

CV Parsing Using NLP

K. Bhavya Sai¹, G. Kavya Sree², S. Sai Soundarya³, C. Sai Pranathi⁴, Y. Durga Bhargavi⁵

B. Tech (IV-CSE) Students, Department of Computer Science and Engineering^{1,2,3,4}

Assistant Professor, Department of Computer Science and Engineering⁵

Ace Engineering College, Hyderabad, Telangana, India

bhavyasaikongara@gmail.com¹, kavyasreeg999@gmail.com², sistlasoundarya@gmail.com³,

pranathisai1711@gmail.com⁴, bhargavi.yarra2022@gmail.com⁵

Abstract: *The objective of the project is to make a CV parsing algorithm using NLP Processing. The algorithm will parse resumes one by one and can create a Candidate Profile supported the talents mentioned within the resume. A corpus is created which contains various skills required for a particular job profile for example: Machine Learning, Data Science etc. Word Embedding's are created from the corpus and these are used to match the talents within the candidate resume with the specified skills. Finally, the candidate profile is made and plotted as a bar chart for better visualization.*

Keywords: NLP Processing, Candidate Profile, Word Embedding's, Candidate Profile.

I. INTRODUCTION

Resume screening is the process of determining whether a candidate is qualified for a role based on his or her education, experience, and other information captured on their resume. It's a form of pattern matching between a job's requirements and the qualifications of a candidate based on their resume. The goal of screening resumes is to decide whether to move a candidate forward – usually onto an interview – or to reject them. Talent acquisition is an important, complex, and time-consuming function within Human Resources (HR). Not only is there a staggering one million people coming into the job market every month, but there is also huge turnover. Most challenging part is the lack of a standard structure and format for resume. Makes short listing of desired profiles for required roles very tedious and time-consuming. Recruiters must be able to properly screen resumes in order to hire the right individual at the right time. The process of deciding whether a candidate is qualified for a position based on his or her qualification, education, work-experience, and other information from their CV is known as resume screening. The importance of efficient and effective resume screening is at the heart of any strong recruitment strategy. The goal of resume screening is to find the best candidates for a position.

In this paper we present the idea of making Resume screening algorithm using NLP which develop a candidate profile that match the required skills and plotted as a bar chart for better visualization.

II. LITERATURE SURVEY

There have been several attempts to automate various aspects of the recruitment process. For example, suggests using techniques like collaborative filtering to recommend candidates matching a job. Describes a method that uses relevance models to bridge the vocabulary divide between job descriptions and resumes. In their method, related job descriptions are identified by matching a given candidate job description with a database of job descriptions. Then, resumes matching those job descriptions are used to capture vocabulary that is not explicitly mentioned in the job descriptions. These methods assume the existence of manually labelled relevance rankings. On the other hand, our method does not assume the presence of relevance rankings for training. In, collaborative filtering measures are combined with content based similarity measures for better ranking. However, most of these studies are performed on synthetic data and not on real world unstructured resumes and job descriptions.

A. NLP Based Extraction of Relevant Resume using Machine Learning:

This technique stated parsing of the resumes with least limit and the parser works the utilization of two or three rules which train the call and address .Scout bundles use the CV parser system for the determination of resumes. As resumes are in amazing arrangements and it has different sorts of real factors like set up and unstructured estimations, meta experiences, etc. The proposed CV parser approach gives the component extraction method from the moved CV's.

B. E-Recruitment System through Resume Parsing, Psychometric Test and Social Media Analysis:

It follows an approach of 4 stages, the first stage was to get the data (resume) and convert them into structured format and then perform the analysis using deep learning techniques. Second step includes the psychometric test where the text mining is used to generate scores for each candidate. In the third step they perform web scraping on various social media sites to get the additional information about the candidates and recommend suitable jobs to them. In the fourth step, the system will recommend the skills and requirements in which the students are lacking and also help them to get recruited in the desired company.

C. Combination of Neural Networks and Conditional Random Fields for Efficient Resume:

Parsing: The techniques used in this category are neural networks and CRF to segment and extract various information from resumes. CNN model is used for segmentation and compared with a Bi LSTM model. A CRF based model is chosen for information extraction and compared with a Bi-LSTM-CNN model. They segmented and extracted several pieces of information from personal, educational and occupational blocks. The results are promising and the output JSON file contains 23 data fields.

D. CV Parser Model using Entity Extraction Process and Big Data Tools:

Here the problem definition was based on designing an automated resume parser system, which will parse the uploaded resume according to the job profile. And it will transform the unstructured resumes into structured format. It will also maintains a ranking system on the resumes. Ranking will depend on the basis of information extracted i.e. according to technical skills, education etc. Here the CV parser is used. CV parsing is such a technique for collecting CV's. CV parser supports multiple languages, Semantic mapping for skills, job boards, recruiter, and ease of customization. Parsing with hire ability provides us accurate results.

III. EXISTING SYSTEM

Existing system is manual screening Manual screening is a much lengthier process than using screening software, as it involves the recruiter reviewing each resume on their own. The recruiters have to screen each and every resume manually and have to find the correct candidates for the required job profile. This may take a long time to proceed and the recruiters may also have a lot of confusion to shortlist the candidates according to the skillset requirement which takes a long time to process.

IV. PROPOSED SYSTEM

To create CV parsing system that produces a candidate profile of the similar skills matched in the corpus. The goal of this project is to create an algorithm that provides the highly accurate results to compare the skills and the CV of the candidate. In this approach we create bigrams of frequently occurred words in the skills corpus and use the algorithm Word2Vec to find the Word Embedding's and create a model to parse the CV's of the candidates. We finally create a bar graph of the skills of the candidates for better visualization.

