CAMELEON 2012

Carlos Ramisch
2012-10-18

Table of Contents

CAMELEON 2012

Second CAMELEON workshop, Grenoble, France, 6-7 september 2012

Participants :

  Speakers: Carlos Ramisch (LIG/UFRGS), Achille Falaise (LIG), Mathieu Mangeot (LIG), Emmanuelle Esperança-Rodier (LIG), Ying Zhang (LIG), Aline Villavicencio (UFRGS)

  Other participants: João Comba (UFRGS), Georges Fafiotte (LIG), Francieli Zanon (UFRGS/LIG), João Cláudio Américo (UFRGS/LIG), Laércio Lima Pilla (UFRGS/LIG), David Rouquet (LIG), Gilles Sérasset (LIG), Mutsuko Tomokiyo (LIG), Rosa Vicari (UFRGS), LingXiao Wang (LIG)

Workshop Presentations

Carlos Ramisch : The CAMELEON comparable corpus

In my presentation, I introduce the CAMELEON comparable corpus. This corpus is a collection of texts in the conference organisation domain. It contains several million words, was crawled from the web and is available in Portuguese, French and English. The goal of the corpus is to support research on automatic management of ontologies related to the corpus. It can help in tasks such as ontology learning, enriching, population, matching and management. Additionally, it can be used for studies on the integration of lexical and ontological resources. These results were published in the ACL 2012 workshop Multilingual Modelling and were obtained as a joint work with Cassia Trojahn, Renata Vieira, Roger Granada, Aline Villavicencio, and Lucelene Lopes.

Achille Falaise : Web site translation with interactive Multilingal Access Gateways

Website translation is traditionally achieved in two ways: by human translation, or by machine translation through a gateway. Human translation quality is reliable but it is slow, costly, and cumbersome to implement. Automatic translation, however, is fast, inexpensive, lightweight to implement, but not reliable. We present here the iMAG (interactive Multilingual Access Gateways) system. It aims to merge this two approaches, by introducing human post-editing over machine translation, in order to control and improve the automatic translation of websites.

Mathieu Mangeot : JeuxDeMots-pt: a serious lexical game for Brazilian Portuguese

JeuxDeMots is a serious lexical game that builds a rich lexical network with specific relations (associated ideas, synonyms, hyponyms, hypernyms, domain, etc.). Each game has two players. A word is proposed and the first player must give associated words. Then, the game is proposed to a second player. If they enter words in common, they obtain points and credits. In the background, a relation is built between these words in common and the word suggested in the game. The first version, French, was launched in 2007. For the CAMELEON project, we decided to launch a Portuguese version in december 2011. We already have some results: 61 players, 20 000 terms and 1600 relations.

Emmanuelle Esperança-Rodier : MT evaluation and multilingualism

My talk will mainly introduce Machine Translation (MT) Evaluation, focusing on MT Evaluation based on tasks and according to user skills. This type of evaluation is an alternative approach to the state of the art MT evaluation which is based on data and entails evaluation bias. I will also talk about multilingualism. Those two concepts are experimented in Cameleon Project.

Ying Zhang : Jibiki-PIVAX, a generic platform for multilingual lexical resources management

In my presentation, I introduce the Jibiki-PIVAX project. The Jibiki platform is an online generic environment for writing, importing, exporting and querying all kinds of dictionaries (microstructures and macrostructures). The Microstructure is the structure of entries, we use CDM (Common Dictionary Markup) pointers to index heterogeneous microstructures without modifying them. The macrostructure is the structure of the different volumes composing a dictionary. We use rich links with type, label and weight between entries of different volumes in order to deal with the different macrostructures (bi-volumes, pivot, etc.). For the PIVAX project, we use a specific macrostructure with several volumes for each language (each volume is used in a specific context like a MT system). Then, there is an axeme volume for each language that gathers the same meaning of entries of the different volumes into one monolingual acception (axeme). The axemes are in turn linked to a pivot volume of axies, interlingual acceptions.

Aline Villavicencio : Natural Language Processing at UFRGS

In her talk, Aline presented the "Neurocognition and Language" research group at UFRGS, and explained their research axes, specially in what concerns language acquisition and the creation and evaluation of language resources.

Future actions

Ph.D. Defense of Carlos Ramish

Carlos will defend his Ph.D. work, tuesday 11 September 2012, 2 pm. at IMAG F018 room, campus of St Martin d'Hères, Grenoble, France.

   The title is : A generic and open framework for multiword expressions treatment: from acquisition to applications

Ph.D. internship of Leonardo Zilio

Leonardo Zilio, student of the Informatics department of UFRGS brazilian university will spend one year from november 2012 at GETALP-LIG laboratory to work on his Ph.D. He will be co-supervised by Mathieu Mangeot and Carlos Ramisch.

Leonardo will work with us on semantic role label annotation for specialised and general-purpose texts. We are specially interested in comparing sets of semantic role labels across languages and to study their impact on multilingual applications such as the construction of rich bilingual predicate dictionaries and machine translation. The integration of the existing annotation into the lexical resources created by GETALP in the CAMELEON project is a parallel goal of the internship. The expenses are financed by the CAMELEON CAPES-COFECUB project.

Third edition of the CAMELEON workshop and shared task

The tentative date is 19/12/2012, and the workshop will gather the team members around a common shared task along the lines of multilingual ontology alignment and/or machine translation using heterogeneous resources.