Italian NLP for Everyone: Resources and Models from EVALITA to the European Language Grid

Valerio Basile*, Cristina Bosco*, Michael Fell*, Viviana Patti*, Rossella Varvara†

*University of Turin, Italy †University of Fribourg, Switzerland

Goals and Outcome
- Broaden the ELG portfolio with linguistic resources for Italian, capitalizing on the work done in the last ten+ years within EVALITA:
  - the evaluation campaign of NLP tools for Italian
  - further support for underrepresented languages
  - Integrate systems and baselines as a pool of web services with a common interface
- Experiments with multi-task learning

The ELG now covers the majority of language technology developed for Italian in the last 10 years: a broad spectrum of linguistic phenomena, language genres and domains (e.g., social media).

Linguistic Resources
- 43 corpora
- 1 lexical/conceptual resource
- Metadata and terms of availability

Models and Systems
- 8 systems:
  - pre-neural models
  - neural networks
  - general neural language models (e.g. AIBERTo)

Experiment: Ensemble Systems

3 models for HS detection
~10 lines of code!
Majority ensemble

Project on ELG: https://live.european-language-grid.eu/catalogue/#!/resource/projects/1397
Project website: http://evalita4elg.di.unito.it/

---

**Goals and Outcome**

- Broaden the ELG portfolio with linguistic resources for Italian, capitalizing on the work done in the last ten+ years within EVALITA:
  - the evaluation campaign of NLP tools for Italian
  - further support for underrepresented languages
  - Integrate systems and baselines as a pool of web services with a common interface
- Experiments with multi-task learning

The ELG now covers the majority of language technology developed for Italian in the last 10 years: a broad spectrum of linguistic phenomena, language genres and domains (e.g., social media).

**Linguistic Resources**

- 43 corpora
- 1 lexical/conceptual resource
- Metadata and terms of availability

**Models and Systems**

- 8 systems:
  - pre-neural models
  - neural networks
  - general neural language models (e.g. AIBERTo)

**Experiment: Ensemble Systems**

3 models for HS detection
~10 lines of code!
Majority ensemble

---

Project on ELG: https://live.european-language-grid.eu/catalogue/#!/resource/projects/1397
Project website: http://evalita4elg.di.unito.it/