การเปรียบเทียบเทคนิคการเพิ่มประสิทธิภาพการค้นคืนกฎหมายที่ดิน

โดย

วัชระ โพธิ์รุ้ง

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2566

A COMPARISON OF TECHNIQUES TO IMPROVE THE LAND LEGAL TEXT RETRIEVAL

BY

WATCHARA PHORUNG

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2024

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

สารนิพนธ์

ของ

วัชระ โพธิ์รุ้ง

เรื่อง

การเปรียบเทียบเทคนิคการเพิ่มประสิทธิภาพการค้นคืนกฎหมายที่ดิน

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

เมื่อ วันที่ 19 มิถุนายน พ.ศ. 2567

ประธานกรรมการสอบสารนิพนธ์ _____
(ผู้ช่วยศาสตราจารย์ ดร.เสาวลักษณ์ วรรธนาภา)

กรรมการและอาจารย์ที่ปรึกษาสารนิพนธ์ _____
(ผู้ช่วยศาสตราจารย์ ดร.วิรัตน์ จารีวงศ์ไพบูลย์)

กรรมการสอบสารนิพนธ์ _____
(ผู้ช่วยศาสตราจารย์ ดร.นุชจรินทร์ อินต๊ะหล้า)

คณบดี _____
(รองศาสตราจารย์ ดร.สุเพชร จิรขจรกุล)

หัวข้อสารนิพนธ์	การเปรียบเทียบเทคนิคการเพิ่มประสิทธิภาพการค้นคืนกฎหมายที่ดิน
ชื่อผู้เขียน	วัชระ โพธิ์รุ้ง
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาสารนิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.วิรัตน์ จารีวงศ์ไพบูลย์
ปีการศึกษา	2566

บทคัดย่อ

ในปัจจุบันมีประเด็นข้อพิพาทหรือความขัดแย้งเกิดขึ้นมากในสังคมประเทศไทย ซึ่งหนึ่งในนั้นเป็นประเด็นข้อพิพาทหรือความขัดแย้งที่เกี่ยวข้องกับกรรมสิทธิ์ในที่ดิน มีการฟ้องร้องร้องเรียน เกิดขึ้นอยู่ตลอดเวลา หลายครั้งที่ปัญหาดังกล่าวนี้ ก่อให้เกิดการทะเลาะวิวาท บางกรณีนำไปสู่การสูญเสียชีวิต ซึ่งส่งผลกระทบต่อทรัพย์สิน ความเป็นอยู่และความปลอดภัยในชีวิตของประชาชน ซึ่งการค้นหาข้อมูลเกี่ยวกับข้องกับข้อพิพาททางกฎหมายโดยการใช้งานอินเทอร์เน็ตเป็นแนวทางหนึ่งในการลดปัญหา อย่างไรก็ตามการใช้ข้อมูลจากอินเทอร์เน็ตนั้นจำเป็นต้องใช้วิจารณญาณในการคิด วิเคราะห์ แยกแยะข้อมูลที่ได้อย่างถี่ถ้วนและใช้ข้อมูลจากหลากหลายแหล่งที่มา เพื่อให้ได้ข้อมูลที่ถูกต้องสมบูรณ์

มีระบบที่ให้บริการสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา การใช้งานระบบดังกล่าวข้างต้นต้องการระบุคำค้นหาแบบเฉพาะเจาะจง (Exact Matching) โดยสามารถใช้เครื่องหมายตรรกะ (Logical Operators) เพื่อช่วยให้ผลลัพธ์ของการค้นคืนคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา มีความถูกต้องตรงกับสิ่งที่ต้องการสืบค้นมากขึ้น แต่มีข้อจำกัดที่ไม่สามารถสืบค้นคำค้นหาจากความหมายคล้ายคลึงกันของข้อความได้

สารนิพนธ์ฉบับนี้มีแนวคิดที่จะทำระบบสืบค้นหาคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาของกฎหมายที่เกี่ยวข้องกับที่ดิน ด้วยชุดข้อมูลข้อมูลที่ได้รวบรวมและจัดทำขึ้นมาใหม่ โดยมุ่งเน้นไปที่การเปรียบเทียบระยะเวลาและประสิทธิภาพในการค้นคืนระหว่างเทคนิคการคัดแยกข้อความ (Text Classification) ร่วมกับการใช้เทคนิคการจับคู่ความหมาย (Semantic Matching) และเทคนิคที่ใช้การจับคู่ความหมายเพียงอย่างเดียว ในการค้นคืนคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาที่เกี่ยวข้องกับกฎหมายที่ดิน

จากผลการทดลองพบว่าการใช้เทคนิคที่ใช้การจับคู่ความหมายเพียงอย่างเดียวให้ประสิทธิภาพการค้นคืนที่ดีกว่าการเพิ่มเทคนิคการคัดแยกประเภทข้อความคำค้น (Text Classification) เมื่อใช้การหาความคล้ายคลึงกันด้วยวิธี Cosine Similarity แต่ใช้ความเร็วในการค้นคืนที่มากกว่า

คำสำคัญ: การคัดแยก, การค้นคืน, กฎหมาย, ระเบียบ, คำสั่ง, คำพิพากษา, ที่ดิน

Independent Study Title	A COMPARISON OF TECHNIQUES TO IMPROVE THE LAND LEGAL TEXT RETRIEVAL
Author	Watchara Phorung
Degree	Master of Science (Computer Science)
Department/Faculty/University	Computer Science Faculty of Science and Technology Thammasat University
Independent Study Advisor	Assistant Professor Wirat Jareevongpiboon, Ph.D.
Academic Year	2024

ABSTRACT

Nowadays, there are many disputes or conflicts in Thai society. One of them is a dispute or conflict related to land ownership. There are lawsuits, complaints all the time. Many times, this problem causes quarrels, some of which lead to loss of life, which affects property, well-being and safety in people's lives. Finding information about legal disputes by using the Internet is one way to reduce the problem. However, using information from the Internet requires critical thinking, analyzing and distinguishing data thoroughly and requires data from a variety of sources to obtain accurate and complete information.

There is a system that can search for judgments, petition orders, and Supreme Court decisions. Using the above system requires specifying specific search terms, using logical operators to assist in restoring the results of judgments, petitions, and Supreme Court decisions to match the content you want to search for more accurately. But, this method is not be able to search based on text similarity.

This paper aims to classify the judgments, petitions, and decisions of the Supreme Court on land related laws from newly collected and prepared datasets. Focusing on the response time and efficiency, This research compares two approaches text classification in combination with semantic matching techniques with semantic

matching techniques only retrieve Supreme Court judgments, orders, petitions and decisions related to land law.

According to the results, it was found that the technique using semantic matching alone provides better retrieval performance when using cosine similarity but worser retrieval speed.

Keywords: classification, infomation retrieval, legal text, the land code

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้จะไม่สามารถสำเร็จลุล่วงไปได้ด้วยดีเลยหากขาดการช่วยเหลือจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.วิรัตน์ จารีวงศ์ไพบูลย์ ที่ช่วยปรับปรุงแก้ไขและให้คำปรึกษาในการทำสารนิพนธ์ฉบับนี้ เพื่อให้สารนิพนธ์ฉบับนี้มีความสมบูรณ์

ขอขอบคุณประธานกรรมการในการสอบสารนิพนธ์ฉบับนี้ ผู้ช่วยศาสตราจารย์ ดร.เสาวลักษณ์ วรรธนาภา และคณะกรรมการในการสอบสารนิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.นุชจรินทร์ อินต๊ะหล้า ที่สละเวลาและให้ข้อเสนอแนะสารนิพนธ์ฉบับนี้เพื่อให้มีความครบถ้วนสมบูรณ์มากยิ่งขึ้น

นอกจากนี้ขอขอบคุณคณะอาจารย์ทุกท่านที่เคยประสิทธิประสาทวิชาความรู้และให้คำปรึกษาตลอดการศึกษารวมถึงครอบครัวและเพื่อนร่วมหลักสูตร วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยธรรมศาสตร์ ที่คอยส่งเสริมและเป็นกำลังใจกันตลอดมา

วัชระ โพธิ์รุ้ง

สารบัญ

สารบัญตาราง

สารบัญภาพ

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ในประเทศไทยมีข้อพิพาททางกฎหมาย ความขัดแย้งที่เกี่ยวกับสิทธิหรือคำร้องเรียนที่เกี่ยวข้องกับที่ดินและเกิดขึ้นอยู่บ่อยครั้งจากหลากหลายสาเหตุ อาทิเช่น เรื่องกรรมสิทธิ์ในที่ดิน ปัญหาที่พบได้บ่อยคือการครอบครองปรปักษ์ การขับไล่ การออกเอกสารสิทธิทับซ้อน การพิพาทแนวเขตที่ดิน ปัญหาที่ดินริมทะเล เรื่องภาระจำยอม หรือเรื่องสัญญาที่เกี่ยวข้องกับที่ดิน ไม่ว่าจะเป็นการซื้อขาย การเช่า การจำนอง เป็นต้น ซึ่งปัญหาเหล่านี้ส่งผลกระทบต่อประชาชนและเศรษฐกิจ ไม่ว่าจะเป็น การสูญเสียที่ดินและสิทธิทำกิน เกิดปัญหาความขัดแย้งและรุนแรงอาจส่งผลไปถึงการทำร้ายร่างกายกัน หรือร้ายแรงจนถึงขั้นเสียชีวิต มีผลกระทบต่อเศรษฐกิจจากการขาดการใช้ประโยชน์ในสิทธิในที่ดินนั้น ทำให้ไม่สามารถทำธุรกรรมใด ๆ ได้ การจัดเก็บรายได้ภาษีลดน้อยลง ส่งผลต่องบประมาณในการบริหารประเทศ และบางกรณีส่งผลกระทบต่อการทำลายทรัพยากรธรรมชาติ โดยข้อพิพาทต่าง ๆ เหล่านี้ได้ถูกส่งคำร้องเรียนส่งเรื่องสู่ศาลเพื่อขอความยุติธรรม คำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาจึงเป็นแนวปฏิบัติที่สำคัญในการแก้ไขปัญหาและเป็นบรรทัดฐานของข้อพิพาททางกฎหมาย ความขัดแย้งที่เกี่ยวกับสิทธิหรือคำร้องเรียนที่เกี่ยวข้องกับที่ดิน

วิธีการเบื้องต้นที่ประชาชนทั่วไปใช้เมื่อเกิดข้อพิพาท คือการค้นหาข้อมูลทางอินเทอร์เน็ต เนื่องจากอินเทอร์เน็ตเป็นส่วนสำคัญในการใช้ชีวิตของผู้คนทั่วโลก โดยใช้ในหลากหลายด้าน การเติบโตของการใช้งานอินเทอร์เน็ตก็รวดเร็วทั้งในด้านจำนวนผู้ใช้ ปริมาณข้อมูลและรูปแบบการใช้งาน ซึ่งอินเทอร์เน็ตเป็นตัวเลือกแรก ๆ ที่ประชาชนส่วนใหญ่เลือกใช้เพราะเป็นวิธีที่สะดวก รวดเร็วและไม่เสียค่าใช้จ่ายในการสืบค้นหาข้อมูล ในปัจจุบันมีระบบรวบรวมข้อมูลเกี่ยวกับข้อพิพาททางกฎหมาย ความขัดแย้งที่เกี่ยวกับสิทธิหรือคำร้องเรียน รวมไปถึงคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยของศาลฎีกาที่เกี่ยวข้องกับที่ดิน สำหรับผู้ที่สนใจ ซึ่งตัวระบบ ฯ สามารถค้นคืนคำพิพากษาเกี่ยวกับที่ดินได้ด้วยวิธีการค้นหาในลักษณะแบบเฉพาะเจาะจง (Exact Matching) โดยสามารถใช้เครื่องหมายตรรกะ (Logical Operators) เพื่อระบุเงื่อนไขในการสืบค้นโดยละเอียด อย่างไรก็ดีระบบ ฯ นั้นจะไม่สามารถค้นหาแบบการจับคู่ความหมาย (Semantic Matching) หรือใช้คำที่มีความหมายคล้ายคลึงกันได้ ส่งผลให้การใช้คำค้นหานั้นต้องใช้คำที่ถูกต้องตรงตามข้อความที่คำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกานั้นมี ทำให้ผู้ใช้งานมีโอกาสได้ผลลัพธ์ที่ไม่ตรงตามที่ต้องการ อีกทั้งการ

ค้นหาด้วยการผสมผสานคำค้นกับเครื่องหมายตรรกะ เพื่อให้ได้ผลลัพธ์ตามที่ต้องการยังเป็นเรื่องที่ยากสำหรับบุคคลทั่วไป

สารนิพนธ์นี้มีแนวคิดที่จะรวบรวมและจัดทำชุดข้อมูลคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาที่เกี่ยวข้องกับที่ดิน โดยใช้เทคนิคการเรียนรู้ของเครื่องเพื่อคัดแยกประเภทข้อความคำค้นหาจากชุดข้อมูลที่จัดทำขึ้นออกเป็นแต่ละประเภทกฎหมายที่เกี่ยวข้อง (Text Classification) แล้วนำคำค้นหานั้นไปค้นคืนคำพิพากษาที่เกี่ยวข้องกับประเภทของกฎหมายนั้น ๆ ด้วยการใช้เทคนิคการค้นหาแบบการจับคู่ความหมาย (Semantic Matching) เพื่อเปรียบเทียบว่าการเพิ่มเทคนิคการคัดแยกประเภทข้อความคำค้นหาจะส่งผลให้ความเร็วและประสิทธิภาพของการค้นคืนดีกว่าการที่ใช้การค้นคืนด้วยเทคนิคการค้นหาแบบการจับคู่ความหมายเพียงอย่างเดียวหรือไม่

1.2 สมมติฐานของงานวิจัย

การค้นคืนด้วยการใช้เทคนิคการค้นหาแบบการจับคู่ความหมาย (Semantic Matching) โดยเพิ่มเทคนิคการคัดแยกประเภทของคำค้นหาโดยเรียนรู้จากชุดข้อมูลคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาของกฎหมายที่เกี่ยวข้องกับที่ดินที่จัดทำขึ้น สามารถช่วยให้ความเร็วและประสิทธิภาพของการค้นคืนดีกว่าการค้นคืนด้วยการใช้เทคนิคการค้นหาแบบการจับคู่ความหมายเพียงอย่างเดียว

1.3 วัตถุประสงค์ของงานวิจัย

1. เพื่อเปรียบเทียบเทคนิคในการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน
2. เพื่อศึกษาและประยุกต์ใช้เทคนิคการคัดแยกประเภทของข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน
3. เพื่อจัดทำชุดข้อมูลข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน
4. เพื่อเสนอเทคนิคระบบค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน ให้มีความถูกต้องและแม่นยำมากขึ้น

1.4 ขอบเขตของงานวิจัย

1. งานวิจัยนี้ใช้ชุดข้อมูลข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน โดยเป็นชุดข้อมูลที่เกี่ยวข้องกับคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาผู้จัดทำได้รวบรวมมาจากแหล่งต่าง ๆ เช่น เอกสารราชการ หนังสือ ไฟล์ดิจิทัล และเว็บไซต์ deka.supremecourt.or.th ที่เป็นเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา และจัดเก็บไว้ในรูปแบบไฟล์ฐานข้อมูล (SQLite)
2. งานวิจัยนี้จะเปรียบเทียบความเร็วและประสิทธิภาพการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดินระหว่าง เทคนิคการคัดแยกประเภทกฎหมายจากคำค้นหาด้วย FastText รวมกับเทคนิคการค้นคืนด้วย roBERTa กับการใช้เทคนิคการค้นคืนด้วย roBERTa เพียงอย่างเดียว

1.5 ข้อจำกัดของงานวิจัย

1. เนื่องจากชุดข้อมูลที่นำมาใช้นั้นเป็นข้อมูลที่มาจากหลากหลายแหล่งที่มาและหลากหลายรูปแบบ ทำให้ใช้วิธีการในการนำเข้าข้อมูลหลากหลายรูปแบบทั้งการดึงข้อมูลจากเว็บไซต์ (Web Scraping) การแปลงข้อมูล PDF ด้วยวิธี OCR และการพิมพ์ข้อมูลจากหนังสือ อีกทั้งการจัดกลุ่มของชุดข้อมูลนี้จัดทำขึ้นโดยผู้วิจัยเอง
2. ชุดข้อมูลนี้ไม่มีเนื้อหาที่เป็นคำสั่ง ระเบียบ กฎกระทรวง ประมวลกฎหมายที่ดิน และพระราชบัญญัติที่เกี่ยวข้องรวมอยู่ด้วย ชุดข้อมูลนี้จะมีข้อมูลคำพิพากษา ประเด็นปัญหาความเห็นกรมที่ดินและถาม-ตอบเท่านั้น

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถนำแนวคิดการเพิ่มประสิทธิภาพการค้นคืนที่นำเสนอ ไปประยุกต์ใช้กับชุดข้อมูลข้อความทางกฎหมายในประเทศไทยที่เกี่ยวข้องกับเรื่องอื่น ๆ ได้
2. สามารถนำผลของงานวิจัยนี้ไปประยุกต์ใช้กับระบบค้นคืนของหน่วยงานต่าง ๆ ที่เกี่ยวข้องกับกฎหมายที่ดิน เพื่อให้บริการในการค้นคืนข้อความกฎหมายที่เกี่ยวข้องกับที่ดินให้กับประชาชนได้

1.7 รายละเอียดสารนิพนธ์

สารนิพนธ์ฉบับนี้ประกอบด้วย 5 บท ได้แก่ บทนำ วรรณกรรมและงานวิจัยที่เกี่ยวข้อง วิธีการวิจัย ผลการวิจัยและอภิปรายผล และสรุปผลการวิจัยและข้อเสนอแนะ

บทที่ 1 บทนำ เป็นบทที่กล่าวถึงที่มาและความสำคัญของสารนิพนธ์ฉบับนี้

บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้องเป็นบทที่กล่าวถึง ทฤษฎีและองค์ความรู้ที่เกี่ยวข้องกับงานวิจัย

บทที่ 3 วิธีการวิจัย เป็นบทที่กล่าวถึงขั้นตอนและวิธีดำเนินการในการวิจัย

บทที่ 4 ผลการวิจัยและอภิปรายผล เป็นบทที่กล่าวถึงผลลัพธ์ของงานวิจัยและวิจารณ์ผลของงานวิจัย

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ บทที่กล่าวถึงในส่วนของการ สรุปผลงานวิจัยและนำเสนอข้อเสนอแนะ

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 กฎหมายที่เกี่ยวกับข้องกับที่ดิน

กฎหมายหลักที่เกี่ยวข้องกับการคุ้มครองสิทธิในที่ดินของประชาชน จัดการที่ดินของรัฐ การรังวัดทำแผนที่ การออกหนังสือแสดงสิทธิในที่ดิน การให้บริการจดทะเบียนสิทธิและนิติกรรมเกี่ยวกับอสังหาริมทรัพย์ การส่งเสริมธุรกิจอสังหาริมทรัพย์

ภาพที่ 2.1
ลำดับศักดิ์ของกฎหมาย

2.1.1 พระราชบัญญัติให้ใช้ประมวลกฎหมายที่ดิน พ.ศ. 2497

เป็นกฎหมายที่ว่าด้วยการจัดสรรที่ดิน กฎหมายว่าด้วยอาคารชุด กฎหมายว่าด้วยช่างรังวัดเอกชน กฎหมายว่าด้วยการเช่าอสังหาริมทรัพย์เพื่อพาณิชยกรรมและอุตสาหกรรม และกฎหมายอื่น ๆ ที่เกี่ยวข้อง

2.1.2 ประมวลกฎหมายที่ดิน

เป็นกฎหมายที่บัญญัติขึ้นเพื่อจัดการเกี่ยวกับที่ดิน โดยเริ่มใช้บังคับใช้ครั้งแรก พ.ศ. 2497 มีการแก้ไขเพิ่มเติมหลายครั้ง ฉบับล่าสุดคือ พ.ศ. 2562

2.1.3 กฎกระทรวง ออกตามความในพระราชบัญญัติให้ใช้ประมวลกฎหมายที่ดิน พ.ศ. 2497

เป็นกฎหมายรองที่ออกตามความในพระราชบัญญัติให้ใช้ประมวลกฎหมายที่ดิน พ.ศ. 2497 มีหน้าที่หลักในการขยายความและกำหนดรายละเอียดเพิ่มเติมเกี่ยวกับประมวลกฎหมายที่ดิน เพื่อให้การบังคับใช้กฎหมายเป็นไปอย่างมีประสิทธิภาพและสอดคล้องกับสถานการณ์ปัจจุบัน

2.1.4 ระเบียบคณะกรรมการจัดที่ดินแห่งชาติ

เป็นระเบียบว่าด้วยการจัดที่ดินเพื่อประชาชน

2.1.5 พระราชบัญญัติอาคารชุด พ.ศ. 2522

เป็นกฎหมายที่บัญญัติเกี่ยวกับอาคารชุด

2.1.6 กฎกระทรวง ออกตามความในพระราชบัญญัติอาคารชุด พ.ศ. 2522

เกี่ยวข้องกับการกำหนดรายละเอียดเกี่ยวกับวิธีการและเงื่อนไขต่าง ๆ ในการจัดการกับอาคารชุด เช่น การจดทะเบียน การออกหนังสือกรรมสิทธิ์ การจดทะเบียนนิติบุคคล ค่าธรรมเนียม และค่าใช้จ่ายต่าง ๆ

2.1.7 พระราชบัญญัติการจัดสรรที่ดิน พ.ศ. 2543

เป็นกฎหมายที่ว่าด้วยการควบคุมการจัดสรรที่ดินให้เป็นไปอย่างมีระเบียบ

2.1.8 กฎกระทรวง ออกตามพระราชบัญญัติจัดสรรที่ดิน พ.ศ. 2543

เป็นกฎหมายที่ว่าด้วยการกำหนดรายละเอียดเกี่ยวกับวิธีการและเงื่อนไขต่าง ๆ ในการจัดการกับการจัดสรรที่ดิน เช่น การขออนุญาต การออกใบอนุญาต การโอนใบอนุญาต การเปลี่ยนแปลงโครงการ การแบ่งแปลงย่อย การรวมแปลง การยกเลิก การเพิกถอน ฯลฯ

