the last three editions. The paper also describes the test suite and discusses the appropriate evaluation measures together with their baselines.

1 Introduction

The task of Recognising Textual Entailment (RTE) [3] aims at deciding whether the truth of a text entails the truth of another text named hypothesis or, in other words, if the meaning of the hypothesis is enclosed in the meaning of the text. The entailment relation between texts is useful for a variety of tasks as, for example, Automatic Summarisation, where a system could eliminate the passages whose meaning is already entailed by other passages; or Question Answering (QA), where the answer of a question must be entailed by the text that supports the correctness of the answer.

Since RTE task has been defined recently, there exists only few corpora for training and testing RTE systems, and none of them are in Spanish. Thus, we planned the development of SPARTE, a corpus for training and testing RTE systems in Spanish, and specially, systems aimed at validating the correctness of the answers given by QA systems. This automatic Answer Validation would be useful for improving QA systems performance and also for helping humans in the assessment of QA systems output.

SPARTE has been built from the Spanish corpora used at Cross-Language Evaluation Forum (CLEF) for evaluating QA systems during 2003, 2004 and 2005. At the end of development, SPARTE contains 2962 hypothesis with a document label and a TRUE/FALSE value indicating whether the document entails the hypothesis or not.

Section 2 describes the development of SPARTE in detail. Section 3 evaluates some features of the corpus. Section 4 discusses and suggests the way of using SPARTE for evaluation purposes. Section 5 is devoted to some other corpora related to RTE. Finally, we give some conclusions and future work.