

บทความวิจัย

ระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ด้วยปัญญาประดิษฐ์ โดยใช้เทคนิคเชิงลึกและดัดคำกลั่นซึ่ง

ปกรณ์ กัลปดี^{1*}, ปิยะวัฒน์ อัฒจักร², อัจฉริยา เหล่าศิริ³

คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี^{1,3}

ศูนย์ความเป็นเลิศด้านปัญญาประดิษฐ์ คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี²

อีเมล: pakorn.k@ubru.ac.th^{1*}, piyawat.a@ubru.ac.th², atchariya.l@ubru.ac.th³

บทคัดย่อ

การวิจัยครั้งนี้เป็นการวิจัยและพัฒนาสารสนเทศ โดยประยุกต์ใช้แบบจำลองสำเร็จรูป ซึ่งมีวัตถุประสงค์เพื่อ 1) พัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย โดยใช้เทคนิค Data Cleansing, Text Embedding, Cosine Similarity และ 2) ศึกษาประสิทธิภาพของระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย กลุ่มตัวอย่างที่ใช้ในการทดสอบเป็นงานวิจัยจำนวน 10 เรื่อง วิเคราะห์ข้อมูลตามเกณฑ์ความคล้ายคลึงเชิงอรรถศาสตร์ที่แสดงเป็นคำร้อยละและการวิเคราะห์เชิงเนื้อหา

ผลการวิจัยพบว่า ระบบสารสนเทศที่พัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชันสามารถตอบสนองความต้องการของผู้ใช้งานได้อย่างมีประสิทธิภาพ โดยกระบวนการเตรียมข้อมูลด้วยเทคนิค Data Cleansing พบว่าช่วยลดข้อมูลรบกวนและเพิ่มความถูกต้องในการประมวลผลอย่างมีนัยสำคัญ อีกทั้งการประยุกต์ใช้เทคนิค Text Embedding และ Cosine Similarity ทำให้ระบบสามารถตรวจจับความซ้ำซ้อนในเชิงบริบทและความหมายเชิงลึกได้แม่นยำกว่าการเปรียบเทียบคำสำคัญแบบเดิม ผลการทดสอบประสิทธิภาพพบว่า ระบบสามารถวิเคราะห์และจัดอันดับความซ้ำซ้อนของหัวข้องานวิจัยได้สอดคล้องตามเกณฑ์ที่กำหนด ดังนั้น ระบบสารสนเทศดังกล่าวจึงมีความเหมาะสมที่จะนำไปประยุกต์ใช้ในระดับบัณฑิต ทั้งในด้านการวางแผนการดำเนินงาน โดยใช้เป็นเครื่องมือสนับสนุนการจัดการข้อมูลงานวิจัยอย่างเป็นระบบ รวมทั้งด้านการสนับสนุนการตัดสินใจ โดยบูรณาการข้อมูลเพื่อเป็นฐานประกอบการตัดสินใจอนุมัติทุนสนับสนุนโครงการวิจัยเรื่องใหม่ ส่งผลให้สามารถบริหารจัดการงบประมาณและทรัพยากรการวิจัยได้อย่างเหมาะสม

คำสำคัญ: ระบบสารสนเทศ, เทคโนโลยีปัญญาประดิษฐ์, เทคนิคเชิงลึก, ดัดคำกลั่นซึ่ง

Received: 11 May, 2026, Revised: 25 Jun, 2026, Accepted: 27 Jun, 2026

* Corresponding author

AN ARTIFICIAL INTELLIGENCE-BASED INFORMATION SYSTEM FOR RESEARCH DUPLICATION DETECTION USING DEEP LEARNING AND DATA CLEANSING TECHNIQUES

Pakorn Kallapadee^{1*}, Piyawat Adthajak², Atchariya Laosiri³

Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University ^{1*,3}

Center of Excellence in Artificial Intelligence, Faculty of Industrial Technology,

Ubon Ratchathani Rajabhat University²

Email: pakorn.k@ubru.ac.th^{1*}, piyawat.a@ubru.ac.th², atchariya.l@ubru.ac.th³

Abstract

This research focused on the design and development of an information system to detect duplication in research proposal submissions. The system was developed by applying a pre-built model and integrating advanced techniques, specifically Data Cleansing, Text Embedding, and Cosine Similarity. The research objectives were to: 1) develop an information system for detecting research duplication, and 2) evaluate the system's performance. The testing process utilized a dataset of 10 research projects, with the collected data analyzed using semantic similarity criteria, expressed as percentages, and content analysis.

The results indicate that the web-based information system effectively meets user requirements. The implementation of Data Cleansing significantly reduced data noise and enhanced processing accuracy. Furthermore, the application of Text Embedding and Cosine Similarity enabled the system to detect duplication based on contextual and semantic meaning with greater precision than traditional keyword matching methods. Performance evaluations demonstrated that the system could accurately analyze and rank research topic duplication in accordance with established standards. Consequently, the system is suitable for organizational implementation, serving as an effective tool for systematic research data management and supporting administrative decision-making for new research funding allocations, thereby optimizing the management of institutional budgets and research resources.

Keywords: Information System, Artificial Intelligence, Deep Learning Techniques, Data Cleansing

บทนำ (Introduction)

ในบริบทปัจจุบันที่โลกก้าวเข้าสู่ยุคแห่งเทคโนโลยี งานวิจัยและนวัตกรรมถือเป็นกลไกสำคัญในการขับเคลื่อนการพัฒนาประเทศในทุกมิติ ทั้งด้านเศรษฐกิจ สังคมและสิ่งแวดล้อม สถาบันอุดมศึกษาในฐานะแหล่งบ่มเพาะทางวิชาการ จึงมีบทบาทสำคัญยิ่งในการสร้างสรรค์ผลงานวิจัยที่มีคุณภาพและสามารถนำสารสนเทศไปประยุกต์ใช้ให้เกิดประโยชน์สูงสุด โดยระบบสารสนเทศเป็นระบบที่รวบรวมสารสนเทศที่เกี่ยวข้องกับการดำเนินงานขององค์กรให้กับผู้ปฏิบัติงานและผู้บริหาร เพื่อสนับสนุนภารกิจที่รับผิดชอบ โดยใช้เทคโนโลยีและอุปกรณ์สมัยใหม่เพื่อสร้างสารสนเทศที่มีประโยชน์ต่อผู้ใช้ เพื่อให้การดำเนินการขององค์กรเป็นไปอย่างราบรื่นและมีประสิทธิภาพ (โอภาส เอี่ยมสิริวงศ์, 2560) อย่างไรก็ตาม ในช่วง 20 ปีที่ผ่านมา จำนวนสถาบันที่พึ่งพาระบบคอมพิวเตอร์ในการดำเนินงานทางอิเล็กทรอนิกส์มีจำนวนเพิ่มขึ้น เนื่องจากองค์กรส่วนใหญ่พึ่งพาเทคโนโลยีในการดำเนินงาน จัดเก็บข้อมูลและนำเสนอข้อมูลให้กับผู้ใช้งาน (พิรพัฒน์ หมั่นจิตร และฐิตาภรณ์ สินจรรย์ศักดิ์, 2568)

