HAREM
An Advanced NER Evaluation Contest for Portuguese
Diana Santos, Nuno Seco, Nuno Cardoso, Rui Vilela

Introduction
• HAREM is the first evaluation contest for Named Entity Recognition (NER) in Portuguese.
• HAREM features several innovative traits and provides a highly configurable open source evaluator.
• There were 10 participants from 6 different countries who submitted 15 official and 6 non-official submissions.

Golden Collection
• 129 documents containing marked named entities
• Documents are from 4 different geographical origins (mainly Portugal and Brazil) and of several genres.
• The categories employed during manual annotation were linguistically motivated.
• Annotation considered the NE in context.

Evaluation Architecture

Results
• New mini-HAREM has already taken place in April 2006.
• Several public resources were made available as a byproduct of HAREM. (SIEMES, REPENTINO, …)

Conclusions
• The architecture deployed in HAREM adds a number of degrees of freedom to experiment with relative to MUC or other NER evaluation contests.
• Resources and new evaluation metrics for NER were deployed.
• The state-of-the-art in NER for Portuguese was measured.

Linguateca
http://www.linguateca.pt