

Parsing Social Media Text

Timothy Baldwin



THE UNIVERSITY OF
MELBOURNE

Overview

- Social media text is noisy (who'd've thought?), and often has real-time processing needs (now you're talkin'!) ... although it's not necessarily the case that everything is equally worthy of parsing (possibility of coarse-to-fine strategy of some type)
- **Objective:** real-time, über-robust parsing of social media data/translation into a semantic representation
- **Applications:** real-time IE/event detection over social media streams, ...

Seed Thoughts

- There is some mileage to be had in pre-processing the texts to “bring the text to the parser” (e.g. [lexical] normalisation, text categorisation, named entity recognition, POS tagging, ...)
- There is also some mileage to be had in building more robustness into the parser proper, incrementally adapting the parse selection model to the data, and also Yi-ing the data using a PCFG “estimate” of the full parser

Discuss ...

- Who cares (within DELPH-IN)?
- How can we make people care outside DELPH-IN?
- What is a sensible semantic representation to map onto (DMRS?), and how can we predict what level of semantic specification is appropriate for a given noise level in the input?
- How fast can we reasonably make it?
- What is the best way forward?