2.1.9 ระเบียบคณะกรรมการจัดสรรที่ดินกลาง

เป็นกฎระเบียบที่ออกโดยคณะกรรมการจัดสรรที่ดินกลาง มีวัตถุประสงค์เพื่อกำหนดวิธีการและเงื่อนไขต่าง ๆ เกี่ยวกับการจัดสรรที่ดิน เพื่อคุ้มครองผู้ซื้อที่ดินจัดสรรและควบคุมการจัดสรรที่ดินให้เป็นไปอย่างมีระเบียบ

2.1.10 กฎหมายแพ่งและพาณิชย์

เป็นกฎหมายที่เกี่ยวข้องกับการดำเนินกิจการทางพาณิชย์ในธุรกิจทั่วไป ซึ่งมีวัตถุประสงค์เพื่อกำหนดกฎและเงื่อนไขในการทำธุรกิจเพื่อให้มีความยุติธรรมและเป็นไปตามหลักการของระบบกฎหมายที่เกี่ยวข้อง

2.2 เทคนิคที่เกี่ยวข้อง

2.2.1 การคัดแยกประเภทของข้อความ (Text Classification)

เทคนิค FastText ถูกพัฒนาและเผยแพร่โดยทีมงานของ Meta Open Source (Facebook Open Source) นั้นความสามารถนอกเหนือจากทำกระบวนการแทนคำแล้ว ยังมีอีกหนึ่งความสามารถสืบเนื่องมากจากกระบวนการก่อนหน้าคือ การคัดแยกประเภทข้อความ (Text Classfication) ซึ่งเทคนิคนี้นิยมใช้ในการแก้ไขปัญหาหลาย ๆ อย่าง เช่น การวิเคราะห์ความรู้สึก (Sentiment Analysis) หรือการตรวจจับสแปม (Spam Detection) เป็นต้น โดยเป้าหมายของการคัดแยกประเภทของข้อความนั้นต้องการแปะป้าย (Labeled) หรือจัดประเภทหมวดหมู่ อย่างน้อยหนึ่งประเภทหรือหลากหลายประเภท ขึ้นอยู่กับชุดข้อมูลนั้น ๆ ซึ่งในงานวิจัยนี้เลือกใช้เทคนิคข้างต้นในการคัดแยกประเภทของคำค้นข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน

2.2.1.1 การแทนคำ (Word Representation)

เป็นวิธียอดนิยมในปัจจุบันสำหรับงานที่เกี่ยวของกับการประมวลผลคำ (NLP) เป็นการนำเสนอการแทนคำ (Word) ให้อยู่ในรูปแบบของเวกเตอร์ (Vector) การแทนคำแบบนี้ทำให้สามารถซ่อนรายละเอียดหรือคุณลักษณะเฉพาะของภาษาเอาไว้ในรูปแบบของตัวเลขได้ ทั้งในเรื่องของการหาคำคล้ายกัน (Analogies) หรือคำที่มีความหมายคล้ายกัน (Sementic) โดยในงานวิจัยนี้ใช้เทคนิค FastText ซึ่งสามารถทำ Word Embedding ได้ มีหลักการแนวคิดมาจาก Word2Vec เป็นการเข้ารหัส (Encode) ของคำแต่ละคำโดยใช้เทคนิค CBOW (Continuous Bag of Word) หรือใช้เทคนิค Skipgram แตกต่างกับการเรียนรู้ Word Vector โดยตรง และงานวิจัยนี้ใช้ Pre-Trained Word Embedding ที่เรียนรู้จากชุดข้อมูลภาษาไทยของ Wikipedia บนเทคนิคโมเดล Skipgram

2.2.2 การค้นคืนข้อมูล (Information Retrieval)

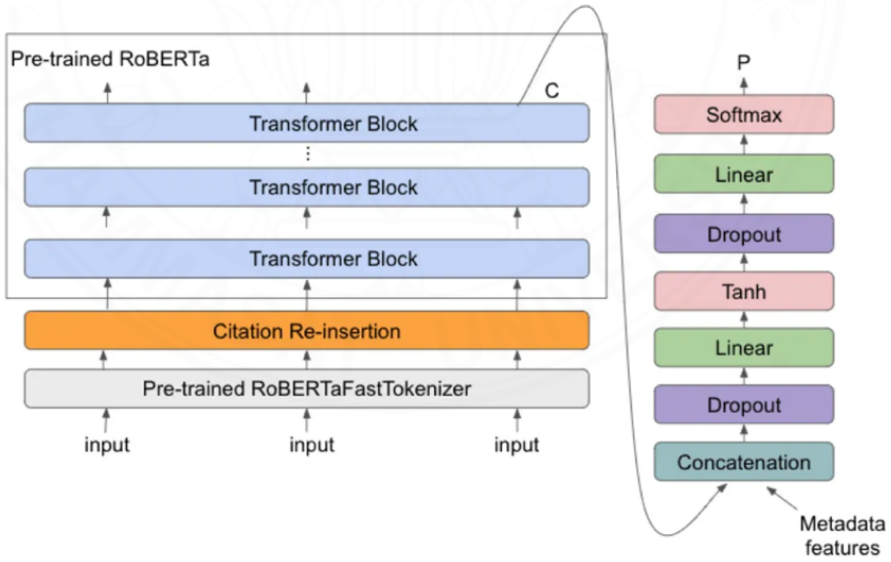
เทคนิคที่สามารถทำให้คอมพิวเตอร์สามารถเปรียบเทียบความคล้ายคลึงกันระหว่างข้อความได้นั้น มีเทคนิคมากมายที่ใช้ในการค้นคืนข้อมูลแต่ละเทคนิคมีข้อดีข้อเสียแตกต่างกันขึ้นอยู่กับชุดข้อมูลและการเตรียมการทำความสะอาดข้อมูล โดยในงานวิจัยนี้ใช้เทคนิค 2 วิธีแล้วนำมาเปรียบเทียบความเร็วและประสิทธิภาพกัน ซึ่งประกอบไปด้วย 2 เทคนิค ดังนี้

2.2.2.1 Robustly Optimized BERT Pretraining Approach (roBERTa)

เป็นโมเดลทางภาษา (Language Model) เป็นโมเดลที่ใช้สถาปัตยกรรมแบบ Transformer ซึ่งมีพื้นฐานมาจากเทคนิค Bidirectional Encoder Representations from Transformers (BERT) โดยได้ปรับแก้เอาขั้นตอน Next-Sentence Prediction ออกและเรียนรู้กับชุดข้อมูลที่มีขนาดใหญ่มากกว่าส่งผลให้เทคนิค roBERTa นั้นมีประสิทธิภาพสูงกว่าเทคนิค BERT อีกทั้งเทคนิค roBERTa ได้ถูกนำมาพัฒนา Finetune ให้สามารถใช้ได้กับหลากหลายภาษารวมถึงภาษาไทย

ภาพที่ 2.2

สถาปัตยกรรมของเทคนิค roBERTa



2.2.2.2 Cosine Similarity

การเปรียบเทียบความคล้ายคลึงกันของเวกเตอร์สองเวกเตอร์ด้วยเทคนิค Cosine Similarity เป็นวิธีการคำนวณ Cosine ของการทำมุมระหว่างเวกเตอร์ 2 ตัว ดังสมการ

$$\cos(\theta) = \frac{A \cdot B}{||A||||B||}$$

โดย A และ B คือเวกเตอร์ที่ต้องการเปรียบเทียบ

· คือการ Dot Product ของเวกเตอร์

$||A||$และ $||B||$ คือการ Normalized เวกเตอร์

ค่าที่ได้จากการคำนวณหาความคล้ายคลึงกันของเวกเตอร์ด้วยวิธี Cosine similarity จะอยู่ในช่วง -1 ถึง 1 หากค่าที่ได้เท่ากับ 1 หมายความว่า A และ B มีความคล้ายคลึงกันมากที่สุด

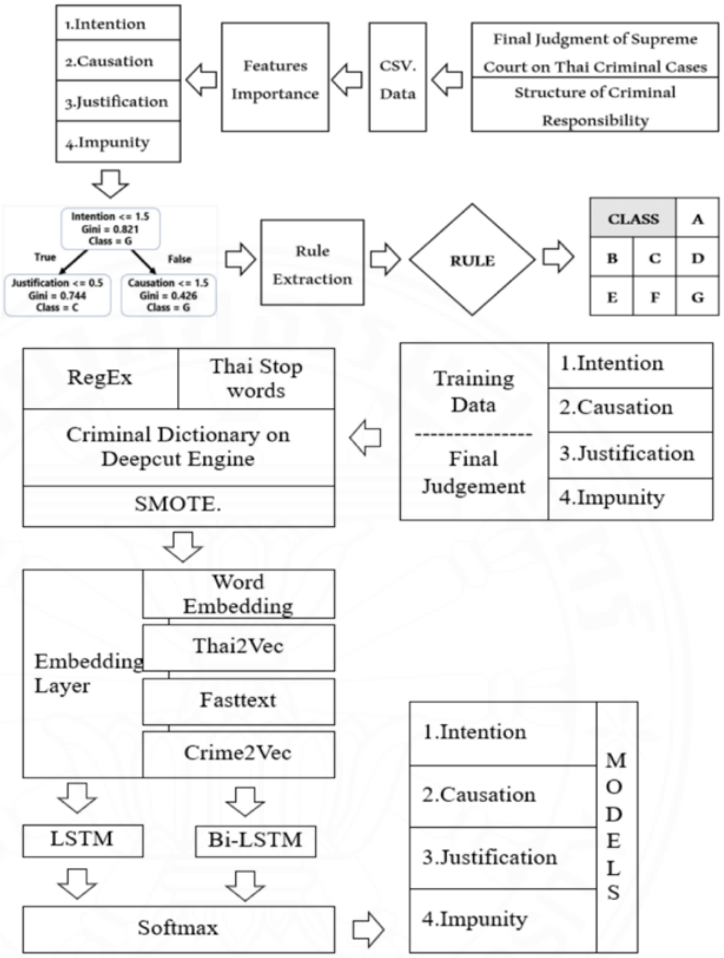
2.3 งานวิจัยที่เกี่ยวข้อง

ในการทบทวนวรรณกรรมของงานวิจัยที่เกี่ยวข้องนั้น เนื่องจากชุดข้อมูลข้อความทางกฎหมายที่ใช้ในงานวิจัยนี้เป็นชุดข้อมูลที่รวบรวมและจัดทำขึ้นมาใหม่ โดยเป็นชุดข้อมูลของคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาที่เกี่ยวข้องกับที่ดินซึ่งเป็นชุดข้อมูลภาษาไทยและยังไม่เคยมีงานวิจัยที่ทำเกี่ยวกับชุดข้อมูลดังกล่าว ดังนั้นจึงเลือกศึกษางานวิจัยที่เกี่ยวข้องกับการรวบรวมและจัดเตรียมข้อมูล เทคนิคการคัดแยกประเภทของชุดข้อมูล (Text Classification) และเทคนิคการค้นคืนชุดข้อมูล (Information Retrieval) ที่มีลักษณะเหมือนหรือคล้ายคลึงกับชุดข้อมูลข้อความทางกฎหมายข้างต้นทั้งภาษาไทยและภาษาต่างประเทศดังนี้

ในงานวิจัยของ Weerayut และ Sukree ในปี 2020 ได้ทำการศึกษาวิเคราะห์ภาษาธรรมชาติของข้อความที่เกี่ยวข้องกับกฎหมายอาญาในประเทศไทย โดยรวบรวมชุดข้อมูลคำพิพากษาศาลฎีกาขึ้นมา ด้วยการดึงข้อมูลจากเว็บไซต์ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา โดยใช้ไลบรารี Beautiful Soap4 ร่วมกับ Regular Expression โดยจะเลือกเฉพาะกฎหมายอาญา มาตราที่ 288 และ 289 ระหว่างปี พ.ศ. 2500 ถึง พ.ศ. 2560 มีข้อมูลประกอบไปด้วยเลขคำที่พิพากษา ข้อเท็จจริง ปัญหาทางกฎหมายและคำตัดสิน และจัดเก็บในรูปแบบไฟล์ csv

ภาพที่ 2.3

สถาปัตยกรรมที่งานวิจัยของ Weerayut และ Sukree (2020) นำเสนอ



เนื่องจากข้อมูลในส่วนของมาตรา 288 มี 2,078 ฎีกา และมาตรา 289 มี 620 ฎีกา รวมจำนวนข้อมูลคือ 2,698 ฎีกา แล้วนำมาทำความสะอาดข้อมูลเพื่อให้ประสิทธิภาพเพิ่มขึ้น โดยจะใช้การตัดคำฟุ่มเฟือย แล้วใช้ไลบรารีของ PyThaiNLP ด้วยตัวคัดคำ Deepcut ร่วมกับเทคนิค Longest Matching จากชุดข้อมูลข้างต้นพบว่าจำนวนของชุดข้อมูลไม่สมดุลกัน (Over-Sampling) จึงแก้ปัญหาด้วยวิธี Synthetic Minority Over-Sampling Technique (SMOTE) ต่อมาในส่วนของการทำ Word Embedding จะใช้เทคนิค Skip-Gram เพราะว่าเหมาะสมกับชุดข้อมูลที่มีจำนวนไม่มาก เพื่อช่วยลดระยะเวลาในการเรียนรู้จึงใช้ Pre-Trained Word Embedding โดยใช้ทั้งหมด 3 แบบ คือ Thai2Vec, Fasttext และ Crime2Vec หลังจากนั้นใช้เทคนิค LSTM และ BiLSTM ร่วมกับ Word Embedding ที่กล่าวมาข้างต้นคัดแยกประเภทของลักษณะของการกระทำหรือเหตุการณ์ที่เกิดขึ้นว่าตรงกับข้อเท็จจริงใด โดยได้ผลลัพธ์ดังตารางที่ 2.1

ตารางที่ 2.1

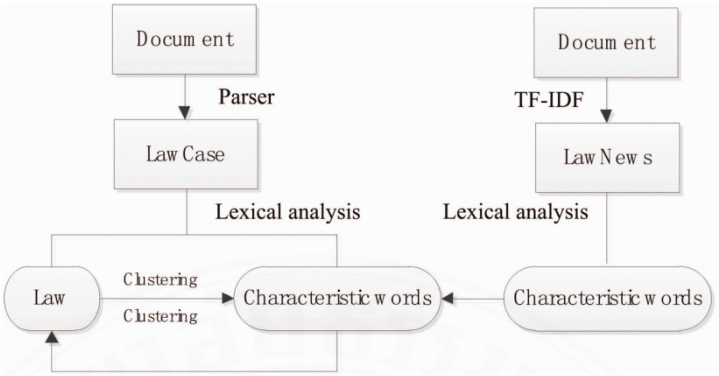
ประสิทธิภาพของเทคนิคที่ใช้จากงานวิจัยของ Weerayut และ Sukree (2020)

Features	Models							
	Intention		*Causation*		*Impunity*		*Justification*	
	A	*F*	*A*	*F*	*A*	*F*	*A*	*F*
LSTM	78.98	78.88	62.02	62.40	75.86	75.98	80.70	80.68
LSTM+ Fasttext	78.98	78.47	75.97	75.50	73.28	72.63	94.74	94.71
LSTM+ Thai2Vec	68.79	68.75	62.02	60.69	33.62	33.40	78.95	78.94
LSTM+ Crime2Vec	75.16	75.01	71.32	71.32	68.97	68.79	92.98	92.87
BiLSTM	81.53	81.09	75.97	75.79	76.72	76.48	91.23	91.13
BiLSTM+ Fasttext	82.17	82.08	77.52	76.91	78.45	78.46	98.25	98.24
BiLSTM+ Thai2Vec	66.88	67.36	74.42	71.30	54.31	52.01	87.72	87.58
BiLSTM+ Crime2Vec	75.80	75.79	76.74	75.81	76.72	76.33	98.25	98.24

ในส่วนของงานวิจัยที่เกี่ยวข้องกับการคัดแยกประเภทข้อความ (Text Classification) ด้วยวิธีต่าง ๆ นั้นพบว่างานวิจัยของ Zhonghao Li (2019) ได้ทำการศึกษาการคัดแยกประเภทของข้อความทางกฎหมายที่เป็นภาษาอังกฤษ โดยใช้เทคนิคเกี่ยวกับ Text Data Mining ซึ่งให้ผลลัพธ์ที่ดีในการแยกคุณลักษณะเฉพาะจากข้อความทางกฎหมาย รวมไปถึงการคัดแยกประเภทตามมาตราและคำพิพากษา โดยใช้เทคนิคการแทนข้อความ (Text Representation) ด้วยวิธี TF-IDF แล้วพัฒนาเทคนิคร่วมกับ Chi-Square Statistics (CHI) ทำให้สามารถเลือกและระบุลักษณะของคำที่จะใช้ให้กับเทคนิค TF-IDF ได้ ซึ่งการปรับปรุงนี้จะช่วยให้คำที่เด่นมีน้ำหนักมากขึ้นและคำเหล่านี้จะช่วยให้การคัดแยกประเภทนั้นมีความแม่นยำมากยิ่งขึ้น

ภาพที่ 2.4

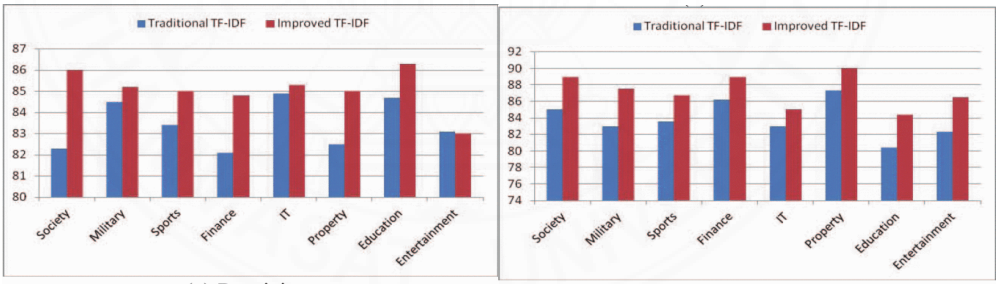
สถาปัตยกรรมที่งานวิจัยของ Zhonghao Li (2019) นำเสนอ



โดยเทคนิคที่ใช้ในการคัดแยกประเภทนั้นจะใช้ SVM เมื่อทำการเปรียบเทียบการคัดแยกประเภทของข้อความทางกฎหมายที่เป็นภาษาอังกฤษ ระหว่าง TF-IDF กับ TF-IDF ที่ได้ปรับปรุงเทคนิคใหม่ พบว่าความถูกต้อง (Precision) และ Recall ของ TF-IDF ที่ปรับปรุงใหม่นั้นสูงกว่า TF-IDF แบบเดิมประมาณ 3 เปอร์เซ็นต์ ดังภาพที่ 2.4

ภาพที่ 2.5

เปรียบเทียบผลลัพธ์ระหว่าง TF-IDF เดิม กับ TF-IDF ที่ได้ปรับปรุงใหม่

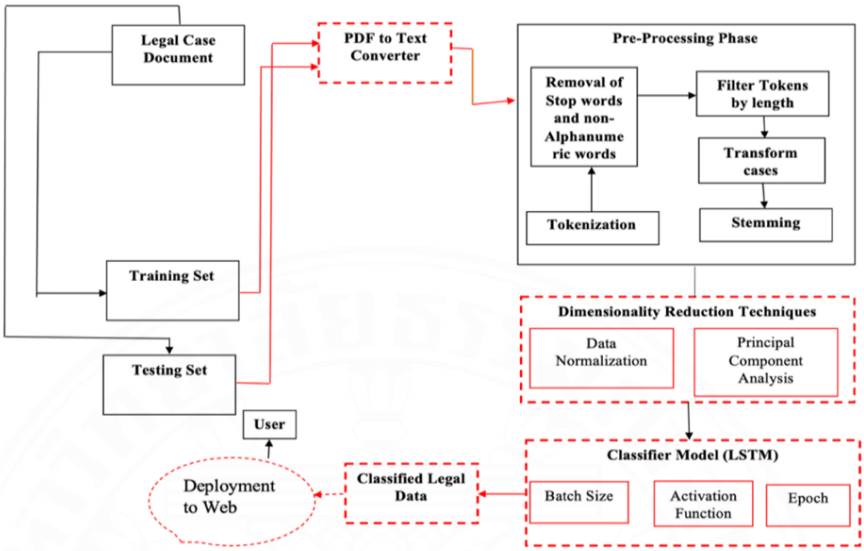


(a) Precision rate (b) Recall rate

อีกงานวิจัยของ Stow และคณะ (2023) นำเสนอเกี่ยวกับการพัฒนาประสิทธิภาพสำหรับการคัดแยกประเภทข้อความทางกฎหมาย โดยใช้เทคนิค Principal Component Analysis (PCA) กับ Long Short-Term Memory (LSTM) ในการเรียนรู้ชุดข้อมูลและคัดแยกประเภท โดยได้ผลลัพธ์ความแม่นยำ (Accuracy) การคัดแยกประเภทข้อความทางกฎหมายกับชุดข้อมูลที่ไม่เคยเห็นมาก่อน (Never-Before-Seen) 99.58 เปอร์เซ็นต์ ซึ่งชุดข้อมูลของงานวิจัยนี้มาจากการรวบรวมไฟล์ดิจิทัลประเภท PDF และแปลงไฟล์ดิจิทัลมาเป็น Text โดยใช้ระบบแปลง PDF-To-Text ซึ่งมีข้อมูลเอกสารเกี่ยวกับข้อความทางกฎหมายทั้งหมด 1,500 เอกสาร แบ่งข้อมูลออกเป็น 6 ประเภท

