

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงการสรุปผลการวิจัย การสกัดนิพจน์ระบุนามสำหรับเหมืองข้อมูลข่าวภาษาไทยตามกิจกรรมของบุคคลและข้อเสนอแนะหรือแนวทางการพัฒนางานวิจัยทางด้านนี้ในอนาคต

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการสกัดนิพจน์ระบุนามสำหรับเหมืองข้อมูลข่าวภาษาไทยตามกิจกรรมของบุคคล โดยใช้บทความข่าวทางด้านการเมืองมาทดสอบและใช้การประเมินผลประสิทธิภาพจากการเปรียบเทียบระบบกับการสกัดด้วยมือมนุษย์ใช้เป็นเกณฑ์มาตรฐาน ซึ่งแบ่งการทำงานออกเป็น 2 โมเดล โมเดลแรกเป็นการสกัดนิพจน์ระบุนาม โดยจะเลือกสกัดเฉพาะนิพจน์ระบุนามประเภทบุคคล จากการศึกษางานวิจัยที่ผ่านมา เทคนิคที่ใช้ในสกัดนิพจน์ระบุนามที่มีความถูกต้องแม่นยำและเหมาะสมกับงานทางด้านการประมวลผลภาษาธรรมชาติก็คือการเรียนรู้ด้วยเครื่องด้วยเทคนิค Conditional Random Fields (CRF) งานวิจัยนี้จะใช้คลังข้อมูลของ BEST 2009 จำนวน 206,695 คำ มาฝึกอบรม (training) ในการสกัดนิพจน์ระบุนามจะต้องฝึกให้ระบบสามารถระบุนิพจน์ระบุนามประเภทบุคคลได้ ดังนั้นงานวิจัยนี้จึงใช้คุณสมบัติของคำบางที่เข้ามาเพื่อฝึกให้ระบบสามารถระบุนิพจน์ระบุนามประเภทบุคคล โมเดลที่สองเป็นการหาความสัมพันธ์ที่เกิดขึ้นของนิพจน์ระบุนามตามกิจกรรมของบุคคลของบทความข่าวทางด้านการเมือง งานวิจัยนี้ได้ให้นิยามของกิจกรรมคือคำที่เป็นคำกริยา จึงใช้เทคนิคสวอธ (swath) เป็นเครื่องมือเข้ามาช่วยเนื่องจากเทคนิคนี้จะมีการกำกับชนิดของคำ ทำให้สามารถคัดกรองคำที่เป็นกิจกรรมได้นอกจากนั้นเพื่อให้ระบบมีประสิทธิภาพในการสกัดนิพจน์ระบุนามตามกิจกรรมของบุคคลมากขึ้น จึงใช้เทคนิคเลกส์โตเข้ามาช่วยในการตัดคำเพื่อลดความกำกวมของชื่อบุคคลและกิจกรรม หลังจากนั้นใช้กฎเข้ามาช่วยเพื่อเพิ่มความถูกต้องมากขึ้นในการสกัดนิพจน์ระบุนามตามกิจกรรมของบุคคล จากผลการทดลอง การสกัดนิพจน์ระบุนามสำหรับเหมืองข้อมูลข่าวภาษาไทยตามกิจกรรมของบุคคลสามารถสรุปได้ว่าการพิจารณาในขอบเขต 10 คำ มีค่าประสิทธิภาพโดยรวมของระบบดีกว่าการพิจารณาในขอบเขต 5 คำ โดยค่าประสิทธิภาพโดยรวมของระบบใน

ขอบเขต 5 คำ จะอยู่ที่ 79.50 และ ค่าประสิทธิภาพโดยรวมของระบบในขอบเขต 10 คำ จะอยู่ที่ 84.68

5.2 ข้อเสนอแนะ

จากงานวิจัยที่ผ่านมา รวมทั้งผลที่ได้จากงานวิจัยฉบับนี้ พบว่า การเพิ่มประสิทธิภาพให้กับระบบการสกัดนิพจน์ระบุนามตามกิจกรรมของบุคคลในส่วนของค่าความระลึกต้องคำนึงถึงบริบทของคำที่ใช้ในการกล่าวอ้างถึงบุคคลอื่นที่ทำการกิจกรรม เนื่องจากบุคคลนั้นไม่ได้ทำการกิจกรรมจริงเป็นแค่การกล่าวถึง นอกจากนี้ ยังพบว่าจำนวนคำที่ใช้ในการฝึกอบรมและขอบเขตของบริบทที่ใช้ในการสกัดก็มีส่วนสำคัญที่ทำให้ประสิทธิภาพโดยรวมต่ำกว่าที่ควรจะเป็น

ในส่วนของงานวิจัยที่สามารถนำไปศึกษาและพัฒนาเพิ่มเติมได้ในอนาคตจากงานวิจัยฉบับนี้หรืองานวิจัยทางด้านนี้ได้แก่

1. พัฒนาในส่วนของนิพจน์ระบุนามโดยเพิ่ม นิพจน์ระบุนามประเภทองค์กร และนิพจน์ระบุนามประเภทสถานที่ เพื่อให้สามารถสืบค้นได้ว่าบุคคลนี้ทำการกิจกรรมอยู่ที่ไหนและองค์กรใดเป็นผู้จัดกิจกรรม
2. พัฒนาเรื่องของเหมืองข้อความในการจัดจำแนกประเภทกิจกรรมที่เกิดขึ้น