

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

คดีความ (Lawsuit) [1] คือเรื่องที่พิพาทหรือกล่าวหากันในทางกฎหมาย ซึ่งต้องดำเนินการตามกระบวนการพิจารณาความตามที่กฎหมายกำหนด ซึ่งคดีความที่เกี่ยวข้องกับชีวิตประจำวันของผู้คนจะมีอยู่ 2 คดีคือ คดีแพ่งและพาณิชย์ (Civil And Commercial Case) และคดีอาญา (Criminal Case) [2]

โดยคดีแพ่งและพาณิชย์ [3] เป็นคดีที่เกี่ยวกับเรื่องส่วนตัวของบุคคล 2 ฝ่าย ที่มีการทำผิดสัญญาหรือโต้แย้งสิทธิกัน เป็นเรื่องที่ได้รับ การโต้แย้งสิทธิ จะฟ้องร้องอีกฝ่ายที่ทำการโต้แย้งสิทธิ หรือทำผิดสัญญา ตัวอย่างคดีแพ่งและพาณิชย์ เช่น คดีกู้ยืมเงิน คดีผิดสัญญา คดีเช่าทรัพย์ คดีตัวเงิน คดีจำนอง คดีซื้อขาย คดีมรดก เป็นต้น

สำหรับคดีอาญา [2] นั้น เป็นคดีที่เมื่อเกิดขึ้นจะกระทบกระเทือนถึงสาธารณชนในบ้านเมือง โดยคดีที่ฟ้องร้องกันเนื่องจากการทำความผิดทางอาญาหรือรับโทษอื่น ๆ ในทางอาญา เช่น ให้ปรับ ให้จำคุกหรือ ให้ประหารชีวิต ตัวอย่างคดีอาญา เช่น คดีทำร้ายร่างกาย คดีลักทรัพย์ คดีชิงทรัพย์ คดีปล้นทรัพย์ คดีฆ่าคนตาย คดีประมาททำให้ผู้อื่นบาดเจ็บ หรือเสียชีวิต คดีรับของโจร เป็นต้น

อย่างไรก็ตาม เมื่อเกิดคดีความก็จะเกิดการร้องทุกข์หรือการฟ้องร้อง ในกรณีที่เป็นเจ้าทุกข์หรือผู้ร้องทุกข์จะเรียกว่า “โจทก์ (Plaintiff)” [4] แต่หากเป็นผู้ที่คาดว่าเป็นผู้กระทำผิดก็จะเรียกว่า “จำเลย (Defendant)” [4] โดยทั่วไปแล้ว การตัดสินวินิจฉัยคดีความจะมี 3 ศาลคือ ศาลชั้นต้น ศาลอุทธรณ์ และศาลฎีกา คดีใดก็ตามหากมีการตัดสินในชั้นศาลฎีกาไม่ว่าผลจะออกมาเป็นเช่นไร ถือว่าคดีนั้นได้สิ้นสุดลงแล้ว [5] สำหรับเป็นคดีที่รู้ผลแล้วว่าใครเป็นฝ่ายผิดฝ่ายถูก เนื่องจากศาลมีคำสั่งวินิจฉัยชี้ขาดหรือพิพากษา (Judge) แล้ว จะเรียกว่า “คดีแดง (Decided case)” [4]

สำหรับผู้ที่สนใจด้านคดีความนั้น คดีแดงกลายเป็นแหล่งศึกษาความรู้และข้อมูลด้านกฎหมายที่สำคัญของประชาชน นักศึกษา หรือแม้แต่พนักงานกฎหมายเอง เพราะในการตัดสินคดีความในแต่ละคดีนั้น ผู้พิพากษาอาจจะมีมุมมองในตัดสินคดีความที่แตกต่างกันไป อย่างไรก็ตาม คดีที่ผ่านการ

ตัดสินแล้วและมีการเผยแพร่ มักจะอยู่ในรูปแบบเอกสารข้อความที่ค่อนข้างยาว ทำให้ยากต่อการทำความเข้าใจ แม้ว่าเอกสารคดีแดงจะใช้คำศัพท์และรูปแบบที่เป็นมาตรฐานก็ตาม ซึ่งก็รวมถึงคดีแดงของประเทศไทยด้วย

ดังนั้นโครงการปริญาพนธ์นี้ จึงนำเสนอการประยุกต์เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ในการสรุปประเด็นสำคัญที่อยู่ในคดีแดง

1.2 วัตถุประสงค์ของโครงการ

นำเสนอการประยุกต์กระบวนการแบบ Extraction-Based Text Summarization ในการสรุปประเด็นสำคัญที่อยู่ในคดีแดงของคดีแพ่งและพาณิชย์ โดยเน้นคดีความเรื่องการซื้อขายที่ดิน

