EMBEDDIA: CROSS-LINGUAL EMBEDDINGS FOR LESS-REPRESENTED LANGUAGES IN EUROPEAN NEWS MEDIA

CONSORTIUM

ACADEMIC PARTNERS

- **Jožef Stefan Institute (SI) - Coordinator**
- University of Ljubljana (SI)
- Queen Mary University of London (UK)
- University of Helsinki (FI)
- University of La Rochelle (FR)
- University of Edinburgh (UK)

NEWS MEDIA INDUSTRY PARTNERS

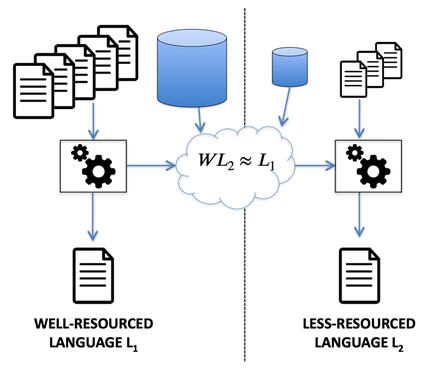
- Trikoder – Styria Media Group (CRO)
- Ekspress Meedia (EE)
- Finnish News Agency STT (FI)

NLP SME

- TEXTA OÜ (EE)

OVERALL GOAL

- Exploit **cross-lingual embeddings & deep neural networks** for **news analysis and generation**
- Empower **less-represented, morphologically-rich EU languages** to benefit from resources and tools for well-resourced languages



ADVANCES FOR NLP

NEW NLP RESOURCES:

- Contextual embedding models (ELMo, BERT) for multiple less-represented languages
- Benchmark datasets and tasks for evaluation of contextual embeddings in multiple languages

NEW GENERAL NLP TOOLS:

- Cross-lingual semantic enrichment (NER, NEL, event detection)
- Multi-lingual dynamic topic model
- Explanation and visualization tools for deep neural networks

GENERAL INFO

- Duration: 01.01.2019 – 31.12.2021
- Budget: 3m EUR
- Call:
  - ICT-29-2018: A Multilingual Next Generation Internet
  - RIA: Domain-specific/challenge-oriented HLT
- **www.embeddia.eu**
- **@embeddiaproject**

ADVANCES FOR NEWS MEDIA

CROSS-LINGUAL ANALYSIS TOOLS:

- Keyword extraction for news articles
- Sentiment & topic analysis for articles and comments
- Automated comment moderation

MULTI-LINGUAL GENERATION TOOLS:

- Article text generation from structured data
- Creative headline generation
- Summarization and reporting

This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 825153

Institut "Jožef Stefan" Ljubljana, Slovenija