University Technology MARA

Investigation of Triple and Evaluation of Re-Ranking Technique in html Document

Ninie Sumarni Binti Abdullah

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, Ninie Sumarni binti Abdullah humbly believe I have delivered this project within the duration given with the product we had expected as we are, undoubtedly learning to become better constantly and continuously.

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

To the person I could never forget from the sweet and enjoy moments, I would like to dedicate my 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. I somehow feel sad as the project going towards an end. She taught us how to write a good quality report using entertaining ways.

Besides that, to my friends who always cheer me and hold my hands to the end of this journey, thank you for helping me. 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 life 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 life. 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 are having a difficulty in selecting and acquiring the valuable information. With existence of semantic search engine, users can simply 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, Triple from html documents are studied. Next, matched html documents are rank according to re-ranking technique. Result show that there is an improvement of the listed html documents in term of relevancy.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.0 Background

The rapid growth in the size of the web user is because people highly prefer to use web service as a source of information provider. This is because the available information in the web is faster and easier to be accessed which is just by type in the certain word or question you desired. In 2001, Liddy stated that web search engines are a well-known component used for an information retrieval (IR) system. Similar to many other IRs, web search engine provides a simple and efficient facility and management that allows us to get information based on particular term (Zhang, Long and Suel, 2008). Web search engine also provides an information or recommendation to user which might be searching for information or trying to make purchase (Feng, Bhargava and Pennock,n.d). As the time passed by, web search engine has become the most popular tools for locating information as the performance of the service for the user is well organized to make the search result to be more useful to the user (Zhang, Long and Suel, 2008). As enormous amount of information are contain in the web and yet they keep expanding, the queries are needed to specify the related information as access tool for user. The search responses after the query or question typed in the search box or text field.

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 provide the best service to the user (Zhang, Long and Suel, 2008). This situation can be improved by having a semantic search engine, which can find document precisely and related to the user query (Youssif, Ghalwash and Amer, 2011). The most popular search engines used by people all over the world are Google and Yahoo yet these search engines do not provide the same mechanism as the semantic search