max planck institut informatik

Problem 1 (Relation Extraction).

Task Description: In this lab, we are working on relation extraction. Our end goal is to be able to extract the following properties for given entity and their Wikipedia abstract:

- Date of Birth
- Nationality
- Alma Mater
- Awards
- Places of Work

In this exercise, we focus on using pattern-based extraction. You can use any tool to pre-process the data, like POS tagging or entity recognition. *Optional:* You can use any other resources to improve your patterns, like dictionaries of relational paraphrases (e.g. RELLY or POLY[1]). You may also use pretrained word embeddings like word2vec or BERT. However, you are **not** allowed to look up relations in existing KBs like DBpedia, Wikidata, YAGO. The properties must be extracted from the provided input text.

Your task is to complete the five property extraction functions called from `your_extracting_function()` inside `run.py`. The output is saved in the `results.csv` file. Store entities as strings and properties as lists. For properties which have no value, an empty list is stored `[]`.

Data and Scripts: Download the `lab05.zip` file from the course website. Extract and rename the folder `Lab05_XXXXX_YourName` by replacing with *your* matriculation number and name. The folder has the following files:

`run.py`	: Calls the property extraction functions and stores the result in `results.csv`.
`input.csv`	: Input file to `run.py` with columns `entity` and `abstract`.
`groundtruth.csv`	: Groundtruth properties for entities in `input.csv` with columns `entity,dateOfBirth,nationality,almaMater,awards,workPlaces`.
`evaluate.py`	: Script to evaluate your extractions.
`run_evaluate.sh`	: Script for executing extraction and evaluation.

You can run and evaluate your program by using: `./run_evaluate.sh`
`run_evaluate.sh` first executes the `run.py` script to extract the properties from the `input.csv` and saves it to `results.csv`. It then executes the `evaluate.py` script to calculate and print the evaluation metrics.

Submission: Once you complete make sure to include the `run.py` file from above and all additional code and files you used to execute `run.py`. If you used any external libraries, please indicate them in a README file. Your final submission **need not** contain the files `input.csv`, `groundtruth.csv`, `evaluate.py` and `run_evaluate.sh`.

Please submit all necessary files, which are compressed into a zip file named:
Lab05_MatriculationNumber_Name.zip
to the email address: **akbc-assignments@mpi-inf.mpg.de** with title of the email: **[AKBC]Lab05_MatriculationNumber_Name**

Deadline: 23:59 30.05.2022 (Monday)

[1] https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/patty/