Team ZeRoS

Reeka Hazarika (UAE) – Free Lancer hazarika.reeka@gmail.com

Swapnil Vishwakarma (India) - SIES Graduate School of Technology

swapnilvishwakarma7@gmail.com

Zyad Al-Azazi (Lebanon) – Lebanese American University

zyadazazi@gmail.com

Specialization: Natural Language Processing

Week 9 Deliverables:

Problem Description:

When companies recruit for any position, they usually end up receiving thousands, if not millions, of resumes. Such a huge number of resumes makes the task of going over all these resumes an extremely difficult and tedious job for HR employees. This made a lot of companies opt for systems that take the necessary information from the candidate after they fill an application with all the required fields. The solution worked greatly for employers; nevertheless, candidates have always found it very illogical to spend tens of hours sharpening their CVs and cover letters only to find out that they must spend another hour or so re-entering all the information they have on their CVs in the designated fields.

Data cleansing and transformation done on the data.

Swapnil Vishwakarma:

Data Cleaning Methods: Worked on expanding contractions, checking null values, removed stop words & lemmatized it.

Reeka Hazarika:

Data Cleaning Methods: Used Tokenization, POS Tag, Stopwords, Lemmatization , TF-IDF & Word2Vec.

Zyad Al Azazi:

Data Cleaning Methods:
- Lowering case of all characters.
- Removing unnecessary characters (including full stops at the end of sentences).
- Removing stop words.
- Tokenizing according to spaces using the split() function.
- Converting numbers from digits to strings.

As for featurization, Tf-Idf was used with n-gram ranging between 1 and 3.

GitHub Repo Link:

zyadalazazi/resume_extraction_team_zeros (github.com)