

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้เป็นการศึกษาการปรับปรุงประสิทธิภาพตัดคำโดยการจดจำนิพจน์ระบุนามโดยมีวัตถุประสงค์เพื่อปรับปรุงการตัดคำในภาษาไทยให้มีประสิทธิภาพมากขึ้น จากผลการทดลองสามารถสรุปผลการศึกษา และได้นำข้อเสนอแนะเพื่อเป็นแนวทางในการศึกษาและพัฒนา งานวิจัยที่ต่อเนื่องหรือเป็นแนวทางในการสร้างสรรค์ผลงานวิจัยด้านอื่นๆ ต่อไป

5.1 สรุปผลการวิจัย

การปรับปรุงประสิทธิภาพการตัดคำโดยการสกัดนิพจน์ระบุนาม เป็นการแก้ไข ปัญหาการตัดคำภาษาไทยจากโปรแกรมที่เล็กส์คือ โปรแกรมแบ่งคำภาษาไทย ซึ่งพัฒนาโดยศูนย์ เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) การเรียนรู้ด้วยเครื่องคอมพิวเตอร์ (Machine Learning) โดยอาศัยหลักการของ Conditional Random Fields (CRFs) ในการเรียนรู้ และใช้คลังข้อมูลของ BEST2009 ขนาด 5 ล้านคำในการฝึกฝนโปรแกรมที่เล็กส์ ปัญหาการตัดคำ ของโปรแกรมที่เล็กส์โดยส่วนใหญ่เกิดมาจากคำที่คอมพิวเตอร์ไม่รู้จัก (Unknown Word) ซึ่ง ปัญหาหลักของ คำที่คอมพิวเตอร์ไม่รู้จักนั่นคือ นิพจน์ระบุนาม เป็นส่วนใหญ่ เพราะนิพจน์ระบุนาม สามารถเกิดขึ้นใหม่ได้ตลอดเวลา ไม่มีกฎตายตัวในการเกิดนิพจน์ระบุนาม โดยเฉพาะนิพจน์ ระบุนามที่มาจากภาษาต่างประเทศ ปัญหาเหล่านี้ทำให้เป็นปัญหาในการตัดคำภาษาไทย การ แก้ไขปัญหานี้ อาจใช้วิธีการแก้ไขหลายวิธี เช่น การสร้างพจนานุกรม นิพจน์ระบุนามยังมีเป็น ปัญหาว่าไม่สามารถสร้างพจนานุกรมได้ครอบคลุมนิพจน์ระบุนามคำที่เกิดขึ้นใหม่ได้แนวโน้มใน การแก้ไขปัญหาคือ การใช้วิธีการเรียนรู้ (Machine Learning) ในการจดจำนิพจน์ระบุนาม

การปรับปรุงประสิทธิภาพการตัดคำโดยการจดจำนิพจน์ระบุนาม ด้วยวิธี Condition Random Fields เป็นวิธีการที่ดีวิธีการหนึ่งในการจดจำนิพจน์ระบุนาม เพื่อนำมาใช้ในการ ปรับปรุงประสิทธิภาพการตัดคำของภาษาไทย จะพบว่าการใช้วิธีทางสถิติอื่นๆ เช่น N-Gram และ Maximum Entropy Markov Models (MEMMs) ที่ทำการเปรียบเทียบการตัดคำเราพบว่า การตัด คำด้วย Condition Random Fields เป็นวิธีที่ดีที่สุดในการตัดคำ และผลของโมเดลการจดจำ นิพจน์ระบุนาม สามารถจดจำได้ดีที่สุดคือบทความกฎหมายได้ถึง 97.41% จะสามารถจดจำได้ดี ที่สุดเพราะเป็นบทความที่มีรูปแบบการเขียนที่เป็นทางการและมีคำขึ้นต้นและคำลงท้ายและการ

นำไปปรับปรุงประสิทธิภาพการตัดคำด้วยโปรแกรม TLex ก็สามารเพิ่มประสิทธิได้ดี จากค่า F1 โดยรวมของ 6 ประเภทบทความ การตัดคำโดยโปรแกรม TLex มีค่าเท่ากับ 92.23 % และใช้โมเดลการปรับปรุงประสิทธิภาพการตัดคำ สามารถเพิ่ม ค่า F1 ได้ถึง 93.80% โดยการปรับปรุงประสิทธิภาพการตัดคำอาจมีปัจจัยที่มีผลต่อการปรับปรุงประสิทธิภาพของการตัดคำโดยการจดจำนิพจน์ระบุนามก็คือ

1. จำนวน นิพจน์ระบุนามในแต่ละบทความที่ทำการปรับปรุงประสิทธิภาพการตัดคำ
2. การมีคำขึ้นต้นหรือคำลงท้ายของนิพจน์ระบุนามเพื่อเพิ่มประสิทธิภาพในการจดจำนิพจน์ระบุนาม
3. นิพจน์ระบุนามที่มีคำหลายคำประกอบขึ้นเป็นนิพจน์ระบุนามหรือ เป็นภาษาต่างประเทศ เช่น คณะปฏิรูปการปกครอง ในระบอบประชาธิปไตย อันมีพระมหากษัตริย์ทรงเป็นประมุข โครงการส่งน้ำและบำรุงรักษาเขื่อนป่าสักชลสิทธิ์ เป็นต้น ในส่วนของภาษาต่างประเทศ เช่น จักรพรรดินโปเลียนที่ 1 เจแปนแอร์ไลน์ กษัตริย์อาเธอร์ เป็นต้น ซึ่งนิพจน์ระบุนามเหล่านี้ยากในการจดจำนิพจน์ระบุนาม

5.2 ข้อเสนอแนะสำหรับการศึกษาวิจัยในอนาคต

ข้อเสนอแนะ สำหรับการศึกษาวิจัยในอนาคต ของการปรับปรุงประสิทธิภาพของการตัดภาษาไทยโดยการสกัดนิพจน์ระบุนามโดยการใช้วิธีการ Conditional Random Fields ในการเรียนรู้

1. อาจเพิ่มจำนวนแกรมในการรู้จำเช่น 9 แกรม 11 แกรม เป็นต้น
2. การนำบริบทข้างเคียงเข้ามาช่วยในการสกัดนิพจน์ระบุนาม เช่นคำว่า กล่าวว่, แดง, เปิดเผยว่, เพื่อช่วยในการจดจำนิพจน์ระบุนามได้มากขึ้น
3. ใช้หน้าที่ของคำ (Part of Speech) เป็น คุณลักษณะ ในการปรับปรุงเพื่อช่วยเพิ่มประสิทธิภาพในการจดจำนิพจน์ระบุนาม