ภาพที่ 2.6

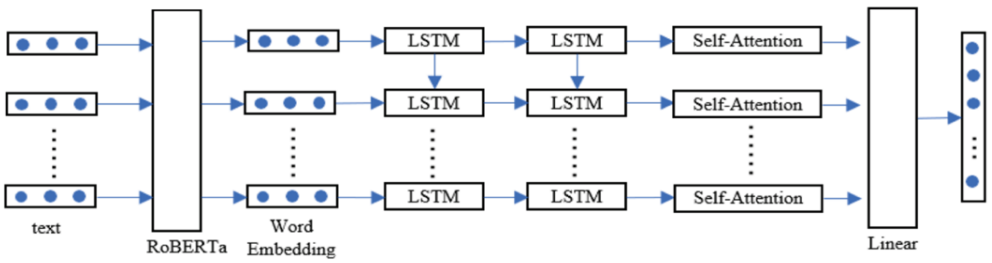
สถาปัตยกรรมที่งานวิจัยของ Stow และคณะ (2023) นำเสนอ



อีกหนึ่งงานวิจัยที่เกี่ยวข้องกับการคัดแยกประเภทข้อความทางกฎหมาย (Text Classification) คืองานวิจัยของ Yang และคณะ (2022) ได้ศึกษาและนำเสนอเทคนิค RoBERTa Text-RNN Base On Self-Attention Mechanism (RTR-SA) สามารถแก้ไขปัญหาในเรื่องของความไม่สมดุลของชุดข้อมูลได้ โดยได้ศึกษาและเปรียบเทียบกับชุดข้อมูลที่มีชื่อว่า CAIL2018-Small Dataset ในชุดข้อมูลนี้ประกอบไปด้วย ข้อกล่าวหาและมาตราของกฎหมายที่เกี่ยวข้อง ซึ่งการทดลองใช้มิติของ Word Vectors ที่ 768 ความยาวของประโยคที่มากที่สุดคือ 330 Learning Rate 1e-5 และใช้ Adam Optimizer เป็นตัวช่วยในการปรับน้ำหนักของโมเดล โดยผลลัพธ์ที่ได้จากการเปรียบเทียบระหว่างเทคนิค CNN, LSTM, Zhanghu และ RTR-SA พบว่าเทคนิค RTS-SA ให้ผลลัพธ์ที่ดีที่สุด

ภาพที่ 2.7

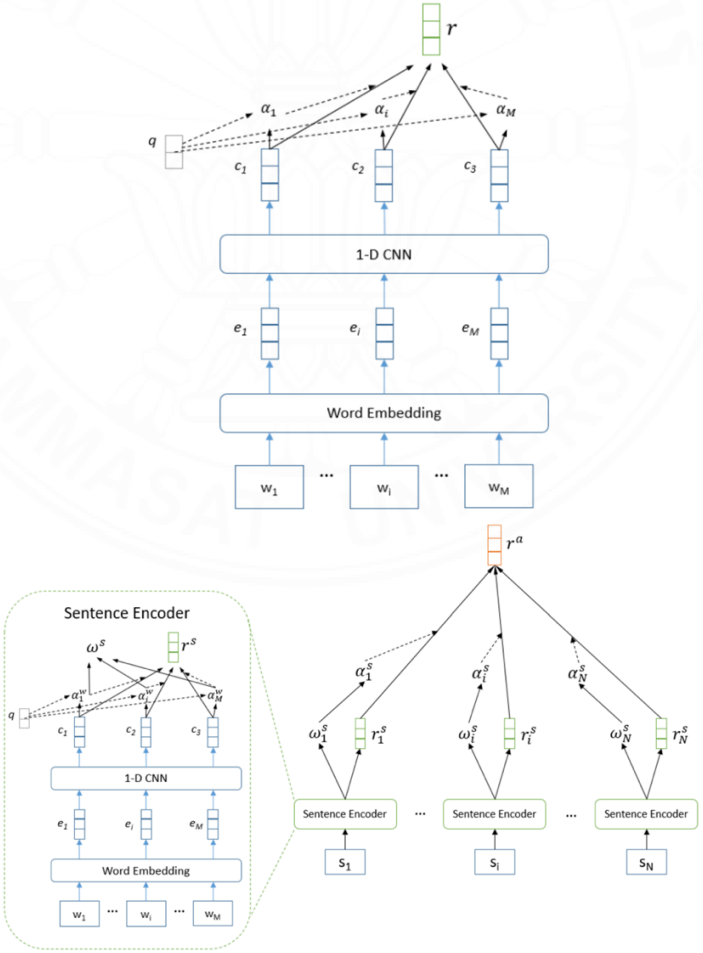
สถาปัตยกรรมที่งานวิจัยของ Yang และคณะ (2022) นำเสนอ



ในส่วนของงานวิจัยที่เกี่ยวข้องกับการค้นคืน (Information Retrieval) ด้วยวิธีต่าง ๆ นั้นพบว่ามีงานวิจัยของ Kein Vu Tran และคณะ (2020) ได้ทำการรวบรวมชุดข้อมูลการถามตอบเกี่ยวกับกฎหมายของประเทศเวียดนาม จำนวน 5,922 รายการ และได้ทำการทดลองเปรียบเทียบการค้นคืนคำตอบ โดยจะเน้นการค้นคืนมาตราของกฎหมายที่เกี่ยวข้องกับคำถามนั้น ๆ โดยจะให้ความสำคัญกับการแทนคำ (Word Representation) เพราะว่าเป็นสิ่งที่มีผลต่อความแม่นยำ (Accuracy) ซึ่งใช้เทคนิค ElasticSearch 2 แบบ คือ BM25 และ TF-IDF, Birch 2 แบบ คือ 256-Frist Words กับ Title และเทคนิคที่นำเสนอในงานวิจัย 2 แบบ คือ Flat Architecture รวมกับ Sentence Encoder และ Hierarchical Architecture รวมกับ Paragraph Encoder ในการทดสอบเปรียบเทียบเทคนิค

ภาพที่ 2.8

สถาปัตยกรรมที่งานวิจัยของ Kein Vu Tran และคณะ (2020) นำเสนอ



โดยวิธีที่นำเสนอนั้นให้ผลลัพธ์ที่ดีที่สุด ดังตารางที่ 2.2 จะเห็นได้ว่าการเลือกที่ใช้เทคนิคที่เหมาะสมกับลักษณะเฉพาะกับข้อความทางกฎหมายนั้นทำให้ประสิทธิภาพที่ได้เหนือว่าเทคนิค ElasticSerch อย่างมากรวมถึงเหนือว่า BERT ด้วย

ตารางที่ 2.2

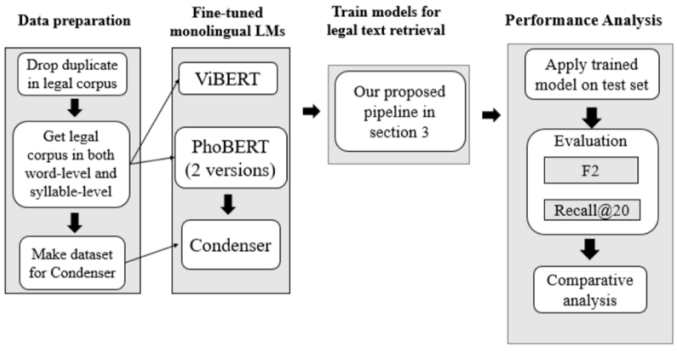
ประสิทธิภาพเทคนิคการค้นคืนที่ใช้จากงานวิจัยของ Kein Vu Tran และคณะ (2020)

System	Recall@20	NDCG@20
ElasticSearch (BM25)	0.357	0.334
ElasticSearch (TF-IDF)	0.478	0.351
Birch (256 first words)	0.763	0.542
Birch (title)	0.783	0.591
Our System I (1000,30,30)	0.798	0.641
Our System I (1000,50,50)	0.811	0.669
Our System II (1000,30,30)	0.801	0.665
Our System II (1000,50,50)	**0.825**	**0.688**

อีกงานวิจัยที่ต่อยอดจากงานวิจัยก่อนหน้านี้ คือการวิจัยของ Pham และคณะ (2022) ได้มีการใช้กระบวนการ Lexical Matching ร่วมกันกับ Semantic Searching เพื่อเพิ่มประสิทธิภาพในการค้นคืนข้อความทางกฎหมายของเวียดนาม งานวิจัยนี้ได้เลือกใช้เทคนิค BM25+ สำหรับกระบวนการ Lexical Matching และใช้เทคนิค Sentence Transformer สำหรับกระบวนการ Semantic Searching ซึ่งนำเสนอการเรียนรู้ของเทคนิค Sentence Transformer ด้วยวิธีการเรียนรู้สิ่งที่ตรงกันข้ามกับสิ่งที่ต้องการ โดยจะเรียนรู้ 3 รอบ แต่ละรอบจะเลือกชุดข้อมูลด้วย BM25+ แล้วดึง Top-K ที่มีคะแนนสูงสุดในรอบการเรียนรู้ก่อนหน้า ตัวอย่างแต่ละชุดข้อมูลจะมีจำนวน K+(จำนวนมาตราที่ถูกต้อง) คู่กับกันมาตราที่ไม่ตรงกับตัวอย่างชุดข้อมูล

ภาพที่ 2.9

สถาปัตยกรรมที่งานวิจัยของ Pham และคณะ (2022) นำเสนอ



โดยภาษาเวียดนามนั้นแตกต่างกับภาษาอังกฤษ งานวิจัยนี้จะเลือกใช้โมเดล Monolingual Pre-Trained ที่เรียนรู้กับชุดข้อมูลภาษาเวียดนามเข้ามาใช้ในงานวิจัยจำนวน 2 โมเดล คือ โมเดล PhoBERT Pre-Training Approach ที่มีพื้นฐานบนเทคนิค RoBERTa และโมเดล ViBERT เรียนรู้ต่อยอดมาจากเทคนิค mBERT ซึ่งผลลัพธ์ของการใช้กระบวนการ Lexical Matching ร่วมกันกับ Semantic Searching ให้ผลลัพธ์ที่ดีกว่าการใช้กระบวนการ Semantic Searching เพียงอย่างเดียวดังตารางที่ 2.3

ตารางที่ 2.3

ประสิทธิภาพเทคนิคการค้นคืนที่ใช้จากงานวิจัยของ Pham และคณะ (2022)

	sqrt(BM25_score)*cos_sim		BM25_score*cos_sim	
System	**Recall@20**	**F2**	**Recall@20**	**F2**
SPhoBERT-base(syl)	0.958	0.715	0.960	0.703
SConPBB(syl)	0.967	0.712	0.965	0.692
SPhoBERT-large(syl)	0.951	0.726	0.955	0.706
SConPBL(syl)	0.953	0.729	0.956	0.710
SViBERT(syl)	0.951	0.695	0.963	0.684
SPhoBERT-base(word)	0.963	0.725	0.964	0.706
SConPBB(word)	0.962	0.719	0.964	0.700
SPhoBERT-large(word)	0.967	**0.741**	**0.970**	0.721
SConPBL(word)	0.960	0.738	0.967	0.718
SViBERT(word)	0.948	0.683	0.950	0.661

อีกหนึ่งงานวิจัยที่เกี่ยวข้องกับการค้นคืน (Information Retrieval) คืองานวิจัยของ Rahman และคณะ (2022) ได้ทำการศึกษาและนำเสนอวิธีการเปรียบเทียบความคล้ายคลึงกันของประโยค (Sentence Similarity) โดยใช้เทคนิคต่าง ๆ 5 แบบ ได้แก่ n_similarity, sorted n_similarity, wmdistance, BERT-based model และ FastText ด้วยชุดข้อมูลคำพิพากษาศาลฎีกาประเทศอินเดีย โดยจะแบ่งเป็นชุดข้อมูลปัจจุบัน (Current Case) กับชุดข้อมูลที่มีมาก่อน (Prior Case) ซึ่งเทคนิค Sentence Similarity คู่กับ Sorted n_similarity ให้ผลลัพธ์ดีที่สุด

ภาพที่ 2.10

สถาปัตยกรรมที่งานวิจัยของ Rahman และคณะ (2022) นำเสนอ

จากการทบทวนวรรณกรรมผู้วิจัยตัดสินใจนำงานวิจัยของ Weerayut และ Sukree (2020) ในส่วนของการรวบรวมและสร้างชุดข้อมูล การทำความสะอาดชุดข้อมูล เทคนิคการแก้ไขปัญหาความสมดุลของชุดข้อมูลมาใช้

ส่วนวิธีการคัดแยกประเภทข้อความ (Text Classification) ด้วยการเทคนิค FastText เทคนิคนี้ยังไม่มีงานวิจัยใดเคยลองใช้มาก่อน และจากการค้นคว้าหาข้อมูลพบว่าเทคนิคนี้มีโมเดล Pre-Trained ที่เป็นภาษาไทย อีกทั้งยังสามารถเรียนรู้การคัดแยกประเภทตามการแปะป้าย (Labeled) ได้อีกด้วย

อีกหนึ่งวิธีการค้นคืน (Information Retrieval) นั้น งานวิจัยส่วนใหญ่ได้ผลลัพธ์การวัดประสิทธิภาพที่สูงกับเทคนิคที่มีพื้นฐานมาจาก roBERTa ผู้วิจัยจึงเลือกใช้เทคนิค roBERTa และเนื่องจากงานวิจัยของ Weerayut และ Sukree (2020) ที่ใช้ชุดข้อมูลคำพิพากษาศาลฎีกาของประเทศไทย ซึ่งเป็นชุดข้อมูลเดียวกันกับที่ผู้วิจัยรวบรวมขึ้นมา โดยงานวิจัยข้างต้นได้ประสิทธิภาพที่ดีกับเทคนิค Skip-Gram สำหรับ Word Embedding ด้วย Fasttext เนื่องจากเป็น Pre-Trained Word Embedding ที่ได้ค่าเฉลี่ยความแม่นยำ (Accuracy) มากที่สุด จึงนำเทคนิค FastText มาร่วมเปรียบเทียบประสิทธิภาพของการค้นคืนด้วย

ตารางที่ 2.4

ตารางเปรียบเทียบงานวิจัย

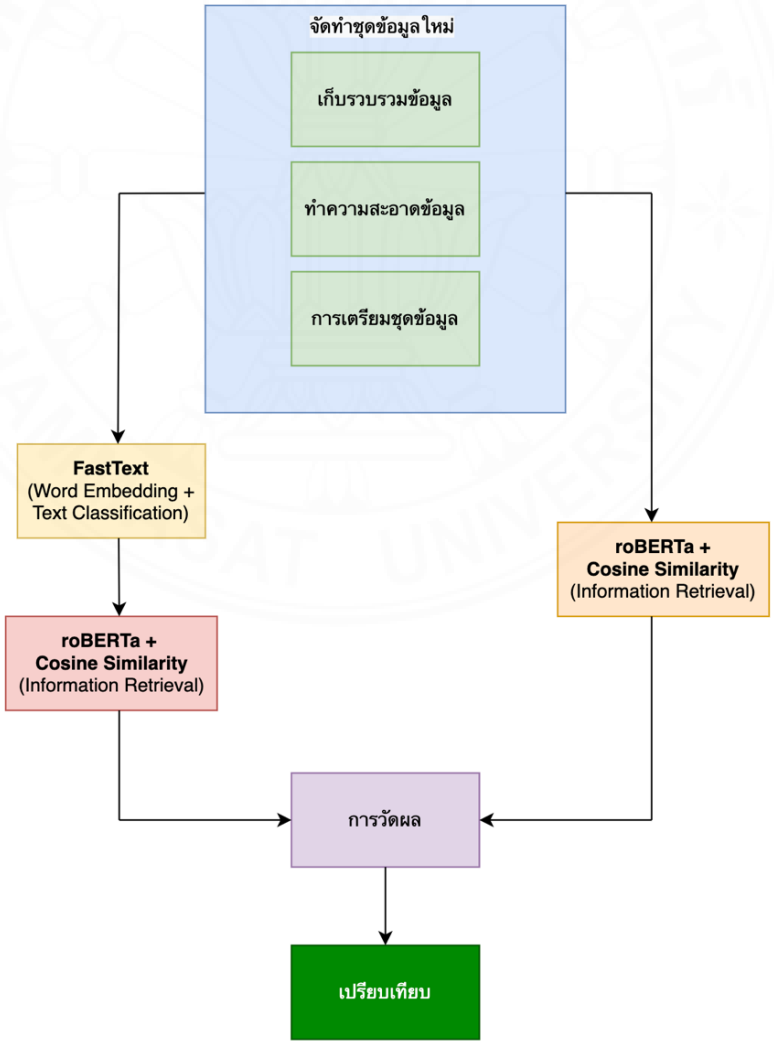
ชื่องานวิจัย	Cosine Similarity	TF-IDF	BM25	Birch	SVM	CNN	Base on BERT	Base on LSTM	Fasttext	Zhanghu	PCA	RTR-SA
An Analysis of Natural Language Text Relating to Thai Criminal Law								✓	✓			
A Classification Retrieval Approach for English Legal Texts		✓			✓							
An Improved Model for Legal CaseText Document Classification								✓			✓	
Text Classification of Judgement Documents Considering Sample Imbalance						✓		✓		✓		✓
Answering Legal Questions by Learning Neural Attentive Text Representation		✓	✓	✓		✓ Add-on Attention mechanism	✓					
Multi-stage Information Retrieval for Vietnamese Legal Texts	✓		✓				✓					
An Intelligent Approach Towards Legal Text-Documents Retrieval	✓						✓		✓			
งานวิจัยนี้	✓						✓		✓			

บทที่ 3
วิธีการวิจัย

3.1 ขั้นตอนของการวิจัย

บทนี้นำเสนอและอธิบายในส่วนของขั้นตอนการวิจัยเพื่อให้ผู้อ่านได้ทราบถึงวิธีที่ใช้ในการรวบรวมและจัดเตรียมชุดข้อมูล เทคนิคที่เลือกใช้ในขั้นตอนการคัดแยกประเภทของคำค้นหาและเทคนิคที่เลือกใช้ในการค้นคืนข้อมูล รวมถึงการประเมินผลการวิจัย

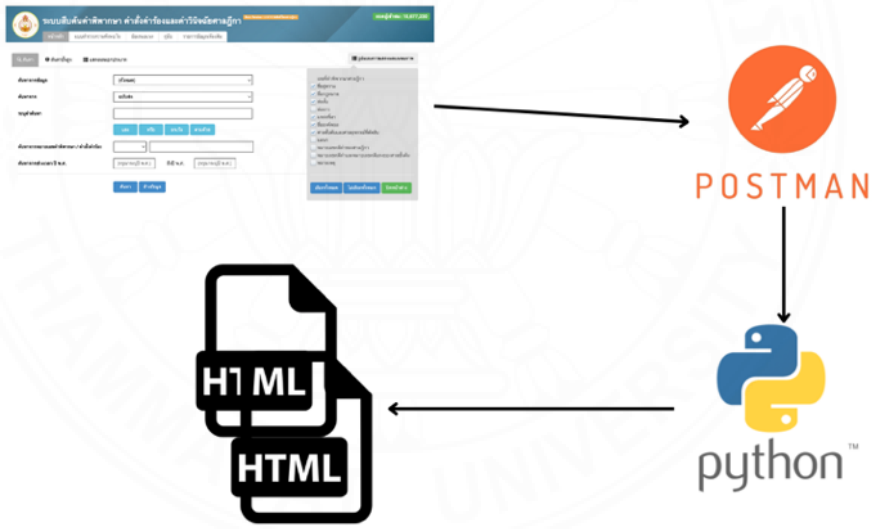
ภาพที่ 3.1
ภาพรวมขั้นตอนของการวิจัย



3.1.1 ขั้นตอนการเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้จะใช้ชุดข้อมูลที่รวบรวมและจัดทำขึ้นมาใหม่โดยจะเป็นการรวบรวมข้อมูลจากหนังสือ ไฟล์ดิจิทัล และเว็บไซต์ deka.supremecourt.or.th เป็นเว็บไซต์ที่รวบรวมคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา ซึ่งจะเลือกเฉพาะข้อมูลที่เกี่ยวข้องกับกฎหมายที่เกี่ยวข้องกับที่ดินเท่านั้น ซึ่งมีขั้นตอนการดึงข้อมูลจากเว็บไซต์ด้วยวิธี Web Scraping ด้วยเครื่องมือ Postman และบันทึกผลลัพธ์ไว้ในไฟล์ HTML ด้วยการเขียนสคริปต์โปรแกรมภาษาไพธอน ดังภาพที่ 3.2 ส่วนข้อมูลที่มาจากหนังสือและไฟล์ดิจิทัล เพื่อลดความผิดพลาดในการรวบรวมชุดข้อมูล ผู้วิจัยเลือกใช้วิธีการพิมพ์ข้อความตามเอกสาร จากนั้นนำมาจัดเก็บไว้ในไฟล์ Text โดยมีรูปแบบการจัดเก็บ ดังภาพที่ 3.3

ภาพที่ 3.2
ขั้นตอนการดึงข้อมูลจากเว็บไซต์



ภาพที่ 3.3

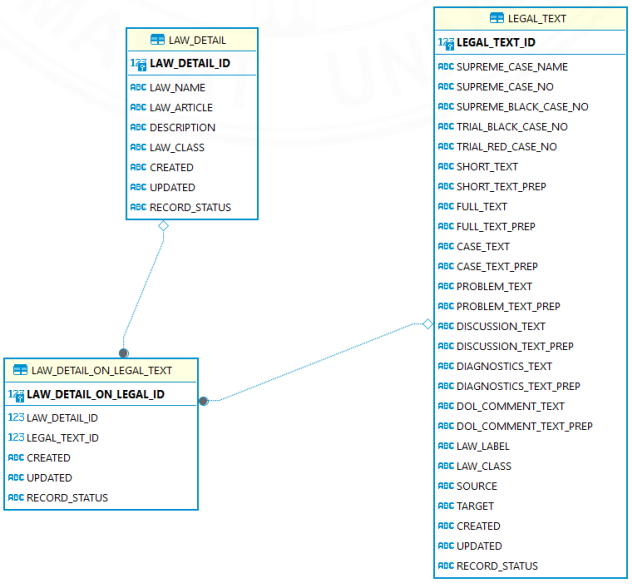
รูปแบบการจัดเก็บข้อมูลจากหนังสือหรือไฟล์ดิจิทัลไว้ในไฟล์ Text

