

**Extracção de relações semânticas. Recursos, ferramentas e estratégias**

**Título** Extracção de relações semânticas. Recursos, ferramentas e estratégias

**Autor/a** Marcos García González

**Directores** Pablo Gamallo Otero

**Tipo** Tese doutoral

**Data de lectura** 15/12/2014

**Lugar de lectura** Universidade de Santiago de Compostela


**Abstract** Relation extraction is a subtask of information extraction that aims at obtaining instances of semantic relations present in texts. This information can be arranged into machine-readable formats, useful for several applications that need structured semantic knowledge. This thesis explores different strategies to automate the extraction of semantic relations from texts in Portuguese, Spanish and Galician. Both machine-learning (distant-supervised and supervised) and rule-based techniques are investigated, and the impact of the different levels of linguistic knowledge is analyzed for the various approaches. Regarding domains, the experiments are focused on the extraction of encyclopedic knowledge, by means of the development of biographical relations classifiers (in a closed domain) and the evaluation of open information extraction systems. In order to implement the extraction systems, several natural language processing tools have been built for the three research languages: from sentence splitting and tokenization modules to part-of-speech taggers, named entity recognizers and coreference resolution systems. Furthermore, several lexica and corpora have been compiled and enriched with different levels of linguistic annotation, which are useful for both training and testing probabilistic and rule-based models. As a result of the work carried out in this thesis, new resources and tools are available for automated processing of texts in Portuguese, Spanish and Galician.

## LIGAZÓNS

 Teseo

## DESCARGAS

 Referencia BibTex

 Descargar versión completa

## PROXECTOS DE INVESTIGACIÓN

CELTIC: Coñecemento Estratéxico Liderado por Tecnoloxías para a Intelixencia Competitiva

## PROGRAMAS CIENTÍFICOS

Intelixencia na web (antigo)

Intelixencia de negocio e na web (antigo)