2024 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2024

ICode - An Unified Competitive Coding Profile Platform

1st Saurav Muke
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
saurav54muke@gmail.com

2nd Samruddhi Ahire
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
samruddhi3915@gmail.com

3rd Geetanjali Kale
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
gvkale@pict.edu

4th Pranali Navghare
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
prnavghare@pict.edu

***Abstract*—Although a software developer is capable of creating excellent software without the necessity for competitive coding. Having their hands in it may increase the project's efficiency and scalability. For every developer, problem solving is a must-required skill that is gained through competitive coding. Many technical recruiters use competitive programming as a barometer to assess a candidate's potential for the post of software engineer, including many prestigious Silicon Valley companies like Google, Facebook,etc. Since there is no such platform that helps students showcase their coding status on every registered platform, the proposed system helps to have such. ICode is a unified platform to help students visualize their coding profile status on a single site at one go. There is no public API that can be used to retrieve the data.from websites; hence, web scraping becomes relevant in this situation. It is one of the strategies of information retrieval.Instead of manually copying data from the browser, with web scraping, it becomes easy to automatically extract the data.without any effort and perform analysis on it. The paper mainly targets the visualization of the data scrapped from competitive coding platforms to create a digital resume and perform analysis based on the scraped data. This is a digital resume.will be a revolution in the modern-day recruitment process since it will contain real-time user's details and coding platform ratings.and their analysis, project information, and digital bridges to platforms like Github and LinkedIn, where recruiters can actually review the users's information in a better way as compared to the old paper resume.**

***Index Terms*—web scraping, retrieval, websites, API, data, visualisation, data analysis.**

I. Introduction

The Internet is gaining popularity since so many networks contain a huge variety of information, along with a system to access that piece of information with our computer, thus helping the World Wide Web (WWW). Information retrieval is the process of finding the required information from a variety of sources, which may consist of the storage process, representation, organisation, investigation, and access. To process data and retrieve information easily and quickly, people are encouraged to develop new technologies.

Web scraping is the procedure of automatically fetching amorphous data from a website and saving it in an organised format. Web scraping is a technique of information retrieval that includes various methods like traditional copy and paste, HTML parsing, DOM parsing, RegEx, etc.

ICode is a unique and cumulative competitive coding profile generation and digital resume platform. It is accumulating users' information from three competitive programming websites: Codeforces, Codechef, and Leetcode. Along with these profiles, it also analyses data from those platforms, displays project information, and acts as a digital resume with all necessary information.

Rather than carrying paper-based resumes, recruiters and students can now rely on ICode to ease the recruitment process and showcase their real-time skills. Even with the help of ICode, the colleges will be able to keep track of the current ratings of students on the provided platforms. Also, the analysed data will help the college with the growth of students and even plan and take subsequent actions. This will help students to show all the necessary data required on their resume in one place with better visualisation and accessibility of information.

The paper is organised as follows: Section II contains a literature review, which gives a brief idea of the survey done before designing the system about the research papers and existing systems, including a study of drawbacks in the existing system and overcoming them in the proposed system. Section III contains the methodology used for developing the proposed system. Section IV contains the results and analysis of the system.Section V includes the Conclusion, which contains the outcomes that are going to be achieved by ICode.

II. Literature Review

A software developer is capable of building amazing software that may not include any need for competitive coding, but having a hand in it may make the project more efficient

979-8-3503-0661-3/24/$31.00 ©2024 IEEE

and scalable [5]. Coding makes a person a better problem solver. There are over 600,000 registered users on CodeForces, 245877 users on CodeChef, and about 20,000,000 on LeetCode.

It becomes necessary to establish platforms that monitor the ratings of those who perform coding since coding is one of the must-do activities of engineers that assures the engineers a job placement [4]. The existing sites lack the feature of having an overall search button, which would allow one to check anyone's rating while having the key that has been considered in ICode.

If a code is insecure, that is, if it contains errors, buffer overflows, or there is a leak of information through it, then it may prove fatal for an organization. Therefore, it becomes mandatory to develop secure systems that are free from vulnerabilities. As per the research, the exploration of competitive programming data is not tested from the perspective of insecure coding practices. Since competitive programming is a must-do for one to get good recruitment, giving contests on coding platforms and getting high ratings is a desire of every individual, for which several non-permitted actions are also taken, which cause plagiarism to the user's account. According to the research, 34.2% of submissions done on Codechef are vulnerable [6].