-กรณี
การจดทะเบียนลงชื่อผู้จัดการมรดกในใบจอง
-ประเด็นข้อหารือ
นาง ก. ยื่นคำขอออกโฉนดที่ดินเฉพาะราย (โดยมิได้แจ้งการครอบครองที่ดิน) จังหวัดตรวจสอบแล้วปรากฏว่า ที่ดินที่ขอออกโฉนดที่ดินรายนี้เป็นที่ดินบางส่วนของที่ดินตามใบจองเลขที่.....
มีชื่อ นาย ข. บิดาของนาง ก. เป็นผู้ขอออกใบจอง ต่อมานาย ข. ถึงแก่กรรม นาง ค. ภรรยาของนาย ข. ซึ่งเป็นผู้จัดการมรดกตามคำสั่งศาล ได้จดทะเบียนลงชื่อผู้จัดการมรดกในใบจอง
และนำที่ดินบางส่วนตามใบจอง ดังกล่าว ไปขอออกโฉนดที่ดินบ้างแล้ว จึงเป็นปัญหาว่าการจดทะเบียนลงชื่อผู้จัดการมรดกในใบจองดังกล่าว เป็นการดำเนินการโดยชอบด้วยกฎหมายและระเบียบวิธีการหรือไม่
-ข้อกฎหมาย
=ประมวลกฎหมายที่ดิน มาตรา 1, 64, 82
-แนววินิจฉัย
=1.ตามมาตรา 82 แห่งประมวลกฎหมายที่ดิน บัญญัติว่า "ผู้ใดประสงค์จะขอจดทะเบียนลงชื่อผู้จัดการมรดกในหนังสือแสดงสิทธิในที่ดินให้ยื่นคำขอ...." การจดทะเบียนตามนัยมาตรา 82
ต้องเป็นการจดทะเบียนในหนังสือแสดงสิทธิในที่ดิน จึงมีปัญหาที่จะต้องพิจารณาว่า ใบจองเป็นหนังสือแสดงสิทธิในที่ดินหรือไม่ ซึ่งเมื่อพิจารณาตามประมวลกฎหมายที่ดินมาตรา 1 คำว่า "สิทธิในที่ดิน"
หมายความว่ากรรมสิทธิ์และให้หมายความรวมถึงสิทธิครอบครองด้วย ส่วนคำว่า "ใบจอง" หมายความว่า หนังสือแสดงการยอมให้เข้าครอบครองที่ดินชั่วคราว ใบจองก็น่าจะเป็นหนังสือแสดงสิทธิครอบครองในที่ดินด้วย
แต่อย่างไรก็ดีตามมาตรา 64 แห่งประมวลกฎหมายที่ดิน ได้บัญญัติว่า "ถ้าโฉนดที่ดิน ใบไต่สวน หนังสือรับรองการทำประโยชน์หรือใบจอง ฉบับสำนักงานที่ดินเป็นอันตราย ชำรุดสูญหาย ให้พนักงานเจ้าหน้าที่ ตามมาตรา 71
มีอำนาจเรียกหนังสือแสดงสิทธิในที่ดินดังกล่าวจากผู้มีสิทธิในที่ดินมาพิจารณาแล้วจัดทำขึ้นใหม่...."จึงพิจารณาได้ว่าใบจองย่อมเป็นหนังสือแสดงสิทธิในที่ดินเช่นเดียวกับโฉนดที่ดิน ใบไต่สวนและหนังสือรับรองการทำประโยชน์
การจดทะเบียนลงชื่อผู้จัดการมรดกในใบจองจึงย่อมกระทำได้ตามนัยมาตรา 82
=2.ตามมาตรา 82 แห่งประมวลกฎหมายที่ดิน ได้บัญญัติว่า ผู้ใดประสงค์จะขอจดทะเบียนลงชื่อผู้จัดการมรดกในหนังสือแสดงสิทธิในที่ดินให้ยื่นคำขอพร้อมด้วยนำหนังสือแสดงสิทธิในที่ดินนั้น
และหลักฐานการเป็นผู้จัดการมรดกมาแสดงต่อพนักงานเจ้าหน้าที่ตามมาตรา 71 ถ้าเป็นผู้จัดการมรดกโดยคำสั่งศาลให้พนักงานเจ้าหน้าที่ดำเนินการจดทะเบียนให้ตามคำขอ สำหรับเรื่องนี้ปรากฏข้อเท็จจริงว่า
นาง ค. เป็นผู้จัดการมรดกของนาย ข. ตามคำสั่งศาล ได้มายื่นคำขอจดทะเบียนผู้จัดการมรดก โดยได้นำใบจองเลขที่....พร้อมทั้งนำคำสั่งศาลมาแสดงต่อพนักงานเจ้าหน้าที่ตามมาตรา 71 แห่งประมวลกฎหมายที่ดิน
ซึ่งพนักงานเจ้าหน้าที่ได้จดทะเบียนลงชื่อนาง ค. ในฐานะผู้จัดการมรดกของนาย ข. ไว้หลังใบจองดังกล่างตามคำขอของนาง ค . แล้วการจดทะเบียนรายนี้ได้ดำเนินการไปถูกต้องแล้ว
-ข้อมูลค้นคว้า
บันทึกกองทะเบียนที่ดิน ที่ มท 0610.1/182 ลงวันที่ 7 มิถุนายน 2539 ตอบข้อหารือกองหนังสือสำคัญ (เรื่องยุติของ สมท. เลขที่ ส.กทม. 6706)

โดยข้อมูลที่รวบรวมมีจำนวนข้อความทางกฎหมายทั้งหมดจำนวน 36,817 ชุดข้อมูล ซึ่งสามารถแบ่งออกมาได้เป็นจำนวนข้อความทางกฎหมายทั้งหมด 51,794 ข้อความที่เกี่ยวข้องกับกฎหมายที่ดิน โดยจัดเก็บลงในฐานข้อมูลเชิงสัมพันธ์ของ SQLite ด้วยการเขียนสคริปต์โปรแกรมภาษาไพธอน นำเข้าข้อมูลที่ได้จากขั้นตอนการรวบรวม ซึ่งชุดข้อมูลนี้จะประกอบไปด้วย ER-Diagram ของตารางชุดข้อมูลและรายละเอียด ดังภาพที่ 3.4 และดังตารางที่ 3.1 – 3.3

ภาพที่ 3.4

ER-Diagram ของตารางชุดข้อมูล



ตารางที่ 3.1

ตารางของชุดข้อมูลประเภทกฎหมายและมาตรา

ลำดับ	แอททริบิวต์	คำอธิบาย	รูปแบบ
1	LAW_DETAIL_ID	รหัสของข้อมูล	INTEGER
2	LAW_NAME	ชื่อกฎหมาย	VARCHAR(50)
3	LAW_ARTICLE	มาตรา	VARCHAR(50)
4	DESCRIPTION	คำอธิบาย	VARCHAR(255)
5	LAW_CLASS	กลุ่มของกฎหมาย	VARCHAR(50)
6	CREATED	วันเวลาที่สร้างข้อมูล	TIMESTAMP
7	UPDATED	วันเวลาที่แก้ไขข้อมูล	TIMESTAMP
8	RECORD_STATUS	สถานะของข้อมูล	VARCHAR(1)

ภาพที่ 3.5

ตัวอย่างข้อมูลในตารางของชุดข้อมูลประเภทกฎหมายและมาตรา

	LAW_DETAIL_ID	LAW_NAME	LAW_ARTICLE	DES		CREATED	UPDATED	RECORD_STATUS
2	1	ประมวลกฎหมายแพ่งและพาณิชย์	62	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
3	2	ประมวลกฎหมายแพ่งและพาณิชย์	63	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
4	3	ประมวลกฎหมายแพ่งและพาณิชย์	66 วรรคหนึ่ง	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
5	4	ประมวลกฎหมายแพ่งและพาณิชย์	419	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
6	5	ประมวลกฎหมายแพ่งและพาณิชย์	1382	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
7	6	ประมวลกฎหมายแพ่งและพาณิชย์	1602	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
8	7	ประมวลกฎหมายแพ่งและพาณิชย์	165 วรรคหนึ่ง	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
9	8	ประมวลกฎหมายแพ่งและพาณิชย์	166	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
10	9	ประมวลกฎหมายแพ่งและพาณิชย์	193/14 (1)	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
11	10	ประมวลกฎหมายแพ่งและพาณิชย์	193/15	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
12	11	ประมวลกฎหมายแพ่งและพาณิชย์	193/30	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
13	12	ประมวลกฎหมายแพ่งและพาณิชย์	653 วรรคหนึ่ง	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
14	13	ประมวลกฎหมายแพ่งและพาณิชย์	1754 วรรคสาม	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
15	14	ประมวลกฎหมายแพ่งและพาณิชย์	420	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
16	15	ประมวลกฎหมายแพ่งและพาณิชย์	1169	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
17	16	ประมวลกฎหมายวิธีพิจารณาความแพ่ง	[NULL]	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
18	17	ประมวลกฎหมายวิธีพิจารณาความแพ่ง	225 วรรคหนึ่ง	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
19	18	ประมวลกฎหมายวิธีพิจารณาความแพ่ง	252	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
20	19	ประมวลกฎหมายอาญา	[NULL]	[NULL]	17	2024-06-02 10:31:33	[NULL]	N
21	20	ประมวลกฎหมายอาญา	227	[NULL]	17	2024-06-02 10:31:33	[NULL]	N
22	21	ประมวลกฎหมายแพ่งและพาณิชย์	1300	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
23	22	ประมวลกฎหมายแพ่งและพาณิชย์	1336	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
24	23	ประมวลกฎหมายแพ่งและพาณิชย์	1719	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
25	24	ประมวลกฎหมายแพ่งและพาณิชย์	1722	[NULL]	5	2024-06-02 10:31:33	[NULL]	N
26	25	ประมวลกฎหมายวิธีพิจารณาความแพ่ง	142	[NULL]	5	2024-06-02 10:31:33	[NULL]	N

ตารางที่ 3.2

ตารางของชุดข้อมูลคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา

ลำดับ	แอททริบิวต์	คำอธิบาย	รูปแบบ
1	LEGAL_TEXT_ID	รหัสของข้อมูล	INTEGER
2	SUPREME_CASE_NAME	ชื่อคำพิพากษา	VARCHAR(50)
3	SUPREME_CASE_NO	หมายเลขคำพิพากษา	VARCHAR(10)
4	SUPREME_BLACK_CASE_NO	หมายเลขคดีดำศาลชั้นต้น	VARCHAR(10)
5	TRIAL_BLACK_CASE_NO	หมายเลขคดีดำศาลชั้นต้น	VARCHAR(10)
6	TRIAL_RED_CASE_NO	หมายเลขคดีแดงศาลชั้นต้น	VARCHAR(10)
7	SHORT_TEXT	คำวินิจฉัยฉบับย่อ	VARCHAR(255)
8	SHORT_TEXT_PREP	คำวินิจฉัยฉบับย่อ (cleaned)	VARCHAR(255)
9	FULL_TEXT	คำวินิจฉัยฉบับเต็ม	VARCHAR(255)
10	FULL_TEXT_PREP	คำวินิจฉัยฉบับเต็ม (cleaned)	VARCHAR(255)
11	CASE_TEXT	กรณี	VARCHAR(255)
12	CASE_TEXT_PREP	กรณี (cleaned)	VARCHAR(255)
13	PROBLEM_TEXT	ประเด็นปัญหา	VARCHAR(255)
14	PROBLEM_TEXT_PREP	ประเด็นปัญหา (cleaned)	VARCHAR(255)
15	DISCUSSION_TEXT	ประเด็นข้อหารือ	VARCHAR(255)
16	DISCUSSION_TEXT_PREP	ประเด็นข้อหารือ (cleaned)	VARCHAR(255)
17	DIAGNOSTICS_TEXT	แนววินิจฉัย	VARCHAR(255)
18	DIAGNOSTICS_TEXT_PREP	แนววินิจฉัย (cleaned)	VARCHAR(255)
19	DOL_COMMENT_TEXT	ความเห็นกรมที่ดิน	VARCHAR(255)
20	DOL_COMMENT_TEXT_PREP	ความเห็นกรมที่ดิน (cleaned)	VARCHAR(255)
21	LAW_LABEL	ประเภทของกฎหมาย	VARCHAR(10)
22	LAW_CLASS	กลุ่มของกฎหมาย	VARCHAR(10)
23	SOURCE	แหล่งที่มา	VARCHAR(50)
24	TARGET	ประเภทของชุดข้อมูล	VARCHAR(1)
25	CREATED	วันเวลาที่สร้างข้อมูล	TIMESTAMP
26	UPDATED	วันเวลาที่แก้ไขข้อมูล	TIMESTAMP
27	RECORD_STATUS	สถานะของข้อมูล	VARCHAR(1)

ภาพที่ 3.6

ตัวอย่างข้อมูลในตารางของชุดข้อมูลคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา

	123	SUPREME_CA	SUPREME_	SHORT_TEXT	SHORT_TEXT_PREP
1	1	คำพิพากษาฎีกาที่	1661-1667/2526	ที่พิพาทเดิมเป็นที่ดินรกร้างว่างเปล่าอันเป็นสาธารณสมบัติของแผ่นดินตามประมวลกฎหมายแพ่งและพาณิ	พิพาทเดิมที่ดินรกร้างว่างเปล่าอันเป็นสาธารณสมบัติแผ่นดินประมวลกฎหมายแพ่งและพาณิชย์มาตรา พร
2	2	คำพิพากษาฎีกาที่	293/2535	โจทก์เข้าจับจองครอบครองที่ดินซึ่งเป็นที่รกร้างว่างเปล่าเนื้อที่ 12 ไร่เศษ หลังจากที่ประกาศใช้ประมวลกฎห	โจทก์จับจองครอบครองที่ดินที่รกร้างว่างเปล่าเนื้อที่ ไร่เศษหลังจากที่ประกาศใช้ประมวลกฎหมายที่ดินแล้ว
3	3	คำพิพากษาฎีกาที่	280/2537	ที่ดินของโจทก์ทั้งห้าเป็นที่ดินมือเปล่า ซึ่งได้ครอบครองทำประโยชน์ต่อเนื่องจากผู้ครอบครองเดิมตลอดมา	ที่ดินโจทก์ห้าที่ดินมือเปล่าครอบครองทำประโยชน์ผู้ครอบครองเดิมประมวลกฎหมายที่ดินบังคับโจทก์ห้าแล
4	4	คำพิพากษาฎีกาที่	1254/2530	คำว่า "ที่ดินรกร้างว่างเปล่า" ตาม ป.พ.พ. มาตรา 1304 (1) ตรงกับคำว่า "ที่ดินรกร้างว่างเปล่าอันเป็นสาธา	ที่ดินรกร้างว่างเปล่ามาตราที่ดินรกร้างว่างเปล่าอันเป็นสาธารณสมบัติแผ่นดินตามพรบว่าด้วยหวงห้ามที่ดิน
5	5	คำพิพากษาฎีกาที่	4431/2550	การจะได้ที่ดินเป็นกรรมสิทธิ์หรือสิทธิครอบครองโดยชอบด้วยกฎหมายนั้นได้มีบัญญัติไว้ในพ.ร.บ. ให้ใช้ ป	ที่ดินกรรมสิทธิ์สิทธิครอบครองโดยชอบด้วยกฎหมายบัญญัติพรที่ดินมาตรา วรรคครอบครองทำประโยชน์
6	6	คำพิพากษาศาลฎีกาที่	9797/2558	การประกาศเขตปฏิรูปที่ดินเพื่อการเกษตรกรรมตาม พ.ร.บ.การปฏิรูปที่ดินเพื่อเกษตรกรรมพ.ศ.2518 มา	ประกาศเขตปฏิรูปที่ดินเกษตรกรรมพรการปฏิรูปที่ดินเพื่อเกษตรกรรมมาตรา สำนักงานการปฏิรูปที่ดินเพื่
7	7	คำพิพากษาฎีกาที่	7216/2542	การโอนที่ดินที่ยังไม่มีใบจองก่อนเวลาที่ทางราชการจะออกใบจองโดยการส่งมอบการครอบครองไม่เป็นการต	โอนที่ดินใบจองก่อนเวลาทางราชการใบจองส่งมอบการครอบครองไม่เป็นการต้องห้ามชัดแจ้งกฎหมายพระ
8	8	คำพิพากษาฎีกาที่	6511/2543	จำเลยฎีกาว่า สัญญาซื้อขายที่ดินพิพาทตกเป็นโมฆะเสียเปล่าตั้งแต่วันทำสัญญาเพราะเป็นที่ดินมือเปล่าซึ่งไ	เลยฎีกาสัญญาซื้อขายที่ดินพิพาทตกเป็นโมฆะเสียเปล่าแต่วันทำสัญญาที่ดินมือเปล่าคำรับรองนายอำเภอท
9	9	คำพิพากษาฎีกาที่	1254/2530	คำว่า "ที่ดินรกร้างว่างเปล่า" ตาม ป.พ.พ.มาตรา 1304(1) ตรงกับคำว่า "ที่ดินรกร้างว่างเปล่าอันเป็นสาธาร	ที่ดินรกร้างว่างเปล่ามาตรา ที่ดินรกร้างว่างเปล่าอันเป็นสาธารณสมบัติแผ่นดินพรว่าด้วยหวงห้ามที่ดินรกร้า
10	10	คำพิพากษาฎีกาที่	5166/2531	ขณะใช้บังคับพระราชกฤษฎีกากำหนดเขตหวงห้ามที่ดินอำเภอปากน้ำโพ อำเภอพยุหรี อำเภอโกรกพระ จัง	บังคับพระราชกฤษฎีกาเขตหวงห้ามที่ดินอำเภอปากน้ำโพอำเภอพยุหรีอำเภอโกรกพระจังหวัดนครสวรรค์พ
11	11	คำพิพากษาฎีกาที่	3107/2539	พระบาทสมเด็จพระมงกุฎเกล้าเจ้าอยู่หัวพระราชทานที่ดินพิพาทให้แก่กองทัพเรือตั้งแต่พ.ศ.2465 และต่อมา	พระบาทสมเด็จพระมงกุฎเกล้าเจ้าอยู่หัวพระราชทานที่ดินพิพาทกองทัพเรือต่อมาพระราชกฤษฎีกาเขตหวง
12	12	คำพิพากษาฎีกาที่	8277/2544	เมื่อที่ดินพิพาทได้มีการออกพระราชกฤษฎีกากำหนดเป็นเขตหวงห้ามสำหรับใช้ในราชการทหารพร้อมด้วย	ที่ดินพิพาทพระราชกฤษฎีกาเขตหวงห้ามสำหรับราชการทหารแผนที่ท้ายพระราชกฤษฎีกาประกาศราชกิจ
13	13	คำพิพากษาฎีกาที่	8033/2553	นายอำเภอท่ายางออกหนังสือรับรองการทำประโยชน์ (น.ส.3 ก.) สำหรับที่ดินพิพาทให้แก่ ม. และย. ในปี 25	นายอำเภอท่ายางหนังสือรับรองการทำประโยชน์สำหรับที่ดินพิพาทปี ที่ดินพิพาทโรงสีบ้านปลูกสร้างชาวบ้า
14	14	คำพิพากษาฎีกาที่	920/2526	หนังสือรับรองการทำประโยชน์มิใช่หลักฐานที่แสดงว่าผู้มีชื่อในหนังสือ เป็นเจ้าของกรรมสิทธิ์ในที่ดินนั้นใน	หนังสือรับรองการทำประโยชน์หลักฐานมีชื่อหนังสือเป็นเจ้าของกรรมสิทธิ์ที่ดินทะเบียนโฉนดที่ดินโจทก์ผู้ซื้
15	15	คำพิพากษาฎีกาที่	148/2539	ใบไต่สวนเป็นหนังสือซึ่งเจ้าพนักงานออกให้เพื่อแสดงว่าเจ้าของที่ดินได้นำรังวัดเพื่อออกโฉนดไม่ใช่เอกสาร	ใบไต่สวนหนังสือเจ้าพนักงานเจ้าของที่ดินรังวัดออกโฉนดเอกสารสำคัญมีชื่อใบไต่สวนมีกรรมสิทธิ์ที่ดินโจท
16	16	คำพิพากษาฎีกาที่	4343/2539	การที่โจทก์มีชื่อในหนังสือรับรองการทำประโยชน์ (น.ส.3) นั้น ประมวลกฎหมายแพ่งและพาณิชย์มาตรา 13	โจทก์มีชื่อหนังสือรับรองการทำประโยชน์ประมวลกฎหมายแพ่งและพาณิชย์มาตรา สันนิษฐานบุคคลรายชื่อ
17	17	คำพิพากษาฎีกาที่	5132/2539	โจทก์กล่าวอ้างว่าบ้านและที่ดินพิพาทเป็นของโจทก์ที่1 กึ่งหนึ่งอีกกึ่งหนึ่งเป็นมรดกของ จ. สามีโจทก์ที่ 1	โจทก์อ้างว่าบ้านที่ดินพิพาทโจทก์กึ่งหนึ่งกึ่งหนึ่งมรดกสามีโจทก์ ตกโจทก์เจ็ดจำเลยฐานะทายาทเท่าๆจำเล
18	18	คำพิพากษาศาลฎีกาที่	9557/2558	ใบเสียภาษีบำรุงท้องที่ (ภ.ท.บ. 5) ไม่ใช่เอกสารที่แสดงว่าผู้ชำระค่าภาษีมีสิทธิในที่ดิน คือ สิทธิครอบครอง	ใบภาษีบำรุงท้องที่ เอกสารชำระค่าภาษีสิทธิที่ดินสิทธิครอบครองกรรมสิทธิ์ที่ดินเอกสารแสดงถึงสิทธิครอบ
19	19	คำพิพากษาฎีกาที่	6492/2550	เจ้าของเดิมครอบครองที่ดินพิพาทมาตั้งแต่ปี2498 ภายหลัง ป.ที่ดิน ใช้บังคับแล้ว เจ้าของเดิมและจำเลยซึ่ง	เจ้าของเดิมครอบครองที่ดินพิพาทปีที่ดินบังคับเจ้าของเดิมจำเลยโอนการครอบครองคุ้มครองสิทธิครอบคร
20	20	คำพิพากษาฎีกาที่	8371/2551	พ.ร.บ.การปฏิรูปที่ดินเพื่อเกษตรกรรมฯมาตรา 36 ทวิ วรรคหนึ่ง ที่กำหนดให้ ส.ป.ก. เป็นผู้ถือกรรมสิทธิ์ใ	พรการปฏิรูปที่ดินเพื่อเกษตรกรรมมาตรา ทวิวรรคกำหนดให้กรรมสิทธิ์ที่ดินอสังหาริมทรัพย์ใดนำไปใช้การ

