

# Evaluating Named Entity Recognition and Disambiguation in News and Tweets

Giuseppe Rizzo<sup>\*1,2</sup>, Marieke van Erp<sup>†3</sup>, and Raphaël Troncy<sup>‡2</sup>

<sup>1</sup>Università di Torino, Turin, Italy

<sup>2</sup>EURECOM, Biot, France

<sup>3</sup>VU University Amsterdam, The Netherlands

Named entity recognition and disambiguation are important for information extraction and populating knowledge bases. Detecting and classifying named entities has traditionally been taken on by the natural language processing community, whilst linking of entities to external resources, such as DBpedia and GeoNames, has been the domain of the Semantic Web community. As these tasks are treated in different communities, it is difficult to assess the performance of these tasks combined.

We conduct a thorough evaluation of the NERD-ML approach [2], an approach that combines named entity recognition (NER) from the NLP community and named entity linking (NEL) from the SW community. We present experiments and results on the CoNLL 2003 English 2003<sup>1</sup> dataset for NER and the AIDA CoNLL-YAGO2<sup>2</sup> dataset for NEL in the newswire domain and on the MSM'13<sup>3</sup> corpus for NER and the Dercynski *et al.* [1] dataset for NEL in the twitter domain.

Our results indicate that combining approaches from the NLP and SW communities improves the performance for NER. As the NEL task is more recent, there is as yet no common agreement on the annotation level to adopt, which makes it difficult to assess the performance of different systems.

## References

- [1] Leon Derczynski, Diana Maynard, Giuseppe Rizzo, Niraj Aswani, Marieke van Erp, Raphaël Troncy, and Kalina Bontcheva. Named Entity Recognition and Linking for Microblogs. *Currently under review*, 2013.
- [2] M. van Erp, G. Rizzo, and R. Troncy. Learning with the Web: Spotting Named Entities on the intersection of NERD and Machine Learning. In *3<sup>rd</sup> Workshop on Making Sense of Microposts (#MSM2013)*, 2013.

---

\*giuseppe.rizzo@di.unito.it

†marieke.van.erp@vu.nl

‡raphael.troncy@eurecom.fr

<sup>1</sup><http://www.cnts.ua.ac.be/con112003/ner>

<sup>2</sup><http://www.mpi-inf.mpg.de/yago-naga/aida/downloads.html>

<sup>3</sup><http://oak.dcs.shef.ac.uk/msm2013>