A Project Report

On

ONLINE APTITUDE EXAM GENERATION SYSTEM USING WEB SCRAPING

Submitted in partial fulfillment of the requirement of
the Degree of

Bachelor of Technology
In
Computer Engineering

Submitted By
Rohan Kadam
Shreyas Varadkar
Nisha Sagar
Atharva Bhingarde

Supervisor
Prof. Sheetal Shimpikar

Department of Computer Engineering

PILLAI COLLEGE OF ENGINEERING (AUTONOMOUS)

New Panvel – 410 206

UNIVERSITY OF MUMBAI

Academic Year 2023-24

DEPARTMENT OF COMPUTER ENGINEERING
PILLAI COLLEGE OF ENGINEERING (AUTONOMOUS)
New Panvel – 410 206

CERTIFICATE

This is to certify that the requirements for the **B.E. Project A** report entitled '**Online Aptitude Exam Generation System using Web Scraping**' have been successfully completed by the following students:

Name	**Roll No.**
Rohan Kadam	B720
Shreyas Varadkar	B755
Nisha Sagar	B743
Atharva Bhingarde	B715

in partial fulfillment of Bachelor of Technology in the Department of Computer Engineering, Pillai College of Engineering, New Panvel – 410 206 during the Academic Year 2023 – 2024.

_____

Supervisor

(Prof. Sheetal Shimpikar)

_____	_____
Head of Department	**Principal**
Dr. Sharvari Govilkar	**Dr. Sandeep M. Joshi**

DEPARTMENT OF COMPUTER ENGINEERING
PILLAI COLLEGE OF ENGINEERING (AUTONOMOUS)
New Panvel – 410 206

PROJECT A REPORT APPROVAL

This **Final Year Project** a report entitled "**Online Aptitude Exam Generation System using Web Scraping**" by **Rohan Kadam, Shreyas Varadkar, Nisha Sagar and Atharva Bhingarde** are approved for the degree of B.E. in **Computer Engineering**.

Examiners:

1. _____

2. _____

Supervisors:

1. _____

2. _____

Chairman:

1. _____

Date:

Place:

Declaration

We declare that this written submission for B.E. Project Report Declaration entitled "**Online AptitudeExam Generation System using Web Scraping**"represents our ideas in our own words and where others' ideas or words have been included. We have adequately cited and referenced the original sources. We also declared that we have adhered to all principlesof academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas / data / fact/ Source in our submission. We understand that any violation of the above will cause disciplinary action by the institute and also evoke penal action from the sources which have thus not been properly cited or from whom paper permission has not been taken when needed.

Project Group Members:

Rohan Suhas Kadam: _____

Shreyas Sandeep Varadkar: _____

Nisha Namdev Sagar: _____

Athrarva Dinesh Bhingarde: _____

Date:

Place:

Table of Contents

Abstract

In today’s fast-paced world, people like things in shortcut – quickest, fastest, easiest, least effort path paradigm – and sometimes creating different aptitude questions sets of questions can be a hectic task for companies while conducting aptitude rounds of employees for various job profiles and also as we know, most of the students and freshers now a days are finding hard to clear aptitude tests too. The online Aptitude test is one of the most important sections in all the Competitive exams, Entrance exams, and Campus Interviews to grab Government jobs and IT Jobs. So, practicing well and enhancing your Aptitude skill becomes essential for new students entering in the engineering sector. Aptitude tests are tests of general intelligence that rate one's cognitive abilities. This test comprises of various subtests that measure one's logical ability, numerical ability, and attention to detail. Over the years, standardized aptitude tests have been introduced in the recruitment space too, and are playing an extensive role in maintaining meritocracy in the technical market. Aptitude tests have become of paramount importance to job seekers in today's competitive job market and it has thus become imperative not to neglect their importance when preparing and seeking for jobs.

So, to reduce the headache of various companies to arrange or set different sets of questions papers and also for students to get a platform where they can brush up their aptitude skills, we would be developing an Online Aptitude Exam Generation System in which question papers would be generated in real time and also the students can practice aptitude questions to gain a idea about what is needed in today's era to gain success in aptitude rounds of various companies.

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Fundamentals

An aptitude test is a quiz intended to assess a person's aptitude for a certain activity or skill set. The underlying premise of aptitude tests is that people have intrinsic talents and limitations as well as a natural tendency towards success or failure in particular domains based on their innate qualities. An aptitude test is used to measure a person's skills by predicting how well they would do in a situation for which they have neither training nor information. Aptitude test questions will encompass categories such as quantitative, verbal, and logical.

1.2 Objectives

The primary objective of this web application is to facilitate an environment that is user-friendly for all users and minimize manual work. The project's objective is to create an effective test and aptitude learning application.

1.3 Scope

The project's scope is fairly broad in terms of knowledge acquisition and information exchange. Several points are:

- Considering that it is a web-based application, it may be used anytime and whenever.
- Both the corporate world and educational institutions will use this application.
- It will help you get a general understanding of your personality.
- Having access to exam results as soon as they become available

1.4 Outline

The report is organized as follows: The introduction is given in Chapter 1. It contains introduction and describes the basic structure of system. This chapter also presents the outline of the objective of the report. Chapter 2 describes the review of the relevant various techniques in the literature systems. It describes the pros and cons of each technique. Chapter 3 presents the Theory and proposed work. It describes the major approaches used in this work. The societal and technical applications are mentioned in Chapter 4. The summary of the report is presented in Chapter 5.

Chapter 2

Literature Survey

2.1 Introduction

Online Aptitude Exam Generation System saves time by automating the process of generating exams for candidates. In recent years, the online exam is been used in a variety of educational measurements and corporate worlds for higher accuracy, easier manipulation, and multi-functional. Though the interface for online exam is found to have good existence, in some circumstances such as many numbers of students, way of attempting the answer, timing, etc, the measurement results still need more smart results for analyzing students' knowledge.

2.2 Literature review:

2.2.1 Questionnaire Generation and Monitoring System

The research done by Sonia Thakur [2021] et al developed a question generator and the monitoring system in which faculties are able to automate the process of creating quizzes or test forms having various options to choose from. The system fetches questions from it's in- house databases which involve questions collected from websites, scanning pdfs and images from which it uses Optical character recognition (OCR) is an electronic service that converts an image into text, and Predicts questions based on given paragraphs or multimedia files by using Questgen AI which is an open-source NLP library that helps to develop easy to question generation algorithms. The exams taken on the system will be in proctored mode. In the future, they are trying to improve the face monitoring module and other processes to make the system faster.[1]

2.2.2 Automated MCQ Generator using Natural Language Processing

Pritam Kumar Mehta1 [2021] et al present a paper that talks about a system that generates questions automatically In the Automated MCQ Generator, questions are generated automatically with the help of NLP. The text of any domain is provided as input to the system which is then summarized using the BERT algorithm. BERT (Bidirectional Encoder Representation from Transformers) is a deep learning-based technique for natural language processing, a pre-trained model from Google. Now the keywords are selected from the summarized text using the Python keyword extractor (PKE) and accordingly, the mapping of a keyword is done with a sentence. Now the main task is relevant distractors.

Distractors are generatedusing the Wordnet approach. Wordnet is an API used to get the correct sense of the word. So goodand relatable distractors are generated. This system solves the problem of manual creation of questions and reduces time consumption and cost.[2]

2.2.3 Automated Exam Question Set Generator Using Utility-Based Agent and Learning Agent

Abd Rahim, and et al, 2020. This research produced an Automated Exam Question Set Generator (AEQSG) using a Utility-Based Agent (UBA) and Learning Agent (LA). Moreover, AEQSG stratifies Bloom Taxonomy (BT) scaling to automate mac's Taxonomy (BT) difficulty level distribution and Genetic Algorithm (GA) to optimize the generation of exam question sets and generate high-quality exam question set that follow educational organization's Standards. The aim of a utility-based agent in AEQSG is to present the users with an option to select actions depending on a user's utility for each generation state. At the same time, the purpose of the learning agent in AEQSG is to learn from its previous exam results.[3]

2.2.4 Automated exam question generator using genetic algorithm

Zalilah Abd Aziz [2017]. This study proposes an automated test question generator to address the issue of creating multiple-choice exam questions. The teacher can auto-generate new exam questions based on the Genetic Algorithm (GA) and six levels of Bloom's Taxonomy to develop high-qualityexam questions that can test the various levels of learners based on Bloom's cognitive domains andeducator-selected chapters. The prototype, which included 500 sample questions, was run 50 timeswith different amounts of chapters for each test case. It achieves a score of 90 percent for the greatestexam question weightage, with a score of 70 percent for the average exam question weightage percentage generated. The lowest weighted percentage of exam questions generated is 40%. The result is influenced by the lesser number of questions in the questions bank for each Bloom's taxonomy level.[4]

2.3 Limitation of existing system or research gap

Although all techniques used in above systems have shown promise in developing Automatic MCQ exam generation systems, there are still some limitations and research gaps in this area. The systems use directory dataset where the exam authority first have to make question entries manually in datasets and then the questions will be appeared on test screen randomly using algorithms.[5] The approach will definitely give results but it is a time and efforts consuming process to create the question datasets.

