TEAM NLP

Nilsu Bozan - bozannilsu@gmail.com - Binghamton University/Istanbul Technical University

Nishchay Vaid - Nishchay89@gmail.com - Rutgers University

Anish Mitra - anishmitra9666@gmail.com - Montana State University

Sukriti Macker - sm11017@nyu.edu - New York University

Specialization: NLP

WEEK10 DELIVERABLES

PROBLEM DESCRIPTION:

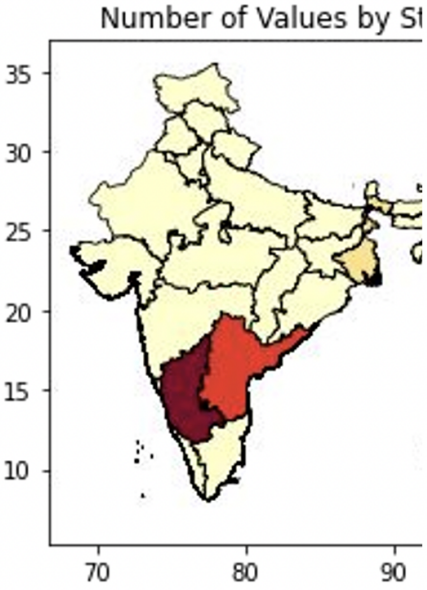
- Parsing resumes with very different formats and inputting the important information into the company database is something that sites with widespread usage such as workday do not do accurately.
- This is frustrating to both job seekers and HR professionals.
- We are trying to ameliorate this problem through this project.
- Implementing NLP techniques, we hope to reduce the difficulties involved in applying for a job and thus streamline the process in the human resources industry.

GITHUB REPOSITORY LINK:

https://github.com/nishchayvaid/Resume-Extraction/tree/main/Resources/Week10

EDA:

LOCATION



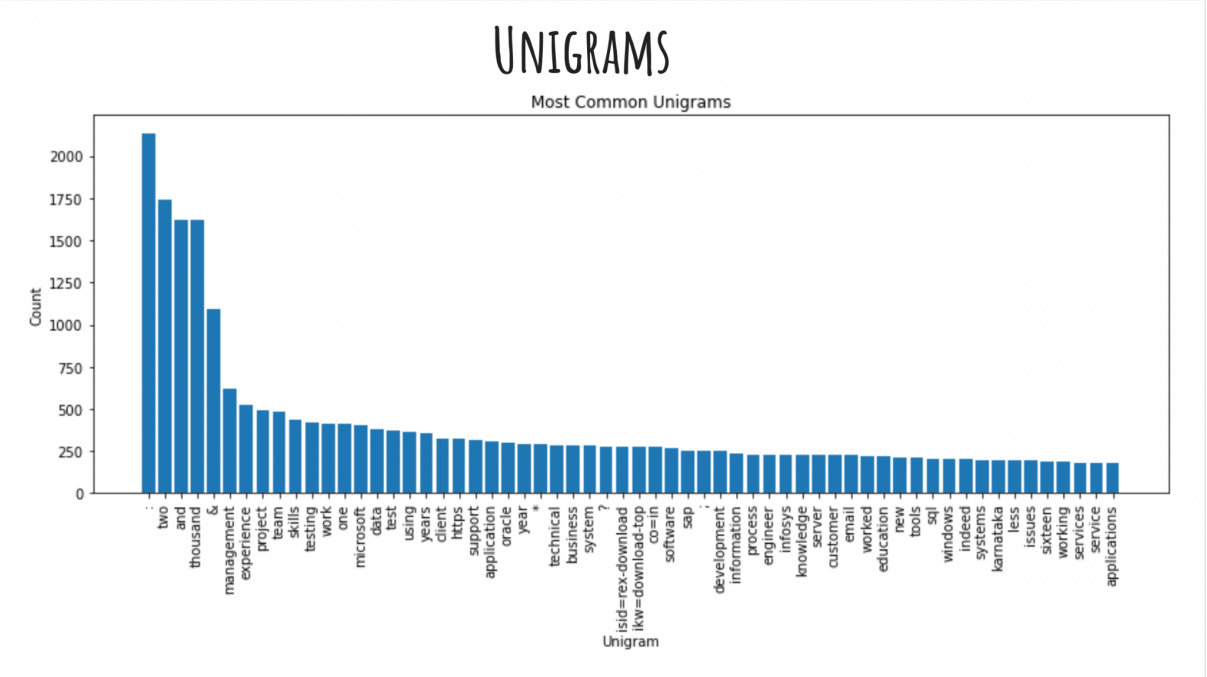
State	Workers in state
Karnataka	70
Andhra Pradesh	50
Maharashtra	49
Delhi	14
West Bengal	14
Chandigarh	6
Tamil Nadu	1

This slide shows the number of workers located in each state and a corresponding visualization on the Indian map.
It shows that most of the workers are from Karnataka.

