

WePS-2 Workshop - Tuesday 21st April 2009

9:30-10:00 **Workshop summary**

10:00-11:00 **Clustering systems**

- *Fuzzy Ants Clustering for Web People Search.*
Els Lefever, Timur Fayruzov, Véronique Hoste and Martine De Cock.
- *Person Name Disambiguation on the Web by TwoStage Clustering.* Masaki Ikeda, Shingo Ono, Issei Sato, Minoru Yoshida and Hiroshi Nakagawa.
- *PolyUHK: A Robust Information Extraction System for Web Personal Names.*
Ying Chen, Sophia Yat Mei Lee and Chu-Ren Huang.

11:00-11:30 coffee break

11:30-12:10 **AE systems**

- *A Two-Step Approach to Extracting Attributes for People on the Web.*
Keigo Watanabe, Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka.
- *Which Who are They? People Attribute Extraction and Disambiguation in Web Search Results.* Man Lan, Yu Zhe Zhang, Yue Lu, Jian Su and Chew Lim Tan.

12:10-13:00

Invited talk: Hugo Zaragoza (Yahoo! Research, Spain)

Entity Search in online collections

Automatic linguistic annotations of text can be used today in a number of ways: to create richer interfaces to the information locked in document collections, to help the user express its information need, and to improve the relevance of the results obtained by the search engine. I will give an overview of our recent work in these three areas, using example applications on online collections such as Wikipedia, financial news and Yahoo! Answers.

13:00-14:30 lunch

14:30-16:00 **Poster session**

Each team will give 90 second introduction

16:00-16:30 coffee break

16:30-17:00

Invited talk: Enrique Amigó (NLP & IR group, UNED, Spain)

Selecting and combining clustering evaluation metrics for WEPS

There is a wide set of evaluation metrics available to compare the quality of text clustering algorithms. In addition, in clustering tasks there is a substantial trade-off between precision and recall oriented metrics which usually depends on a clustering threshold parameter stated in the algorithm. Therefore, selecting and combining the most appropriate metrics for system optimization is not a trivial issue. I will describe the suitability of BCubed metrics and the combining criterion UIR (Unanimous Improvement Ratio) according to formal properties and empirical results over WEPS data.

17:00-17:15

Invited talk: Paul McNamee (Johns Hopkins University, USA)

The Knowledge Base Population track at TAC 2009

A new evaluation will take place at the Text Analytics Conference that involves cross-document entity coreference, relation discovery, and question answering. Participants are expected to add facts to a reference knowledge base, created from Wikipedia pages with infoboxes. Systems will be assessed on several sub tasks: (1) linking entities to KB nodes (or determining that no appropriate node exists); (2) filling in missing slot values for a given entity; and, (3) where appropriate, slot values must be linked to nodes in the KB.

17:15-18:00 **Panel: Future of WePS**