e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

THE OFFICE - A SKILL FINDER WEBSITE

Nikhil Sontakke*[1], Shivansh Rastogi*[2], Shriraj Sonawane*[3], Nishant Tekale*[4], Ayush Wagh*[5]
*[1,2,3,4,5]Vishwakarma Institute OF Technology, Pune, India.

ABSTRACT

This project uses the potential of DOM parsing, Web Scraping and data cleaning to provide statistical overview of the skills required in any particular domain . A good resume is a much needed documentation for everyone seeking a job. Having the right set of skills in the resume helps an individual to get the desired job. The main aim of the project is to help users know about the skills required in a job. It will provide a platform which would minimize the time invested in researching skills which are not needed for the job. It will ease the user's work. This project also aims at providing users with a skillset which is dependent not only on the job title but also the skills which match one particular job profile and are local to a region.

Keywords: Data Science, Python, Django, Html, Css, Beautiful Soup, Web Application, Web Scraping, Job Skills.

I. INTRODUCTION

Finding the right job plays an important part in setting up a good career. Sometimes the job pays well enough but there is no job satisfaction. To get the perfect job, different skills must be acquired. Skills play a critical role in our resume. As a matter of fact, the number of skills doesn't have much significance compared to a quality skill. This model provides the user with insights on which skill is most required in the industry, this model not only gives results depending on the job profile but also based on the region in which the person is interested to work .The website aims to help users get information about the skills needed for any specific job mentioned by the user. The GUI of the website is carefully designed without using dark and appealing colors which ensures a better screen time. Previous studies imply that the use of vibrant and appealing colors decrease the visual appeal and perceived usability. Also, use of saturation is perceived as having negative effects on viewers. Pie charts provide an easy to understand picture, also provides a data comparison for the user at a glance to give an immediate analysis or to quickly understand information. The application will provide all the information related to skills in csv format and graphical pie chart format so that it is easy for the user to understand the job requirements.

II. OBJECTIVE

The objectives of proposed project are as follows:

A. To provide a platform where skills are compared and if and additional skills is required it may be highlighted to the user

B. To minimize the time wasted in learning unnecessary skills which are not even needed for the job profile

C. Graphs and charts provide a graphical overview of what skill is in demand in the specific job category

III. LITERATURE REVIEW

1. As a single-topic search engine, it is also an example of vertical search. Indeed, it is currently available in over 60 countries and 28 languages. In October 2010, Indeed.com passed Monster.com to become the highest-traffic job website in the United States. The site aggregates job listings from thousands of websites, including job boards, staffing firms, associations, and company career pages. They generate revenue by selling premium job posting and resume features to employers and companies hiring. In 2011, Indeed began allowing job seekers to apply directly to jobs on Indeed's site and offering resume posting and storage
2. Web scraping is a technique of data-collection from the world wide web. Web Scraping is the most suitable way of collecting real-time data from the world wide web. There are various methods of data collections ranging from manual human copy-paste to HTML parsing to DOM parsing to vertical aggregation , to semantic annotation recognizing. Out of these, this project has implemented HTML parsing , HTML parsing is a way to capture strings from a particular , elements from the client-side scripts .
3. Data cleaning is required while processing data which has many unwanted characters or strings . Data cleaning model which is used in this project is mostly based on REGEX which stands for regular expression. REGEX is a library in python which provides sufficient tools, in the form of functions/methods for

eliminating unwanted characters and ommiting text with keywords .Data cleaner checks for duplicate values and performs duplicate deletion.

4. Data visualization is an integral part of projects involving data analysis or as a matter of fact any project involving or dealing with data requires data visualisation for giving an appealing yet understandable look to the results. Pie charts are one of the most used data visualization methods . It helps users to understand the result in a matter of seconds. Which other methods such as tabular form or bar graph may require more attention and pre-requisite understanding to analyze the result.

IV. DESIGN

The proposed model is a website with a working dynamic backend and interactive frontend. The backend is designed upon the django framework using python programming language and the frontend is designed using HTML and CSS. When the user enters the job title and job location into the skillfinder web page the form will relay this information to appropriate functions in the backend responsible for procuring the data required for the job then the procured data will be cleaned and processed and at last the results will be displayed in the form of pie chart with an downloadable excel file which can be used for future reference by the user

