

Evaluating Entity Linking: An Analysis of Current Benchmark Datasets and a Roadmap for Doing a Better Job

Filip Ilievski¹ Pablo Mendes², Heiko Paulheim³, Julien Plu⁴, Giuseppe Rizzo⁴,
Felix Tristam⁵, Marieke van Erp,¹ and Jörg Waitelonis⁶

¹ VU University Amsterdam

² IBM Research USA

³ University of Mannheim

⁴ EURECOM

⁵ CITEC, Bielefeld University

⁶ Hasso-Plattner-Institut, Universität Potsdam

Abstract. Entity linking has become a popular task in both natural language processing and semantic web communities. However, we find that the benchmark datasets for entity linking tasks do not accurately evaluate entity linking systems. In this paper, we aim to chart strengths and weaknesses of current benchmark datasets and sketch a roadmap for the community to devise better benchmark datasets.

An extended abstract that followed from this discussion was submitted to LREC 2016.

Keywords: natural language processing, semantic web, entity linking