EasyChair Preprint

№ 10088

Review on Resume Analyzer

Kritika Joshi, Manvi Jain, Ritika Ranjan and Reet Prasad

May 12, 2023

Review on Resume Analyzer

Prof. Kritika Joshi

Dept. Of computer Engineering DITU,

DIT University Dehradun, India

kritika.joshi@dituniversity.edu.in

Manvi Jain

Dept. Of computer Engineering DITU,

DIT University Dehradun, India

1000011961@dit.edu.in

Ritika Ranjan

Dept. Of computer Engineering DITU,

DIT University Dehradun, India

1000012803@dit.edu.in

Reet Prasad

Dept. Of computer Engineering DITU,

DIT University Dehradun, India

1000013961@dit.edu.in

Abstract—Resume Analyzer is an NLP template that can extract information such as skills, university, degree, name, phone, job title, email, other social media links, nationality , etc. whatever their structure and calculate the resume score according to the job requirements and skillset. To create such an NLP model capable of extracting various information from the CV, we need to train it on an appropriate dataset. We all know that creating datasets is difficult if we label manually. In order to reduce the time required to create the dataset, we used various techniques and libraries in Python, which helped us to identify the required information from the CV.

Keywords :- CV classification, Resume score, NLP

Introduction

In Recruitment, the best candidates are hired from a wide range of applicants with different skills relevant to a particular job profile. Companies with a job opening receives many emails from different candidates each day. Recently, more than thousands of e-recruitment sites had developed by different developer , using various methods to identify job seekers for the required job application company. The basic purpose of these sites is to show candidates the results for which category they best match by summarizing their resume/resume based on the keywords used there.

A data-driven approach automates resume processing, giving recruiters more

time to screen promising candidates. It uses ML and NLP to fetch the necessary data and calculate ranking scores. The authors created a resume analysis and reporting tool using ML and NLP.

NLP is a part of computer science and artificial intelligence that deals which helps us to interaction between computers and human language. Natural language is the language which humans used to communicate with each other.
NLP can be defined as a group of methods to fech grammatical structure and decode the input provided to perform useful tasks.

Therefore, natural language is generated by the rules of target language. NLP is useful in the areas of learning system-assisted instructions and DB interfaces because it provides opportunities for increased communication between systems .

Literature Review

NLP

NLP is a part of computer science and artificial intelligence that deals which helps us to interaction between computers and human language. It defines the processes of interaction between computers and human language.

Lexical analysis - The experimental phase of the compiler. The modified source code is language posessor provides source code written in sentence form. The parser deletes the comment and spaces from the given code and splits that grammar into strings of tokens.
Syntactic Analysis - It describes the structure of the data. The module consists of hierarchies of expression, the smaller one being primitive symbols and the larger one being sentences. The module can be visualized as a tress comprising of nodes which represents expression. Primitive symbols can represented by the values in the nodes. Root words represent sentences.
Semantic Analysis - Semantic analysis can be defined as the study of the semantics, structure and meaning of speech. This process links syntactic structures to the overall writing hierarchy at the clause, sentence, paragraph, and sentence levels. This has to do with their language independent meaning.[1]

Text Mining

It is the process of extracting data from the text, it is the same as text analysis, which can be defined as the extraction of high quality necessary information from the text. 4244 Good source of information is often derived 4244 from pattern and trend formulas by methods such as statistical 4244 model training. Based on Horspool and Karp-Rabin algorithms, the enhanced unique keyword pattern matching algorithm. Karp-Rabin's algorithm is based on hash method instead of comparing characters, which is also the advantage of algorithm.the week point of this takes a long time when there is has a long pattern.[7]

K-Means Clustering
1) K is an unsupervised machine learning algorithm.
2) K-means divides the given data into k groups.
where k is the number of clusters.
3) The clustering algorithm regroups the scores into k
Groups.[6]

Methods and approaches

Raw data is transformed into an understandable format using data preprocessing which is a data mining process.

