PANIMALAR
ENGINEERING COLLEGE
An Autonomous Institution, Affiliated to Anna University, Chennai
A Christian Minority Institution
(JAISAKTHI EDUCATIONAL TRUST)
Approved by All India Council for Technical Education

Department of Computer Science and Engineering

EXPEDITING HR MANAGEMENT VIA DYNAMIC E-PORTFOLIO BY EMPLOYING WEB SCRAPPER

JEEVAN AKSHAY B	(211419104113)
KEVIN CHRISTOPHER A	(211419104140)
MOHAMMED ABDULLAH	(211419104167)

Guide Name
Mr. KARTHIKEYAN A, B.E., M.TECH.,

Coordination Name
Dr. Senthil Kumar M.C.A., M.Phil., M.B.A.,M.E., Ph.D.,

INTRODUCTION

- 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.
- There are technologies such as web scraping which is an automated method to obtain large amounts of data from websites.
- Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications.
- 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.

Objective Of The Project

- 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 .

LITERATURE REVIEW

Author	Title	DOI/ Reference	Findings	Publised
Geunseong Jung, Sungjae Han, Hansung Kim, Kwanguk Kim	Extracting the Main Content of Web Pages Using the First Impression Area	https://doi.org/10.1109/ACCESS.2022.3229080	Extracting the main content from a web using web crawlers and browser reader modes.	14 December 2022
Siew Lee Chang & Muhammad Kamarul Kabilan	Using social media as e-Portfolios to support learning in higher education: a literature analysis	https://doi.org/10.1007/s12528-022-09344-z	Use of (Social Media) SM-based e-Portfolios' potentials to support and facilitate students' learning and development.	16 November 2022
Asih Zunaidah	Meaningful Online Learning with e-Portfolio: University Students' Perspectives	https://doi.org/10.1109/ICET56879.2022.9990059	During emergency online classes, students were engaged in the learning process while utilizing digital technology through e-portfolio.	23 December 2022

PROBLEM STATEMENT

- The job recruiters generally refer to the resume for getting summarized view of candidate’s profile.
- This is generally shown in the printout by the candidate to the recruiter. Thus it has the limitation of just summarizing the information.
- These days, jobs with quality income require the candidates appearing for the interview to apply online.
- Portfolio hosted on web can serve as a good means to tell more about the projects which had been completed by the candidate.
- Other possible alternatives include online social media sites LinkedIn. The LinkedIn app is meant to connect people with similar interests and jobs.
- Thus it is not specialized for displaying the projects and experience in internships.
- The samples of projects can be displayed via the website. Generally, resume is meant to be a summary of the candidate’s profile.
- On the other hand, portfolio is meant to be more descriptive about the candidate. The achievements of candidates might contain some extra information about the learnings and how the issues were resolved.

DEVELOPMENT ENVIRONMENT

SOFTWARE REQUIREMENTS

The following are the packages and libraries required to build the application:

1. Node.js – version (16.15.0 LTS)
2. React-router-dom@6
3. Firebase (OAuth, Storage)
4. ExpressJS (REST API)
5. Socket.IO (Socket Programming for sharing real-time content)
6. Puppeteer (To launch browser in headless mode)

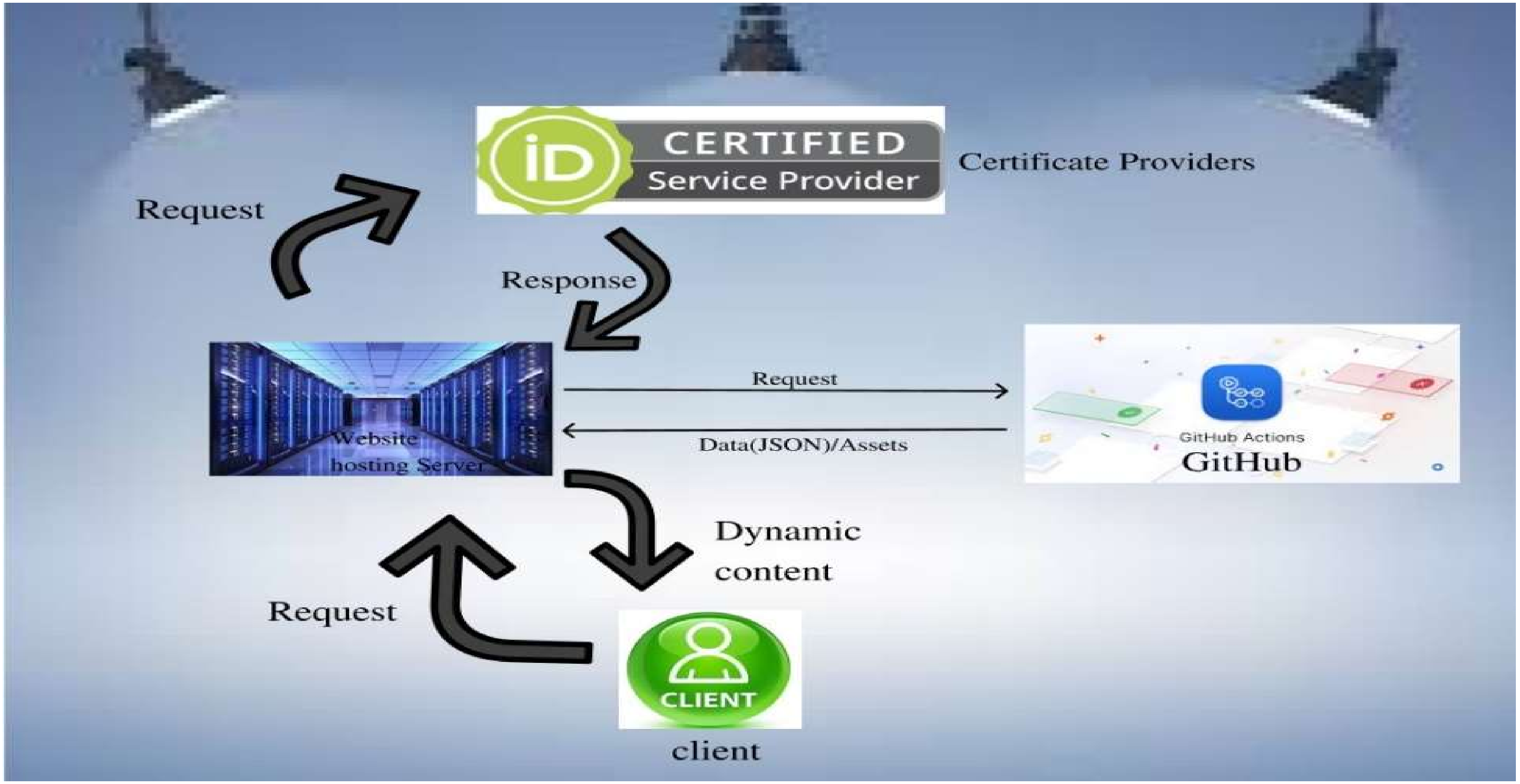
Other npm packages included in the application as dependencies are as follows:

