

Resume Parser

A
Project Report
Submitted for the Partial Fulfillment
Of B.Tech Degree
in
INFORMATION TECHNOLOGY
by

Himanshu Pandey (1805213024)

Deependra Kumar (1805213019)

Rajneesh Gupta (1805213044)

Under the supervision of

Dr. Upendra Kumar

Mr. Ajit Shukla



Department of Computer Science and Engineering

Institute of Engineering and Technology

Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh

Contents

DECLARATION	i
CERTIFICATE	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
LIST OF FIGURES	v
LIST OF TABLES	vi
1. INTRODUCTION	1
2. LITERATURE REVIEW	4
3. METHODOLOGY	9
3.1 Problems	9
3.2 Modules	9
3.2.1 Extraction of Text from Pdf Files	10
3.2.2 Extracting Field from Text	10
3.2.3 Extracting Skills from Resume	11
3.2.4 Saving to Database	11
3.3 Implementation	12
4. EXPERIMENTAL RESULTS	13
5. CONCLUSIONS	18
6. REFERENCES	19

Declaration

We hereby declare that this submission is our own work and that, to the best of our belief and knowledge, it contains no material previously published or written by another person or material which to a substantial error has been accepted for the award of any degree or diploma of university or other institute of higher learning, except where the acknowledgement has been made in the text. The project has not been submitted by us at any other institute for the requirement of any other degree.

Submitted by: - Date:

(1) Name: Himanshu Pandey

Roll No.: 1805213024

Course: Bachelor of Technology

Branch: Information Technology

Signature:

(2) Name: Deependra Kumar

Roll No.: 1805213019

Course: Bachelor of Technology

Branch: Information Technology

Signature:

(3) Name: Rajneesh Gupta

Roll No.: 1805213044

Course: Bachelor of Technology

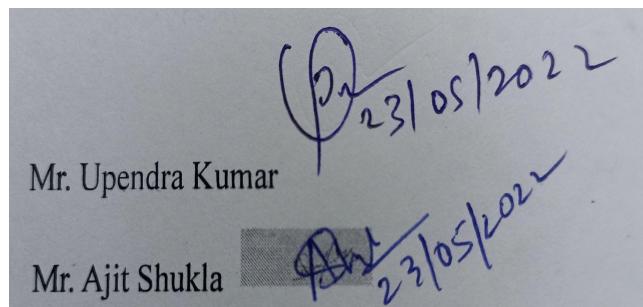
Branch: Information Technology

Signature:

Certificate

This is to certify that the project report entitled "Resume Parser" presented by Himanshu Pandey, Deependra Kumar and Rajneesh Gupta in the partial fulfillment for the award of Bachelor of Technology in Information Technology, is a record of work carried out by them under my supervision and guidance at the Department of Computer Science and Engineering at Institute of Engineering and Technology, Lucknow.

It is also certified that this project has not been submitted at any other Institute for the award of any other degrees to the best of my knowledge.



Department of Computer Science and Engineering
Institute of Engineering and Technology, Lucknow

Acknowledgement

We would like to extend our sincere thanks and thanks to our managers Dr. Upendra Kumar and Mr. Ajit Shukla who gave us this wonderful opportunity to work on a project called "Resume Parser". Our supervisors have been very helpful in completing this task. We would like to thank our managers for their diligence, patience, insightful comments, helpful information, practical advice and endless feedback that we have always been very helpful during the completion of this work. We learned a lot of new things while working on this project.

Secondly we would like to thank our family and friends who have assisted us so much in completing the work in a timely manner. We also extend our gratitude to the Department of Computer Science and Engineering Center for Engineering and Technology, Lucknow for supporting our project vision.

Thanks for all your encouragement!

Abstract

Online-fundamentally based rental gatherings play a significant capacity inside the employing channel with the quick blast of the net. Nowadays, almost every association or office imparts its assignment necessities to different online employing stages. There are north of 1000 work prerequisites transferred in accordance with minutes on Monster.com. a web-based townhouse might be extremely valuable in saving time for managers and workers. It permits work searchers to transport their letters to two or three representatives simultaneously without going to the workplace and recoveries a gathering of laborers' time making arrangements and paint returns. In the meantime, there are likewise numerous sites that capacity as an external association administration among work searchers and business undertaking people, to overhaul a great deal of those sites. For instance, LinkedIn.com has accumulated north of 300 million confidential CVs transferred through clients. In view of the creating measure of information, the procedure of accurately dissecting each resume is a top notch issue, which has drawn in the eye of specialists.

In reality, ambling elephants are uncovered through the animosity of surging smaller people. Most updates aren't written related to an inescapable configuration or a particular layout record. For this situation, the arrangement of restart information isn't entirely dependable. diminish the degree of outcome in suggesting staff who address the limit of the undertaking's issues and set aside some margin for faculty to do similar work. with a reason to improve the presentation of movement coordinating, investigating a powerful way to suit competitors with up-and-comers is basic and indispensable. Essentially, the resumption of mining assists with making a client model for the townhouse stage.

As far as its expressions of purpose, the resume private measurements have the accompanying attributes. To begin with, work searchers revise their books in a spread of approaches, but numerous reliefs incorporate typical text blocks, which incorporate non-openly available reports,

contacts, tutoring, and process records. 2d, confidential resumes rate a record-stage series shape, that is divided between unique things in the relating printed content block for each CV.

List of Figures

2.2 Visual Flowchart of proposed resume system

3.1 System Architecture

4.1 Input Format

4.2 Input Format

4.3 Output Format

List of Tables

2.1 Related Implemented Work

Chapter 1

Introduction

1.1 Overview

As of late, more electronic rental apparatuses have been created to continue data creation. Albeit fundamental speculations and procedures of web information extraction are accessible, the greater part of the electronic employing devices are as yet battling to handle text and coordinate up-and-comers with work prerequisites. There are three principal special cases for the resume in past exploration, which incorporate a catchphrase based search strategy, a standard based approach, and a semantic-based approach. Since continue subtleties are rare, it is one method for accomplishing the objective of match-fixing via looking for watchwords. Supported by the news website admin discharge strategy, a couple of rules-based expulsion techniques treat the restart text as a site page and concentrate point by point realities in view of the DOM tree structure. In the last option case, the scientists regarded the replication work as an issue for separating the semantic-based element. A few scientists utilize successive naming methods or text-isolating procedures to foresee the class markings of each line. Notwithstanding, the majority of these techniques depend vigorously on progressively work level construction data in restart text as well as a lot of marked information. As a matter of fact, the investigation of text yield models frequently depends on marked/characterized information by a human master. Furthermore, the more mastery and time required the recording system, the more costly it is to record information. Moreover, there might be limitations on the number of information occasions a solitary expert that can name in a potential manner.

