UNIVERSITI TEKNOLOGI MARA

PERFORMANCE EVALUATION OF INVERTED FILES, B-TREE AND B+ TREE INDEXING ALGORITHM ON MALAY TEXT

SUHANAH BINTI ROSNAN

Dissertation submitted in fulfillment of the requirements
for the degree of
Master of Computer Science

**Faculty of Computer and Mathematical
Science**

January 2016

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This topic has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

In the event that my dissertation be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and agree be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

Name of Student   :   Suhanah binti Rosnan

Student I.D. No.   :   2014204182

Programme   :   Master of Computer Science

Faculty   :   Faculty of Computer and Mathematical Sciences

Dissertation Tittle   :   Performance Evaluation of Inverted Files, B-tree and

                                B+tree Indexing Algorithm on Malay Text

Signature of Student   :   ....

Date   :   January 2016

ABSTRACT

These worlds face a challenge of big data. Indexing is a part of Information Retrieval that helps to overcome the results yield on certain information. Like nowadays, people tend to have everything is seconds and faster. Thus, a study on indexing techniques that tend to have a better performance in retrieving documents in terms of time had been done. The indexing techniques used in this project by comparing their performance are Inverted Files Indexing, B-Tree Indexing and B+ Tree Indexing. This research uses 500 Malay text documents as the test collection. The indexing technique chosen has a slight difference in its data structure. Thus the ranking process uses the same method. The performance of the indexing technique was measured and compared in its retrieval time taken. The behaviors of each indexing technique are different and the searching methods are different too. There are five experiments done in this paper that helps to determine which of the techniques are said have a good performance. The experiments done are on the time taken to index the terms or keywords to its data structure which are either linked list as in Inverted Files or B-Tree as implemented in B-Tree and B+ Tree indexing technique. In addition, experiments on the time taken to retrieve the documents were done by having a variable as the condition. All the results and discussion are explained and discuss deeper in this paper.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS