Towards Better Entity Linking Evaluation

Henry Rosales-Méndez. Department of Computer Science, University of Chile

supervised by Aidan Hogan and Barbara Poblete





2 Are the Entity Linking systems targetting the same entities?

Example annotations produced by four EL systems: AIDA (A), Babelfy (B), DBpedia Spotlight (D) and TAGME (T).

 $\underbrace{\text{Joseph Walter "Joe" Jackson, was a steelworker at } \underbrace{\text{U.S. Steel.}}_{B|D} \underbrace{\text{Steel}}_{A|B|T}. \text{ In an interview } \underbrace{\text{D}|T}$

with Martin Bashir for the 2003 documentary Living with Michael Jackson, $\frac{A|D|T}{D}$

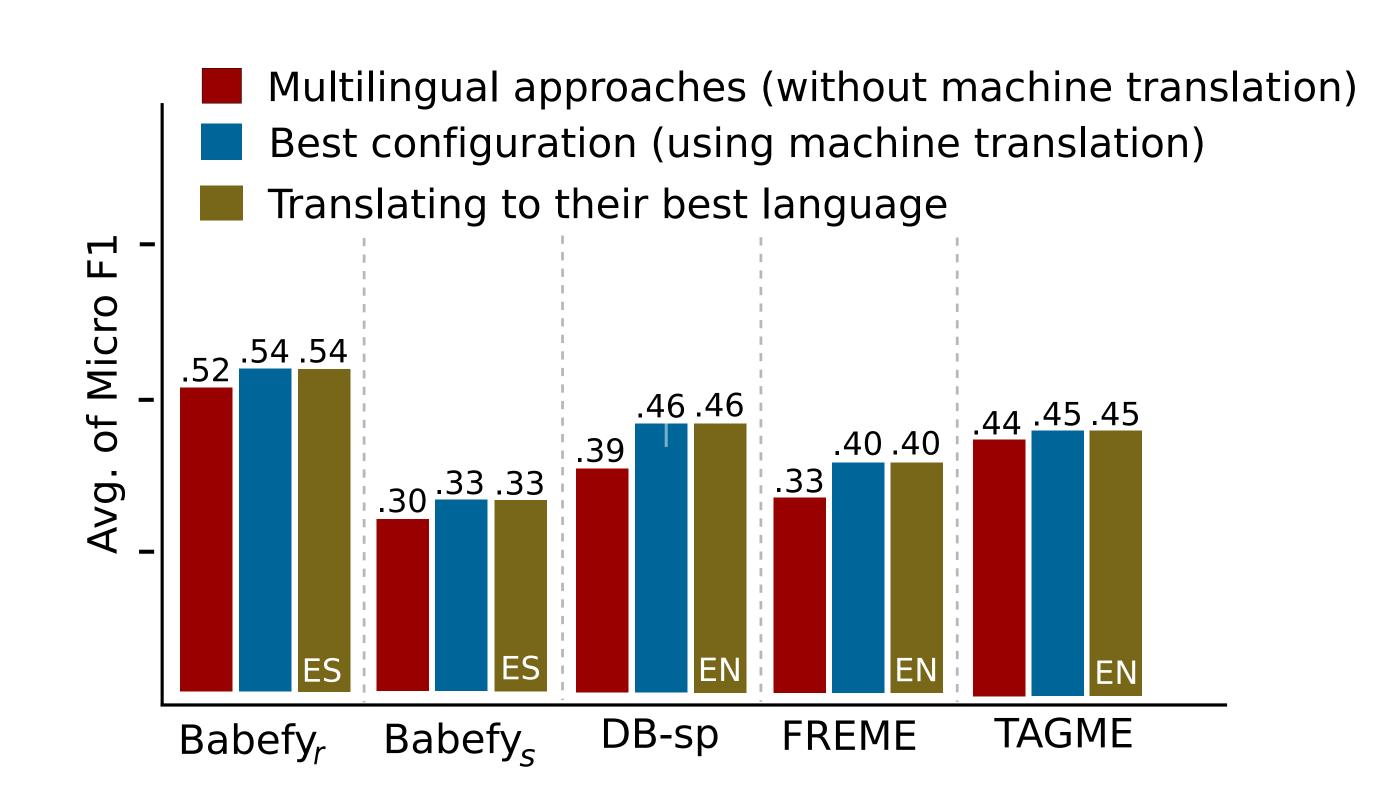
 $\frac{\text{Jackson recalled that } \underline{\text{Joe}}}{A|B} \text{ often sat in a } \underline{\text{chair}}_{T} \text{ with a } \underline{\frac{\text{belt}}{D|T}} \text{ in his } \underline{\frac{\text{hand}}{T}} \text{ as he and}$

his siblings rehearsed.

3 Are the Entity Linking datasets enough?

Dataset	Languages		
AIDA/CoNLL-Complete	EN		
KORE50	EN		
IITB	EN		
ACE2004	EN		
AQUAINT	EN		
MSNBC	EN		
DBpedia Spotlight	$\mathbf{E}\mathbf{N}$		
N3-RSS 500	${ m EN}$		
Reuters 128	${ m EN}$		
Wes2015	${ m EN}$		
News-100	DE		
Thibaudet	FR		
Bergson	FR		
SemEval 2015 Task 13	EN,ES,IT		
DBpedia Abstracts	DE,EN,ES,FR,IT,JA,NL		
MEANTIME	EN,ES,IT,NL		
$VoxEL_R$	DE,EN,ES,FR,IT		
$VoxEL_S$	$\mathrm{DE}, \mathrm{EN}, \mathrm{ES}, \mathrm{FR}, \mathrm{IT}$		

Could we not simply focus on supporting one language in the EL system and translate the input text to that language?



Results of the Relaxed Version of VoxEL

5 Are the NIF tools enough for handling benchmark datasets?

QRTool	NIF-Dataset Creation	X Demo	X Source Code	
BENGAL	Automatic NIF Creation	│ ⊘ Demo	Source Code	
Eaglet	NIF-Dataset Validation	X Demo	Source Code	
GERBIL	Benchmark Framework	⊘ Demo	Source Code	Visualization
Orbis	Benchmark Framework	X Demo	X Source Code	✓ Visualization
NIFify	NIF-Dataset Creation / Va Demo Source Co			

6 Proposed Approach

The present Ph.D. work proposes to address a variety of open questions regarding the evaluation of EL systems. In this direction, the following research questions are being or will be addressed:

- 1. What should Entity Linking link?
 - (a) How can we define the goal of the EL task?
 - (b) Is consensus possible on the definition of an "entity"?
 - (c) If not, how can we define benchmark EL datasets and what metrics can we use to reflect the lack of consensus?
- 2. How well do EL systems perform in multilingual settings?
 - (a) How can we compare EL performance across languages?
 - (b) How do EL systems perform for different languages?
 - (c) Why do results differ across languages?
 - (d) Are multilingual EL approaches necessary with recent improvement in machine translation?
- 3. How can we adapt and configure EL techniques for different applications?
 - (a) What are the specific applications for EL?
 - (b) What are the different settings that can be considered?
 - (c) Which techniques work best under what assumptions?
 - (d) How can existing EL techniques be best configured to meet the needs of a particular application setting?



Acknowledgements The work of Henry Rosales-Méndez was supported by CONICYT-PCHA/Doctorado Nacional/2016-21160017. The work was also supported by the Millennium Institute for Foundational Research on Data (IMFD) and by Fondecyt Grant No. 1181896.