KNOWLEDGE DISCOVERY FOR INTERESTING PLACES FOR TOURISTS IN JOHOR BAHRU, MALAYSIA

ATAE REZAEI AGHDAM

UNIVERSITI TEKNOLOGI MALAYSIA

Replace this page with form PSZ 19:16 (Pind. 1/07), which can be obtained from SPS or your faculty.

Replace this page with the Cooperation Declaration form, which can be obtained from SPS or your faculty.

KNOWLEDGE DISCOVERY FOR INTERESTING PLACES FOR TOURISTS IN JOHOR BAHRU, MALAYSIA

ATAE REZAEI AGHDAM

A dissertation submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Information Technology-Management)

Faculty of Computing
Universiti Teknologi Malaysia

JUNE 2013

I declare that this dissertation entitled *"Knowledge Discovery For Interesting Places For Tourists In Johor Bahru, Malaysia"* is the result of my own research except as cited in the references. The dissertation has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	
Name	:	Atae Rezaei Aghdam
Date	:	July 31, 2013

This dissertation is dedicated to my beloved mother, father and sister who dedicated their life for my success and never failed to give me every support.

ACKNOWLEDGEMENT

Praises to God for giving me the patient, strength and determination to go through and complete my study. I would like to express my deep and sincere gratitude to my supervisor, Dr Alex Tze Hiang Sim. His wide knowledge and his logical way of thinking has been of great value for me. I would like to express my appreciation to my family and my friend who helped me during my study.

ABSTRACT

Nowadays, Tourists are presented with a lot of online recommendation options before traveling. They often get confused in choosing specific places to travel and this is a time consuming process among all tourists across the globe. In this project, we crawled tourist profiles and interesting places in Johor Bahru from www.tripadvisor.com to discover clusters of customers with a different profiles, customers behavior, important feedback by tourists and useful knowledge in order to recommend appropriate places to tourists. This research includes two steps; in the first step, we clustered and applied ARM technique to uncover important knowledge about tourists and interesting places by Weka machine learning software. In clustering part we applied EM and K-Means algorithm and in association rules mining we used Apriori algorithm to find the rules between items in dataset. In the second step, we coded tourist's comments, which are about interesting places in Johor bahru through Nvivo software. Results showed that, tourists could be clustered according to their preferences for instance, local people are not satisfied with the price of food in Legoland moreover, they prefer to travel with spouse and family with young children but foreigners like to travel with friends or business colleagues. Also, Legoland is one of the fix options for all male tourists aged between 25 to 34. Furthermore, Nvivo outputs shows that, Legoland has some affirmative and negative points. Local tourists believed that, assets of Legoland outweigh liabilities but some foreigners such as; Chinese and New Zealanders considered negative points like foods price and long queues. We believe that our two steps of analysis are powerful and results can be useful for tourism industry regarding to attract great bulk of tourists.

ABSTRAK

Cara menggunakan teknologi maklumat untuk mengurus dan mendapatkan pengetahuan dari data kini adalah satu daripada kajian yang ketara. Kini, industri pelancongan tumbuh secara mendadak. Dengan meningkatkan taraf hidup di dunia dan kemajuan teknologi dalam pengangkutan, aktiviti melancong menjadi bahagian penting dalam kehidupan. Baru-baru ini satu istilah baru teknologi wujud dalam industri IT yang dinamakan perlombongan data. Perlombongan data adalah proses menganalisis data dari pelbagai aspek dan ekstrak pengetahuan yang berguna daripada data mentah. Dalam projek ini, kita akan cuba untuk menganalisis data pelancong dan tempat-tempat menarik di Johor Bahru, Malaysia dan cuba untuk mendapatkan pengetahuan yang tersembunyi seterusnya mengesyorkan tempat-tempat yang sesuai untuk pelancong berdasarkan pengetahuan yang dikeluarkan. Pertama kita akan mengumpul data dari web oleh web crawler tertentu dan yang kedua kita akan menganalisis data dan akhirnya cuba untuk mengesyorkan tempat-tempat menarik yang sesuai untuk pelancong.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

KD	–	Knowledge Discovery
DM	–	Data Mining
KDP	–	Knowledge Discovery Process
KDD	–	Knowledge Discovery Database
ARM	–	Association Rules Mining
EM	–	Expectation Maximization
	–	

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Obviously, we live in technology era and today many people use technological equipment in each field such as e learning, e-commerce, e-government and etc. Nowadays, people almost can use Internet everywhere to do their tasks. Recently a new term of technology exists in IT industry which name is data mining. Data mining is the process of analyzing data from different aspects and extracts the knowledge from raw data. This technology provide opportunities for users to analysis the data from different kind of dimension and extract the hidden pattern from huge amount of data. Technically, data mining is the processing of finding relationship or patterns among dozens of fields in a database.

