

Universiti Teknologi MARA

**Web-Based RIG Performance Reporting System
Using Web Scraping Technique**

Siti Rasyiqah Binti Muhamad Zaki

**Thesis submitted in fulfilment of the requirements for
Bachelor of Computer Science (Hons.) Computer Science
Faculty of Computer and Mathematical Sciences**

January 2019

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks goes to my supervisor, Dr Azlan Ismail that has help me during the process and guide me to make this report perfect. Without his help and knowledge, it would be impossible to finish the project within time. Thank you to my lecturer, Dr Zainura Hj Idrus and Dr Marina Ismail for all the useful guidance in completing this project.

Special appreciation also goes to my beloved parents for their support and encouragement in the process of completing the research. There are not enough words to describe my feeling towards them.

Last but not least, I would like to give my gratitude to my dearest friends and classmates for helping me a lot in this research.

ABSTRACT

Data is a piece of information that exists in almost everywhere in the Internet and it is represented in either in the form of unstructured, semi-structured or structured data. In cases like these, when people want to copy certain data they do extraction from the web to the readable file. This system will be using scraping technique that can retrieve information from multiple sources and display it in the table form. The requirement for scraping the data from different sources can be done by using Python language and tools like BeautifulSoup or Pandas. BeautifulSoup can read and parse html and xml files by getting the certain data that needed. Meanwhile, Pandas purpose can clean and is a data analysis tools. The sources target limited to UiTM lecturer's publication which is Pure and Google Scholar. This web-based is important as it can reduce in extracting the data manually. There are many tools for collecting the data in the Internet; however, this research can help in the use of data collection method that can make the extraction easier.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	xii
CHAPTER ONE: INTRODUCTION	
1.1 Introduction	1
1.2 Background of Study	1
1.3 Problem Statement	2
1.4 Research Objectives	3
1.5 Research Scope	3
1.6 Research Significant	4
1.7 Conclusion	4

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter explains the research background of the project. This section also describes the problem statement, research objectives, research significances and research scope.

1.2 Background of Study

Data analytics often describe as collection of large volumes of data. The fast-paced development of both academia and industry for advanced technologies has made big data analytics as one of the most focused areas (Ghofrani, He, Goverde, & Liu, 2018). The application areas of big data tools, and techniques is vast (Eckroth, 2018). Data is often categorized, analysed and stored to draw conclusions and patterns.

Data displayed by most websites can only be viewed in web browsers (Mahto & Singh, 2016). The contents from the web page cannot be saved into file. The only choice left is to manually copy and paste from the web and continuously do the extraction until the work is done. It could take more time to finished collecting data in the internet. That is why the modern world right now has introduces the ways to collect data in much more simple and easier than the traditional methods.

Extracting data means to retrieve the information from data sources that mostly from unstructured data and in other types of data sources to perform data representation in tables, graph and chart. Most of the data being extracted from