ตารางที่ 3.3

ตารางจับคู่ของชุดข้อมูลในตารางที่ 3.1 กับ 3.2

ลำดับ	แอททริบิวต์	คำอธิบาย	รูปแบบ
1	LAW_DETAIL_ON_LEGAL_ID	รหัสของข้อมูล	INTEGER
2	LAW_DETAIL_ID	รหัสของข้อมูลคำพิพากษา	INTEGER
3	LEGAL_TEXT_ID	รหัสของข้อมูลกฎหมาย	INTEGER
4	CREATED	วันเวลาที่สร้างข้อมูล	TIMESTAMP
5	UPDATED	วันเวลาที่แก้ไขข้อมูล	TIMESTAMP
6	RECORD_STATUS	สถานะของข้อมูล	VARCHAR(1)

ภาพที่ 3.7

ตัวอย่างข้อมูลในตารางจับคู่ของชุดข้อมูลในตารางที่ 3.1 กับ 3.2

	LAW_DETAIL_ON_LEGAL_ID	LAW_DETAIL_ID	LEGAL_TEXT_ID	CREATED	UPDATED	RECORD_STATUS
1	1	7,983	1	2024-06-02 10:32:01	[NULL]	N
2	2	7,983	2	2024-06-02 10:32:01	[NULL]	N
3	3	7,983	3	2024-06-02 10:32:01	[NULL]	N
4	4	7,983	4	2024-06-02 10:32:01	[NULL]	N
5	5	7,983	5	2024-06-02 10:32:01	[NULL]	N
6	6	7,983	6	2024-06-02 10:32:01	[NULL]	N
7	7	7,984	7	2024-06-02 10:32:01	[NULL]	N
8	8	7,984	8	2024-06-02 10:32:01	[NULL]	N
9	9	7,985	9	2024-06-02 10:32:01	[NULL]	N
10	10	7,985	10	2024-06-02 10:32:01	[NULL]	N
11	11	7,985	11	2024-06-02 10:32:01	[NULL]	N
12	12	7,985	12	2024-06-02 10:32:01	[NULL]	N
13	13	7,985	13	2024-06-02 10:32:01	[NULL]	N
14	14	1,812	14	2024-06-02 10:32:01	[NULL]	N
15	15	1,812	15	2024-06-02 10:32:01	[NULL]	N
16	16	1,812	16	2024-06-02 10:32:01	[NULL]	N
17	17	1,812	17	2024-06-02 10:32:01	[NULL]	N
18	18	1,812	18	2024-06-02 10:32:01	[NULL]	N
19	19	2,491	19	2024-06-02 10:32:01	[NULL]	N
20	20	2,491	20	2024-06-02 10:32:01	[NULL]	N

3.1.2 การทำความสะอาดข้อมูล (Data Cleansing)

ในขั้นตอนการทำความสะอาดข้อมูลนี้ ผู้วิจัยทำความสะอาดข้อมูลทั้งหมดเพื่อลดข้อผิดพลาดในการวิจัยทำให้ได้ผลการทดลองถูกต้องแม่นยำมากยิ่งขึ้นโดยใช้ไลบรารีของ PyThaiNLP ในการดำเนินการขั้นตอนการทำความสะอาด ดังนี้

- เปลี่ยนตัวเลขไทยเป็นตัวเลขอารบิก
- ลบคำฟุ่มเฟือย (Stop Words) และคำไม่สำคัญต่าง ๆ เช่น ชื่อ โจทย์ จำเลย ยศ ตำแหน่ง วันที่ เลขที่คำสั่ง ชื่อจังหวัด แขวงตำบล เขตอำเภอ รวมถึงคำที่เจอบ่อย ๆ ในทุก ๆ เอกสาร
- ลบข้อความในส่วนที่เป็น URL ออกจากข้อความ
- ลบข้อความที่เป็นคำสั่ง HTML ออกจากข้อความ
- ลบช่องว่างที่มากเกินกว่าปกติออกจากข้อความ
- ลบเครื่องหมายวรรคตอนและเครื่องหมายพิเศษออกจากข้อความ
- ลบตัวเลขออกจากข้อความ

ตารางที่ 3.4

ตัวอย่างข้อมูลที่ผ่านการทำความสะอาดข้อมูล

ก่อนทำความสะอาดข้อมูล	หลังทำความสะอาดข้อมูล
คู่ความท้ากันว่า ถ้าจำเลยนำหนังสือซื้อขายที่ดินหรือสำเนาที่อำเภอรับรองมาจากอำเภอได้ โจทก์ยอมแพ้ ถ้าไม่มีหนังสือดังกล่าวที่พิพาทตกเป็นของโจทก์ เมื่ออำเภอมีหนังสือแจ้งมาว่าค้นไม่พบ คู่ความท้ากันใหม่ว่าให้ศาลเอาหนังสือตอบนี้ไปประกอบกับคำท้าเดิมแล้ววินิจฉัยว่าที่พิพาทจะเป็นของฝ่ายใด ศาลวินิจฉัยว่าตามหนังสือตอบของอำเภอฟังไม่ถนัดว่ามีหนังสือซื้อขายที่พิพาทหรือไม่ จึงชี้ขาดให้แพ้ชนะกันตามคำท้าไม่ได้ เมื่อต่างแถลงไม่สืบพยานคดีนี้ หน้าที่นำสืบตกแก่โจทก์ โจทก์จึงต้องแพ้	คู่ความท้าจำเลยหนังสือซื้อขายที่ดินสำเนาอำเภอมาจากอำเภอโจทก์ยอมแพ้ หนังสือพิพาทตกเป็นของโจทก์อำเภอหนังสือแจ้งค้นคู่ความท้าศาลหนังสือตอบท้าเดิม วินิจฉัยว่าที่พิพาทใดศาลวินิจฉัยหนังสือตอบอำเภอฟังถนัดหนังสือซื้อขาย พิพาทชี้ขาดแพ้ชนะท้าแถลงสืบพยานคดีหน้าที่นำสืบตกโจทก์โจทก์แพ้

3.1.3 การเตรียมข้อมูล (Data Preprocessing)

ในขั้นตอนการเตรียมข้อมูล เนื่องจากข้อมูลที่นำมาใช้ในงานวิจัยนี้รวบรวมมาจากหลายแหล่งที่มาทั้งหนังสือ ไฟล์ดิจิทัล และเว็บไซต์ โดยในขั้นตอนนี้จะแบ่งออกเป็น 5 ส่วน คือ การแปะป้ายประเภท (Labeled) กับชุดข้อมูล การแบ่งชุดข้อมูลสำหรับใช้เป็นชุดข้อมูลทดสอบ การหาข้อความทางกฎหมายที่คล้ายคลึงกันกับชุดข้อมูลที่ใช้ทดสอบ การตัดคำ (Word Segmentation) และการจัดการชุดข้อมูลที่ไม่สมดุลกัน

3.1.3.1 การแปะป้ายประเภท (Labeled)

เนื่องจากชุดข้อมูลมีจำนวนกฎหมายต่าง ๆ จำนวนมาก เนื่องจากแต่ละกฎหมายมีจำนวนปีที่ประกาศใช้ที่แตกต่างกันด้วย ผู้จัดทำจึงจำแนกประเภทของกฎหมายออกเป็นหมวดหมู่เดียวกัน จากนั้นทำการเลือกชุดข้อมูลที่ไม่เกี่ยวเนื่องกันในแต่ละประเภทแล้วนำชุดข้อมูลเหล่านั้นมาแปะป้ายประเภท โดยสามารถแบ่งประเภทออกได้เป็น 6 ประเภท ดังตารางที่ 3.4

ตารางที่ 3.5

การแปะป้ายประเภทของชุดข้อมูล โดยแบ่งตามประเภทกฎหมาย

ชื่อคลาส	ชื่อกฎหมาย
A	ประมวลกฎหมายแพ่งและพาณิชย์
B	ประกาศกระทรวง
C	ประมวลกฎหมายที่ดิน
D	ประมวลกฎหมายอาญา
E	ประมวลรัษฎากร
F	พระราชบัญญัติ

3.1.3.2 การแบ่งชุดข้อมูลสำหรับใช้เป็นชุดข้อมูลทดสอบ

เนื่องจากงานวิจัยนี้ต้องแบ่งชุดข้อมูลออกเป็น 2 ชุดข้อมูล คือ ชุดข้อมูลสำหรับการคัดแยกประเภทคำค้นข้อความทางกฎหมาย และชุดข้อมูลสำหรับทดสอบการค้นคืนข้อความทางกฎหมาย โดยการแบ่งชุดข้อมูลแรกจะแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ ชุดข้อมูลสำหรับเรียนรู้ ชุดข้อมูลสำหรับตรวจสอบ และชุดข้อมูลสำหรับทดสอบ ดังภาพที่ 3.9 ซึ่งจะแบ่งในอัตราส่วน 70:15:15

ภาพที่ 3.8

ภาพชุดข้อมูลของการคัดแยกประเภทคำค้นข้อความทางกฎหมาย

__label__F ที่ดินพิพาทพระราชกฤษฎีกาเขตหวงห้ามสำหรับราชการทหารแผนที่ท้ายพระราชกฤษฎีกาประกาศราชกิจจานุเบกษา
ประชาชนทราบต่อมากรมธนารักษ์ขึ้นทะเบียนที่ดินพิพาทที่ราชพัสดุดังนั้นประชาชนทราบที่ดินพิพาทที่ดินหวงห้ามเข้าไปยึดถือครอบครอง
เว้นแต่ได้รับอนุญาตพนักงานเจ้าหน้าที่จำเลยทั้งสองอ้างว่าเจตนาเข้าไปยึดถือครอบครองที่ดินพิพาทห้วงห้ามจำเลยทั้งสองมีความผิดฐานเข้าไปยึดถือครอบครองที่ดินรัฐฟ้อง
__label__F นายอำเภอท่ายางหนังสือรับรองการทำประโยชน์สำหรับที่ดินพิพาทปี ที่ดินพิพาทโรงสีบ้านปลูกสร้างชาวบ้าน
เข้าไปประโยชน์ใดที่ดินพิพาทลักษณะสาธารณะเป็นเหตุให้จำเลยโดยสุจริตสิทธิครอบครองที่ดินพิพาทจำเลยค่าตอบแทนเงินถึง บาท
ซื้อที่ดินพิพาทข้อเท็จจริงรับฟังได้จำเลยครอบครองที่ดินพิพาทเจตนาที่จะบุกรุกที่ดินอันเป็นสาธารณประโยชน์ต่อมาจำเลยทราบหนังสือ
นายอำเภอท่ายางที่ดินพิพาทส่วนหนึ่งศาลาสระน้ำหนองขนานสาธารณประโยชน์สาธารณสมบัติแผ่นดินหลักฐานยันกันหนังสือรับรองการทำประโยชน์
เอกสารสิทธิที่ดินพิพาทจำเลยออกจากที่ดินพิพาทเข้าไปยึดถือครอบครองครอบครองที่ดินพิพาทเจตนาจำเลยในมาแต่ต้นสิทธิครอบครองที่ดินพิพาท
ถือได้ว่าจำเลยเจตนาบุกรุกที่ดินพิพาทจำเลยที่ดินพิพาทเจตนาเดิมการกระทำความผิดตั้งแต่แรกจำเลยเจตนาบุกรุกทราบนายอำเภอท่ายางที่ดินพิพาทสาธารณสมบัติแผ่นดิน
__label__C หนังสือรับรองการทำประโยชน์หลักฐานมีชื่อหนังสือเป็นเจ้าของกรรมสิทธิ์ที่ดินทะเบียนโฉนดที่ดินโจทก์ผู้ซื้อที่ดินอ้างประมวลกฎหมายแพ่งและพาณิชย์มาตรา
__label__C โจทก์มีชื่อหนังสือรับรองการทำประโยชน์ประมวลกฎหมายแพ่งและพาณิชย์มาตรา สันนิษฐานบุคคลรายชื่อทะเบียนสิทธิครอบครอง
จำเลยอ้างว่าจำเลยซื้อที่ดินมาจากเจ้าของที่ดินเดิมครอบครองทำประโยชน์ที่ดินพิพาทจำเลยภาระการพิสูจน์
__label__C โจทก์อ้างว่าบ้านที่ดินพิพาทโจทก์กึ่งหนึ่งกึ่งหนึ่งมรดกสามีโจทก์ ตกโจทก์เจือจำเลยฐานะทายาทเท่าๆจำเลยต่อสู้บ้านที่ดินพิพาท
จำเลยเป็นเจ้าของแต่ผู้เดียวข้อตกลงทายาทจำเลยครอบครองสงบเจตนาเป็นเจ้าของ ปีสิทธิครอบครองเป็นเรื่องจำเลยอ้างว่าบ้านที่ดินพิพาทแบ่งปันตก
จำเลยจำเลยมีชื่อหนังสือรับรองการทำประโยชน์รับประโยชน์ข้อสันนิษฐานประมวลกฎหมายแพ่งและพาณิชย์มาตราจำเลยพิสูจน์จำเลยส่วนแบ่งทรัพย์สิน
ข้อตกลงทายาทชอบครอบครองโดยสุจริตอันเป็นข้อเท็จจริงจำเลยยกขึ้นอีกด้วยหนังสือรับรองการทำประโยชน์เอกสารสิทธิอันเป็นเอกสารราชการ
เจ้าพนักงานมีชื่อหนังสือรับรองการทำประโยชน์ครอบครองทำประโยชน์ที่ดินถือได้ว่าอสังหาริมทรัพย์ทะเบียนที่ดิน
__label__C ใบภาษีบำรุงท้องที่ เอกสารชำระค่าภาษีสิทธิที่ดินสิทธิครอบครองกรรมสิทธิ์ที่ดินเอกสารแสดงถึงสิทธิครอบครองกรรมสิทธิ์ที่ดินเป็นไปตามที่ดินมาตรา
หนังสือรับรองทำประโยชน์โฉนดที่ดินการกระทำการใบภาษีบำรุงท้องที่จำเลย กระทบโจทก์ยึดถือที่ดินเอกสารสิทธิจำเลยหกมอบให้เป็นหลักประกันเข้าร่วม
โครงการปลูกพืชพลังงานทดแทนโจทก์โจทก์เจตนาชำระค่าภาษีบำรุงท้องที่ใบภาษีบำรุงท้องที่ ท่อนมอบให้เจ้าของที่ดินติดต่อชำระภาษีบำรุงท้องที่จำเลย
แจ้งความใบภาษีบำรุงท้องที่มอบให้โจทก์ยึดถือหายแล้วไปออกใบแทนโจทก์เสียหายโจทก์ผู้เสียหาย

ส่วนชุดข้อมูลทดสอบการค้นคืนข้อความทางกฎหมายนั้น จะเลือกข้อมูลจากชุดข้อมูลที่มีในหัวข้อประเด็นข้อปัญหาหรือกรณีปัญหาเป็นข้อมูลทดสอบการค้นคืน โดยมีจำนวนทั้งหมด 103 ข้อความ ซึ่งตัวอย่างของชุดข้อมูลทดสอบคำค้นคืนข้อความทางกฎหมาย ดังภาพที่ 3.10

ภาพที่ 3.9

ภาพตัวอย่างของชุดข้อมูลทดสอบคำค้นคืนข้อความทางกฎหมาย

การจดทะบียนแลงชื่อผู้จัดการมรดกขึ้นใจอง	จดทะบียนลงชื่อผู้จัดการมรดกขึ้นจอง
ขอจดทะบียนลงชื่อผู้จัดการมรดกตามคำสั่งศาลคัวงเคยมาเป็นเวลาสิบเป็นสิ้น	จดทะบียนลงชื่อผู้จัดการมรดกคำสั่งศาลคว่างเคยเป็นเวลาสิบเป็
ขอจดทะบียนแปลี่ยนผู้จัดการมรดก เนื่องจากศาลได้มีคำสั่งให้เพิกถอนผู้จัดการมรดกซึ่งได้จดทะบียนลงชื่อผู้จัดการมรดกและโอนมรดกลงชื่อตนแองเป็นผู้รับมรดกไว้แล้ว	จดทะบียนผู้จัดการมรดกศาลคำสั่งเพิกถอนผู้จัดการมรดกจดทะบียนลงชื่อผู้จัดการมรดกโอนมรดกลงชื่อผู้รับมรดก
ผู้จัดการมรดกตามคำสั่งศาลอันขอจดทะบียนลงชื่อผู้จัดการมรดกที่ไม่ถือเจ้าของคำยังไม่มีชื่อไม่ถือเป็นที่ดิน	ผู้จัดการมรดกคำสั่งศาลอันขอจดทะบียนลงชื่อผู้จัดการมรดกโฉนดที่ดินเจ้านครมีชื่อโฉนดที่ดิน
ศาลมีคำสั่งแต่งตั้งผู้จัดการมรดกย้อน	ศาลคำสั่งแต่งตั้งผู้จัดการมรดกย้อน
ศาลตั้งผู้ยาร์มีผู้จัดการมรดก	ศาลตั้งผู้ยามีผู้จัดการมรดก
ศาลตั้งผู้จัดการมรดกสองคนแต่ผู้จัดการมรดกเพียงคนเดียวมาอ้นขอจดทะบียนลงชื่อผู้จัดการมรดกตามคำพิพากษาตามยอมทีกระทำโนฐานะผู้จัดการมรดก	ศาลผู้จัดการมรดกสองคนผู้จัดการมรดกคนเดียวอ้นจดทะบียนลงชื่อผู้จัดการมรดกคำพิพากษาฐานะผู้จัดการมรดก
ตัดสำแดงจดทะบียนลงชื่อผู้จัดการมรดกตามพินัยกรรม	ตัดสำแดงจดทะบียนลงชื่อผู้จัดการมรดกพินัยกรรม
ผู้จัดการมรดกขอแก้ไขชื่อบิดามารดาของเจ้ามรดก	ผู้จัดการมรดกชื่อบิดามารดาเจ้ามรดก
ผู้จัดการมรดกตามคำสั่งศาลตัดสำแดงการจดทะบียนลงชื่อผู้จัดการมรดกตามพินัยกรรม	ผู้จัดการมรดกคำสั่งศาลตัดสำแดงจดทะบียนลงชื่อผู้จัดการมรดกพินัยกรรม
จดทะบียนผู้จัดการมรดกที่โฉนดที่ดินผู้ยอมไม่สามารถนำโฉนดที่ดินฉบับเจ้าของมาได้	จดทะบียนผู้จัดการมรดกโฉนดที่ดินโฉนดที่ดินฉบับเจ้าของ
ขอจดทะบียนลงชื่อผู้จัดการมรดกและโอนมรดกตามคำสั่งศาลที่ดินผู้ไม่ถึงที่สุด	จดทะบียนลงชื่อผู้จัดการมรดกโอนมรดกคำสั่งศาลคดีถึงที่สุด
ขอจดทะบียนลงชื่อผู้จัดการมรดกโฉนดที่ดินที่เป็นเสิมรดและยังไม่ปรากฏชื่อเจ้าของมรดกเป็นผู้ถือกรรมสิทธิ์	จดทะบียนลงชื่อผู้จัดการมรดกโฉนดที่ดินเสิมรดชื่อเจ้าของมรดกกรรมสิทธิ์
ผู้จัดการมรดกยังไม่ได้จดทะบียนลงชื่อที่ดินโฉนดที่ดินได้ทำสัญญาประนีประนอมยอมความ	ผู้จัดการมรดกจดทะบียนลงชื่อโฉนดที่ดินแห่งสัญญาประนีประนอมยอมความ
ผู้จัดการมรดกโอนที่ดินให้เป็นที่สาธารณประโยชน์	ผู้จัดการมรดกโอนที่ดินสาธารณประโยชน์
ศาลมีคำสั่งไว้ในคำร้องของผู้จัดการมรดกว่า "การจัดการมรดกให้เป็นไปตามคำสั่งและกฎหมาย"	ศาลคำสั่งคำร้องผู้จัดการมรดกเป็นไปตามคำสั่งกฎหมาย
ผู้จัดการมรดกขอโอนมรดกให้แก่บุคคลอื่นไม่ใช่ทายาทตามพินัยกรรม	ผู้จัดการมรดกโอนมรดกบุคคลทายาทพินัยกรรม
ผู้จัดการมรดกขอจดทะบียนโอนมรดกให้กับทายาทที่อยู่ในคำดับถัดลงไปโดยทายาทคำดับต้นไม่มีชีวิตอยู่	ผู้จัดการมรดกขอจดทะบียนโอนมรดกทายาทที่อยู่คำดับถัดทายาทคำดับต้นไม่มีชีวิต
ผู้จัดการมรดกได้ดำเนินการขอทายาทเป็นผู้ถือกรรมสิทธิ์ตามรายงานกระบวนการพิจารณาของศาลที่โฉนดที่ดิน ซึ่งได้จดทะบียนโอนมรดกให้กับทายาทไปแล้ว	ผู้จัดการมรดกขอได้ดำเนินการขอทายาทกรรมสิทธิ์รายงานกระบวนการพิจารณาศาลโฉนดที่ดินจดทะบียนโอนมรดกทายาท
ผู้จัดการมรดกขอโอนมรดกให้กับบุคคลของเจ้ามรดก โดยมิได้โอนเมื่อกับทายาทโดยชอบของมรดกซึ่งถึงแก่กรรมภายหลังเจ้ามรดก	ผู้จัดการมรดกโอนมรดกบุคคลเจ้ามรดกโอนทายาทโดยชอบมรดกถึงแก่กรรมเจ้ามรดก
ผู้จัดการมรดกตามพินัยกรรมขอจดทะบียนโอนมรดกให้กับผู้รับพินัยกรรมจำนวนเนื้อที่ดินเมื่อมากว่าที่ผู้ทำพินัยกรรมกำหนดไว้	ผู้จัดการมรดกพินัยกรรมจดทะบียนโอนมรดกผู้รับพินัยกรรมจำนวนเนื้อที่ดินผู้ทำพินัยกรรม
ผู้จัดการมรดกนำสำแนาพินัยกรรมซึ่งศาลแพ่งชั้นไว้เป็นเอกสารที่เกาพิจารณาคดีแล้วผู้จัดการมรดกมายื่นโอนมรดกให้แก่ทายาท	ผู้จัดการมรดกสำแนาพินัยกรรมศาลแพ่งเอกสารพิจารณาคดีผู้จัดการมรดกโอนมรดกทายาท
การขอรับโอนมรดกสิ่งปลูกสร้างโดยมิผู้จัดการมรดก	ขอรับโอนมรดกสิ่งปลูกสร้างผู้จัดการมรดก
ผู้จัดการมรดกขอเพิ่มชื่อทายาทในระยการจดทะบียนโอนมรดกซึ่งได้จดทะบียนเสร็จแล้ว	ผู้จัดการมรดกชื่อทายาทรายการจดทะบียนโอนมรดกจดทะบียน
ผู้จัดการมรดกขอจดทะบียนเนื้อที่ดินของเจ้าถึงกรมสิทธิ์ขาม (ไม่มีคำตอบแทน) ในที่ดินของเจ้ามรดกโดยอ้างว่าเป็นเหตุที่ดินที่ทำมาหาได้ร่วมกันขณะอยู่กินฉันฉันที่สามีภรรยา	ผู้จัดการมรดกจดทะบียนเนื้อที่ดินของเจ้ามรดกถึงกรมสิทธิ์ตอบแทนที่ดินเจ้ามรดกอ้างว่าพรั่งมีเท่ามาหาได้ร่วมกันฉันที่สามีภรรยา
จดทะบียนผู้จัดการมรดกตามคำสั่งศาลถึงบางส่วนและขอจดทะบียนโอนที่ดินให้แก่ทดแองตามคำพิพากษาตามยอมซึ่งได้ทำให้กับทายาทเจ้ามรดก	จดทะบียนผู้จัดการมรดกคำสั่งศาลบางส่วนจดทะบียนโอนที่ดินคำพิพากษาทำทายาทเจ้ามรดก
ผู้จัดการมรดกขอจัดแบ่งกรรมสิทธิ์ร่วมที่โฉนดที่ดินซึ่งตนมีชื่อถือกรรมสิทธิ์ร่วมอยู่ด้วย	ผู้จัดการมรดกจัดแบ่งกรรมสิทธิ์ร่วมโฉนดที่ดินมีชื่อกรรมสิทธิ์ร่วม
ผู้จัดการมรดกตามคำพิพากษาตามยอมบางคนไม่มาดำเนินการ	ผู้จัดการมรดกคำพิพากษาบางคนดำเนินการ

