

## NATURAL LANGUAGE PROCESSING APPLIED ON A JOB PLATFORM

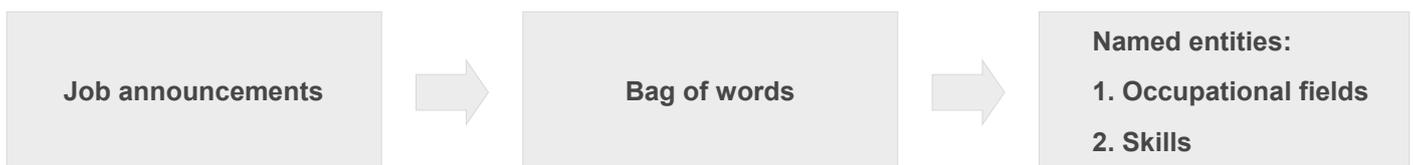
### PROJECT GOAL

Analyzing of job market trends or finding an appropriate candidate is always a time consuming task. Thus, searching an entity of interest results in hundreds or even thousands of outcomes. The aim of the project was to develop a model that extracts information of occupational fields and employees' skills according to job advertisements.

The implementation was divided into two steps: web crawling of a job portal to sample the data of more than 100 companies and creation of a deep learning solution to recognize the targeted named entities of the 30 DAX members.

### PROVIDED DATA

The project is based on a data set of posted job announcements obtained by the web crawling.



### CHALLENGES

Firstly, lots of job platforms prohibit web crawling. Additionally, the used portal made permanent changes of a source code what interferes crawling as well.

During training of an algorithm a sufficient amount of words with wide meaning spectrum was detected. As far only entities with specific meaning were of interest, other entities were filtered out.

### APPLIED METHODS

A structured data set contained descriptions of positions, names of companies, locations and dates of publication was created by the Python-based web crawler. The database was available for download as a CSV file.

The occupational fields and skills were labeled with the help of the Dataturks tool and downloaded as a JSON format.

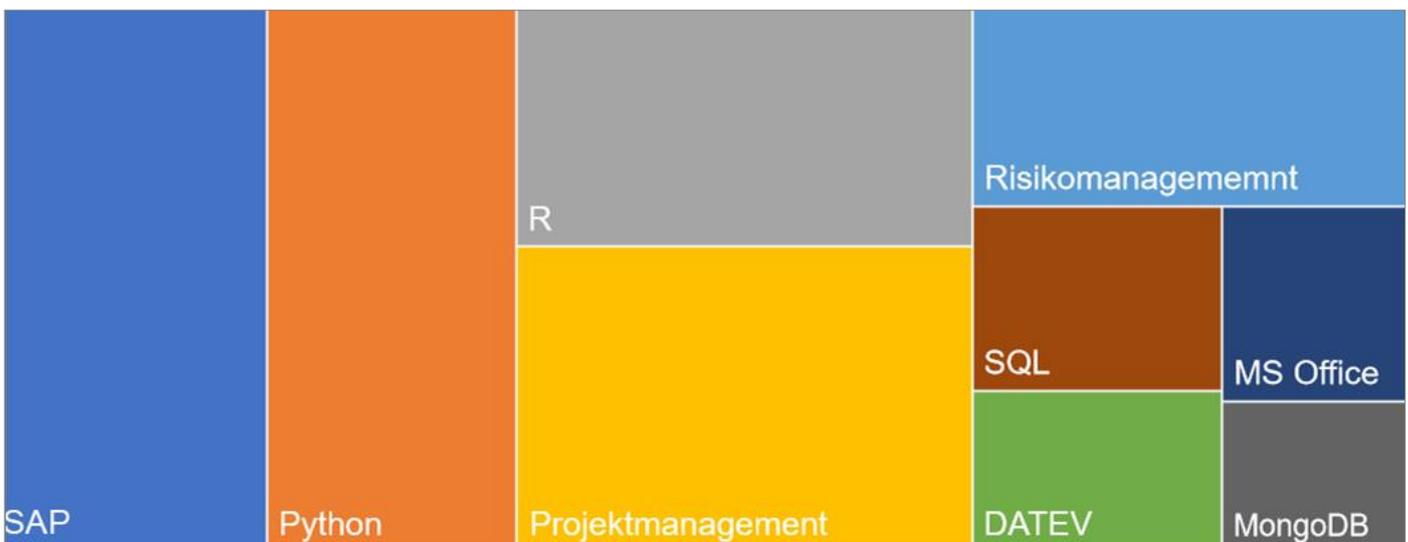


The labeled entities were divided into training and test data sets. By means of the spaCy library for natural language processing the algorithm was trained and the trained model was saved to be loaded anytime for the subsequent analysis. The test data set was evaluated by sklearn metrics with high precision (95.1%). Finally, the 30 DAX members were web crawled as well and the named entities corresponded to the occupational fields and skills were stored. Since both single words and phrases were labeled, couple of German stop words such as “und”, “sowie” and “auf” appeared. These stop words were removed in order to get just the meaningful entities.

### PROJECT OUTCOME

The obtained results are presented in a user-friendly dashboard. The user can select a company, skills or occupational fields depending on individual needs. Filtering by companies will provide a diagram where sizes of rectangles correspond to the number of opened positions in that company.

### SEARCHED SKILLS AT ADIDAS



By selecting a specific skill or experience, a relational amount of opened positions in all companies will be displayed. Further analysis as combined skills can be carried out as well.

The created algorithm can be extended to scan the huge amount of candidates resumes a company obtains.

The results presents a solution to economize companies' resources for market research.

### LAB EXPERIENCE