- @mui/material
- @emotion/react
- @emotion/styled

HARDWARE REQUIREMENTS

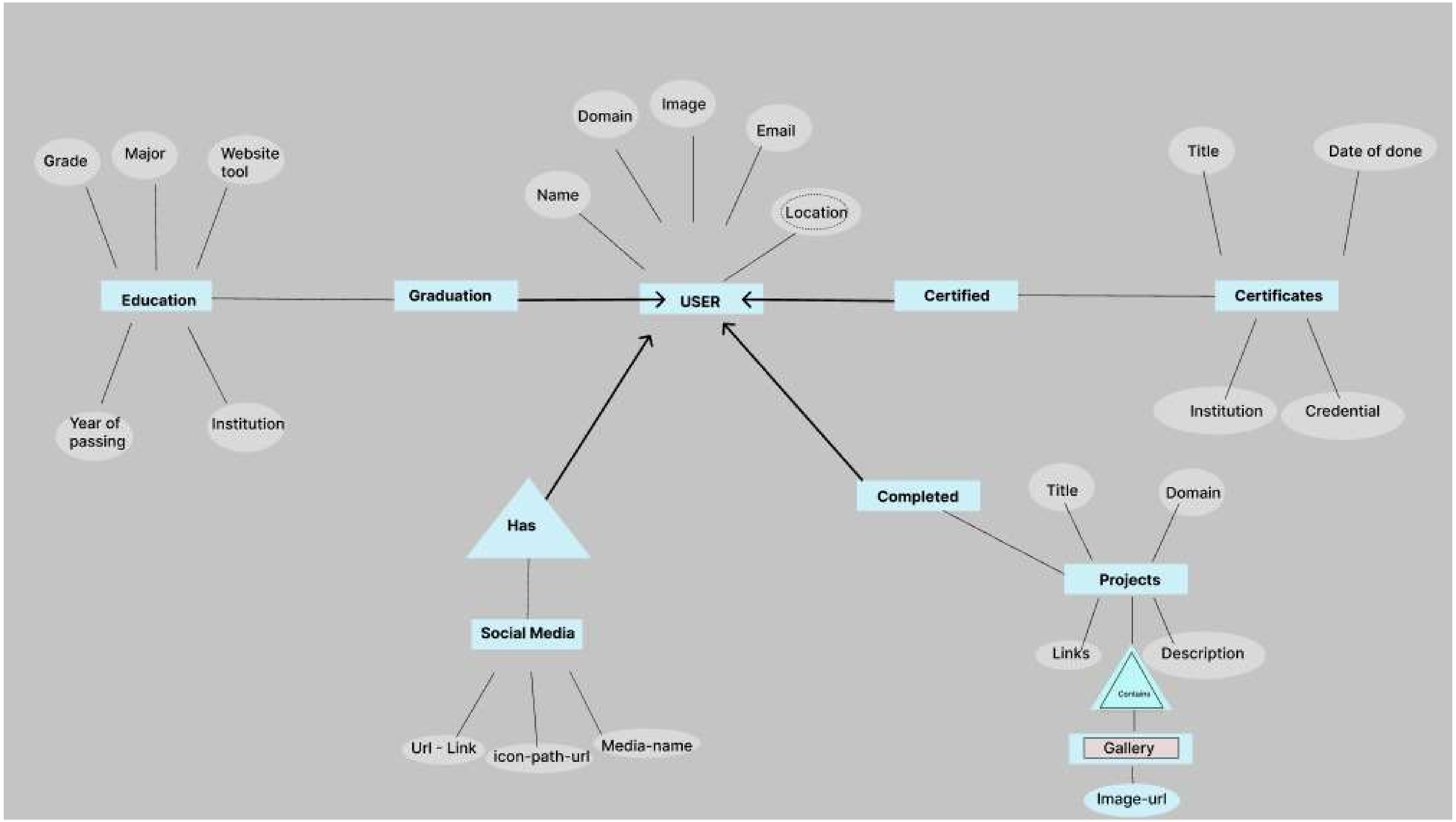
1. 8 GB RAM (To run puppeteer)
2. Network interface to connect to internet,
   - Download Speed: 25 Mbps
   - Upload Speed: 5 Mbps.