3.1.3.3 การหาข้อความทางกฎหมายที่คล้ายคลึงกันกับชุดข้อมูลที่ใช้ทดสอบ

เนื่องจากชุดข้อมูลมีจำนวนมากทำให้ผู้จัดทำไม่สามารถจับคู่ข้อความได้ด้วยตาเปล่า ดังนั้นผู้จัดทำจึงเขียนสคริปต์โปรแกรมภาษาไพธอนขึ้นมาใช้ในการหาคีย์เวิร์ด (Keywords) ของชุดข้อมูลทดสอบ เพื่อช่วยกรองชุดข้อมูลที่มีความคล้ายคลึงกันกับชุดข้อมูลทดสอบแล้วใช้การอ่านช่วยในการเลือกชุดข้อมูลที่มีความคล้ายคลึงกันอีกครั้งนึง จากชุดข้อมูลทดสอบทั้งหมด 103 ข้อความ

ภาพที่ 3.10

ภาพตัวอย่างคีย์เวิร์ด (Keywords) ของชุดข้อมูลทดสอบ

```
['ผู้จัดการมรดก', 'ใบจอง'],
['ผู้จัดการมรดก', 'คำสั่งศาล', 'ล่วงเลย'],
['ผู้จัดการมรดก', 'ศาล', 'คำสั่ง', 'เพิกถอน', 'มรดก', 'ผู้รับมรดก'],
['ผู้จัดการมรดก', 'คำสั่งศาล', 'คัดค้าน', 'พินัยกรรม'],
['ผู้จัดการมรดก', 'มรดก', 'คำสั่งศาล', 'คดี', 'ถึงที่สุด'],
['ผู้จัดการมรดก', 'จดทะเบียน', 'โฉนดที่ดิน'],
['ผู้จัดการมรดก', 'ที่ดิน', 'สาธารณประโยชน์'],
['ผู้จัดการมรดก', 'มรดก', 'ทายาท', 'พินัยกรรม'],
['ผู้จัดการมรดก', 'มรดก', 'ทายาท', 'ชีวิต'],
['ผู้จัดการมรดก', 'พินัยกรรม', 'มรดก', 'ทายาท'],
['ขอรับ', 'มรดก', 'สิ่งปลูกสร้าง', 'ผู้จัดการมรดก'],
['ผู้จัดการมรดก', 'ทายาท', 'รายการ', 'มรดก'],
['ทางด่วน'],
['ผู้จัดการมรดก', 'ขายฝาก', 'ขยาย', 'ไถ่'],
['ผู้จัดการมรดก', 'โอน', 'มรดก', 'สัญญา', 'ทายาทโดยธรรม'],
['ศาล', 'ผู้จัดการมรดก', 'โอน', 'ที่ดิน', 'คำพิพากษา', 'ทายาท'],
['ผู้จัดการมรดก', 'คำสั่งศาล', 'มรดก', 'ที่ดิน', 'พินัยกรรม', 'เจ้ามรดก'],
['ผู้จัดการมรดก', 'มรดก', 'ผู้รับพินัยกรรม', 'พินัยกรรม', 'ที่ดิน']
```

3.1.3.4 การตัดคำ (Word Segmentation)

เนื่องจากชุดข้อมูลของงานวิจัยนี้เป็นภาษาไทยทำให้เทคนิคที่ใช้ในการตัดคำนั้นแตกต่างกับภาษาอังกฤษ เพราะว่าภาษาไทยมีรูปแบบการเขียนคำเรียงติดกัน ส่วนภาษาอังกฤษใช้ช่องว่างในการแบ่งคำ ผู้วิจัยใช้ไลบรารีของ PyThaiNLP เข้ามาช่วยในการตัดคำภาษาไทยเพื่อให้ได้ข้อมูลในระดับคำ (Word Level)

ภาพที่ 3.11

ภาพตัวอย่างการตัดคำภาษาไทยในระดับคำ (Word Level)

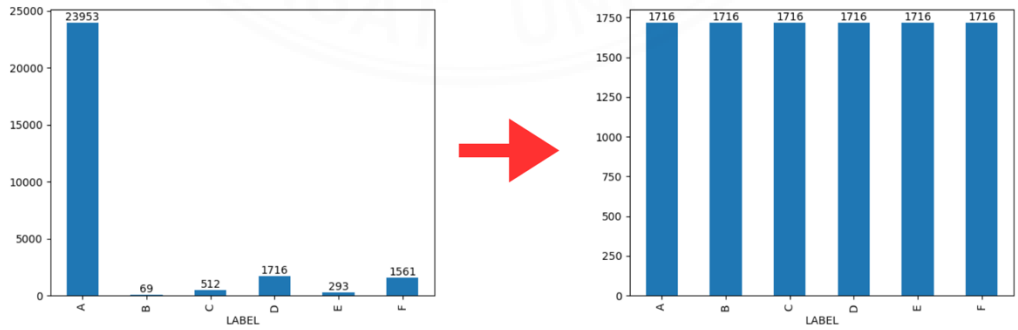
['บ้าน', 'บางสวน', 'ปลูก', 'รุกล้ำ', 'เข้าไป', 'ที่ดิน', 'ชายทะเล', 'อันเป็น', 'สาธารณสมบัติ', 'แผ่นดิน', 'ประชาชน', 'เนื้อที่', ' ', 'ตารางเมตร', 'ปลูกสร้าง', 'บิดา', 'มารดา', 'ถึงแก่กรรม', 'บ้าน', 'ที่ดิน', 'ตกเป็น', 'สิทธิอาศัย', 'บ้าน', 'สร้าง', 'รั้ว', 'สังกะสี', 'ปิดกั้น', 'ทางเดิน', 'ทางสาธารณะ', 'เดิน', 'ชายทะเล', 'พฤติการณ์', 'ที่ดิน', 'ปลูก', 'บ้าน', 'ล้อมรั้ว', 'ปิดกั้น', 'ทางเดิน', 'โดยสุจริต', 'เข้าไป', 'ยึดถือ', 'ครอบครอง', 'ที่ดิน', 'ชายทะเล', 'โดยมีเจตนา', 'ที่จะ', 'ฝ่า', 'ฝืนกฎหมาย', 'มีความผิด', 'ประมวลกฎหมาย', 'ที่ดิน', 'มาตรา', ' ', 'ทวี']

3.1.3.5 การจัดการชุดข้อมูลที่ไม่สมดุลกัน

เนื่องจากชุดข้อมูลมีจำนวนของประเภทที่ถูกแปะป้ายไม่เท่ากัน จึงแก้ปัญหาด้วยวิธี Synthetic Minority Over-Sampling Technique (SMOTE) โดยใช้ไลบรารีของ PyThaiNLP เข้ามาช่วยในการจัดการปัญหานี้

ภาพที่ 3.12

ภาพเปรียบเทียบจำนวนประเภทที่ถูกแปะป้ายก่อนและหลังการทำจัดการชุดข้อมูลที่ไม่สมดุลกันด้วยวิธี SMOTE



3.1.4 การคัดแยกประเภทคำค้น (Text Classification)

ในขั้นตอนการคัดแยกประเภทคำค้นงานวิจัยนี้จะใช้เทคนิค FastText เข้ามาใช้ในการคัดแยกประเภทของคำค้นออกเป็นตามมาตราของกฎหมายนั้น ๆ โดยจะใช้เครื่องมือคัดแยกประเภทข้อความ (Text Classification) ซึ่งสามารถแปะป้ายประเภท (Labeled) ได้หลายประเภทกับชุดข้อมูล

3.1.4.1 การแทนคำ (Word Representation)

ในขั้นตอนการแทนคำงานวิจัยนี้มุ่งเน้นการนำเทคนิค FastText เข้ามาใช้การแทนคำในข้อความทางกฎหมาย ซึ่งจะเรียกใช้ผ่านไลบรารีของตัว FastText โดยจะใช้ Pre-Trained Word Embedding ที่เรียนรู้มากจากชุดข้อมูลคำใน Wikipedia มีวิธี Skipgram ที่ทำนายคำเป้าหมาย (Target Word) ซึ่งสุ่มเลือกคำที่ใกล้เคียงคำเป้าหมายเข้ามาช่วยคำนวณเวกเตอร์ (Vector)

ภาพที่ 3.13

ภาพตัวอย่างการแทนคำ (Word Representation) ด้วย FastText

```
[ 4.73565370e-01,  1.04950450e-01, -6.93905830e-01,  2.98824817e-01,
       -3.82763177e-01,  9.46009278e-01, -3.60264003e-01,  5.23279011e-01,
       -1.34528625e+00, -6.78838670e-01, -4.92536396e-01,  2.18731537e-01,
       -1.66903806e+00, -2.38320380e-01,  1.21652193e-01, -3.13156754e-01,
        5.69900513e-01, -9.56381142e-01,  6.62509143e-01,  6.45947993e-01,
        2.49388620e-01, -7.42312133e-01, -9.05490696e-01,  4.21261519e-01,
        1.07003562e-01,  8.29634130e-01, -4.24962968e-01,  4.50886279e-01,
       -8.51219773e-01, -1.26601756e+00,  9.37195480e-01, -2.03667015e-01,
        1.99637860e-01, -1.93090096e-01,  6.51754081e-01,  1.08801961e+00,
       -7.29206502e-01, -1.26811057e-01, -2.08517045e-01,  7.34837890e-01,
       -7.77198255e-01,  1.53251186e-01,  4.78443094e-02,  2.89684147e-01,
       -1.12137213e-01, -7.00212955e-01, -3.57979506e-01,  7.51801610e-01,
        5.35786033e-01,  1.58239830e+00, -1.10869122e+00, -1.36996901e+00,
        2.44233876e-01, -2.23869830e-01,  9.07898486e-01, -4.96345937e-01,
        4.67276387e-02, -6.72606766e-01, -8.72952878e-01, -1.02734196e+00,
        1.59449494e+00, -5.27537227e-01,  4.97109950e-01,  7.72192776e-01,
        6.84210658e-01, -4.53926742e-01, -7.85745904e-02, -1.48486269e+00,
       -1.99238092e-01,  1.02756095e+00,  1.75976086e+00, -3.37539256e-01,
        1.75211489e+00, -1.26092720e+00, -1.36475980e+00, -1.34664640e-01,
       -3.45073074e-01,  7.28911281e-01,  1.43626308e+00, -5.48493445e-01,
        1.11577094e-01,  1.20426345e+00,  8.98719966e-01,  5.50150335e-01,
       -1.57686427e-01,  1.32553732e+00,  2.09392881e+00, -4.99095678e-01,
       -3.33379000e-01, -2.55076289e-01,  1.17414856e+00,  5.37380457e-01,
       -8.50972831e-01,  9.60495830e-01, -2.21023232e-01, -1.01083541e+00,
       -1.47845614e+00, -1.49994537e-01, -1.32388866e+00,  1.35620725e+00,
       -1.85225713e+00,  2.39798951e+00, -1.31460834e+00,  1.40602493e+00,
        7.96215832e-01,  7.05522224e-02,  8.57894838e-01,  3.51459771e-01,
        1.49210370e+00, -1.15480702e-02, -8.39724004e-01, -3.37804973e-01,
```

3.1.5 การค้นคืนข้อมูล (Information Retrieval)

ในส่วนนี้เป็นส่วนของการค้นคืนข้อความทางกฎหมาย จากที่ได้ทบทวนวรรณกรรมในบทที่ 2 ผู้วิจัยได้นำเทคนิค roBERTa มาใช้ในการค้นคืนคำพิพากษาศาลฎีกา เพราะมีประสิทธิภาพที่ดีผ่านการทดลองมาแล้วกับชุดข้อมูลที่มีลักษณะคล้ายคลึงกับชุดข้อมูลของงานวิจัยนี้ โดยหลังจากการแทนคำกับข้อความแล้วใช้ Cosine Similarity เพื่อคำนวณคะแนนความคล้ายคลึงกันระหว่างข้อความและคำพิพากษา

ซึ่งการในงานวิจัยนี้มีการเลือกใช้โมเดล WangchanBERTa ที่เป็น Pre-trained Thai Language Model ซึ่งเป็นโมเดล Monolingual Language Model มีพื้นฐานบนเทคนิค roBERTa โดยโมเดลมีการเรียนรู้จากชุดข้อมูลขนาดใหญ่ที่เป็นภาษาไทย งานวิจัยนี้เลือกใช้โมเดล WangchanBERTa ที่ผ่านการ Fine-tuned ด้วยเทคนิคต่าง ๆ ทั้งหมด 4 โมเดล ดังตารางที่ x.x ซึ่งแต่ละโมเดลมีประสิทธิภาพที่ดีมากกว่าโมเดลเดิม ด้วยการ Fine-tuned โมเดล โดยมีทั้งการเพิ่มเทคนิคในชั้น Word Representation โดยเพิ่มเทคนิค Simple Contrastive Learning of Sentence Embeddings (SimCSE) หรือเพิ่มเทคนิค Self-Supervised CrossView Training (SCT) ทำให้โมเดลมีประสิทธิภาพดีขึ้นในการทำ Sentence Embeddings รวมถึงการเพิ่มประสิทธิภาพของโมเดลด้วยการ Fine-tuned กับชุดข้อมูลอื่นเพิ่มเติมด้วยชุดข้อมูล Sanook news หรือชุดข้อมูล Thai Wikipedia โดยในงานวิจัยจะเปรียบเทียบและเลือกใช้โมเดลที่ให้ประสิทธิภาพดีที่สุดในการหาข้อความที่คล้ายคลึงกันมากที่สุดด้วยการใช้ Cosine Similarity กับชุดข้อมูลที่ใช้ในงานวิจัยนี้

ตารางที่ 3.6

โมเดล WangchanBERTa ที่ใช้ในงานวิจัย ซึ่งผ่านการ Fine-tuned ด้วยเทคนิคต่าง ๆ

ชื่อโมเดล	เทคนิคการ Fine-tuned
simcse-model-wangchanberta-base-att-spm-uncased	SimCSE
simcse-model-wangchanberta-finetuned-sanook-news	SimCSE, Sanook news
SCT-model-wangchanberta	SCT
simcse-model-wangchanberta	SimCSE, ThaiWikipedia

3.1.6 การวัดผลการวิจัย

ในขั้นตอนการวัดผลการทดลอง ผู้วิจัยเลือกวิธีการวัดประสิทธิภาพทั้งหมด 3 วิธีการ ดังนี้

3.1.6.1 Precision

สำหรับการเปรียบเทียบผลของการคัดแยกประเภทกฎหมายของคำค้นข้อความทางกฎหมายว่าสามารถคัดแยกประเภทต่อผลลัพธ์ที่ทำนายทั้งหมด ดังสมการ

$$Precision = \frac{TP}{TP + FP}$$

สำหรับการเปรียบเทียบผลของการค้นคืนข้อความทางกฎหมายว่าสามารถค้นคืนข้อความทางกฎหมายต่อผลลัพธ์ที่ค้นคืนได้ทั้งหมด ดังสมการ

$$Precision = \frac{relevant\ retrieved}{retrieved\ items}$$

3.1.6.2 Recall

สำหรับการเปรียบเทียบผลของการทำนายว่าคัดแยกประเภทกฎหมายของคำค้นข้อความทางกฎหมายได้ถูกต้องต่อผลลัพธ์ที่เป็นจริงทั้งหมด ดังสมการ

$$Recall = \frac{TP}{TP + FN}$$

สำหรับการเปรียบเทียบผลของการค้นคืนข้อความทางกฎหมายของคำค้นข้อความทางกฎหมายได้ถูกต้องต่อผลลัพธ์ที่เกี่ยวข้องกับคำค้นทั้งหมด ดังสมการ

$$Recall = \frac{relevant\ items\ retrieved}{relevant\ items}$$

3.1.6.3 F2-Score

เป็นการเปรียบเทียบผลของการทำนายหรือการค้นคืนข้อความทางกฎหมาย โดยใช้ค่าเฉลี่ยฮาร์มอนิก ซึ่งจะใช้ค่า Precision และ Recall เป็นการให้ความสำคัญกับค่า Recall เนื่องด้วยการค้นคืนคำพิพากษา ควรได้ข้อมูลที่เกี่ยวข้องครอบคลุมกับคำค้นหามากที่สุด โดยมีสูตรดังสมการ

$$F2 = \frac{5 * Precision * Recall}{4 * Precision + Recall}$$

3.1.6.4 ความเร็วในการค้นคืนข้อมูลทางกฎหมาย

นอกจากการหาประสิทธิภาพของการทำนายหรือค้นคืน งานวิจัยนี้ได้วัดในส่วนของการทดสอบความเร็วระหว่าง วิธีการคัดแยกประเภทร่วมกับการค้นคืน และการค้นคืนเพียงอย่างเดียว โดยจะวัดเวลาที่ใช้ในกระบวนการทำงานของวิธีการข้างต้น

3.2 เครื่องมือที่ใช้ในการวิจัย

ในการจัดเตรียมตลอดจนการวัดผลชุดข้อมูล ผู้วิจัยได้พัฒนาด้วยภาษา Python Version 3 บน Google Colab ซึ่งเป็นบริการ Software as a Service (SaaS) ของ Google สำหรับใช้งานเขียนและเรียกใช้ Python ในเบราว์เซอร์ โดยจะใช้ GPU (Graphics Processing Unit) เข้ามาช่วย ในการประมวลผลการเรียนรู้โมเดล

ตารางที่ 3.7

สรุปรายละเอียดการใช้เทคนิคและเครื่องมือที่ใช้ในการทดลอง

ชื่อเครื่องมือ	คำอธิบาย
PyThaiNLP	ไลบรารีภาษา Python สำหรับเตรียมชุดข้อมูลที่เป็นภาษาไทย
FastText	ไลบรารีภาษา Python สำหรับทำ Word Embedding และ Text Classification
Sentence-transformers	ไลบรารีภาษา Python สำหรับเรียกใช้โมเดล roBERTa และทำ Word Representation
Pytorch	ไลบรารีภาษา Python สำหรับเก็บและแปลงตัวแปรที่ได้จากการทำ Word Representation ให้สามารถใช้ GPU ในการประมวลผล
SQLite3	ไลบรารีภาษา Python สำหรับจัดเก็บชุดข้อมูลคำพิพากษา ฯ และการแปะป้าย (Labeled)
Numpy	ไลบรารีภาษา Python สำหรับสร้างสมการคำนวณค่าความคล้ายคลึงกันของข้อความ (Cosine Similarity)
Pandas และ Matplotlib	ไลบรารีภาษา Python สำหรับทำสถิติและวัดผลการวิจัย