Retrieval of data is not possible directly because they do not provide any APIs. Web scraping is considered one solution because it allows the user to retrieve data from public websites that don't have APIs or provide only limited access to the data [1].

The techniques and tools of web scraping are provided by Renita Crystal Pereira. Which was a solution for facing several complexities, as extracting data is not that simple. The collected data is correct, has better integrity, and is consistent, which is guaranteed by these strategies [2]. Even though there are a few problems that are faced by these functional techniques, the measurement level of the scrapper varies with the units of measurement of the original source file, which makes it difficult to interpret the data.

Web scraping is a remedial attack all the time. When the company places its data on the internet, there is probably a high chance of coping with the data and utilising it from other points of view, making the organisation unaware of this task; hence, a lot of protection mechanisms have been considered while introducing the techniques of web scraping [2]. Among the methods of web scraping like traditional copy and paste, HTML parsing, and DOM parsing, one of the most popular is web scraping through RegEX, which is also called regular expression. It is used to match patterns, and a library in Python called 're' is used to perform this match pattern task [8].

The solutions to the problem of web scraping are essentially aimed at transforming the highly complex data obtained through websites into a structured format that would be stored in a central database and examined well. As per the survey, web scraping solutions have a remarkable impact on the result of the cause. The approach of extracting the data easily from the web pages with web scraping was projected by Sameer Padgham. This method proves time-saving, enhances the quality of data significance, and also helps to scrape data from numerous websites with a minimum amount of human interference [2]. Legally, the scraped data may be useful for the development of databases, research purposes, and many similar activities. However, illegal scraping can be stopped by using safe web scraping and effective methods. Web scraping can be considered a blessing to the human race and, if used wisely, can help in their advancement.

This sector effectively advances a shared aim with an ambitious project that still needs innovations, conceptual comprehension, the mission of object models, artificial intelligence, and communication between humans and computers. Through web scraping, extracting the data dynamically or gathering world-wide web information becomes easy.

WebScrapingAPI can be called a Rest API, which helps the user scrape any HTML page while handling javascript rendering, IP rotations, and CAPTCHA. APIs like Zenscrape, ScraperAPI, ScrapingBee, ScrapingAnt, Simplescraper, Scraper's Proxy, ScarpeGoat, and ScrapingMonkey are some of the WebScrapingAPI [3]. Sometimes these APIs are also called web crawlers, which are used to 'scrape' data from the data available on public websites from the Internet. A good example of this would be Google using it to determine all the results of searches. Web developers who wish to use data from the available websites are the perfect contenders to use these APIs.

The need for Web ScraperAPI is supreme because it allows developers to compile all the existing data sets into a single resource, which accordingly eliminates the high-priced replication of effort. All these APIs are subsidised and drawn parallel in numerous developer programming languages and SDKs like Node.js, PHP, Python, Ruby, Java, C#,.Net, and Curl [7].

Scraping data from websites available on the Internet is legal, as long as the scraped data is not used for any malicious purpose or to attack the scraped website's organisation directly. To avoid these harmful uses, some protection regulations are provided in countries like California and European Union states. Yet it is not provided in the US, but the union of various laws and regulations at the state level often protects it. Therefore, it becomes mandatory to not scrape personal, recognisable information. Even if it is done, businesses can mask and protect their organisations through data-enhancing technologies [7].

Electronic human resource management can improve the efficiency and effectiveness of the recruitment and selection processes [9]. Recommendations for digital resume design include using clear and concise language, formatting the resume in a consistent manner, and tailoring the resume to a specific job opening [9].

The data must be scraped at a reasonable rate in order to avoid overloading the website's server. The copyrighted material must not be scraped, and the data should be handled in a responsible and ethical way [10].

III. Methodology

Competitive coding is now one of the most important aspects of the lives of students for cracking placements. Showcasing the coding profiles on different websites becomes a bit tedious task. Also, from an institutional point of view, keeping track of these details becomes impossible. This problem is solved by bringing an interface that will show all the details of the user in one place and generate a profile that will be showcased to everyone.

The major objectives of ICode are:

1) To provide a unified platform for competitive coders: ICode is a unique platform that aims to generate a profile of a user containing current ratings at Codechef, Codeforces, and Leetcode in a single place.
2) To showcase the user's profile to their peers, which will include: With the help of a username, the registered users can view their profile as well as their peer's profile by entering it in the search user feature provided in ICode.
3) To save the user's time and effort when visiting a specific website and checking for a profile, the user gets all the information. at once on ICode, thus saving time and effort.
4) To provide a platform where colleges can keep track of the coding status of students and encourage them to do so.
5) To create a digital resume platform for coding-related showcasing of skills.

