

Problem 1 (Entity Typing). In this lab, we are working on entity typing. We provide a sample file, `train.tsv`, with the following format:

```
entity-mention [tab] list-of-types [tab] sentence
```

Based on this data set, build a typing model that outputs a list of types for a given entity mention. Your model can be built using any kind of methods such as pattern-based, rule-based, supervised or a combination, with the following constraints:

- You are **allowed** to use POS tags and dependency parses, and pre-trained word embedding dictionaries (e.g. GloVe) and WordNet (only the type system. You are **not allowed** to lookup an instance to get its types there).
- You are **not allowed** other resources, in particular lookups in existing KBs/encyclopedias, such as Wikidata, Wikipedia or Yago, entity embeddings such as Wikipedia embeddings, and pre-trained entity-typing models, including named entity recognition (NER) tools of existing libraries.
- If using supervised approaches, the model **has to be trained only on the given dataset**.

Evaluation

We provide sample evaluation data, `test.tsv` & `test-groundtruth.tsv`, and a script for evaluating your output. The files `test.tsv` and `test-groundtruth.tsv` have the following format, respectively:

```
id [tab] entity-mention [tab] sentence  
id [tab] ground-truth-types
```

Name your submission, `run.py`. It must be callable from the command line as follows:

```
python run.py test.tsv results.tsv
```

The first parameter is an input file of sentences (like `test.tsv`), and the second parameter is a filename where your predictions are stored, in same format as `test-groundtruth.tsv`:

```
id [tab] predicted-types  
for example: 1 [tab] ['person', 'musician'], or, 1 [tab] [], if it cannot predict the types.
```

To evaluate your results, run the script file, `evaluate.py`, by using the following command:

```
python evaluate.py results.tsv test-groundtruth.tsv
```

You can run typing and evaluation at the same time using file `run_evaluate.sh`:

```
./run_evaluate.sh test.tsv results.tsv test-groundtruth.tsv
```

Your submitted files must include all necessary code and files, especially the typing file `run.py`, and your trained models if you use supervised approaches. If you used any external libraries, please indicate them in a README file.

Note: In grading, we only run the bash file.

```
./run_evaluate.sh our-testfile.tsv results.tsv our-testfile-groundtruth.tsv
```

Please submit all necessary files, which are compressed into a zip file named:

Lab04_MatriculationNumber_Name.zip

to the email address: `cxchu@mpi-inf.mpg.de` with title of the email: **[IE]Lab04_MatriculationNumber_Name**

Deadline: 23:59 16.11.2019 (Saturday)