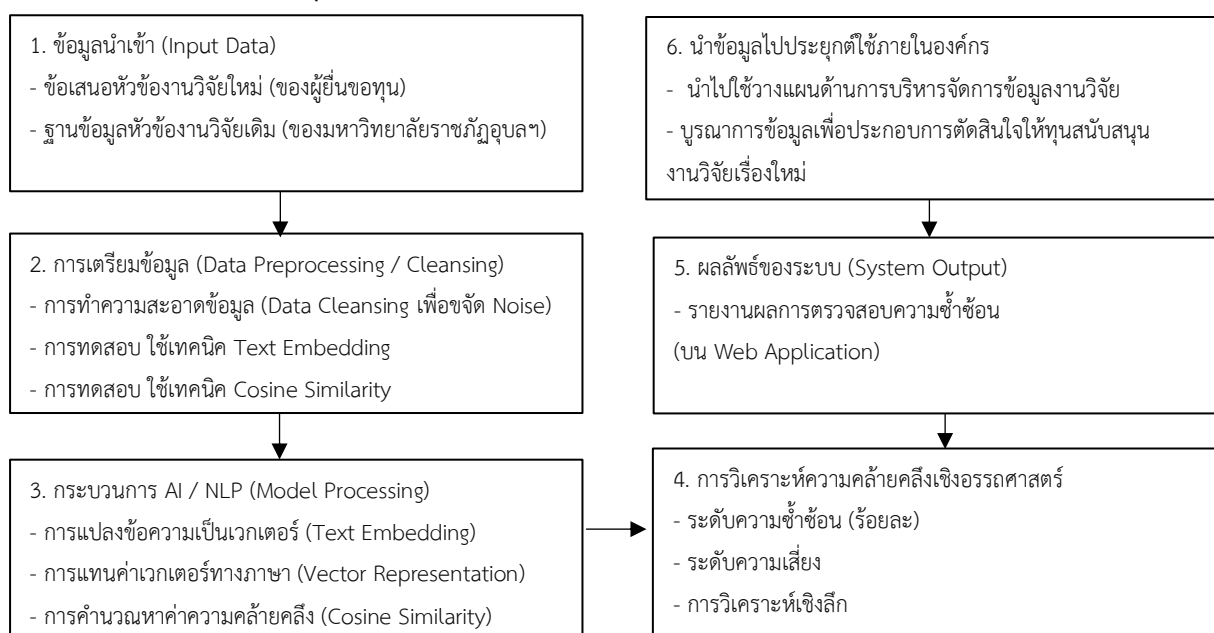
ระบบสารสนเทศเป็นกลไกที่สำคัญต่อการยกระดับสมรรถนะการดำเนินงานขององค์กร โดยทำหน้าที่หลักในการประมวลผลข้อมูลและนำผลลัพธ์ที่ได้มาประยุกต์ใช้ในการวางแผนเชิงยุทธศาสตร์ ซึ่งไม่เพียงแต่ช่วยเพิ่มประสิทธิภาพในการทำงาน แต่ยังสร้างความได้เปรียบทางการแข่งขันอย่างยั่งยืน ด้วยเหตุนี้ จึงมีการนำเทคโนโลยีคอมพิวเตอร์มาพัฒนาระบบสารสนเทศเพื่อสนับสนุนกระบวนการจัดเก็บ เรียกใช้และประมวลผลข้อมูล ซึ่งช่วยลดภาระงานและระยะเวลาในการปฏิบัติงานได้อย่างมีนัยสำคัญ ครอบคลุมไปถึงการบริหารจัดการภายในสถาบันการศึกษาที่เริ่มมีการนำเทคโนโลยีดิจิทัลมาประยุกต์ใช้อย่างแพร่หลาย ในบริบทของสถาบันอุดมศึกษา ภารกิจหลักที่สำคัญคือการส่งเสริมและสนับสนุนให้คณาจารย์ผลิตผลงานวิจัยและสร้างองค์ความรู้ใหม่ เพื่อถ่ายทอดประโยชน์สู่ผู้เรียน ชุมชนและสังคมโดยรวม สถาบันอุดมศึกษาจึงมุ่งเน้นกระตุ้นให้เกิดการดำเนินงานวิจัยอย่างต่อเนื่อง เพื่อให้เกิดประโยชน์สูงสุดต่อพันธกิจดังกล่าว ส่งผลให้หลายสถาบันเล็งเห็นการนำระบบสารสนเทศมาใช้เป็นเครื่องมือในการบริหารจัดการงานวิจัย (สุรพงษ์ วิริยะ และคณะ, 2567) อย่างไรก็ตาม การบริหารจัดการงานวิจัยจำเป็นต้องอาศัยระบบที่มีประสิทธิภาพและกลไกการตรวจสอบที่รัดกุม เนื่องจากปัญหาความซ้ำซ้อนของการเสนอหัวข้องานวิจัยถือเป็นข้อจำกัดสำคัญที่ส่งผลกระทบโดยตรงต่อประสิทธิภาพในการจัดสรรงบประมาณและขีดขวางการต่อยอดองค์ความรู้ใหม่ หากปราศจากกระบวนการตรวจสอบที่แม่นยำและเป็นระบบ ย่อมนำไปสู่ความสูญเสียทรัพยากรโดยเปล่าประโยชน์และลดทอนคุณค่าทางวิชาการของงานวิจัย

จากการศึกษาสภาพการดำเนินงานของสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏอุบลราชธานี ซึ่งเป็นหน่วยงานพันธกิจหลักในการส่งเสริมและบริหารจัดการงานวิจัยระดับท้องถิ่นและระดับชาติ พบว่ากระบวนการบริหารจัดการในปัจจุบันยังคงเผชิญข้อจำกัดในหลายมิติ โดยเฉพาะการขาดแคลนระบบสารสนเทศแบบบูรณาการที่มีประสิทธิภาพในการเชื่อมโยงข้อมูลระหว่างนักวิจัย บุคลากรสายสนับสนุนและผู้บริหาร รวมไปถึงการไม่มีระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ข้อจำกัดเหล่านี้ส่งผลให้การเข้าถึงข้อมูล การสืบค้นหัวข้องานวิจัยย้อนหลัง ตลอดจนการติดตามสถานะการดำเนินงานผ่านระบบเครือข่ายเป็นไปอย่างยากลำบากและต้องพึ่งพาระบบเอกสารสิ่งพิมพ์ซึ่งก่อให้เกิดความสิ้นเปลืองทรัพยากร ทั้งนี้ ปัญหาหลักที่จำเป็นต้องได้รับการแก้ไขอย่างเร่งด่วน คือ ความซ้ำซ้อนในการตรวจสอบความซ้ำซ้อนของหัวข้องานวิจัยที่เสนอใหม่เทียบกับฐานข้อมูลเดิม ซึ่งปัจจุบันยังคงตรวจสอบโดยใช้ทรัพยากรบุคคล (Manual Review) ที่มีข้อจำกัดทั้งด้านเวลาและปริมาณงาน เนื่องจากการบันทึกข้อมูลและตรวจทานข้อมูลแบบเดิมยังไม่สามารถนำข้อมูลมาวิเคราะห์และใช้ในการวางแผนได้ อีกทั้ง การค้นหาข้อมูลที่เกี่ยวข้องกับงานวิจัยในปัจจุบันยังไม่สามารถที่จะตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ซึ่งการนำระบบสารสนเทศเข้ามาช่วยในการบริหารจัดการจะทำให้สามารถใช้ประโยชน์จาก

ข้อมูลงานวิจัยได้อย่างมีประสิทธิภาพยิ่งขึ้น ประกอบกับปัจจุบันการเขียนงานวิจัยมีการดัดแปลงรูปประโยคหรือการใช้คำพ้องความหมาย (Synonyms) ที่มีความหมายเชิงอรรถศาสตร์ (Semantic Meaning) คล้ายคลึงกัน ส่งผลให้การสืบค้นจากคำสำคัญ (Keyword Search) แบบเดิมไม่สามารถตรวจจับความซ้ำซ้อนในระดับความหมายเชิงลึกได้อย่างมีประสิทธิภาพ ซึ่งอาจนำมาสู่ปัญหาการคัดลอกผลงานวิจัย สอดคล้องกับการศึกษาของ กรธัช อยู่สุข (2557) ที่ระบุว่าการเผยแพร่งานวิจัยซ้ำซ้อนในประเทศไทยเป็นประเด็นจริยธรรมสำคัญที่ควรเร่งแก้ไข แต่อย่างไรก็ตาม ในปัจจุบันมหาวิทยาลัยราชภัฏอุบลราชธานียังไม่มีระบบสารสนเทศขึ้นมาเพื่อแก้ปัญหาี้ จากการทบทวนวรรณกรรมในอดีต อาทิ การศึกษาประสิทธิภาพระบบสารสนเทศเพื่อการจัดการงานวิจัย มหาวิทยาลัยราชภัฏเพชรบุรี (กมลดา เรื่องอร่าม, 2566) และมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร (เมธิกา พ่วงแสง และวิสุตา วรรณห้วย, 2562) รวมถึงการศึกษาล่าสุดของมหาวิทยาลัยราชภัฏเชียงราย (เดชชนะ ศรีรัตนลัม และธิดารัตน์ สุขประภากรณ์, 2568) จะชี้ให้เห็นเด่นชัดว่า การพัฒนาระบบดังกล่าวช่วยเพิ่มประสิทธิภาพการบริหารจัดการงานวิจัย ลดความซ้ำซ้อนในกระบวนการทำงาน ส่งเสริมความร่วมมือระหว่างภาคส่วนและสนับสนุนการตัดสินใจเชิงนโยบายอย่างยั่งยืน ผู้วิจัยจึงมุ่งแก้ปัญหาดังกล่าวโดยการจัดสร้างระบบสารสนเทศเพื่อเข้ามาจัดการข้อมูลงานวิจัยให้ก้าวสู่ระบบดิจิทัล

