Development of a Web Archival System for Universities in Nigeria

Mutiat A. Ogunrinde[1] and Abdullateef A. Adeleye [2]
Department of Mathematical and Computer Sciences, Fountain University, Osogbo, Osun State, Nigeria.12
bogunrinde@gmail.com1, adeleyeabdullateef@gmail.com

Received: 12-DEC-2023; Reviewed: 18-MARCH-2024; Accepted: 27-SEPT -2024
https://dx.doi.org/10.4314/fuoyejet.v9i3.6

ABSTRACT- Web archiving is the process of gathering up data that has been recorded on the World Wide Web, storing it, ensuring the data is preserved in an archive, and making the collected data available for future research. A Web archive system is a traditional subject for preserving web content for the future, and its importance is getting more significant due to the explosive growth of web content. As good as it is, Universities in Nigeria do not have their intellectual properties put together on the same platform, making accessibility difficult for students and researchers. The primary purpose of this work is to develop a central repository website capable of assisting universities in Nigeria to upload their final-year project so that students worldwide can access it and know if such projects are already in existence. The web archival system was developed using the Django framework (Python, HTML/CSS) and MySQL as a database server. The output of this website allows different Universities to add their past project to make it easier for Students to search if their projects are already in existence and have more knowledge about how projects are being done. This work will give Nigerian Universities more visibility and promote interoperability.

Keywords: Web, Archive, Website, Project, Documents, Collection

1. INTRODUCTION

A website is a collection of files and related resources accessible worldwide and organized under a particular domain name. (Zhongke & Jun, (2023), Anthony, (2013)) Typical files found on a website are HTML documents with their associated graphic image files (Gif, Jpegs, etc.), scripted programs (in Perl, PHP, Java, etc.), and similar resources. The site's files are usually accessed through hypertext or hyperlinks embedded in other files. (Tutorials point, 2017). The Web is a precious source of information, available to all and sundry, without the exception of universities. Most Government and Private Institutions are involved in archiving parts of their intellectual materials on the web for various purposes, such as making them freely managed and kept for use by the general public, historians, researchers, and future generations (Anthony, 2013).

Web archiving is gathering information posted on the Internet, preserving it, ensuring that it is maintained, and making the gathered information accessible for upcoming research (Costa *et al*., 2017; Niu, 2018), which has begun since 1996.

The Internet achieved, and several national libraries have initiated and observed Web archiving practices since 2001 (Niu, 2018). Since its inception in 2001, the International Web Achieving Workshop (IWAW) has offered a forum for knowledge sharing and idea exchange. International collaboration on building standards and Open-Source technologies for online archives has been dramatically enhanced by the later formation of the International Internet Preservation Consortium (IIPC) in 2003. More and more libraries and archives will inevitably have to deal with the issues of Web archiving due to these advancements and the expanding amount of human culture that is created and recorded on the Web.

Some businesses preserve their web material with basic technologies and procedures. In-depth archiving of culturally significant Web information is a project that includes participation from national libraries, national archives, and numerous parties and organizations. Organizations that must archive their web information for regulatory, legal, historical, or commercial reasons have access to commercial Web archiving tools and services. The Internet Archive, which intends to preserve a complete World-Wide Online archive, is the largest web archiving group currently crawling the Web.

Currently, in Nigerian universities, there is no central repository for all projects done at Nigeria University, be it at the Undergraduate and Postgraduate level, so many students don't know if their project is already in existence in other universities and to what extent. Similarly, there is little or no exchange of information between students of different universities, which has created a severe bottleneck. This paper aimed to develop a web-based central repository system capable of assisting Universities in Nigeria in putting together their final year project so that students worldwide can access it for intellectual purposes and know if their topics have already been done and to what the degree to which they covered. The remaining part of this paper covers related works, the methodology used, provides an overview of web archiving, explains the methods employed, examines various difficulties experienced while web archiving, and offers potential solutions.

2. 2.0 REVIEW OF RELATED WORKS

Web archiving frameworks are often evaluated based on the quality and scalability of their

*Corresponding Author: bogunrinde@gmail.com

Section B- ELECTRICAL/COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Ogunrinde A. and Adeleye A. (2024), Development of a Web Archival System for Universities in Nigeria. FUOYE Journal of Engineering and Technology (FUOYE**JET), 9(3), 406-416.**
https://dx.doi.org/10.4314/fuoyejet.v9i3.6

preserved records, according to Martin *et al*. (2019). The authors argue that the prevalence of dynamic web content poses a significant challenge to scale-optimized crawler-based systems such as the Internet Archive. Web recorder tool and other human-driven services provide high-quality archival captures, however they are not scalable. Scalability and archival quality are balanced by the Memento Tracer framework, which was introduced in the study. It evaluated the extent and archival significance of its operation while describing the concept and design. The quality was found to be comparable to or better than that of reputable archiving systems, and its overhead when used in large quantities was manageable.