In order to achieve this, Icode was developed. It uses the user's information to generate a digital resume that includes real-time coding data that is retrieved from relevant websites, projects and any live links that may be available, portfolio websites, and other websites that are readily accessible, such as GitHub, LinkedIn, etc. Anyone can use the unique ID they created during registration to search the user's profile. Anybody who requests to view the user's profile can obtain it by sharing his profile ID. Likewise, the institution has the ability to log in and keep an eye on the students that are affiliated with it. Additionally, the institute will have access to the generated analysis of the student body that is available. There is no direct API available to retrieve user data from the coding website for display purposes. These are public websites where anyone can view the profile directly; however, each platform requires a unique username. Therefore, "web scraping" technology is used to retrieve the information.

Fig. 1 shows the system architecture of the proposed system. The system mainly has three major modules:

A. Web Application

Users engage with the system through the web application or web interface that is accessible to them. The web application is developed using ReactJs, a JavaScript native library known for crafting intuitive user interfaces, along with Bootstrap, an open-source CSS repository that contributes to the application's visual aesthetics and responsiveness.

The user interface is designed with simplicity in mind, allowing users to seamlessly navigate through various features. Notable functionalities include Search User, Login, Register, Update Profile, and Dashboard. Users can effortlessly access these features through straightforward navigation within the interface.

This web application is designed to ensure compatibility with a wide range of browsers that support JavaScript. Users can conveniently interact with the system, enjoying a user-friendly experience facilitated by the combination of ReactJs and Bootstrap.

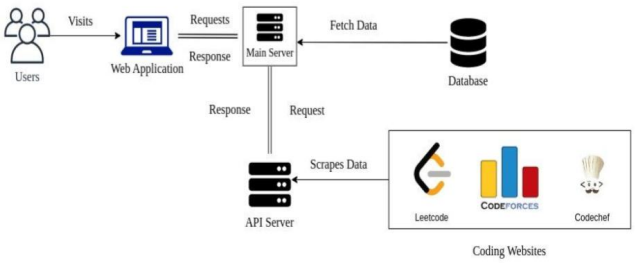

Fig. 1. System Architecure

The digital profile of the user is showcased through this web application. Different users will use different parts of the web application according to their roles and priorities. The systems are used by three different types of users. They are listed below:

1) Students and Developers: Students and developers are the users who will be creating their profiles to showcase real-time stats. First, the user needs to register by filling out certain details, as shown in Fig. 2. Once the user is registered, the unique ID generated can be shared across all users in order to showcase the profile. Other users and recruiters can see the profile by using the same unique ID.

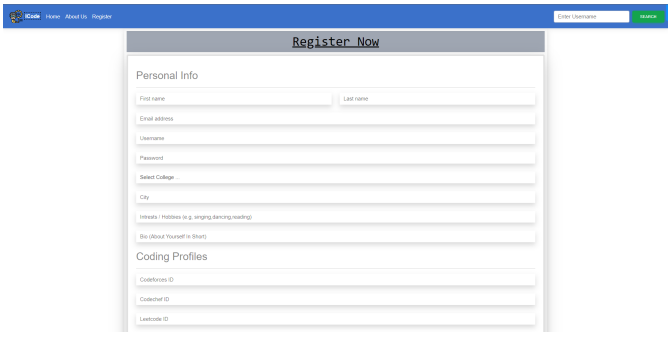

Fig. 2. User Registration

2) College: College administrators will be able to keep track of the students registered in their college. First, the college needs to login with the login credentials provided. After successfully logging in, they will be redirected to the College Dashboard as shown in Fig.3, where they will be able to see the list of students in their college registering on respective platforms. Also,

the rank-wise list will be showcased, and the placement status of the students will be indicated in the dashboard.

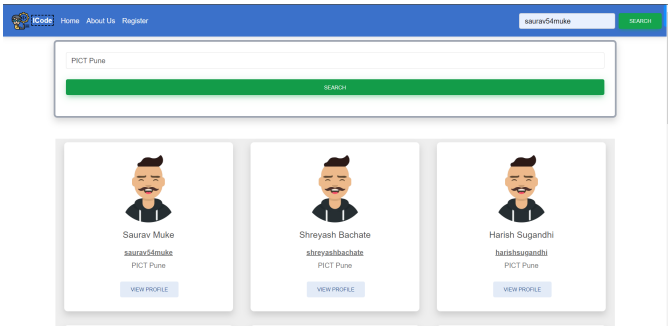

Fig. 3. College Dashboard

3) Recruiters: Recruiters will see the profile of the applicant and check the real-time data. With the help of the username of the user, recruiters can see the real-time coding status, as shown in Fig. 4, by just searching the username in the search bar provided. If the username is proper, he or she will be redirected to the profile page.

