

ELTeC Data Analysis, Representation of the (Geo-)Entities and Interlinking with Knowledge bases (Wikidata) in Python

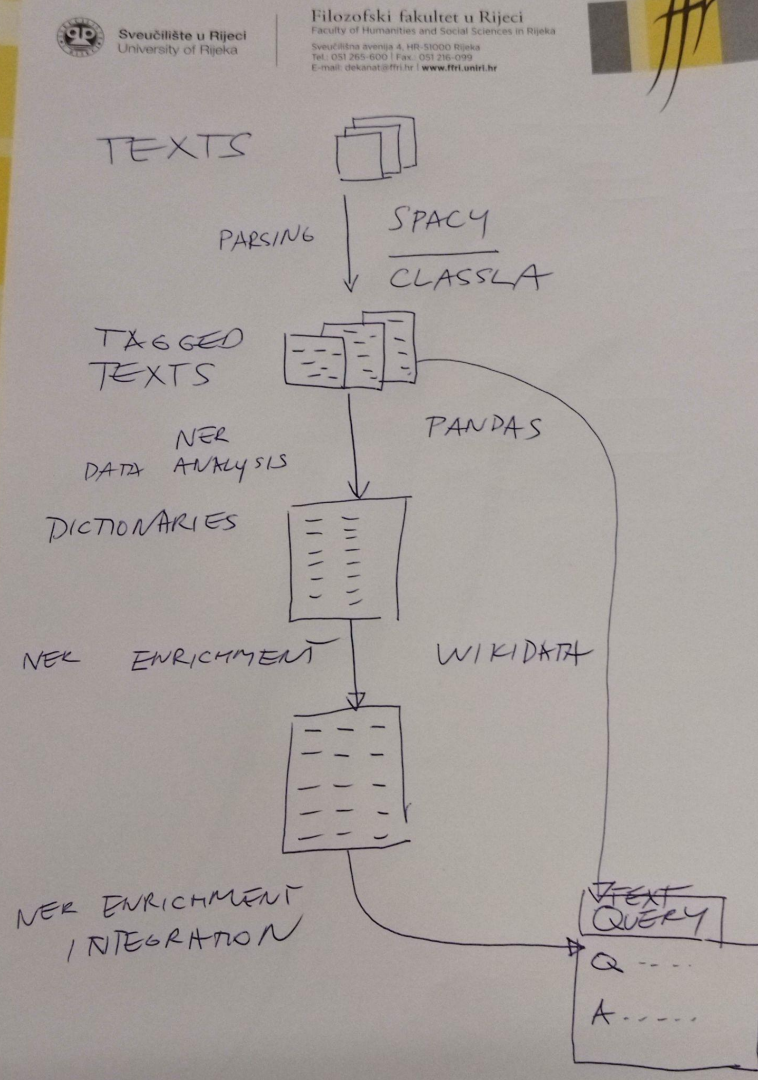
Benedikt Perak

Faculty of Humanities and Social Sciences,
University of Rijeka
bperak@uniri.hr

Agenda:

Create a Python-based pipeline for:

- a) ELTeC extraction
- b) multilingual parsing of texts using standard libraries (SpaCy, Classla),
- c) information enrichment for NE tagged structures,
 - i) NE data analysis on language specific corpora
 - ii) implementation of automated NE geolocation and visualisation (using Neonatim, Folium)
 - iii) extraction of information from Wikidata
- d) implementing information enrichment for creating contextual queries.



Resources

This presentation:

<https://docs.google.com/presentation/d/16xZZqjbl4SF9DBSQX3XUS3dym9CgPPXWw2Ze2JylfEc/edit?usp=sharing>

Google Colab:

<https://colab.research.google.com/drive/1BgZBBXVX6gOU5qn8k3wuUy4Z0rqnM--V?usp=sharing>

Shared OneDrive folder:

https://uniri-my.sharepoint.com/:f:/g/personal/bperak_uniri_hr/Es9uqLbatzVBipWEWbu2ESEB_MXEhjbImQldsl83zy4raQ?e=Thju2k