

**Title:** CompanyName2Vec: Company Entity Matching Based on Job Ads

**Project:** CEM

**Author:** Ran Ziv

**Date:** September 2021

Contains:

— — — — —

- /code/ - directory with an archive of the CEM project
  - Project contains several directories:
    - This Readme file
    - jobAdsProcessing - contains fingerprinting and job ads corpus processing jobs
    - model - contains the method building blocks
      - model\_emb.py - primary executable, includes environmental settings. Loads the input data, generates the model and calls the evaluation process
      - evaluate\_emb.py - contains the evaluation process. Calls the index\_emb.py
      - Index\_emb.py - builds an index for evaluation purposes
      - datasetFiltering.py - prepare job ads corpus and save it to input directory. includes filtering capabilities for testing purposes
      - utils - contains several utilities, like t-test calculator, fuzzy distance test function, data export/import functions, etc.
  - /Data/cem-input-export.zip - compressed directory with input directories and files for the CEM project
  - /Data/cem-output-export.zip - compressed directory with output directories and files for the CEM project

Instructions:

— — — — —

- Download and extract files from both Code and Data directories
- Make sure python 3.6.8 or compatible version of it is installed and all required packages are installed
- Edit environment variables in model\_emb.py
- Execute model\_emb.py
- Results will be printed to stdout