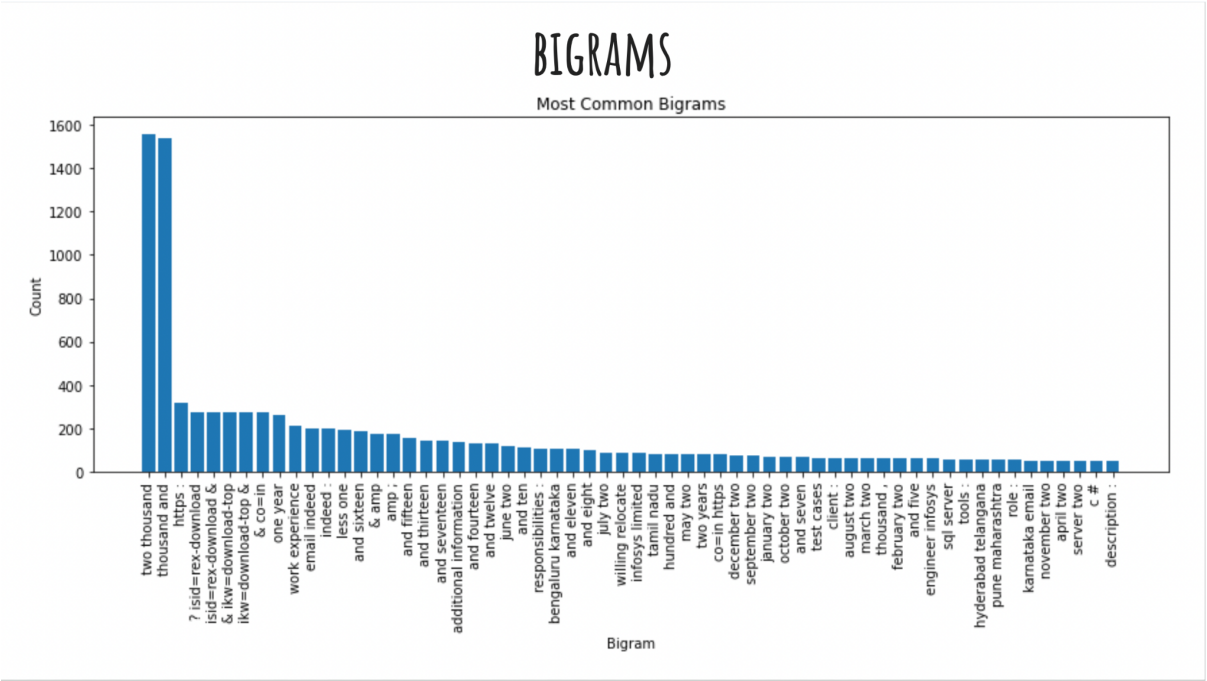
WORD CLOUD

A word cloud of the most common words appearing in the resumes.

Our observations emphasized that 'application', 'team', 'management', 'year', 'project', 'client', 'service' are most common words used in the resumes.

