IEEE SG WIE AG Technical Talk

New Reordering and Modeling Approaches for Statistical Machine Translation & Plagiarism Detection

Thursday 3 March 2011

6:30pm -7:30pm
(Refreshment will be provided from 6pm)

The IEEE Singapore Women-In-Engineering (WIE) Affinity Group has the pleasure to invite you to attend the following technical talk.

Speaker: Dr. Marta Ruiz Costa-jussà, Universitat Politècnica de Catalunya (UPC, Barcelona).

Synopsis

This talk is divided in two main parts. The first part is in the area of Statistical Machine Translation. Particularly, we will present a reordering model which consists of generating weighted reordering hypotheses using the same powerful techniques of SMT systems in order to undo the source language structure and to make it more similar to the target language structure. Therefore, the translation challenge is divided into two steps: predicting the order of the words in the target language and substituting these words in the target language. In order to infer new reorderings that were not learnt during training, the SMR system uses word classes instead of words themselves. In order to correctly integrate the SMR and SMT systems, both are concatenated, by using a word graph. This approach is an elegant and efficient reordering approach that is capable of achieving significantly improved translation in the target language.

In the second part, in the area of information retrieval, we will present our participation in the CLEF 2010 plagiarism detection lab. The objective of the task refers to detecting plagiarized sections in a suspicious document and the corresponding source sections in a given set of source documents. In a first step, we build an information retrieval system based on SOLR/Lucene, segmenting both suspicious and source documents in smaller texts. We perform a search based on bag-of-words which provides a first selection of plagiarized texts. In a second step, we implemented a sliding window approach that computes cosine distances between overlapping text segments from both the source and suspicious documents in a pair wise basis. As a result, a similarity matrix between text segments is obtained, which is smoothed by means of low-pass 2-D filtering. From the smoothed similarity matrix, plagiarized segments are identified by using image processing techniques.

Speaker's Biography

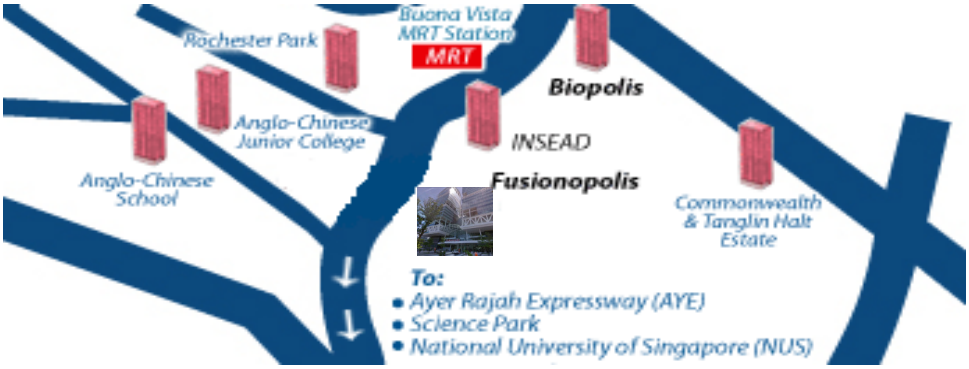
Marta Ruiz Costa-jussà is a Telecommunication's Engineer by the Universitat Politècnica de Catalunya (UPC, Barcelona). She received her PhD from the UPC in 2008. She has worked in LIMSI-CNRS (Paris) both in Automatic Speech Recognition and Statistical Machine Translation. She has taught Engineering and Informatics undergraduate courses in the UPC, in the Universitat Oberta de Catalunya and in the Universitat Pompeu Fabra. She has been scientific consultant of several translation companies such as TaWithYou and UniversalDoctor. Currently, she has a 'Juan de la Cierva' government-fellowship in Fundació Barcelona Media, where she is working on Machine Translation, Cross-Language Information Retrieval and Opinion Mining. She has participated in several European and Spanish national projects. She has been part of more than 10 program committees of international conferences. Her research publications include more than 15 international scientific journals and 40 international conferences.

Free Registration: For catering purposes please register before **Wednesday, March 2nd, 2011** at the following link http://ieeeps.org/ems1/regforevent.php?event_id=231

Contact person: IEEE SG WIE secretary at 62482248/92960231 or ieeesgwie@gmail.com

Location: Fusionopolis, Franklin at Level 11, Connexis South Tower, 1 Fusionopolis Way, Singapore 138632

Map: http://maps.google.com.sg/maps?hl=en&q=singapore+138632&um=1&ie=UTF-8&hq=&hnear=138632&gl=sg&ei=MKBcTZeUCZH4sAOnp_nrBQ&sa=X&oi=geocode_result&ct=image&resnum=1&ved=0CBkQ8gEwAA



GETTING TO FUSIONOPOLIS

Drop-off and Parking at Fusionopolis

Drop Off - Fusionopolis' driveway along Ayer Rajah Ave

Parking

• On Ayer Rajah Ave, turn left at first junction, **Fusionopolis Way**

• Turn left at next junction, **Portsdown Road**

• Go straight and turn left at a small lane, **Fusionopolis Plc**

• Head straight down to Basement Carpark; public parking lots are on levels **B4 - B6**.

By Mass Rapid Transit (MRT) service

Commuters may take East-West train line and alight at Buona Vista MRT station (EW 21). Walk to North Buona Vista Road (opposite Buona Vista MRT) and transfer to bus no. 91. Alight at the second bus stop which is on Ayer Rajah Avenue, after INSEAD.

Shuttle buses are also available for commuters from the MRT station. For more information on the routes and timings, please click http://www.thebamboospa.com/pdf/Fusionopolis%20Shuttle%20Bus.pdf.

By public bus services from Buona Vista MRT to Fusionopolis

(Board from the side of Biopolis / Ministry of Education building)

• Service 91 stops in front of INSEAD, which is next to Fusionopolis.

• Services 92, 95, 198 and 200 ply North Buona Vista Road. Alight two bus stops down at North Buona Vista Flyover (opposite Ayer Rajah Industrial Estate). It is just a 2-minute walk to Fusionopolis from that bus stop.