



สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
School of Computer Engineering , Faculty of Engineering , King Mongkut's Institute of Technology Ladkrabang

[Main](#) [Curriculum](#) [Subject](#) [Project](#) [Faculty](#) [Webboard](#) [Hall of Fame](#) [Alumni](#) [Admission](#) [About Us](#)

วิชาโครงการ > รายละเอียดโครงการ

หัวข้อ (ภาษาไทย) :	การวิเคราะห์การตัดคำภาษาไทย
หัวข้อ (ภาษาอังกฤษ) :	THAI WORD SEGMENTATION ANALYSIS
อาจารย์ที่ปรึกษา :	ดร. วัชร วัชรวิริยะ
รายนามนักศึกษา :	43010375 วราภรณ์ สกวรัตน์กุล 43010488 สุพัตรา ตรีเกษตรศาสตร์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบประสิทธิภาพของหลักการตัดคำภาษาไทยที่มีอยู่ โดยพิจารณาประสิทธิภาพของโปรแกรมในหลาย ๆ ด้านร่วมกัน เช่น ความเร็ว ความถูกต้อง ทรัพยากรที่ใช้เพื่อให้ผู้ใช้สามารถนำรูปแบบการตัดคำที่เหมาะสมมาประยุกต์ใช้กับงานที่ต้องการ ในการทำงานวิจัยนี้ ผู้วิจัยได้ศึกษาการตัดคำและคลังข้อความด้วยวิธีการต่าง ๆ แล้วนำมาสร้างเป็นโปรแกรม และทำการทดสอบประสิทธิภาพในสภาวะที่มีความใกล้เคียงกันมากที่สุด หลังจากนั้นจึงทำการสรุปวิเคราะห์ประสิทธิภาพในด้านต่าง ๆ ทั้งนี้ผู้วิจัยได้ทำการเขียนโปรแกรมที่ใช้ในการทดลองประสิทธิภาพโดย MATLAB version 6.5 และข้อมูลและข้อมูลที่ได้ทำการทดลองได้แก่ คลังข้อความ และพจนานุกรม ได้นำมาจากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) จากผลการวัดประสิทธิภาพพบว่า การตัดคำโดยใช้พจนานุกรมมีความถูกต้องมากที่สุด การตัดคำโดยกฎมีความถูกต้อง, การใช้ทรัพยากรและเวลาในการตัดคำน้อยที่สุด การตัดคำโดยใช้คลังข้อความมีการใช้ทรัพยากรมากที่สุดและใช้เวลาในการตัดคำนานที่สุด

Abstract

The purpose of this thesis is to analyze and compare the performance of the Thai word segmentation programs in three ways ,speed ,accuracy and resource. Several method of Thai word segmentation was studied and implemented in the same environment using based on one of the analyze a corpus name Orchid to make sure that the comparison are merit. The performance testing programs are implemented using MATLAB version 6.5 .Data and testing data ,analyzed sentence from various source, 25,000 sentences ,and dictionary of 30,000 are provided from NECTEC. Experimental result show that the dictionary method (longest pattern matching), gives the most accurate rate output. While the rule pattern requires minimum memory using but a poor accurate rate. The Corpus segmentation occupy the large memory and also require extensively computing power.

Department of Computer Engineering , Faculty of Engineering , King Mongkut's Institute of Technology Ladkrabang
Ladkrabang Bangkok 10520 Thailand. Tel. +66 2 329 8341 - 2 Fax.+66 2 329 8343
Copyright © iCE. All Rights Reserved. Developed by iCE Team.