INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1748)
Available online at: https://www.ijariit.com

Scraping of Job Portal

Shubham B. Gulik
shubhamgulik80@gmail.com
Datta Meghe College of Engineering, Navi Mumbai, Maharashtra

Akash R. Gharat
aakashrgharat172@gmail.com
Datta Meghe College of Engineering, Navi Mumbai, Maharashtra

Jayesh L. Choudhary
Jayeshchoudhary9769@gmail.com
Datta Meghe College of Engineering, Navi Mumbai, Maharashtra

Sujata Kolhe
sujata.kolhe@dmce.ac.in
Datta Meghe College of Engineering, Navi Mumbai, Maharashtra

ABSTRACT

Scraping of Job Portal" is an online Job Searching Web application, a web application through which job seekers can register and apply for jobs. A job seeker must find jobs through advertisements, college fairs, job fairs etc., and the employers must put in much effort to find the right candidate for a vacant position. This application addresses such shortcomings and is a convenient platform for job seekers to find and apply for jobs. Candidates can search for jobs in any field through advanced search techniques. Basically, our scraper scraps the data from various job portals and then provides job options to the user or job seeker according to their needs such as location, salary, working time, technical skills etc. They can upload their basic information related to job search which is stored in a database. Candidates can use this portal without any geographical barrier, from any part of the world. Also, our portal will notify the registered users through email notifications which consists of basic information about jobs and also the apply link for that job.

Keywords— *Web Application, Web Scraping, Job Portal, Apply Link, Email Notifications, Django, Beautiful soup, SMTP, Authentication.*

1.INTRODUCTION

Whether entering the job market for the first time or re-entering after a break or switching career, job search is a challenging task. But how about tools/applications, making this tedious process look friendly, systematic and easy to reach out, to employers or candidates. Searching and landing up with a job portal is a tedious process for a job seeker. This project is aimed at making such challenges much easier despite the geographic location of either the job seeker or the Company. Although a job search portal doesn't guarantee a job offer, it is the best place for potential candidates and employers to get connected and know more about each other. "Scraping of Job Portal" is an online web application which is a web scraper which scraps the suitable jobs based on the user's profile. It is a simple, efficient, convenient and systematic application. This portal enables candidates looking for jobs to register themselves with the website, look up for different jobs according to their qualifications and apply for those jobs conveniently. After applying, our web application sends a HTTP request to the URL of the webpage the user wants to access such as timesjobs.com. The server responds to the request by returning the HTML content of the webpage. For this task, we will use a third-party HTTP library for python requests. Once we have accessed the HTML content, we are left with the task of parsing the data. Since most of the HTML data is nested, we cannot extract data simply through string processing. One needs a parser which can create a nested/tree structure of the HTML data. Now, all we need to do is navigate and search the parse tree that we created, i.e.tree traversal. For this task, we are using another third-party python library, Beautiful Soup. It is a Python library for pulling data out of HTML and XML files. Thus, by using web scraper technique our web application notifies the user with best suitable job options with the help of email notification which also includes basic information of the particular job. A scraping of a job portal web application comes to the rescue at this point where a lot of meaningful time can be saved. The entire process of a job scraping as well as job searching is speeded up. Manual processes get replaced by automated processes. Candidates are just a few clicks away to get connected. Another advantage is once the candidate is registered and applies for a position, his/her information stays with the database for both the present and future use for available positions. Also our system will send the email notification to the registered users which consists of basic information about jobs and also the apply link for that job.

1.1 Objective

The objective of this project is to create a web application prototype that can combine web scraping process, aggregation,

and dashboard by using Data and Data given by the user while registering on web application. This project is to develop an online job search for the job seekers who can get multiple jobs from different job portals. The system is an online application that can be accessed throughout the organization and outside as well with proper login provided. Here user will be able to see the various jobs suggestions from multiple job portals in our scraping portal.