2.4 Literature Summary

The literature survey highlights the potential of different technologies in developing automatic question generation However, there are still some limitations and research gaps in this area such in all the existing systems the question needs to be manually entered by exam authorities. Future research should focus on addressing these challenges by using the web scraping techniques which will reduce time and efforts required to generate manual question sets.

SN	Paper	Author	Advantages and Disadvantages
1.	Questionnaire Generation and Monitoring System	Sonia Thakur, Smith Gajjar, Hrithik Malvani, Jai Soneji, Mrs. Mannat Doultani [2021]	Advantages: Automate the process of creating quizzes or test forms having various options to choose from. The exams taken on the system will be in proctored mode. Disadvantages: Some of the Modules in the system are process intensive, hence they require a high computing power like the question generation module has lots of processing before predicting questions that cannot be executed on commodity hardware.
2.	Automated MCQ Generator using Natural Language Processing	Pritam Kumar Mehta, Prachi Jain, Chetan Makwana, Dr. C M Raut [2021]	Advantages: The proposed system creates automated questions with the help of NLP that reduces human intervention and it is a cost and time-effective system. The system uses BERT, Wordnet, and PKE technologies for processing. Disadvantages: The existing system is based on Google's BERT Model, and the accuracy of the system will increase in the future as the performance of the model is improved. The model takes time to improve its accuracy.

3.	Automated Exam Question Set Generator Using Utility-Based Agent and Learning Agent	Abd Rahim, Ma. Stella Tabora Domingo, Mohamed Farid Noor Batcha, and Zalilah Abd Aziz [2020]	Advantages: The proposed system is Automated Exam Question Set Generator using Utility Based Agent, Learning Agent, Bloom Taxonomy (BT) Scaling, and Genetic Algorithm (GA). Disadvantages: In this system, the faculty have to enter questions manually for the specified limit, it'll be a time and effort consuming process.
4.	Automated exam question generator using genetic algorithm	Abd Rahim, Zalilah Abd Aziz, Rose Hafsah Ab Rauf, and Noratikah Shamsudin [2017]	Advantages: Present an automated examquestion generation of multiple choice exam questions using GA and Bloom's Taxonomy. Disadvantage: The question bank which will be used by system to create questions has to be created by teachers manually. The efforts required to make questions bank for large number of students is more.

Table 2.1 Summary of literature review

Chapter 3

Proposed System

3.1 Overview

Aptitude test is a way for employers to assess a candidate's abilities through a variety of different testing formats. Aptitude tests will test your ability to perform tasks and react to situations at work. The probability of student failing aptitude exam is high. We are proposing an innovative mechanism which allows exams authority to conduct an online aptitude examination using web scraping.

3.1.1 Existing System Architecture

This brings up the age-old debate between storing information in databases and storing it on paper.

The following issues arise when information is kept in hard copy papers.

- **Lack of space** – It becomes a problem in itself to find space to keep the sheets of paper being generated as a result of the ongoing discussion. The documents being generated are too important to be ill-treated.
- **Filing poses a problem –** Filing the documents categorically is a time consuming and tedious exercise.
- **Filtering is not easy –** It becomes hard to filter relevant documents for the irrelevant ones if the count of the same crosses a certain manageable number.
- **Reviewing becomes time-consuming –** All the process done manually at the centers and all the records are maintained on the papers. So, the maintenance of the record is very difficult in the departments and as well as it’s very difficult for the workers to check the record.
- **Lots of Manual work:** In existing system teacher have to enter questions manually one by one which require too much efforts and time.

3.1.2 Proposed System Architecture

The previous sections discussed the strengths and weaknesses of the existing system. In order to achieve better domain results, we are proposing an innovative mechanism which allows exams authority to conduct an online aptitude examination using web scraping. The proposed system will have various modules for example creating a paper, Online Paper Assessment etc. We will be web scrapping. The questions will be selected and filtered ahead according to various types or levels of

aptitude. It will create an automated real time data set as the questions will be fetched in real time from different websites and from them the papers will be generated. So,the system will be dependent on machines. Here there is minimum or no human intervention for thesaid process. The questions scrapped fetched would be stored in a customized dataset and then fromthat the paper would be generated. Once the paper has been selected dataset for the set questions along with the options and correct answers then the examination authority can select the pattern thatis the number of questions and minimum passing marks and same will be informed to the candidate.The candidate will attempt the exam then he will get instant result whether he passed or not. If anyorganization wants same system to be used by their own dataset for example one of the organizationor one of the companies has its own dataset the same can be achieved through upload module. And the same questions will be fetched from the uploaded CSV file.

Our application can also be used by students of various educational institutions to practice aptitude questions and to brush up with their aptitude skills which could help them to excel in aptitude exams.

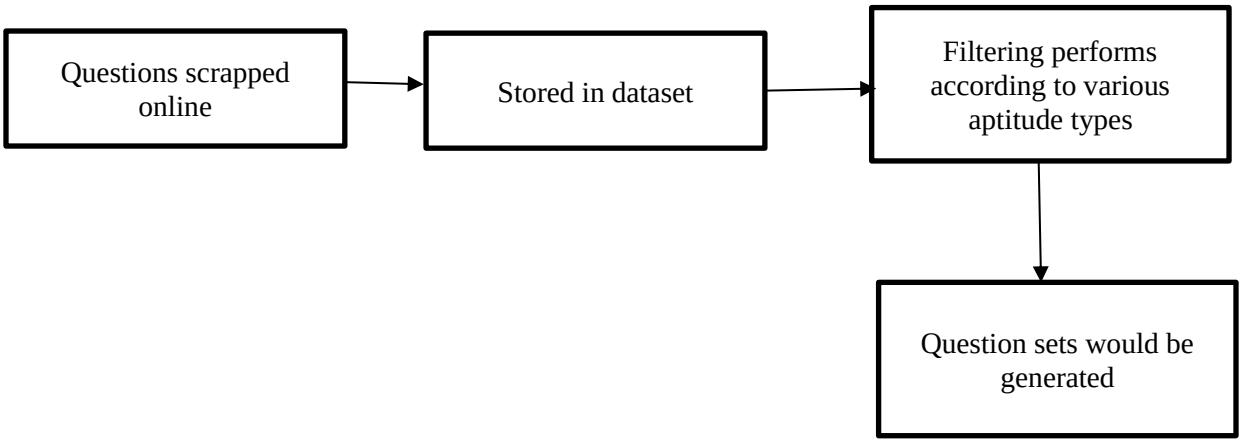


Fig. 3.1 Proposed system architecture

3.2 Implementation Details

This section provides information about the implementation specifics. The system utilizes web scraping to extract questions, which are then stored in a database. These questions are subsequently retrieved randomly from the database and assigned to students for testing. Students can then attempt the test. Further details about the web scraping process are explained below.

3.2.1 Techniques used

A. Web Scrapping

Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique used to automatically extract large amounts of data from websites and save it to a file or database. The data scraped will usually be in tabular or spreadsheet format.

Data displayed by websites can only be viewed using a web browser. Most websites do not allow you to save or download this data. If you need the data, the only option is to manually copy and paste the data - a very tedious job which can take many hours or days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.

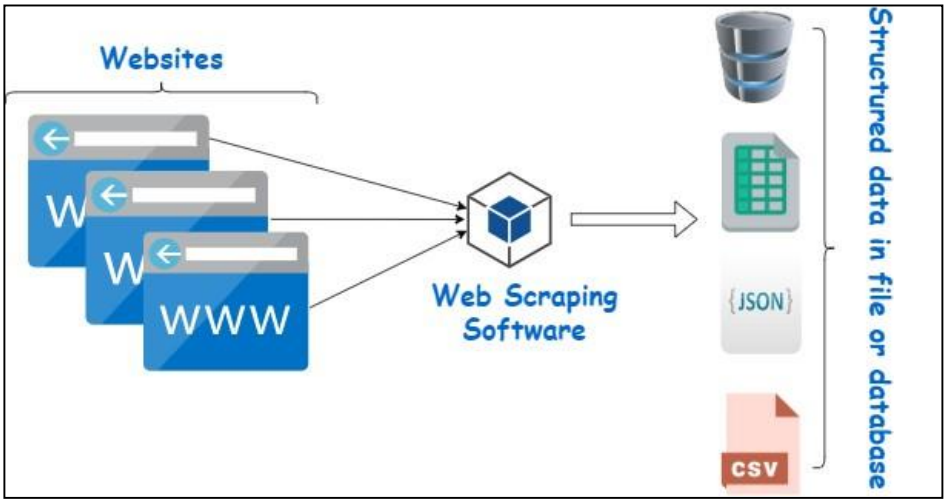


Fig. 3.2 working of web scrapper

A web scraping software will automatically load, crawl and extract data from multiple pages of websites based on your requirement. It is either custom built for a specific website or one which can be configured to scrape data from any website. With the click of a button, you can easily save the data displayed by websites to a file in your computer.

