



Italian legislative text classification for Gazzetta Ufficiale

M. Rovera, A. P. Aprosio, F. Greco, M. Lucchese, S. Tonelli, A. Antetomaso



Project (IPZS & FBK)

Gazzetta Ufficiale (GU): official organ for legislative acts of the Italian Republic.

Current workflow: manually indexing based on IPZS subject index

Project goals

- 1) ***Automatizing the indexing/classification process*** of legal acts in the GU
- 2) Guiding the ***migration process*** of IPZS ***towards*** the adoption of european standard taxonomy ***EuroVoc***



Dataset

Name	Institution/Topic	N. of samples*	N. unique labels
<i>Serie Generale</i>	Central/Periph. administrations	363,989	1003
<i>Serie speciale 1</i>	Constitutional Court	16,485	566
<i>Serie speciale 2</i>	European Community	2,285	227
<i>Serie speciale 3</i>	Regional Affairs	25,926	610
<i>Serie speciale 4</i>	Public Exams	-	-
<i>Serie speciale 5</i>	State Contracts	-	-

*All data are referred to the time span 1988 - 2021.



Methods and evaluation

Large scale **multi-label** text classification on act's titles - static vs. contextual embeddings

<i>Metric</i>	FastText	BERT
Precision (micro)	0.791	0.816 (+0.025)
Recall (micro)	0.803	0.832 (+0.029)
F1 (micro)	0.797	0.824 (+0.027)
F1 (macro)	0.271	0.285 (+0.014)
F1 (weighted)	0.779	0.807 (+0.028)



Adoption of EuroVoc (ongoing)

Migration from the IPZS subject index to european standard taxonomy EuroVoc

Challenges

- **different structure** and **size** of the two taxonomies
- **different annotation practices** adopted so far

Experiments

Semi-supervised data-driven ontology alignment using multiple dataset (GU and EurLex)



Thank you for the attention!

Italian legislative text classification for Gazzetta Ufficiale

M. Rovera, A. P. Apro시오, F. Greco, M. Lucchese, S. Tonelli, A. Antetomaso