

## บทคัดย่อ

วิทยานิพนธ์นี้นำเสนอวิธีการปรับปรุงประสิทธิภาพของการตัดคำภาษาไทยโดยใช้เทคนิคการจดจำนิพจน์ระบุนามใช้วิธี Conditional Random Fields (CRFs) ในการเรียนรู้เพื่อสร้างโมเดล การจดจำนิพจน์ระบุนาม ใช้คุณลักษณะคำขึ้นต้นและคำลงท้ายของนิพจน์ระบุนามที่สำคัญได้แก่ นาย,นาง,นางสาว,คณะ,มหาวิทยาลัย เป็นต้น ในขั้นแรกจะทำการเปรียบเทียบการจดจำของระดับคำใน Conditional Random Fields ในระดับแกรมของระดับคำมี 3, 5 และ 7 แกรม โดยคลังข้อมูลที่ใช้ในการเรียนรู้และทดสอบโมเดล คือ BEST 2010 (Benchmark for Enhancing the Standard of Thai language processing) ซึ่งเป็นคลังข้อมูลมาตรฐานที่ใช้สำหรับการประมวลผลภาษาธรรมชาติ ผลการทดลองพบว่าการใช้จำนวน 7 แกรม สามารถจดจำนิพจน์ระบุนามได้ดีที่สุด หลังจากนั้นนำโมเดลการจดจำนิพจน์ระบุนามมาปรับปรุงประสิทธิภาพการตัดคำ โดยนำผลที่ได้จากการตัดคำภาษาไทยของโปรแกรม Thai Lexeme Analyser (TLex) ซึ่งใช้เทคนิค Conditional Random Fields นำมาปรับปรุงประสิทธิภาพ จากการทดลองพบว่า การตัดคำด้วยโปรแกรม Thai Lexeme Analyser (TLex) ทำการทดลองกับ 37 บทความ มีจำนวน 72,000 คำ ผลการทดลองโดยวัดประสิทธิภาพ F1 มีค่ารวมโดยเฉลี่ย 92.23% และหลังจากการปรับปรุงประสิทธิภาพการตัดคำโดยการจดจำนิพจน์ระบุนามสามารถช่วยเพิ่มประสิทธิภาพการตัดคำ F1 เท่ากับ 93.80 % หรือเพิ่มขึ้น 1.57%