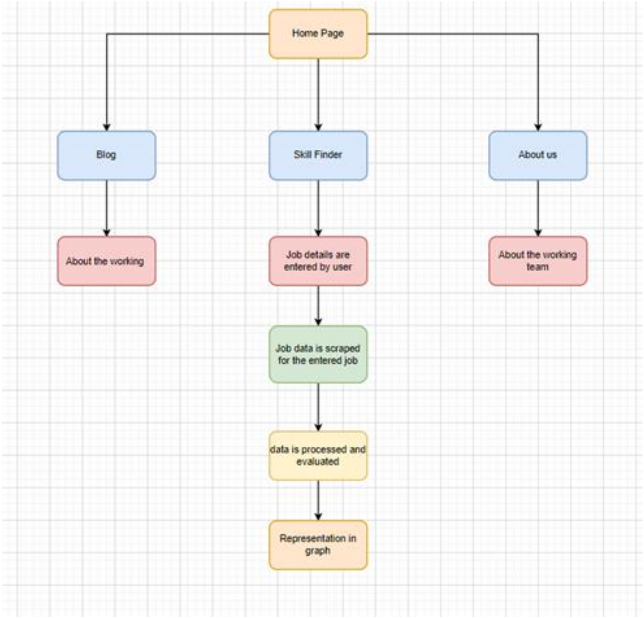

fig 1. - Design flow of the project

V. IMPLEMENTATION

The project is implemented on a web application platform using django framework in association with python programming language with the help of many python packages to assist. The main libraries used in our project are requests, bs4, pandas, matplotlib, csv Requests package is a python library which helps us to interact with the internet and send https requests to websites. We need this library to send http request to indeed and get the data from the website for further processing Python is used as the main programming language in this project. It is a popular language widely used in various different applications nowadays. It is used in development of many websites and provides the programmer an easy way to code and understand the concepts and implement them. Django is used as a backend framework in our web application. It is a python framework which is widely used in web development. Django 's primary goal is to provide simplified creation of complex, database-driven websites. The Django framework emphasizes “reusability” and “pluggability” of components, less code, low coupling, rapid development, and the principle of don't repeat yourself. Python is used throughout the django development, setting up the environment, programming files, and preparing data models. Django provides an optional administrative create, read, update, and delete interface which is generated dynamically through introspection and configured via admin models. Front-end development is implemented using HTML and CSS language which both when used effectively gives us a better user interface to interact with. HTML is used for making the frame of the webpage and to define all the elements and different aspects of the webpage. Whereas CSS is used to design this HTML defined webpage and make it look appealing

and interesting to interact with. Web Scraping in our project is achieved through "Beautiful Soup" which is a python library used for scraping data from different web pages. Beautiful Soup makes it easy to scrape information from web pages. It sits atop an html or xml parser in our case it sits atop a html parser and provides pythonic idioms for iterating, searching, and modifying the parse tree. We use html parser The data scraped through the Beautiful Soup library is raw data with so many unwanted characters so before working and processing the data we have to clean this data. This operation of data cleaning is procured through "re" python library which provides all the necessary functionalities related to data cleaning. After data cleaning the data is processed and only necessary items from the data are kept. We extract only the skills from the description of the job. Also some unnecessary data is removed from the temp object and the data is passed through csv writer to write the data into csv file for further use so that it is simple for us to work with a unified csv format and parse through it. This final CSV file is also made available for the user to download for future use. Pandas is a flexible, fast, powerful and easy to use open source data manipulation and analysis tool. It is built atop of the Python programming language. We are using the pandas library to work with CSV files and to handle data frames in our project. Matplotlib is a comprehensive library for making static, animated, and interactive visualizations in python. matplotlib produces publication-quality figures in different varieties of hardcopy formats and interactive environments across platforms. matplotlib is utilized in python scripts, the python and ipython shell, web application servers, and varied graphical interface toolkits.

VI. RESULTS AND DISCUSSION

The website when hosted upon a local port we are presented with a homepage of the website. We can navigate into different pages through the navigation menu. There are 5 different pages in the website which are homepage, about us page, contact us page, skill finder page and blogs page.

Homepage : This page provides just the basic idea of the website and its functionalities with a feature to signup for newsletter and links to different social media platforms

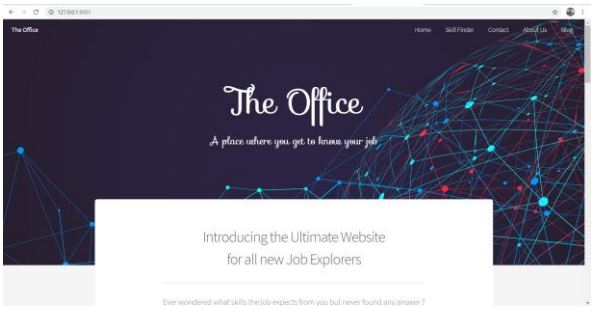

fig 2. - Homepage

About us page : This page provides the information about the developers of the website

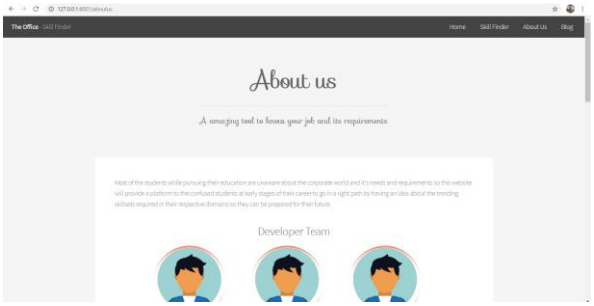

fig 3. - About Us page

Contact us page : This page gives a platform for users to contact the developer in case of any problems related to use of service by sending and email from the website

