UIP Centralized Database

Kong Mun Jun

Bachelor of Computer Science(Honours) in Data Science

Supervisor: Dr Lim Siew Mooi

Abstract

In Malaysia, university information are rarely standardized and some even fragmented. Hence this system aimed to build a centralized database to convenient students, universities, researchers and industries able to conveniently discovered the information they needed. I have contributed in every milestone including data gathering, data cleaning, data sorting and User-Interface of the Database.
In the end, the base of the database have successfully built up to make further advance improvement in the future.

Problem Statement

1. Some university data are not real time information, not standardized and even fragmented
2. Students and researchers hardly find the best choice of universities for their study and researches.
3. Enterprises need to have a strong connection with the educators to ensure the enhanced learning scope and contents.

Objectives

1. To crawl various valid information from the university, government agency and industry websites into csv format.
2. To create a system that acts as a centralised database to gather pieces of universities information.
3. To provide convenient, efficient and accurate visualisation results to the third party.

Design and Methodology

Evolutionary Prototype Model are used in this project, UIP Centralized Database, as this model allow this project to be continuously build and refining till the best condition of the system are developed.

Hardware and Software requirements are quite common in this project. Basic desktop, mouse and keyboard are only required for hardware while python 3.8 and html5 will be used as Software which also easily to be installed.

Target Market for this project are listed below:

1. Universities researchers
2. Academic community
3. Centres of Excellence of Universities
4. Government agencies and Ministries
5. SMEs and Corporation in 15 industry Groups
6. Investors
7. Undergraduates and Postgraduate students

Construction and Testing

Visual Studio Code and Google Colab are used to construct the whole project to process the development of data cleaning and user-interface. Google colab mainly used for data sorting and data cleaning with python3.8 while Visual Studio Code is used to construct the database user-interface with the support of HTML5, CSS and JavaScript.

ParseHub also used to assist in Data Scraping module by scraping from any website and store the data into another tools known as Microsoft Excel with csv format.

User-Interface were created in a simplistic style by showing the features of the module obviously. Information in the module also went through every process of beautification before storing into database. White Box testing and Black Box testing also used to refine the module in the website to avoid bugs being published.

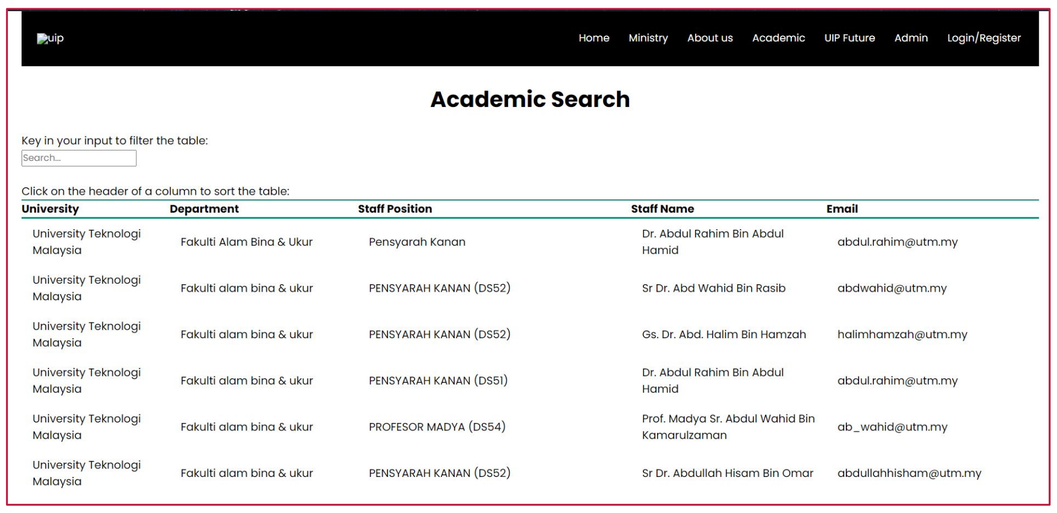

Figure 1: Academic Module Webpage

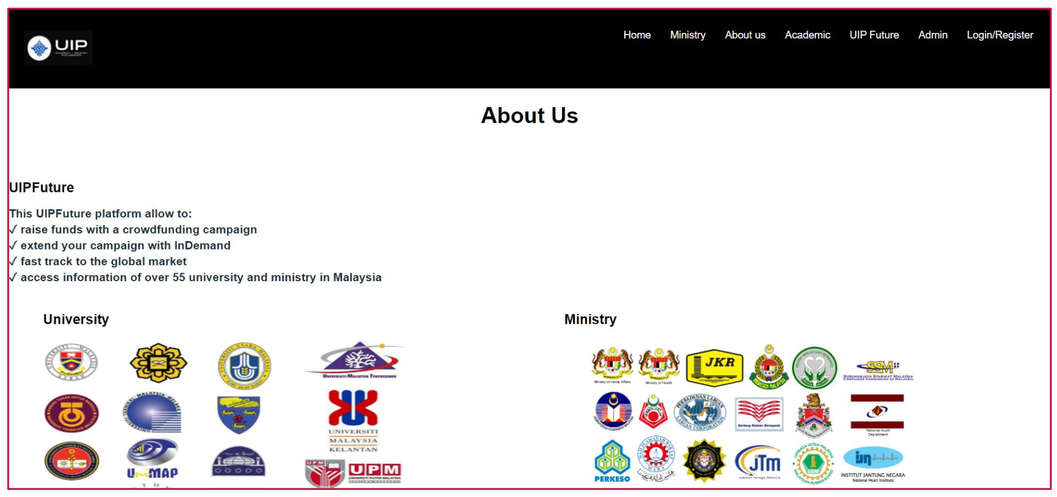

Figure 2: Main Page UIP Database

Contribution

This project is the first ever database that centralize whole education industry information into one. It can contribute from business industry to academic industry.

Conclusion

In the nutshell, initial objective in this project are achieved and base of the design in the project were built. However, more advance feature or design need to use to refine the project. Fortunately, every output required for this project are successfully shown out to the users.