SYSTEM ARCHITECTURE



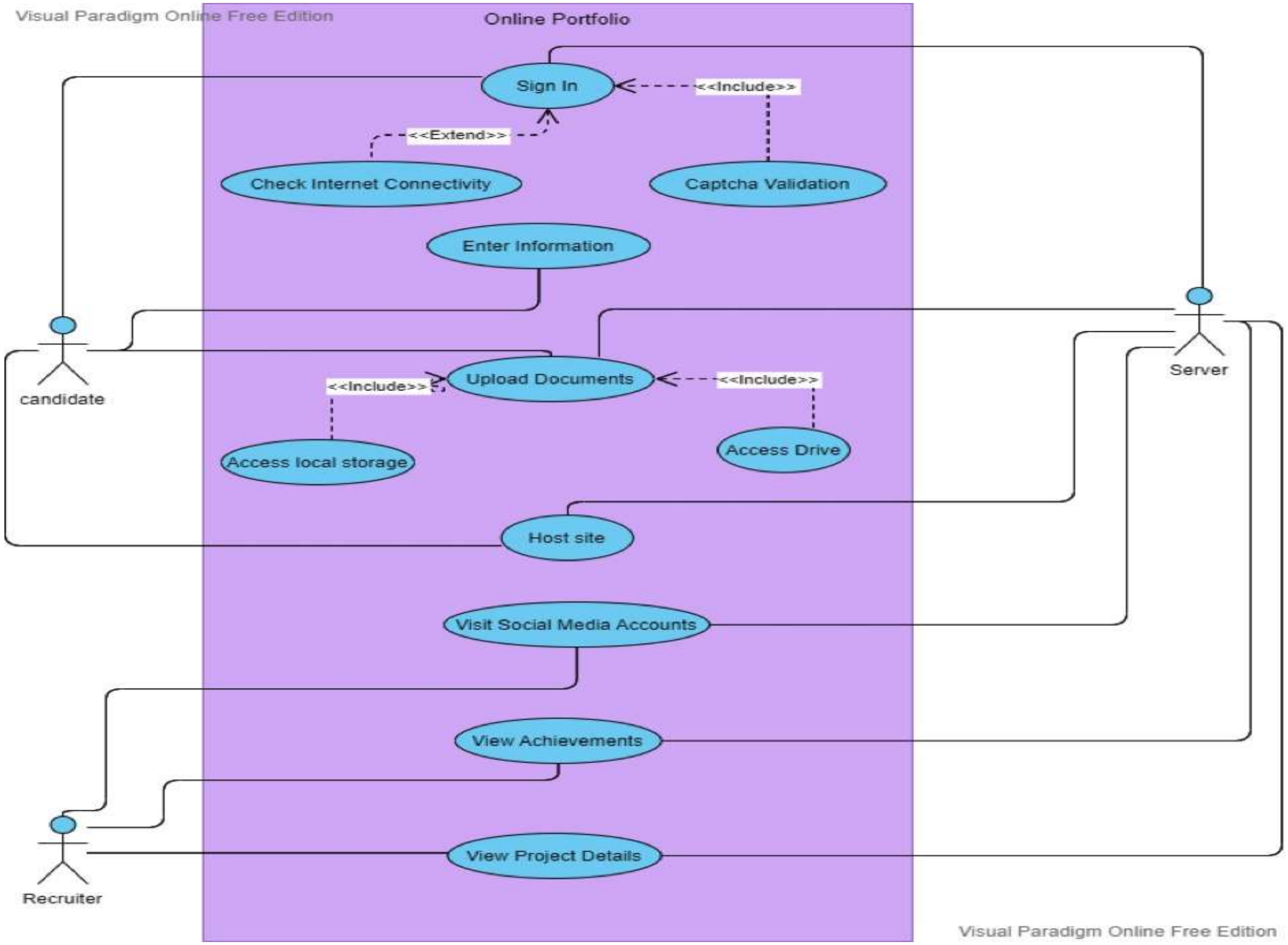
SYSTEM DESIGN

E-R Diagram



SYSTEM DESIGN

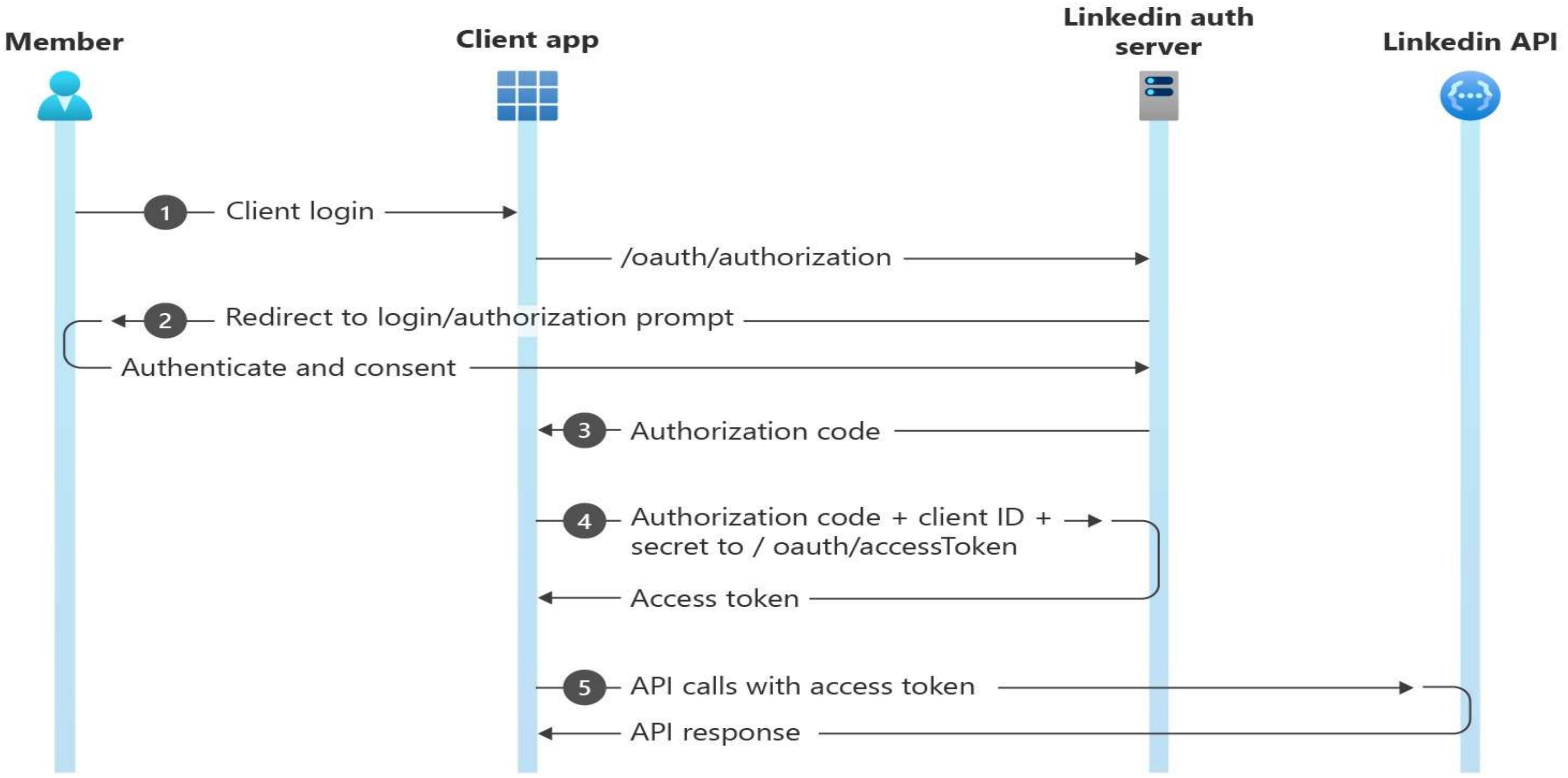
USE CASE DIAGRAM



SYSTEM DESIGN

SEQUENCE DIAGRAM

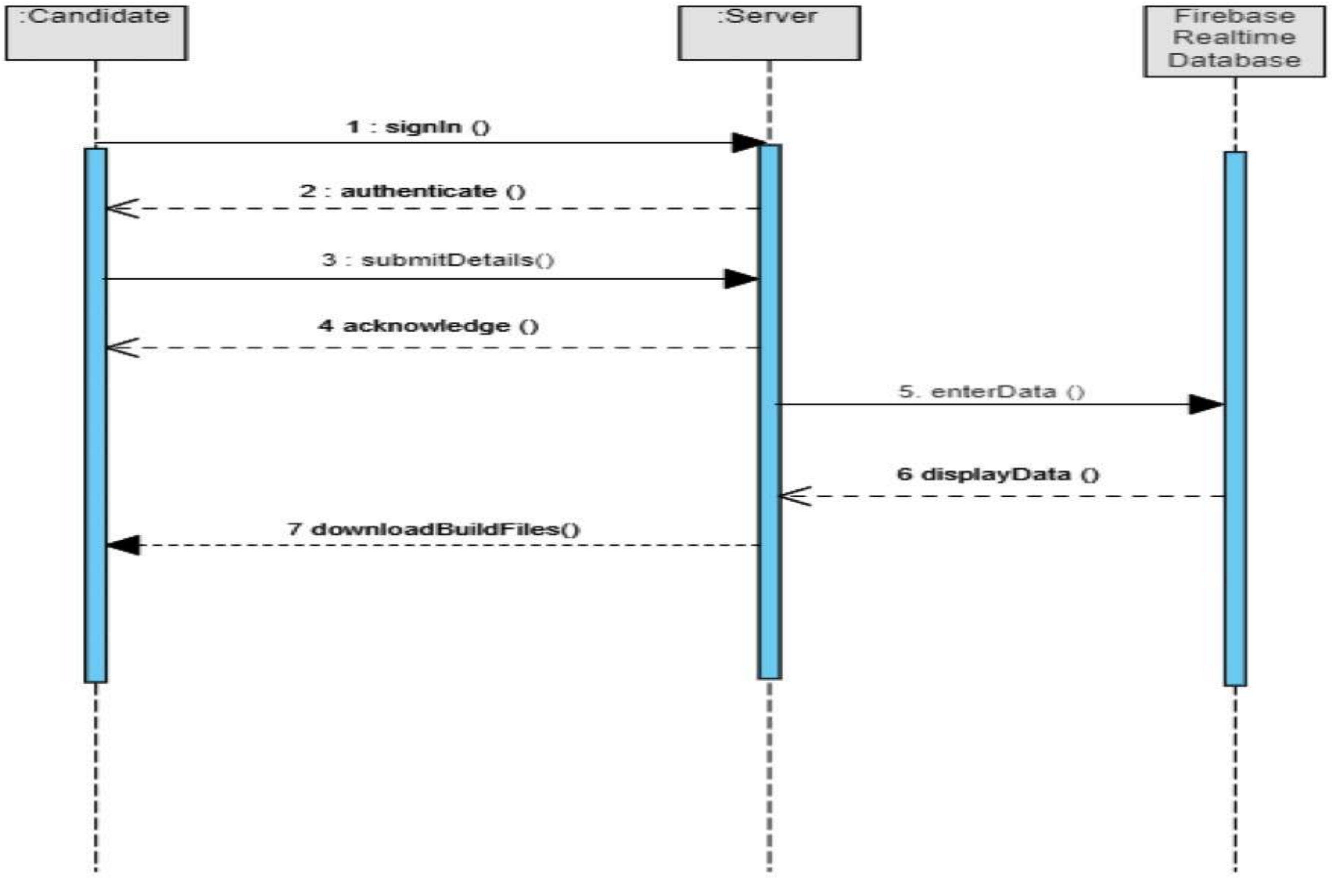
For Authentication (From Linkin API):



SYSTEM DESIGN

SEQUENCE DIAGRAM

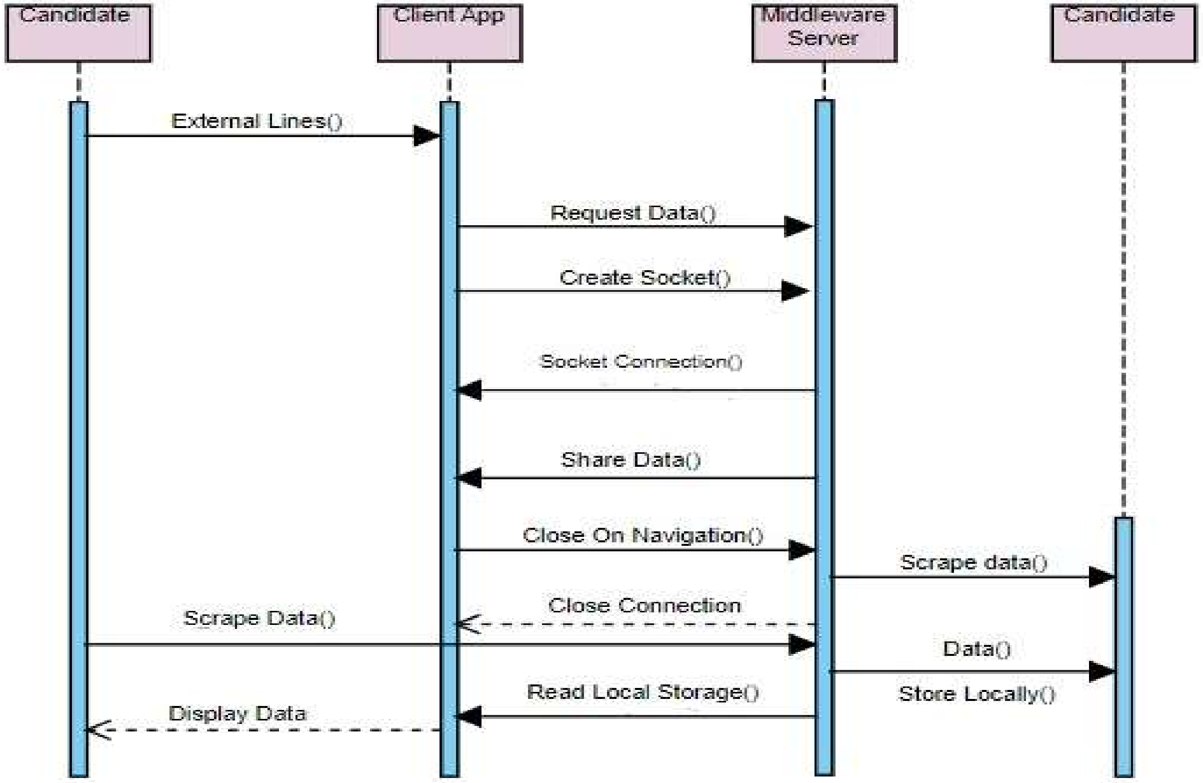
For storing and retrieving data



SYSTEM DESIGN

SEQUENCE DIAGRAM

Dynamically display Github and related data & scrape data from linkedIn



MODULE DESRIPTION

FIREBASE

LOCATION:

```
Root-folder
    |--firebase.js
    |--src
        |--App.js
        |--firebase-config
```

FUNCTION:

- To perform authentication for existing user and to register a new user.
- To store and retrieve data from Realtime Database.
- To host the website in the server

DEPENDENCIES:

- Firebase CLI
- NPM modules:
  - @firebase/app
  - @firebase/auth
  - @firebase/database

MODULE DESRIPTION

WEB SCRAPPER CLIENT

LOCATION:

```
Root-folder
    |--src
          |--Components
                 |--Callback
                 |       |--CallbackHandler.js
                 |--Form
                 |        |--Form.js
                 |--Projects
                          |--Projectpage.js
                          |--Model.js
```

FUNCTION:

- To receive and display images from middleware server in projects.
- To scrape data from linkedIn

DEPENDENCIES:

- NPM modules:
  - @express
  - @puppeteer
  - @socket.io-client

MODULE DESRIPTION
MIDDLEWARE SERVER

LOCATION:

Root-folder

|--*index.js*

|--screenhot.png

|--*screen.shooter.js*

FUNCTION:

- Screenshot.png is used for store the screenshots taken by puppeteer, which is later sent to the client
- This process is achieved with established socket connection.

