International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 03 | Mar 2022 www.irjet.net p-ISSN: 2395-0072

Resume Parser with Natural Language Processing

Abdul Wahab.S[1], Dr.M.N. Nachappa[2]

[1]PG student, Department of Computer Application, JAIN University, Karnataka, India
[2]Head of Department, Department of Computer Application, JAIN University, Karnataka, India

---

Abstract - *As a result of the online recruiting system's progress. Candidates can effortlessly upload their resumes to the job application website. As a result, a significant number of people are impacted. There are resumes being mailed in. The human resource department has suffered as a result. Recruiting new employees is a difficult task for the department, as is filtering through a large number of candidates. Additionally, candidates that submit resumes come in a variety of shapes and sizes. Fonts, font sizes, colours and other kinds of writing. For example, human resource departments are in a difficult condition. in analyzing all of the resumes sent by candidates and selecting the best candidate for the job. As a result, in this project, I propose resuming the parser using natural language. processing in order to assist the human resources department or recruiter in gathering precise information from the resume Natural language processing is used to scan resume content, locate keywords, and group them into sectors based on their relevance. based on their keywords, and ultimately, depending on keyword matching, give the most appropriate CV to the company. The user must first create an account on the website and upload a resume. The resume parser collects all necessary information from the resume and automatically fills out a form for the user to proofread. The resume is saved in our NoSQL database after the user confirms, and it is ready to be seen by employers. In addition, the user receives a JSON and PDF version of their resume.*

***Key Words*: online recruiting system, mail, best candidate, Scan resume, NoSQL database**

1.INTRODUCTION

Corporate businesses and recruitment agencies review a big number of resumes every day. This isn't a job for humans. There is a need for an intelligent automated system that can extract all of the key information from unstructured resumes and convert them all to a similar structured format that can then be ranked for a specific job position. Names, email addresses, social media accounts, personal websites, years of work experience, employment experience, years of education, education experiences, publications, credentials, volunteer experiences, and years of service keywords and the CV cluster (ex: computer science, human resources, etc.) are among the information parsed. The parsed data is subsequently saved in a database (in this case, NoSQL) for future use. In contrast to other types of unstructured data (ex: email bodies, web page contents, etc.), those are the most common file types utilized by job applicants. As a result, to extract all of the information from unstructured resumes and a range of data sources, an automated intelligent system based on natural language processing is necessary. Converting all resumes to a comparable structured format and selecting only the information that is relevant to screening, such as name, job, education, years of experience, work experience, certificates, email, phone number, and so on, is the process for parsing resumes. Following that, the structured resume data will be parsed and saved in a database for future use. Each set comprises information on the person's contact information, employment experience, and educational background. Despite this, resumes might be difficult to decipher. This is due to the fact that the kinds of information, the sequence in which they are presented, and the literary style in which they are written differently. They can also be written in a variety of formats. ".txt", ".pdf", ".doc", ".docx", ".odt", ".rtf", and so on are some of the most frequent. The model cannot rely on the order or kind of data to effectively and efficiently extract data from various types of resumes.

2. LITERATURE REVIEW

Using Text Processing as a Resume Analyzer This genuine review outlines an excellent Company Recommender System that employs text mining and machine learning techniques to assist recruiters in selecting the best candidate for a given position. Candidates' resumes are sorted according to the company's needs when they upload them. The organization can utilizes the rating to choose the best candidates. The methods and model for this post will be given in four steps: gathering resumes and searching for keywords in the resume text's information base. Then, based on a rating score, candidates are ranked and categorized. In addition, this system may extract new keywords from resumes in order to broaden the knowledge base. In the IT recruitment process, information from Polish resume documents is extracted automatically. This review examines and discusses automated information retrieval for the recruitment process in the IT business. The suggested solution uses a multi-module system to deal with low-resource language dictionaries and intricate linguistic linkages in Polish. Entity recognition is the most useful method for assessing CVs, and it is used in this research. It's a semi-semantic text analysis that only recognizes particular terms. It's an essential phase in getting the text's information content ready for processing.

This project's data sets are separated into two categories. The first is a GitHub dataset of 200 resumes that include names (first and last name) and positions to apply for. Other databases include global university and skills, for example.

