UniversitiTeknologi MARA

Automated Agriculture Terms Dictionary Using String Parsing

NUR SYAHIDA BINTI AHMAD ZUBAIDILLAH

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

JULY 2014

ACKNOWLEDGEMENT

Alhamdulillah, praise and thank to Allah because of His Almighty and His utmost blessing, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor, Dr. NurSuriati Binti Jamil because guide me from starting until the end. Special thanks to my FYP lecturer Dr. Siti Salwa Binti Salleh for teaching this subject.

A special appreciation also goes to my beloved parents Ahmad Zubaidillah Bin Othman and . Thanks for giving support.

Last but not least, I would like to give my gratitude to my dearest friend in CS230. Thank you for your helps.

ABSTRACT

This paper proposed the task of generating terms dictionary in the Agriculture domain. Its purpose is to generate only important terms using in Agriculture fields from Hypertext Markup Language (.html) file and Word Document (.doc) file as the input. The goal is to create a prototype which can automatically read the string from this type of file and then extracts their terms. The limitation of the previous manual work for generating terms are known to be labor-intensive, limitation of applicability and time consuming. Terms extracted are the smallest fragments of texts in documents, rather than the entire documents that contain the texts. After analyzing the characteristics of terms in Agriculture Dictionary, we propose a string parsing method to deal with the task. We also design and develop a parser on identified parsing method. To test the functionality of the prototype, compared between the actual terms from document and the generated terms is determined. The final prototype achieved 100% accuracy between actual terms and generated terms from html document and achieves 86.27% accuracy from word document file. However, to extract terms is difficult because of the type of files that cannot easily be read by the default library package. For future work, Automated Agriculture Terms Dictionary will be fixed to get better result.

Table of Contents

CONTENTS **PAGE**