DEPENDENCIES:

- NPM modules:
  - @puppeteer
  - @socket.io

MODULE DESRIPTION

WEBSCRAPER SERVER

LOCATION:

```
Root-folder
    |--data
          |--data.js
    |--index.js
```

FUNCTION:

- Query Selectors are used to select the content from the web page and store it in the files under data directory.
- Express server listens on a port for request to provide the data from files
- This content is retrieved from the data directory when a GET request is made.
- Promises are used to execute asynchronously

DEPENDENCIES:

- NPM modules:
  - @puppeteer
  - @socket.io
  - @express

MODULE DESRIPTION

REACT BROWSER ROUTER

LOCATION:

```
Root-folder
    |--src
        |--App.js
        |--Components
              |--Home
                   |--Home.js
```

FUNCTION:

- There are two routers present in the application.
- The one present in App.js handles the login, registration and form page navigation.
- The browser router present in Home.js handles the navigation when the user navigates through the portfolio.

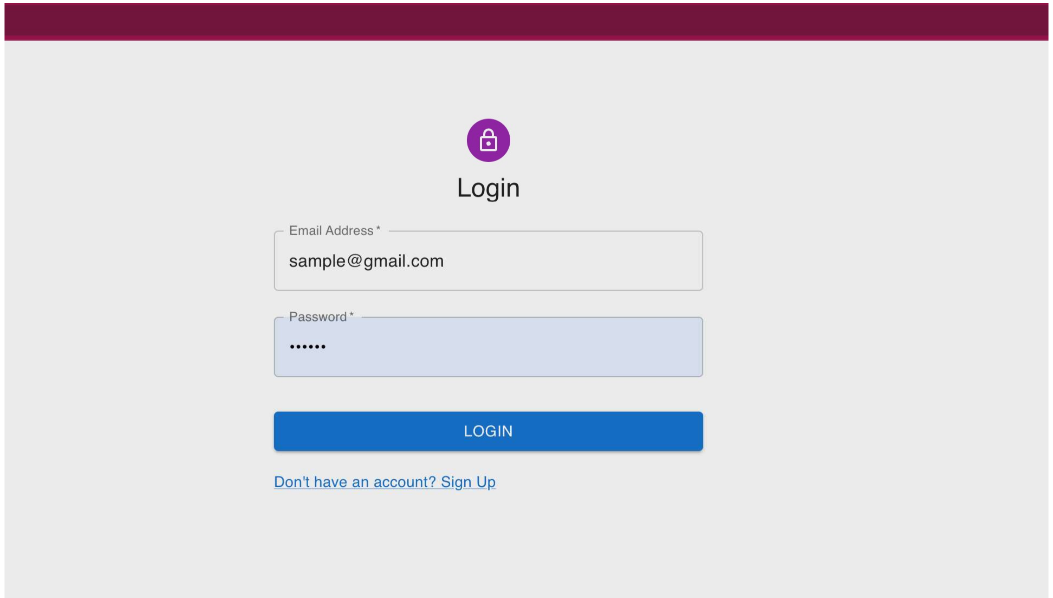
DEPENDENCIES:

- NPM modules:
  - @browser-router

TESTING

TEST CASE ID	TEST SCENARIO	TEST CASE	PRE-CONDITON	TEST STEPS	TEST DATA	EXPECTED RESULT	POST CONDITION	ACTUAL OUTCOME	STATUS (PASS/ FAIL)
1.	Verify username detail	Enter Credentials	Should have a valid portfolio account	1.Enter username 2.Click submit	<Valid login username>	Verify login access with correct credentials	Navigate to form data page	Logged in page is shown	Pass
2	Verify information	Enter invalid credentials	Should have a valid portfolio account	1.Enter Username 2.Enter password 3. Click submit	<Invalid credentials	Error message	Stay in same page	Error message	pass
3.	Verify password detail	Enter Password	Should have a valid portfolio account	1. Enter Pasword 2. Wait for 200ms for auto change function to complete	<Valid password >	Verify whether it is minimum of 6 characters	Show data page	Error message is hidden	Pass
4	Click the submit button for the form	Click submit Button	Should have filled both username and password	1.Enter required fields. 2.Enter submit	<Valifd button field>	Check whether button is enabled	Show login page	Button is enabled	Pass