An Augmented Reality (AR) interface was presented by Matsushita (2017) and is intended to be used during archeological research. In order to enable regular discussions via a 3D virtual field site that users can visit, examine, and incorporate information into over the Internet, the study developed a 3D archiving system that automatically refreshes and shares information on outcomes. In this context, an augmented reality (AR) is defined as a user interface that enhances mobile device access to the archiving system on-site. Using SFM (structure from motion), a photo taken on the site can be consistently linked to the pre-registered photo sets in the archiving system. The archived data is then quickly layered on the photo, resembling the virtual site representation that users see on desktop PCs.

Furthermore, Purmono *et al*. (2019) used SFM (Structure from Motion) and the Perspective-n-Point (PnP) problem to demonstrate an image taken there. This project intends to build a system of acceptance for new students using a website-based technique to assist potential new students in the topic and make it easier to get information on new student registration. When accepting new pupils, Al Amanah Vocational School Pasar Kemis continues to manage data in traditional ways. Following that, the registration data is entered into a book, and reports are made by hand, recapitulating the data, resulting in archival loss and a lengthy report-creation procedure. Prospective learners arrive at the institution directly and fill out the registration form.

Similarly, Web archive systems are considered standard techniques for archiving web content in the future by Hwang *et al*. (2020). It is difficult to discern the original from the content because content integrity cannot be guaranteed. A web archive system's shortcomings make it susceptible to a claim of content integrity.

Additionally, Lupfer *et al*.'s (2019) study looks into new media as a way to enhance how student teams construct and arrange artifacts while doing design. While some design artifacts, like previous work, reference photos, and code framework repositories, are ready-made, other types, like storyboards, mock-ups, prototypes, and user study findings, are created independently. We looked at the methods computer science students used to gather information, put together, and present their team-based design projects using free-form web curation. Based on our investigation using mixed qualitative methodologies, we discovered that their creative processes of contextualization and communication relied heavily on their use of space and size. In multiscale design curation, prefabricated and handmade design artifacts are gathered, assembled as elements in a continuous space, and layers of visual scale are used for communication, creation, thinking through, exhibiting (presenting), and documenting the design process. A constructivist approach is realized through multiscale design curation, which emphasizes the importance of design process representation. The open and unstructured nature of student curations helps to prevent early formality and encourages introspection during iterative design processes. By using visual chunking to capitalize on human spatial cognition, multiscale design curation facilitates collaborative articulation work in integrated space as well as creative activities.

In a similar vein, Niu (2019) stated that the process of producing a web archive comes with a number of problems. To help future practitioners overcome these challenges, library and information institutions should make sure that their curricula include education in web archiving techniques and abilities. The author performed a thorough literature review before creating a web archiving course. This paper reports the findings and offers the author's perspective on some of the approaches used, such as organizing and describing preserved web resources using typical archive management principles and theories.

In a project published in 2018, Lee & Wang compare the technical infrastructure, access models, and collection development plans of national libraries across the globe to analyze web archiving projects. It highlights recurring patterns and obstacles that national libraries encounter to preserve and make available web-based resources for scholarly inquiry and cultural heritage purposes.

Additionally, Wong (2020) explores the potential and difficulties of using archived web content in empirical research and using web archives as repositories for research data. The study examines web-based data access, analysis, and interpretation techniques with implications for multiple academic fields.

Li (2019) examines the best techniques for capturing and conserving online content, technological developments, and the function of web archiving in digital preservation initiatives. The study

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

emphasizes the crucial cooperation between organizations and stakeholders to guarantee the sustained availability of web-based resources.

3. METHODOLOGY

Agile methodology has become one of the most popular approaches in software development, valued for its flexibility, collaboration, and adaptability. Unlike traditional linear models, Agile emphasizes iterative progress through short cycles known as sprints. Each sprint typically lasts two to four weeks and aims to deliver a functional piece of software, enabling teams to release new features frequently and gather feedback quickly.

Agile methodology was employed to realize the objective of this study. This makes the development work closely with stakeholders to prioritize tasks based on user needs. This ensures that the product being developed remains aligned with the study's goals and customer expectations. Instead of a rigid plan, Agile allows teams to adapt to changes in requirements, making it easier to respond to shifting market demands or evolving project goals.

The steps taken were as follows:

1. Requirements Gathering: In this phase, the requirements were defined, and the time and effort required for the project were also discussed. The information gathered was analyzed to determine the software boundaries and technical feasibility.
2. Design the Requirements: Following the Requirements Gathering, the researcher worked with stakeholders design the system architecture through defined requirements. The system flowchart and UML diagram were used to determine how the system will be incorporated into your existing University system through the Library with other users.
3. Develop: The implementation followed the Model, View, and Controller (MVC) model. It was done in phases, and the product will undergo different stages of improvement using simple and minimal functionality.
4. Test: This system was tested on the developer's side by checking its performance, and reported bugs during this phase were fixed.
5. Deployment: the product is released to the user to interact with it and give feedback.

