



Departamento de Computación

UNIVERSIDADE DA CORUÑA

On **Friday, May 2nd, 2014, 17:00**,
at the **Aula de Graos Víctor Gulías**
(**Facultade de Informática, Campus de Elviña, A Coruña**),

Prof.

Yue Zhang

(**Singapore University of Technology and Design, Singapore**)

will give an **invited talk** on:

Joint word segmentation, POS-tagging and parsing of Chinese using word structure annotations and a shift-reduce algorithm

Abstract:

Chinese sentences are written as continuous sequences of characters without explicit boundaries between words. On the other hand, words do exist in Chinese, and play an important role in the comprehension of sentences. Traditional Chinese NLP takes word segmentation as a first step, breaking Chinese sentences into a sequence of words before higher-level tasks are performed. In syntactic processing, a typical pipeline include segmentation, POS-tagging and parsing.

However, there are two limitations of such pipelined methods. First, they suffer from error propagation. Second, syntactic information, which can help improve segmentation, is not used in the segmentation step. More interestingly, Chinese words themselves have structures, which can be treated as a syntactic tree over characters. Head nodes from the character trees can be useful indicators of the syntactic roles of words in a sentence. How to leverage such information is an interesting research question.

In this talk, I present our on-going research on joint segmentation and syntactic processing. I begin by talking about joint segmentation and POS-tagging, for which external syntactic information (POS) helps segmentation. Then I introduce joint segmentation, POS-tagging and parsing by using manually annotated word internal structures, which allows both internal (word structure) and external (POS and syntax) information to be used to improve segmentation. A left-to-right incremental learning and search framework is used in all tasks to give highly competitive accuracies with linear time complexity.

Bio:

Yue Zhang is currently an assistant professor at Singapore University of Technology and Design. Before joining SUTD in July 2012, he worked as a postdoctoral research associate in University of Cambridge, UK. Yue Zhang received his DPhil and MSc degrees from University of Oxford, UK, and his BEng degree from Tsinghua University, China. His research interests include natural language processing, machine learning and artificial intelligence.



Organised by:

Grupo de Lengua y Sociedad de la Información (www.grupolys.org)

Carlos Gómez Rodríguez (cgomezr@udc.es)

Supported by:

Xunta de Galicia – Consolidación e estruturación de unidades de investigación competitivas: Grupos con Potencial de Crecemento (CN2012/008)