Universiti Teknologi MARA

The Effectiveness of Thesaurus Query Approach in Document Retrieval

Noor Sharoni Binti Arifin

Thesis submitted in fulfillment of the requirements for
Bachelor of Science (Hons) Information Technology
Faculty of Information Technology And Quantitative Sciences

April 2005

DECLARATION

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

APRIL 1, 2005

NOOR SHARONI BINTI ARIFIN
2003664396

ABSTRACT

This project incorporates the thesaurus approach in documents retrieval. In information retrieval, the common information retrieval model is by using traditional simple query as the query task. An important issue is that most query language use the context and the structure of the text to find relevant documents. In that sense, the system may fail to find relevant answer in order to get the appropriate document. Beside that, limitation of traditional query can result in losing a lot of the relevant document. But, using thesaurus techniques in searching, it will find the words that contain the same meaning with user query. Using thesaurus can help user to get the result in shorter time. It is because user can use this approach to find other relevant document. Users do not necessary to think the other word and then query in query form in order to get the result. This project is constructed to develop a prototype for the document retrieval which considering the thesaurus query and retrieval approach and evaluate the effectiveness. The query language that has been used in this project is Visual Basic. NET and for the database is Microsoft SQL Server 2000. An experiment is conducted by using direct query as the benchmark, to evaluate the effectiveness of thesaurus query. 50 of FTMSK official letters is taken as the test document and 10 queries are been used to the retrieval system. From this project, the result of the experiment shows that thesaurus query is effective but not efficient compared to direct query.

TABLE OF CONTENTS