This study focuses on knowledge discovery from tourism industry sector. In fact, we will try to analyze the data of tourists and find the hidden pattern in data to suggest interesting places to tourists according to their preferences or any different aspects. In the other word, this study will link the data mining to tourism industry for increasing the level of tourist satisfaction, and guide them to choose appropriate option to travel. In essence, this research concentrates on analyzing the data of interesting places in Johor Bahru in Malaysia and recommend this places to tourists regarding to their information and preferences. In addition our target data will extract from the web.

As a result, users will enable to find that which place they are looking for, in a shoeret time.

1.2 Problem Background

The number of tourists increased dramatically during recent years in Malaysia. From 5.5 million arrivals in 1998, tourist arrivals in Malaysia have more than quadrupled to 22 million in 2008 and 25.3 million in 2012. So by growing the number of tourists the demands of tourists increased as well. Usually tourists face with numerous choices and it is difficult for them to choose the proper location to go. Consequently, they need to spend more time to choose their pleasant places. This is time consuming for tourists. Hence, if we can provide opportunity to guide tourists for matching best items to select among their choosing process, it will be helpful for them.

Another problem is that, there is a vast amount of unstructured data on the web and we will come across with numerous data in future because every year our data volume increases. In addition, we need to analyze the data and extract knowledge from it. Data mining techniques will perform this analyzing process. Therefore; our significant question is how we can discover this hidden pattern. There are two ordinary data mining approaches, which are, clustering and classification. In this project we need to investigate these two techniques and choose the suitable one to apply in our study.

Clustering is one of the techniques that use in data mining and the function of this technique is that split data elements into groups of same objects. Clustering techniques able to make strong association between members into groups. There are four types of clustering algorithms such as: exclusive, overlapping, hierarchical and probabilistic. Based on our needs, we can choose these techniques. In addition, clustering has discovery tools to uncover patterns in data that are not obvious before

this. In a word from all these point discussed above, clustering focuses on find some elements into groups based on similarities. . In addition, we will try to use some other data mining algorithms to discover useful knowledge from our database.

Our main challenges in this project are divided into two sections. First, collect the information from the web and save it into database and second one is analyzing the collected data and discover the hidden pattern regarding to tourists information.

1.3 Problem Statement

The most challenging part in this project is the collecting and analyzing the data. Our database consists of quantitative and qualitative data and we need to analyze both of them to extract precise pattern to recommend to tourist. Some tourists face with some problems during the processing of choosing best places to travel for instance, they get confuse in choosing the place they want to go. There are some websites such as: agoda.com or tripadvisor.com that suggest places to tourists and they uses some recommendation techniques for recommend to tourist but their recommendation items sometimes not precious and incompatible with their preferences. In essence, we will try to analyze tourist information from different perspective and extract the hidden knowledge from this data and find out the tourists behavior and interests to recommend best places for them according to extracted knowledge.

1.4 Project Objectives

Objectives of this study consists of:

1 To collect the information about interesting places in Johor Bahru, Malaysia and information about tourists from the web by acquiring a task specific web crawler.

2 To analyze qualitative collected data by using Nvivo software.

3 To analyze and discover hidden knowledge through data mining techniques.

4 To recommend appropriate places to tourists based on discovered knowledge.

1.5 Project Scope

This study will focus on interesting places in Johor Bahru in Malaysia and the database will include locations information and tourist information such as: age, country, gender, goal of travel, title, comment and so on. In this study we will use crawler for gathering the data from the tripadvisor.com, N-vivo software for validating the discovered knowledge and Weka data mining software for analyzing the data.

1.6 Project Importance

In recent years, the number of tourists is increased and the travel and tour agencies will face with a numerous demands of tourists from all across the globe hence, enhancement of tourists satisfaction, is the important issue. Basically, in this study we well try to provide more facilities to tourists and give them a suitable choice to make better decision according to their preferences and interests.

Another important issue of this project is gathering the data from the web and pre-processing the data and analyzes it to uncover the hidden pattern in this data. This project is not design just for Johor Bahru or Malaysia actually this study can be used for analyzing the information of every place of all across the globe that mentioned in tripadvisor.com. Furthermore, it will be useful for tourist association in Johor Bahru in Malaysia and also it would be helpful for researchers who are study about tourism industry and data mining.

1.7 Summary

This chapter discussed about role of data mining and recommendation systems in tourism industry. The main aim of this project is knowledge discovery from the data that stored in the tripadvisor.com. We will try to use data mining techniques to analyze our data and extract the knowledge from it to guide tourists for selecting the best location to go. It will improve the tourists decision-making process.