This paper centers around the proposed yield calculation to restart the genuine result. Our

contributions are as per the following. We recommend a two-venture novel arrival of the data calculation. New sentence linguistic structure data, Writing Style, for each line in the CV book is based on this page. Utilized for acquiring arbitrary information ID data. We give strong execution approval of the proposed extraction calculation.

1.2 Motivation

Work Demand is viewed as one of the best compassionate undertakings to accomplish, so to speak. Our proposed model is essentially to separate information and measurements from the resume and to rank the resume in light of organization inclinations related with its requirements utilizing Natural Language Processing (NLP) strategies. Investigating and assessing the resume makes the employing system simpler and more effective. The resume holds different minutes of information inside it and any trustworthy supervisor needs to separate this information, for example, instruction, data, project, address and so on. For a particular work and utilizing the NLP technique, the expected subtleties will be isolated and the formal precise restart will be continued and will be re-recorded by the arrangement of organization abilities and faculty abilities in the CV gave.

Corporate organizations and employing offices process different positions consistently. This isn't crafted by men. A robotized insightful framework is required that can remove all the significant data from irregular records and convert everything into the very formal organization that can be supplanted by a particular work. scholarly experience, distributions, declarations, volunteer encounters, watchwords lastly a resume assortment (for example software engineering, HR, and so forth.). The refreshed data is then put away on a site (NoSQL for this situation) for sometime in the future. In contrast to other arbitrary information (for example email body, page content, and so forth), the restart is stopped. Data is put away in various sets. Each set contains information about private contact, work data or instructive subtleties. Regardless of this, the resumes are challenging to investigate. This is on the grounds that they fluctuate in the kind of data, their construction, way of composing, and so forth. What's more, they can be written in an assortment of ways. Other normal ones incorporate '.txt', '.pdf', '.doc', '.docx', '.odt', '.rtf' and so forth. To isolate information from various kinds of restart proficiently and really, the model shouldn't rely upon the request or sort of information.

1.3 Objectives

Examine data on continue utilizing normal language handling, track down catchphrases, order them in light of their watchwords lastly show the main congruity for the business in view of catchphrase coordinating. To start with, the client transfers the resume to the web stage. The examiner dissects all the necessary data from the resume and naturally fills in the structure for the client to check. When the client has checked, the resume is put away on our NoSQL site fit to be shown to managers. Likewise, the client gets their continuation in both JSON design and pdf.

1.4 Related Works

To get applicable writing on e-business and information mining from restarting, we summed up the past examination strategies and painstakingly chose the most important articles for our exploration. As far as acknowledged highlights, there are three kinds of famous techniques for recovering data from past examination, which can be portrayed as follows.

The primary gathering of strategies takes the arrival of catchphrases with thought. Just certain information was chosen to channel progression streams. The two of them expect to accelerate the presentation of occupation searchers. A portion of the key inquiries were made to channel the resume set with the goal that it could assist with further developing staff effectiveness. Albeit these sorts of techniques are not difficult to utilize, unrefined text content carries an excessive number of sounds to the file, prompting low precision and inadmissible quality outcomes.

The second gathering of strategies depends on the DOM (Document Object Model) tree structure, where labels are inner hubs and nitty gritty text, connection, or pictures, leaf hubs. Ji et al. they proposed a marker tree calculation, in which they found and eliminated a piece that was divided among website pages with a similar format, and the fundamental text was saved. Furthermore, a few strategies separate data about Regex rules in HTML pages. Jsoup and Apache POI can be utilized to dissect the resume that follows a particular layout document. Jsoup is a Java library for true HTML handling. Gives a simple to-utilize interface for extricating

and controlling information in light of DOM engineering. Furthermore, POI is a helpful Java library for working with Office document, which centers around removing record content. It is not difficult to make an arrangement to remove data from those CVs that follow a specific layout record.

Chapter 2

Literature Review

A. Towards an Information Extraction system based on ontology to match resumes and jobs[1]:

Proposed Methodology is Ontology based continue parser for finding Resume. The proposed cosmology driven data extraction framework, called ORP, will be worked on a few million English and Turkish resumes for switching the resumes over completely to an ontological format. The in general goal of the proposed ORP framework depends on an idea matching undertaking and ontological principles for English and Turkish continues that give semantic examination of information and parse related data like insight, elements, and business and training data from a resume. Here conversation is to computes rate fulfillment rely on work insight, schooling and so on.

B. A Job Title Classification System for the Online Recruitment Domain[2]:

The Proposed Methodology is Carotene order framework for Online Recruitment. The engineering of Carotene, a task title grouping framework right now underway at CareerBuilder. Carotene has a classifier design made out of a SVM-kNN overflow as well as a bunching part that is utilized in scientific categorization discovery. We thoroughly analyze Carotene with a third - party occupation characterization framework called Autocoder as well as an early variant of Carotene that depended on a level classifier engineering. Trial results show that Carotene better arrangement results with better degree of granularity. Here conversation is to involves SVM and KNN strategy in Carotene design.

C. Comparison research on text Preprocessing methods on Twitter Sentiment Analysis[3]:

Procedure is N-grams model, prior limit model, Naive Bayes Classifier. We direct a movement of preliminaries using four classifiers to check the feasibility of a couple pre-taking care of strategies on five Twitter datasets. Preliminary outcomes show that the departure of URLs, the removal of stop words and the ejection of numbers irrelevantly impact the introduction of classifiers; besides, superseding nullification and expanding contractions can additionally foster the gathering accuracy. Accordingly, taking out stop words, numbers, and URLs is fitting to decrease uproar yet doesn't impact execution. Subbing invalidation is suitable for feeling assessment. We select fitting pre-dealing with procedures and component models for different classifiers for the Twitter feeling course of action task. Here discussion is to use Large Volumes of data.

D. Architecture of efficient word processing using Hadoop MapReduce for big Data Applications[4]:

Proposed Methodology is Hadoop- Map Reduce, Hadoop Distributed File System. MapReduce is an extremely famous equal handling system for huge scope information examination which has turned into a viable technique for handling monstrous information by utilizing bunches of PCs. Somewhat recently, how much clients, administrations and data expanded quickly, yielding the large information examination issue for administration frameworks. To stay aware of the rising volume of datasets, it requires proficient logical ability to process and examine information in two stages. They are planning and lessening. Among planning and lessening stages, MapReduce requires a rearranging to worldwide trade the moderate information produced by the planning. In this paper, it is proposed an original rearranging system to

empower productive information development and diminish for MapReduce rearranging with number of continuous words and their include in the word processor. Here conversation is to utilize Time superseding technique

E. Text mining Research[5]:

Proposed Methodology is text mining utilizing Natural Language Processing and grouping. Here we plan to increment proficiency with text mining apparatuses in the extraction point. The space explicit applications are more appropriate for text examination. This is vital and it works on the exhibition of the data recovery framework. This examination paper breaks down the exhibition of seven open source tokenization instruments. From this examination, a few devices have perused text records just and thought about the set number of characters. By investigating every one of the actions, contrasted with different apparatuses Nlp Dotnet tokenizer gives the best result. Regularly, the Tokenization cycle can be utilized for the reports which are written in English and French, in light of the fact that these dialects utilize void area to isolate the words. Tokenization cycle couldn't be utilized in that frame of mind, for instance, Chinese, Thai, Hindi, Urdu, Tamil, and so on.

