www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 05, May 2023, pp : 267-271

e-ISSN : 2583-1062

Impact Factor : 5.725

BUILDING A JOB PORTAL USING WEB SCRAPPING

Vishnu Priya N P[1], Nehaa Shri A[2], Mary Valentina Janet A[3], MubeenFathima G[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Agni College of Engineering, Chennai, Tamil Nadu India.

[2,3,4] Students, Department of Computer Science and Engineering, Agni College of Engineering, Chennai, Tamil Nadu India.

DOI: https://www.doi.org/10.58257/IJPREMS31247

ABSTRACT

Job portals are online platforms that facilitate the matching of job seekers with potential employers. Web scraping, which involves extracting data from websites, is commonly used to gather information from job portals to gain insights into the current job market. Through web scraping of job portals, software tools or bots can extract data such as job titles, industries, required skills, and locations of job vacancies. This information can be analyzed to provide valuable insights into the job market, including the most in-demand skills, industries that are hiring the most, and the regions with the most job opportunities. Web scraping of job portals can also provide job seekers with useful information about job titles and roles offered by recruiters, along with the skills required for a particular job. This information can help job seekers to better understand the job market and the skills they need to acquireto increase their employability. For instance, if web scraping reveals that programming skills are in high demand, a job seeker can take courses to improve their programming skills and increase their chances of getting a job in the IT industry. However, it is important to note that web scraping of job portals can be unethical and illegal if done without permission from the job portal or in violation of the portal's terms of service. Some job portals may prohibit data scraping, and any attempt to scrape data from these portals may lead to legal action. Therefore, it is crucial to conduct web scraping of job portals in a responsible and ethical manner, ensuring that any data extraction is done with appropriate permissions and within legal boundaries.

Keyword: Web Application, Web Scraping, Job Portal, ApplyLink, Email Notifications, Django, SMTP, Authentication

1. INTRODUCTION

Searching for a job can be a daunting task, whether it's a first-time job seeker, someone re-entering the workforce after a break, or someone looking to switch careers. Fortunately, there are tools and applications available that make this process more manageable and user- friendly for both job seekers and employers, regardless of their location. Job search portals connect potential candidates and employers and provide a platform for them to learn more about each other. However, finding the right job portal can be a time-consuming process for job seekers. To make this process easier, we have developed an online web application called "Scraping of Job Portal." A web scraper uses a user's profile to find suitable job opportunities. This application is simple, efficient, convenient, and systematic. Job seekers can register on the website, search for jobs according to theirqualifications, and apply for them with ease. Once the user applies, our web application sends a HTTP request to the webpage's URL, such as timesjobs.com. The server responds by returningthe HTML content of the page. To parse this data,we use the Python requests library for HTTP and Beautiful Soup for parsing HTML.With our web scraper, job seekers can receive emailnotifications about suitable job options, includingbasic information about the job. This process speeds up job searches, replacing manual processes with automated ones. Once a candidate registers and applies for a position,their information stays in the database for future use. Additionally, our system sends email notifications to registered users with information about available jobs and the apply link for each job. Overall, our "Scraping of Job Portal" web application simplifies the job search process, making it more user-friendly and efficient for both job seekers and employers.

2. LITERATURE SURVEY

Renita Crystal Pereira et. al., Web scraping involves summarizing, employing techniques and utilizing tools to extract data from websites, but it's not a straightforward process. The goal is to ensure that the extracted data is accurate, reliable, and maintains integrity despite the vast amount of information available that can be difficult to manage and store. However, there aresome issues associated with functional techniques, such as the potential damage caused to websites due to excessive web scraping.

Kaushal Parikh et. al., A machine learning- based approach for detecting web scraping has been proposed, which can be particularlybeneficial for research-focused companies.Web scraping has always been a challenging security issue, as any data placed on the internet can potentially be copied, pasted, and used by others without the knowledge ofthe

original company. Despite the existence of various protective measures, some are stillbeing overlooked.

