Universiti Teknologi Mara

Enhancement of Rules-Application-Order (RAO) Stemming Algorithm Based On the First Character of Malay Word

Edatul Muliana binti Ghazalli

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

November 2005

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries, which have been duly acknowledge.

NOVEMBER 21, 2005

EDATUL MULIANA BINTI GHAZALLI
2003657552

ABSTRACT

Stemming is important thing to improve retrieval effectiveness. Stemming is used to reduce the size of indexing file for relevancy of document retrieval. Stemming is technique to truncate the word into the root word that will reduce vocabulary size and improve recall. The Malay affixes consist of four different types such as prefix, prefix-suffix, suffix and infix. An effective and powerful of Malay stemmer is it just not to move the suffixes rules only but it must remove all four types of affixes. Without removing all the affixes, the stem cannot be effectively used to index of Malay documents. So in order to get the best order of morphological rule for effective and powerful stemmer the researcher has to find out the best order of morphological rule to stem Malay words based on first character for each alphabet. This project involves the use of two combinations simultaneously. The words that could not stem correctly by the first combination of best order which is primary will shift to alternative combination of best order of morphological rule. The results of experiment B, which is enhance project is better than experiment A, which is Rules-Application-Order (RAO) by Fatimah (1995) because that algorithm has successfully stemmed all word begin with alphabet "A" until "Z" that extracted from Quran documents..

CONTENTS