F. Survey on Text Mining and its Techniques[6]:

Proposed Methodology is Text Mining Process, NLP. Here the point is to get helpful pertinent data from grimy information. Text mining strategy is essentially utilized for removing designs from unstructured information. Different procedures for productively performing text mining are talked about in this paper. So in this paper, our emphasis is essentially on how message is to be mined. We have likewise examined the course of text mining, its applications, benefits and faults. Text reports are considered as semi-organized or unstructured configuration. PCs have not that much capacity to effortlessly separate phonetic examples when contrasted with people. Yet, the PC can deal with text at high velocity and in huge volume. Along these lines, text digging becomes helpful for PCs to analyze unstructured information. This procedure utilizes quantities of calculations for changing over unstructured text into valuable examples.

F. A CV Parser Model using Entity Extraction Process and Big Data Tools[7]:

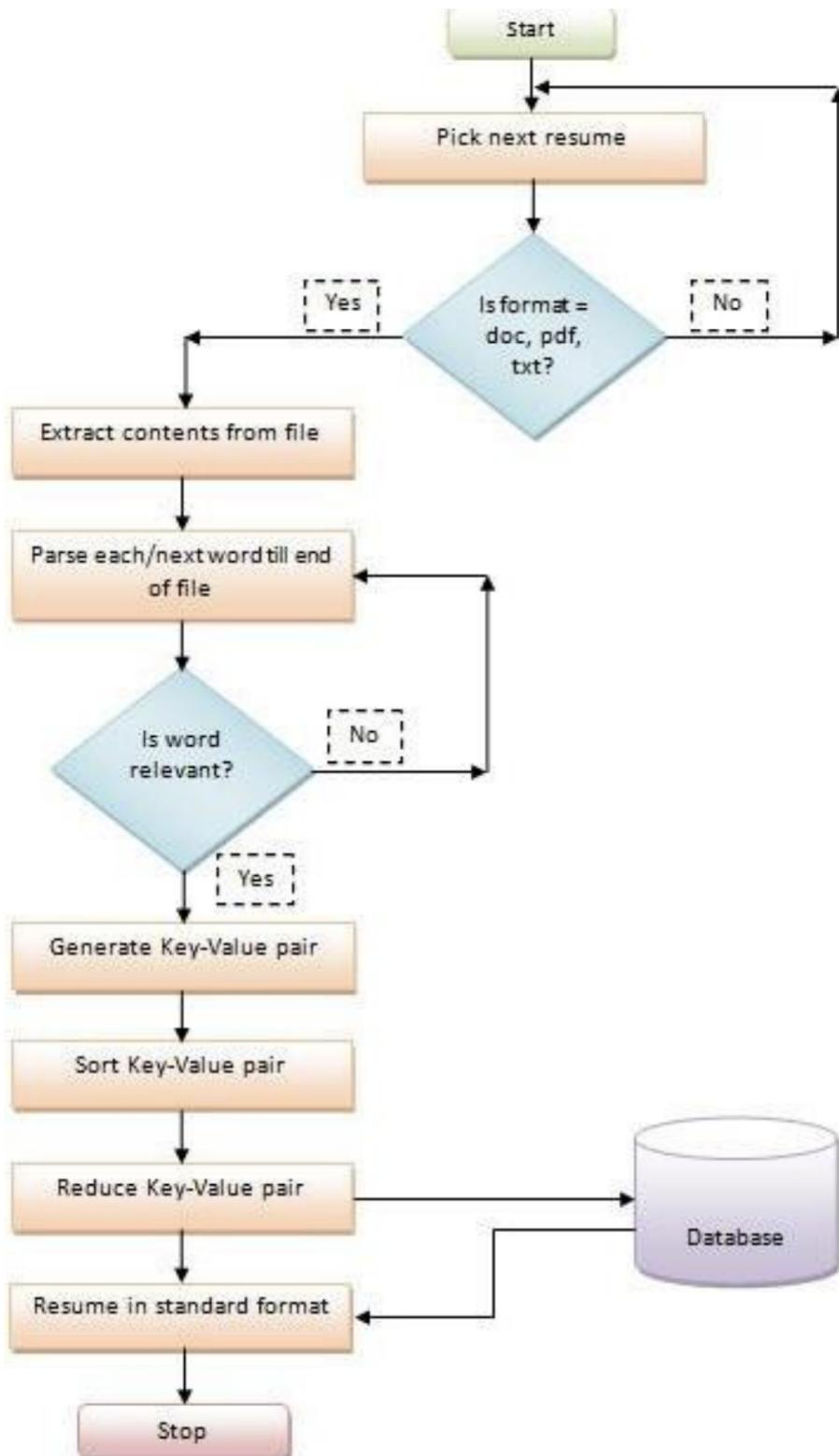
The informational collections of enormous information are huge and complex in nature. In large information, text examination has various viewpoints i.e element extraction, idea extraction, article extraction, microformat extraction and so forth. Presently, everybody looks for occupations through internet based work portal by transferring resumes. Every one of the resumes are not in structured design. A few resumes are unstructured and furthermore semi-organized. Our concern definition depends on planning a computerized continue parser framework, which will parse the transferred continue as per the work profile. Also, it will change the unstructured resumes into organized design. It will likewise maintain a positioning framework on the resumes. Positioning will rely upon the premise of data separated i.e as indicated by specialized abilities, instruction and so on. In this way, numerous products have been acquainted with tackle such enormous databases. CV parsing is such a method for gathering CV's. CV parser upholds various dialects, Semantic planning for abilities, work sheets, enrollment specialist, simplicity of customization. Parsing with enlist capacity gives us exact outcomes. Its innovation speeds up for requesting resumes as indicated by its sorts and configurations. Its mix makes clients. Enrollment specialist companies utilize CV parser procedure for determination of resumes. As resumes are in various arrangements and it has various kinds of information like organized and unstructured information, meta information and so on. The proposed CV parser strategy gives the element extraction technique from the transferred Cv's.