Erin Farley et. al., The purpose of this presentation is to introduce the concept of web scraping to law enforcement researchers and explain how this technique operates. Recently, there has been a growing trend of investigators in criminal justice using web crawling for their work. While web scraping is commonly viewed as a method for data collection to facilitate analysis and research, it requires technical skills in computer science that social science researchers may not possess. Developing a web scraper necessitates a high level of proficiency in programming languages such as R or Python. Despite this, the literature review reveals a significant gap in understanding the potentialbenefits of web scraping for stakeholders in various fields, including healthcare, social media, e-commerce, and job portals. For instance, data extracted from job portals canprovide insights into the most critical skills required by both recruiters and job seekers. Therefore, this study aims to bridge this knowledge gap by exploring the various dimensions of web scraping and data analysisacross different industries.

3. EXISTING SYSTEM

The job market offers numerous opportunities for job seekers through various job portals. However, the process of finding a suitable job can be daunting, as candidates are required to register on multiple job portals and provide extensive personal information, including their name, email ID, Aadhar number, and mobile number, among others. This can be a time- consuming and tedious task, with each registration taking around 15 minutes.

1. PROPOSED SYSTEM

Our job portal offers a streamlined registrationprocess where users are only required to provide essential information such as their fullname, mobile number, and preferredpassword for their profile, select their skills from a dropdown list, and choose job titles andlocations of their interest. Once the user completes the registration, they can accessnumerous job options from multiple job portals in one convenient location. With our portal, users can save significant time that would otherwise be spent on lengthy registrations at each individual job portal.

2. DIRECTORIES OF MODULES

2.1MODULE 1: WEB SCRAPPING ENGINE

Web scraping is a valuable technique for extracting data from websites. However, it can be a challenging and time-consuming process, requiring knowledge of programming languages and data mining techniques. Our web-scraping engine offers a simple and efficient solution for users to extract data from various websites. Withadvanced algorithms and technologies, our platform ensures that the collected data is reliable and accurate. Our platform is designed with a user-friendly interface and customizable web scrapers, making it easy for users to configure the scrapers based on their specific requirements. The platform supports various data formats, including HTML, CSV, JSON, and XML, enabling users to store and analyze the extracted data with ease. To ensure that the data is always up-to-date, our platform features a built-in scheduler that allows users to schedule data extraction at specific times. Our platform also supports authentication methods such aslogin credentials and cookies, enabling users to scrape data from websites that require user authentication. Our web-scraping engine ishighly scalable and capable of handling large volumes of data efficiently. Users can easily manage and organize the scraped data,enabling them to make informed decisions and gain valuable insights from the collected data. Overall, our web-scraping engine provides a reliable and efficient solution for users looking to extract data from various websites. With our platform, users can streamline their data collection process,saving time and resources while obtainingaccurate and reliable data.

2.2MODULE 2: DATABASE

In web scraping, one of the most critical aspects is the management and organization of the vast amounts of data collected from different websites. This is where databases play a vital role in the process. Databases offer a robust platform to store and manage the extracted data in an organized and structured manner. With databases, users can search, analyze, and update the collected data seamlessly, without the need for manual intervention. One of the significant advantages of using databases in web scraping is the ability to consolidate data from different sources into a single location. This makes it easier to manage the data, analyze it more effectively, and gain insights into the market trends. Databases also enable users to automate data extraction and management processes, saving time and effort. By scheduling the extraction of data from specific websites and automatically adding it to the database, users can streamline the data collection and analysis processes. Moreover, databases can help users quickly locate specific data by indexingand sorting the information. This can be especially helpful in identifying patterns and trends in the data, leading to informed decisions. In conclusion, databases are a crucial component of web scraping. They offer a robust platform for organizing and managing large amounts of data, enablingusers to streamline their data collection and analysis processes, leading to more informeddecision-making.