จากประเด็นปัญหาดังกล่าว ผู้วิจัยจึงพัฒนาระบบสารสนเทศเพื่องานวิจัยในรูปแบบเว็บแอปพลิเคชัน (Web Application) ด้วยเทคโนโลยีปัญญาประดิษฐ์ โดยมุ่งเน้นการประยุกต์ใช้วิทยาการการเรียนรู้เชิงลึก (Deep Learning) ผ่านเทคนิค Text Embedding ในการแปลงข้อความเป็นเวกเตอร์และตัดสินความซ้ำซ้อนผ่านค่าความคล้ายคลึง (Cosine Similarity) ควบคู่กับการใช้เทคนิคการทำความสะอาดข้อมูล (Data Cleansing) เพื่อขจัดข้อมูลที่ไม่สมบูรณ์หรือข้อมูลรบกวน (Noise) ออกจากฐานข้อมูลสำหรับกระบวนการวัดผล ซึ่งการประเมินรอบด้านนี้จะช่วยยกระดับมาตรฐานการบริหารจัดการงานวิจัยของสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏอุบลราชธานี ให้ก้าวสู่การเปลี่ยนผ่านทางดิจิทัล (Digital Transformation) อย่างเป็นรูปธรรม เพื่อสร้างความเชื่อมั่นว่าการบริหารจัดการงานวิจัยจะเป็นไปอย่างมีประสิทธิภาพและมีความพร้อมที่จะสนับสนุนการสร้างองค์ความรู้ใหม่ที่ปราศจากความซ้ำซ้อนอย่างเป็นรูปธรรม

กรอบแนวคิดการวิจัย (Conceptual Framework)



ภาพประกอบที่ 1 กรอบแนวคิดการวิจัย

วัตถุประสงค์ของการวิจัย (Objective of the Research)

1. เพื่อพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย โดยใช้เทคนิค งานวิจัย โดยใช้เทคนิค Data Cleansing, Text Embedding, Cosine Similarity
2. เพื่อศึกษาประสิทธิภาพของระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย

เอกสารและงานวิจัยที่เกี่ยวข้อง (Related Literature and Research)

แนวคิดของระบบสารสนเทศ

ระบบสารสนเทศ (Information System) หมายถึง ระบบที่ได้รับการออกแบบเพื่อทำหน้าที่ในการรวบรวม ประมวลผล จัดเก็บ วิเคราะห์และเผยแพร่สารสนเทศ โดยมีวัตถุประสงค์ที่กำหนดไว้เฉพาะเจาะจงในเชิงปฏิบัติการ ประกอบด้วยองค์ประกอบพื้นฐานที่สำคัญ 4 ส่วนหลัก ได้แก่ ปัจจัยนำเข้า (Input) กระบวนการแปรสภาพหรือการประมวลผล (Process) ผลลัพธ์ (Output) และข้อมูลป้อนกลับ (Feedback) โดยผลผลิตที่ได้รับจากการดำเนินงาน คือ สารสนเทศ (Information) ซึ่งนิยามถึงข้อมูลที่ผ่านกระบวนการประมวลผลและการวิเคราะห์อย่างถูกต้อง ครบถ้วนและเหมาะสมสำหรับการนำไปประยุกต์ใช้เพื่อสนับสนุนการบริหารจัดการและการตัดสินใจภายในองค์กร (สุรพงษ์ วิริยะ และคณะ, 2567)

แนวคิดของปัญญาประดิษฐ์

ปัญญาประดิษฐ์ (Artificial Intelligence: AI) เปรียบเสมือนกับอัลกอริทึม (Algorithms) แต่อย่างไร อัลกอริทึม เกิดขึ้นก่อนปัญญาประดิษฐ์และมีการนำไปใช้อย่างแพร่หลาย โดยรากศัพท์ของคำว่า "อัลกอริทึม" นั้น มีที่มาจากชื่อของ Mohammed ibn Musa al-Kharizmi นักคณิตศาสตร์ชาวเปอร์เซียในช่วงคริสต์ศตวรรษที่ 9 และมีความหมายอ้างอิงถึง ชุดคำสั่งเฉพาะเพื่อการแก้ไขปัญหา ส่วนปัญญาประดิษฐ์ หมายถึง การจำลองสติปัญญาที่เป็นคุณลักษณะเฉพาะของมนุษย์โดยระบบคอมพิวเตอร์ เป็นเทคโนโลยีที่เอื้อให้เครื่องจักรสามารถจำลองทักษะของมนุษย์ ปัญญาประดิษฐ์สามารถปฏิบัติงานภายใต้ สภาพแวดล้อมที่มีความซับซ้อนได้ (Sheikh, Prins & Schrijvers, 2023) สอดคล้องกับ Russell and Norvig (2021) ที่กล่าวว่าระบบคอมพิวเตอร์สามารถทำงานเลียนแบบความสามารถทางสติปัญญาของมนุษย์ได้ เช่น การเรียนรู้ การรับรู้ การแก้ไขปัญหาและการตัดสินใจ

งานวิจัยที่เกี่ยวข้อง

สุรพงษ์ วิริยะ และคณะ (2567) พัฒนาระบบสารสนเทศเพื่อสนับสนุนการบริหารจัดการงานวิจัยในรูปแบบเว็บแอปพลิเคชัน โดยประยุกต์ใช้บูตสตรัปเฟรมเวิร์ก (Bootstrap Framework) ร่วมกับภาษาพีเอชพี (PHP) และบริหารจัดการฐานข้อมูลด้วยระบบมายเอสคิวแอล (MySQL) ในส่วนของการประเมินคุณภาพระบบ คณะผู้วิจัยได้ดำเนินการทดสอบฟังก์ชันการทำงานแบบแบ็กบ็อกซ์ (Black-Box Testing) เพื่อตรวจสอบความถูกต้องของการประมวลผลข้อมูลนำเข้าและแสดงผลลัพธ์ในทุกส่วนประสานผู้ใช้ ผลการศึกษาพบว่า ระบบสารสนเทศที่พัฒนาขึ้นสามารถตอบสนองความต้องการของผู้ใช้ได้อย่างมีประสิทธิภาพ สามารถนำไปประยุกต์ใช้เพื่อวางแผนการบริหารจัดการงานวิจัยได้อย่างเป็นรูปธรรม โดยผลการประเมินประสิทธิภาพของระบบโดยรวมอยู่ในระดับมาก

ภวิกา อ้วนละมัย และคณะ (2568) พัฒนาระบบแนะนำหนังสือโดยใช้เทคนิคการกรองตามเนื้อหา (Content-Based Filtering) ร่วมกับการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ในกระบวนการตัดคำ จากนั้นจึงคำนวณค่าความถี่ของคำด้วยวิธี TF-IDF และแปลงข้อมูลข้อความให้อยู่ในรูปเวกเตอร์ตัวเลขด้วยเทคนิค Label Encoding ผลการวิจัยพบว่า ระบบแนะนำหนังสือดังกล่าวมีความถูกต้องในภาพรวมร้อยละ 92.97 เมื่อพิจารณาประสิทธิภาพ

ตามดัชนีวัดความคล้ายคลึง พบว่า วิธี Cosine Similarity ให้ค่าความถูกต้องสูงสุดที่ร้อยละ 92.92 รองลงมาคือวิธี Adjusted Cosine Similarity ที่ร้อยละ 92.50 ผลการวิจัยนี้จึงช่วยยืนยันถึงประสิทธิภาพของการเลือกใช้วิธีวัดความคล้ายคลึงที่เหมาะสมในการพัฒนาระบบแนะนำแบบอิงเนื้อหาได้อย่างมีนัยสำคัญ

รัชพิชญ์ ชำนาญกิจ และฐิติรัตน์ ศิริบรรรัตนกุล (2565) พัฒนาระบบต้นแบบสำหรับการตรวจสอบข่าวปลอมภาษาไทยโดยประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ในกระบวนการสกัดคุณลักษณะของข้อมูล เพื่อนำไปเปรียบเทียบกับฐานข้อมูลข่าวปลอม โดยระบบจะคัดเลือกและแสดงผลพร้อมข้อมูลที่มีความคล้ายคลึงกับข้อความนำเข้า (Input) สูงสุด 3 อันดับแรก จากการทดลองกับชุดข้อมูลทดสอบจำนวนทั้งสิ้น 120 ข่าว (ครอบคลุมทั้งข่าวจริงและข่าวปลอม) พบว่า ระบบต้นแบบสามารถจำแนกข่าวปลอมได้อย่างมีประสิทธิภาพ โดยมีค่าความแม่นยำ (Precision) เท่ากับร้อยละ 91.84 และค่าความระลึก (Recall) เท่ากับร้อยละ 75.00

วรพงศ์ บำรุงศรี (2564) พัฒนาระบบประมวลผลภาษาธรรมชาติเพื่อตรวจจับการกลั่นแกล้งบนสื่อสังคมออนไลน์ในบริบทภาษาไทย โดยประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นกลไกหลักในการคัดเลือกและวิเคราะห์ข้อความจากการโต้ตอบบนแพลตฟอร์มทวิตเตอร์ (Twitter) ในส่วนของการประเมินประสิทธิภาพของแบบจำลอง คณะผู้วิจัยได้ใช้ดัชนีวัดผล ได้แก่ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเอฟเมเชอร์ (F-Measure) ผลการวิจัยพบว่า แบบจำลองที่พัฒนาขึ้นมีประสิทธิภาพและสามารถทำนายผลลัพธ์เพื่อตรวจจับการกลั่นแกล้งได้อย่างถูกต้อง โดยมีค่าความแม่นยำรวมเท่ากับร้อยละ 84.23

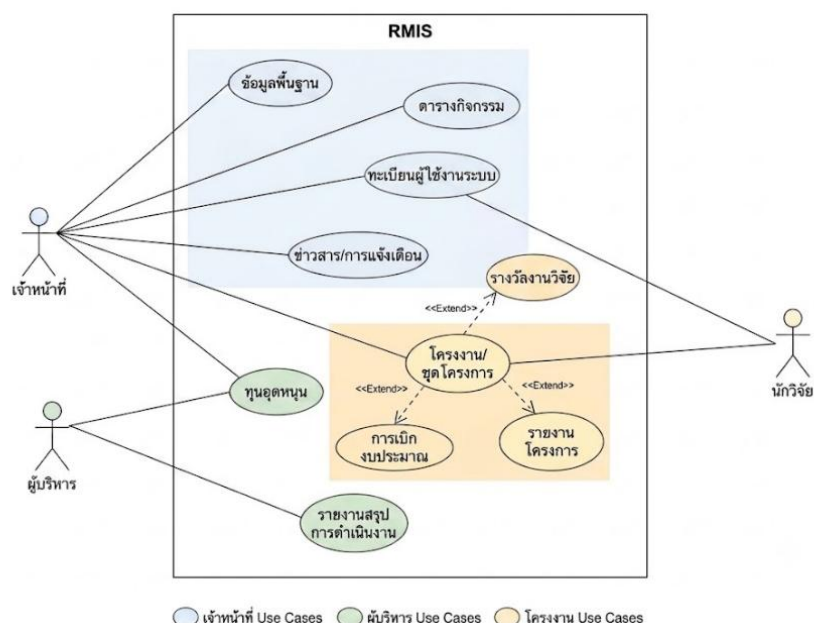
วิธีวิจัย (Research Methodology)

1. ผู้วิจัยศึกษาแนวทางการจัดทำระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัยจากเอกสาร งานวิจัยที่เกี่ยวข้อง รวมถึงการเก็บรวบรวมข้อมูลบริบทการทำงานด้านการบริหารงานวิจัยของสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏอุบลราชธานี

2. การสร้างเครื่องมือวิจัย ผู้วิจัยดำเนินการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย โดยใช้แนวคิด Data Cleansing, Text Embedding, Cosine Similarity เพื่อใช้ในการพัฒนาระบบไม่ให้เกิดการซ้ำซ้อนของหัวข้อวิจัยและช่วยเพิ่มประสิทธิภาพในการบริหารจัดการระบบ ดังนี้

2.1 การวิเคราะห์ความต้องการใช้ระบบ โดยดำเนินการศึกษาและวิเคราะห์ปัญหาจากกระบวนการทำงานในปัจจุบัน (As-is Process) พร้อมทั้งสำรวจความต้องการของผู้ใช้ (User Requirements) เพื่อนำมาประมวลผลและกำหนดฟังก์ชันการทำงานของระบบสารสนเทศให้สอดคล้องกับความต้องการดังกล่าว ตลอดจนกำหนดขอบเขตของระบบ (System Scope) ให้มีความชัดเจน

2.2 การออกแบบระบบ ผู้วิจัยดำเนินการออกแบบเว็บแอปพลิเคชัน ออกแบบฟังก์ชันการใช้งานต่าง ๆ ให้สอดคล้องกับวัตถุประสงค์ของการทำงานและสามารถตอบสนองความต้องการของผู้ใช้ ทั้งนี้ ผู้วิจัยจำลองฐานข้อมูลงานวิจัยในระบบสารสนเทศ จำนวน 200 เรื่องเพื่อเป็นต้นแบบในการวิเคราะห์ข้อมูล รวมทั้งใช้วิธีการวิเคราะห์และออกแบบระบบเชิงวัตถุ (Object-oriented Analyst and Design : OOAD) ซึ่งเป็นแนวคิดของ Gamma et al. (1995) โดยนำแผนภาพต่าง ๆ ที่อยู่ใน Unified Modeling Language (UML) ซึ่งเป็นโมเดลมาตรฐานที่ใช้ในการออกแบบโปรแกรมเชิงวัตถุ มาแสดงให้เห็นถึงองค์ประกอบและการทำงานของระบบก่อนที่จะถูกนำไปพัฒนาใช้งานจริง โดยแผนภาพแบบจำลองของระบบที่นำมาใช้รายละเอียดดังภาพประกอบที่ 2

