On Zero-shot and Few-shot transfer for different scripts for NER

Supervisor: Silvia Severini
Examiner:
BSc, MSc, Open: Open
General Topic Area: Named Entity Recognition, Multilinguality.
Summary: Named Entity Recognition (NER) is the task of delimiting and categorizing mentions to entities in text. Deep neural networks achieve state-of-the-art results for NER, but require large amounts of annotated data for training that could be scarce for certain languages [1].
In this thesis, you will analyze the zero-shot and few-shot [2] performance of NER models and investigate which factors influence the transfer (e.g., language structure, families, …). This requires analysys of the SOTA to formulate hypotheses about what helps and what hinders knowledge transfer. We will build and test neural NER models and analyze results to verify the hypotheses.
Additionally, you may want to investigate the role of romanization in NER knowledge transfer, similar to [3] that analyze romanization in Machine Translation and show that it can be effective when applied to the target side.
Prerequisites: Experience with Python, ideally experience with machine learning libraries

[1] Nasar, Z., Jaffry, S. W., & Malik, M. K. (2021). Named Entity Recognition and Relation Extraction: State-of-the-Art. *ACM Computing Surveys (CSUR)*, *54*(1), 1-39.
[2] Lauscher, Anne, et al. "From zero to hero: On the limitations of zero-shot cross-lingual transfer with multilingual transformers." arXiv preprint arXiv:2005.00633 (2020).
[3] Amrhein, Chantal, and Rico Sennrich. "On Romanization for model transfer between scripts in neural machine translation." arXiv preprint arXiv:2009.14824 (2020).