3.1 SYSTEM FLOWCHART

The diagram Figure 1 illustrates how data are communicated between the users, admins, and the software.

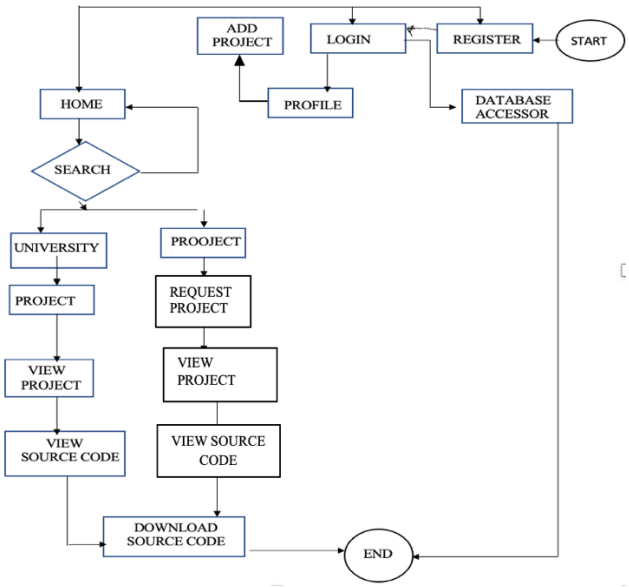

Figure 1: SYSTEM FLOWCHAR

3.2 SYSTEM ARCHITECTURE

Below is the system architecture, which shows a graphical representation of the system, views, and database. It also shows the connections and interactions that can be carried out in the system, where users search for or collect data requested from the software backend.

The system architecture shows how users (the student and the Librarian) send requests to the system. The whole system consists of the Front End (which is what the users see) and the back end that house the repository. The details is shown in Figure 2.

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

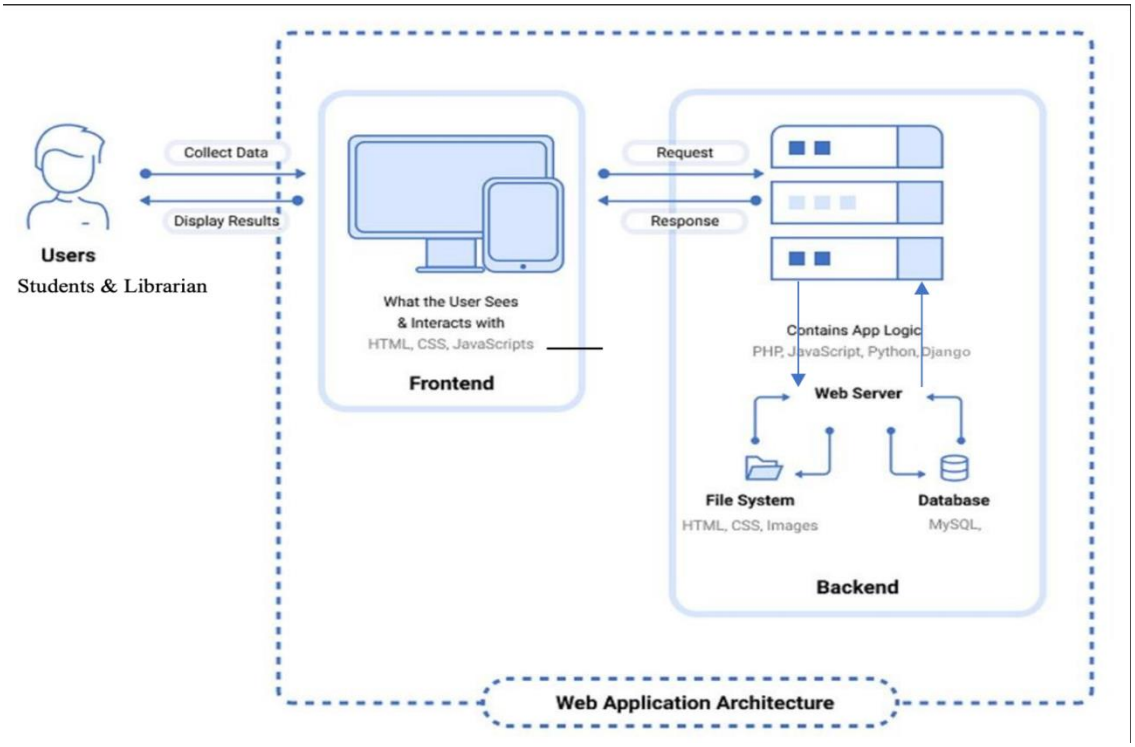

Figure 2: SYSTEM ARCHITECTURE

3.3 DATA FLOW DIAGRAM (DFD):

A Data Flow Diagram (DFD) maps out the flow of information for any process or system. Figure 3 shows the data flow diagram of the system from when the user login or visited the site till when the resources needed are gotten. It uses different symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination.

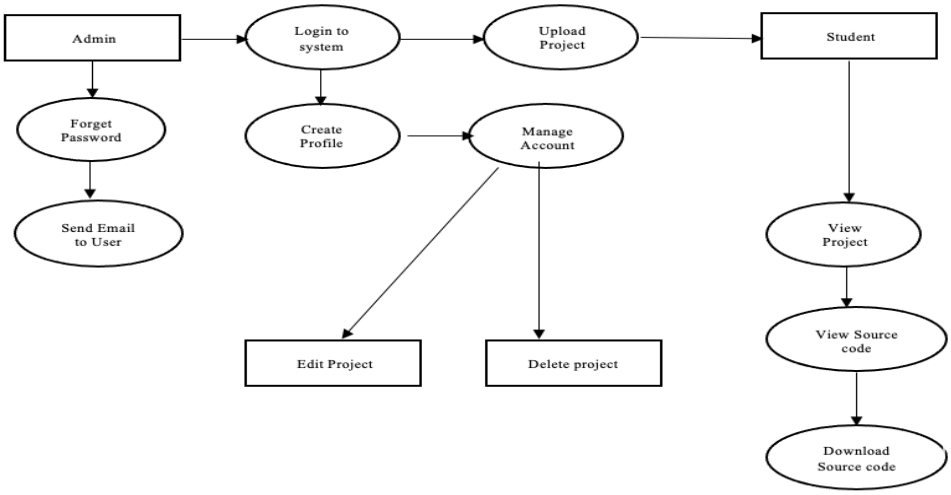

Figure 3: DATAFLOW DIAGRAM

3.4 SYSTEM USE CASE DIAGRAM

The Use Case diagram depicts the interactions between the users and the software system. It shows the functions of the system in chronological order and identifies the specific users that interact with them as illustrated in figure 4.

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

Figure 4: SYSTEM USE CASE DIAGRAM

6. IMPLEMENTATION & RESULTS

The implementation was done following the Model, View, and Controller (MVC) model. The whole implementation was in phases. This work has produced and deployed a web archival system, .

4.1 MAIN FEATURES

4.1.1 SEARCH BAR FOR PROJECTS

The search bar was developed so that users can simply input some keyword(s) related to the topic like the University name, project name, skill name

where Universities in Nigeria can upload their past and present projects so students all over the world can access it from their comfort zones without been physically available in the University.

or University where the project is being done for their specific needs. This decision gives the system a more organized and refined feel and increases efficiency and ease of use. Below is a view of the search engine, which is the first page to be seen after logging in. Figure 5 to 8 shows the details.