3.2.2 Use Case Diagram

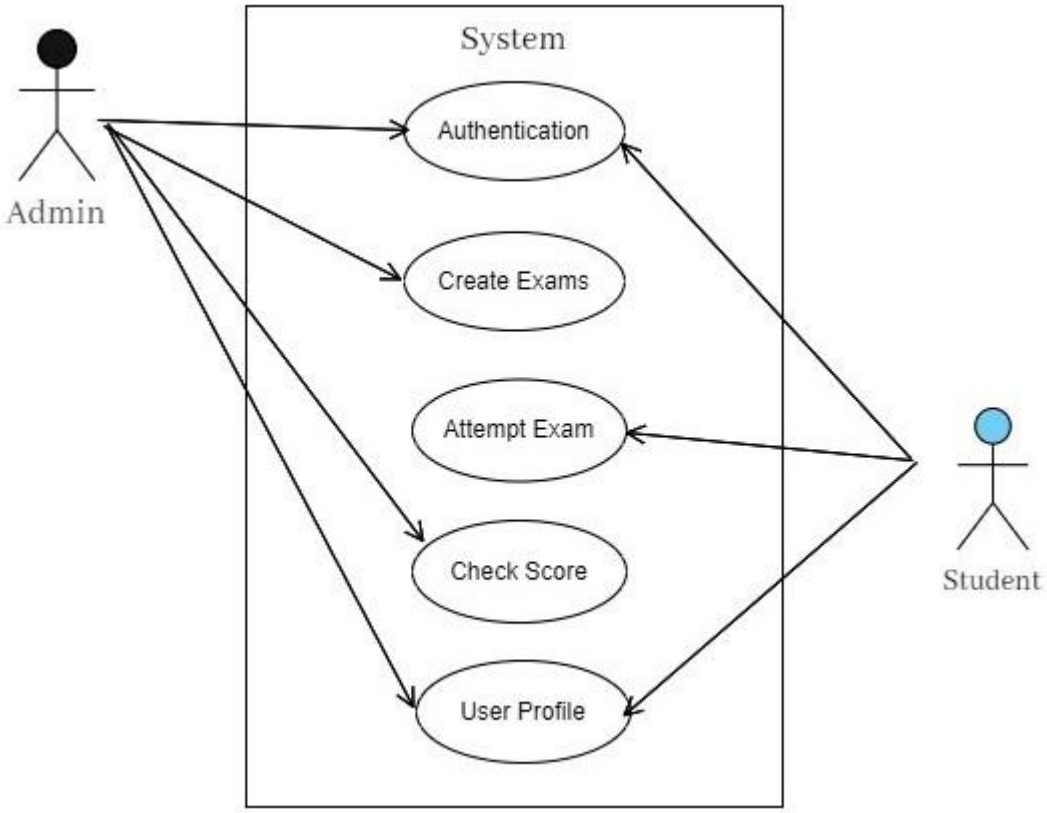


Fig. 3.3 Use Case Diagram

3.2.3 Hardware and Software Specifications

The experiment setup is carried out on a computer system which has the different hardware and software specifications as given in Table 3.2 and Table 3.3 respectively.