2. SPECIFICATION

Domain: Web Scraping
Hardware: i5 Processor
Operating System: Windows 10
Technology: Python3, HTML, CSS
Environment: Anaconda
IDE: PyCharm
Libraries: bs4, requests, url open, SMTP, send mail

3. METHODOLOGY

Step 01
Sending a HTTP request to the URL of the webpage we want to access. For example Fresher Jobs Portal (https://www.freshersworld.com/jobs) The server responds to the request by returning the HTML content of the webpage. For this task, we will use a third-party HTTP library for python requests.

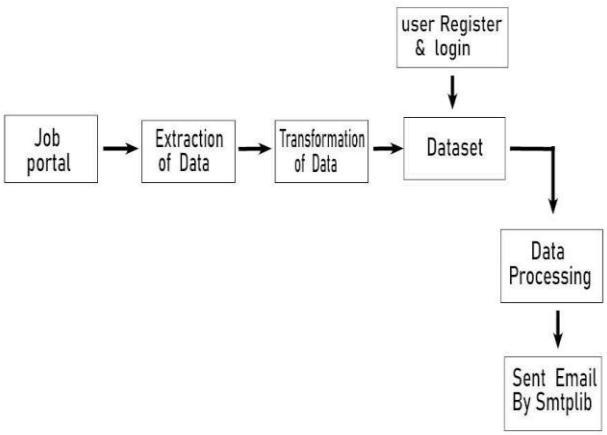
Step 02
Once we have accessed the HTML content, we are left with the task of parsing the data. Since most of the HTML data is nested, we cannot extract data simply through string processing.
One needs a parser which can create a nested/tree structure of the HTML data.

Step 03
Now, all we need to do is navigating and searching the parse tree that we created, i.e. tree traversal. For this task, have used another third-party python library, Beautiful Soup. It is a Python library for pulling data out of HTML and XML files.

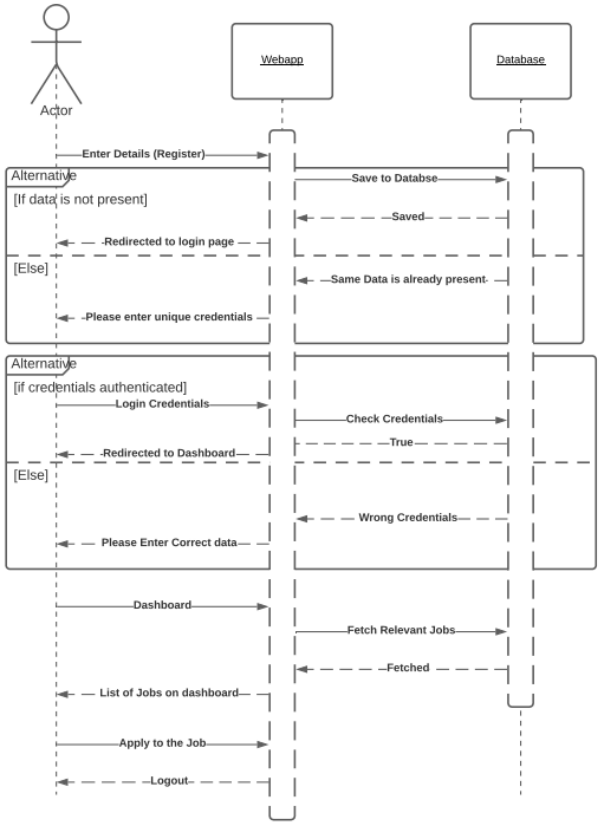
Step 04
Once the data is extracted, it will be stored in our MondoDB database. It will be fetched according to the registered user's requirement.

Step 05
We have used SMTP protocol for sending email notifications to the user as per the best suited jobs. Our system will sends the email to the user twice a day.

