e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705058 |

A Framework for Text Categorization using NLP

Rutuja Shridhar Bachhav, Dr. Monika Deshmukh

M.Tech(CTIS) Student, Dept. of Computer Science and Engineering, Sandip University, Mahiravani, Nashik, Maharashtra, India

Assistant Professor, Dept. of Computer Science and Engineering, Sandip University, Mahiravani, Nashik, Maharashtra, India

ABSTRACT: Recruitment in the IT sector has been on increased in recent times. Software companies are on the hunt to recruit raw talent right from the colleges through job fairs. And through job fairs, companies receive thousands of resumes from job seekers. It is difficult to identify the good match between the qualifications of the candidate and match the skill that a company seeks by examining each resume. The Resume Classifier tries to find the resumes for any job/university interview more robust by doing information extraction approach based on the data of previously selected and rejected candidates. The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, stemming and filtering the content of the data. By using Phrase Matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter.

KEYWORDS: Text Categorization, E-Recruitment, Word / Sentences Similarity, Natural Language Processing, Phrase Matcher, Preprocessing, Word Order Similarity between Sentences, Resume Score Calculation.

I. INTRODUCTION

The recruitment in the Information Technology sector has seen an exponential increase in recent times. Companies recruit thousands of young talents, right from the college every year through campus fairs it is difficult to identify the good match be- tween the qualifications of the candidate and match the skill that a company seeks by examining each resume for HR department in any organization. To address this challenge, many companies have shifted for using e-recruiting platforms. These plat- forms reduce the cost, time and effort required for manually processing and screening applicant resumes. These systems employ different methods and approaches to address the challenges associated with screening, matching, and classifying candidate resumes. The different that made to make the process easy Although these approaches produce high precision ratios in finding candidates to fill a vacancy, they give less attention to the run time complexity of the matching process i.e. every job offer will be matched with every resume in the corpus instead of matching resumes that are only related to their occupational category. The proposed resume Classifier tries to find the resumes for any job/university interview more robust by doing information extraction approach based on the data of previously selected and rejected candidates. The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, lemmatizing and filtering the content of the data. By using phrase matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter.

.

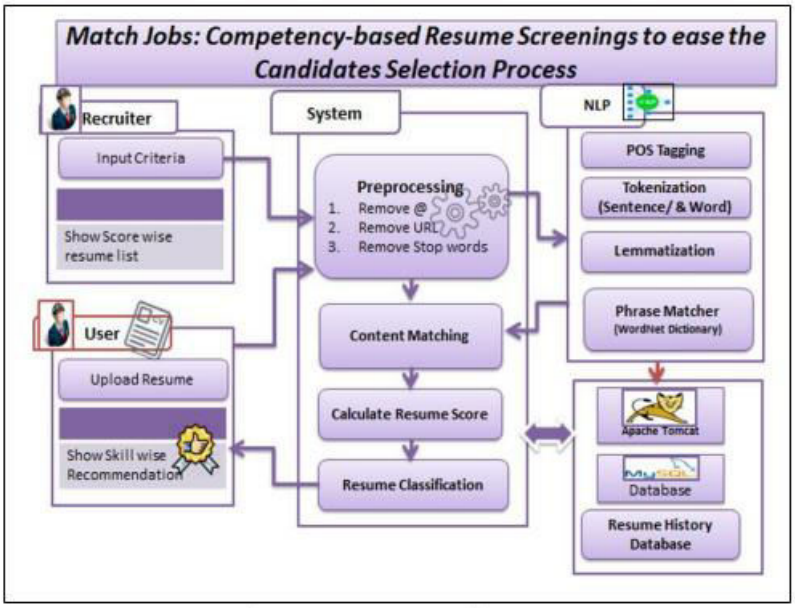
II. RELATED WORK

The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, stemming and filtering the content of the data. By using Phrase Matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705058 |



Upload Resume: User can upload the resume to the system. The system preprocesses the input to Remove@, Remove URL Remove Stop words to get the fine data and extracts the tag word from the data. The resume should have the skill mentioned.

Preprocessing: The system can Understand Each Word from all the resumes using Natural Language Processing (NLP). It can apply different techniques. for understanding the sentence and word. They can analyze the words using two different ways like,

Sentence & Word Understanding

1. Sentence Tokenization

In this techniques system can divides the sentence into several tokens. It split the large raw text into several sentence to get more meaningful information out. For e.g.

"All work and no play make jack a dull boy, all work and no play". The above sentence is divided into sentence like,

2. POS Tagging

This algorithm is used for detects if the word token is noun, verb, adjective. • POS Tagging in which a word is assigned in accordance with its syntactic functions. In English the main parts of speech are noun, pronoun, adjective, determiner, verb, preposition, adverb, conjunction, and interjection.

3. Word Tokenization

This technique the sentence or data can split into several words. For e.g." All work and no play make jack a dull boy, all work and no play". This sentence split into word like,

4. Word Lemmatization

Lemmatization is a more methodical way of converting all the grammatical/inflected forms of the root of the word. Lemmatization uses context and part of speech to determine the inflected form of the word and applies different normalization rules for each part of speech to get the root word (lemma).

5. Word Similarity

By using this technique, the system can find the similar words. We use the WordNet dictionary for finding the synonyms. WordNet Dictionary-

Rule			Example		
SSES	→	SS	caresses	→	caress
IES	→	I	ponies	→	poni
SS	→	SS	caress	→	caress
S	→		cats	→	cat

WordNet is a combination of dictionary and thesaurus. It groups English words into sets of synonyms called

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705058 |

synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members.