บทที่ 4
ผลการวิจัยและอภิปรายผล

เนื้อหาในบทนี้นำเสนอผลการทดลองการคัดแยกและผลการค้นคืนข้อความทางกฎหมายพร้อมทั้งอภิปรายผลที่ได้จากการทดลอง โดยใช้เทคนิค FastText ในการคัดแยกประเภทคำค้นหาข้อความทางกฎหมาย เพื่อหากฎหมายที่เกี่ยวข้องกับคำค้น ร่วมกับเทคนิค roBERTa ในการทำ Word Representation เปรียบเทียบกับการใช้เทคนิค roBERTa ในการทำ Word Representation คู่กับการใช้เทคนิคการหาความคล้ายคลึงกันของข้อความด้วย Cosine Similarity เพื่อค้นคืนข้อความทางกฎหมาย โดยใช้ผลลัพธ์ Precision, Recall, F2-Score และความเร็วในการค้นคืนข้อความทางกฎหมาย เข้ามาเปรียบเทียบประสิทธิภาพ

4.1 ผลการทดลอง

4.1.1 เปรียบเทียบประสิทธิภาพการค้นคืนข้อความทางกฎหมายระหว่างเทคนิค roBERTa โมเดล WangchanBERTa ที่ผ่านการ Fine-tuned 4 โมเดล

ผลจากการทดลองเปรียบเทียบประสิทธิภาพการค้นคืนข้อความทางกฎหมายด้วยโมเดล WangchanBERTa ที่ผ่านการ Fine-tuned ด้วยเทคนิคต่าง ๆ พบว่าโมเดล SCT-model-wangchanberta ที่หาค่าความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมายด้วยเทคนิค Cosine Similarity ที่ 0.5 ให้ผลลัพธ์ Recall สูงที่สุด ดังตารางที่ 4.1 ดังนั้นในการทดลองของงานวิจัยนี้จะใช้โมเดลดังกล่าวข้างต้นในการทดลองทั้งหมด

ตารางที่ 4.1

ตารางเปรียบเทียบประสิทธิภาพการค้นคืนข้อความทางกฎหมาย

ชื่อโมเดล	Precision	Recall	F2-Score
simcse-model-wangchanberta-base-att-spm-uncased (Cosine Similarity 0.6)	0.42	0.29	0.385
simcse-model-wangchanberta-finetuned-sanook-news (Cosine Similarity 0.6)	**0.531**	0.4	**0.498**
SCT-model-wangchanberta (Cosine Similarity 0.5)	0.182	**0.747**	0.214
simcse-model-wangchanberta (Cosine Similarity 0.6)	0.497	0.335	0.453

4.1.2 FastText และ roBERTa

ผลจากการทดลองโดยใช้เทคนิคการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText โดยแบ่งออกเป็น 6 คลาส ให้แต่ละคลาสแทนประเภทของกฎหมาย ดังตารางที่ 4.2

ตารางที่ 4.2

คลาสของการคัดแยกประเภทของข้อความตามกฎหมาย โดยแบ่งตามประเภทกฎหมาย

ชื่อคลาส	ชื่อกฎหมาย
A	ประมวลกฎหมายแพ่งและพาณิชย์
B	ประกาศกระทรวง
C	ประมวลกฎหมายที่ดิน
D	ประมวลกฎหมายอาญา
E	ประมวลรัษฎากร
F	พระราชบัญญัติ

ผลการทดลองการคัดแยกคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เป็นดังตารางที่ 4.3

ตารางที่ 4.3

ตารางแสดงผลการทดลองคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText

ตัววัดผล	ผลลัพธ์
Precision	0.952
Recall	0.952
F2-Score	0.952

หลังจากคัดแยกคำค้นของความทางกฎหมายแล้ว จะนำข้อความทางกฎหมายใน Corpus ที่เกี่ยวข้องกับกฎหมายที่คัดแยกได้ มาทดลองการค้นคืนข้อความทางกฎหมายโดยใช้เทคนิค roBERTa ร่วมกับเทคนิค Cosine Similarity ในการหาความคล้ายคลึงกันของข้อความ ซึ่งผลการทดลองที่ได้เป็น ดังตารางที่ 4.4

ตารางที่ 4.4

ตารางผลลัพธ์การค้นคืนข้อความทางกฎหมายด้วยเทคนิค roBERTa ร่วมกับ Cosine Similarity

ตัววัดผล	ผลลัพธ์
Precision	0.13
Recall	0.55
F2-Score	0.15
ความเร็วเฉลี่ยในการค้นคืน	0.05 วินาที

4.1.3 roBERTa

ผลจากการทดลองโดยใช้เทคนิคการค้นคืนข้อความทางกฎหมายด้วยเทคนิค roBERTa ในการทำ Word Representaion ร่วมกับเทคนิค Cosine ในการหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมายเป็น ดังตารางที่ 4.5

ตารางที่ 4.5

ตารางผลลัพธ์การค้นคืนข้อความทางกฎหมายด้วยเทคนิค roBERTa ร่วมกับ Cosine Similarity

ตัววัดผล	ผลลัพธ์
Precision	0.18
Recall	0.75
F2-Score	0.21
ความเร็วเฉลี่ยในการค้นคืน	0.15 วินาที

4.1.4 เปรียบเทียบผลการทดลองการค้นคืนข้อความทางกฎหมายระหว่างเทคนิค FastText ร่วมกับ roBERTa และ roBERTa

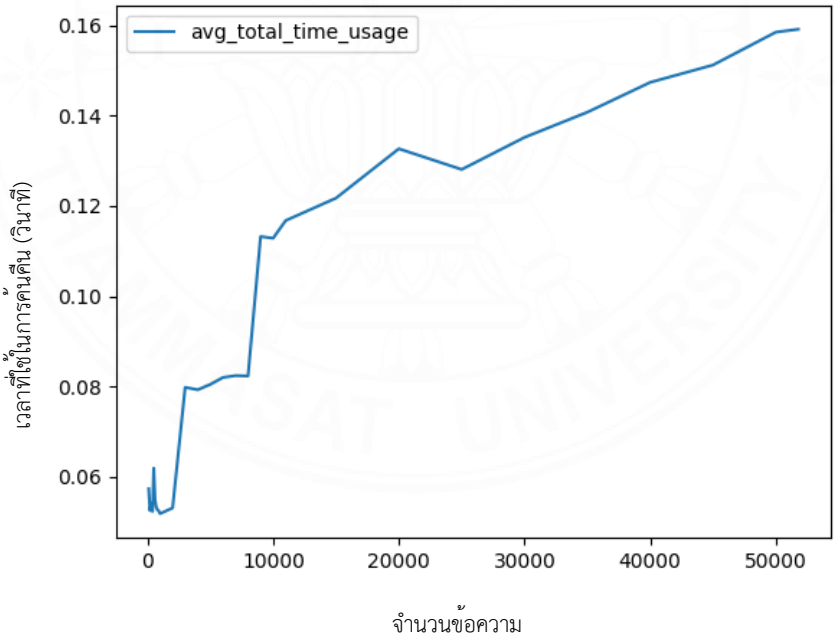
ตารางที่ 4.6

ตารางเปรียบเทียบผลลัพธ์การค้นคืนข้อความทางกฎหมาย

	FastText และ roBERTa	roBERTa
Precision	0.13	**0.18**
Recall	0.55	**0.75**
F2-Score	0.15	**0.21**
ความเร็วเฉลี่ยในการค้นคืน	**0.05 วินาที**	0.15 วินาที

ภาพที่ 4.1

ภาพกราฟแสดงเวลาที่ใช้ในการค้นคืนข้อความทางกฎหมายกับจำนวนข้อมูลข้อความทางกฎหมาย



4.2 อภิปรายผลการทดลอง

จากการทดลองค้นคืนข้อความทางกฎหมายด้วยการเพิ่มเทคนิคการคัดแยกประเภทคำค้นของข้อความทางกฎหมาย โดยจากผลการทดลองในตารางที่ 4.2 พบว่าการคัดแยกประเภทคำค้นของข้อความทางกฎหมายออกเป็น 6 ประเภทตามตารางที่ 4.1 ได้ค่า Precision ที่ 0.95 ค่า Recall 0.952 จะเห็นได้ว่าผลของการคัดแยกประเภทข้อความทางกฎหมายได้ผลลัพธ์ที่ค่อนข้างสูง จากนั้นนำการคัดแยกประเภทข้อความทางกฎหมายนี้ไปหาความคล้ายคลึงกันของคำค้น จากนั้นค้นคืนข้อความทางกฎหมายได้ผลลัพธ์ตามตารางที่ 4.3 พบว่าได้ค่า Precision ที่ 0.13 ค่า Recall ที่ 0.55 ค่า F2-Score ที่ 0.15 และเวลาที่ใช้ในการค้นคืนเฉลี่ย ที่ 0.05 วินาที

จากการทดลองการค้นคืนข้อความทางกฎหมายโดยใช้เทคนิค roBERTa เพียงอย่างเดียวผู้วิจัยพบว่าผลการทดลองในตางรางที่ 4.4 ได้ค่า Precision ที่ 0.18 ค่า Recall ที่ 0.75 F2-Score ที่ 0.21 และเวลาที่ใช้ในการค้นคืนเฉลี่ย ที่ 0.15 วินาที

จากสมมุติฐานที่ตั้งไว้ของงานวิจัยว่า การค้นคืนด้วยการใช้เทคนิคการค้นหาแบบการจับคู่ความหมาย (Semantic Matching) โดยเพิ่มเทคนิคการคัดแยกประเภทของคำค้นหาโดยเรียนรู้จากชุดข้อมูลคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกาของกฎหมายที่เกี่ยวข้องกับที่ดินที่จัดทำขึ้น สามารถให้ประสิทธิภาพในด้านความเร็วของการค้นคืนดีกว่าการค้นคืนด้วยการใช้เทคนิคการค้นหาแบบการจับคู่ความหมายเพียงอย่างเดียว แต่จากการทดลองพบว่าการใช้เทคนิคการค้นคืนด้วย roBERTa เพียงอย่างเดียวนั้น ให้ประสิทธิภาพด้านความแม่นยำและความถูกต้องของการค้นคืนที่ดีกว่า การใช้เทคนิคการคัดแยกประเภทคำค้นหาของข้อความทางกฎหมาย แต่ในด้านความเร็วในการค้นคืนข้อความทางกฎหมาย การที่เพิ่มเทคนิคการคัดแยกประเภทของคำค้นข้อความทางกฎหมายให้ผลลัพธ์ที่ดีกว่า ซึ่งอาจจะวิเคราะห์ได้ว่าชุดข้อมูลที่ใช้ทดสอบการค้นคืนข้อความทางกฎหมายนั้นมีประเภทของกฎหมายที่เกี่ยวข้องมากกว่าหนึ่งประเภท ส่งผลให้คลังข้อมูลของข้อความทางกฎหมายที่ใช้สำหรับการค้นคืนมีจำนวนลดลงหลังจากที่ผ่านการคัดแยกประเภทกฎหมาย ทำให้ประสิทธิภาพความถูกต้องและความแม่นยำของการค้นคืนข้อความทางกฎหมายลดน้อยลง

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 สรุป

ในปัจจุบันการมีสื่อออนไลน์เข้ามามีบทบาทในการเผยแพร่ข้อมูล กระจายข่าวสารต่าง ๆ ได้อย่างรวดเร็วและเป็นวงกว้าง ทำให้ได้เห็นว่าปัญหาข้อพิพาทในเรื่องต่าง ๆ นั้นมีอยู่มากมายหลากหลายประเด็นปัญหา ซึ่งปัญหาข้อพิพาทที่เกี่ยวข้องกับกรรมสิทธิ์ในที่ดินก็มีมาอย่างต่อเนื่องและยาวนาน ซึ่งบางกรณีก็กลายเป็นเรื่องราวปานปลาย การจัดการปัญหาเหล่านี้สามารถนำแนวทางการปฏิบัติที่เคยมีมาแล้วในอดีตเข้ามาปรับใช้หรืออ้างอิงได้ มีแพลตฟอร์มระบบสืบค้นในปัจจุบันที่ให้ประชาชนสามารถเข้าไปสืบค้นคำพิพากษาได้ ด้วยการใส่คำค้นหรือใส่เงื่อนไขในการสืบค้น ช่วยให้ประชาชนสามารถนำคำพิพากษาเหล่านั้นมาใช้เป็นแนวทางปฏิบัติเพื่อลดผลกระทบจากปัญหาที่เกิดขึ้นได้ แต่การสืบค้นต้องใช้คำในการสืบค้นที่ตรงกับเนื้อหาในคำพิพากษาและตั้งเงื่อนไขให้ถูกต้อง ซึ่งอาจจะเกิดปัญหาความไม่สะดวกในการสืบค้น หรือได้ข้อมูลที่ไม่ตรงกับปัญหาที่เกิดขึ้น

สารนิพนธ์ฉบับนี้จึงมีแนวคิดในการช่วยหาทางแก้ปัญหาข้อพิพาทโดยการค้นคืนข้อความทางกฎหมายที่มีลักษณะคล้ายคลึงกับข้อพิพาทนั้น ๆ เพื่อช่วยเป็นแนวทางในการจัดการแก้ไขปัญหาข้อพิพาทที่เกิดขึ้น โดยนำเทคนิคการประมวลผลทางภาษาเข้ามาช่วยในการค้นคืนข้อความทางกฎหมายเพื่อหาความคล้ายคลึงกันของข้อพิพาทกับข้อความทางกฎหมายที่เกี่ยวข้องกับเรื่องที่ดิน โดยเปรียบเทียบเทคนิคการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText ร่วมกับเทคนิคการทำ Word Representation ด้วย roBERTa และหาความคล้ายคลึงกันของข้อความทางกฎหมายด้วย Cosine Similarity เปรียบกับกับการใช้เทคนิคการค้นคืนด้วยเทคนิคการทำ Word Representation ด้วย roBERTa หาความคล้ายคลึงกันของข้อความทางกฎหมายด้วย Cosine Similarity เพียงอย่างเดียว เพื่อหาเทคนิคที่เหมาะสมโดยพิจารณาจากผลลัพธ์ที่ได้จากการทดลอง

โดยผลการทดลองผู้วิจัยพบว่าการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน นั้นวิธีการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa และหาความคล้ายคลึงกันด้วยของคำค้นกับข้อความทางกฎหมายด้วย Cosine Similarity ได้ผลลัพธ์ดีกว่าการใช้เทคนิคคัดแยกประเภทของคำค้นข้อความทางกฎหมาย แล้วใช้เทคนิคค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa ถึงแม้ว่าความเร็วของการค้นคืนข้อความทางกฎหมายด้วยการใช้สองเทคนิคควบคู่กัน แต่ในด้านประสิทธิภาพการใช้เทคนิคการค้นคืนข้อความทางกฎหมายเพียงอย่างเดียวได้ผลลัพธ์ที่ดีกว่ามาก

5.2 ปัญหาและข้อเสนอแนะ

ปัญหาที่ชุดข้อมูลที่ได้จากเว็บไซต์ deka.supremecourt.or.th นั้นเป็นข้อมูลที่ใช้วิธีการ web scraping ในการทำชุดข้อมูล ทำให้ใช้เวลานานในการจัดเตรียมข้อมูลให้เหมาะสมกับการนำมาใช้งาน ทั้งจัดการในเรื่องการหาข้อกฎหมาย การดึงในส่วนของข้อความย่อสั้นและข้อความฉบับเต็ม ส่วนการนำเข้าข้อมูลด้วยวิธีพิมพ์พบปัญหาในการเคาะวรรคและเรื่องการสะกดคำ ทำให้ต้องทำความสะอาดข้อมูลและแก้ไขคำให้เรียบร้อยก่อนนำข้อมูลมาใช้ในการทดลอง

ปัญหาที่ชุดข้อมูลบางส่วนไม่ได้อ้างถึงตัวกฎหมายที่เกี่ยวข้องกับเรื่องนั้น ๆ โดยตรงแต่ใช้การอ้างอิงคำพิพากษาศาลฎีกาแทน ทำให้ไม่สามารถนำชุดข้อมูลเหล่านี้มาใช้ในงานวิจัยได้

ปัญหาที่พบในงานวิจัยในส่วนของการจัดเตรียมข้อมูล เนื่องจากผู้วิจัยต้องทำการจับคู่ข้อมูลคำค้นที่ใช้ทดสอบกับชุดข้อมูลที่จัดทำขึ้นมาทั้งหมดเอง ด้วยการใช้วิธีหา keyword ในชุดข้อมูลกับคำค้น ทำให้บางคำค้นนั้นมีคู่ที่คล้ายคลึงกันจำนวนไม่มาก ดังนั้นควรรวบรวมชุดคำค้นขึ้นมาใหม่โดยให้มีคู่ที่คล้ายคลึงกันมากกว่านี้จะช่วยให้ได้ผลลัพธ์การทดลองที่ชัดเจนขึ้น

ปัญหาเรื่องการตัดคำด้วยการใช้ PyThaiNLP ผู้วิจัยพบว่าการตัดคำในบางส่วนนั้นให้ผลลัพธ์การตัดคำที่ไม่ถูกต้อง ซึ่งส่งผลต่อขั้นตอนการจัดเตรียมข้อมูล และส่งผลต่อการทำ Word Representation ได้ไม่ดีเท่าที่ควร ดังนั้นถ้าปรับปรุงการปรับคำโดยเรียนรู้โมเดลจากชุดข้อมูลที่เกี่ยวข้องกับข้อความทางกฎหมายจะทำให้ได้ผลลัพธ์ที่ดีขึ้น

ปัญหาในส่วนของโมเดล Pretrained-Model roBERTa ที่เลือกใช้ในงานวิจัยนั้นเรียนรู้จากข้อความที่อยู่ใน Wikipedia ซึ่งในการวิจัยนี้ข้อความจะเป็นในเชิงกฎหมายและไม่มี Pretrained-Model ที่เรียนรู้ข้อความที่เกี่ยวข้องกับด้านกฎหมาย ทำให้การทำ Word Representation ได้ผลลัพธ์ที่ไม่ดีทำที่ควร

เนื่องจากไม่มีงานวิจัยที่เกี่ยวข้องกับการค้นคืนคำค้นหาที่เกี่ยวข้องกับกฎหมายที่ดิน ทำให้ไม่สามารถเปรียบเทียบได้ว่าเทคนิคของงานวิจัยนี้นั้นเลือกใช้เทคนิคที่เหมาะสมหรือไม่ หรือสามารถนำเทคนิคอื่น ๆ ที่เหมาะสมมากกว่ามาปรับใช้ เพื่อให้ได้ผลลัพธ์ที่ดี

ข้อเสนอแนะและแนวทางที่จะพัฒนาต่อในอนาคตหลังจากที่ผู้วิจัยได้จัดทำสารนิพนธ์ฉบับนี้ สามารถทำได้โดยพิจารณาถึงข้อจำกัดของงานวิจัยในปัจจุบันแล้วนำมาปรับปรุงให้เหมาะสมมากยิ่งขึ้น เช่น การจัดทำชุดข้อมที่เกี่ยวกับข้อความทางกฎหมายโดยเฉพาะและ Transfer Learning โมเดลที่ใช้ในการทำ Word Representation เพื่อให้โมเดลสามารถตัดคำและทำ Word Embedding ได้ดีมากขึ้นกับชุดข้อมูลที่เป็นข้อความทางกฎหมาย เพื่อช่วยให้การหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมายนั้นมีประสิทธิภาพมากยิ่งขึ้น

รายการอ้างอิง

หนังสือและบทความในหนังสือ

ชัยชาญ สิทธิวิวัชธรรม, กรมที่ดิน (2562). แผนภูมิกฎหมายที่ดิน (ฉบับปรับปรุงแก้ไขใหม่) (พิมพ์ครั้งที่ 1). สำนักมาตรฐานการทะเบียนที่ดิน

สำนักมาตรฐานการทะเบียนที่ดิน, กรมที่ดิน (2551). แนวคำวินิจฉัยกรมที่ดิน เรื่องการจดทะเบียนสิทธิเกี่ยวกับอสังหาริมทรัพย์ซึ่งได้มาโดยทางมรดก (พิมพ์ครั้งที่ 1). สำนักมาตรฐานการทะเบียนที่ดิน

สำนักมาตรฐานการทะเบียนที่ดิน, กรมที่ดิน (2556). คู่มือแนวคำวินิจฉัยปัญหาการจดทะเบียนสิทธิและนิติกรรมเกี่ยวกับอสังหาริมทรัพย์ตามประมวลกฎหมายที่ดิน เล่ม 2 (พิมพ์ครั้งที่ 1). สำนักมาตรฐานการทะเบียนที่ดิน

สำนักมาตรฐานการทะเบียนที่ดิน, กรมที่ดิน (2562). คู่มือแนวคำวินิจฉัยปัญหาการจดทะเบียนสิทธิและนิติกรรมเกี่ยวกับอสังหาริมทรัพย์ตามประมวลกฎหมายที่ดิน เล่ม 3 (พิมพ์ครั้งที่ 1). สำนักมาตรฐานการทะเบียนที่ดิน

บทความวารสาร

Weerayut Krungklang., Sukree Sinthupinyo. (2020). An Analysis of Natural Language Text Relating to Thai Criminal Law. doi: 10.1109/ECAI50035.2020.9223143

Zhonghao Li. (2019). A Classification Retrieval Approach for English Legal Texts. doi: 10.1109/ICITBS.2019.00059

Zhaoxu Yang, Jike Ge, Tingkai Hu, Wencheng Yu, Yujie Zheng, Yan Dong. (2022). Text Classification of Judgement Documents Considering Sample Imbalance. doi: 10.1109/iciea54703.2022.10006051

May TamaraStow, ChidiebereUgwu, Laeticia Onyejegbu. (2023). An Improved Model for Legal Case Text Document Classification. European Journal of Electrical Engineering and Computer Science, doi: 10.24018/ejece.2023.7.2.509

Phi Manh Kien, Ha-Thanh Nguyen, Ngo Xuan Bach, Vu Tran, Minh Le Nguyen, Tu Minh Phuong. (2020). Answering Legal Questions by Learning Neural Attentive Text Representation. doi: 10.18653/V1/2020.COLING-MAIN.86

Nhat-Minh Pham, Ha-Thanh Nguyen, Trong-Hop Do. (2022). Multi-stage Information Retrieval for Vietnamese Legal Texts. doi: 10.48550/arxiv.2209.14494

Md. Mushfiqur Rahman, Zahin Azmaeen, Mithila Arman, Md Latifur Rahman, Md Moinul Hoque. (2022). An Intelligent Approach Towards Legal Text-Documents Retrieval. doi: 10.1109/iccit57492.2022.10054858