2.3 MODULE 3: JOB SEARCH ENGINE

A job search engine is a powerful tool thatsimplifies the process of job hunting by collecting job listings from various sources, including job boards, company websites, and recruitment agencies. With the help of a job search engine, job seekers can access numerous job opportunities in one place, without having to visit multiple job portals. The job search engine works by crawling through different websites and job boards, collecting relevant job listings, and organizing them based on the job seeker's search query. The collected data is then displayed on a single platform, making it easier for job seekers to browse and apply for relevant job listings. Job search engines also offer several advantages over traditional job search methods. For instance, they save job seekers a significant amount of time by eliminating the need to manually search for job openings on different websites. Instead, users can specify their job requirements and get instant access to joblistings that match their qualifications. Another benefit of using a job search engine is that it provides access to a broader range of job opportunities. This is because job search engines often include job listings from smaller or niche job boards that may not be easily accessible otherwise.Job search engines also offer features such as customized job alerts, which notify job seekers of new job listings that match their search criteria. This ensures that job seekers are aware of new job opportunities as soon as they become available. In conclusion, a job search engine is an excellent tool for simplifying the job search process. It provides job seekers with access to numerous job opportunities in one place, saves time, and offers customized job alerts. With these benefits, job seekers can find their dream job more efficiently and effectively.

2.4MODULE 4: USER INTERFACE

The user interface (UI) of a job portal plays a crucial role in providing a seamless and enjoyable experience for job seekers. It encompasses all the elements that a user interacts with while searching for job opportunities on the platform, including the website's layout, design, and navigation features. A well-designed UI should be intuitive, easy to use, and visually appealing toattract and retain users. Job seekers should beable to easily find relevant job listings and filterthem based on their preferences, such as location, industry, experience level, and job type. The UI should also include clear and concise job descriptions that provide detailedinformation about the job responsibilities, requirements, and qualifications. Another important aspect of the UI is the registrationand profile creation process. This should be a seamless experience for users, with minimal input required to complete the process. The platform should also offer features for users to customize their profiles, upload their resumes, and track their application status. Additionally, the UI should provide easy access to the platform's support and helpfeatures, such as FAQs, chat support, or emailsupport. This ensures that users can quickly and easily get assistance whenever they needit. Overall, a well-designed UI for a job portalis essential to creating a positive user experience, attracting, and retaining jobseekers on the platform.

2.5 MODULE 5: JOB RECOMMENDATIONENGINE

A job recommendation engine is a technological tool that utilizes machine- learning algorithms to provide personalized job suggestions to job seekers based on their preferences and behavior. The engine can analyze job seeker's search history, resume, and other user behavior to provide the most accurate job recommendations. It can also match job descriptions and requirements with the job seeker's skills and qualifications to find the best-suited job opportunities. To be effective, the job recommendation engineshould be reliable, accurate, and continuously updated to adapt to changes inthe job market and user behavior. The enginecan also benefit recruiters and employers by providing a pool of qualified candidates, thereby enhancing the efficiency of the recruitment process. Overall, the job recommendation engine can significantly improve the job search experience for job seekers and streamline the recruitment process for employers and recruiters.

2.6 MODULE 6: ANALYTIC ENGINE

The analytic engine is a versatile and effectivetool that enables businesses to process large volumes of data and derive valuable insights to gain a competitive edge in various industries such as finance, healthcare,marketing, and e-commerce. By utilizing complex algorithms and data models, the engine can analyze both structured and unstructured data in real-time, allowing organizations to respond quickly to changes inthe market and customer behavior. In addition to identifying patterns, trends, and correlations, the analytic engine can also optimize operations, enhance efficiency, and reduce costs by identifying areas for improvement. It can further assist in predicting future trends and events, enabling organizations to proactively adjust their strategies and operations accordingly. However, to harness the full potential of the analytic engine, a robust infrastructure and a skilled team of analysts are necessary to interpret the data and provide actionable insights.Overall, the analytic engine is a valuable tool that empowers organizations to make data- driven decisions and improve operationalefficiency, ultimately leading to a competitive advantage.

