Named Entity Recognition (NER) in Resumes

SAKI Exercise 2

Challenge

To extract named entities (labeled individual data) from an existing data set of resumes using an NER / NLP library (Flair)

Approach

- Implement an NER algorithm for the resume data using the Flair-NLP library
- Run and evaluate the algorithm and its performance on ColabGPU

Data Set

Fictional resumes are available in this directory for download.

All entries in this data set consists of

- the original content and
- hand-labeled entities.

The data set should be split into training, test, and validation data sets.

Technology

- Programming technology
  - Python 3.6 (or later compatible versions)
  - Jupyter notebook
- ML / AI technology to use
  - Flair-NLP https://research.zalando.com/welcome/mission/research-projects/flair-nlp/
  - It uses state-of-the-art approaches, is widely used, and is well-documented
- ColabGPU
  - Assuming you don’t have a data center humming under your desktop

Literature

[YB18] Yadav, V., & Bethard, S. (2018, August). A survey on recent advances in named entity recognition from deep learning models. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 2145-2158).