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

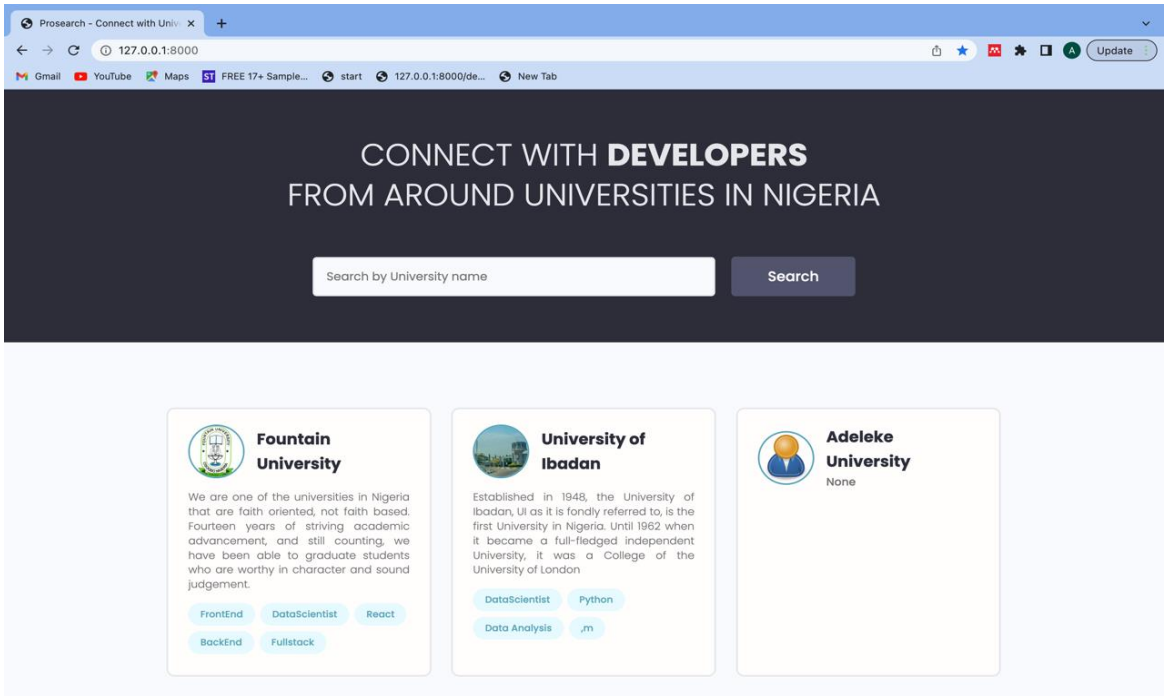

Figure 5: Search engine for Universities

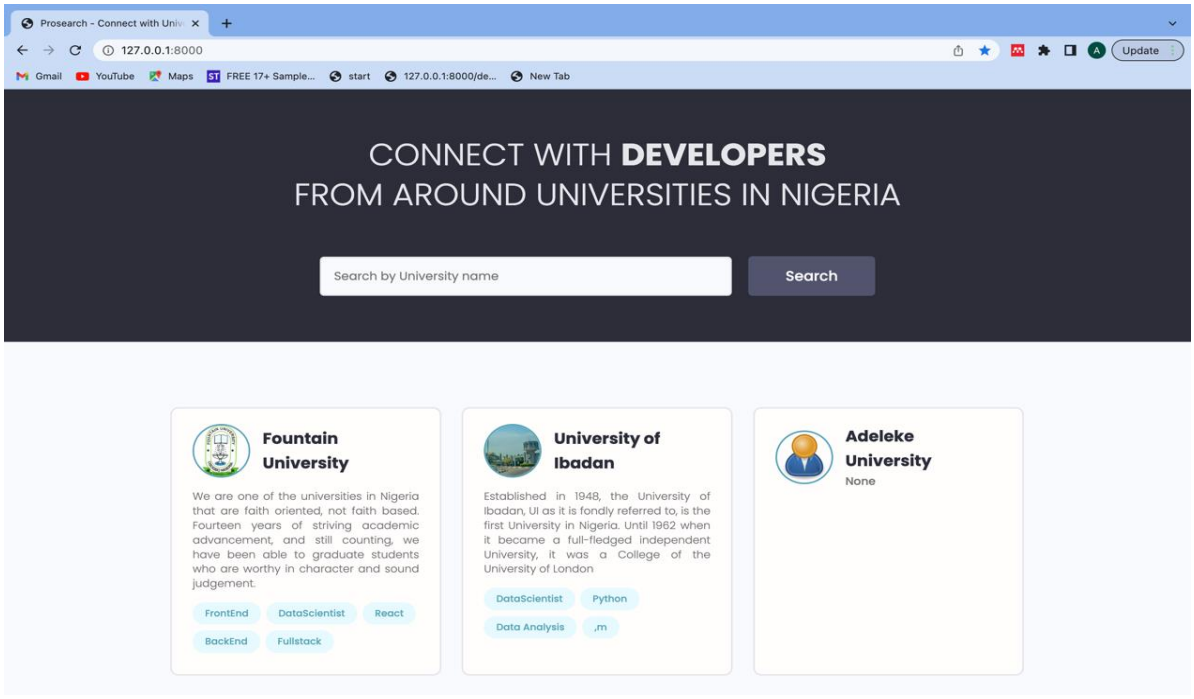

Figure 6: Search Engine for project

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

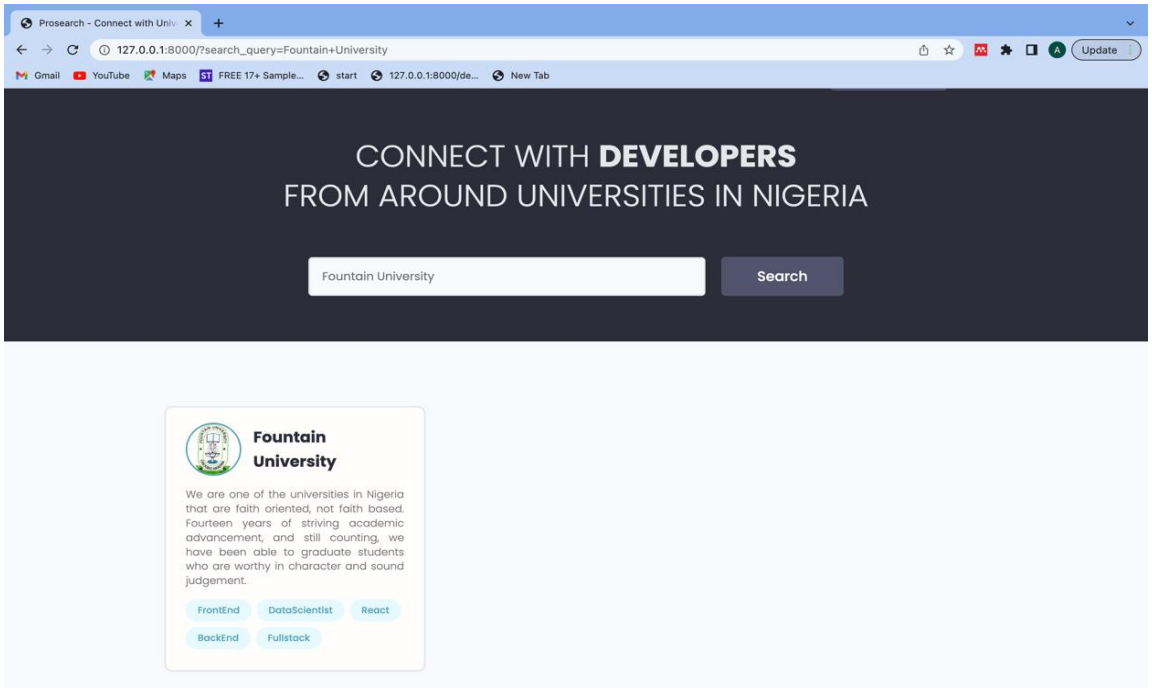

Figure 7: Search Engine result for university

Figure 7 shows the search result after a 'Fountain University' search.