Paper Title	Techniques/Tools Objectives	Challenges
Architecture of Efficient Word Processing using Hadoop Map Reduce for Big Data Applications [6]	Hadoop Map Reduce, Hadoop Distributed File System To count the number of consecutive words and repeating lines	Time Consuming Method

Resume Parser: Semi-structured Chinese document analysis [22]	SVM classification model Extract informations from semi-structured Chinese resume format	Limited comparable dataset.
---	--	-----------------------------

Related Implemented Work

Fig.2.1



Flowchart of proposed resume system

Fig. 2.2

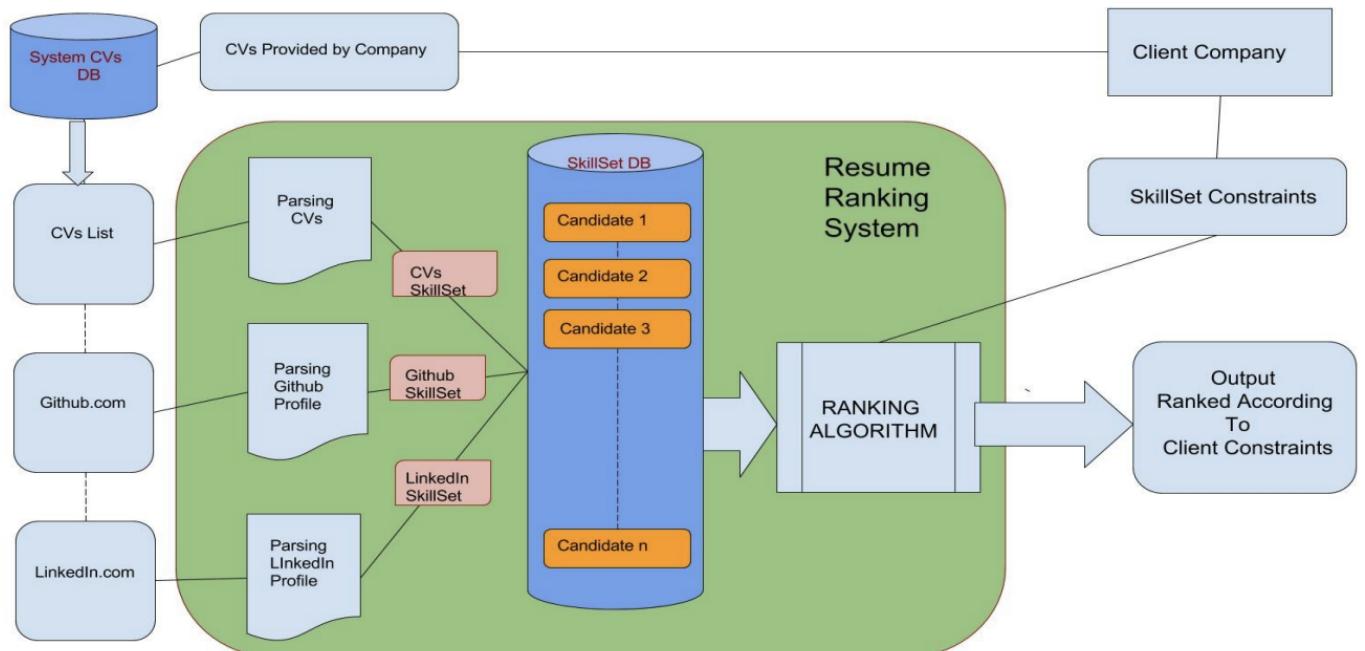
Chapter 3

Methodology

3.1 Problem

The problem is that current employment is not very flexible and efficient and time-saving. The current recruitment process is very complex and time consuming which forces candidates to complete all their skills and knowledge. The HR team also needs additional staff to process re-nominations. So that inspired me to build a flexible and automated solution.

3.2 Modules



System Architecture

Fig.3.1

1. Extraction of Text From PDF files

In this module we use pdfminer.six.

PdfMiner.six : Pdfminer.six is a local area (community) with the first PDFMiner fork. It is a device for separating data from PDF records. It centers around acquiring and investigating text information. Pdfminer.six removes text from the page straightforwardly into PDF source code. It can likewise be utilized to find the specific area, textual style or text tone.

Assembled measured with the goal that every part of pdfminer.six can be effectively altered. You might have the option to utilize your interpreter or your device which utilizes the force of pdfminer.six for purposes other than text examination.

```
1 pip install pdfminer.six
```

2. Extracting fields from Text

In this module, we will discuss about the nltk and numpy module.

NLTK : NLTK is a fundamental stage for building Python undertakings to work with human language data. Gives easy to-use objections to more than 50 associations and word reference gadgets like WordNet, as well as a summary of text-based libraries, token, venturing, checking, stepping, division, and semantic thinking, NLP dynamic current libraries, and conversation dynamic talk.

Numpy : NumPy is an essential software engineering bundle for Python. A Python library gives a multi-faceted cluster object, an assortment of articles recognized (like covered exhibits and frameworks), and an assortment of quick moving practice rehearses for similar individuals, including math, rationale, duplicity, sifting, determination, I/O. , Fourier clear advances, essential line variable based math, fundamental numerical working, arbitrary recreation and significantly more.

```
1 pip install nltk
2 pip install numpy # (also required by nltk, for running the following code)
3
```

3. Extracting skills from the resumes

Well, you went very well. This is the stage where things get tough. Extracting writing skills is also a very challenging task to increase accuracy.

Extraction of Entity: We can use NLP(Natural Language Processing) for parsing a text into paragraphs and sentences. This kind of text analysis is difficult in other programming languages. Because human language can be rich So that computer languages can not capture & encode the amount of information.

Annotation process: We need to create annotators for words and sentences. Annotators are created by functions which load the underlying Java libraries. These functions then mark the places in the string where words and sentences start and end. The annotation functions are themselves created by functions.

Tokenization: This is the very first step of the text pre-processing method. We have taken a text file in Python and broken it into words and sentences. These are called tokenization, as we are breaking up the text into units of meaning, called tokens.

Keyword searching: We take one resume as an input and search on the basis of different keywords.

4. Saving to DataBase

In this module we use **MongoDb** Database

MongoDB: MongoDB is a text-based NoSQL data set utilized for high-volume information capacity. Rather than involving tables and lines as in an ordinary relationship site, MongoDB utilizes assortments and reports. The records contain key matches that are the essential unit of information in MongoDB. Assortments contain record sets and capacity identical to related site tables.

Every site contains assortments containing text. Each record can shift with an alternate number of fields. The size and content of each report might vary from each other.

The construction of the record is a lot of in accordance with the manner in which engineers fabricate homerooms and their items in their different programming dialects. Specialists will frequently say that their classes are not lines and segments but rather have an unmistakable design with key worth matches.

