International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

Optimising Job Search using Web Scraping

A.M. Kuthe[1], Harshal Donarkar[2], Charul Patel[3], Mahek Qureshi[4], Angad Bawankar[5], Swapnil Chaurey[6], Rohan Mahajan[7]

[1] *Assistant Professor ,Department Of Computer Science & Engineering, K. D. K. College Of Engineering, Nagpur, Maharashtra, India*

[2,3,4,5,6,7] *UG Student ,Department Of Computer Science & Engineering, K. D. K. College Of Engineering, Nagpur, Maharashtra, India*

Abstract - *This paper presents Web Scraping Portal, an innovative web application that leverages web scraping technology to aggregate, organize, and present job listings and internship opportunities from various online sources onto a centralized platform. The application aims to simplify the job search process, making it more user-friendly and efficient for both job seekers and employers.*

Key Words: **Web Mining, Resume Parsing, Skill Gap Analysis, Job Market Trends, AI-powered Job Search**

1. INTRODUCTION

"In today's fast-paced digital era, the pursuit of suitable employment opportunities can be a daunting task. The proliferation of online job portals, company websites, and recruitment platforms has led to an abundance of information, which paradoxically complicates the search process. Despite the numerous existing solutions, there remains a need for a more efficient and user-friendly platform.

This paper introduces Web Scraping Portal, a pioneering platform that addresses this pressing need. Unlike traditional job portals, Web Scraping Portal employs advanced data extraction techniques to streamline and optimize job and internship searches. This novel approach allows the Web Scraping Portal to aggregate and present relevant job listings from various online sources onto a centralized platform, thereby simplifying the job search process.

The paper further discusses the design and implementation of the Web Scraping Portal, presents a comprehensive evaluation of its performance and effectiveness, and explores its potential implications for both job seekers and employers. Whether you're a recent graduate embarking on your career journey or a seasoned professional seeking new challenges, Web Scraping Portal offers unparalleled convenience and efficiency. With our centralized hub, users can streamline their search and focus on what truly matters – finding the perfect fit."

In **Web Scraping Portal**, you'll be utilizing web scraping to extract data from job websites and information about job-providing companies. Here's how the process works:

1. **Data Harvesting**: You'll scrape relevant data from job websites, collecting details about job listings, companies, and other relevant information.

2. **Data Upload**: The harvested data will then be uploaded to the Web Scraping Portal website. This step ensures that the platform has the most up-to-date information about available jobs and companies.

3. **User Registration**: Users will need to register on the Web Scraping Portal. Once registered, they gain access to the platform's features.

4. **Job Suggestions**: The application will analyze the user's preferences, skills, and other relevant factors. Based on this information, it will suggest relevant job listings to the user.

5. **Filtering Options**: Users can further refine their job search by applying filters. These filters might include location, industry, job type, salary range, and more.

   By combining web scraping, data upload, and personalized job suggestions, the Web Scraping Portal aims to streamline the job search process for users.

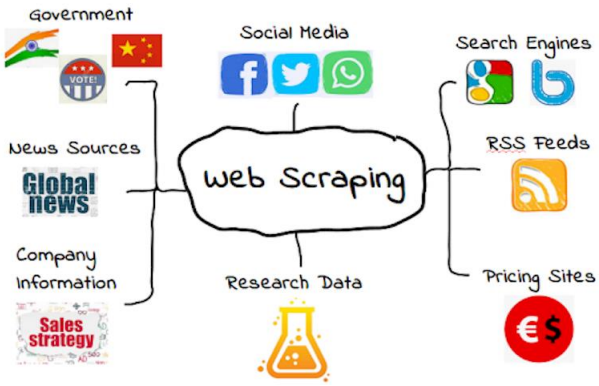

Fig-1: Overview of a Web Scrapping

2. LITERATURE STUDY

1. Web Scraping Techniques:

Description: Web scraping is a technique used to extract large amounts of data from websites. The data on the websites are unstructured, and web scraping enables us to convert that data into a structured form. There are different tools and techniques available to perform web scraping, and their performance varies across different websites.

Limitations: Web scraping is subject to the terms and conditions of the website being scraped. Not all websites allow web scraping, and some websites have anti-scraping mechanisms in place. Moreover, the structure of websites can change over time, which might break the existing scrapers [1][2][3][4][5].