6. Sentence Similarity

By using this technique, the system can find the similar sentence.

III. PROPOSED WORK

The proposed resume Classifier tries to find the resumes for any job/university interview more robust by doing information extraction approach based on the data of previously selected and rejected candidates. The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, lemmatizing and filtering the content of the data. By using phrase matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter. The recruitment in the Information Technology sector has seen an exponential increase in recent times. Companies recruit thousands of young talents, right from the college every year through campus fairs it is difficult to identify the good match be- tween the qualifications of the candidate and match the skill that a company seeks by examining each resume for HR department in any organization. To address this challenge, many companies have shifted for using e-recruiting platforms. These plat- forms reduce the cost, time and effort required for manually processing and screening applicant resumes. These systems employ different methods and approaches to address the challenges associated with screening, matching, and classifying candidate resumes. The different that made to make the process easy Although these approaches produce high precision ratios in finding candidates to fill a vacancy, they give less attention to the run time complexity of the matching process i.e. every job offer will be matched with every resume in the corpus instead of matching resumes that are only related to their occupational category. The proposed resume Classifier tries to find the resumes for any job/university interview more robust by doing information extraction approach based on the data of previously selected and rejected candidates. The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, lemmatizing and filtering the content of the data. By using phrase matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter.

IV. RESEARCH METHODOLOGY

The Resume Classifier tries to find the resumes for any job/university interview more robust by doing information extraction approach based on the data of previously selected and rejected candidates. Software project estimation is form of problem solving here. The complex software is hard to estimate hence it is divided into smaller pieces. The estimation of project will be correct only when the estimation of size of the project is correct. In the context of project planning size refers to quant able outcome of project. Here, the direct approach is selected and hence, the size is estimated in Line of Codes. The feasibility study comprise of an initial investigation into personnel will be required. Feasibility study will enable us to make informed and straightforward choice at crucial points while developing phase. All projects are feasible given unlimited times and resources. But, the development of computer-based system is more likely to be plagued to scarcity of resources. It is both essential and prudent to evaluate the feasibility of project at earliest possible time. The System extracts the information from the resume. Then Natural language processing (NLP) technologies are used for parsing, tokenizing, stemming and filtering the content of the data. By using Phrase Matcher, we can calculate the score of the particular resume based on the recruiter information and suggest lacking skills to the users and recommend top resume to recruiter.

V. RESULTs

- Firstly, download the eclipse Luna tar file from the www.eclipse.org Website.
- Then extract tar setup of eclipse Luna to run the application.
- Open the extracted folder of eclipse Luna to run the application.
- Click on the icon of eclipse and run the application.
- Register Login to the system.
- When you logged in as an Admin, you will be having access to manage the users in the system.
- When you logged in as a User, you will be having access to write and compile the code in different programming languages.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705058 |

Test Case ID	Test Case Name	Test Steps			Test Case Status (P/F)	Test Priority	Defect severity
		Description	Expected	Actual			
1	Upload Resume	User Login to the system first	Student can upload the resume to the system	Same as Expected	Pass	High	Low
2	Apply Pre-processing	Data Pre-processing is applied to clean the data	Word Understanding using Natural Language Processing (NLP)	Find the similar sentences	Pass	High	High
3	Resume Score calculation	Apply TF-IDF Technique	Score of the resume is calculated	Classify the resume	Pass	High	High
4	Resume Recommendation	Calculate Score for each and every resume	Display all the Shortlisted Resume	Same as Expected	Pass	High	High
5	Skill based Recommendation	Based on the rejection recruiter gives the suggestion	User can improve their skills according to the market conditions	Same as Expected	Pass	High	High

Table 1: Test Cases

VI. CONCLUSION

Due to the constant growth in online recruitment, job portals are starting to receive thousands of resumes. It is difficult to identify the good match between the qualifications of the candidate and match the skill that a company seeks by examining each resume manually. The proposed system helps in classifying resumes by word/ sentences understanding. The system extracts information from resume, by using natural language processing system understands the meaning of the data. Resume score is calculated by using phrase matcher based on resume information and classify the resume. The system also helps to reduce human workload for resume classification providing quick process of sorting resumes.

REFERENCES

1. Vishnunarayanan.R1, Shreekrishna Prasad2, Krishnan.A. N3, Palanivel.S4, Uma- makeswari. A, “Resume classification using analytic hierarchy process and keyword extraction”, International Journal of Pure and Applied Mathematics Volume 115 No.7.
2. Pradeep Kumar Roya, Sarabjeet Singh Chowdharyb, Rocky Bhatia “A Machine Learning approach for automation of Resume Recommendation system”, International Conference on Computational Intelligence and Data Science (IC- CIDS 2019).
3. Luiza Sayfullina, Eric Malmi, Alexander Jung, Yiping Liao,” Domain Adaptation for Resume Classification Using Convolutional Neural Networks”, July 2017.
4. Ankita Satish Vaidya and Pooja Vasant Sawant (2015), “Resume Analyzer an Automated Solution to Recruitment Process”, International Journal of Engineering and Technical Research (IJETR), Volume-3, Issue-8.
5. Maryam Fazel-Zarandi1, Mark S. Fox2, “Semantic Matchmaking for Job Recruitment: An Ontology-Based

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705058 |

Hybrid Approach", at: https://www.researchgate.net/publication/265922198.

6. Renuka S. Anami, Gauri R. Raoo," Automated Profile Extraction and Classification with Stanford Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-4 Issue-7, December 2014.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com