

1. Introduction

This project is involved in big data analytic to help universities and industries in creating a linked connection between Universities, Industries and Investors. With the connections, the published researches and uncommercialized patents from the universities able to be seen by the investors and industries to turn these universities works become commercialised in the community. The data sets used by this projects involves the details of Universities, the governments, and et cetera.

This project includes functions such as data crawling from various websites, segmenting raw data into data clusters and insert into an online database. These functions are handling by 2 groups which involve with 7 different teammates.

2. Research Statement

With these loads of universities in Malaysia, it takes time for third party such as the teenagers, parents and lecturers to explore and do researches of the universities with every aspects such as the Courses, Subject material, Social activities, Portfolios and et cetera. However, it is found guilty that the information that we can find in websites of most universities are counted as fragmented information which are all broken pieces of data that are not close together. Some universities even irresponsible to update their university information to the latest such as the alumni of the university and facilities developments. Therefore, these fragmented data became the core problem of third party having difficulties in getting true real-time information.

This core problem, the fragmented data issues in every university had effected two main phenomena in the society. The phenomena could easily discovered are the third party had hard time to produce a better choice in targeting the most suitable university for every individual. Feedbacks of every subject from every student will always collect from the universities after every semester. However, the feedbacks are disclosed by every universities which actually the feedbacks are very vital to the third party as they able to feel the true learning environment from the feedbacks of the students. Websites and advertising will always show the best from their facilities but feedbacks will show the true feeling towards the surrounding. Thus, It is possible for students to remain anonymous and providing feedbacks in the surveys or forums to show the true experience received from the universities

The second phenomena is the connection between the industry, universities and investors is weak which happens to be categorized as part of the fragmented data. It is to be believed that every university has numerous published researches and non-commercialized patents that need more engagement and adoption by the industries. Moreover, technology is getting more contact with the methods in teachings of educators which change the norms in students' learning experience. Hence, industries need to have a strong connection with the educators to ensure the enhanced learning scope and contents to be learnt by the students in universities.

3. Research Question

- 1) What is the quantity and the quality of the sources that can be collected from every university?
- 2) What will be the main focus of the students when making a decision?
- 3) Will the platform be accurate for the academic researchers when making a decision?
- 4) How to retrieve information from institutions that insist in disclosing their information?
- 5) How fast will the system respond and finish analysing the results?

4. Research Aims

The project aims is to provide a convenient platform for academic researchers' universities researches and uncommercialised patents bring into the real life and various industries to be commercialized.

The project sub-objectives are as below:

- a. To improve the skills of web scraping websites and extract from codes instead of displays.
- b. To bring up the work and contributions of educators' researches.
- c. To simplify the workflow and figure out the effectiveness of the platform.

5. Research Objectives

The Objectives of the project is obvious as below:

- a. To crawl various information from websites into excel
- b. To create a software which act as a centralized database to store every information related to the universities.
- c. To extract all data from excel into a centralized online database
- d. To provide a platform to gather all informations of universities
- e. To provide a convenient, efficient and accurate visualisation results to the third party.

6. Research Scope

In this project, UIPFuture will be developed and during the development phase, several works and scope in this research will be focused as follow:

a) Dataset

The dataset will be used in this project are the set of data of every important information in every higher level study institutions, industries and government agencies.

b) Programming Language

The programming language that we will be using are HTML, CSS, JavaScript, Python, Php and Laravel. HTML, CSS and JavaScript will be used to assist in building user interface; Php and Laravel will be used to build the database; Python will be used to scrap data and analysis.

7. Research Significance

The finding of this research study will be reflected to the community and beneficial for the education community such as the universities, industries and the investors as this research will be seeking linked relationship and collaboration towards each of the universities. Academic researchers will also find easy and convenient to do their own finding as well as fund can be easily funded when the researches is easily found by investors in a platform.

8. Research Timeline

Mile Stone	Mile Stone Goal	Deadline
Web Scrapping	Scrap all the related information of Malaysia Universities and produce output in .csv format.	
Database development	Push all the scraped information into one central database with auto search and filter functions.	
Database UI/UX development	Development of UI/UX for the database to produce information needed and data flowchart.	
Dashboard Design and Development	Design and development of the dashboard of the RoboAdvisor AI-based system.	
Live feeds scraping	News scraping for latest development/trend	

9. Project Structure