Data Cleaning: Processes, like removing noisy data and resolving all the inconsistencies, cleanses the data.
Data Integration: Statistics together with numerous representations are clustered together and the clashes between the information are looked after.
Data Transformation: Data are classified, combined, and accepted.
Data Reduction: The purpose of data reduction is to contract structure in the data warehouse.
Data Discretization: In this process, the product is divided by the variable, reducing the number of values of the non-interfering properties.
Tokenization: Tokenization is the function of decoding sequential tokens and detailed information. It removes with some characters, such as punctuation, cuts off the remaining units called characters.
Stemming: Stem formation is the process of reducing changes and generating root word from different similar words which have the same meetings.It is desirable to map the reference form to the same stem, even if the root is not valid.
Parts of speech tagging: In corpus linguistics, the process of labeling text (body) whose definition and meaning (such as its relation to adjacent words in a sentence, sentence, or paragraph) appear to be part of speech is called part (part) of speech labeling. speech tagging or POST). It is also referred to as part of grammatical or speech confusion.

Chunking: : Fragmentation, also known as shallow parsing, is the precision of speech and sentences. A noun, verb, adjective, etc. We can use part of speech to say whether it is or not, but we cannot get information from it about the sentence in the sentence.[1]

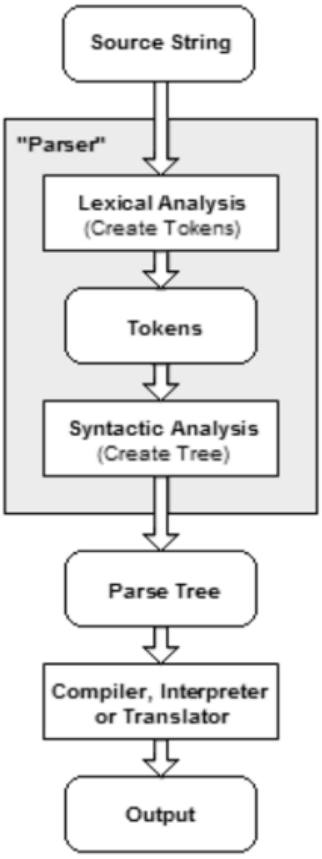
NLP

NLP is a part of computer science and artificial intelligence that deals which helps us to interaction between computers and human language. It defines the processes of interaction between computers and human language.

- **LEXICAL ANALYSIS** The experimental phase of the compiler. The modified source code is language posessor provides source code written in sentence form. The parser splits the grammar into chains of tokens, removing words or spaces from the source code.

- **SYNTACTIC ANALYSIS** Syntactic analysis determines the structure of knowledge. Architecture consists of a hierarchy of sentences, the smallest being the historical symbol and the largest sentence. We can view the architecture as a tree whose nodes represent sentences. Values stored in nodes represent significant characters. The root word represents the sentence.For example - parse tree
- **SEMANTIC ANALYSIS :** semantic analysis can be defined as the study of the

semantics, structure and meaning of speech. This system links the syntactic structure to the entire literary hierarchy consisting of clauses, clauses, clauses, and clauses. This is about their simple meaning in the language.[3]



K-Means Clustering

Following steps are involved in grouping:
Let A = {a1,a2,a3,……..,an} be the set of data points, B = {b1,b2,…….,bc} be the center.
1) Choose group position "c" randomly.
2) Evaluate the distance between all data points from group points.
3) Put the point of data in the center where the smallest distance is between all the cluster centers.
4) Calculate the center of the new cluster again using data points in the cluster
5) Calculate the length between every point of data again and the newly calculated center of the cluster.
6) If no content is edited, stop it, otherwise go back from step

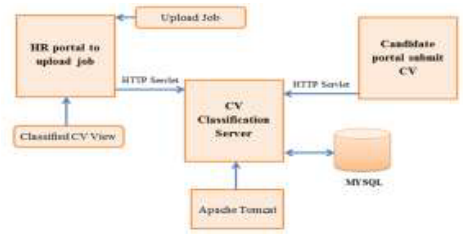
System Structure
The application process is based on the client server architecture of the system and the client will use the java servlet for communication. The front end of the system will use java swing components to build and the back end of the system will use java to build. The database system will contain relational MYSQL databases. As shown in the picture above, the system will have the following three modules:
1) HR / Manager Module
2) Candidates Module
3) Web Server

