

บทที่ 5

สรุปและอภิปรายผลการทดลอง

ในบทนี้จะเป็นการสรุปภาพรวมของการสกัดสาระสำคัญของเอกสารคดีความ ตลอดจนจนถึงปัญหาและอุปสรรค และรวมถึงข้อเสนอแนะที่ได้ทำการรวบรวมไว้ดังนี้

5.1 สรุปผลและอภิปรายผล

โครงงานฉบับนี้เป็นโครงงานเรื่องการสกัดสาระสำคัญในรูปแบบประโยคในแต่ละส่วนของเอกสารคดีความ ซึ่งมีอยู่ 4 ส่วนคือ ข้อพิพาทของคดีความ ข้อเท็จจริงของคดีความ ข้อวินิจฉัยของคดีความ และคำตัดสินของคดีความ จากนั้นจะนำเอกสารที่ได้มาเข้ากระบวนการสรุปความซึ่งมีหลักการดังนี้

ขั้นตอนในการดำเนินงานเพื่อสรุปความ ขั้นแรกจะเป็นการรวบรวมเอกสารเกี่ยวกับคดีความ เรื่องการซื้อขายที่ดิน ซึ่งรวบรวมมาจากเว็บไซต์ของศาลฎีกา และเอกสารจะถูกจัดเก็บให้อยู่ในรูปแบบ XML

ขั้นตอนการแยกองค์ประกอบของเอกสาร ขั้นตอนนี้เป็นการแยกเอกสารคดีแดงออกเป็น 4 ส่วนคือ Dispute, Fact, Decision และ Judgment เพื่อนำแต่ละส่วนที่ได้มาทำการสรุปความ

ขั้นตอนการสกัดประโยคสำคัญเพื่อการสรุปความ จะเป็นการนำเอาเอกสารที่รวบรวมไว้มาผ่านกระบวนการตัดคำ และทำการแยกประโยค โดยคำเหล่านั้นสามารถแสดงให้อยู่ในรูป Vector Space Model เพื่อแสดงความสัมพันธ์ระหว่างคำกับเอกสาร จากนั้นจะให้ค่าน้ำหนักด้วย tf-idf เพื่อใช้ในการให้ค่าน้ำหนักคำที่อยู่ในเอกสารซึ่งจะใช้เป็นตัวแทนของเอกสารนั้น จากนั้นจะนำเอกสารที่รวบรวมไว้มาวิเคราะห์หาค่าความคล้ายคลึงด้วยอัลกอริทึม BM25+ และจะใช้การสกัดประโยคด้วยการวิเคราะห์ความคล้ายคลึงของข้อความแบบใช้ค่าเทรชโฮลด์ (Threshold-based Text Similarity) โดยในการเปรียบเทียบประโยค จะเปรียบเทียบระหว่างประโยคที่เป็นการสรุปความโดยผู้เชี่ยวชาญและประโยคในเอกสารที่กำลังจะทำการสรุปความแบบอัตโนมัติ หากประโยคใดในเอกสารที่กำลังพิจารณา มีความเหมือนหรือสอดคล้องกันกับประโยคที่เป็นการสรุปความโดยผู้เชี่ยวชาญ ก็จะถูกสกัดออกมาเป็นประโยคของการสรุปความ

สำหรับขั้นตอนการวัดประสิทธิภาพในโครงการนี้จะใช้ การวัดค่าความระลึก ROUGE-N และ ROUGE-L โดยค่าความระลึกจะเป็นอัตราส่วนของเอกสารที่จัดกลุ่มได้จากเอกสารทั้งหมดที่มีอยู่

5.2 ปัญหาและอุปสรรคในการดำเนินงาน

5.2.1 ปัญหาเกี่ยวกับการตัดคำ

เนื่องจากการใช้การตัดคำแบบเปรียบเทียบคำที่ยาวที่สุด (Longest Matching) ซึ่งจะเปรียบเทียบกับพจนานุกรม และคำในคติความบางคำอาจจะไม่ปรากฏในพจนานุกรม จึงทำให้การตัดคำทำงานได้ไม่มีประสิทธิภาพมากพอ เพราะจะทำการตัดคำออกมาเป็นคำที่มีความยาวมากที่สุด ซึ่งอาจจะทำให้นำไปเปรียบเทียบแล้วไม่พบคำนั้น ๆ ได้

5.2.2 ปัญหาเกี่ยวกับขอบเขตของประโยค

เนื่องจากประโยคในภาษาไทยจะไม่มีจุด (Period) บอจุดสิ้นสุดของประโยคเหมือนกับประโยคในภาษาอังกฤษ จึงทำให้การตัดประโยคในภาษาไทยมีข้อผิดพลาดเกิดขึ้น เนื่องจากบางประโยคที่ตัดมาได้อาจจะมี ความหมาย หรือบางประโยคก็อาจจะไม่มีความหมาย

5.2.3 ปัญหาเกี่ยวกับการแยกองค์ประกอบของเอกสาร

เนื่องจากคำ หรือ วลีสำคัญที่ใช้ในการแยกองค์ประกอบของเอกสารบางคำมักปรากฏอยู่ในทุกส่วนของเอกสาร จึงอาจจะทำให้การแยกองค์ประกอบของเอกสารมีประสิทธิภาพในการแยกองค์ประกอบได้ค่อนข้างน้อย ซึ่งเมื่อทำการแยกองค์ประกอบไม่ได้ก็จะส่งผลให้ทำการสกัดสาระสำคัญมีข้อผิดพลาด หรือไม่สามารสกัดสาระสำคัญได้

5.3 ข้อเสนอแนะ

1. ควรรวบรวมคำศัพท์ที่เกี่ยวข้องให้ได้มากที่สุดหรือครอบคลุมเรื่องที่ศึกษามากที่สุด เพื่อให้ได้โมเดลมีประสิทธิภาพมากที่สุด
2. ประสิทธิภาพของโมเดลการสกัดประโยคสำคัญเพื่อสรุปความ จะขึ้นอยู่กับเอกสารที่ใช้สร้างโมเดลและความถูกต้องของเอกสารที่ใช้ในการสร้างโมเดล

ดังนั้นการสกัดสาระสำคัญจากเอกสารคติความในเรื่องของการซื้อขายที่ดิน ควรจะมีจำนวนคำศัพท์ที่จำเป็นสำหรับการสกัดสาระสำคัญปริมาณไม่น้อยจนเกินไป และตรงกับเรื่องที่ศึกษามากที่สุด