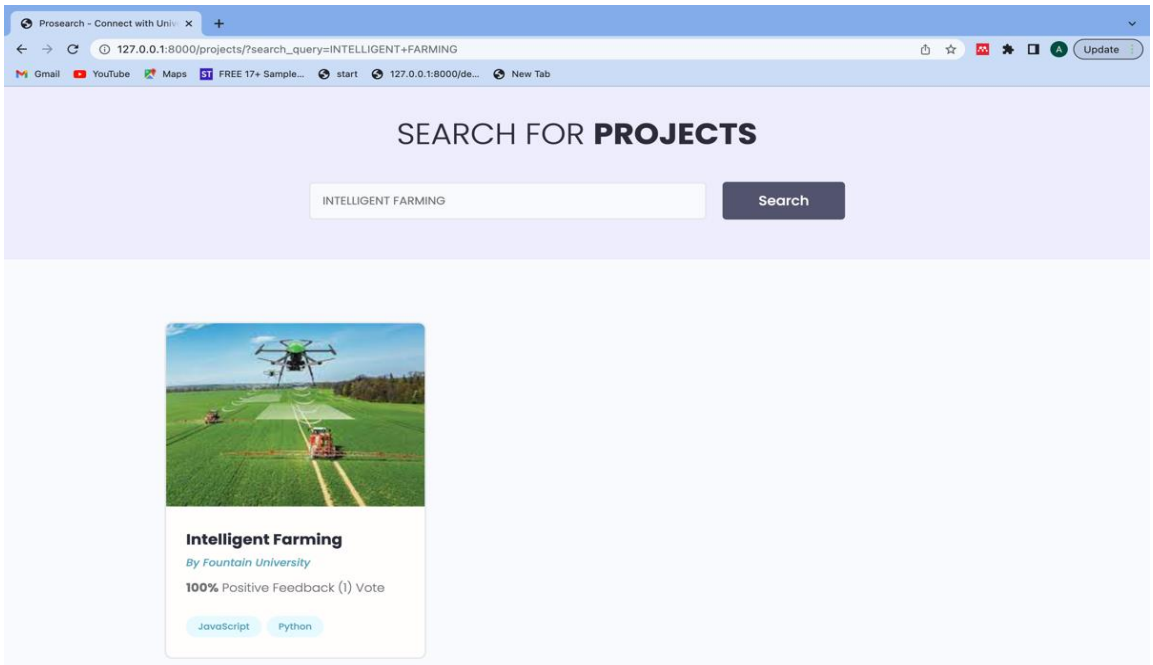

Figure 8: Search engine results for the project

4.1.2 UPLOADING PROJECTS

The interface was designed for users (researchers, enabling admins, lecturers, or Liberians of the Universities) to upload projects to the repository. The users type the project title, choose the file by clicking on choose file, and fill out the description. The details are shown in Figure 9.