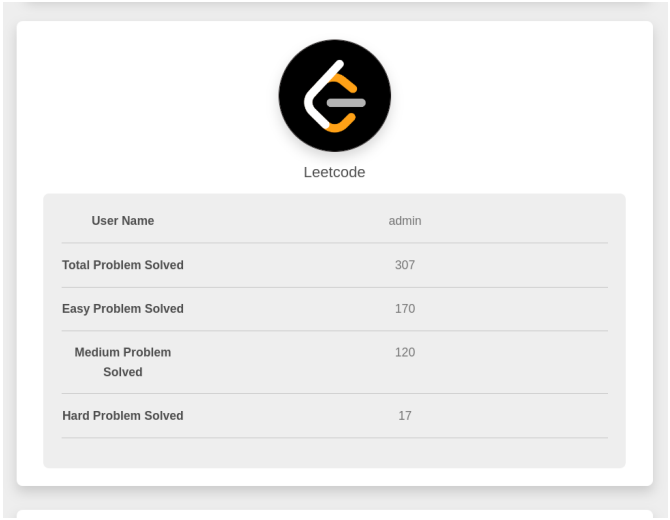

Fig. 4. User Profile - Leetcode

B. Application Server

It is a web server that manages all the user queries and gives back the required data in response. It is built using NodeJs and ExpressJs. Its major functions include login, registration, data storage, data retrieval, etc. It interacts with the API server in order to fetch real-time coding status data.

The front-end sends the request to the application server, and based on the request, the application server sends the information back to the application server. The application server is associated with a NoSQL database, which stores the data in NoSQL format.

C. API Server

It is the most crucial component of the application, which is responsible for fetching data from websites. It uses the web scraping technique for the retrieval of information. This server is built using Flask. For the scraping of data, the BeautifulSoup library in Python is used.

The API server is a standalone server that treats requests based on two parameters: the coding platform name and the username of the user for that particular website. It uses HTML parsing techniques to extract the information from the website. It only extracts relevant information, groups all the data together, and returns a JSON response.

IV. Results

After thoroughly applying statistical and logical techniques, we have analysed the following results: The platform was tested for over 520 plus students, and we took their coding platform IDs and evaluated them on our website. The desired objectives of the proposed system were met after its implementation. A Google Form was circulated among the students in the college to take their valid details, including the student's first and last name, email, enrollment number, Codechef ID, Codeforces ID, and Leetcode ID, which helped to make entries in ICode.

A username was given to the students, and the checking was done by them to check whether the output was correct or not. The user profile is viewed by entering the username in the search box provided and even in the college section, where all the users are registered. Students under a specific college can be viewed by entering their college name in the provided select college field. After allowing the students to access ICode through their systems, they were guided to keep the below five perspectives in their minds and provide feedback to our team.

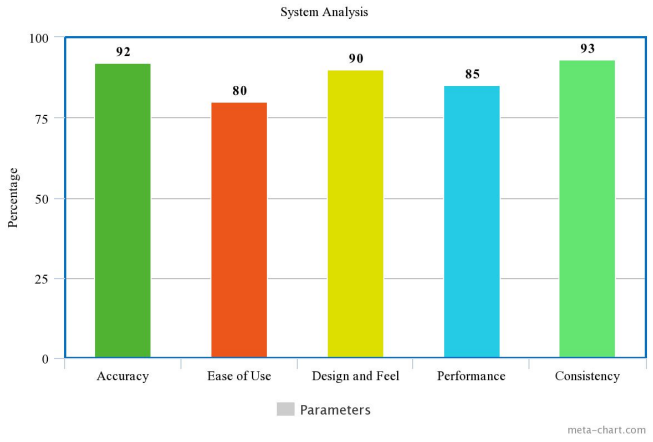

Fig. 5. System Analysis

Following parameters are taken into consideration for system analysis:

1. Accuracy: Checking was done based on how accurate the data is. Because more conventionality will result in more accurate data. The original data was checked with the resulting data from the website. After analysing the data, the results obtained were accurate, which means that after cross-checking the details from the three platforms with the obtained results,

the attributes were found to be correct, as were the platform's details and ICode's details.

