

Universiti Teknologi MARA

**Extracting Triples from Web
Document**

Nur Afiqah Binti Mazlan

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

January 2014

ACKNOWLEDGEMENT

Bismillahirrahmanirrahim

In the name of Allah, The most Gracious and the most Merciful

Firstly, I would like to express my gratitude to Allah with His graciousness for giving me a time to complete this project paper on time. The completion of this report is through relentless cooperation of several persons in giving me guidance.

My greatest appreciation goes to my project supervisor PM Dr. Nurazzah Abd. Rahman. She has given me such a good idea that has opened my minds to what is currently needed in the project with her strong knowledge in Information Retrieval. Without her endless advice and support, I would have not completed this work.

Secondly, my heart appreciation also goes to Prof Dr. Zainab Abu Bakar, Puan Khairul Nurmazianna Ismail and all LRGS group's members from UiTM, UKM, UPM, UIA and UMS for their invaluable helps, assistance and advice.

Special appreciation also goes to my beloved parents and my family for their unconditional love and support, without whom it would have been impossible to triumph till the end and accomplish the entire project.

Last but not least, I would like to give my gratitude to my dearest classmate who provided useful advice and resources when I met problem in doing this dissertation.

ABSTRACT

Nowadays Semantic Web technologies become a most popularity in the world. According to Berners-Lee, he said that a web of data that can be processed directly and indirectly by machines. In the world, all human need the information and they can search information through internet. But when the process of searching for specific information become more complex, it is very hard to minimize the number of unrelated documents for the output specific searching. Therefore, the meaning of the sentences is ignored and no extraction or annotation tasks are involved. This research will be use the technique of Semantic Web to convert web document into triples. Semantic Web uses the triples data or RDF data to reduce the unrelated document only exactly document will be post in internet. Next Semantic Web also can be used to evaluate the effectiveness from of match triples from given a query. This research used collection a web document of durian for data collection to show that the effectiveness triples of match from given a query. This research can show the performance of the document when the data of durian convert to triples. In semantic web, the user can search the durian data more efficiency and effectiveness because, the Semantic Web is more flexible data model and make the data independent of its internal representation.

Keyword: Semantic Web, Resource Description Framework (RDF), SPARQL, Allegrograph, Triples Store.

TABLE OF CONTENTS

CONTENTS	PAGE
SUPERVISOR'S APPROVAL	i
DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER ONE: INTRODUCTION	
1.1 Research Background	1
1.2 Problem Statement	3
1.3 Research Objectives	4
1.4 Project Scope	4
1.5 Research Significance	5
1.6 Research Element	5
1.7 Research Question	5
1.8 Stakeholder	6
1.9 Possible Solution	6
1.10 Expected Outcome	6
1.11 Summary	6
1.12 Structure	7

CHAPTER ONE

INTRODUCTION

This introduction gives a briefly description about research background that explain problem background and rationality for the research study. It is also explained about the problem statement, research objective, scope, significance, research element, stakeholder, and possible solution.

1.1 Research Background

As recently as the 1990s, studies showed that most people realized that the most important of archiving and finding information from other people rather than from information retrieval system. With the existence of the computer, it become more possible ways to store large amount of information and finding useful information from such collections become necessity. Moreover, the field has matured considerably and several IR systems are used on everyday basis by a wide variety of users. According to Baeza-Yates and Ribeiro-Neto(1999), information retrieval(IR) has changed a lot for recent year expansions of World Wide Web and the advent of modern and inexpensive graphical user interface and mass storage. A lot of research has been spent a lot of time on developing retrieval system for use through the web.

An information retrieval (IR) is develop to analysis process and store sources of information and retrieves those that match a particular user's requirements. Since information retrieval and database system each of this handle different types of data, some database system has a problem which is not encountered in database systems. Information retrieval has found many applications such as on-line library catalogue system, on-line document management systems (Yates, 1999). Furthermore, the field of information retrieval has come a long way in the last forty years and has enables easier and faster information discovery.