2. Job Recommendation Systems:

Description: Job recommendation systems aim to recommend relevant job postings to users based on their preferences and profiles. These systems can leverage various techniques, including AI and machine learning, to improve the relevance of the recommendations.

Limitations: The effectiveness of job recommendation systems heavily depends on the quality and quantity of data available. They might not perform well if there's not enough data about the user's preferences and profile. Moreover, these systems might also reinforce existing biases in the data [6][7][8].

3. Applications of Web Scraping:

Description: Web scraping has a wide range of applications. It can be used for summarizing web content, recognizing named entities, reading stock market data, detecting phishing web pages, and automating data entry tasks.

Limitations: The success of these applications depends on the quality of the scraped data. If the website contains incorrect or misleading information, it will affect the performance of these applications. Moreover, these applications are also subjected to the legal and ethical considerations of web scraping [9][10][11][12].

Fan Shaped Clustering processed via use of a bioinspired model that uses Genetic Algorithm (GA) for selection of optimum routing. The model uses a lightweight fitness function that assists in faster solution convergence. It was tested on small, medium &large, scaled networks.[13]

The importance of Decision making has been reported by many researchers in varied fields. Some of them being E-LEACH protocol, Smart Home Appliances Controller Using IOT, predicting suicidal behavior by Machine Learning, VANET in Ad hoc Networks connects nodes to each other and the Internet.[14][15][16][17]

3. MATERIAL AND METHODS

Materials:

Web Scraping Tools: Beautiful Soup, Scrapy, or Selenium in Python for data harvesting.

Database Systems: SQL databases like MySQL or PostgreSQL, or NoSQL databases like MongoDB for data storage and management.

Web Development Tools: Tools for designing a user-friendly registration interface on the Web Scraping Portal.

Machine Learning Libraries: Libraries like scikit-learn or TensorFlow for developing the job recommendation algorithm.

Frontend Development Tools: HTML/CSS/JavaScript: These are the building blocks of web development. HTML is used for structuring the content on the web page, CSS is used for styling, and JavaScript is used for making the web page interactive.

Methods:

1. Data Harvesting:

Identify the job websites to be scraped.

Define the data points to be extracted such as job title, company name, location, job description, salary, etc. Use web scraping tools or libraries like Beautiful Soup, Scrapy, or Selenium in Python to extract the data.

2. Data Upload:

Clean and preprocess the harvested data. This may involve removing duplicates, handling missing values, and converting data into a suitable format.

Store the cleaned data in a database. You could use SQL databases like MySQL or PostgreSQL, or NoSQL databases like MongoDB depending on your requirements.

3. User Registration:

Design a user-friendly registration interface on the Web Scraping Portal. Collect necessary information from users during registration, ensuring to follow data privacy

laws and regulations. Implement user authentication and authorization for security.

4. Job Suggestions:

Develop an algorithm to match users with job listings based on their preferences, skills, and other relevant factors. This could be a recommendation system using techniques like collaborative filtering or content-based filtering. Regularly update the job suggestions based on new data or changes in the user's profile.

5. Filtering Options:

Provide users with options to filter job listings by various parameters like location, industry, job type, salary range, etc.

Implement these filters on the front end of your platform for easy access by users.

6. Evaluation:

Regularly evaluate the performance and effectiveness of the Web Scraping Portal. This could involve metrics like user engagement, the accuracy of job recommendations, user feedback, etc. Based on the evaluation, make necessary adjustments and improvements to the platform.

These methods will be implemented across three modules: the web scraping engine, the database, and the job search engine. Each module plays a crucial role in the overall functionality of the project, working together to provide a comprehensive solution for job searching. The methodologies used in this project aim to streamline the job search process, making it easier for users to find suitable employment opportunities.