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

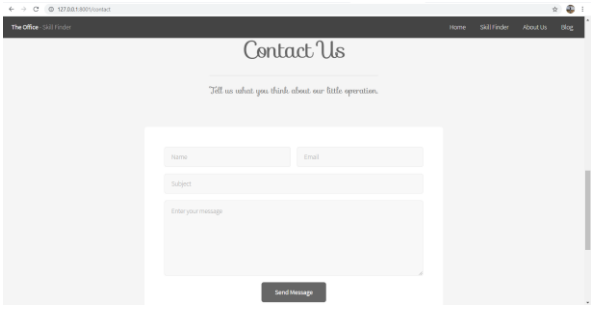

fig 4. - Homepage

Skill finder page : This page provides an interface for the user to give his job title and job location preferences and use the main application.

In this page there is a form in which the user enters his job title and job location for which he wants to know the skills required for, specifies the branch of his degree and validates captcha and then clicks the search button.

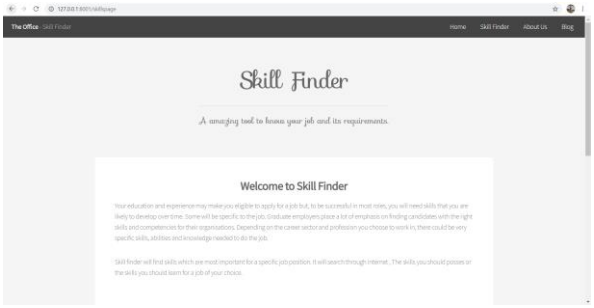

fig 5. - Skill finder page

Then the website relays this information to backend processing and after all the results are generated the website will redirect to a new page where users will be provided with a pie chart depicting the skills percentage for a better visualization of the shares of skills from the extracted skills.

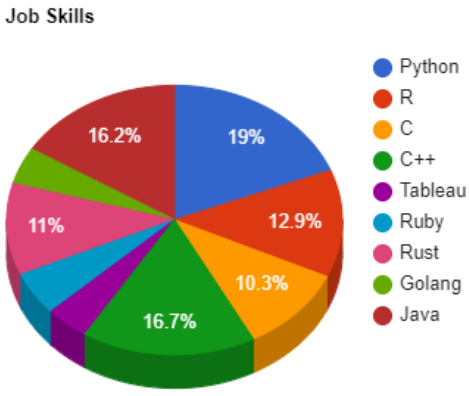

fig 6. - Output Pie chart

Above pie chart shows the result of one of the trial runs where inputs were provided: 'Job Title' as 'Software Engineer', 'Region' as 'California'. The pie chart depicts that Python R Golang are the most favoured language for eligibility in this California for software developer aspirants.

VII. CONCLUSION

THIS PROJECT HAS BEEN SUCCESSFUL IN ACHIEVING ALL THE OBJECTIVES. SCRAPING TIME OF THE WEB SCRAPER MODEL IS OPTIMISED WITH LIMITED RESOURCES. DATA CLEANER USES ONLY BASIC REGEX FUNCTIONS TO EXTRACT USABLE DATA FROM THE RAW DATA WHICH IS PARSED BY THE WEB SCRAPER. DATA IS VISUALIZED IN THE FORM OF PIE CHART AND TABULAR FORM.

FUTURE SCOPE

A. The module can be implemented over a mobile application for easy accessibility

B. The Data visualization can be more interactive

C. The program algorithms can be more optimized

ACKNOWLEDGEMENTS

We would like to express our special thanks to our professors Milind Patwardhan and Vijay Gaikwad who gave us the golden opportunity to do this wonderful project , which also helped us in doing a lot of Research and we came to know about so many new things. We are really thankful to them.

VIII. REFERENCES

[1] Beautiful soup documentation - https://www.crummy.com/software/BeautifulSoup/bs4/doc/

[2] Pandas documentation - https://pandas.pydata.org/docs/

[3] Django documentation - https://docs.djangoproject.com/en/3.2/

[4] Web scarping technology, EIJO Journal of Engineering, Technology and Innovative Research (EIJO – JETIR) [ISSN : 2455 - 9172], http://eijo.in/journals/article_detail/232

[5] Benchmarking of Web scraping data on various types of Industries Supporting Transport Sector, https://ijsdr.org/papers/IJSDR190862.docx

[6] S.C.M de S Sirisuriya2015, A comparative study on web scraping. Proceeding of 8th international research conference, KDU

[7] List of Web Harvester, Data Scraper, Web Scraping Software and Tools, n.d.WebData Scraping,http://webdata-scraping.com/webscraping-software/