Lines (or texts as they are brought in MongoDB) don't need a predefined outline. All things considered, fields can be made on the fly.

The information model found inside MongoDB permits you to address portion connections, keep similar individuals, and other complex designs without any problem.

Versatility - MongoDB conditions are extremely hazardous. Organizations all over the planet have characterized assortments some of which utilize 100+ hubs with a great many texts inside the site.

3.3 Implementation : First take the Resumes set as input, then use the pdfminer.six module extract the text from the pdf files and then use nltk to split text into paragraphs and sentences. This type of text analysis is difficult for some editing languages. Because human language can be rich In order for computer languages to capture and encode information.

We need to build annotations of words and sentences. Annotations are created by tasks that load the Java libraries below. These activities mark the places in the series where words and sentences begin and end. Annotation functions themselves are created by functions.

This is the first step in the process of processing a text. We took a text file via Python and broke it into words and sentences. This is called tokenization, as we divide text into descriptive units, called tokens, and then take the resumes as input and search on the basis of different keywords and use the same algorithms (brute force) to determine what percentage of skills. they correspond to the database of required skills and are calculated accordingly.

Formula for calculating score = (number of skills matched / number of input skills)*100

Chapter 4

Experimental Results

 Deependra  Deependra	DEEPENDRA KUMAR		(+91)6387000252 deependra2199@gmail.com
EDUCATION Lucknow, India Bachelor of Technology	Institute of Engineering and Technology Information Technology	Aug 2018 - May 2022	CGPA: 8.65
<hr/>			
EXPERIENCE Software Engineer, Intern • Created a Dashboard in Grafana for metric / health analysis of application and uses prometheus as a datasource which scrape all the default exposed metrics From Kubernetes application running in Minikube k8s cluster. • Created Rest API's for K8s cluster for interaction with kubernetes resources like Pods and integrated swagger -ui for api documentation and written Unit-Test cases • Tech: Docker Kubernetes Prometheus Grafana Promql python Flask	Zeta Suite (Direct)	17 Jan 2022-Present	
Software Engineer, Intern • Developed an Internal application for Log Analysis with ElasticSearch (DB as a destination) Kibana(Visualization) Kinesis fireHose (data streaming) and S3(as a backup) in AWS. • Developed an Application of simple calculator in AWS for basic Understanding of LAMBDA ,API GATEWAY ,SNS,SQS,IAM • Troubleshoot, Fixed and Tested API bugs, also reduced the Latency of code by changing the Logic and Asyncio • Developed Docker-image for specific versions of libraries and python for aws Lambda for removing Limitation of Runtime of python versions and size of File. • Tech: AWS Serverless Python MongoDB ELK stack Docker	Karza Technologies Pvt. Limited	28 June 2021 -28 Sep 21	
<hr/>			
TECHNICAL PROJECTS Weather App • An Application that gives the weather update of a particular location and it uses an API which takes a location name and it generates longitude and latitude of that location. After that it uses another API for forecasting the data. • Tech: Node.js Express HTML CSS API's .			
A Todo App • An App in which user can performing various operations like as add the task and update the task and delete the task • CRUD API where users can perform create, retrieve, update, delete operations. • Tech: Node.js Express Mongodb HTML CSS API's .			
Service2u Application • This is an app in which any person enter the location and app will show all the garage with many filters like which type of vehicle's garage owner providing service and nearest to the entered location and lowest possible price(not -exact) for the service. • Tech: Node.js Express Mongodb HTML CSS API's .			
<hr/>			
ACHIEVEMENTS • Highest-5 Star at codechef . • Highest Rating 1577(specialist) at codeforces . • Highest 1775 at Leetcode • Global Rank 59 (Div. 2) in Codechef Sept19 Long Challenge • Global Rank 144(DIV. 1) in Codechef Feb20 Long Challenge • Global Rank 962 in Round(A) Kickstart 2020 • Solved 1000+ problems on codechef,codeforces,hackerearth,hackerrank,leetcode,spoj . Stopstalk			
<hr/>			
TECHNICAL SKILLS • Strong Areas: Data Structures And Algorithms • Languages: C++ C Javascript python Java SQL • Technologies/Frameworks : Node.js Express Flask Mongodb Git Docker Kubernetes Prometheus Grafana			

 Phimanshu1234 phimanshu_2000	HIMANSHU PANDEY	Phone : (+91) 7704058035 pandey101299@gmail.com
EDUCATION Lucknow 226021, India Bachelor of Technology	Institute of Engineering and Technology Information Technology	Aug 2018 - May 2022 CGPA: 8.73
EXPERIENCE		
Software Development Engineering, Intern	LIDO Learning	1 July, 2021 - 31 October
<ul style="list-style-type: none"> > Implemented Cash payment validation by generating cron-based random security key and using that key the bank will call the validation API and after validate, deposit, the cash and send us a payment Receipt. > Implemented Refund-request generation features after completing all refunds and for that write Lambdas using GraphQL and Python. > Implemented various graphql queries, postgresql queries, lambda and views for various features like multi currency support, loan down payment percentage changes for specific providers, issue credits in loans. > Implemented Discounts coupon features for Middle-East Region students. > Techs - GraphQL Hasura Python postgresql Apollo server Reactjs Git Postman. 		
Software Development Engineering, Intern	Slice	3 Jan 2022- Present
<ul style="list-style-type: none"> > Implemented Various REST API's for admin Panel to manage users and their activities. > Implemented Various API for Crypto Services for Users Panel and also write Various SQL queries. > Implement Various Events Analytics and also perform Amplitude Testing. > Implement Authentication service for Admin and also tested all deployed API's through POSTMAN. > Techs - Postgresql Node.js Express.js MongoDB Typescript Git AWS Postman. 		
TECHNICAL PROJECTS		March 2021 - April 2021
Task-Manager API		
<ul style="list-style-type: none"> > A Task-Manager Backend API to give users a way to keep records of their To-do tasks developed in Nodejs. > A secure Registration and login system was created using Bcrypt and Json web tokens. > Implemented various features like pagination, sorting, filtering and sending emails through Sendgrid API. > Techs - Nodejs Express MongoDB Multer JWT. 		
Real-time chat Application		Jan 2021 - Feb 2021
<ul style="list-style-type: none"> > Developed full-stack real-time chat application in Node.js and Socket.io and hosted servers using Express.js. > Location sharing is enabled as the pinned location on Google Maps and chats automatically hide any profanity. > Added ability to remember messages when a new user joins an open room or return to old room. > Techs - Node.js Express.js Socket.io Mustache HTML CSS. 		
ACHIEVEMENTS		
<ul style="list-style-type: none"> > Ranked 1 in CodeMate 2021 conducted by NIT Jalandhar. (View) > Ranked 14 in GEEK BATTLE 2020 conducted by SSEC. (View) > Ranked 59(Global) in (Div.2) in Codechef Sept19 Long Challenge. > Ranked 36(Global) in Hackerearth October circuits 2020 and 74(Global) in Hack the Interview VI Asia Pacific. > Ranked 166 in Hackerrank Hackfest 2020 over 1000+ participants. (View) > Ranked 181 in Hackerearth Coding Challenge 2020 over 2000+ participants. (View) > Ranked 252 in CodeAgon(Codonation) over 47000+ participants. (View) > Ranked 980 in Google kickstart Round A 2020 and 1837 in Google Codejam Round 1B 2020 over 10000+ participants. > Highest Rating 1994(4-star) at Codechef and Highest Rating 1762 at Hackerearth. > Highest Rating 1591(Specialist) at Codeforces and 1894(Knight) at leetcod. > Qualified Facebook Hackercup upto Round 2. > Solved 1500+ Problems at Codeforces, Codechef, Leetcode, GFG, Hackerearth, Hackerrank, Spoj e.t.c (View) 		
TECHNICAL SKILLS		
<ul style="list-style-type: none"> > Strongest Areas : Data Structures And Algorithms . > Languages : C++ C Javascript Python SQL HTML CSS. > Technologies/Frameworks : Node.js Express.js Postgresql MongoDB Mysql GraphQL Git Postman Heroku. > Certifications : Hackerrank Intermediate Problem Solving. 		