Processor	2 GHz Intel
HDD	180 GB
RAM	2 GB

Table 3.2 Hardware details

Operating System	Windows 7
Programming Language	Python
Database	MongoDB

Table 3.3 Software details

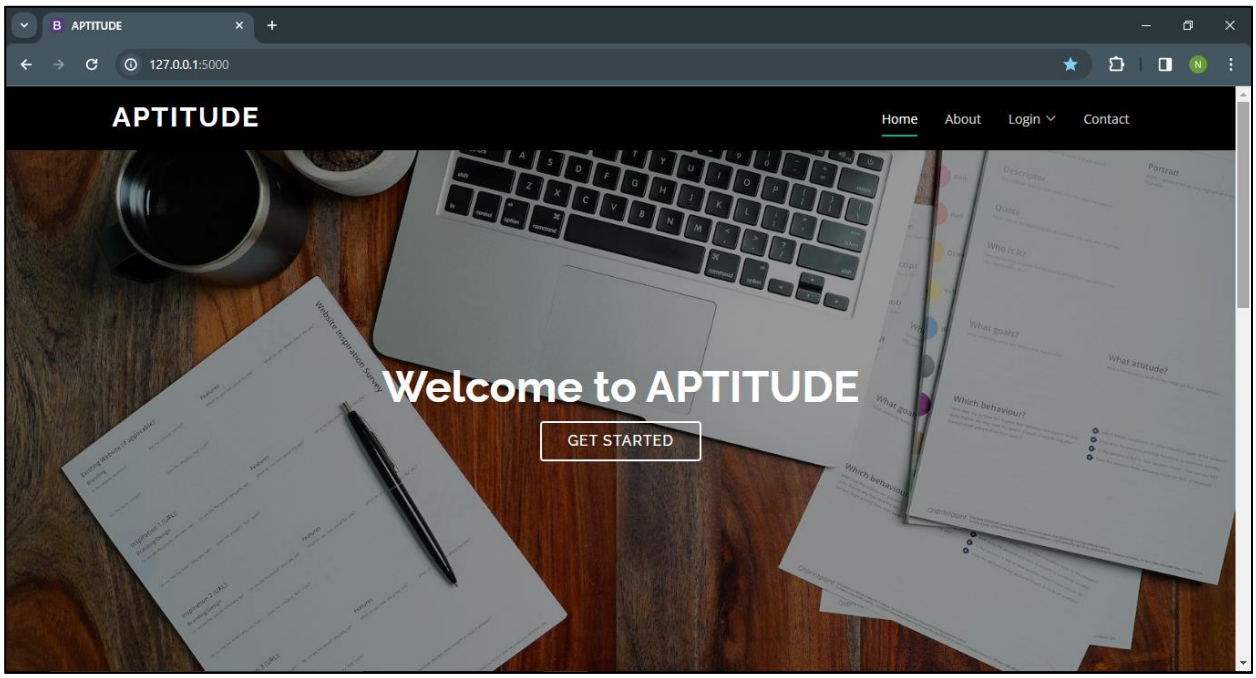
Chapter 4

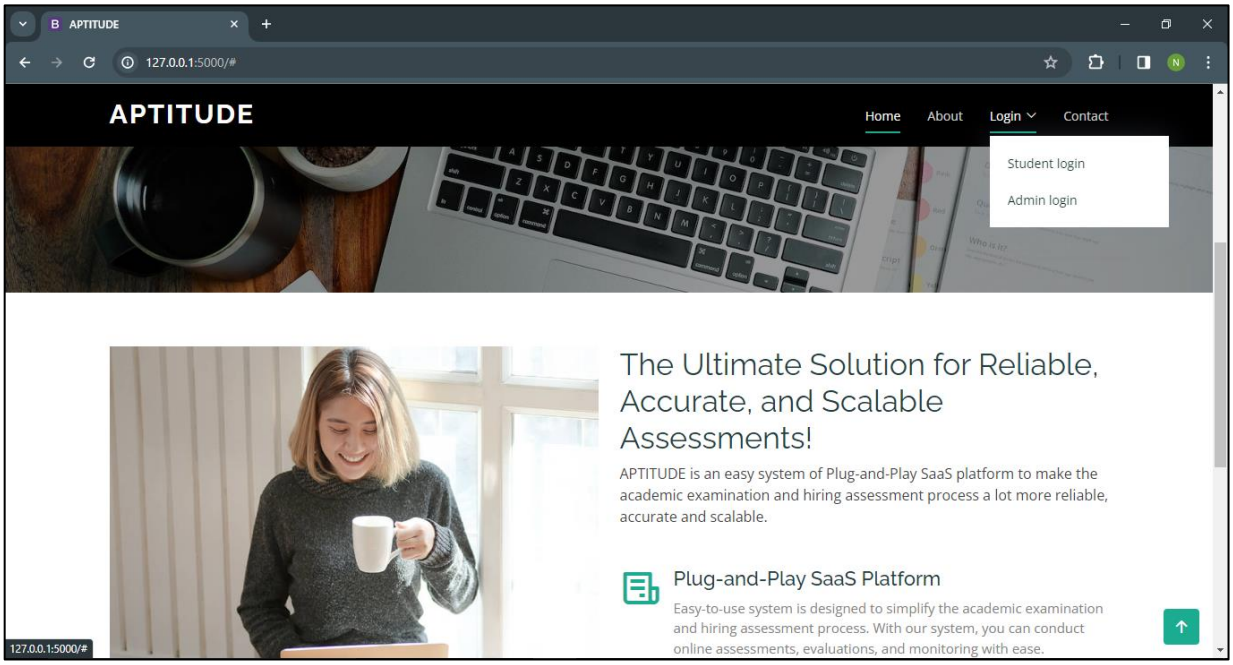
Results and Discussion

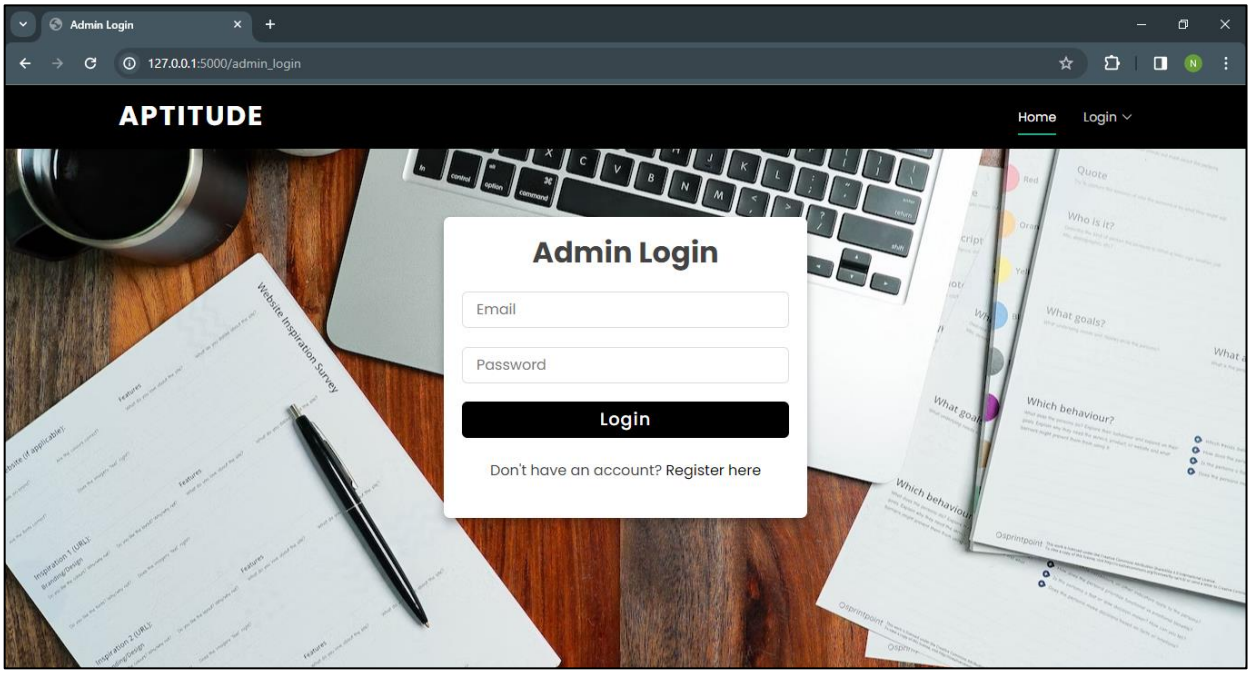
4.1 Dataset Used

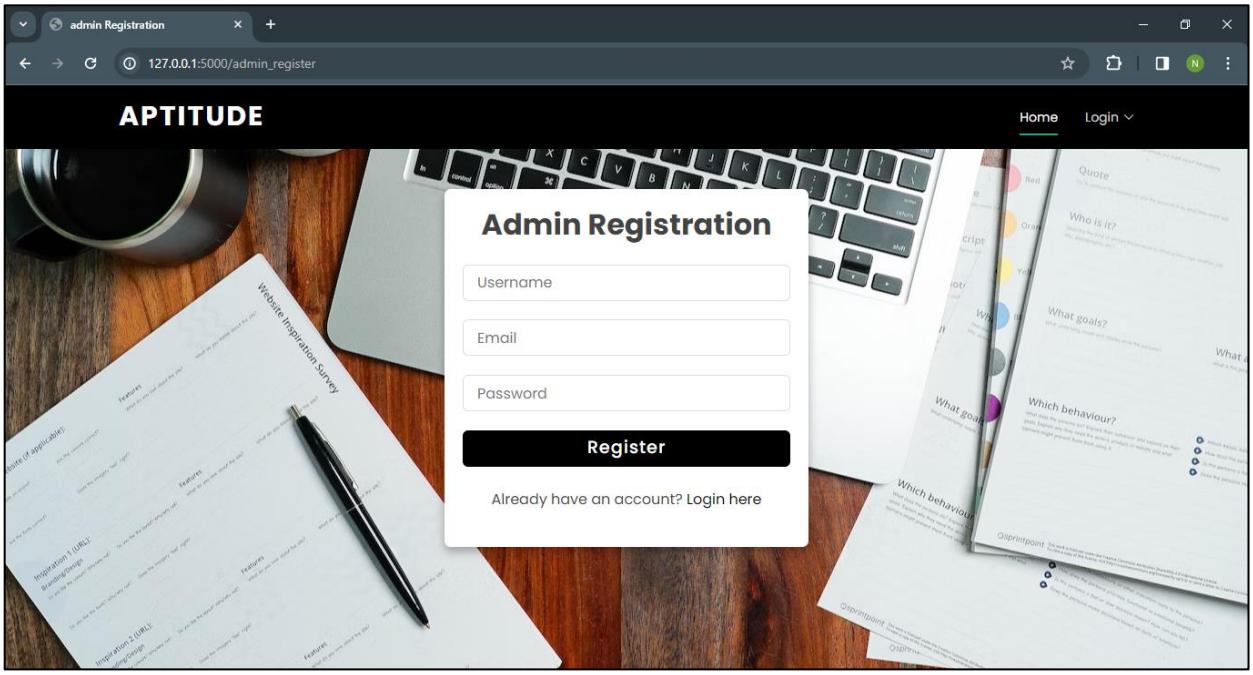
Real time dataset will be created from the question extracted from different websites using web scrapping and crawling. The questions will be fetched from the websites and stored into the CSV files and will be used as dataset.

