# Documentation

# TAaDS

# Winter 23/24

# Team 03

# Topic 07 (Skill Identification in NER)

## Given:

A file (jobs_crawl_filter_jobsuche_en.csv) of job offers.

## Sought:

A representation of skill identification in Name Entity Recognition for all jobs.

## Installation: Dependecies is describe in README.md file.

## Method:

- Process text from raw csv file.
- Extract job title and description by using conditions (6978 unique job founded).
- Train our NER to extract skills for the description according to their position.
- To prepare training data, we use ner-annotator [1] to annotate text for spaCy NER model.
- Train new entity and label as skills and identify all skills.

| TEXT | START | END | LABEL |
|---|---|---|---|
| Telemarketing | 1 | 15 | Skill |
| Python | 44 | 54 | Skill |

- For further and get more specific NER as Hard skill & soft skill, we use a open source library, named SkillNER [2,3].
- With this library we successfully classify soft skills(English language), hard skills(Python) and certificate(Engineering professional).

## Limitation (Problems with NER identification):

- Working with Unstructured Text data always difficult. If We apply this custom Train model in a text where no skill remains, it could lead to wrong identification.

- For this particular task, skill identification, no training data available. No Training data available. We had to prepare manually training data by annotating.
- In this work we didn't perform any model evaluation. Also it is difficult evaluate or analysis of correctness of final outcomes if we have large text data.
- It would be nicer if we able to label the separated skill. For example
    - Python → Programming skills
    - Working as a SW developer → Experience

  Unfortunately, we did not able label like that.

## Result with custom trained NER model:

| **Custom Trained Model** |
|---|
| linchpin in the electrical sector |
| Oberschwaben as an electrician |
| electronics technician |
| service technician |
| several years of |
| MSR environment |

## Result with SkillNER library:

| **SkillNer** |
|---|
| electrical |
| electronics service |
| electronics service |
| automation |
| develop personally |
| addition |

## References:


1. https://tecoholic.github.io/ner-annotator/
2. AnasAito/SkillNER: A (smart) rule based NLP module to extract job skills from text
3. Mining and mapping soft skills from any text – ScienceDirect