

Universiti Teknologi MARA

**Hiking Gears Comparing System
(HGCS) Using Web Scraping Technique**

Azrul Nazrin Bin Hamdan

**Thesis submitted in fulfilment of the requirements for
Bachelor of Information Technology (Hons.)
Information Systems Engineering
Faculty of Computer and Mathematical Sciences**

January 2017

STUDENT'S DECLARATION

I certify that this report and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

.....
AZRUL NAZRIN BIN HAMDAN
2014715323

JANUARY 10, 2017

ABSTRACT

This project is about the development of Hiking Gears Comparing System (HGCS) for people who love to do hiking activity around Malaysia. Hiking is one of the favorite outdoor activities in Malaysia. One of the factors of people tend to carry out hiking activity is because the Malaysia landform that has many hills and mountain. So, people in Malaysia like this challenging activity to fulfill their interest and care about their healthiness. In order to carry out this activity, hikers need to use the suitable equipment or gears. Hiking gears consist of hiking backpacks, hiking footwear, hiking clothing, tents and survival kit. Hikers normally buys the hiking gears either walk in at the shop or buy through online. The problems that have been identified when the hikers want to buy the hiking gears, most of them do not know the best websites that promote the best products based on the price and the functionality. They need to open many websites to make a survey for the products that they want to buy. HGCS used the web scraping technique to extract the data from the websites that provide hiking gears. This technique has been identified in the knowledge acquisition phase. From the extracted data, HGCS will compare the products based on price, product category and functionality. This project applied the waterfall model methodology. The phases included in this project is knowledge acquisition, requirement gathering and analysis, design and implementation. Findings and analysis was done to ensure that the objectives of this project are tally with the phases in methodology. For the future work, this system may include some others outdoor activities gears, for example, rock climbing gears, cycling gears or kayaking gears. In addition, extract data from more websites that sell the hiking gears by using the web scraping technique.

TABLE OF CONTENT

CONTENTS	PAGE
SUPERVISOR’S APPROVAL	ii
STUDENT’S DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENT	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER ONE: INTRODUCTION	
1.1 Background of Study	1
1.2 Problem Statement	3
1.3 Aim	3
1.4 Research Objectives	3
1.5 Research Scope	4
1.6 Research Significance	4
1.7 Expected Result	4
1.8 Summary	5
CHAPTER TWO: LITERATURE REVIEW	
2.1 Overview of Hiking Activities	6
2.2 Hiking Gears and Equipment	7
2.2.1 Backpacks	7
2.2.2 Sleeping Bag	8
2.2.3 Hiking Footwear	8
2.2.4 Survival Kit	9
2.2.5 Tent	9
2.3 Web Page Data Extraction	10
2.3.1 Web Page Data Extraction by Web Scraping	10
2.3.2 Web Page Data Extraction by Web Crawler	11
2.3.3 Web Page Data Extraction by Web Mining	11

2.3.4	Discussion	12
2.4	Methodology	13
2.4.1	Systems Development Life Cycle (SDLC)	13
2.4.2	Agile Methodology	14
2.4.3	Spiral Methodology	15
2.4.4	Summary of System Development Methodologies	17
2.4.5	Discussion	18
2.5	Related Works	18
2.5.1	Trivago	19
2.5.2	HotelsCombined	19
2.6	Summary	20

CHAPTER THREE: METHODOLOGY

3.1	Introduction of Methodology	21
3.2	Knowledge Acquisition	24
3.3	Requirement Gathering and Analysis	24
3.4	Design	26
3.4.1	Interface Design	27
3.4.2	Database Design	27
3.5	Implementation	28
3.6	Hardware and Software Requirement	28
3.6.1	Hardware Requirement	28
3.6.2	Software Requirement	29
3.7	Summary	31

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1.	Knowledge Acquisition	32
4.1.1	Web Scraping Technique	33
4.2	Requirement Gathering and Analysis	34
4.2.1	Conducted Interview Session	35
4.2.2	Questionnaire	36
4.2.3	Documented Requirements	41
4.3	Design	45
4.3.1	Choose Suitable Design Pattern	46
4.3.2	Construct Component Diagram	46
4.3.3	Construct the Design Class Diagram	47