

WG3: Hybrid Parsing of MWEs

Polish Academy of Sciences

17th September 2013

Outline

- Givens
 - Objectives and Challenges
 - Members
- Discussion points
 - WG Leader/Vice-Leader
 - Refining objectives, challenges, title
 - Workplan issues

WG3 Objectives

- Increasing the efficiency and accuracy of parsing methods, especially hybrid ones.
- Paving the way towards using widely accessible unannotated data in order to improve grammars and models based on annotated data.

WG3 Challenges/Outcomes

- Cope with long-distance relations and discontinuities in probabilistic parsing.
- Integrate high-quality language resources into probabilistic parsing (e.g. MWE-oriented reranking).
- Enhance knowledge-based parsing of MWEs with probabilistic scores.
- Enhance supervised methods (using scarce annotated corpora) with unsupervised ones (using unannotated corpora).

Current Members

- Croatia: Zeljko Agic (ESR P)
- Portugal: Antonio Branco (K)
- France: Matthieu Constant (P)
- Israel: Yaakov HaCohen-Kerner (P)
- Greece: Stella Markantonatou (F K)
- Spain: Montserrat Marimon (ESR F K)
- Malta: Michael Rosner (P/K)
- Sweden: Joakim Nivre (P/K)

Current Membership

- Countries: 8
- ESR: 2/8
- M/F: 2/8
- Knowledge Based: 3/8
- Stats/Prob/ML: 3/8
- Hybrid: 2/8

Candidate Leader/Vice-Leader

- Mike Rosner
- Matthieu Constant
- Other candidates??

Refining the Scope/Title?

- Why this might be needed
 - Small numbers (so far)
 - Perceived overlap/confusion with WG2
- Suggestion: include wider processing issues than parsing/include application areas
 - including in particular MT?
 - including other applications e.g. topic extraction, sentiment analysis, summarisation,
- Hybrid parsing **and processing** of MWEs?

Possible Research Issues

- Knowledge-based enhancements to data-driven methods
- Data-driven enhancements to knowledge-based methods
- Hybrid methods for MWE annotation
- Resources:
 - Hybrid resource production
 - existing ML terminology databases
- Evaluation of hybrid-methods
- Translation
 - parsing, transfer, generation
 - language models and translation models
- Other application areas

Further Research Issues

Workplan

- Y1:
 - Fix objectives/outcomes
 - Organise workshop for Athens meeting
 - State of the art report
 - Resource catalogue
 - Specify STSMs
- Y2:
 - Elaboration of objectives set in Y1
 - STSMs

Workplan Year 1

- contrastive state-of-the-art survey (subcommittee)
- detailed scientific program (subcommittee)
- workshop (subcommittee)
- website contents:
 - collect/publish resources
- mailing lists and internal website
- workshop proceedings, pubs & tech docs
- annual report