University Technology MARA

Extraction of Triple and Evaluation of Exact Matching using Keywords on Durian Html Document

Nur Fathihah Binti Ismail

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

July 2014

ACKNOWLEDGEMENT

Bismillahirahmanirahim and Alhamdulillah, praise to Allah of His full blessings, I Nur Fathihah binti Ismail humbly believe I have delivered this project within the duration given with the product I had expected as I are, undoubtedly learning to become better constantly and continuously.

Therefore, with this golden opportunity, I would firstly love to give my outmost appreciation to my dear supervisor; Prof. Dr. Zainab binti Abu Bakar who had been the only person to allow us to engage with the project. I could not learn as much as we could if it would not because of her. Not to forget, she had also given us moral support and endless encouragement throughout my project term. May Allah returns her our favor with luxury and prosperity.

To the person I could never forget from the sweet and enjoy moments, I would like to dedicate our special thanks to my project coordinator, beloved Dr Siti Salwa binti Salleh for her guidance to the course with many, countless approaches of lifting our spirit up. Me somehow feed sad as the project going towards an end. She taught us how to write a good quality report using entertaining ways.

Besides that, the important is my team mate Ninie Sumarni binti Abdullah and Nabila Huda binti Mazlan, thank you for helping me to finish our project. Also not forget to my friends who always cheer us and hold our hands to the end of this journey. I hope to see each other in future so that I could recall what I had been through to complete a final year project.

As to my family, I, truly never could have come this far without them. I owe them with my lives because they had contributed so much that I could not be able to finish it. Their support is my essence to move forward not just in this project, but in my lives. Thank you so much.

ABSTRACT

Search engine is used to retrieve information from the World Wide Web. However there is a deluge amount of documents retrieved by a given query from the user. Users have difficulty to select relevant documents. With semantic search engine, user can acquire relevant document that is sufficient enough that answer user query. Research has been shown that semantic search engine performs better than available search engine. However to build a semantic search engine requires representing deluge information in triples. In this project, Semantic Search Engine for Durian is built. Firstly, documents from html documents related to Durian and query are collected. Triples from selected documents are extracted and saved. Secondly, matching using query words procedures is built. Finally the matching procedure is evaluated using the search engine known as King. The result shows that the matched documents are the relevant to the queries collected.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.0 Background

Internet connection makes the people in worlds connect each other as soon as possible. The power of fingertips can get a lot of knowledge by using search engine. The seekers can express their search the query textually (Adelfio et al., 2011). According to Goodwin, 2011 said in reality now, have growing drastically the information on public web sites. The abundance of content of World Wide Web is useful, even of millions. Even though, the browse are so simply but a lot of information useful that searcher can get (Arasu et al., 2011). According to Lawrence and Giles (1999), less than two years the size of the Web have doubled and to continue for the next two years the projected will growth rate. The search responses after the query or question typed in the search box or text field. In 2004, Giunchiglia et al. stated that operator takes graph for instance structures and produces mappings between the nodes of the graphs that correspond semantically to each node that is match. Actually, a match is important operator has a lot of famous application domains, for example shema/ontology integration, data warehouse and XML message mapping. The key intuition in semantic matching is to exploit the model-theoretic information, where is organize in the node and the structure of graphs (Giunchiglia and Shvaiko, 2003).

Despite the increase of information, the number of search engine has also increases to the respect of time (Lawrance and Gilles, 1999). Therefore , in order to find the most suitable and precise information over billions of information that match with user query, search engine need to have an advance mechanism in order to generate high-quality matching, ranking, personalization and be able to give the most relevant throughput. These features