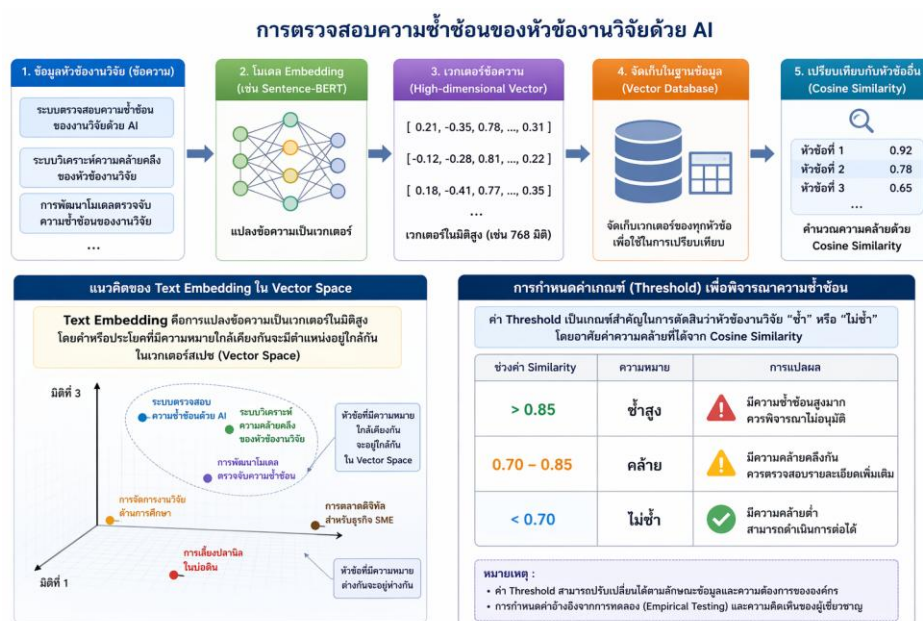


ภาพประกอบที่ 2 Use case diagram ของระบบระบบสารสนเทศเพื่องานวิจัย

2.3 การเตรียมฐานข้อมูลสำหรับใช้ตรวจสอบ ผู้วิจัยทำ Data Cleansing เพื่อเตรียมฐานข้อมูลสำหรับใช้ตรวจสอบความซ้ำซ้อนของหัวข้องานวิจัยด้วย AI เป็นกระบวนการสำคัญที่มุ่งเน้นการปรับปรุงคุณภาพข้อมูลดิบให้มีความถูกต้อง สม่าเสมอและพร้อมสำหรับการประมวลผล โดยเริ่มจากการกำจัดข้อมูลที่ผิดพลาด ไม่สมบูรณ์หรือไม่เกี่ยวข้อง เช่น อักขระพิเศษ ข้อมูลซ้ำหรือรูปแบบข้อความที่ไม่เป็นมาตรฐาน ปรับรูปแบบข้อความ (Normalization) ตัดคำภาษาไทย (Tokenization) ลบคำที่ไม่จำเป็น (Stopwords) และจัดการคำพ้องความหมายเพื่อให้ข้อมูลมีความสอดคล้องกัน กระบวนการเหล่านี้ช่วยลดความคลาดเคลื่อนของข้อมูลและเพิ่มประสิทธิภาพของโมเดล AI เนื่องจากข้อมูลที่สะอาดจะส่งผลให้การวิเคราะห์มีความแม่นยำมากขึ้น (เอนก รุ่งนาไร่, 2566) หลังจากนั้น ข้อมูลจะถูกแปลงให้อยู่ในรูปแบบโครงสร้าง JSON ซึ่งเหมาะสำหรับการส่งผ่าน API และจัดเก็บในฐานข้อมูลของระบบ AI โดยภายใน JSON ประกอบด้วยข้อมูลสำคัญ เช่น หัวข้องานวิจัยเดิม ข้อความที่ผ่านการทำความสะอาด รายการคำ (tokens) และเวกเตอร์ (embedding) ที่ได้จากโมเดล NLP เพื่อนำไปใช้คำนวณความคล้ายคลึงด้วย Cosine Similarity เมื่อมีข้อมูลใหม่เข้าสู่ระบบจะถูกประมวลผลในรูปแบบเดียวกันและนำไปเปรียบเทียบกับข้อมูลในฐานข้อมูล ทำให้สามารถตรวจสอบความซ้ำซ้อนของหัวข้องานวิจัยได้อย่างมีประสิทธิภาพ รองรับ การประมวลผลข้อมูลขนาดใหญ่ได้อย่างเหมาะสม

2.4 การทดสอบใช้เทคนิค Text Embedding คือ การแปลงข้อความเป็นเวกเตอร์ในมิติสูง (High-dimensional vector) คำหรือประโยคที่มีความหมายใกล้เคียงกันจะมีตำแหน่งอยู่ใกล้กันในเวกเตอร์สเปซ (Vector Space) จะถูกนำเข้าสู่มอเดล Embedding เพื่อแปลงเป็นเวกเตอร์ จากนั้นจะถูกจัดเก็บไว้ในฐานข้อมูลเพื่อนำไปใช้เปรียบเทียบกับหัวข้ออื่น ๆ (Mikolov, Chen, Corrado & Dean, 2013) ที่ผู้วิจัยจำลองฐานข้อมูลงานวิจัยในระบบสารสนเทศ จำนวน 200 เรื่อง เพื่อเป็นต้นแบบในการวิเคราะห์ข้อมูลไว้แล้ว ส่วนการกำหนดค่าเกณฑ์ (Threshold) เป็นขั้นตอนสำคัญในการตัดสินใจว่าหัวข้องานวิจัย

“ซ้ำ” หรือ “ไม่ซ้ำ” โดยอาศัยค่าความคล้ายที่ได้จาก Cosine Similarity > 0.85 ซ้ำสูง $0.70 - 0.85$ คล้าย < 0.70 ไม่ซ้ำ (Manning, Raghavan & Schütze, 2009) ดังรายละเอียดภาพที่ 3



ภาพประกอบที่ 3 การตรวจสอบความซ้ำซ้อนของงานวิจัยโดยใช้เทคนิค Text Embedding

2.5 การติดตั้งและการบำรุงรักษาระบบ (System Deployment and Maintenance) ผู้วิจัยดำเนินการติดตั้งเว็บแอปพลิเคชันบนเครื่องแม่ข่าย (Server) เพื่อเปิดให้ผู้ใช้ใช้งานจริง นอกจากนี้ ยังได้กำหนดแนวทางการบำรุงรักษาระบบอย่างต่อเนื่อง โดยหากตรวจพบข้อผิดพลาดหรือปัญหาจากการใช้งานจะดำเนินการตรวจสอบ วิเคราะห์หาสาเหตุและปรับปรุงแก้ไขระบบให้สามารถทำงานได้อย่างมีประสิทธิภาพต่อไป

3. การวิเคราะห์ข้อมูล

3.1 การวิเคราะห์และประมวลผลระบบสารสนเทศ ผู้วิจัยไม่ได้ใช้สถิติเชิงพหุคูณศาสตร์ แต่วิเคราะห์ด้วยข้อมูลทางคอมพิวเตอร์ผ่านกระบวนการคัดกรอง 2 ขั้นตอนหลัก ได้แก่

1) การวิเคราะห์เพื่อทำความสะอาดข้อมูล (Data Cleansing Analysis) เป็นการจัดการข้อมูลข้อเสนอหัวข้องานวิจัยที่เป็นข้อความดิบ (Raw Text) โดยนำเทคนิคเทคนิคทอร์มัลไลเซชัน (Text Normalization) และการตัดคำภาษาไทย (Tokenization) มาใช้วิเคราะห์เพื่อกำจัดข้อมูลรบกวน (Noise) อักขระพิเศษและลบคำที่ไม่จำเป็น (Stopwords) รวมถึงการวิเคราะห์คำพ้องความหมาย (Synonyms) เพื่อแปลงข้อมูลให้อยู่ในโครงสร้างมาตรฐานเดียวกันในรูปแบบ JSON

2) การวิเคราะห์ความคล้ายคลึงเชิงอรรถศาสตร์ (Semantic Similarity Analysis) เป็นการนำข้อความที่ผ่านการทำความสะอาดแล้วเข้าสู่แบบจำลองสำเร็จรูป (Pre-trained Model) เพื่อแปลงค่าเป็นเวกเตอร์ในมิติสูง (High-dimensional Vector) ด้วยเทคนิค Text Embedding แล้วจึงนำไปคำนวณเปรียบเทียบกับฐานข้อมูลเดิมของมหาวิทยาลัยราชภัฏอุบลราชธานีด้วยฟังก์ชัน Cosine Similarity เพื่อจำแนกและระบุระดับความซ้ำซ้อนของหัวข้องานวิจัยตามเกณฑ์การวิเคราะห์ความคล้ายคลึงเชิงอรรถศาสตร์ ดังรายละเอียดตารางที่ 3

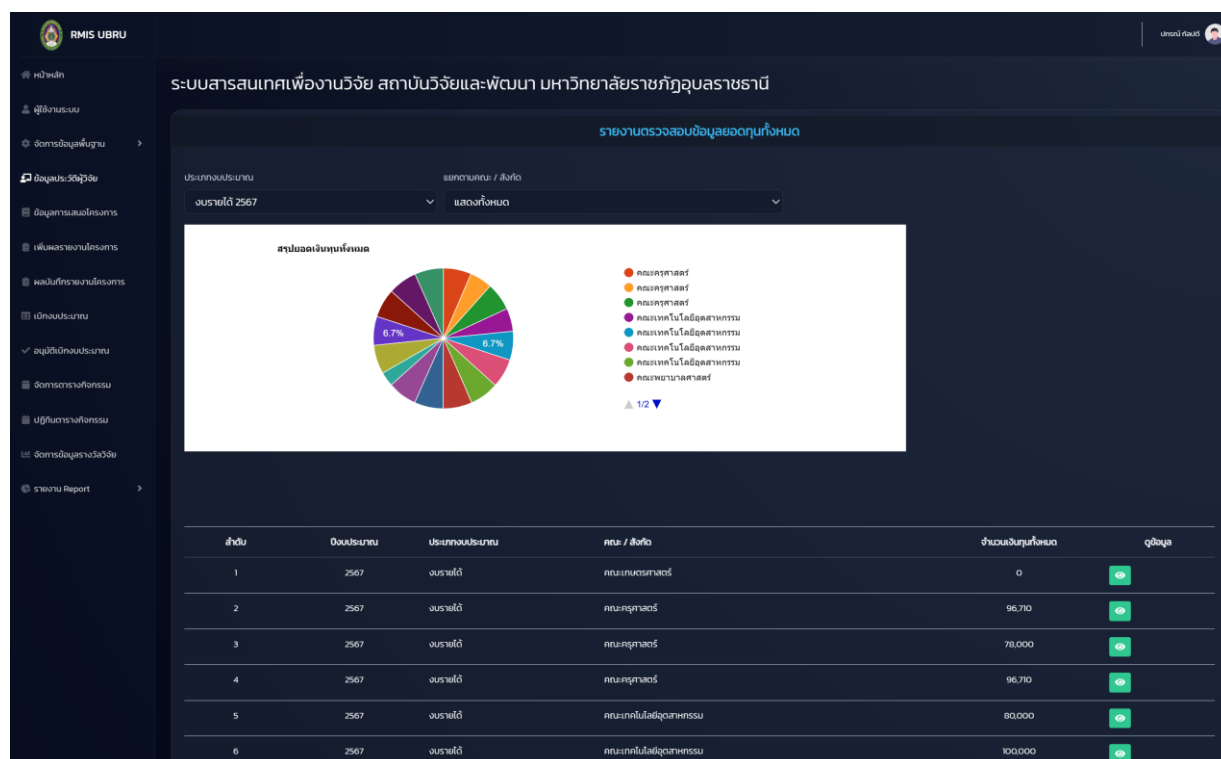
ตารางที่ 3 เกณฑ์การวิเคราะห์ความคล้ายคลึงเชิงอรรถศาสตร์