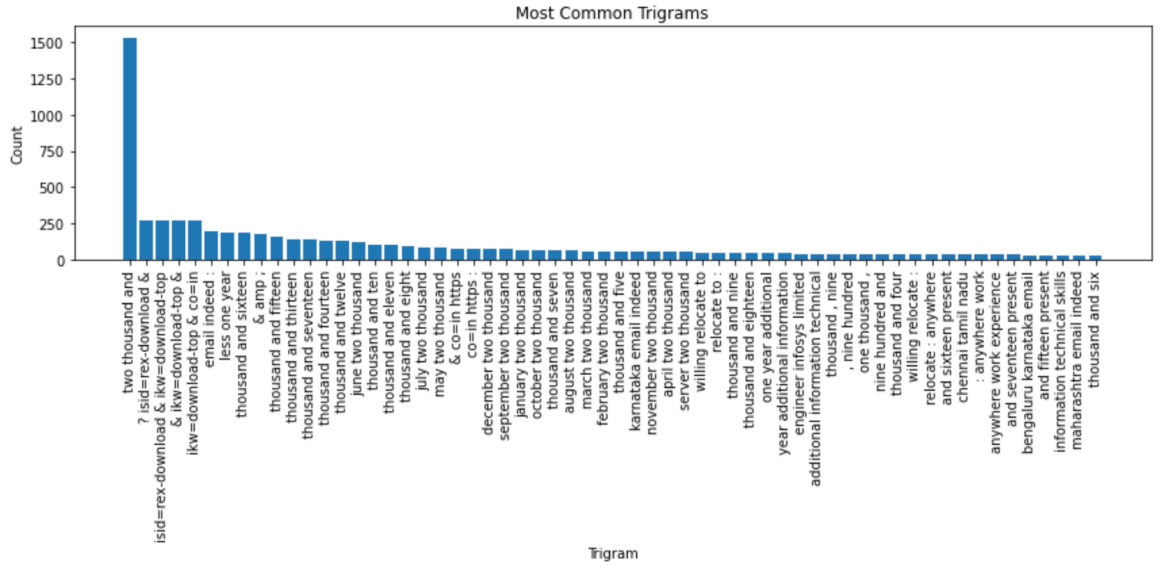


Bar plot illustrates that monst commonly used unigrams are ‘:’, ‘two’, ‘and’, ‘thousand’



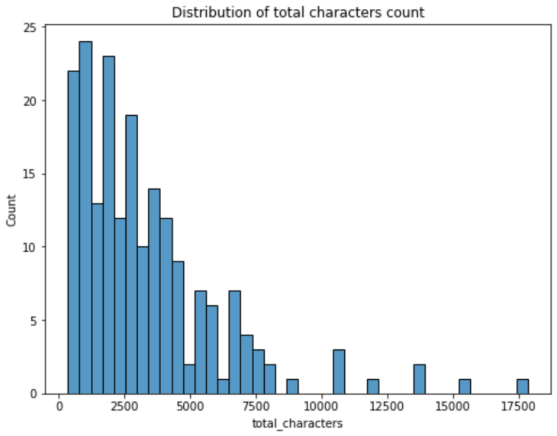
Bar plot illustrates that monst commonly used bigrams are, ‘two thousand’, ‘thousand and’

TRIGRAMS



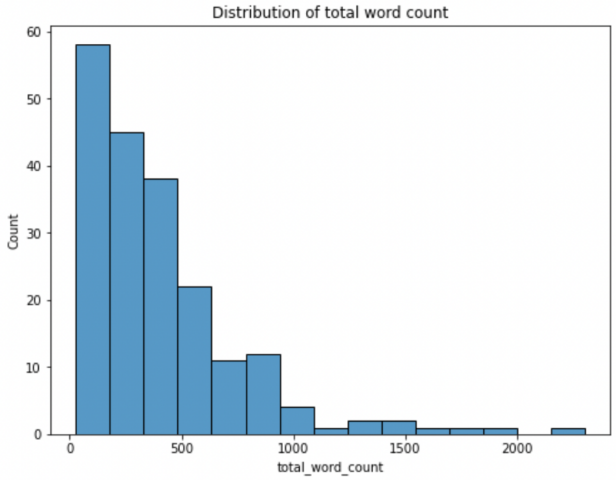
Bar plot illustrates that monst commonly used trigram is ‘two thousand and’

CHARACTER COUNT



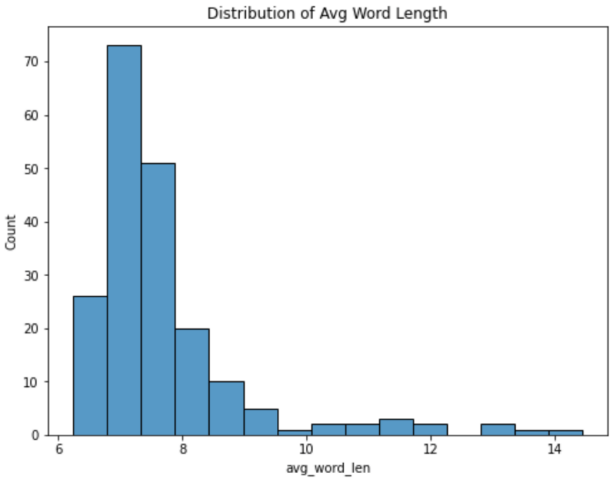
The average number of characters in the resumes is 3337 ± 2850 words. The resume with the fewest words is 337 while the highest is 17866 words.

