Eastern European model and dataset collection.
Language similarity and distillation effects transferability across languages.

**CONTRIBUTIONS**

- **Dataset and model index**
  - We present a broad index of existing Eastern European language resources published as a GitHub repository.

- **Benchmark datasets**
  - We provide hand-crafted cross-lingual datasets for five different semantic tasks, compiled by processing data from various sources into the same format.

- **Baselines**
  - We perform several experiments with the existing multilingual models on our datasets to define performance baselines.

**EENLP Index on GitHub**

- https://github.com/altsof/EENLP
- 90+ datasets, 45+ models

**CONCLUSIONS**

- Multilingual pre-training is crucial for cross-lingual transfer learning.
- Distillation decreases the ability for cross-lingual transfer learning.
- Similarity of languages matters for the cross-lingual transfer learning quality.
- XLM-RoBERTa generally dominates over the rest of the tested models.

Alexey Tikhonov, Alex Malkhasov, Andrey Manoshin, George Dima, Réka Cserháti, Md. Sadek Hossain Asif, Matt Sárdi