1.3 ขอบเขตของโครงการ

- 1) นำเสนอการประยุกต์กระบวนการแบบ Extraction-Based Text Summarization ในการสรุปประเด็นสำคัญที่อยู่ในคดีแดงของคดีแพ่งและพาณิชย์ โดยเน้นคดีความเรื่องการซื้อขายที่ดิน
- 2) Extraction-Based Text Summarization ที่ใช้ในโครงการฯ นี้ จะเป็นการสรุปความแบบการสกัดสาระสำคัญในรูปแบบประโยคในแต่ละส่วนของเอกสารคดีความ ซึ่งมีอยู่ 4 ส่วนคือ ข้อพิพาทของคดีความ (Dispute) ข้อเท็จจริงของคดีความ (Fact) ข้อวินิจฉัยของคดีความ (Decision) และคำตัดสินของคดีความ (Judgment)
- 3) คดีแดงที่นำมาใช้ในโครงการฯ นี้ นำมาจากเว็บไซต์ของศาลฎีกา <http://deka.supremecourt.or.th/>
- 4) ในการทดสอบการสรุปความ จะทดลองกับชุดข้อมูลคดีแดงจำนวน 60 คดี โดยคดีเหล่านี้ โดยใช้ข้อมูลในการเป็นผลเฉลย 10 คดี ใช้ในการทดสอบเพื่อกำหนดค่า Threshold 10 คดี และทำการทดสอบ 30 คดี
- 5) มีผู้เชี่ยวชาญ (ทนายความ) อย่างน้อย 1 ท่าน ได้ทำการสรุปเอาไว้สำหรับเป็นผลเฉลย
- 6) การประเมินผลลัพธ์ของการสรุปความจะประเมินด้วยค่าความระลึก (Recall) ในการประเมินการแยกองค์ประกอบของเอกสาร และค่า ROUGE-N และ ROUGE-L ในการประเมินข้อความที่สรุปได้

- 7) ใช้การวิเคราะห์ความคล้ายคลึง (Similarity Analysis) ในการเปรียบเทียบความคล้ายคลึงระหว่างประโยค และใช้ค่าเทรชโฮลด์ (Threshold-Based Analysis) ในการเลือกว่าประโยคนั้นๆ จะนำมาใช้ในการสรุปความหรือไม่

1.4 นิยามศัพท์เฉพาะ

คดีแดง (Decided case) ในโครงการงานปริญญาโทครั้งนี้คือคดีแดงที่ถูกตัดสินในศาลฎีกา เป็นคดีแดงในกลุ่มคดีแพ่งและพาณิชย์ ในเรื่องของการซื้อขายที่ดิน

Passage คือ ส่วนของประโยคในเอกสารที่จะใช้ในการพิจารณาค่าความคล้ายคลึงเมื่อเทียบกับส่วนของเอกสารที่เป็นข้อสรุปความต้นแบบ โดย Passage ในโครงการงานนี้จะเป็น 1 ประโยค, 2 ประโยค จนถึง 5 ประโยค

Extraction-Based Text Summarization คือกระบวนการที่จะทำการสรุปความ โดยการสกัดใจความที่คาดว่าจะมีสำคัญจากเอกสารต้นฉบับออกมา และนำมารวมกันเพื่อเป็นข้อความสรุปย่อของเอกสาร โดยไม่ได้สนใจความสละสลวยของข้อความที่สรุปออกมา

Threshold-based Text Similarity คือ ค่าระดับความคล้ายคลึงที่จะพิจารณาประโยคในเอกสารว่า ควรจะถูกคัดเลือกให้อยู่ในเนื้อหาของสรุปความหรือไม่

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้กระบวนการแบบ Extraction-Based Summarization ในการสรุปประเด็นสำคัญที่อยู่ในคดีแดงของคดีแพ่งและพาณิชย์ โดยเน้นคดีความเรื่องการซื้อขายที่ดินที่เป็นภาษาไทย

1.5.2 บุคคลทั่วไปและนักกฎหมายสามารถทำความเข้าใจได้ง่ายยิ่งขึ้นจากการสรุปความภาษาไทย

1.6 อุปกรณ์และเครื่องมือที่ใช้ในการดำเนินงาน

1.6.1 ฮาร์ดแวร์

เครื่องคอมพิวเตอร์ 2 เครื่อง ที่มีคุณสมบัติดังต่อไปนี้

1. Asus

- Intel(R) Core (TM) i3-3110M CPU @ 2.40GHz
- RAM 8 GB, SSD 480 GB
- NVIDIA GEFORCE 610M 2GB
- 64 bit Operating System

ตารางที่ 1.1 แผนการดำเนินงาน (ต่อ)

กิจกรรม	เดือน												
	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	
7. นำเสนอ โครงการ													

Computer Science Department
Faculty of Informatics, Maharakham University