Sudhanshu Ranjan	 Sudhanshu Ranjan codemail.sudhanshu@gmail.com
Education	
B.Tech in Information Technology , Institute of Engineering and Technology, Lucknow (19-23) 12th PCM (90%) , Central Board for Secondary Education 2019 10th (9.6 CGPA) , Central Board for Secondary Education 2017	
Experience	
Softpro India computer Technologies Pvt. Ltd. Intern Remote Sep '20 – Nov '20 <ul style="list-style-type: none"> • This Training Program helps trainees to get familiar with the Advancements taking place in the industry • Working On IOT, Raspberry Pi, Python, Java, C++, C Problem Solving. 	
LawtsApp.com UI Design Intern Jan '21 - March '21 <ul style="list-style-type: none"> • Worked closely with the UI Development Team. • Set Up a Process that includes the initial Brainstorming for the requirements of Users through different UX techniques like Surveys, User Based Feedback, Storyboarding 	
Head of Visual Design @ Entrepreneurship Cell, IET Lucknow	
Technical Projects	
Weather Application <ul style="list-style-type: none"> • An Application that gives the weather update of a particular location. • Displays Current weather • Forecast weather for upcoming week • It uses Mapbox API which takes a location name and generates longitude and latitude of that location, which uses by Weatherstack API which generates forecasting of that location <p>Tech Stack used : React JS Axios Html Css Js Deployed Link : Weather Web App</p>	
Food Catalog website using Zomato API <ul style="list-style-type: none"> • Keyword based searching • Every Result Redirects directly to restaurant page. • Responsive and Functional UI • It uses Zomato Web API, which takes a query as input and renders the best result in the locality. • Rendering using React functional components <p>Tech Stack used : React JS Axios Html Css Js Deployed Link : Eats App</p>	
Technical Skills	
Languages : C++ C Javascript SQL HTML CSS Frameworks/Technologies : ReactJs Axios MongoDB MySQL Git Postman	



EDUCATION

Lucknow 226021, India
Bachelor of Technology

PRIYA SINGH

Phone : (+91) 7394083919
singhpriya78460@gmail.com

PROJECTS

ComBett Web Application

- May 2022 - June 2022
- A common Platform that builds a community of helping each other in every aspect of software engineering, be it coding, development, placement preparation by providing right resources and direct interactions among the community.
 - Implemented a secure Registration and login system was created using **Bcrypt** and **Json web tokens**.
 - Added chat features using **Socket.io** so that users can easily chat with required skills experts.
 - Implemented **Interview Experience Dashboard** so that any users can share and read interview experience from there.
 - Implemented various features like **pagination, sorting, filtering** so that users can filter the content according to their preferences.
 - **Tech-stack-** Nodejs | Express | MongoDB | Multer | JWT | Socket.io.

Task-Manager API [\(View\)](#)

April 2022 - May 2022

- A Task-Manager Backend API to give users a way to keep records of their To-do tasks developed in Nodejs.
- A secure Registration and login system was created using Bcrypt and Json web tokens.
- Implemented various features like pagination,sorting, filtering and sending emails through Sendgrid API.
- **Tech-stack-** Nodejs | Express | MongoDB | Multer | JWT.

Manage-Budget [\(View\)](#)

Nov 2021- Dec 2021

- A Chrome Extension that manages various expenses and help us to maintain and regulate a budget.
- Implemented various features like setting spend limit, reset the limit and sending notifications if spendings is greater than the spend limit .
- **Tech-stack-** Javascript | CSS | HTML .

Movie_Recommendation [\(View\)](#)

May 2021- June 2021

- Movie_Recommendation is a project which helps in selection of top rated movies and its genre detection on the basis of emotion provided by the user.
- **Tech-stack-** Python | Beautiful Soup | Tkinter .

ACHIEVEMENTS

- **4 star (Rating: 1845)** rating on Codechef.
- Secured Global rank of **207** out of **14000+** participants in Codechef November Long Challenge.
- Ranked **357**(Global) out of **20000+** participants in Codechef September Long Challenge.
- Ranked **1080, 1259, 1281** among **11000+** participants in round Google Kickstart **Round B(2022) & Round G,H(2021)**.
- Ranked **1288** among **36000+** participants in SnackDown 2021-Online Qualifiers.
- Global rank **475** among 3000+ participants in **Google Code Jam 1/0 2022**.
- Secured **1365** rank out of **3505** teams in ACM-ICPC India Regionals, Online Round 2020.
- Ranked **122**(Global) in Coderita2.0 organized by Geeks for Geeks.
- Ranked **347** in Atcoder Beginner Contest 210.
- Solved **1000+** Problems at **Codechef, Atcoder, LeetCode, Codeforces, Geeks for Geeks, SPOJ**.

TECHNICAL_SKILLS

- **Strongest Areas :** Data Structures And Algorithms .
- **Languages :** C | C | Java | JavaScript | Python | Python | SQL | HTML | CSS .
- **DB :** MySQL | PostgreSQL | DBMS | OS | C | C++ .
- **Technologies/Frameworks :** Node.js | Express.js | MongoDB | MySQL | Git | Postman | Heroku .
- Certifications: **Hackerrank Intermediate Problem Solving**.