SCREENSHOTS



Login

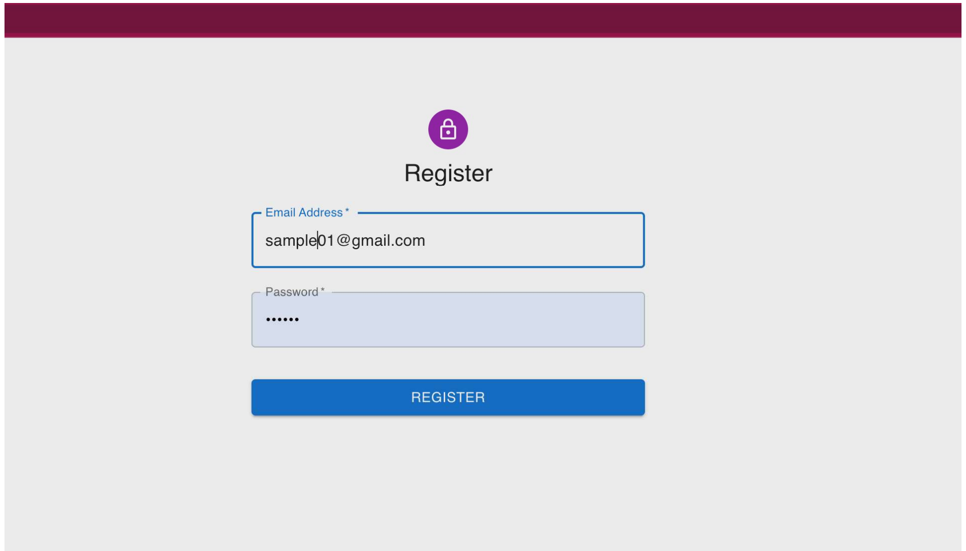
Email Address *

sample@gmail.com

Password *

••••••

Fig 1. Registration Screen



Register

Email Address *

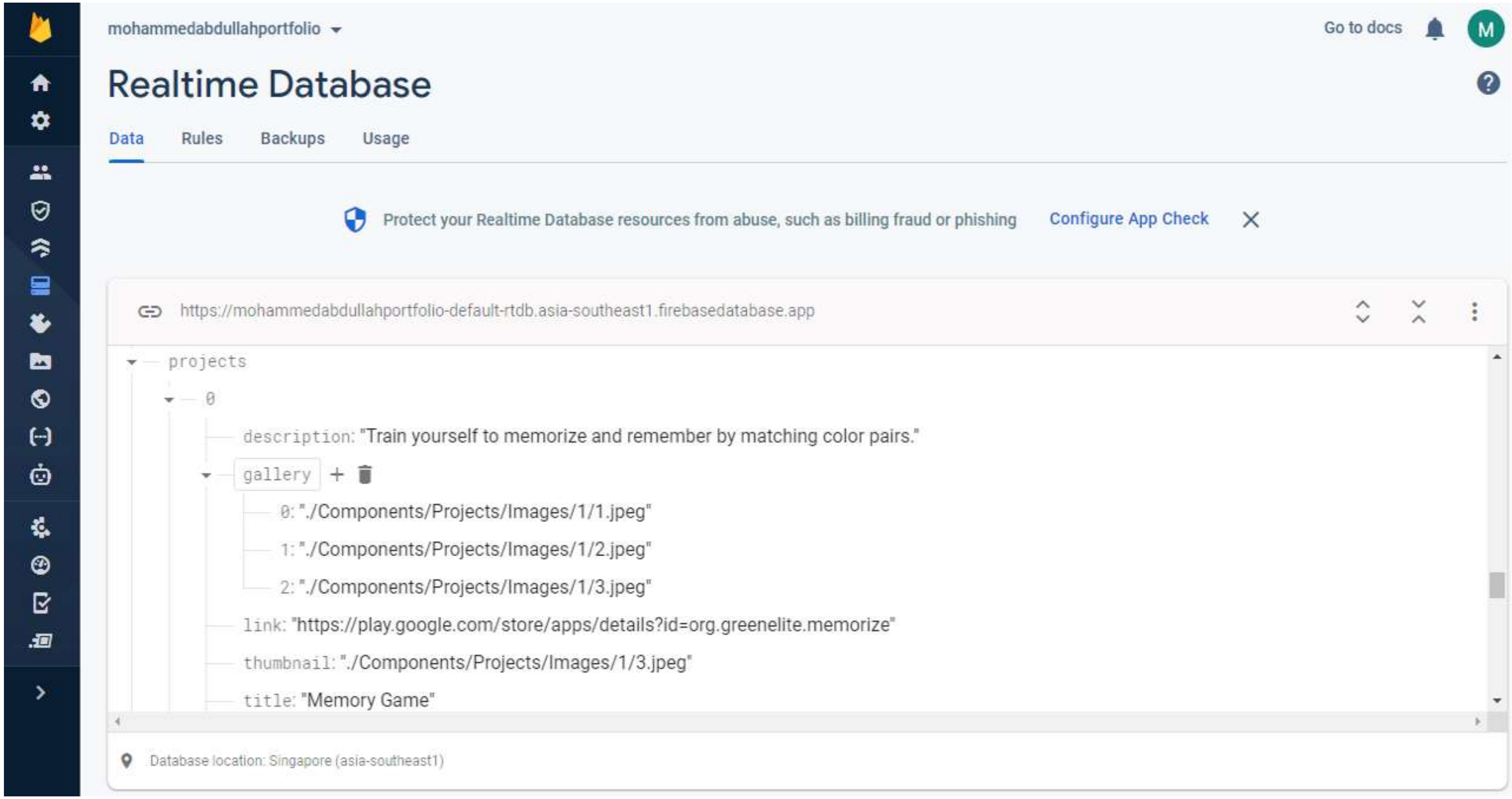
sample01@gmail.com

Password *

••••••

Fig 2. Login Screen

SCREENSHOTS



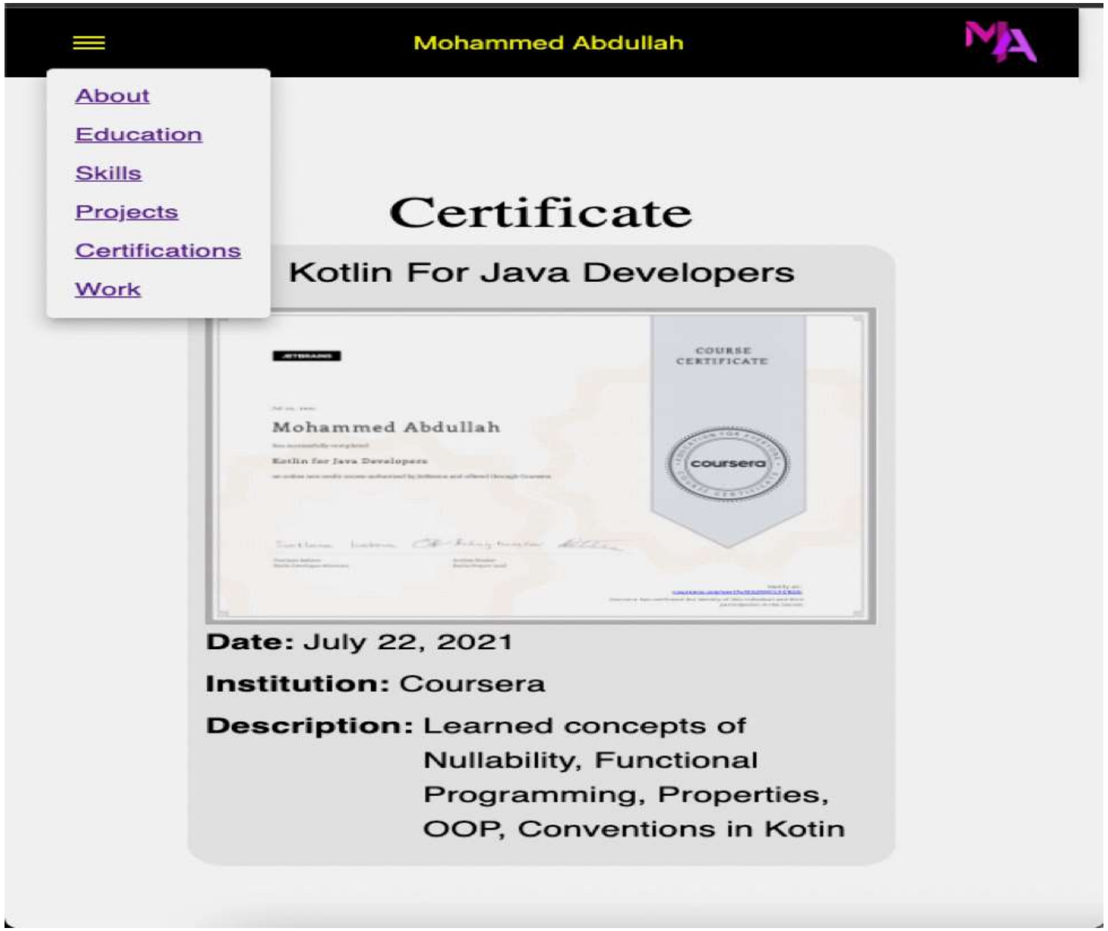
mohammedabdullahportfolio
Go to docs
M
Realtime Database
Data
Rules
Backups
Usage
Protect your Realtime Database resources from abuse, such as billing fraud or phishing
Configure App Check
https://mohammedabdullahportfolio-default-rtdb.asia-southeast1.firebasedatabase.app

```
projects
  0
    description: "Train yourself to memorize and remember by matching color pairs."
    gallery
      0: "./Components/Projects/Images/1/1.jpeg"
      1: "./Components/Projects/Images/1/2.jpeg"
      2: "./Components/Projects/Images/1/3.jpeg"
    link: "https://play.google.com/store/apps/details?id=org.greenelite.memorize"
    thumbnail: "./Components/Projects/Images/1/3.jpeg"
    title: "Memory Game"
```

