

Universiti Teknologi MARA

**Restaurant Review System with
Data Visualization and Text Search
using TF-IDF Algorithm**

Syazwina Binti Abd Hamid

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

July 2020

SUPERVISOR APPROVAL

FoodBytes: RESTAURANT REVIEW SYSTEM WITH DATA VISUALIZATION AND TEXT SEARCH USING TF-IDF ALGORITHM

By

**SYAZWINA ABD HAMID
2017412182**

This thesis was prepared under the supervision of the project supervisor, Nurulhuda Zainuddin. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for degree of Bachelor of Computer Science (Hons.).

Approved by

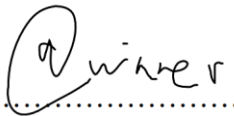


.....
Mdm. Nurulhuda Zainuddin
Project Supervisor

JULY 10, 2020

STUDENT DECLARATION

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

A handwritten signature in black ink, appearing to read 'Syazwina', written over a horizontal dotted line.

SYAZWINA ABD HAMID
2017412182

MAY 29, 2020

ABSTRACT

A single bad review can poison the mind of reader. That shows how online reviews are playing an important asset for decision-making. Today, restaurant review system allow feedback of customers based on personal opinion and experience to be shared with others. However, high availability of reviews made it impractical for user to read them all where users found it frustrating to go thru each one for enough information. Besides that, the text in reviews convey better information than the rating alone thus encourages deeper study on the review content. Furthermore, current approach is lacking and needs new improvement to yield better results when user search for restaurant. With that being said, feature for easy access of review through searching and to provide analysis on these data would be highly beneficial to many, from investor, business owner and Internet user. Hence, this project focuses on the integration of information retrieval with data visualization in order to produce valuable insights. Term Frequency-Inverse Document Frequency (TF-IDF) algorithm is performed for text search, where it satisfies multiple keywords at a time and the search result for review is based on relevancy score. The implementation of TF-IDF will have to go through preprocessing such as tokenization, convert to lowercase, removal of special characters and stop words and lastly, stemming. Preprocessing aims to clean the review text before the score is calculated which boost the performance of information retrieval. Meanwhile, for visualization, D3.js library is used with aim to promote interactivity and enhance user experience when using the system. Visualizations used in the system are circular bar plot, word cloud and bar chart where each visualization conveys different purpose. For dataset, the data was scrapped entirely from TripAdvisor website and it only considers review data and restaurant data. Various features are deliver from this system such as login, logout, sign up as member, post new review, view submitted review or even delete reviews, search of restaurant, view restaurants, read reviews of restaurant and lastly, view overall analysis for insights where top 10 restaurants are ranked based on its average rating and total number of review. Outcome gained from functionality and usability testing indicate that all functions in the system work perfectly and received promising feedback from users. The future work of this project is to improve data collection to build better dataset, implement sentiment analysis and provide recommendation for restaurants.

TABLE OF CONTENTS

CONTENTS	PAGES
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS	xiv

CHAPTER ONE: INTRODUCTION

1.1	Background of Study	1
1.2	Problem Statement	2
1.3	Project Objectives	4
1.4	Project Scopes	4
1.5	Significant of Project	5

CHAPTER TWO: LITERATURE REVIEW

2.1	Online Restaurant Review	6
2.1.1	Review Data	7
2.2	Information Retrieval	8
2.3	Keyword Extraction Algorithm	10
2.3.1	Text Rank	11
2.3.2	Rapid Automatic Keyword Extraction (RAKE)	12
2.3.3	Term Frequency-Inverse Document Frequency (TF-IDF)	13
2.3.4	Previous Research	15
2.4	Data Visualization	17
2.4.1	Common Graph	18