V. WORKFLOW

1. **Skill Set Input:** The skill set is a document/corpus that defines the skills required for particular job profile e.g. Machine learning, Data Science etc.
2. **Preprocessing:** Preprocessing is the stage where the skills document is cleaned by removing the unnecessary details, wide spaces, punctuations from the provided skills document.
3. **Creating Bigrams:** Bigrams are the words that occur together almost all the time. Example: Machine Learning etc. These are created using Phrases.
4. **Creating Word Embedding's:** The Genism library provides a simple API to the Google word2vec algorithm which is used to create Word embedding's. Word embedding's is one of the most popular representation of document vocabulary. It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc.

5. **Extracting Resumes:** Resumes are stored in a folder and extracted one by one using PyPDF library and returned as a sequence of string. The string is pre-processed and is further processed to create candidate profiles.
6. **Candidate profile generation:** Spacy's Phrase matcher is used to match the array (obtained from word2vec) with the extracted text and the candidate profile is generated and visualised as a graph.

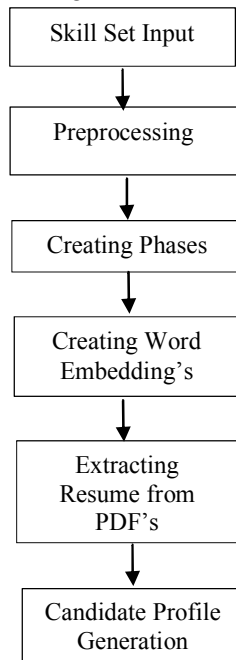


Figure 1: Workflow

VI. RESULT

As mentioned in the above procedure first we take the input of the skills and the CV as input in the form of a PDF document. The Following are the results

6.1 Input

Sample CV document screen shot:

Academic Qualifications:				
Course	Institution	University/ Board	Year of passing	Percentage / CGPA
B.Tech (CSE)	ACE Engineering College	JNTUH	2018-2022	79.5
Intermediate	Sri Gayathri Junior College	Telangana Board of Intermediate Education	2016-2018	95.6
SSC	PACE School	Telangana Board of secondary Education	2016	92

Technical Skills:

- **Languages:** C, C++, java Basic, Python
- **Database:** MySQL
- **Web Designing:** HTML, PHP
- Microsoft EXCEL ,Microsoft Word

Workshops Attended:

- Certified Microsoft Office Specialist for Office Word 2016
- Certified Microsoft Office Specialist for Office Excel ® 2016
- Participated in National level C Programming Contest (Formula C) conducted by ACE Engineering College during 30 – 31 August 2019.
- Participated in National level Python Programming League (PPL) conducted by ACE Engineering College during 30 – 31 August 2019.
- Participated in a two day workshop on "Future on Drone Technology and Its Applications" during 14 – 15 February 2020 conducted by ACE Engineering College in association with ARIES Solutions Pvt. Ltd. Hyderabad.
- Completed the Boot Camp ENLITE 2020 On Innovation and Entrepreneurship held

Figure 2: Sample Input Screen Shot



6.2 Sample Output:

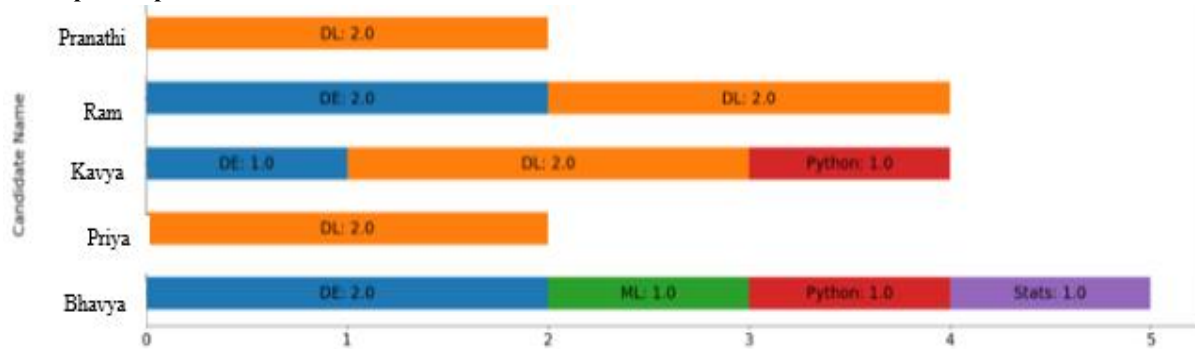


Figure 3: Sample Output

VII. CONCLUSION

This CV Parsing technique aims to be used for screening candidate profiles, to reduce an organization’s time-to-hire. CV Parsing is the process of providing a bar graph view of the candidate skills set so that it could be easy for the recruiter for recruiting process of candidates.

ACKNOWLEDGMENT

We would like to thanks to our guide Mrs.Y.Durga Bhargavi and Mrs. S.Kavitha for their continuous support and guidance. Due to his guidance, we can complete our project successfully. Also, we are extremely grateful to Dr. M. V. VIJAYA SARADHI, Head of the Department of Computer Science and Engineering, Ace Engineering College for his support and invaluable time

REFERENCES

- [1]. <https://course.spacy.io/en/chapter1>
- [2]. <https://stackabuse.com/python-for-nlp-vocabulary-and-phrase-matching-with-spacy/>
- [3]. <https://towardsdatascience.com/understanding-nlp-word-embeddings-text-vectorization-1a23744f7223>
- [4]. <https://radimrehurek.com/gensim/models/phrases.html>
- [5]. <https://machinelearningmastery.com/develop-word-embeddings-python-gensim/>
- [6]. <https://www.shanelynn.ie/word-embeddings-in-python-with-spacy-and-gensim/>