

EVALITA4ELG

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

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: the evaluation campaign 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

G1 Linguistic Resources

- Surveying** the corpora published by the organizers of EVALITA tasks and their terms of availability
- Collecting** the resources and restructuring the data in order to standardize their format and adhere to the ELG guidelines

G2 Systems and Baselines

- New **service functions** for Italian based on
 - i) the best models from EVALITA task participants
 - ii) general Transformer language models
- Enrich** the ELG platform implementing new baselines for a broad variety of tasks.
- Multi-task **benchmark** platform for Italian

Team

Coordinator



UNIVERSITÀ DEGLI STUDI DI TORINO
Dept. of Computer Science
University of Turin



Viviana Patti (PI)



Valerio Basile



Cristina Bosco



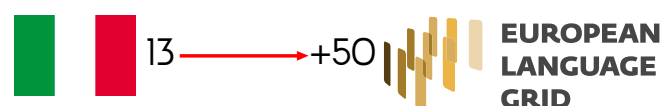
Rossella Varvara



Michael Fell

Outcome

The ELG platform will cover the majority of language technology developed for Italian in the last 10 years



language resources + services

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



Running since 2007

6 tasks at EVALITA 2016

10 tasks at EVALITA 2018

14 tasks at EVALITA 2020



Github repository
collecting data
since 2016

Schedule and Milestones

September 2020 → August 2021

Experiment (M1-M8)

T.1.1 Surveying tasks and models, data collection

T.1.2 Developing Web services

T.1.3 Cataloguing available EVALITA systems

Integration (M8-M12)

T.2.1 Anonymizing resources

T.2.2 Defining and harmonizing licenses

T.2.3 Integrating software

Subcontractor



Andrea Bolioli



Alessio Bosca

Project coach: Khalid Choukri (ELRA)

Project on ELG: <https://live.european-language-grid.eu/catalogue/#/resource/projects/1397>

Project website: <http://evalita4elg.di.unito.it/>