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

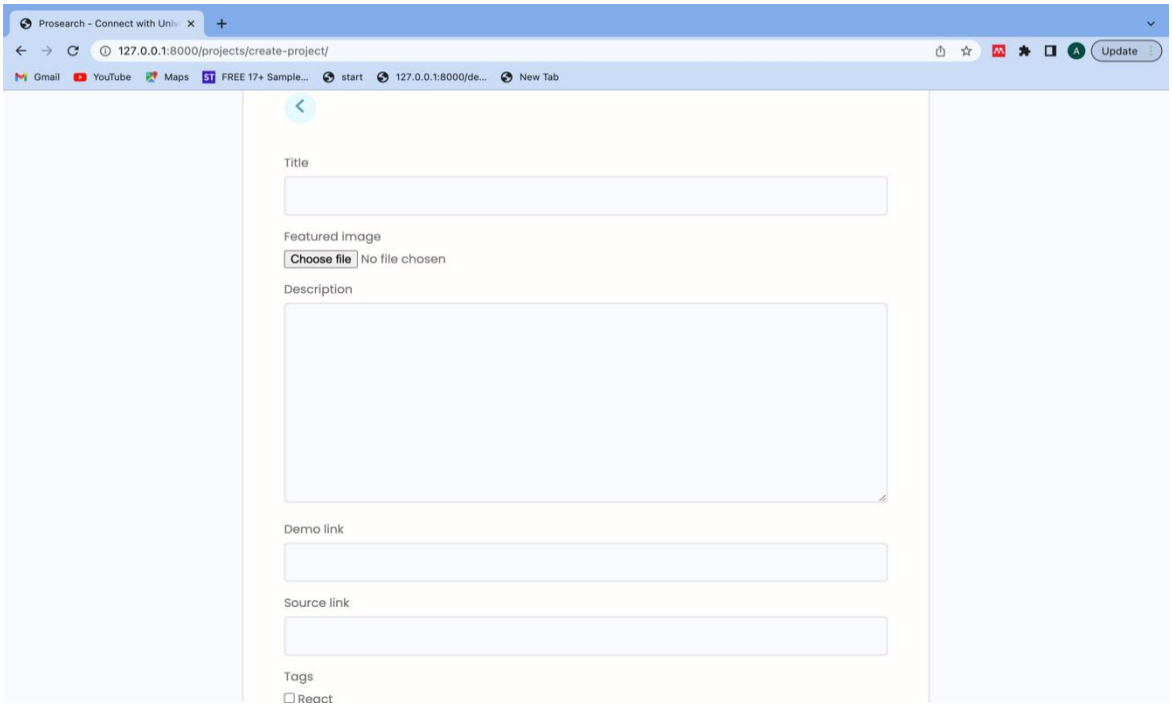

Figure 9: Add-Project page

4.1.3 VIEWING PROJECTS

The main reason for this project is to allow students to view projects and know how it has been done. Therefore, a necessary feature of this system is for them to access the documents on the website and have the option of downloading the source code if necessary. The system not only allows the students to see the project abstract and methodology directly on the website, but it enables them to download the source code for personal use. Below can be found an image of the view for showing the project abstract and methodology and downloading the source code.

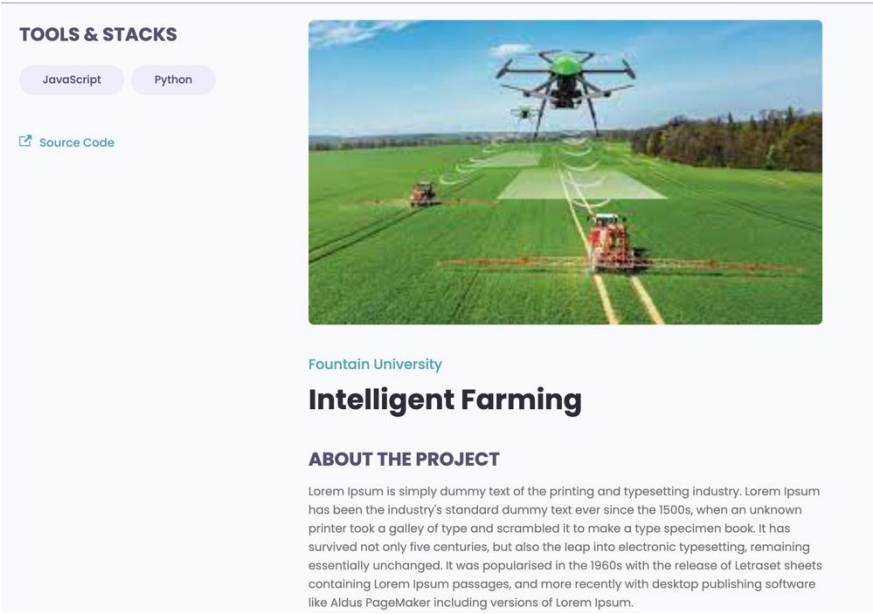

Figure 4.6: Viewing the project abstract and methodology

The picture above shows the project, which consists of the project abstract and methodology and the link to the source code in case students want to download it for personal or future use.

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

7. TESTING AND EVALUATION RESULTS

Testing is one of the most critical aspects of developing good software. It helps to validate and verify whether all functional requirements are met. Testing is essential for making sure that the software quality is assured. This was carried out by running test data through the system to ensure that it works the way it's supposed to. It was also used to sniff out any vulnerabilities in the system. All vulnerabilities are then isolated and reviewed from the rest of the system. This section will discuss all testing procedures for the components of the system, and a resulting test report will be included.