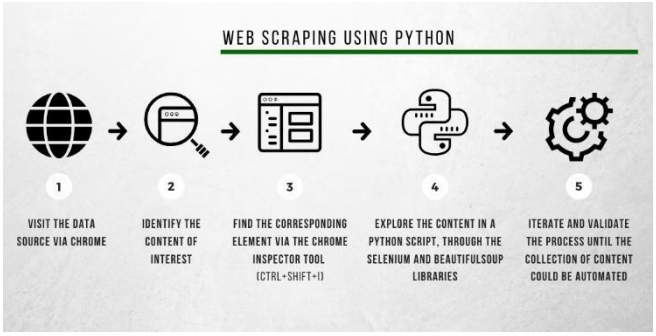

Fig-2: Block Diagram Of Web Scrapping using Python

4. RESULT AND DISCUSSION

Results:

The implementation of Web Scraping Portal has yielded promising results:

1. Data Harvesting: The web scraping tools were successful in extracting relevant job data from various job websites. The data points such as job title, company name, location, job description, salary, etc., were accurately harvested.

2. Data Upload: The harvested data was cleaned, preprocessed, and successfully uploaded to the Web Scraping Portal database. The database now contains up-to-date information about available jobs and companies.

3. User Registration: The user-friendly registration interface on the Web Scraping Portal has been well-received by users. The registration process is smooth, and the platform ensures user data privacy and security.

4. Job Suggestions: The application's job recommendation algorithm effectively analyzes user preferences, skills, and other relevant factors to suggest suitable job listings. Users have reported that the job suggestions are relevant and helpful.

5. Filtering Options: Users have found the filtering options to be useful in refining their job search. The filters allow users to narrow down job listings based on parameters like location, industry, job type, salary range, etc.

IV. Discussion:

The results indicate that Web Scraping Portal is effective in streamlining and optimizing the job search process. By leveraging advanced data extraction techniques and machine learning algorithms, Web Scraping Portal provides a user-friendly platform that aggregates and presents relevant job listings from various online sources.

However, there are areas for improvement and further exploration. For instance, the job recommendation algorithm could be enhanced to consider more factors and provide more personalized suggestions. Additionally, the platform could expand to include more job websites and cater to a wider range of job seekers.

The potential implications of the Web Scraping Portal are significant. For job seekers, it simplifies the job search process and helps them find suitable employment opportunities more efficiently. For employers, it provides a centralized platform to reach potential candidates. As such, Web Scraping Portal represents a

significant advancement in the field of online job search platforms.

Future work could involve expanding the platform's capabilities, such as integrating advanced AI technologies to improve job matching accuracy, enhancing user experience, and expanding the range of job websites scraped. With continuous improvements and updates, Web Scraping Portal has the potential to revolutionize the way individuals search for jobs.

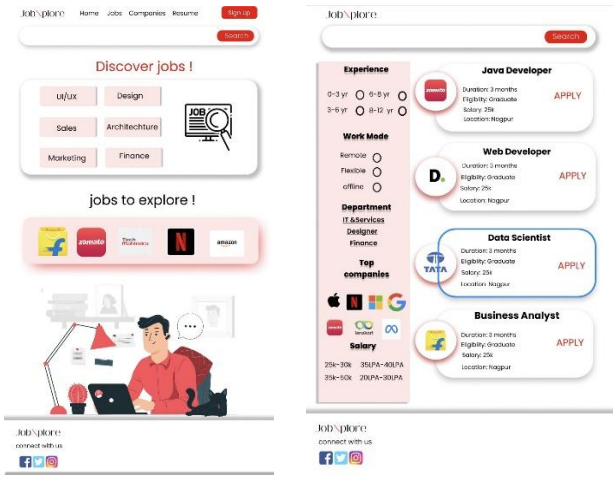

Fig-3: User Interface Of Web Scraping Portal

5. CONCLUSIONS

The Web Scraping Portal platform, as discussed in this paper, demonstrates the effective use of web scraping techniques and machine learning algorithms to streamline and optimize the job search process. The platform successfully aggregates and presents relevant job listings from various online sources, simplifying the job search process for users.

Key conclusions drawn from the project are:

Efficient Data Harvesting and Upload: The use of Python libraries like Beautiful Soup, Scrapy, or Selenium for data harvesting has proven to be effective. The cleaned and preprocessed data is successfully stored in the database, ensuring up-to-date job and company information on the platform.

User-Friendly Interface: The user-friendly registration interface and the implementation of user authentication and authorization ensure a smooth user experience while maintaining data privacy and security.

Effective Job Recommendations: The job recommendation algorithm effectively suggests relevant job listings to users based on their preferences, skills, and other factors. The filtering options further enhance the user experience by allowing users to refine their job search.