Database location: Singapore (asia-southeast1)

Fig 3. Store data in Firebase BaaS

SCREENSHOTS



Certificate

Kotlin For Java Developers

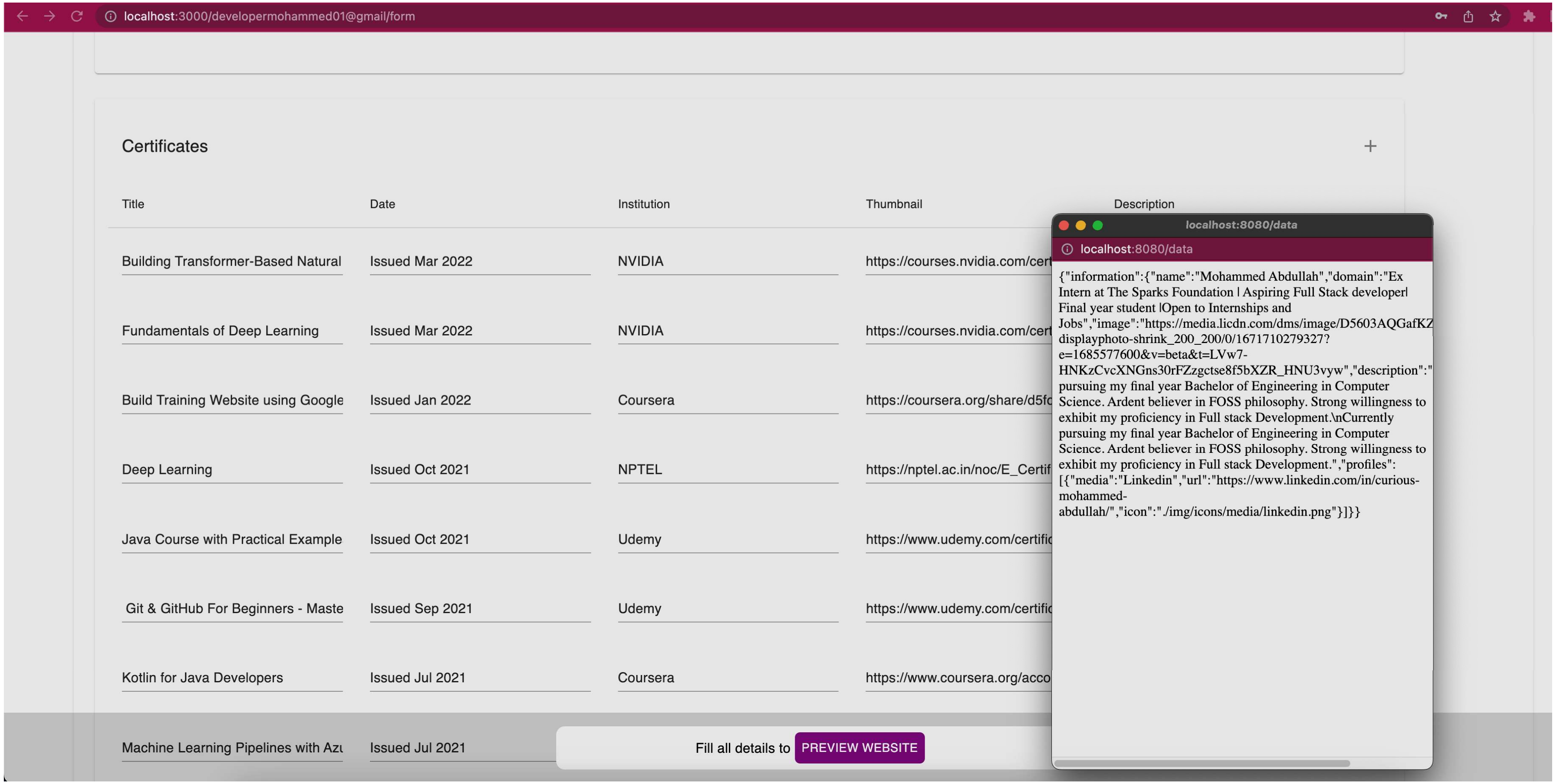
Date: July 22, 2021

Institution: Coursera

Description: Learned concepts of Nullability, Functional Programming, Properties, OOP, Conventions in Kotin