Features to be tested

The following aspects of the system were focused on during the testing of this system:

- Home page
- Register page
- Search engine
- Login page
- Logout button
- Viewing/downloading source code
- Upload page

5.1.1 TEST SUITE (FOR UNIT TESTING, INTEGRATION TESTING, AND SYSTEM TESTING)

Table 1: Test-case 001 (Admin Login

ADMIN LOGIN TEST	RESULT
Test Case Summary	Ensuring admins can login
Prerequisite to testing	- The system must be up running - Admins must be registered - Uninterrupted internet connection
Test procedure	- Open the website in the browser - Navigate to the login page - Login
Test Data	- Email - Username - Password
Expected result	The user should be able to log in successfully.
Actual Result	The user logged in successfully
Status	Passed

Table 2: Test Case 002 Logout Function:

LOGOUT FUNCTION	RESULT
Test Case Summary	Ensuring user can Logout
Prerequisite	- Website is Running - Uninterrupted internet connection - Users are already logged in
Test Procedure	- Login - Click log out
Test Data	Logout
Expected Result	-User should be able to logout
Actual Result	User was able to logout
Status	Passed

Table 3: Test Case 003: Search Bar for Project.

SEARCH Bar for Project.	RESULT

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

Test Case Summary	Ensuring the search bar works and produces the desired results
Prerequisite	-Website is Running -Uninterrupted internet connection
Test Procedure	-Type keyword(s) in search bar for project and click search
Test Data	-Login -Key word "Intelligent Farming"
Expected Result	List of related project topics in the database were listed
Result	Passed

Table 4: Test Case 004 Viewing Project

PROJECT VIEWING	**RESULT**
Test Case summary	User should be able to view project
Test procedure	- Search for a project - Click on the project - Click source code link to download them if required
Test data	Key word 'Intelligent Farming'
Expected result	User should be able to view project
Actual result	User was able to view project
Status	Passed

Table 5: Test Case 005. Adding Project

ADD PROJECT CASE	**RESULT**
Test case summary	Admins should be able to upload project
Prerequisite	- Logged in as admin -Uninterrupted Internet - Add project from profile
Test Procedure	- Log in - Navigate to "Add project" from profile - Fill form - Submit Form
Test data	New Project
Expected result	Admin should be able to add project
Actual result	Admin was able to add project successfully
Result	Passed

Table 6: Test Case 006 Home page

HOME PAGE TEST	**RESULT**
Test Case Summary	The homepage should open when the URL is searched

https://dx.doi.org/10.4314/fuoyejet.v9i3.6 engineering.fuoye.edu.ng/journal

.Prerequisite	- Uninterrupted internet - Browser is running
Test procedure	- Search for the local host
Test data Expected result	- The homepage should open after searching
Actual result	The homepage successfully opened
Status	Passed

8. CONCLUSION:

The work has explored and discussed the importance of technology and online resources within the educational system around the world. Web archival system is one of the most secure methods of preserving data for future research and the successful implementation of this project will bring relief to students because it will allow students to have access to resources from other Universities from the comfort of their computers without having to be physically available in the University.

9. REFRENCES:

Anthony Adoghe, Onasoga, Kayode, Ike Dike U., & Ajayi, Olujimi (2013). Web Archiving: Techniques, Challenges, and Solutions. *International Journal of Management & Information Technology*, 5(3), 598-603. https://rajpub.com/index.php/ijmit/article/view/760

Costa, Miguel, Gomes, Dniel, & Silva, Mario J. (2017). The evolution of web archiving. International Journal on Digital Libraries, 18(3), 191–205. https://doi.org/10.1007/s00799-016-0171-9

Purmono, Cahyono A., Purmono, **Bayu** & Oganda, **Fitra I** (2019). Design of Information System in Admission of New Students Based on Web in SMK Al Amanah. Artist Transactions on Management (ATM), 3(2), 159–167. https://doi.org/10.33050/ATM.V3I2.1034

Lupfer Nic, Kerne Andruid, Linder Rhema, Fowler Hannah, Rajanna Vijay, Carrasco Matthew, &Valdez Alyssa (2019). Multiscale Design Curation: Supporting Computer Science Students' Iterative and Reflective Creative Processes, C&C '19 June 23–26, San Diego, CA, USA 38(3) 178-182.

Martin Klein, Shankar Balakireva, Lyudmila Harihar, Herbert, Sompel V.(2019). The Memento Tracer Framework: Balancing Quality and Scalability for Web Archiving, arXiv:1909.04404v1 **[cs.DL],** 18, 7(3), 21-23

Matsushita, R. (2017). On-site AR Interface with Web-Based 3D Archiving System for Archaeological Project.

Niu, j. (2019). *An* Overview of Web archiving. Retrieved from Dlib.org: www.dlib.org/dlib/march12/niu/03niu1.html, https://doi.org/10.24297/IJMIT.V5I3.760

Hamilton, T. (2019). Agile methodology. Agile methodology: What is the agile model in software testing, 11(1), 78-82.

Hwang, H. C., Shon, J. G., & Park, J. S. (2020). Design of an enhanced web archiving system for preserving content integrity with blockchain. Electronics (Switzerland), 9(8), 1–13. https://doi.org/10.3390/ELECTRONICS9081255

Zhongke Wang, Jun Li (2023), Design of University Archives Business Data Push System Based on Big Data Mining Technology, *International Journal of Advanced Computer Science & Applications*, Vol 14, Issue 11, p42. Dio: 10.14569/ijacsa.2023.0141105