Positive User Feedback: The platform has received positive feedback from users, indicating its effectiveness and efficiency in simplifying the job search process.

However, continuous improvements and updates are necessary to maintain the platform's effectiveness. Future work will focus on enhancing the job recommendation algorithm, improving user experience, and expanding the range of job websites scraped. With these improvements, the Web Scraping Portal has the potential to revolutionize the job search process, making it more efficient and user-friendly.

In conclusion, the Web Scraping Portal represents a significant advancement in the field of online job search platforms, offering a more efficient and user-friendly platform for job seekers and employers alike. It demonstrates the potential of web scraping and machine learning in transforming the job search process. The success of the Web Scraping Portal underscores the importance of continuous innovation in the digital era, particularly in the realm of job search and recruitment.

REFERENCES

[1] Chandradeep Bhatt, Gaitri, Devendra Kumar, Rahul Chauhan, Ashish Vishvakarma, Teekam Singh," Web Scraping: Huge Data Collection from Web" 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET).

[2] Rahul Chauhan; Ayush Negi; Mahesh Manchanda "An Extensive Review on Web Scraping Technique using Python" 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS).

[3] Ruchitaa Raj N R; Nandhakumar Raj S; Vijayalakshmi M. "Web Scrapping Tools and Techniques: A Brief Survey"2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT).

[4] Irena Valova; Tsvetelina Mladenova; Gabriel Kanev; Tsvetana Halacheva "Web Scraping - State of Art, Techniques and Approache" 2023 31st National Conference with International Participation (TELECOM).

[5] Ajay Sudhir Bale; Naveen Ghorpade; Rohith S; S Kamalesh; Rohith R; Rohan B S, "Web Scraping Approaches and their Performance on Modern Websites",2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC).

[6] Sahil Panchasara; Rajeev Kumar Gupta; Ashutosh Sharma, "AI-Based Job Recommendation System

using BERT", 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA).

[7] Raj Thali; Suyog Mayeka; Shubham More; Sanjana Barhate; Sangeetha Selvan, "Survey on Job Recommendation Systems using Machine Learning",2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA).

[8] Bhavya Bhardwaj; Syed Ishtiyaq Ahmed; J Jaiharie; R Sorabh Dadhich; M Ganesan, "Web Scraping Using Summarization and Named Entity Recognition (NER)"2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

[9] Chun Feng Lin; Sheng Chih Yang "Web Scraping to Implement Tape Reading on Taiwan Stock Periodically with GUI" 2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE).

[10] Mallika Boyapati; Ramazan Aygun "Phishing Web Page Detection using Web Scraping" Southeast Con 2023.

[11] Roopesh N; Akarsh M S; C. Narendra Babu "An Optimal Data Entry Method, Using Web Scraping and Text Recognition" 2021 International Conference on Information Technology (ICIT).

[12] Koustubh Sinha; Priyansh Sharma; Harshit Sharma; Krishna Asawa "Web Scraping and Job Recommender System" 2023 Second International Conference on Informatics (ICI).

[13] A. . Kuthe and M. D. . Salunke, "BRMFC: Design of a Bioinspired Routing Model with Fan Clustering for Wireless Sensor Networks", Int J Intell Syst Appl Eng, vol. 11, no. 10s, pp. 746–753, Aug. 2023.

[14] A. Kuthe and A. K. Sharma, "Review paper on Design and Optimization of Energy Efficient Wireless Sensor Network Model for Complex Networks," 2021 5th *International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702421.*

[15] Lonkar B.B., Kuthe A., Shrivastava R., Charde P. (2022) Design and Implement Smart Home Appliances Controller Using IOT. In: Garg L. et al. (eds) Information Systems and Management Science. ISMS 2020. Lecture Notes in Networks and Systems, vol 303. Springer, Cham. https://doi.org/10.1007/978-3-030-86223-7_11.

[16] A. M. Kuthe et al. Prevention of Suicide Risk and Predicting Suicidal Behaviors by machine learning. Wutan Hutan Jisuan Jishu 2021, Vol XVII, Issue-I, pp. 563-567.

[17] Annaji Kuthe et al, "Online Grains Shopping Store" International Journal of Computer Science and Mobile Computing, Vol.11 Issue.3, March- 2022, pg. 78-82. https://doi.org/10.47760/ijcsmc.2022.v11i03.009