Fig 4. Supported Mobile Responsiveness using MUI

SCREENSHOTS

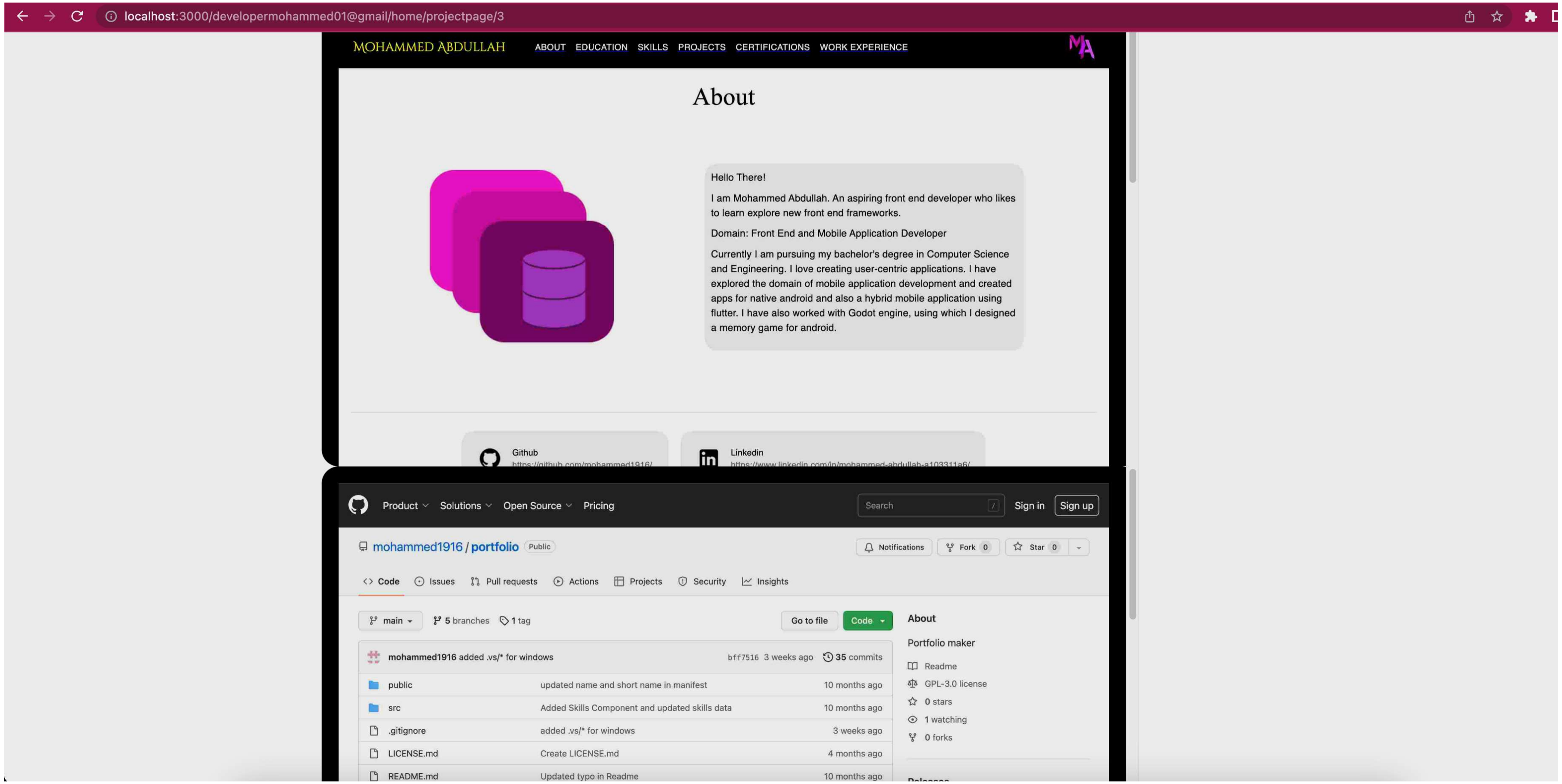


Title	Date	Institution	Thumbnail	Description
Building Transformer-Based Natural	Issued Mar 2022	NVIDIA	https://courses.nvidia.com/cert	
Fundamentals of Deep Learning	Issued Mar 2022	NVIDIA	https://courses.nvidia.com/cert	
Build Training Website using Google	Issued Jan 2022	Coursera	https://coursera.org/share/d5fc	
Deep Learning	Issued Oct 2021	NPTEL	https://nptel.ac.in/noc/E_Certif	
Java Course with Practical Example	Issued Oct 2021	Udemy	https://www.udemy.com/certific	
Git & GitHub For Beginners - Maste	Issued Sep 2021	Udemy	https://www.udemy.com/certific	
Kotlin for Java Developers	Issued Jul 2021	Coursera	https://www.coursera.org/acco	
Machine Learning Pipelines with Azu	Issued Jul 2021			

```
{"information":{"name":"Mohammed Abdullah","domain":"Ex
Intern at The Sparks Foundation | Aspiring Full Stack developer|
Final year student |Open to Internships and
Jobs","image":"https://media.licdn.com/dms/image/D5603AQGafKZ
displayphoto-shrink_200_200/0/1671710279327?
e=1685577600&v=beta&t=LVw7-
HNKzCvcXNGns30rFZzgctse8f5bXZR_HNU3vyw","description":"
pursuing my final year Bachelor of Engineering in Computer
Science. Ardent believer in FOSS philosophy. Strong willingness to
exhibit my proficiency in Full stack Development.\nCurrently
pursuing my final year Bachelor of Engineering in Computer
Science. Ardent believer in FOSS philosophy. Strong willingness to
exhibit my proficiency in Full stack Development.","profiles":
[{"media":"Linkedin","url":"https://www.linkedin.com/in/curious-
mohammed-
abdullah/","icon":"./img/icons/media/linkedin.png"}]}}
```

Fig 5. Auto-fill using scrapped data

SCREENSHOTS



About

Hello There!

I am Mohammed Abdullah. An aspiring front end developer who likes to learn explore new front end frameworks.

Domain: Front End and Mobile Application Developer

Currently I am pursuing my bachelor's degree in Computer Science and Engineering. I love creating user-centric applications. I have explored the domain of mobile application development and created apps for native android and also a hybrid mobile application using flutter. I have also worked with Godot engine, using which I designed a memory game for android.

mohammed1916 / portfolio Public

About

Portfolio maker

mohammed1916 added .vs/* for windows	bff7516 3 weeks ago	35 commits
public	updated name and short name in manifest	10 months ago
src	Added Skills Component and updated skills data	10 months ago
.gitignore	added .vs/* for windows	3 weeks ago
LICENSE.md	Create LICENSE.md	4 months ago
README.md	Updated typo in Readme	10 months ago

Fig 6. Dynamic viewing of content using socket connection and puppeteer
Top: Project hosting
Bottom: GitHub ReadMe.md file

CONCLUSION

- The system makes convenient the process of maintaining the portfolio by maintaining concurrency of data with other websites referenced by the web scraper.
- Any person without the knowledge of the domain can work with the project due to the presence of documentation.
- The application thus makes use of the technology of backend as a service efficiently, which provide a way to scale the project with any new data.
- Authentication of the website ensures that no unauthorized users access the write access grant of the data.

The significant issues addressed by this system relate to the following:

- Dynamic fetching of data.
- Real-time updates of the information in referenced websites
- Easy verification of certificates with active available credentials
- Updated information
- No maintenance required, i.e., initial setup of data is sufficient.

PUBLICATION

Our Proposed Paper :

- Name: Journal of Survey in Fisheries Sciences

- Scopus Index: https://www.scopus.com/sourceid/21100905326

- Karthikeyan A, Mohammed Abdullah, B Jeevan Akshay & Kevin Christopher A. (in press). "Expediting HR Management Via Dynamic E-Portfolio by Employing Web Scrapper," Journal of Survey in Fisheries Sciences, Vol. 10 No. 4S (2023): (Special Issue4). http://sifisheriessciences.com/journal/index.php/journal/article/view/1172/1184

REFERENCES

[1] 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.

[2] Chang, S.L., Kabilan, M.K. Using social media as e-Portfolios to support learning in higher education: a literature analysis. *J Comput High Educ* (2022). https://doi.org/10.1007/s12528-022-09344-z

[3] 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.