Rajneesh Gupta

Email: jmd1000rajneesh@gmail.com

Github

LinkedIn

Phone: +91-8887982315

EDUCATION

Institute of Engineering and Technology

Bachelor of Technology in Information Technology- 9.02 / 10

Lucknow, India

Aug 2018 - May 2022

EXPERIENCE

Backend Developer Intern

Airblack

June 2021 - Sept 2021

- Worked on developing Restful API's and integrating them with **Retrool** for automating the various requests from the member experience team thus reducing dependency on the tech team.
- Worked on adding different functionality like **showcase** notification, **videoFAQ**, **Course Completion** flow, **polls**, **Chat Reaction** in the system and testing them for various use cases.
- Worked on **Refund Flow** by integrating the APIs of Razorpay and Razorpayx.
- Worked on resolving Tech dependent Member issues , doing Root Cause Analysis and then debugging it.
- Tech Stack** - Node.js | MongoDB | TypeScript | Express.js | Firebase

PROJECTS

Camping Site

Mar 2021-Apr 2021

- Design and Develop a camping site where users can **add, delete** campgrounds and can give **review** and **rating**.
- Authentication and authorisation** is added to ensure that only authenticated users can add, delete and review. This feature is implemented through **passport JS**.
- Also users can delete only their reviews and campgrounds which they add, not of other users.
- Tech Stack**- Node.js | MongoDB | Express.js | JavaScript | CSS | HTML | Bootstrap

Car Game

Jan 2021-Feb 2021

- Implemented **Car Game** in which the user has to prevent his car from colliding with other cars or getting out of track.
- Score is calculated on the basis of driving speed and time duration of survival.
- Used **pygame** library for **game visualisation**.
- Tech Stack**- Pygame | Python

Color Guessing Game

Sept 2020-Oct 2020

- Developed a **game** in which we have to **guess the color** of given **rgb** code.
- Used **timers** for finding whether color is guessed correctly and disappears the color if guessed incorrectly.
- Two levels in the game are implemented, medium and hard.
- Random colors are generated through the `Math.random()` function of javascript.
- Tech Stack**- HTML | CSS | Javascript

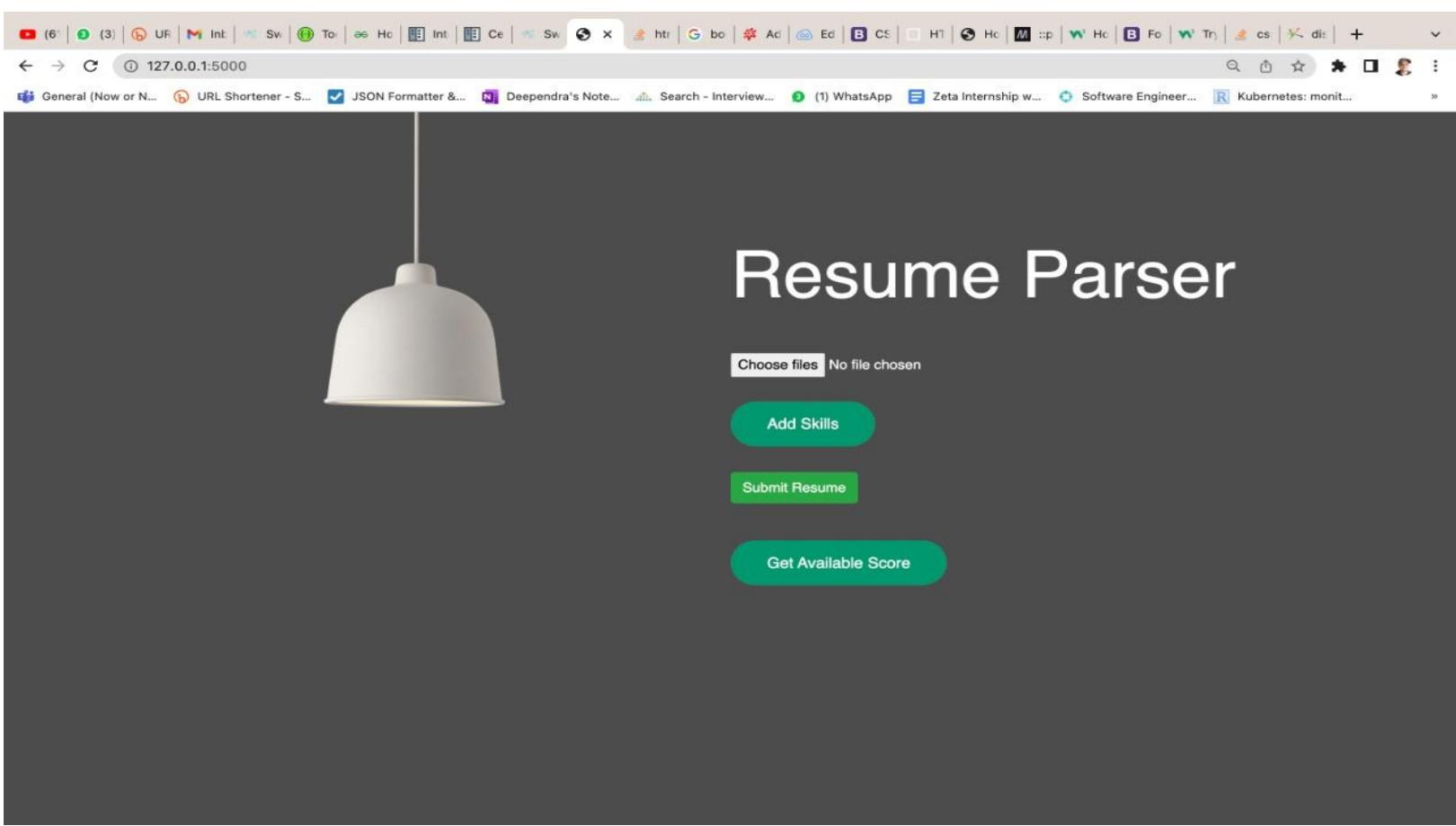
ACHIEVEMENTS

- 4* coder on **CodeChef** with the highest rating of **1957**.
- Solved **600+** problems in **Codechef, SPOJ, Codeforces, Leetcode (stopstalk)**.
- Ranked **107/15k+** participants in **Codechef February Long Challenge 2020**.
- Ranked **161/300+** participants in **Codechef September Cookoff 2020**.
- Ranked **2171/37k+** participants in **Codeforces Round 686 2020**.
- Ranked **2361/10000+** participants in **Google Kickstart 2020**.