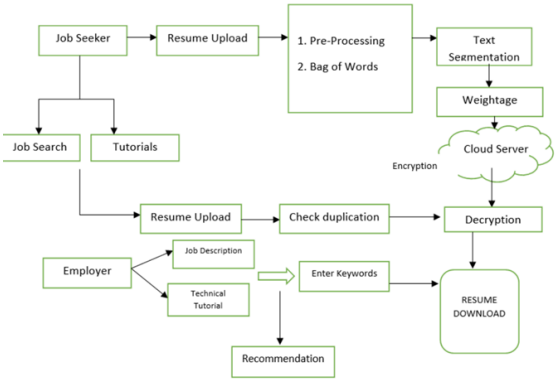
HR/Admin Module:
This module works on the HR or Admin side. This will be the desktop software and its front end will be built using java swing components. The module will be well protected with a username and password, the password will be protected with using the SHA-1 algorithm. Administrators can edit task and assign the weight and submit the task as needed. Manager
will get approval for all resumes submitted by candidate, where k-means clustering algorithm will be used.
Candidate module:
This module will be run by the candidate. This will be a desktop software whose front end will be built using java swing components. Candidates need to register in the system first and then they will see the interface and there will answer and choose different ways where

will be part of the candidate's resume. Candidates who send CV will learn their CV scores.

Web Server:

web server will be created using java servlet and will run using Apache Tomcat application server. The Web server will be responsible for all request and response operations on the system. It will manage the file and provide the necessary information to the administrator.[6]





The proposed architecture uses NLP to make candidate and employer roles easy and fast. The main modules on the candidate side are eliminating duplications in different resumes, score analysis of resumes using NLP, job listing and tutorials, selection of the candidate, and scheduling the interview based on the resume scores from multiple resumes, when the required skills are matched with the job profile then that will automatically recommend to the candidate.

Employers can register and log in to Portal to recruit qualified job seekers. It uses collaborative filtering to make an accurate list of candidates for this alert and the requested process.

In the existing system a job seeker uploads the same resume for n no times for various job postings which increases the cost of encryption and storage

To improve storage, we use deduplication technology to avoid storing the same data multiple times. For this technology, we use the **Proactive Copy Control Method (PRCR) algorithm** to detect duplication of stored data. The use of the natural processing technique is also advantageous among job seekers and employers

saves time. At work, web browsers is used to fetch job descriptions and requirements.

on the applicant side, after the return is sent, stop the message filter and the segmentation will be done.

After a division of words, points are given according to education, job, skill, good character and degree frequency. Finally, our proposal made recommendations for the next generation of that should work for the next level of education. We use the **AES encryption algorithm** to protect the personal data of job seekers and protect themselves. For research and analyzing, we use open cloud storage to make data accesible worldwide.

CV Recomendation System

The standard recommendation is to accept the job description and job description as a guide and provide a list of available jobs which matches to the job description. This is done in two ways.

Content-Based Recommender

Bearing in mind that this is the case for confirmation of similar information, we hire employees based on recommendations from employers that match job descriptions to the reference point and recommend the n best (n configurable)

CVs to employee comparisons. The prototype takes the cleaned resumes and job descriptions and merges them into a single file, then calculates the cosine similarity between the job descriptions and the resumes.

k-Nearest Neighbours: In this prototype, k-NN is used to determine the resume nearest to the jd, that is, the closest to the job description. First, to achieve a similar scale for JD and CV, we use an open-source library known as"gensim" that generates the content of the given text in a limited language. So, to get similar words for JD and CV, this library is used to generate the content of JD and CV, and it is used to find similar CVs like k-NN, JD.[5]