3.1 Data Flow Diagram



3.2 Sequence Diagram



4. IMPLEMENTATION AND DEVELOPMENT

REGISTRATION AND LOGIN:

Generally, in today's world, there are plenty of job portals which provides job opportunities to the job seekers. But user have to register all of the job portals to find the best suited job for them based on various parameters by providing lot of details such as email id, name, aadhar number, mobile number etc. Thus, it is the time consuming and lengthy process as one job portal takes approx 15 minutes for registration. In our job portal, user have to provide only basic information for registration such as Full Name, Mobile Number, Password (for your profile), Your skills(by selecting from drop down list), List of locations, List of job titles. By completing the registration process, user will be able to see the lot of job options from multiple job portals at one place. In turn our portal reduces the time of the lengthy registration at every job portal.

Drop Down List of Skills, Locations and Job Titles: In this dropdown list, the options will gets automatically updated by the scrapper based on the scrapped data.

Update User Profile: Also, user will be able to add or delete list of skills, locations and job lists according to their convenience.

Email notification: Our job portal will notify the user about newer job based on the top suited jobs by comparing the existing job profile of the user. Thus, it will reduce the lengthy process of job searching and user will be able to apply for the best suited just by clicking on the link.

Newer Jobs: Our scrapper will scrape the job portals once a day. Thus, our database will get updated after every 24 hours. Hence user will be able to see the newer jobs as per their existing profile.

Delete Old Jobs: Old jobs from the database will be automatically deleted based on their expiry date. Thus, user will be able to see the newer jobs and not the expired ones.

Robo3T GUI: In this graphical user interface, we(owner) can see the data in three formats namely text, table and tree mode. It will helps us to perform various operations on the database. We can add or delete columns or rows by simply clicking on the add or delete button.

4. CONCLUSIONS

We have developed this scraping model which will rescue at that point where a lot of meaningful time can be saved. The entire process of a job scraping as well as job searching is speeded up. Manual processes get replaced by automated processes. Candidates are just a few clicks away to get connected.

Some of the major challenges faced was in understanding the scraping concepts and implementing them in the application. By dealing with it, we have developed the best web scrapper using python based beautiful soup (bs4) library. So the user will be able to see the best job opportunities from multiple job portals (Assan Jobs, Freshers World, etc.) at a single place that is our job portal. We have successfully added the best functionality i.e. the email notifications. This was the major problem we had faced while developing this scrapper model. By dealing with it, we have developed notification model using SMTP protocol and scheduler.

Thus, the registered user will be able to avail the functionalities such as easy registration process, scrapping the best jobs, getting the email notifications daily, updating their jobs profiles as per requitements. In turn, it will reduce the leghty registration process on every job portal.

5. FUTURE WORK

There is ample scope of enhancement and adding functionalities to this application. This application can be extended to send automated interview scheduling through acceptance/rejection of Resume. The application can have a job recommendation system based on the frequent search results of different users. There can be a feedback or review section for the application. Also, unlike the current job description page, user can view job description in a separate page with one click on the job description. The User functionality can be extended to give the user options to save the job and later apply, to upload multiple documents. The application can be more scalable by extending the search functionality based on country, city or area. While this application meets the basic requirements of a job portal eliminating few of the traditional challenges faced like time, money and effort, it can be extended to make the application more dynamic and robust. The User Interface can be made more attractive and user friendly.

6. ACKNOWLEDGEMENT

We would like to thank our Project Guide Prof. Dr. Sujata Kolhe for believing in us, our capabilities and keeping her faith on us that we would be able to finish the project on time, guiding us through all the major decisions and above all believing in our abilities. Her constant guidance, encouragement and valuable feedback led to the successful completion of this project. Above all, thank you God for all the opportunities that you have given us.

7. REFERENCES

[1] Vivek Kumar Sehgal1, “Job Portal-A Web Application for Geographically Distributed Multiple Clients”, 1 Department of CSE and ICT Jaypee University of Information Technology, Waknaghat, Solan, H.P (INDIA)M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[2] ”Ghazvinian, Holbert, Viswanathan. “Simple WebScraping.”Internet:https://seanholbert.wor dpress.com/2011/07/15/scrappy-simple-web scraping/, Jun. 2015”

[3] “Datahen."3 Advantages of web scraping for yourenterprise"Internet:https://www.datahen.c om/3-advantages-web-scraping enterprise/,May.17,2017”