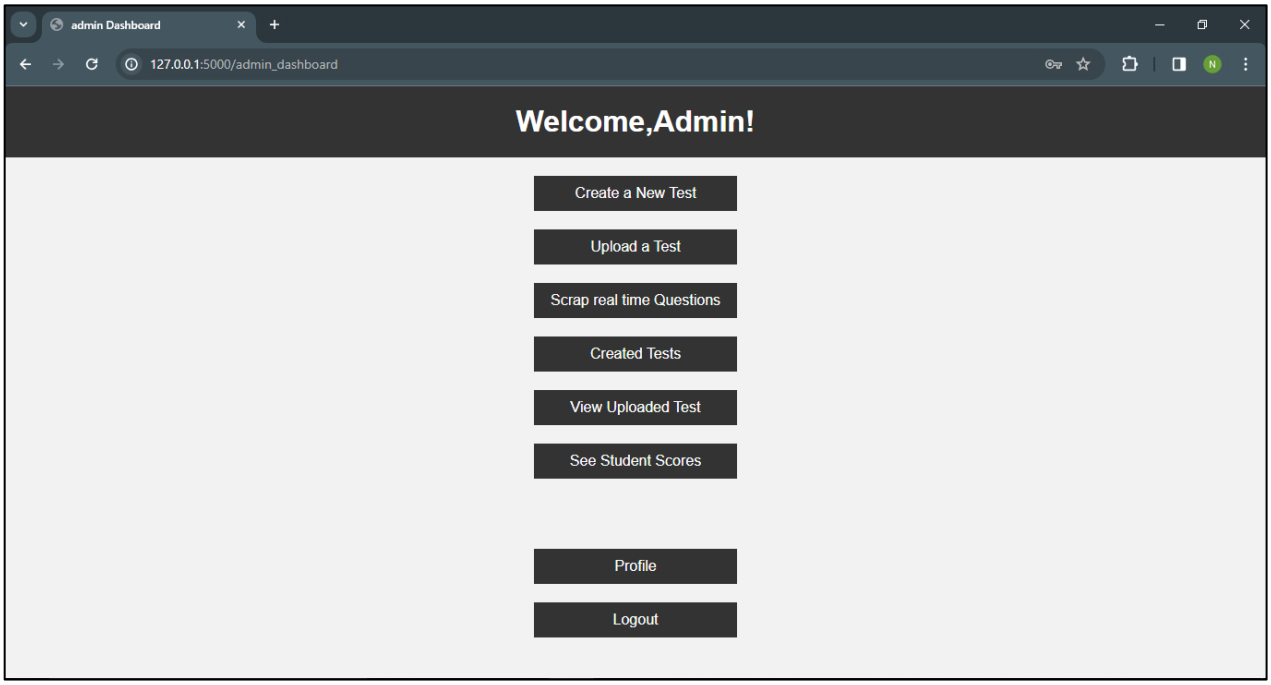
4.2 Results

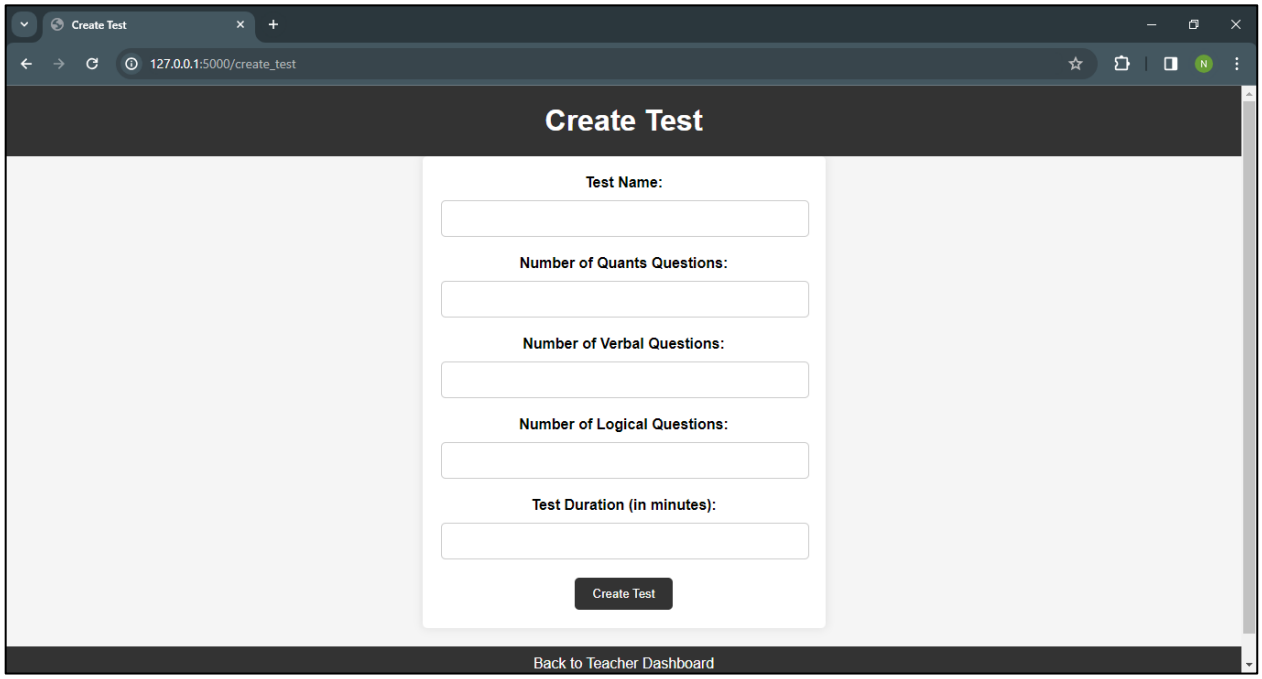


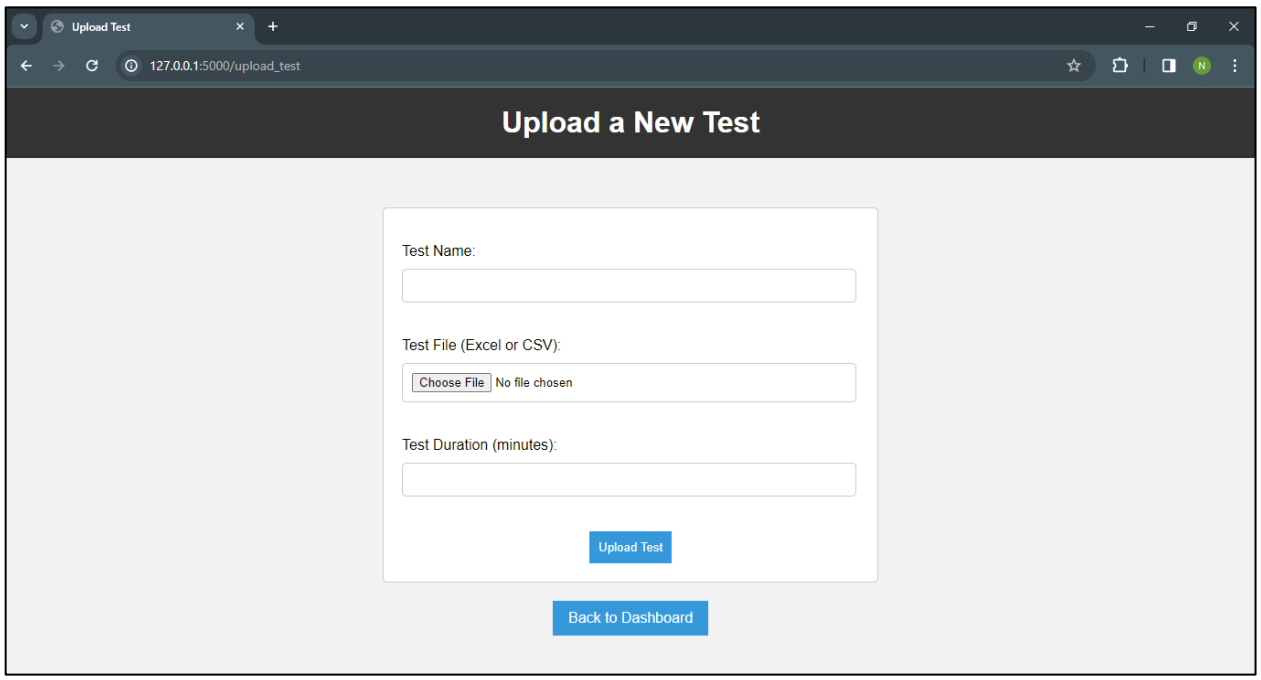

APTITUDE
127.0.0.1:5000/#
APTITUDE
Home
About
Login
Contact
Student login
Admin login
The Ultimate Solution for Reliable, Accurate, and Scalable Assessments!
APTITUDE is an easy system of Plug-and-Play SaaS platform to make the academic examination and hiring assessment process a lot more reliable, accurate and scalable.
Plug-and-Play SaaS Platform
Easy-to-use system is designed to simplify the academic examination and hiring assessment process. With our system, you can conduct online assessments, evaluations, and monitoring with ease.
127.0.0.1:5000/#

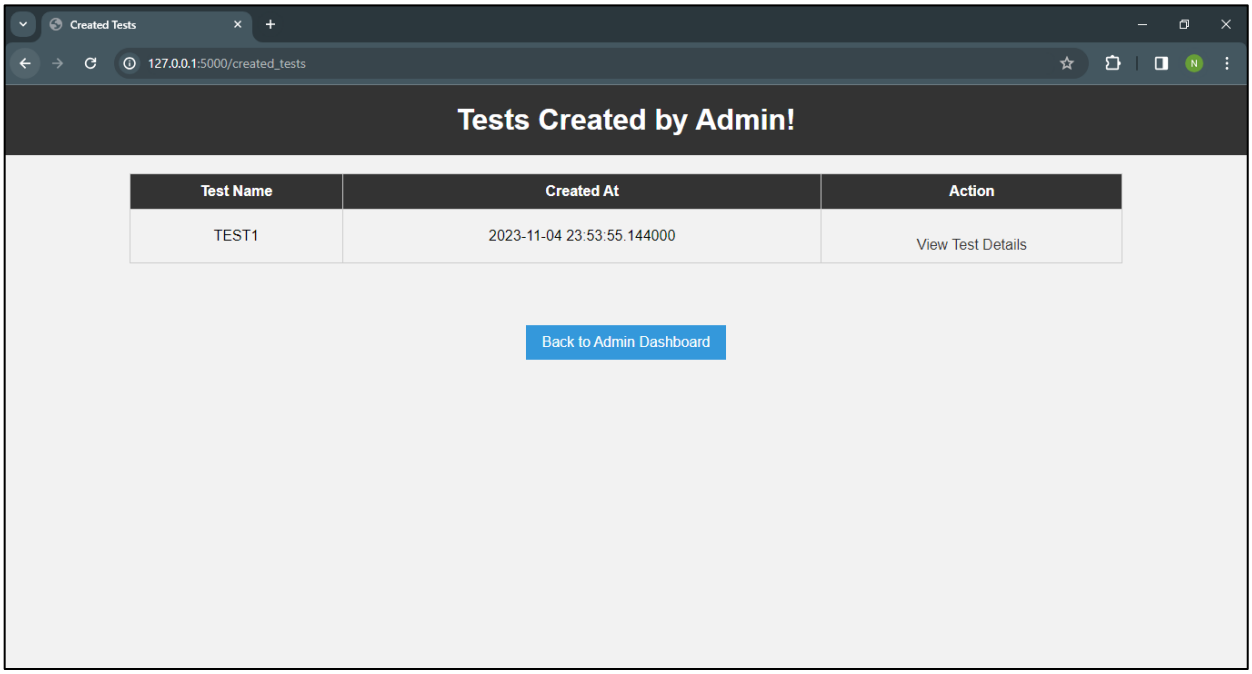

Admin Login
127.0.0.1:5000/admin_login
APTITUDE
Home
Login
Admin Login
Email
Password
Login
Don't have an account? Register here


