

INTRODUCTION

Introduction

Over the years, technology has advanced to a stage where we are able to make everything organized into a single platform with different components that have been able to solve many real-life problems. Data is one of the most important factors in the advancement of technology in the modern world. Modern technology is dependent on the data. It is without a doubt that technology has simplified our lives and has made changes in the way people used to work or think before. Without proper data collection, no system can be developed whether it is related to technology, science, or any other relevant fields. Data is extracted and refined using certain techniques and methods. Web Scraping is one such technology. Web Scraping is the process involving the extraction of data from a source using automated software and tools. It has grown more popular over the years since its utilization boosts the businesses of different organizations or individuals. In recent years, it has helped different organizations and individuals gain access to useful information that can be used for a wide range of tasks like marketing, research, analysis, etc. For businesses looking to gain a competitive edge in the industry, web scraping has become an essential tool for them.

Web scraping-related technology has advanced significantly over time. Web scraping was a labour-intensive manual procedure in the early days of the internet that entailed manually copying and pasting data from webpages into spreadsheets. The technology evolved in tandem with the growth and complexity of the internet. Various organizations and companies use this technology to analyse their data over the years and work on improving their system in the future. For universities like UTM which is a research-intensive institute, it is very important to keep track of the publications made by the staff of different faculties. One of the main faculties of UTM is the faculty of computing. Web Scraping is one of the very good approaches to keeping track of the publications made by the faculty of computing staff. With this approach, we can analyse the research output and evaluate the current state. A dashboard created with

the data collected from web scraping will be more insightful for viewing the outputs of research and publications made by the staff from the faculty of computing,

In a nutshell, a dashboard can assist faculty members with keeping track of faculty publications, identifying patterns in research output, and automating the process of updating publication data by giving them a comprehensive and up-to-date perspective of research output.

Problem Background

One of the cornerstones of UTM is the Faculty of Computing. The staff of the faculty of computing not only excels in academic teaching but also in research. University Teknologi Malaysia (UTM) is a research-intensive university. The Faculty of Computing at the Universiti Teknologi Malaysia (UTM) has many faculty members who publish their research papers, articles, and other academic works. Keeping track of all the publications from each faculty member can be a challenging task, particularly when this information is dispersed over numerous websites and databases. It is important to keep track of the publications to do an analysis of the results from the staff of faculty of computing staff. Through this approach, the faculty members can keep track of their latest research updates and can analyse them accordingly to make better progress in the future. It can be challenging to obtain an accurate picture of the faculty's research output because of discrepancies in data and inaccuracies caused by manual reliance on individual faculty members. Through this, faculty members may risk missing out on funding or partnership opportunities as well as possible avenues for research advancement without a thorough and current perspective of research output.

Project Aim

The purpose of this project is to develop a web-based dashboard system that will highlight publications, indexed publications, non-indexed publications, and citations to improve the decision-making process and can assist faculty members with

keeping track of faculty publications, identifying patterns in research output, and automating the process of updating publication data by giving them a thorough and current perspective of research output.

Project Objectives

Setting up the objectives is a key step for developing any project. The objectives of the project are given below:

- a) To gather and analyse data on the publication output of faculty members in the Faculty of Computing at UTM.
- b) To design and explain architectural models, database models, and user interface of the dashboard.
- c) To do requirement analysis and design the system requirements.
- d) To develop a Web-Based Dashboard system according to the proposed design.
- e) To test and evaluate the Dashboard system to check if it matches the user requirements gathered from the stakeholders.

Project Scope

The scope of the web scraping project for the faculty of computing staff journal publication at UTM includes the collection and analysis of publication data from various sources and the creation of a user-friendly dashboard. The project will prioritize data security and privacy while also enabling the faculty to track their own publication output, identify research strengths, and inform decision-making.

- (a) Web scraping data from various sources including the UTM website, Google Scholar, and other academic databases.
- (b) Data Duplication will be handled by applying suitable deduplication techniques and algorithms depending on the variance of the dataset. Manual review and verification will also be done to eliminate duplicate entries of data.

- (c) Creating a dashboard that allows users to view and analyse publication data based on various criteria such as publication year, journal, and citation count.
- (d) The project will use appropriate web scraping and data visualization tools based on the requirements and constraints of the project.
- (e) The project will be a web-based system that can be accessed through a desktop or a personal computer with Windows, MacOS and, a stable internet connection.

Project Importance

The main idea of this project is to provide real-time information on faculty activities using web scraping techniques. The main aim of this project is to give an idea to improve the decision-making process by providing accurate and timely information to the administration and faculty. In today's world, many organizations use the data acquisition technique to make future decisions on the improvement of the system. The type of data that will be collected and included are publications, indexed publications, non-indexed publications, and citations. This sort of data is important for analysing the effectiveness of the research published by the faculty of computing staff.

With web scraping, the analysis of the information gathered can be used to identify trends, patterns, and insights to make informed decisions. The Dashboard developed during the project can be used for a variety of purposes, including monitoring faculty performance, identifying areas for improvement, tracking research activity, and fostering collaboration among faculty. For example, a dashboard can provide information on the number of papers each faculty authored, the number of indexed and non-indexed publications, and the number of citations received. This data will help to identify highly productive faculty members and will guide the allocation of R&D resources.

Additionally, the Dashboard can be used to monitor the status of grant applications. This information can facilitate effective communication and collaboration among faculty members, leading to successful research projects. The dashboard will also provide information about other resources such as income generation, training conducted, and conferences organized. This information will help the faculty members to work on their professional development and can increase the visibility of goals set by the faculty. The administration can also allocate resources more effectively. The dashboard will highlight the networking opportunities which will help the faculty members to recognize the collaboration standards and work on them.

In summary, the project is an important attempt to boost decision-making processes, foster collaboration among faculty members, and enable effective resource management. This project uses web scraping techniques to collect real-time information from various sources that can be used to identify trends, patterns, and insights.

Report Organization

Chapter 1 provides a summary of the project's background, including the introduction, problem statement, project aim, objectives, scopes, project importance, and report structure. Chapter 2 elaborates on the literature review for this project, which will be used to develop a dashboard containing research publications using web scraping. Chapter 3 describes the methodology of the system development process as well as the hardware and software required to develop the project. All the system's requirement analyses and designs are shown in Chapter 4. The system's implementation (code) and testing are discussed in Chapter 5 to ensure quality. Finally,

chapter 6 is the conclusion, which includes a review of the conclusions as well as suggestions for further studies.

