

WG3: Workplan

Polish Academy of Sciences

18th September 2013

Outline

- Members
- Objectives
- Title
- Workplan

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)
- Turkey: Gulsen Eryiğit (P/K F ESR)

Current Membership

- Countries: 9
- ESR: 3/9
- M/F: 3/9
- Knowledge Based: 3/9
- Stats/Prob/ML: 3/9
- Hybrid: 3/9
- check reps. from WG2, WG4
- email mike.rosner@um.edu.mt

Research Issues

- Elaboration and evaluation of hybrid processing of MWEs i.e.
 - *Knowledge-based enhancements to data-driven methods*
 - *Data-driven enhancements to knowledge-based methods*

with respect to:

- Parsing
- Resource production, annotation, exploitation
- Translation
 - parsing, transfer, generation, decoding
 - language models and translation models
- Other MWE-relevant application areas

WG3 Title

- Hybrid parsing of MWEs
- Hybrid [and multilingual] [approaches to] [processing|parsing] [and translation] of MWEs
 - Hybrid processing of MWEs
 - Hybrid approaches to processing and translation of MWEs
 - Hybrid and multilingual approaches to processing of MWEs

Workplan Y1

- Determine and publish WG objectives/ outcomes/title
- Identify subWGs for parsing, resources, MT
- Organise workshop for Athens meeting
 - Form subcommittee (3)
- State of the art report on MWE hybrid methods and resources (2)
- Resource catalogue; explore relationship with METANET (1) and CLARIN (1)
- STSM proposals
- Website contents (1)

Workplan Y2

- Elaboration of objectives set in Y1
- Execution of STSMs; new STSMs
- Explore and instantiate relationships with other WGs, particularly WG2 and WG4

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.
- [MT/ML]

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).

Candidate Leader/Vice-Leader

- Mike Rosner
- Matthieu Constant

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?

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