SKILLS

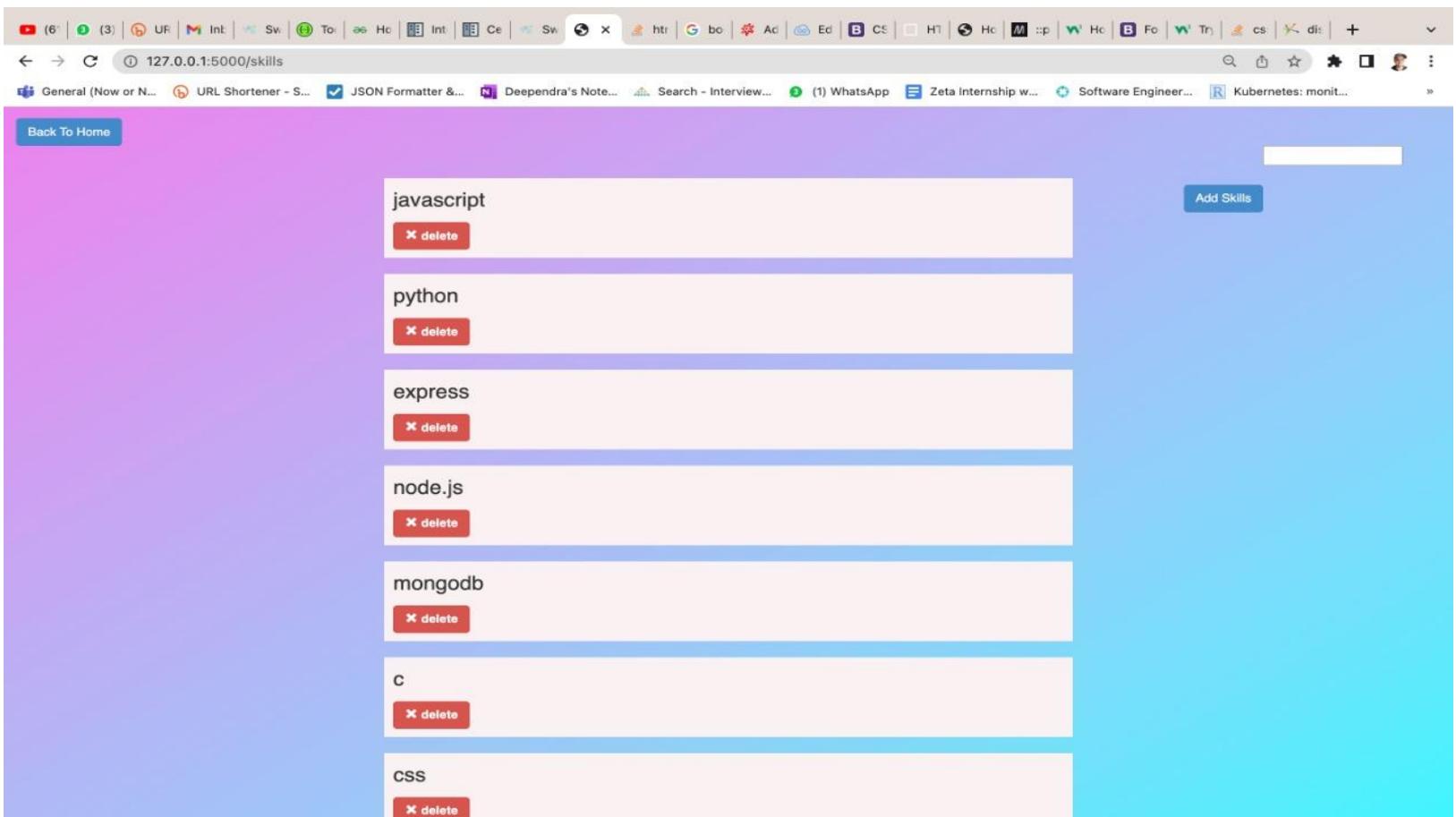
- Strongest Area** : Data Structures and Algorithms
- Languages** : C | C++ | JavaScript
- Technologies** : Web Scraping | HTML | CSS | Bootstrap | Node.js | MongoDB | Express.js | Git | Postman

Input Format



Input Format

Fig 4.1



Input Format

Fig 4.2

Output Format

Rank	Email	Mobile No.	Matching Percentage	Action
1	singhpriya78460@gmail.com	+91 7394083919	75.0%	X delete
2	pandey101299@gmail.com	+91 7704058035	75.0%	X delete
3	jmd1000rajneesh@gmail.com	+91 8887982315	62.5%	X delete
4	codemail.sudhanshu@gmail.com	+916355704015	62.5%	X delete
5	deependra2199@gmail.com	+91)6387000252	62.5%	X delete

Output Format

Fig 4.3

Here in output we showed the Ranking according to matching to requirements of skills , as it showed in output Format first we showed mail of user and mobile no. and then showed how many percentage skills are matching to the requirements skills and Ranked according to that.

Chapter 5

5.1 Conclusion

Our approach to making the work of companies and candidates easier and more efficient. In fact, our goal is to simplify the hiring process. This program will provide quality applicants to companies. Bad practices and prejudices in the system will be reduced. Based on information on state of the art technology resumes will be listed in order.

The main future of our project is to analyze the re-launch of various applications and websites such as LinkedIn, GitHub, Naukri.com, etc. In the future, this system could be developed into a wide variety of devices where extensive psychometric tests will be added. As a future project, we can expand the data set and improve the performance of the proposed system.

5.2 Future Works

- I. Making the model using real time values and work in real time.
- II. Then calculate the accuracy of the new model and compare it with this model and see if there is little or no difference.
- III. Training the model for parsing the resume of other languages other than English.

References

1. Çelik, Duygu, et al. "Towards an information extraction system based on ontology to match resumes and jobs." 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops. IEEE, 2013.
2. Javed, Faizan, et al. "Carotene: A job title classification system for the online recruitment domain." 2015 IEEE First International Conference on Big Data Computing Service and Applications. IEEE, 2015.
3. JIANQIANG, Z., AND XIAOLIN, G. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access* 5 (2017), 2870–2879.
4. MANDAL, B., SETHI, S., AND SAHOO, R. K. Architecture of efficient word processing using hadoop mapreduce for big data applications. In *Man and Machine Interfacing (MAMI)*, 2015 International Conference on (2015), IEEE, pp. 1–6.
5. VIJAYARANI, S., AND JANANI, M. R. Text mining: open source tokenization tools—an analysis. *Advanced Computational Intelligence* 3, 1 (2016), 37–47.
6. JADHAV, A. M., AND GADEKAR, D. P. A survey on text mining and its techniques. *International Journal of Science and Research (IJSR)* 3, 11 (2014).
7. <https://www.mecs-press.org/ijites/ijites-v10-n9/IJITCS-V10-N9-3.pdf>