ค่าความคล้ายคลึง Cosine Similarity	การวิเคราะห์ผล
> 0.85	ความซ้ำซ้อนสูง
0.70 - 0.85	มีความคล้ายคลึง
< 0.70	ไม่มีความซ้ำซ้อน

สรุปผลการวิจัย (Research Conclusions)

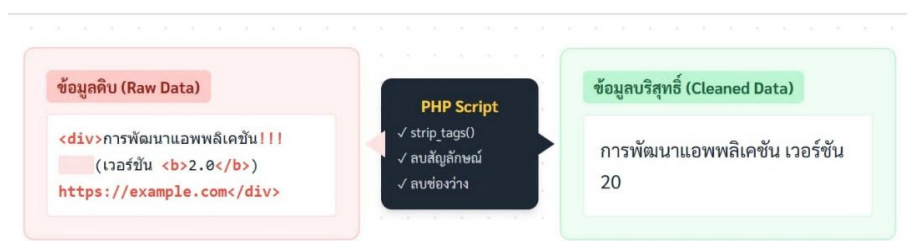
ผลการศึกษาตามวัตถุประสงค์การวิจัยข้อที่ 1 พัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้อ สามารถสรุปผลการวิจัยได้ดังต่อไปนี้

1. ผลการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ผู้วิจัยได้ดำเนินการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้อ โดยกำหนดฟังก์ชันการใช้งาน การจัดการข้อมูลพื้นฐาน ข้อมูลประวัติผู้วิจัย ข้อมูลการเสนอโครงการ เพิ่มผลรายงานโครงการ ผลบันทึกรายงานโครงการ เบิกงบประมาณ อนุมัติเบิกงบประมาณ จัดการตารางกิจกรรม ปฏิทินตารางกิจกรรม จัดการข้อมูลรางวัลวิจัย ส่วนฟังก์ชันรายงาน Report ประกอบด้วย รายงานตรวจสอบข้อมูลยอดทุนวิจัยและรายงานระบบประเมินความซ้ำซ้อนของโครงร่างงานวิจัย ระบบสารสนเทศดังกล่าวได้ถูกนำมาใช้เพื่อสนับสนุนการบริหารจัดการงานวิจัยให้การทำงานมีประสิทธิภาพสอดคล้องกับสถานการณ์ปัจจุบัน พร้อมทั้งนำไปใช้วางแผนด้านการบริหารงานวิจัยของอาจารย์ ทำให้การบริหารงานในองค์กรมีประสิทธิภาพมากขึ้น ดังรายละเอียดภาพที่ 4 ตัวอย่างรายงานที่ได้จากการพัฒนาระบบสารสนเทศ



ภาพประกอบที่ 4 รายงานตรวจสอบข้อมูลยอดเงินทุน

2. ผลการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัยโดยใช้เทคนิค Data Cleansing พบว่ากระบวนการเตรียมและทำความสะอาดข้อมูล (Data Cleansing) ที่ผู้วิจัยพัฒนาขึ้นโดยใช้ภาษา PHP ก่อนส่งข้อมูลเข้าสู่กระบวนการของปัญญาประดิษฐ์สามารถลบข้อมูลขยะหรือสิ่งรบกวน (Noise) เช่น แท็ก HTML (HTML Tags), อักขระพิเศษ, สัญลักษณ์ที่ไม่จำเป็น ตลอดจนการจัดการช่องว่างส่วนเกิน ช่วยให้ชุดข้อมูลงานวิจัยมีความบริสุทธิ์มากขึ้น ทำให้การประมวลผลมีความแม่นยำในการวิเคราะห์ความซ้ำซ้อนมากขึ้น ดังนั้น จึงสามารถลดข้อผิดพลาดจากการประเมินของ AI และลดขนาดของข้อมูล (Token) ที่ส่งไปประมวลผล ทำให้ระบบสามารถทำงานและตอบสนองต่อผู้ใช้งานได้อย่างรวดเร็วมากยิ่งขึ้น ภาพจำลองการประมวลผลด้วย Data Cleansing ดังรายละเอียดภาพที่ 5



ภาพประกอบที่ 5 จำลองผลการประมวลผลด้วย Data Cleansing

3. ผลการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย โดยนำเทคนิค Text Embedding มาประยุกต์ใช้ผ่านการทำงานของโมเดลภาษาปัญญาประดิษฐ์ พบว่าเทคนิคนี้ช่วยให้ระบบสามารถแปลงข้อมูลตัวอักษรที่อยู่ในรายละเอียดของงานวิจัย เช่น ชื่อเรื่องและคีย์เวิร์ด บทคัดย่อ ขอบเขตการศึกษา ของโครงร่างงานวิจัยที่เสนอใหม่ให้กลายเป็นตัวแทนข้อมูลเชิงเวกเตอร์ (Vector Representations) ได้อย่างมีประสิทธิภาพ ผลลัพธ์ที่ได้คือ ระบบไม่ได้เปรียบเทียบข้อมูลแบบจับคู่คำต่อคำ (Keyword Matching) แบบระบบเดิม แต่สามารถเข้าใจบริบท (Context) และความหมายเชิงลึก (Semantic Meaning) ของข้อความได้ เช่น ระบบสามารถรับรู้ได้ว่าคำว่า AI และ "ปัญญาประดิษฐ์" มีความหมายเดียวกัน ทำให้การตรวจจับความซ้ำซ้อนของการสลับคำ การใช้คำพ้องความหมายหรือการเขียนเรียบเรียงใหม่มีความแม่นยำสูงขึ้น

4. ผลการพัฒนาระบบสารสนเทศสำหรับตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย โดยใช้เทคนิค Text Embedding, Cosine Similarity จากผลการพัฒนาระบบสารสนเทศ พบว่าการนำหลักการทางคณิตศาสตร์อย่าง Cosine Similarity มาใช้ในการวัดระยะห่างระหว่างเวกเตอร์ของหัวข้องานวิจัยที่เสนอใหม่กับฐานข้อมูลงานวิจัยเดิม ทำให้ระบบสามารถแปลงผลการวิเคราะห์เชิงความหมายออกมาเป็น "ค่าร้อยละ (Percentage)" ที่จับต้องได้ ผลการทดสอบระบบแสดงให้เห็นว่าการคำนวณด้วยวิธีนี้สามารถจัดอันดับ (Ranking) งานวิจัยที่มีความคล้ายคลึงกันมากที่สุดได้อย่างถูกต้องตามลำดับความสำคัญ ช่วยให้ผู้ใช้งานมีเกณฑ์ตัวเลขที่ชัดเจนประกอบการตัดสินใจพิจารณาอนุมัติหัวข้อและสามารถตั้งเกณฑ์มาตรฐานของระบบได้ เช่น หากค่าความคล้ายคลึงเกินร้อยละ 80 ระบบจะแจ้งเตือนความเสี่ยงระดับวิกฤต ตัวอย่างรายงานผลการตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ดังรายละเอียดภาพที่ 6



ไม่เข้าซ้อน

เข้าซ้อนต่ำ

เข้าซ้อนสูงมาก

ภาพประกอบที่ 6 ตัวอย่างรายงานผลการตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย

การศึกษาตามวัตถุประสงค์ของการวิจัยข้อที่ 2 เพื่อศึกษาประสิทธิภาพของระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย จากการทดสอบการใช้งานระบบสารสนเทศโดยนางงานวิจัยจำนวน 10 เรื่อง มาวิเคราะห์ความซ้อนตามเกณฑ์การวิเคราะห์ความคล้ายคลึงเชิงอรรถศาสตร์จากตารางที่ 3 เพื่อประเมินประสิทธิภาพของระบบสารสนเทศและแสดงผลสถิติพื้นฐานด้วยคำร้อยละ ผลการทดสอบพบว่าระบบสารสนเทศที่พัฒนาขึ้นสามารถตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัยได้ทั้งหมด 10 เรื่อง แสดงว่าสามารถตอบสนองต่อวัตถุประสงค์หลักในการตรวจสอบความซ้ำซ้อนของหัวข้องานวิจัยได้อย่างมีประสิทธิภาพ ทั้งในด้านความถูกต้องของการประมวลผล (Accuracy) ความรวดเร็วในการสืบค้นข้อมูลและความง่ายในการใช้งานที่มีความเป็นมิตรกับผู้ใช้ ด้านฟังก์ชันการทำงานระบบสามารถประยุกต์ใช้เทคนิค Data Cleansing และ Text Embedding ได้ ส่งผลให้การเปรียบเทียบความคล้ายคลึงของหัวข้อวิจัย (Cosine Similarity) มีความน่าเชื่อถือและลดความผิดพลาดที่เกิดจากการจัดการข้อมูลดิบได้

