ELG Pilot Project:
Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools
EVALITA4ELG
Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG platform

Core team:

- Viviana PATTI (PI)
- Valerio BASILE
- Cristina BOSCO
- Michael FELL
- Rossella VARVARA

Institution:
University of Turin, Computer Science Department
Content-Centered Computing group: http://di.unito.it/ccc

CELLI Language Technology
Italian Language Technology enterprise: https://www.celi.it/en/

EVALITA4ELG
Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools for the ELG platform

The Italian language is underrepresented in the European Language Grid (ELG) platform, currently including few Language Technology services and corpora - mostly partial corpora and multilingual dependency trees, focused on texts filtered by standard terms and syntagms. The goal of EVALITA4ELG is to enable the ELG users to access the resources and models for the Italian language produced over the years in the context of the EVALITA evaluation campaign. Our aim is to build the catalogue of EVALITA resources and tasks ranging from traditional tasks like POS tagging and parsing to recent and popular ones such as sentiment analysis and hate speech detection on social media, and integrate them in the ELG platform. The project includes the integration of state-of-the-art NLP services into the ELG platform, accessible as web services.

http://evalita4elg.di.unito.it/
Main goals:

- Broaden the ELG portfolio with more than 50 linguistic resources for Italian, capitalizing on the work done in the last + ten years within EVALITA (http://www.evalita.it):
  - the biannual event of AILC for the evaluation of NLP tools for Italian
  - further support for underrepresented languages
  - sharing of knowledge within the NLP community

- Integrate systems and baselines as a pool of web services with a common interface

- Experiment with cross-lingual and cross-domain tasks, multi-task learning
Main tangible outcome: the ELG platform will cover the majority of language technology developed for Italian in the last 10 years (resources + services)

13 → 50 language resources

Broader spectrum of linguistic phenomena, language genres and domains (e.g., social media)

Reproduction of results

Multi-task benchmark platform for Italian