

# Job Scraper: Digitizing Your Job Findings

Rishika Vaidya<sup>1</sup> Shikha Thakur<sup>2</sup> Nidhi Laad<sup>3</sup> Srishti Laad<sup>4</sup>

<sup>1,2,3,4</sup>Department of Information Technology

<sup>1,2,3,4</sup>Acropolis Institute of Technology and Research, India

**Abstract** — This project solely focuses on solving the paramount need of new age that is job finding by connecting the employers with seekers with astonishing a technology. It works in mainly two parts scraper and crawler each doing its part with the help of python libraries. It closely relates to web indexing an information extracting technique used by multiple search engines to index data on web using human programmed bots. Currently there are multiple software gadgets available that aim to apply scraping techniques to personalize your website.

**Keywords:** Job Scraper, Job Findings,

## I. INTRODUCTION

Scraper is a tool which is being used for gathering job postings information online in a programmatic manner. This automated way of extracting data from web helps people get job data efficiently and build a resourceful job database by integrating various sources into one. Data scraping is a process of importing data from websites into files or spreadsheets. It is used to extract data from the web, either personal use by the scraping operator or to reuse the data on other websites. There are numerous software applications for automating data scraping. The online job market has many person hiring activities. This is especially true when most cities around the globe face rounds of lock down and more job shifts to a remote mode since covid-19. In this sense scraping tools helps not only institutions and organizations but also job seekers.

According to reports made by gallop 51% of employees keep an eye on new opportunities online recruiting resources give rise to businesses of job boards and aggregator websites, this kind of aggregator website is really making money.

### A. Job Data Uses:

- There are numerous uses of job data in today’s market by employees and as well as employers it serves as basic data for job market survey. It provides data for aggregators to fuel their research.
- It collects data for analyzing trends and contemplating over labour market by both employees and employers.
- It tracks competitors open position and compensation so that a seeker can simply get a leg up in the competition.

- Staffing agencies scrape job boards to keep their job databases up to date.

This is just tip of iceberg and it is evident from above data that the job scraper simply is more relevant with todays need and is very effective in building a resourceful job database by integrating various data sources into one through erudite libraries of python.

### B. Methods to Scrape Jobs:

There are numerous methods which can be simply used for scraping job data from various websites and there are copious libraries in python which are very serviceable for this some of which are:

- Using a scraping tool: Technologies has been advancing and so are our needs given this we can now simply automate the scraping process. There are many software’s that are designed so that even non-technical people can fetch data from web. These are budget friendly and quite easy to use.
- Hiring a scraping service: There are certain companies which provides what is generally known as “scarping services”, some of which are Scraping hub, Data hen etc. They work in a customized way and take your requests into account to get the suitable results for you as per need.
- Your own job scraper: doing web scraping in house with your own scraper and resources comes with lots of perks. Job scraping is a niche process that requires high technical skills so if you are using this method you need to be savvy in this field.

### C. Challenges:

Although web scraping can be extremely useful but it can frazzle you out because of the challenges that lies in journey.

- Gathering job data from multiple sources can be very risky and arduous because of the privacy policy of many aggregator sites such as indeed LinkedIn, GitHub etc.
- Your IP may get banned by frequently visiting these sites as it breaches their terms of safety.
- Depth layer searching is tough.
- The anti-scraping techniques are quite popular always implemented by large websites to prevent scraping bots from collecting information from then.
- High cost for job crawler building and maintaining.

## II. LITERARY SURVEY

This is comparison between different python libraries used in this project.

Factors	Beautifulsoup	Scrapy	Selenium
Extensibility	Suitable for low-level complex projects	Best choice for large or complex projects	Best for projects dealing with core JavaScript
Performance	Pretty slow compared to other libraries while performing a certain task	Rapid processing due to the use of asynchronous system calls	Can handle up to some level but not as much as Scrapy

Ecosystem	It has a lot of dependencies on the ecosystem	It has a flexible ecosystem making it easy to integrate with proxies and VPNs	It has a good ecosystem for the development
-----------	---	---	---

– All the previous solutions proposed

Name of proponent	Solution proposed
RENITAL crystal Pereira et al	Offered a description of online scraping approach and tools.
Kaushik Parikh	Offered a description of web scraping detection using machine learning.
Sameer Padghan	Proposed a method for extracting data from online pages in order to make scraping easier
Anand Sircar	Offered web scraping is a new approach

### III. OUR SOLUTION

In this project we simply tackle our current problem of job findings and mainly focus on its digitisation we along with job scraper also focus on time scheduling as it can become frustrating at times to regularly scrape the data along with it we work on periodic scheduling with an appraiser with simply works as a very useful reminder in a occupied day of a student.

In the above proposed solutions jobs were not the soul focus but rather just promotion of web scrapping but our project solely focusses on job scraping.

The main theme of our project varies from the previously proposed solutions as it is student oriented rather than mass as a paramount.

#### REFERENCE

- [1] S.D.S. Siri Suriya, "A comparative study on web scraping," 8th International Research Conference, KDU, p. 135-140, November 2015.
- [2] R.B. Mbah, M. Rege, and B. Misra, "Discovering job market trends with text analytics," in 2017 International Conference on Information Technology (ICIT).
- [3] S. Munzert, C. Rubbia, P. Mainer, and D. Nyhuis, Automated data collection with R: A practical guide to web scraping and text mining. John Wiley & Sons, 2014.
- [4] J. Ward, Instant PHP web scraping. Pact Publishing Ltd, 2013.
- [5] SOURCE: <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>
- [6] L, Richardson, "Beautiful soup," Jan 2020.