

Spanish Question Answering Evaluation

Anselmo Peñas, Felisa Verdejo and Jesús Herrera

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

Abstract. This paper reports the most significant issues related to the launching of a Monolingual Spanish Question Answering evaluation track at the Cross Language Evaluation Forum (CLEF 2003). It introduces some questions about multilingualism and describes the methodology for test suite production, task, judgment of answers as well as the results obtained by the participant systems.

1 Introduction

Evaluation forums as the Text REtrieval Conference (TREC¹), NTCIR project² or the Cross-Language Evaluation Forum (CLEF³) have shown their capability to stimulate research, to establish shared working lines, and to serve as a meeting point for their respective communities. These forums permit the comparison of different systems evaluated under the same conditions. Thus, some evidences about which are better approaches can be extracted. In this kind of evaluation, test suites must be produced to serve as the common evaluation exercises for every system under competition. Test suites generation requires a considerable effort that is justified in such evaluation forums. At the end, these test suites remain as a very valuable resource for future systems evaluation.

Question Answering (QA) research has been promoted and evaluated in such way since TREC-8 in 1999. Now, the Cross-Language Evaluation Forum (CLEF 2003) has brought new challenges: to consider different languages than English and to perform translangual QA [3]. The UNED NLP Group (Spanish Distance Learning University), as Spanish member of the CLEF consortium, is responsible for the Spanish test suite generation in all the QA tasks that involve Spanish, and is also responsible for the results assessments when Spanish take part as target language. We report here the most significant issues related the Monolingual Spanish evaluation task which has been launched in CLEF 2003.

Sections 2 and 3 describe the usual methodology for a QA evaluation based on systems comparison and introduce the challenge of multilingualism. Sections 4 and 5 describe the production of the Spanish test suite. The task, the assessment process and the results for the first monolingual Spanish QA evaluation are described in sections 6, 7 and 8 respectively.

¹ <http://trec.nist.gov>

² <http://research.nii.ac.jp/ntcir/index-en.html>

³ <http://www.clef-campaign.org>