จากการทดสอบพบว่าระบบสารสนเทศมีประสิทธิภาพและมีความแม่นยำอย่างยิ่งในการคัดกรอง เปรียบเทียบและวิเคราะห์ความซ้ำซ้อนของงานวิจัย นอกจากนี้ ระบบสารสนเทศสามารถวิเคราะห์เชิงลึกแยกแยะส่วนที่เหมือนและส่วนที่แตกต่างได้อย่างเป็นรูปธรรม โดยตรวจสอบความซ้ำซ้อนของชื่องานวิจัย คีย์เวิร์ด บทคัดย่อ ขอบเขตการศึกษา แสดงผลความซ้ำซ้อนเป็นคำร้อยละของการวิเคราะห์ผลระดับความเสี่ยง รวมทั้งสามารถวิเคราะห์เชิงลึกเพื่อเปรียบเทียบและวิเคราะห์ความซ้ำซ้อนประเด็นที่งานวิจัยใหม่มีความซ้ำซ้อนแตกต่างจากงานวิจัยฉบับเดิมได้อย่างเป็นรูปธรรม ดังนั้น แสดงให้เห็นว่าระบบสารสนเทศนี้มีประสิทธิภาพ เหมาะสมที่จะนำไปประยุกต์ใช้ในการบริหารจัดการงานวิจัยในหน่วยงานระดับอุดมศึกษา เพื่อช่วยลดความซ้ำซ้อนของหัวข้อวิจัยและยกระดับคุณภาพของงานวิจัยในองค์กรได้อย่างเป็นรูปธรรม แสดงดังตารางที่ 4

ตารางที่ 4 ผลการทดสอบประสิทธิภาพของระบบสารสนเทศ

อันดับ	รหัส	ชื่องานวิจัย	ความเข้าซ้อน (ร้อยละ)				การวิเคราะห์ผล	ระดับความเสี่ยง	การวิเคราะห์เชิงลึก
			ชื่อเรื่องและ คีย์เวิร์ด	บทคัดย่อ	ขอบเขต การศึกษา	ความเข้าซ้อน ในภาพรวม			
1	DB-042	ฤทธิ์ต้านเชื้อราก่อโรคราก เน่าโคนเน่าของสารสกัด สมุนไพรไทย	90.00	85.00	60.00	82.00	ความเข้าซ้อนสูง	เข้าซ้อนสูง มาก	ประเด็นเข้าซ้อน: ใช้วิธีการสกัดด้วยเอทานอลและทดสอบกับเชื้อรา Fusarium spp. เหมือนกันทุกประการ และใช้สมุนไพรหลักคือ "สะเดา" ตรงกัน จุดที่แตกต่าง: งานวิจัยเก่าไม่ได้ใช้ "บอระเพ็ด" ร่วมด้วย และทำในพืชตระกูลส้ม ไม่ใช่ทุเรียน
2	DB-118	การใช้สารสกัดจาก บอระเพ็ดเพื่อยับยั้งเชื้อรา ก่อโรคในพืชเศรษฐกิจ	80.00	75.00	40.00	75.00	มีความคล้ายคลึง	เข้าซ้อนสูง	ประเด็นคล้ายคลึง: ใช้สารสกัด "บอระเพ็ด" เป็นตัวแปรต้น และมุ่งเป้าไปที่การ กำจัด "เชื้อรา" เหมือนกัน จุดที่แตกต่าง: งานใหม่มีความแปลกใหม่ที่นำ "สะเดา" มาผสมรวมด้วย และ เจาะจงที่ทุเรียนอินทรีย์โดยเฉพาะ
3	B-087	ประสิทธิภาพของสารสกัด ชีวภาพกำจัดเชื้อราใน แปลงทุเรียนอินทรีย์จังหวัด จันทบุรี	50.00	60.00	95.00	68.00	มีความคล้ายคลึง	เข้าซ้อน บางส่วน	ประเด็นเข้าซ้อน: ขอบเขตงานวิจัย (แปลงทุเรียนอินทรีย์ จังหวัดจันทบุรี) เข้าซ้อนเกือบ 100% จุดที่แตกต่าง: งานวิจัยเก่าใช้สารสกัดจากว่านหางจระเข้ ไม่ใช่สะเดาและ บอระเพ็ด ถือเป็นองค์ความรู้ใหม่ในพื้นที่เดิม
4	DB-015	การใช้ชีวภัณฑ์จากสารสกัด บอระเพ็ดเพื่อควบคุม ศัตรูพืชในสวนผลไม้	60.00	50.00	40.00	55.00	มีความคล้ายคลึง	เข้าซ้อน บางส่วน	มีความคล้ายคลึงในส่วนของการใช้สารสกัด "บอระเพ็ด" ในบริบทของสวนผลไม้ แต่เป้าหมายงานวิจัยต่างกันอย่างสิ้นเชิง โดยงานเดิมเน้นควบคุมแมลงศัตรูพืช แต่งานใหม่นำยับยั้งเชื้อรา
5	DB-143	ผลของสารสกัดสะเดาต่อ การเปลี่ยนแปลงทาง ชีวเคมีของเชื้อราสาเหตุ โรคพืช	50.00	60.00	10.00	42.00	มีความคล้ายคลึง	เข้าซ้อนต่ำ	เนื้อหามีความเกี่ยวข้องกันในเชิงตัวแปร (สะเดาและเชื้อรา) แต่งานเดิมเป็นการ ทดลองระดับห้องปฏิบัติการ (In-vitro) เพื่อดูค่าทางชีวเคมี ไม่ได้นำไป ประยุกต์ใช้ในแปลงเกษตรจริง

ตารางที่ 4 ผลการทดสอบประสิทธิภาพของระบบสารสนเทศ (ต่อ)

อันดับ	รหัส	ชื่องานวิจัย	ความเข้าช้ (ร้อยละ)				การวิเคราะห์ผล	ระดับความเสี่ยง	การวิเคราะห์เชิงลึก
			ชื่อเรื่องและ คีย์เวิร์ด	บทคัดย่อ	ขอบเขต การศึกษา	ความเข้าช้ ในภาพรวม			
6	DB-092	การจัดการโรครากเน่าในพืชตระกูลส้มด้วยวิถีเกษตรอินทรีย์	40.00	30.00	60.00	35.00	มีความคล้ายคลึง	เข้าช้ต่ำ	บริบทภาพรวมตรงกันคือ "การแก้ปัญหาโรครากเน่าในเกษตรอินทรีย์" แต่วิธีดำเนินการต่างกัน งานเดิมเน้นใช้วิธีการปรับปรุงดินและการตัดแต่งกิ่ง แทนการใช้สารสกัดสมุนไพร
7	DB-008	ประสิทธิภาพของสารสกัดสะเดาต่อการเจริญเติบโตของพืชผักสวนครัว	40.00	20.00	20.00	28.00	มีความคล้ายคลึง	เข้าช้ต่ำ	มีเพียงตัวแปรต้นคือ "สารสกัดสะเดา" ที่ตรงกัน แต่วัตถุประสงค์เพื่อเป็นฮอร์โมนเร่งการเจริญเติบโตและใช้กับกลุ่มเป้าหมายคือพืชผักสวนครัว ซึ่งไม่เกี่ยวข้องกับขอบเขตงานใหม่
8	DB-176	การศึกษาเปรียบเทียบตัวทำละลายในการสกัดสารสำคัญจากบอระเพ็ด	30.00	30.00	5.00	22.00	ไม่มีความเข้าช้	ปลอดภัย (ไม่เข้าช้)	เป็นงานวิจัยสายวิทยาศาสตร์เคมีวิเคราะห์(หาค่า Extraction Yield) ไม่มีบริบทของการนำไปใช้ทางการเกษตรหรือการกำจัดเชื้อรา
9	DB-033	ระบาดวิทยาของเชื้อรา Fusarium spp. ในดินแปลงเกษตรภาคตะวันออก	20.00	20.00	40.00	15.00	ไม่มีความเข้าช้	ปลอดภัย (ไม่เข้าช้)	ศึกษาพฤติกรรมและการกระจายตัวของเชื้อรา Fusarium ในภาคตะวันออก แต่ไม่มีเรื่องการทดลองใช้สารสกัดสมุนไพรเพื่อควบคุมเชื้อรา
10	DB-199	การประเมินคุณภาพดินและธาตุอาหารในแปลงปลูกทุเรียนอินทรีย์	10.00	10.00	50.00	10.00	ไม่มีความเข้าช้	ปลอดภัย (ไม่เข้าช้)	มีเพียงคำว่า "แปลงปลูกทุเรียนอินทรีย์" เท่านั้นที่ตรงกัน แต่วิธีดำเนินการวิจัยและการวัดผล (วัดแร่ธาตุในดิน) แตกต่างกันโดยสิ้นเชิง

