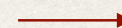


European Language Grid – META-FORUM 2020
Viviana Patti (Università Degli Studi di Torino, Italy)

ELG Pilot Project: Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools

EVALITA4ELG



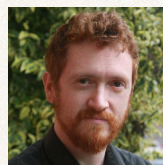
EUROPEAN
LANGUAGE
GRID

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/ccg>



di.unito.it



CELI Language Technology

Italian Language Technology

enterprise: <https://www.celi.it/en/>



Andrea BOLIOLI



Alessio BOSCA

Home Partners

EVALITA4ELG

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

The Italian language is underrepresented in the European Language Grid platform, currently including few Language Technology services and corpora - mostly parallel corpora and multilingual dependency treebanks, focused on texts featured by standard forms and syntax. 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 LT services into the ELG platform, accessible as web services.

Grav was </>



Khalid Choukri
Project coach

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

EVALITA4ELG



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

❖ 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>):

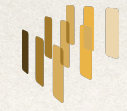
the biannual event of AILC for the evaluation of NLP tools for Italian

- 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

EVALITA4ELG



EUROPEAN
LANGUAGE
GRID

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

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.

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



EUROPEAN
LANGUAGE
GRID

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

Reproduction of results

Multi-task benchmark platform for Italian