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. (2019). ROBERTA: A robustly optimized BERT pretraining approach. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1907.11692

Lalita Lowphansirikul, Charin Polpanumas, Nawat Jantrakulchai, Sarana Nutanong (2021). WangchanBERTa: Pretraining transformer-based Thai Language Models. arXiv (Cornell University). https://arxiv.org/pdf/2101.09635

Tianyu Gao, Xingcheng Yao, Danqi Chen (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings. arXiv (Cornell University). https://arxiv.org/pdf/2104.0882

Peerat Limkonchotiwat, Wuttikorn Ponwitayarat, Lalita Lowphansirikul, Can Udomcharoenchaikit, Ekapol Chuangsuwanich, Sarana Nutanong (2023). An Efficient Self-Supervised Cross-View Training For Sentence Embedding. arXiv (Cornell University). https://arxiv.org/pdf/2311.03228

สื่ออิเล็กทรอนิกส์

ระบบสืบค้นคำพิพากษา คำสั่งคำร้องและคำวินิจฉัยศาลฎีกา. สืบค้นเมื่อวันที่ 8 เมษายน 2567 จาก https://deka.supremecourt.or.th

FastText Library for Efficient Text Classification and Representation learning. . Retrieved April 10, 2024, from https://fasttext.cc/

Fine-Tuning Pre-Trained Language Models. Retrieved April 15, 2024, from https://huggingface.co/

ภาคผนวก

ภาคผนวก ก
รายละเอียดการทดลอง

ในส่วนนี้เป็นส่วนของการแสดงรายละเอียดในการทดลองในการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน โดยเริ่มจากขั้นตอนการเก็บรวบรวมชุดข้อมูล และผลการทดลองโมเดลการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText รวมถึงผลการทดลองการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa โดยใช้ Pretrained-Model Wangchanberta ที่ผ่านการ Finetuned ด้วยเทคนิคต่าง ๆ และเว็บไซต์การค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน โดยการใช้เทคนิคของงานวิจัยนี้

ก.1 ผลการทดลองการเรียนรู้ของโมเดลการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastTex ในแต่ละรอบด้วย n-gram ที่ต่างกัน

ก.1.1 การทดลองการเรียนรู้ของโมเดลในรอบที่ 1 เมื่อใช้ n-gram ที่ 1

ตารางที่ ก.1

ตารางแสดงผลการทดลองการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เมื่อใช้ n-gram ที่ 1

ตัววัดผล	ผลลัพธ์
Precision	0.952
Recall	0.953
F2-Score	0.952

ก.1.2 การทดลองการเรียนรู้ของโมเดลในรอบที่ 2 เมื่อใช้ n-gram ที่ 2

ตารางที่ ก.2

ตารางแสดงผลการทดลองการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เมื่อใช้ n-gram ที่ 2

ตัววัดผล	ผลลัพธ์
Precision	0.905
Recall	0.907
F2-Score	0.905

ก.1.3 การทดลองการเรียนรู้ของโมเดลในรอบที่ 3 เมื่อใช้ n-gram ที่ 3

ตารางที่ ก.3

ตารางแสดงผลการทดลองการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เมื่อใช้ n-gram ที่ 3

ตัววัดผล	ผลลัพธ์
Precision	0.902
Recall	0.903
F2-Score	0.902

ก.1.4 การทดลองการเรียนรู้ของโมเดลในรอบที่ 4 เมื่อใช้ n-gram ที่ 4

ตารางที่ ก.4

ตารางแสดงผลการทดลองการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เมื่อใช้ n-gram ที่ 4

ตัววัดผล	ผลลัพธ์
Precision	0.897
Recall	0.898
F2-Score	0.897

ก.1.5 การทดลองการเรียนรู้ของโมเดลในรอบที่ 5 เมื่อใช้ n-gram ที่ 5

ตารางที่ ก.5

ตารางแสดงผลการทดลองการคัดแยกประเภทคำค้นของข้อความทางกฎหมายด้วยเทคนิค FastText เมื่อใช้ n-gram ที่ 5

ตัววัดผล	ผลลัพธ์
Precision	0.896
Recall	0.896
F2-Score	0.896

ก.2 ผลการทดลองการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa กับ Pretrained-Model Wangchanberta ที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased โดยอาศัยวิธีการหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมาย

ก.2.1 การทดลองรอบที่ 1 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.9

ภาพที่ ก.1

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.9 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.9
---------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.01
AVG F2-Score:                   0.05
```

ก.2.2 การทดลองรอบที่ 2 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.8

ภาพที่ ก.2

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.8 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.8
---------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.026
AVG F2-Score:                   0.13
```

ก.2.3 การทดลองรอบที่ 3 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.7

ภาพที่ ก.3

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.7 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.7
---------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.125
AVG F2-Score:                   0.625
```

ก.2.4 การทดลองรอบที่ 4 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.6

ภาพที่ ก.4

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.6 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.6
---------------------------------------
AVG PRECISION:                  0.42
AVG RECALL:                     0.29
AVG F2-Score:                   0.385
```

ก.2.5 การทดลองรอบที่ 5 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.55

ภาพที่ ก.5

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.55 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.55
---------------------------------------
AVG PRECISION:                  0.219
AVG RECALL:                     0.397
AVG F2-Score:                   0.241
```

ก.2.6 การทดลองรอบที่ 6 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.52

ภาพที่ ก.6

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.52 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

```
cosine similarity threshold:    0.52
---------------------------------------
AVG PRECISION:                  0.173
AVG RECALL:                     0.454
AVG F2-Score:                   0.197
```

ก.2.7 การทดลองรอบที่ 7 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.5

ภาพที่ ก.7

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.5 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค base-atts-spm-uncased

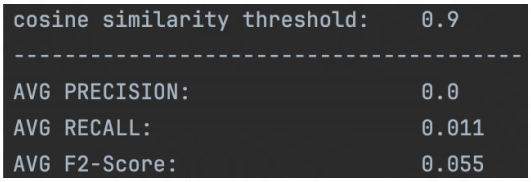
```
cosine similarity threshold:    0.5
---------------------------------------
AVG PRECISION:                  0.15
AVG RECALL:                     0.485
AVG F2-Score:                   0.174
```

ก.3 ผลการทดลองการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa กับ Pretrained-Model Wangchanberta ที่ผ่านการ Finetuned กับชุดข้อมูล sanook-news โดยอาศัยวิธีการหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมาย

ก.3.1 การทดลองรอบที่ 1 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.9

ภาพที่ ก.8

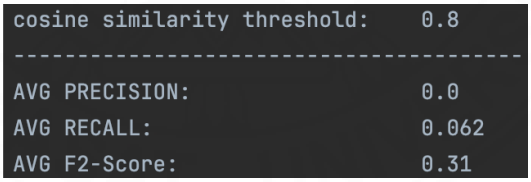
ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.9 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news



ก.3.2 การทดลองรอบที่ 2 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.8

ภาพที่ ก.9

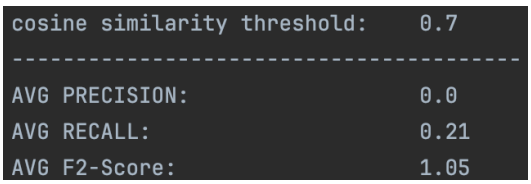
ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.8 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news



ก.3.3 การทดลองรอบที่ 3 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.7

ภาพที่ ก.10

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.7 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news



ก.3.4 การทดลองรอบที่ 4 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.6

ภาพที่ ก.11

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.6 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news

```
cosine similarity threshold:    0.6
---------------------------------------
AVG PRECISION:                  0.531
AVG RECALL:                     0.4
AVG F2-Score:                   0.498
```

ก.3.5 การทดลองรอบที่ 5 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.55

ภาพที่ ก.12

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.55 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news

```
cosine similarity threshold:    0.55
---------------------------------------
AVG PRECISION:                  0.381
AVG RECALL:                     0.489
AVG F2-Score:                   0.399
```

ก.3.6 การทดลองรอบที่ 6 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.52

ภาพที่ ก.13

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.52 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news

```
cosine similarity threshold:    0.52
---------------------------------------
AVG PRECISION:                  0.321
AVG RECALL:                     0.548
AVG F2-Score:                   0.35
```

ก.3.7 การทดลองรอบที่ 7 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.5

ภาพที่ ก.14

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.5 ด้วยโมเดลที่ผ่านการ Finetuned ชุดข้อมูล sanook-news

```
cosine similarity threshold:    0.5
----------------------------------------
AVG PRECISION:                  0.286
AVG RECALL:                     0.573
AVG F2-Score:                   0.318
```

ก.4 ผลการทดลองการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa กับ Pretrained-Model Wangchanberta ที่ผ่านการ Finetuned ด้วยเทคนิค SCT โดยอาศัยวิธีการหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมาย

ก.4.1 การทดลองรอบที่ 1 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.9

ภาพที่ ก.15

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.9 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.9
----------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.043
AVG F2-Score:                   0.215
```

ก.4.2 การทดลองรอบที่ 2 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.8

ภาพที่ ก.16

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.8 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.8
----------------------------------------
AVG PRECISION:                  0.949
AVG RECALL:                     0.311
AVG F2-Score:                   0.673
```

ก.4.3 การทดลองรอบที่ 3 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.7

ภาพที่ ก.17

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.7 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.7
----------------------------------------
AVG PRECISION:                  0.863
AVG RECALL:                     0.575
AVG F2-Score:                   0.784
```

ก.4.4 การทดลองรอบที่ 4 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.6

ภาพที่ ก.18

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.6 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.6
----------------------------------------
AVG PRECISION:                  0.301
AVG RECALL:                     0.705
AVG F2-Score:                   0.34
```

ก.4.5 การทดลองรอบที่ 5 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.55

ภาพที่ ก.19

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.55 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.55
----------------------------------------
AVG PRECISION:                  0.219
AVG RECALL:                     0.737
AVG F2-Score:                   0.255
```

ก.4.6 การทดลองรอบที่ 6 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.52

ภาพที่ ก.20

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.52 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.52
---------------------------------------
AVG PRECISION:                  0.193
AVG RECALL:                     0.744
AVG F2-Score:                   0.227
```

ก.4.7 การทดลองรอบที่ 7 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.5

ภาพที่ ก.21

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.5 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค SCT

```
cosine similarity threshold:    0.5
---------------------------------------
AVG PRECISION:                  0.182
AVG RECALL:                     0.747
AVG F2-Score:                   0.214
```

ก.5 ผลการทดลองการค้นคืนข้อความทางกฎหมายด้วยการใช้เทคนิค roBERTa กับ Pretrained-Model Wangchanberta ที่ผ่านการ Finetuned ด้วยเทคนิค simcse โดยอาศัยวิธีการหาความคล้ายคลึงกันของคำค้นกับข้อความทางกฎหมาย

ก.5.1 การทดลองรอบที่ 1 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.9

ภาพที่ ก.22

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.9 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.9
---------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.01
AVG F2-Score:                   0.05
```

ก.5.2 การทดลองรอบที่ 2 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.8

ภาพที่ ก.23

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.8 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.8
----------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.036
AVG F2-Score:                   0.18
```

ก.5.3 การทดลองรอบที่ 3 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.7

ภาพที่ ก.24

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.7 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.7
----------------------------------------
AVG PRECISION:                  0.0
AVG RECALL:                     0.149
AVG F2-Score:                   0.745
```

ก.5.4 การทดลองรอบที่ 4 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.6

ภาพที่ ก.25

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.6 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.6
----------------------------------------
AVG PRECISION:                  0.497
AVG RECALL:                     0.335
AVG F2-Score:                   0.453
```

ก.5.5 การทดลองรอบที่ 5 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.55

ภาพที่ ก.26

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.55 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.55
----------------------------------------
AVG PRECISION:                  0.301
AVG RECALL:                     0.429
AVG F2-Score:                   0.32
```

ก.5.6 การทดลองรอบที่ 6 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.52

ภาพที่ ก.27

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.52 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

```
cosine similarity threshold:    0.52
----------------------------------------
AVG PRECISION:                  0.261
AVG RECALL:                     0.48
AVG F2-Score:                   0.287
```

ก.5.7 การทดลองรอบที่ 7 เมื่อกำหนดค่าความคล้ายคลึงกันที่ 0.5

ภาพที่ ก.28

ภาพผลการทดลองการค้นคืนข้อความทางกฎหมาย ที่ค่าความคล้ายคลึง 0.5 ด้วยโมเดลที่ผ่านการ Finetuned ด้วยเทคนิค simcse

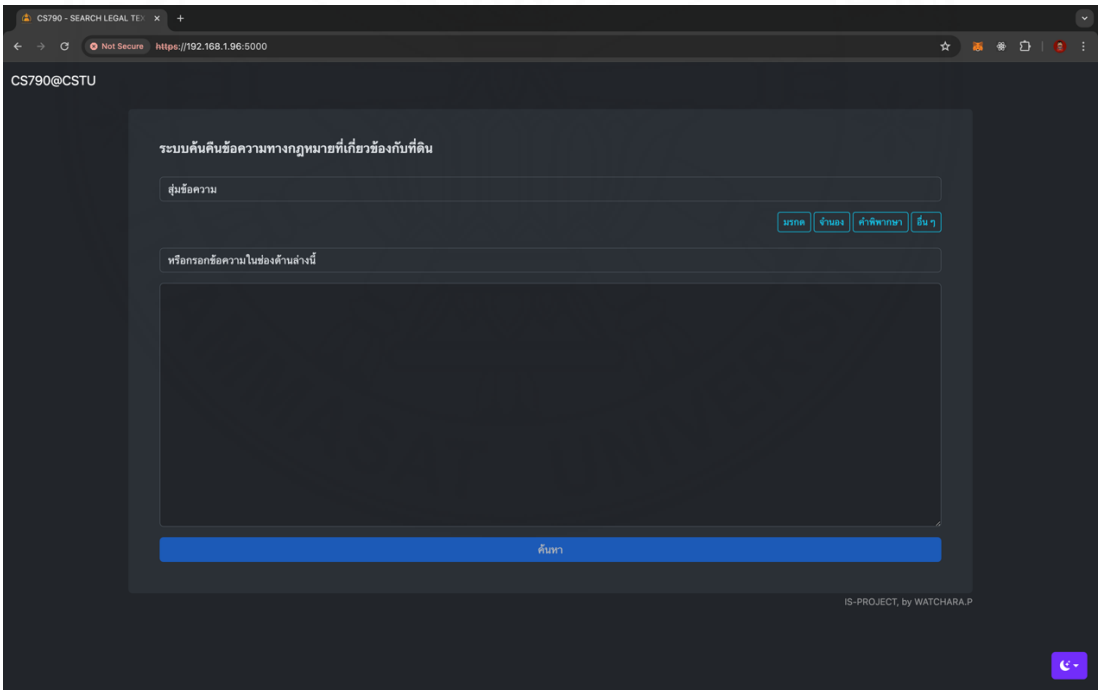
```
cosine similarity threshold:    0.5
----------------------------------------
AVG PRECISION:                  0.24
AVG RECALL:                     0.514
AVG F2-Score:                   0.269
```

ก.6 เว็บไซต์ตัวอย่างระบบการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน ด้วยเทคนิคที่ใช้ในงานวิจัยนี้

ผู้จัดทำได้สร้างเว็บไซต์ตัวอย่างสำหรับระบบการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน โดยใช้เทคนิคในงานวิจัย โดยผู้ใช้สามารถกรอกข้อความในกล่องข้อความหรือสามารถสุ่มข้อความทางกฎหมายที่เกี่ยวกับเรื่องต่าง ๆ ได้ด้วยการกดที่ปุ่ม มรกด, จำนอง, คำพิพากษา, หรือ อื่น ๆ จากนั้นกดปุ่มค้นหา ซึ่งระบบจะแสดงผลลัพธ์การค้นคืนข้อความทางกฎหมาย โดยแสดงระยะเวลาที่ใช้ในการค้นคืนข้อความทางกฎหมาย และแสดงผลลัพธ์ของการค้นคืนทั้งหมด 20 อันดับแรก โดยเรียงลำดับคะแนนค่าความคล้ายคลึงกันด้วยเทคนิค Cosine Similarity จากมากไปน้อย

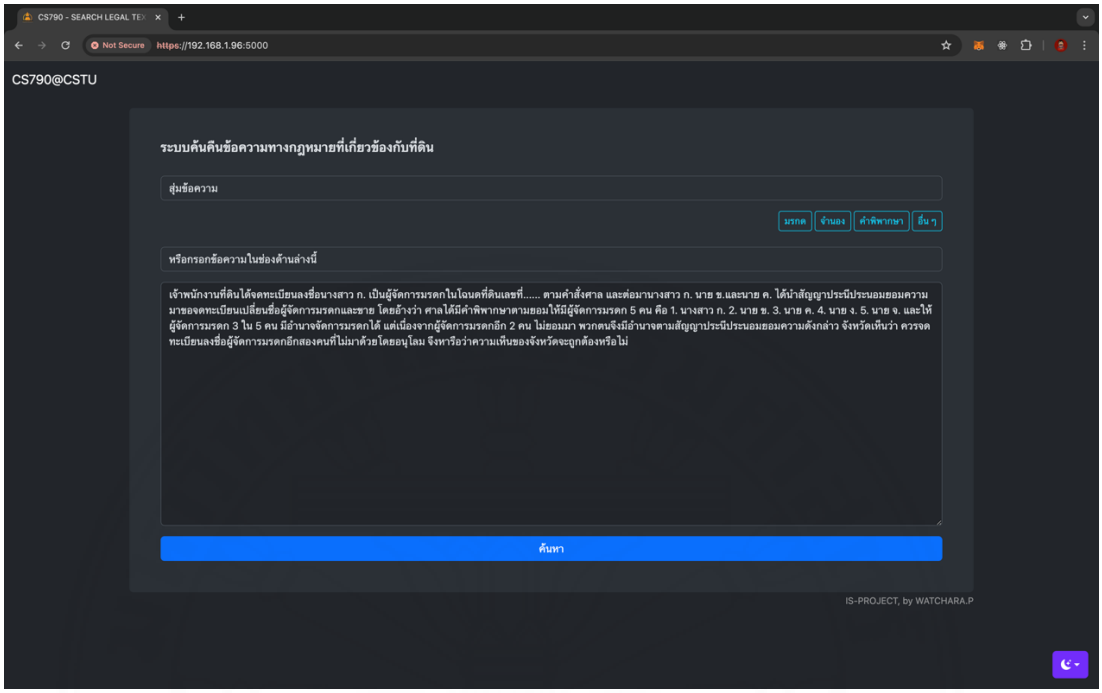
ภาพที่ ก.29

ภาพหน้าแรกของเว็บไซต์ตัวอย่างระบบการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน



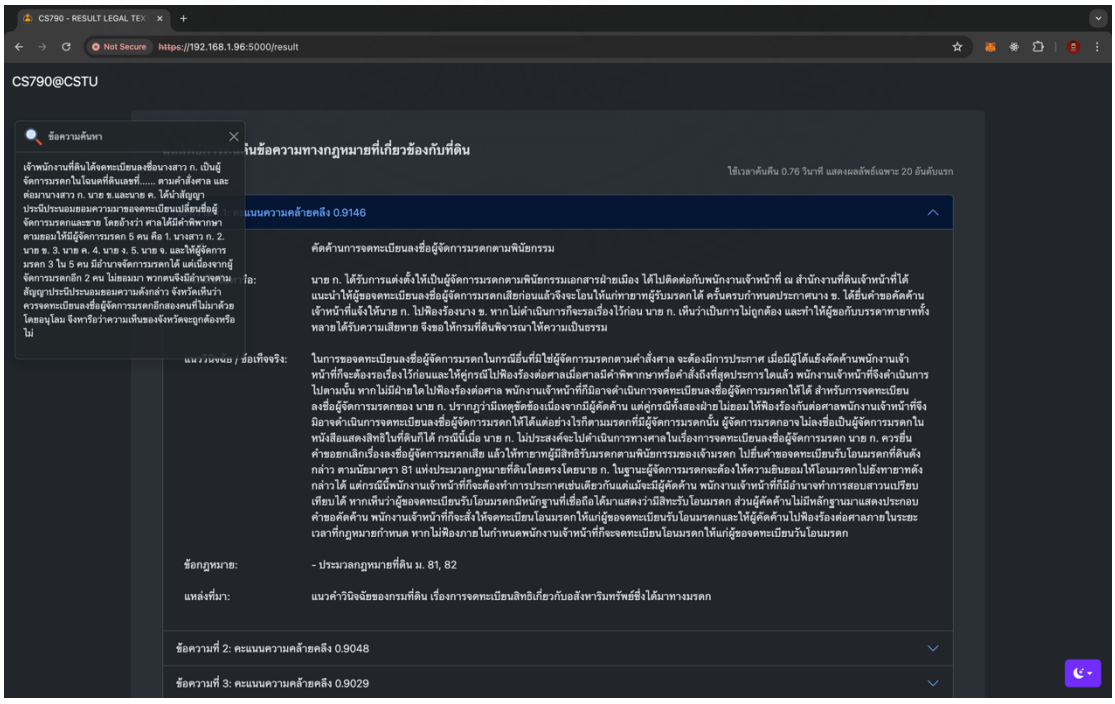
ภาพที่ ก.30

ภาพตัวอย่างข้อความสุ่มขึ้นมาจากระบบการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน



ภาพที่ ก.31

ภาพผลลัพธ์ที่ได้การระบบการค้นคืนข้อความทางกฎหมายที่เกี่ยวข้องกับที่ดิน



ประวัติผู้เขียน

ชื่อ	วัชระ โพธิ์รุ้ง
วุฒิการศึกษา	ปีการศึกษา 2562: ปริญญาตรี สาขาวิศวกรรมคอมพิวเตอร์ วิศวกรรมศาตรบัณฑิต มหาวิทยาลัยเชียงใหม่
ตำแหน่ง	นักวิชาการคอมพิวเตอร์ สำนักเทคโนโลยีสารสนเทศ กรมที่ดิน