admin Registration
127.0.0.1:5000/admin_register
APTITUDE
Home
Login
Admin Registration
Username
Email
Password
Register
Already have an account? Login here

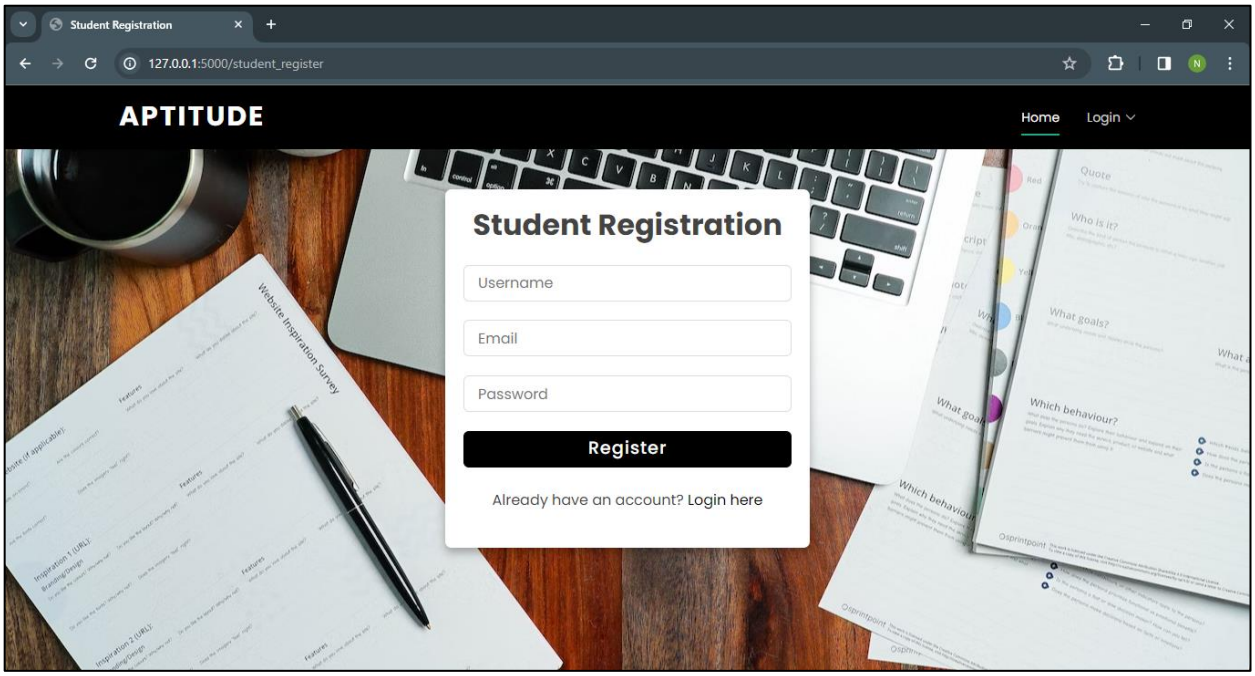

admin Dashboard
127.0.0.1:5000/admin_dashboard
Welcome,Admin!
Create a New Test
Upload a Test
Scrap real time Questions
Created Tests
View Uploaded Test
See Student Scores
Profile
Logout

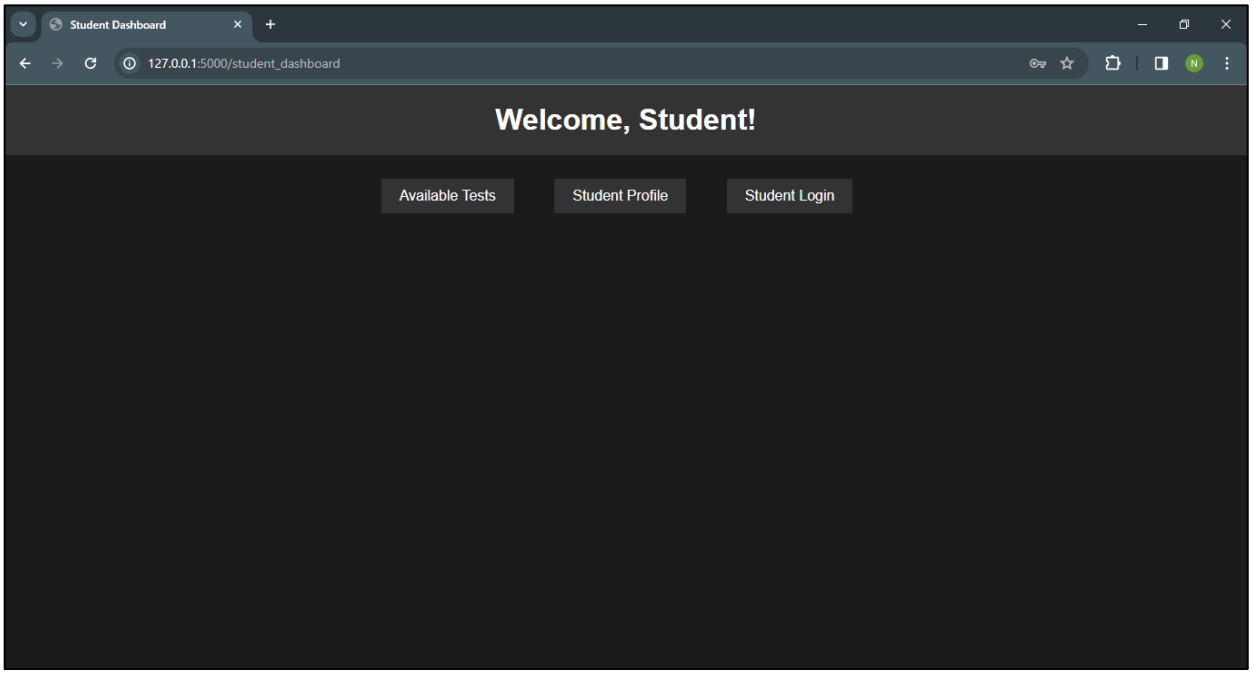

Create Test
127.0.0.1:5000/create_test
Create Test
Test Name:
Number of Quants Questions:
Number of Verbal Questions:
Number of Logical Questions:
Test Duration (in minutes):
Create Test
Back to Teacher Dashboard

