

## **Preface: Recent Advances in Natural Language Processing**

As an attempt to cover the constant academic progress made within the field of computational linguistics and natural language processing in Spanish, the Postgraduate Program of Linguistics of the University of Costa Rica organized the first *Colloquium on Natural Language Processing* that took place in February 2016. The following papers were specially selected to conform the present issue of the CLEI Electronical Journal.

In the first article, “Comparison of Two Forced Alignments Systems for Aligning Bribri Speech” written by Rolando Coto Solano and Sofía Flores Solórzano, the authors made a comparison between the FAVE-align and the EasyAlign systems in order to detect which system provided a more accurate alignment for the Bribri speech of the Chibchan language.

The second article, “Trending Topic Extraction Using Topic Models and Bitern Discrimination” by Minor Eduardo Quesada Grosso, Edgar Casasola and Jorge Antonio Leoni de Leon, deals with the mining and exploitation of data in Twitter. The authors make use of the Topic Model Bitern Bursty algorithm. Then, they create a graph from the co-occurrence of terms so that term discrimination can reduce the amount of processing required to identify trending topics.

The next paper, “Evaluation of Named Entity Recognition Algorithms in Short Texts” by Raquel Fonseca and Edgar Casasola, calculates the precision of two of the Entity Recognition algorithms in a corpus taken from The SemEval-2015 Aspect Based Analysis conference. This study analyses the suitability of processing a large amount of text generated in the social media.

In the fourth paper, “Evaluation of Potential Spanish Text Markers on Social Posts as Features for Polarity Classification”, written by Edgar Casasola Murillo, Antonio Leoni de León, Gabriela Marín Reventós, the authors make a sentiment analysis through lexical markers that indicated emphasis or denoted emotion in a Spanish corpus from Facebook postings. After normalizing their corpora and identifying the markers, the authors make use of the TASS 2015 collection for the polarity classification task.

The fifth article, “Automatic parameterization of Support Vector Machines for Short Texts Polarity Detection”, written by Aurelio Sanabria and Edgar Casasola, makes a comparison between the Grey Wolf Optimizer and the Grid Search –two algorithms for parameter optimization– on support vector machines using radial basis functions. This allows the authors to identify polarity on short Spanish texts from Twitter.

As a special guest for this issue, we also present the article “Dictionaries of Mexican Sexual Slang for NLP” by Roberto Villarejo Martínez, Noé Alejandro Castro Sánchez and Gerardo Sierra. In this paper the creation of a morphological and a semantic dictionary is discussed. Addressing the double entendre of Mexican sexual slang, the

authors base their morphological dictionary on Freeling Library; whereas, for the semantic dictionary, they employ synsets from WordNet in order to provide representative meanings for the dictionary.

We hope these articles may provide the reader with a wider scope for the current investigations that are being lead within the computational linguistics and natural language process fields as well as their implementation on current social and media problems.

*Gerardo Sierra and César Aguilar*  
*Invited Editors*