Collaborative and Automatic Methods for the Multilingualisation of Lexica and Ontologies

Carlos Ramisch
2013-11-08

# Table of Contents

Welcome to the website of the Cameleon Project! Collaborative and Automatic Methods for the Multilingualisation of Lexica and Ontologies.

This is a 4-year international cooperation project that aims at creating, reinforcing and continuing academic exchanges between French and Brazilian researchers in the domain of multilingual lexica and ontologies. It is co-financed by CAPES (Brazil) and by COFECUB (France), as project CAPES-COFECUB number 707/11. The project started in 2011 and is expected to be concluded by the end of 2014.

This project is part of the France-Brazil international laboratory LICIA .
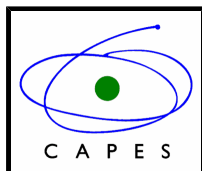
Keywords

- Natural Language Processing
- Web Semantic
- Lexical Resources
- Ontologies
- Multilingualism
- Machine Translation
- Information Retrieval
- Lexical Acquisition
- Ontology Matching
- Multiword Expressions

Project Goals

Scientific goals

The goal of this project is to investigate, propose, experiment, apply and validate automatic and collaborative techniques for the development of lexical and ontological resources that can be useful in the context of multilingual applications, particularly for French, Portuguese and English.

Therefore, it aims at the investigation of methods for acquiring linguistic information for the construction of lexical resources, integrating multilingual lexica and ontologies, focusing on collaborative and automatic techniques. In the former, volunteer contributors can use a platform to edit dictionary entries and to create links, online via a Web browser. Analogously, multilingual applications can access and contribute automatically to the lexical resources stored on the platform through an API. The latter, i.e. automatic construction of resources, is based on the extraction of lexical information from textual corpora, using empirical/statistical evidence and machine learning techniques.

The integration of automatic and collaborative methods has several advantages because they are somehow complementary. On the one hand, collaborative methods could use automatically generated data as a starting point, thus saving time and effort when creating a new instance (for a new language/domain/language pair). On the other hand, data-driven methods produce noisy results that should be later filtered by human experts. The use of collaborative platforms seems the most natural environment for post-editing automatically extracted lexical and ontological resources.

Formation goals

1. Strengthening of links between distinct groups and universities
2. Sharing of knowledge and resources between groups and universities
3. Experience of distinct research communities for students and researchers, contributing for a richer formation
4. Experience of a distinct cultural environment for students and researchers, contributing for a better integration and awareness of citizens of these communities
5. Publications of the results in academic publications

Resource goals

- Serious lexical game

- Lexical network
- Lexical database management system
- Automatic multilingual multiword expressions identification
- Multilingual ontology alignment