Word count



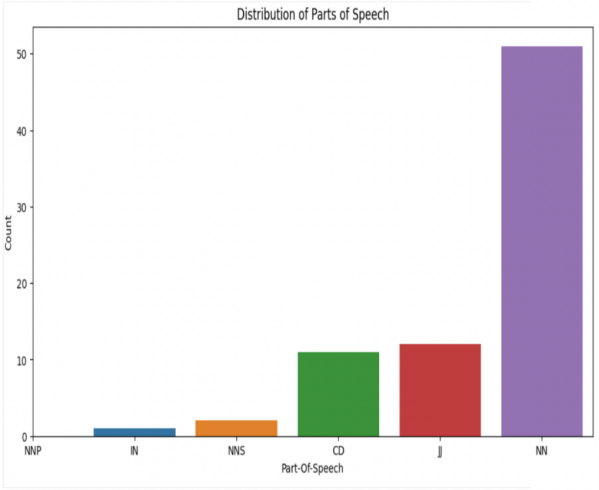
The average number of words in the resumes is 406 ± 362 words. The resume with the fewest words is 28 while the highest is 2801 words.

Average word length in resumes



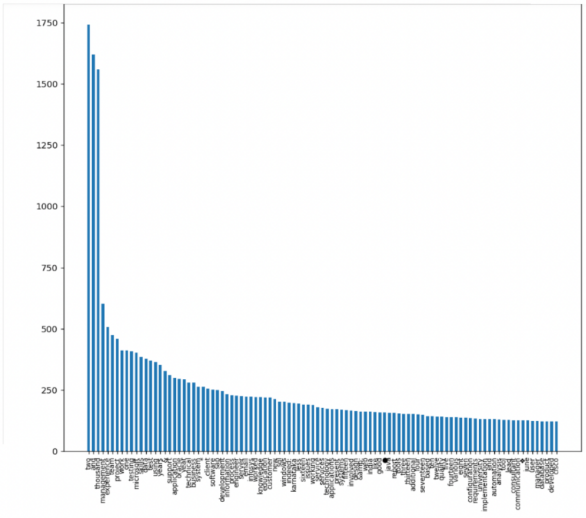
The average length of the words in the resumes is 7.70 ± 1.32 characters. The shortest average word length is 6.24 while the highest is 14.4 words.

Parts of speech analysis



When specifying the resume details, singular nouns dominate the description of each candidate. This showcases. This in turn can highlight the use of concrete details in most resumes to satisfy ATS format.

Frequency Distribution of 100 Most common words



1. "Experience" ranked as the fifth most commonly used word, emphasizing its importance in the job market over education.
2. Management experience is listed as the fourth most sought-after skill by employers.

3. Microsoft and Oracle are the most frequently mentioned companies on applicants' resumes.

FINAL RECOMMENDATION

Through the analysis of resume data, several key findings have emerged. The number of workers located in each state in India shows that Karnataka has the highest concentration of workers. Furthermore, the analysis of commonly used unigrams, bigrams, and trigrams reveals frequently occurring words and phrases in resumes, such as "two thousand" and "thousand and."

Additionally, statistical measurements provide insights into the length and composition of resumes. The average number of characters and words in resumes indicate an overall resume length of 3337 ± 2850 characters and 406 ± 362 words. The average length of words in resumes is approximately 7.70 ± 1.32 characters.

The analysis also highlights the prominence of singular nouns in describing candidate details, indicating the use of concrete information to meet ATS format requirements. Furthermore, the data shows that "experience" is a highly valued attribute in the job market, ranking higher than education. Management experience is identified as the fourth most sought-after skill by employers. Microsoft and Oracle emerge as the most frequently mentioned companies on applicants' resumes.

By leveraging these insights and implementing NLP techniques, it is expected that the project will contribute to improving the accuracy and efficiency of resume parsing, ultimately enhancing the job application process and benefiting both job seekers and HR professionals in the human resources industry.