Conclusion

Resume Analyzer reduces effort and time in the CV shortlisting process and increases efficiency and accuracy. Considering all the features, the analyzer benefits the job seekers to land into a good job as well as the recruiters to hire the candidates with the required and quality skills. Recruiters hardly give a fair amount of time analysing a resume as they have thousands of resumes to study deciding whether it's valuable continuing the recruitment process with the applicant. The feelings and mental state of a person comes in between providing a non biassed judgement. Thus this system provides a non biassed approach to improve the recruitment process by giving a data-driven method to process cv automatically and provide recruiters more time to only examine candidates which are deserving.
As per the candidate point of view, it reduces the need for them to manually check their resumes and gives them confidence in their resumes.On the basis of the score generated from the analyser ,one can make the required changes in their respective resumes to get the best score using the analyzer, increasing their chances of being shortlisted by the companies.

REFERENCES

[1] Resume Parser with Natural Language Processing Satyaki Sanyal1 , Souvik Hazra2 , Soumyashree Adhikary3 , Neelanjan Ghosh 4 School of Electronics Engineering1, 3 School of Electrical Engineering2, 4 KIIT University, India

[2] Using an Electronic Resume Analyzer Portal (eRAP) to Improve College Graduates Employability Graduates EmployabilityNathan Green Michelle (Xiang) Liu Department of Information Technology, Data Science, and Cybersecurity Marymount University Arlington Virginia 22207, USA

[3] Resume Analyzer Using Text Processing B.Kelkar1, R.Shedbale2 , D.Khade3, P.Pol4, A.Damame5

[4]International Conference on Computational Intelligence and Data Science (ICCIDS 2019) A Machine Learning approach for automation of Resume Recommendation system Pradeep Kumar Roya,∗ , Sarabjeet Singh Chowdharyb, Rocky Bhatiab aVellore Institute of Technology, Vellore, TN, India bAdove, Noida, Uttar Pradesh, India

[5] Proposed System for Resume Analytics

[6] Automated CV Classification using Clustering Technique Prof. Sagar More1, Bhamare Priyanka2, Mali Puja3, Kachave Kalyani41Prof. Sagar More, Dept. of Computer Engineering, GCoE, Dhule(MS) India 2,3,4BE Computer

[7] JOB PORTAL RESUME EVALUATON SYSTEM USING TEXT MINING AND NATURAL LANGUAGE PROCESSING Jaichandran R1 , Leelavathy S1 , Kanaga Suba Raja3 , Pranav Kumar1 , Sailendra Kumar Majhi1 and Vikas Singh Thakur1

[8] Design of Automated Resume Extraction System Using Horspool and Karp-Rabin Algorithms in Text Mining P.Gajalakshmi*, Dr.V. Ramesh** *(Department of CSA, SCSVMV University, Kanchipuram) **(Assistant Professor, Department of CSA, SCSVMV University, Kanchipuram)]

[9] AN AUTOMATED RESUME SCREENING SYSTEM USING NATURAL LANGUAGE PROCESSING AND SIMILARITY Chirag Daryania, Gurneet Singh Chhabrab, Harsh Patelc, Indrajeet Kaur Chhabrad, Ruchi Patele

[10] A Resume Analyzer Application for Matching Candidates with Job Requirements, Using A Parsing Algorithm Abdiaziz Abdi Ali1, Farhan Hersi Olow2, Abdinasir Abdullahi Dahir3, Abdifitah Mohamed Hassan4, Mohamed Ahmed Mohamud5

[11] The Resume Research Literature: Where Have We Been and Where Should We Go Next? Stephen D. Risavy1 1 Lazaridis School of Business & Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada

[12] Data-driven HR Resume Analysis Based on Natural Language Processing and Machine Learning Tim Zimmermann Leo Kotschenreuther Karsten Schmid

[13] CV Analysis Using Machine Learning Authors: Avisha Anand, Mr. Sandeep Dubey

[14] Resume Analyzer An automated solution to recruitment process Ankita Satish Vaidya, Pooja Vasant Sawant

[15] REVIEW ON NATURAL LANGUAGE PROCESSING IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.3, No.1, February 2013