www.ijprems.com
editor@ijprems.com

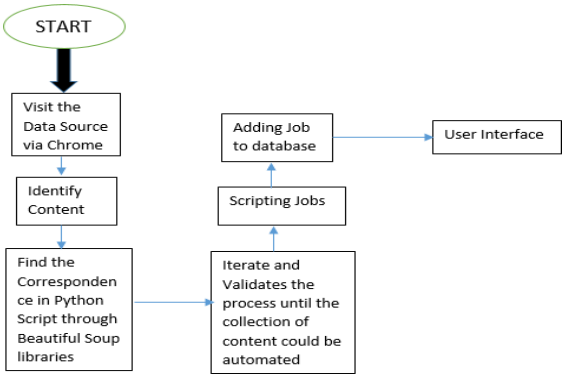
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 05, May 2023, pp : 267-271

e-ISSN : 2583-1062

Impact Factor : 5.725

4. WORKFLOW DIAGRAM



5. RESULTS AND DISCUSSION

By using web-scraping techniques, job portals can improve the accuracy of their job listings data. Web scraping can automate the process of collecting job listings data and reduce the risk of human errors that can occurduring manual data entry. Sometimes there will be some difficulty in login page due to heavy traffic when used by multiple users. Users can face difficulty were they wont get the suitable jobs when not specified correctly. In that case, we are providing multiple job options were the user can explore by filtering according to the skills andqualification.

6. SCREENSHOTS

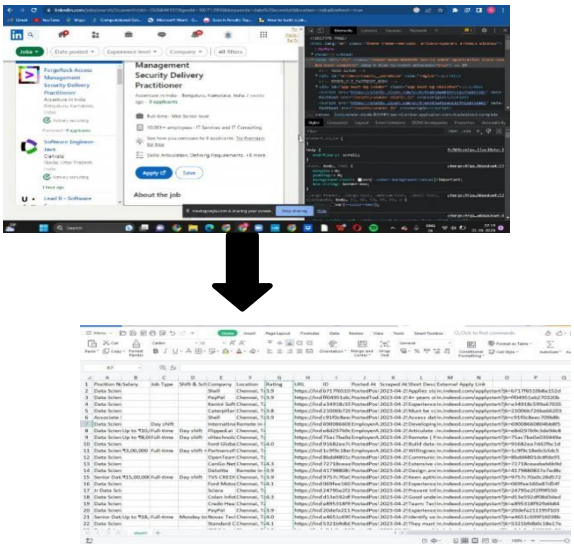


Figure 8.1 Extracting Information from various websites and stored in excel sheet

