Expediting HR Management Via Dynamic E-Portfolio by Employing Web Scrapper

Karthikeyan A[a], Mohammed Abdullah[b], B Jeevan Akshay[b], Kevin Christopher A[b]
[a]Assistant Professor, Dept of CSE, Panimalar Engineering College, Chennai, India
[b]UG Scholar, Dept of CSE, Panimalar Engineering College, Chennai, India

ABSTRACT:
Be it promotion or recruitment, HR management is all about giving the right people the right role. Resume printed on paper can serve as an excellent medium to communicate about oneself to the job recruiter. These days jobs with quality income require the candidates appearing for the interview to apply online. However, portfolio hosted on web can aid a student to express projects and achievements. The samples of projects can be displayed through the website.
To make portfolio more robust and dynamic with real-time content updates, this paper introduces the concept of employing a web scrapper to fetch project details from websites like GitHub and certificates from certificate providers such as Coursera.
Keywords: Web Scrapper, HR Recruitment, Dynamic Portfolio, Data Extraction

I. INTRODUCTION:

A. SIGNIFICANCE OF PORTFOLIO

When a company realizes the necessity to increase assets in the company for delivering the services or products as per demand, it might consider recruitment of employees as one of the options to fulfil the need. Thus, recruitment in such situations is necessary to be quick and only the highly qualified candidates are expected to be hired.
In such a scenario, to eliminate or to select a candidate, resume is the first step for shortlisting. Most companies expect a resume to be a tailored work of presenting oneself suitable for the job. This has a caveat of ignoring any achievements outside the job profile. On the other hand, portfolio is meant to be more descriptive about the candidate. The achievements of candidates might present the viewer of portfolio [1] with useful information such as what new things were learned and how issues were resolved during projects or events. This helps in standing out of the crowd, giving the HR strong motive to make decision on selecting the candidate [2].
To make the process of generating a portfolio more dynamic, web scraper can be employed. Web scrapper is basically a piece of code which runs to extract data [3] from the World Wide Web and saves it to a file system or database the purpose of retrieving or analyzing it at later stages.

B. WEB SCRAPPING:

The process of scraping data from the Internet can be divided into two sequential steps; acquiring web resources and then extracting desired information from the acquired data. Specifically, a web scraping program starts by composing a HTTP request to acquire resources from a targeted website. This request can be formatted in either a URL containing a GET query or a piece of HTTP message

containing a POST query. Once the request is successfully received and processed by the targeted website, the requested resource will be retrieved from the website and then sent back to the give web scraping program. The resource can be in multiple formats, such as web pages that are built from HTML, data feeds in XML or JSON format, or multimedia data such as images, audio, or video files. After the web data is downloaded, the extraction process continues to parse, reformat, and organize the data in a structured way [4].

II. EXISTING SYSTEM:

A. PORTFOLIO GENERATORS

Few of the portfolio generators [5] are Adobe Express and Canva [6]. These websites provide a personalized portfolio by following a four-step process. First step is to organize one's samples. This might include assets of previous work and certificates. The second step is to create a layout with predefined impressive and customizable templates. The third step is to choose a theme which can be found in the preset selections. This includes highlighting one's preferred field of work and domain of passion. Finally, one can share the portfolio to be displayed online with anyone across the world.

B. SYSTEMS WITH PORTFOLIO FUNCTIONALITY

On the other side of portfolio generators are social networking sites such as LinkedIn [7], and job seeking sites such as Naukri.com [8], and recruiting platforms such as Superset can function in a similar fashion to display one's portfolio.

C. LIMITATIONS

Any updates made in projects is not reflected back in the portfolio, making the process of maintaining the portfolio as a difficult task. The following is the summary of limitations of existing systems:

- Lack of concurrency in data
- Static Nature of data, i.e., the data is does not change until the user manually makes changes.
- Inextensible, i.e., limited information can be highlighted which is only useful for preliminary stage of filtering the candidates and not at the stage where HR tries to understand strengths of the candidate

III. PROPOSED SYSTEM:

A. CORE SYSTEM

The main concept of our proposed system is to utilize web scrappers to its optimal use. The web scrapper in our system will scan foe two types of websites. One is the website which consists of repositories and artifacts of candidates work. The other type of URL to scan is the certificate credentials link. Examples for the former type of websites, include GitHub and bitbucket.

The web scrapper will scan the website for ReadMe.md files and display the data in the portfolio site. If there are any assets, then it will provide a link to that website. Therefore, this information will provide credibility of the candidate for the project to the HR recruitment manager. The later form of URL's scanned by the web scrapper will result in either responses in the form of images or credential data, which might include but not limited to certificates signed using GPG keys [9] that can be verified using any of the available tools such as gpgtools [10].

B. ARCHITECTURE:

The web scrapper is present in the website, hosted and accessible via internet. When a user views the website, the user's device behaves as a client and the device hosting

the website behaves as a server. During the process of web scrapping, where the web scrapper scans the external websites, our website with web scrapper behaves as a client requesting to the external websites, which behave as servers, for structured responses [11].