2. Ease of Use: It's the criterion for the contentment of the user who is using the website. The platform tends to be convenient to the user since, on the navigation bar, links to home, register, login, and about us are provided, making it easy for him or her to use the website efficiently. One can also view the profiles of their peers if they know the username. It even becomes simple for colleges to directly check their students' profiles using the given search button on the navigation bar.

3. Design and feel: The UI design is straightforward, and the user can easily make use of the website. Keeping the golden rules of HCI in mind, the website is designed accordingly. Consistency is maintained. Even the banners on the home page encourage the user about the need for competitive coding and make ICode attractive. Error handling is done to avoid errors while the user is using the site. In order to reduce the short-term memory load of the user, the UI is built in such a way that the user need not remember where the functionalities are on the site since everything is displayed on the home page itself. The logout button helps the user sign off from the website. Since a professionally designed website helps to make a positive first impression on end users and even facilitates easy access and navigation for the users, the design and feel are taken into consideration by Icode.

4. Performance: Accessibility and other important metrics of the website are dependent on the performance of the system. There is always a strong correlation between user experience and website performance; for this, ICode performance must be efficient. After the user logs into the website to see his or her profile, the site at least takes an average of two to three seconds to display the details. The major concern of users while using any website is the time the site takes to respond to a query. If the response time is less and the details are accurate, then it satisfies the user; otherwise, the user is left disappointed. On ICode, the user doesn't need to wait for a long time to get a correct response since all the constraints are taken into consideration. Even the profile details coming as output are valid and consistent. After testing, it was determined that the website was efficient enough.

5. Consistency: The major concern of ICode was to give its users the current ratings, which are up-to-date. It means if the user has given a contest on any of the platforms and his or her rating is increased in the contest based on the number of questions solved, then after the contest gets finished, the current rating on the platform should increase, may it be Codechef, Codeforces, or Leetcode, and simultaneously it gets reflected on ICode. On the contrary, if the rating decreases, then it gets decreased on the platform and on ICode too. Therefore, depending on the number of questions solved on these 3 platforms in the contest, if more questions are solved than in the previous contest, the rating will increase; if less questions are solved, the rating will decrease; and side by side, the current rating will be reflected on ICode.

V. Conclusion

This system will be beneficial for both students and college level which will help the students to keep a track of their profile and simultaneously college will be able to know the Competitive coding level of their students which will help to analyse how much more efforts do they need to put for gaining importance of coding in students. It will also save time and efforts of both students as well as the college. Also being able to see the progress of their peers the students will also stay motivated.Also it will revolutionise the recruitment process and will benefit recruiters for better analysing the profile of users.

References

[1] Kurniawati, D., Triawan, D. (2017, November). Increased information retrieval capabilities on e-commerce websites using scraping techniques. In 2017 International Conference on Sustainable Information Engineering and Technology (SIET) (pp. 226-229). IEEE.

[2] Singrodia, V., Mitra, A., Paul, S. (2019, January). A review on web scrapping and its applications. In 2019 international conference on computer communication and informatics (ICCCI) (pp. 1-6). IEEE.

[3] Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., Mbaye, S. N. (2019, December). Web scraping: state-of-the-art and areas of application. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 6040-6042). IEEE.

[4] Nair, P. R. (2020). Increasing employability of Indian engineering graduates through experiential learning programs and competitive programming: Case study. Procedia Computer Science, 172, 831-837.

[5] Raman, R., Vachharajani, H., Achuthan, K. (2018). Students motivation for adopting programming contests: Innovation-diffusion perspective. Education and Information Technologies, 23(5), 1919-1932.

[6] Das, D., Mathews, N. S., Chimalakonda, S. (2022, June). Exploring Security Vulnerabilities in Competitive Programming: An Empirical Study. In Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering (pp. 110-119).

[7] Chapagain, A. (2019). Hands-On Web Scraping with Python: Perform advanced scraping operations using various Python libraries and tools such as Selenium, Regex, and others. Packt Publishing Ltd.

[8] Thivaharan, S., Srivatsun, G., Sarathambekai, S. (2020, September). A survey on python libraries used for social media content scraping. In 2020 International Conference on Smart Electronics and Communication (ICOSEC) (pp. 361-366). IEEE.

[9] Furtmueller, E., Wilderom, C., Tate, M. (2011). Managing recruitment and selection in the digital age: e-HRM and resumes. Human Systems Management, 30(4), 243-259.

[10] Krotov, V., Silva, L. (2018). Legality and ethics of web scraping.