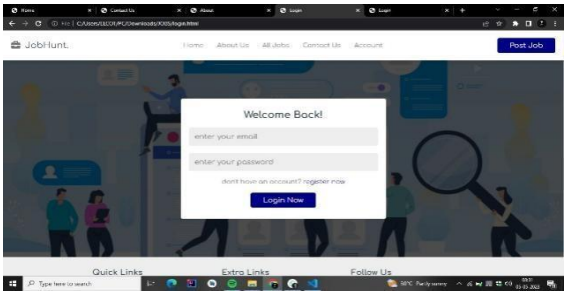


Figure 8.2 login Page

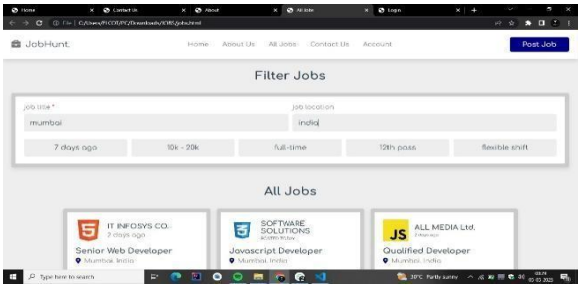


Figure 8.3 Filtering the Job Options

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 05, May 2023, pp : 267-271

e-ISSN : 2583-1062

Impact Factor : 5.725

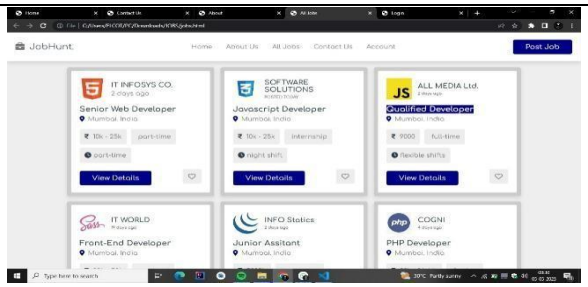


Figure 8.4 Showing Various Job Opportunities

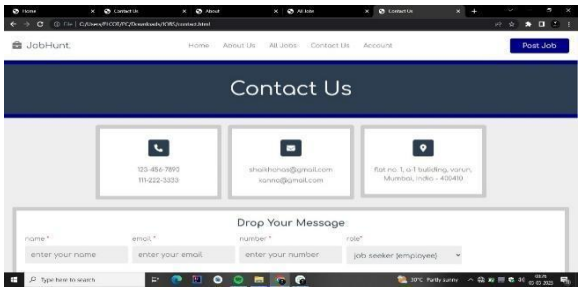


Figure 8.5 Entering the details of the User

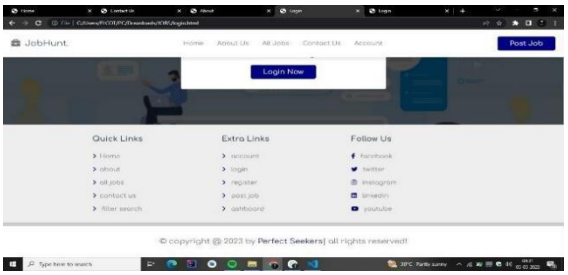


Figure 8.6 Giving Various options like job posting,checking user account, filtering the jobs

7. CONCLUSION

Our scraping model streamlines the job searchprocess, allowing candidates to save valuable time. Through automation, we have eliminated the need for manual searching andmade it easier for users to connect with potential employers. We faced several challenges while developing the model, particularly in understanding and implementing scraping concepts. However, weovercame these obstacles by leveraging the Python-based Beautiful Soup (bs4) library, which enabled us to scrape job listings from multiple portals (such as Assan Jobs and Fresher's World) and display them on our platform. One of the key features we added is email notifications, which was a major challenge during the development process. We addressed this challenge by implementinga notification model using the SMTP protocol and a scheduler. Registered users can take advantage of several benefits, including a simple registration process, easy access to the best job opportunities, daily email notifications, and the ability to update their job profiles as needed. Our platformstreamlines the registration process on multiple job portals, saving users significant time and effort.

8. REFERENCES

[1] Ghazvinian,Holbert, Viswanathan. "Simple Web Scraping. "Internet:https://seanholbert.wordpress com/2011/07/15/scrappy-simple-webscraping/, Jun.2015"

[2] S. d. S. Sirisuriya, "A comparative study on web scraping," 8th International Research Conference, KDU, p. 135–140, November 2015.

[3] R. B. Mbah, M. Rege, and B. Misra, "Discovering job market trends with textanalytics," in 2017 International Conference on InformationTechnology (ICIT).

[4] S. Munzert, C. Rubba, P. Meißner, and D. Nyhuis, Automated datacollection withR: A practical guide to web scraping and text mining.John Wiley & Sons, 2014.

[5] J. Ward, Instant PHP web scraping. PacktPublishing Ltd, 2013.

[6] F. Suleman, "The employability skills of higher education graduates: insights into conceptual frameworks and methodological options

[7] A. Radermacher and G. Walia, "Gaps between industry expectationsand the abilities of graduates," in Proceeding of the 44th ACM technicalsymposium on Computer science education, 2013