Table -1: Number of datasets for each entity.

Entity	Number of data
Name	205
Designation	473
University	829
Skills	1,249

This research uses Named Entity Recognition, a branch of Natural Language Processing that analyses enormous amounts of unstructured human speech.
NER extraction is the first stage in extracting information and topic modelling. The algorithm examines the paragraph in its entirety and highlights the text's most important entity elements. You can use Stanford NER or Spacy for this project because the resume text is an unstructured text into predetermined categories.
Regular expressions, as well as regular expressions in scripts, were utilized in this project. A regular expression is a string of special characters that represents a search pattern by matching a character pattern to the search string.
A. Text conversion from PDF and DOC files
This project uses the PyMu PDF library to convert PDF files to text format, and the python-docx library to convert Doc, Docx files to text format.
B. Recognizing Named Entities (NER) Getting a name (both first and last name) and a designation. This project's train dataset is in the PKL (Pickle) format. Pickle is a Python module that serializes objects so that they can be saved to a file and then reloaded when the program calls them. Then, for the training model, Named Entity Recognition (NER) is employed because the purpose of this project is to use a tagged dataset to locate and classify unstructured resume material into specified categories.
C. Regular Expressions We can extract the name of the university by using regular expressions to look for terms like University, School, College, Institute, and so on in university names. Then look for all the characters that are in the vicinity of those keywords. Obtaining a degree or educational background by using regular expressions to look for keywords in university names such as Bachelor of, Master of, Doctor of, Degree, and so on. Then look for all the characters that are in the vicinity of those keywords. The ability to get information out of a situation. The first step is to clean the data by removing stop words and punctuation from the text. Stop words are a group of phrases that are often used in a language but contain little relevant information.
entire resume text. Then, in the talents database, look for each token (.csv file).
Resume processing is also constrained by ethical concerns. As a result of this approach, the only input will be text. As a result, this method is only appropriate for screening certain positions. A graphic designer job or any design job that requires a visual sample of the work, an image as evidence of work, and examination of the resume's beauty and colour. For example, may not be suited for this approach. The bias of this system appears to be leading businesses to lose staff.

3. CONCLUSION

Due to the advancement of the internet recruiting system, a considerable number of resumes were submitted. As a result, the human resource department or company faces a hurdle in hiring new personnel and assessing a huge quantity of applicants. As a result, by utilizing an automated intelligent system based on natural language processing, this technology has aided employers. This system can successfully convert many resume formats to text format and retrieve certain key information. The proportion of similarity between the applicant's résumé and the job description can also be determined by comparing the two. This approach can help the human resources department or the company review resumes prior to conducting interviews and choosing the best candidate for the job. We were able to convert several resume formats to text and extract crucial information from there. We were also able to harvest terms from several social networking sites, such as Stack Overflow and LinkedIn, and detect similarities between them, allowing us to define the resume's genre (e.g: Computer science, Management, Sales, human resource, etc). Future work will include grading resumes and analyzing information about candidates obtained from social networking sites such as Facebook and Twitter in order to make more accurate and authentic decisions about whether or not to offer the candidate a job. Our strategy is to make employers' and candidates' jobs easier and more efficient. Our main goal is to make the hiring process easier. The approach will offer the companies with high-quality applications. The process's unfair and discriminatory practices will be curtailed. The resumes will be sorted in order based on the information provided in the form of technical skills.

REFERENCES

[1] What is resume parsing: Retrieved from, https://www.smartrecruiters.com/resources/glossary/resume- parsing/

[2] NLP Based Resume Parser using BERT in Python: Retrieved from, https://www.pragnakalp.com/case-study/nlp-resume-parser-bert- python/

[3] NLP based resume parser in Python (Beta): Retrieved from, https://demos.pragnakalp.com/resume-parser/.

[4] Literature Reviews - Automated extraction of information from Polish resume documents in the IT recruitment process: Retrieved from, https://www.sciencedirect.com/science/article/pii/S18770509 .

[5] World University Ranking 2016: Retrieved from, https://data.world/hhaveliw/world-university-ranking-2016?fbclid=IwAR01WBDbntwc7K3NRkHpc1XCp8WcESQEV MR2zXCXD8R31f-NTwJv1DZ7mWY