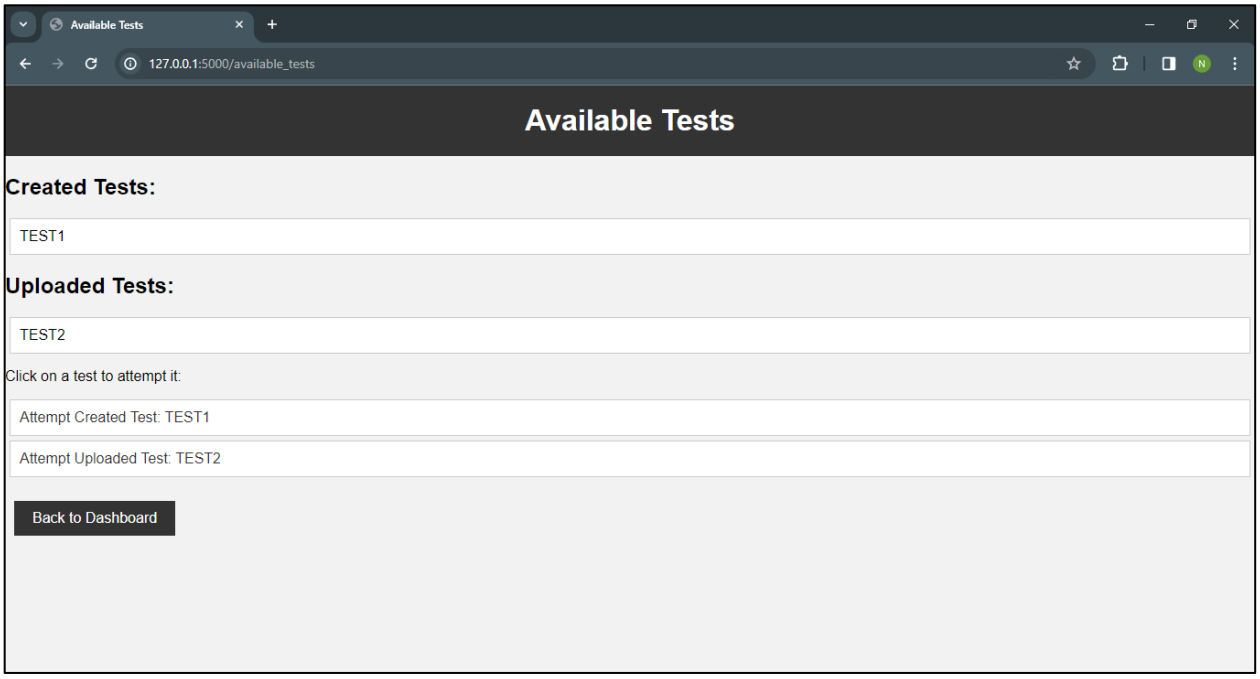

Upload Test
127.0.0.1:5000/upload_test
Upload a New Test
Test Name:
Test File (Excel or CSV):
Choose File
No file chosen
Test Duration (minutes):
Upload Test
Back to Dashboard