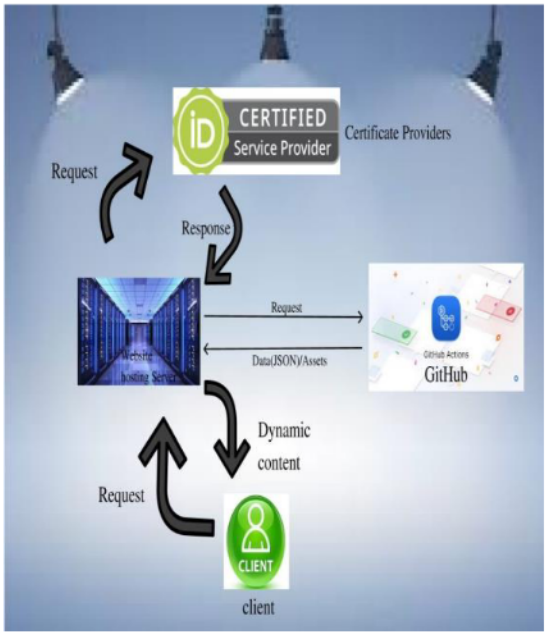

Figure 1. Proposed Architecture

The external websites which our web scrapper will scan, are the certificate providers and websites where one can display their work such as GitHub for developers and dribble for designers. [Figure:1]

C. PROCESS:

The user creates an account if the user is visiting the website for the first time. Otherwise, the user logs in to the website by providing the credentials as username and password. The data is fetched from the database and provided in the dashboard. If the user is new to the website, then the user must fill the data and submit to store in the database. This is where our project proposed that the links which are entered for websites such as GitHub are scrapped for content when anyone visits the portfolio website. Similarly, verification of certification can be accomplished. This is because, the certificate image from certificate providers such as Coursera and Udemy will only be provided in the response when a valid request is made. Similarly for certificate providers such as EdX, credentials that can be verified using a public key will be transmitted to the portfolio website. Thus, each time the user visits the website; the above-mentioned content is fetched dynamically.

D. IMPLEMENTATION

React was used for front-end and firebase was used as BaaS. The login and registration functionality were implemented using firebase. Puppeteer was used to implement web scrapping running on the NodeJS HTTP web server. Data was exchanged in the JSON format unless it was image or other assets. For processing of data, packages were installed using NPM package manager.

IV. FINDINGS:

The following is the comparison of our proposed dynamic e-portfolio against the existing systems with factors dealing with user accessibility and data authenticity:

S. No	Factors	HR Recruitment Sites				
		Proposed System	Existing System			
1.	Type	With Web Scrapper	Portfolio Generator	Social Networking Site	Job Site	Pool Off-campus Sites

2.	Website considered for evaluation	Our implementation as described in proposed system	Adobe Express	LinkedIn	Naukri.com	Superset
3.	UI Customization	×	✓	×	×	×
4.	Dynamic Fetching of data each time the page refreshes	✓	×	×	×	×
5.	Real-time information	✓	×	×	×	×
6.	Automatic updating of data	✓	×	×	×	×
7.	Verification of credentials from certificate providers [12]	✓	×	✓	×	✓
8.	Experience/Recommendations/ comments by oneself/colleagues/ certificate issuer on artifacts/ courses	✓	×	✓	×	×

V. CONCLUSION:

The system makes convenient the process of maintaining the portfolio by maintaining concurrency of data with other websites referenced by the web scraper. The significant issues addressed by this system relate to the following:

- Dynamic fetching of data
- Real-time updates of the information in referenced websites
- Verification of certificates
- Updated information
- No maintenance required, i.e., initial setup of data is sufficient.

REFERENCES:

1. Yan, Li, et al. "Well-designed Electronic Portfolio as a Facilitator for EFL Engineering Learners Reading Skill." 2022 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE). IEEE, 2022.
2. Luchoomun, Dharmadeo, Joe McLuckie, and Maarten van Wesel.

"Collaborative e-Learning: e-Portfolios for Assessment, Teaching and Learning." Electronic Journal of e-Learning 8.1 (2010): pp21-30.

3. G. Jung, S. Han, H. Kim, K. Kim and J. Cha, "Extracting the Main Content of Web Pages Using the First Impression Area," in IEEE Access, vol. 10, pp. 129958-129969, 2022, doi: 10.1109/ACCESS.2022.3229080.
4. Zhao, Bo. "Web scraping." Encyclopedia of big data (2017): 1-3.
5. A. Zunaidah, "Meaningful Online Learning with e-Portfolio: University Students' Perspectives," 2022 8th International Conference on Education and Technology (ICET), Malang, Indonesia, 2022, pp. 228-232, doi: 10.1109/ICET56879.2022.9990059.
6. Gehred, Alison Paige. "Canva." Journal of the Medical Library Association: JMLA 108.2 (2020): 338.
7. Zide, J., Elman, B. and Shahani-Denning, C. (2014), "LinkedIn and recruitment: how profiles differ across occupations", Employee Relations, Vol. 36 No. 5, pp. 583-604. https://doi.org/10.1108/ER-07-2013-0086.
8. Swami, Sanjeev. "Job Search at Naukri. com: Case Study of a Successful Dot-Com Venture in India." Contemporary Research in E-Marketing, Volume 2. IGI Global, 2005. 58-87.
9. Shi, Yingnan, et al. "Knowledge pricing structures on MOOC platform-A use case analysis on edX." PACIS. 2018.
10. Milliff, Aidan. "Data security in practitioner-academic partnerships: an agenda for improvement." (2020).
11. Sirisuriya, De S. "A comparative study on web scraping." (2015).
12. Young, Jeffrey R. "Coursera announces details for selling certificates and verifying identities." chronicle 1 (2013): 14.