

บทที่ 1

บทนำ



ความเป็นมาและความสำคัญของปัญหา

การเขียนประโยคภาษาไทยเป็นการเขียนติดต่อกันโดยไม่มีการใช้เครื่องหมายวรรคตอนเป็นส่วนใหญ่ ซึ่งก็มีการแบ่งวรรคตอนด้วยช่องว่างบ้างเป็นบางครั้งจึงทำให้มีความแตกต่างจากการเขียนประโยคภาษาอังกฤษ ที่มีการแบ่งแยกคำแต่ละคำด้วยช่องว่างอย่างชัดเจน ลักษณะเช่นนี้ทำให้การประมวลผลภาษาไทยด้วยคอมพิวเตอร์มีความจำเป็นที่จะมีวิธีการให้คอมพิวเตอร์มีความสามารถที่จะเรียนรู้ขอบเขตของคำ เพื่อแก้ปัญหาเรื่องการตัดคำในการประมวลผลคำภาษาไทยให้ถูกต้องตามหลักภาษาศาสตร์

การพัฒนาขั้นตอนวิธีการตัดคำภาษาไทย ได้มีการพัฒนามาเป็นเวลานานจากหน่วยงานต่างๆ ซึ่งแต่ละวิธีการตัดคำมีขั้นตอนการตัดคำและประสิทธิภาพที่ได้จากการตัดคำมีความแตกต่างกันตามลักษณะของงานที่จะนำไปใช้ โดยจะยังคงมีปัญหาคำที่ตัดคำจากข้อความที่มีความกำกวมและปัญหาการตัดคำคำที่ไม่ปรากฏในพจนานุกรม ในงานวิจัยนี้จะพิจารณาเฉพาะการตัดคำจากข้อความที่มีความกำกวมเท่านั้น ซึ่งแต่ละขั้นตอนวิธีการตัดคำที่ได้มีการพัฒนาขึ้นมา นั้น จะทำการเลือกการตัดคำแบบใดแบบหนึ่ง ซึ่งบางวิธีก็จะเป็นการตัดคำที่ถูกต้อง บางวิธีก็เป็นการตัดคำที่ทำให้ได้ความหมายของข้อความผิดไป เช่น ข้อความ ตากลม สามารถตัดคำได้เป็น ตาก ลม กับ ตา กลม ในงานวิจัยนี้เรียก ตาก ลม ว่าเป็นหนึ่งคำกำกวม ถ้าใช้วิธีแบบเลือกคำที่ยาวที่สุด (longest matching) ทำการตัดคำจากข้อความนี้จะได้เป็น ตาก ลม ถ้าใช้วิธีการตัดคำให้จำนวนน้อยที่สุด (maximal matching) ก็จะได้ตัดคำได้เป็น ตาก ลม เช่นเดียวกัน ซึ่งที่จริงแล้วจะเป็นการตัดคำแบบใดนั้นดูได้จากคำบริบท หรือคำที่อยู่ติดกันกับข้อความนั้น เช่นถ้าอยู่ในข้อความ ปลาทองตัวนี้มีตากลมโตกว่าตัวนั้นมีคำว่า มี อยู่ใกล้เคียงก็มีความน่าจะเป็นสูงที่จะตัด ตากลม เป็น ตา กลม แต่ถ้าตัดคำออกมาเป็น ตาก ลม ก็จะทำให้ข้อความมีความหมายผิดไป

ดังนั้นในงานวิจัยนี้จะนำคุณลักษณะทางไวยากรณ์มาใช้ในการตัดคำกำกวม ซึ่งคำกำกวมจะหมายถึงแบบหนึ่งที่ได้จากการตัดคำจากข้อความกำกวม ที่สามารถตัดคำออกได้เป็นหลายแบบ โดยลักษณะทางไวยากรณ์ที่นำมาใช้คือ คำบริบทและคำที่อยู่ก่อนหน้าและหลังคำกำกวมมาเป็นคุณลักษณะในการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัยคือระบบ FOIL เพื่อจะได้อนุประโยคผลลัพท์ของคุณลักษณะหรือกฎของแต่ละคำกำกวม ที่จะนำไปช่วยการตัดคำกำกวม

วัตถุประสงค์

1. ศึกษาวิธีการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย คือระบบ FOIL เพื่อนำไปประยุกต์ใช้ในการตัดคำกำกวมในข้อความภาษาไทย
2. พัฒนาโมดูลย่อยที่สามารถตัดคำกำกวมจากการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย เพื่อดูว่าการเรียนรู้คุณลักษณะต่างๆของคำกำกวมที่ได้จากการเรียนรู้ด้วยโปรแกรมตรรกะเชิงอุปนัย จะสามารถช่วยการตัดคำกำกวมได้ดีขึ้นหรือไม่

ขอบเขตของการวิจัย

1. ใช้วิธีการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย คือระบบ FOIL เป็นเครื่องมือในการหาคุณลักษณะของแต่ละคำกำกวม
2. พัฒนาโมดูลย่อยที่สามารถตัดคำกำกวมจากการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย

ขั้นตอนการวิจัย

1. ศึกษาวิธีการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย
2. ทำการหาคุณลักษณะต่างๆของแต่ละคำกำกวมจากการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัย
3. พัฒนาโมดูลย่อยที่สามารถตัดคำกำกวมจากการเรียนรู้คำกำกวมโดยใช้ระบบโปรแกรมตรรกะเชิงอุปนัย
4. ทดสอบโมดูลย่อยที่ได้พัฒนาขึ้น
5. สรุปผลการวิจัย

ประโยชน์ที่ได้จากการวิจัย

1. ได้โมดูลย่อยที่สามารถตัดคำกำกวมจากการเรียนรู้คำกำกวมโดยใช้ระบบโปรแกรมตรรกะเชิงอุปนัย
2. แนวทางการปรับปรุงขั้นตอนวิธีการตัดคำกำกวม