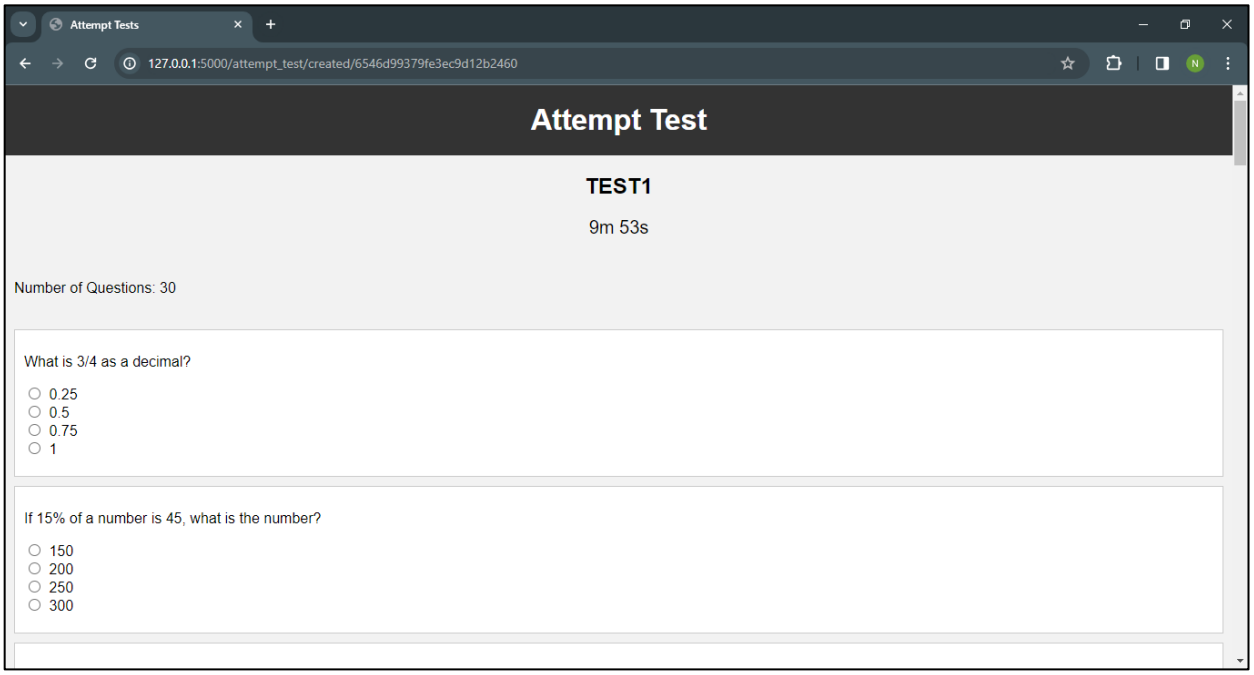


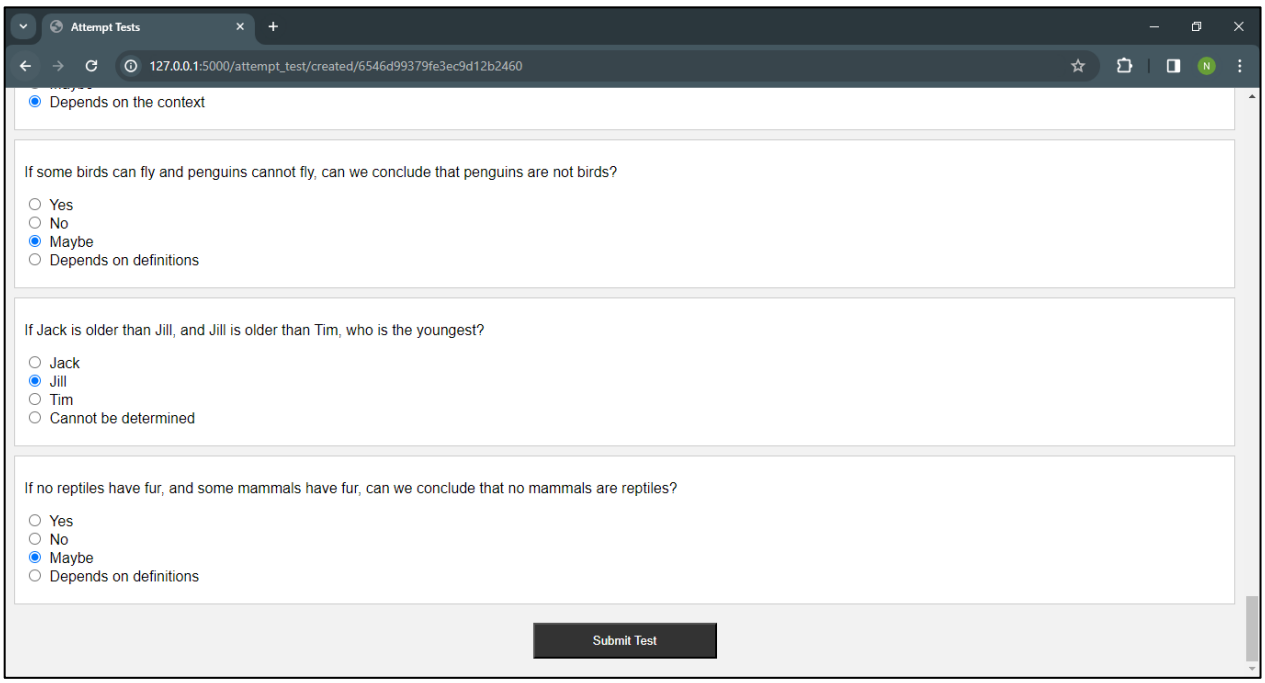
Student module-

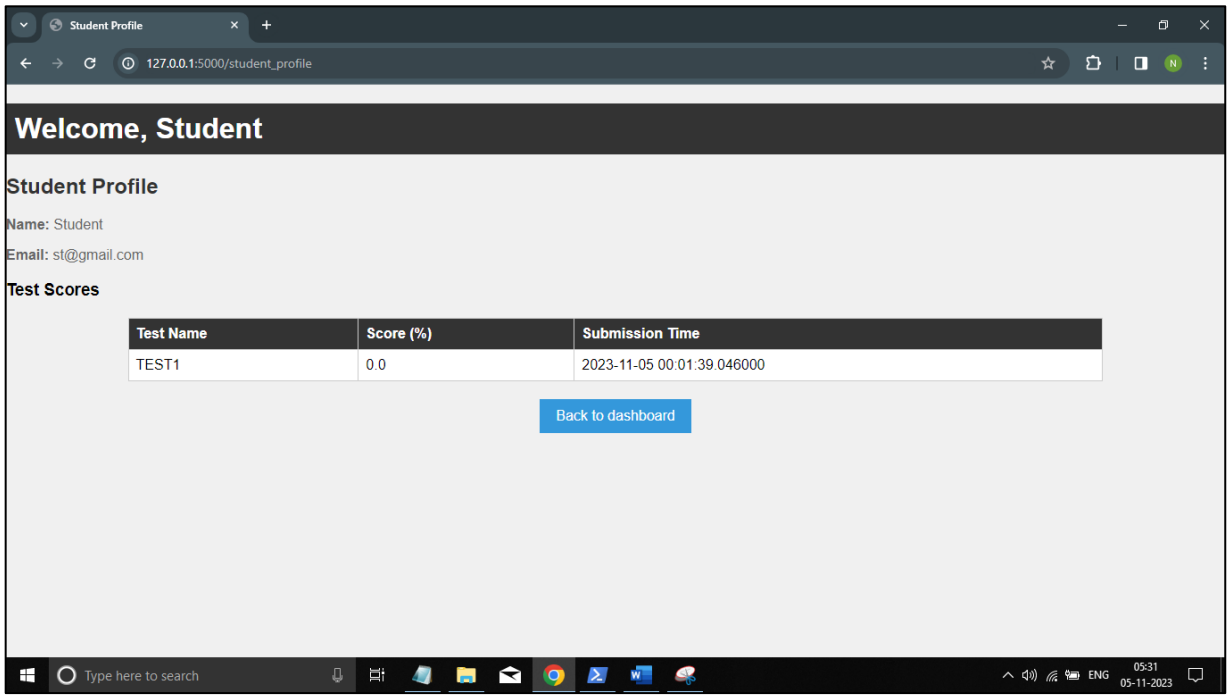


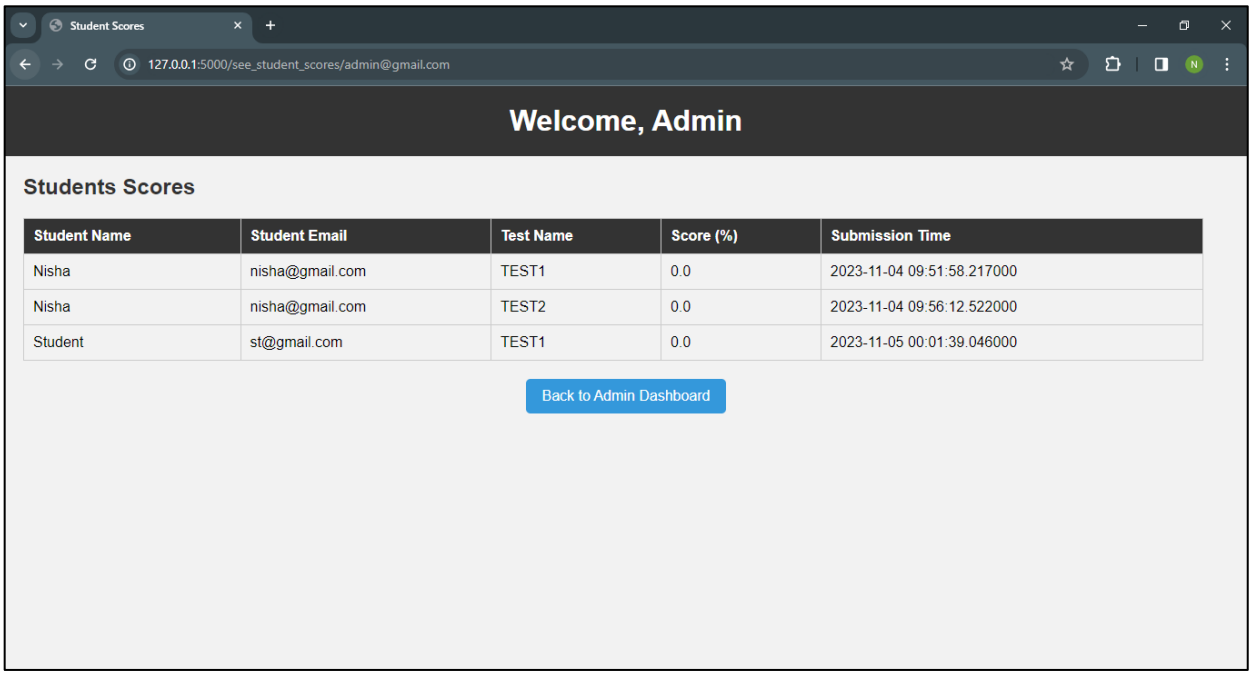

Student Dashboard
127.0.0.1:5000/student_dashboard
Welcome, Student!
Available Tests
Student Profile
Student Login


Available Tests
127.0.0.1:5000/available_tests
Available Tests
Created Tests:
TEST1
Uploaded Tests:
TEST2
Click on a test to attempt it:
Attempt Created Test: TEST1
Attempt Uploaded Test: TEST2
Back to Dashboard


Attempt Tests
127.0.0.1:5000/attempt_test/created/6546d99379fe3ec9d12b2460
Attempt Test
TEST1
9m 53s
Number of Questions: 30
What is 3/4 as a decimal?
0.25
0.5
0.75
1
If 15% of a number is 45, what is the number?
150
200
250
300


Attempt Tests
127.0.0.1:5000/attempt_test/created/6546d99379fe3ec9d12b2460
Depends on the context
If some birds can fly and penguins cannot fly, can we conclude that penguins are not birds?
Yes
No
Maybe
Depends on definitions
If Jack is older than Jill, and Jill is older than Tim, who is the youngest?
Jack
Jill
Tim
Cannot be determined
If no reptiles have fur, and some mammals have fur, can we conclude that no mammals are reptiles?
Yes
No
Maybe
Depends on definitions
Submit Test


Student Profile
127.0.0.1:5000/student_profile
Welcome, Student
Student Profile
Name: Student
Email: st@gmail.com
Test Scores
Test Name
Score (%)
Submission Time
TEST1
0.0
2023-11-05 00:01:39.046000
Back to dashboard
Type here to search
ENG
05:31
05-11-2023


Student Scores
127.0.0.1:5000/see_student_scores/admin@gmail.com
Welcome, Admin
Students Scores
Student Name
Student Email
Test Name
Score (%)
Submission Time
Nisha
nisha@gmail.com
TEST1
0.0
2023-11-04 09:51:58.217000
Nisha
nisha@gmail.com
TEST2
0.0
2023-11-04 09:56:12.522000
Student
st@gmail.com
TEST1
0.0
2023-11-05 00:01:39.046000
Back to Admin Dashboard

Chapter 5

Summary

This report provides an exploration of various methods, and let's make it more relatable by connecting it to a practical scenario. Imagine your part of a group working on a project, and your goal is to develop an online platform for generating aptitude exams in real-time. Instead of relying on a static set of questions, the system would dynamically fetch questions as needed.

To accomplish this, we would leverage web scraping techniques to retrieve questions from various sources. This dynamic approach not only ensures that the exams remain up-to-date but also minimizes the risk of question paper leaks, which can be a significant concern in the education sector. Furthermore, this innovative system would reduce the workload for examiners who typically spend a lot of time creating question sets manually. Instead, the system would automate the process, generating diverse and relevant question sets efficiently. This not only improves the quality of the exams but also frees up valuable time for educators to focus on other essential tasks.

References

[1] Thakur, S. and Malvani, H. (2021) “Eklavya - Questionnaire generation and Monitoring system.”

[2] Mehta, P. and Jain, P. (2021) “Automated MCQ Generator using Natural Language Processing,” *International Research Journal of Engineering and Technology (IRJET)* [Preprint].

[3] Rahim, A. and Domingo, M.S.T. (2020) “Automated Exam Question Set Generator Using Utility Based Agent and Learning Agent.”

[4] Rahim, A. and Aziz, Z.A. (2017) “Automated exam question generator using genetic algorithm,” *ResearchGate* [Preprint]. Available at: https://www.researchgate.net/publication/326361013_Automated_exam_question_generator_using_genetic_algorithm (Accessed: 2023).

[5] M. Sultan, S.S. and Abdel-Fattah, M.A. (2021) “A Framework for Automatic Exam Generation based on k-means and Genetic Algorithm,” *International Journal of Computer Applications* [Preprint].

Acknowledgement

We would like to express our special thanks to **Prof. Sheetal Shimpikar**, our mini project guide whoguided us through the project and who helped us in applying the knowledge that we have acquired during the semester and learning new concepts.

We would like to express our special thanks to **Dr. Sharvari Govilkar** the H.O.D of our Computer Engineering department who gave us the opportunity to do this mini project because of which we learned new concepts and their application.

We are also thankful to our mini project coordinator **Prof. Sheetal Shimpikar** along with other faculties for their encouragement and support.

Finally, we would like to express our specials thanks of gratitude to Principal **Dr. Sandeep Joshi** who gave us the opportunity and facilities to conduct this mini project.

Rohan Kadam
Shreyas Varadkar
Nisha Sagar
Atharva Bhingarde