อภิปรายผล (Discussion)

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาระบบสารสนเทศและศึกษาประสิทธิภาพของระบบสารสนเทศเพื่อตรวจสอบความซ้ำซ้อนของการเสนอหัวข้องานวิจัย ผลการวิจัยสามารถอภิปรายได้ดังนี้

1. ประสิทธิภาพของการพัฒนาระบบสารสนเทศและการคัดกรองข้อมูล ผลการทดสอบประสิทธิภาพของระบบสารสนเทศ พบว่าระบบที่พัฒนาขึ้นสามารถจำแนกความซ้ำซ้อนของหัวข้องานวิจัยได้อย่างมีประสิทธิภาพ โดยในภาพรวมระบบสามารถแยกแยะความแตกต่างระหว่าง "หัวข้องานวิจัยที่ซ้ำซ้อน" กับ "หัวข้อที่มีเพียงความคล้ายคลึงของคำสำคัญแต่ต่างวัตถุประสงค์" ได้อย่างชัดเจน เช่น รหัสงานวิจัย DB-042 กรณีของงานวิจัยที่เกี่ยวข้องกับเชื้อรา *Fusarium* spp. ระบบสามารถระบุได้ว่างานวิจัยที่มีเป้าหมายเป็นพืชเศรษฐกิจชนิดต่างกันหรือใช้วิธีการทดลองที่แตกต่างกัน เช่น ระดับห้องปฏิบัติการเทียบกับการนำไปใช้จริง เป็นงานวิจัยที่ไม่ซ้ำซ้อนกัน สอดคล้องกับแนวคิดของ เอนก รุ่งนาไร่ (2566) ที่ระบุว่ากระบวนการ Data Cleansing ที่ช่วยลด Noise และความคลาดเคลื่อนของข้อมูล ส่งผลให้การประมวลผลของโมเดล AI มีความแม่นยำสูงขึ้น

2. ความแม่นยำในการวิเคราะห์ความหมายเชิงลึก (Semantic Similarity) จากการทดสอบพบว่าการใช้เทคนิค Text Embedding ร่วมกับ Cosine Similarity ช่วยให้ระบบสามารถเข้าใจความหมายเชิงลึกมากกว่าการใช้ Keyword Matching แบบเดิม ซึ่งเป็นจุดเปลี่ยนสำคัญในการตรวจจัดการ "ตัดแปลงข้อความ" หรือ "ใช้คำพ้องความหมาย" (Synonyms) ของนักวิจัย ผลการวิเคราะห์ในตารางที่ 4 ยืนยันว่าระบบสามารถจัดอันดับ (Ranking) ความคล้ายคลึงได้อย่างถูกต้องและสอดคล้องกับงานวิจัยของ ภวิกา อ้วนละมัย และคณะ (2568) ที่พบว่าวิธีการวัดความใกล้เคียงด้วย Cosine Similarity ให้ค่าความถูกต้องในการแนะนำข้อมูลสูงถึงร้อยละ 92.92 ซึ่งเป็นเทคนิคที่เหมาะสมและมีประสิทธิภาพสูงในการพัฒนาระบบที่ต้องวิเคราะห์ความสัมพันธ์เชิงเนื้อหา

ผลการศึกษานี้สอดคล้องกับงานวิจัยของ สุรพงษ์ วิริยะ และคณะ (2567) ที่พัฒนาเว็บแอปพลิเคชันเพื่อสนับสนุนการบริหารจัดการงานวิจัยและพบว่าระบบสามารถตอบสนองความต้องการของผู้ใช้ได้เป็นอย่างดีและอยู่ในระดับมาก นอกจากนี้ยังสอดคล้องกับงานวิจัยของ ธัชพิชญ์ ชำนาญกิจ และฐิติรัตน์ ศิริบรรรัตนกุล (2565) รวมถึง วรพงศ์ บำรุงศรี (2564) ที่ชี้ให้เห็นว่าการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (NLP) ในการสกัดคุณลักษณะข้อมูล สามารถจำแนกหรือระบุข้อความที่ผิดปกติหรือซ้ำซ้อนได้อย่างแม่นยำ (Precision > 80%) ซึ่งการศึกษานี้ได้ต่อยอดแนวคิดดังกล่าวโดยการนำมาประยุกต์ใช้กับฐานข้อมูลหัวข้องานวิจัยในระดับอุดมศึกษา เพื่อแก้ปัญหาความซ้ำซ้อนเชิงจริยธรรมตามที่กรธ อยู่สุข (2557) ได้นำเสนอไว้ การบูรณาการระบบสารสนเทศเข้าสู่ระบบดิจิทัลอย่างเต็มรูปแบบนี้ จึงเป็นกลไกสำคัญในการยกระดับมาตรฐานการบริหารงานวิจัยของสถาบันอุดมศึกษาให้มีความแม่นยำ ลดภาระงานของเจ้าหน้าที่ (Manual Review) และส่งเสริมความโปร่งใสในการจัดสรรงบประมาณได้อย่างเป็นรูปธรรม

ข้อเสนอแนะ (Suggestion)

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การใช้งานเป็นเครื่องมือคัดกรองเบื้องต้น (Screening Tool) สถาบันวิจัยและพัฒนาควรนำระบบสารสนเทศนี้ไปบูรณาการเป็นขั้นตอนแรกในการรับข้อเสนอโครงการวิจัย เพื่อลดภาระของเจ้าหน้าที่และกรรมการพิจารณาในการตรวจสอบด้วยบุคคล (Manual Review)

1.2 การปรับปรุงฐานข้อมูลอย่างต่อเนื่อง เพื่อให้การทำงานของโมเดล AI (Text Embedding) มีความแม่นยำ ผู้ดูแลระบบควรมีการนำเข้าข้อมูลหัวข้องานวิจัยใหม่ ๆ เข้าสู่ฐานข้อมูล (Vector Database) อย่างสม่ำเสมอ

1.3 การจัดอบรมเชิงปฏิบัติการ ควรมีการจัดอบรมการใช้งานเว็บแอปพลิเคชันให้กับกลุ่มนักวิจัยและบุคลากรที่เกี่ยวข้องเพื่อให้เข้าใจเกณฑ์การคำนวณค่าความคล้ายคลึง (Cosine Similarity) และสามารถแปลผลความซ้ำซ้อนได้อย่างถูกต้อง

2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

2.1 การวิเคราะห์ความคล้ายคลึงข้ามภาษา (Cross-lingual Similarity) ควรมีการพัฒนาหรือปรับแต่งโมเดล NLP ให้สามารถตรวจจับความซ้ำซ้อนของการเสนอหัวข้อที่เป็นภาษาต่างประเทศเทียบกับภาษาไทยได้ เพื่อป้องกันการแปลงงานวิจัยเดิมมานำเสนอใหม่

2.3 การใช้โมเดลภาษาขนาดใหญ่ (Large Language Models: LLMs) ควรศึกษาแนวทางการนำ LLMs ที่มีความซับซ้อนมาช่วยในการประเมิน "บริบทของปัญหา" หรือ "ระเบียบวิธีวิจัย" เพื่อให้ระบบสามารถวิเคราะห์ความซ้ำซ้อนในเชิงแนวคิดได้ลึกซึ้งยิ่งขึ้น

เอกสารอ้างอิง (References)

- กมลดา เรืองอร่าม. (2566). การพัฒนาและปรับปรุงระบบสารสนเทศเพื่อการจัดการงานวิจัย มหาวิทยาลัยราชภัฏเพชรบุรี. *วารสารวิชาการมหาวิทยาลัยราชภัฏเพชรบุรี*, 13(3), 12-18.
- กรธัช อยู่สุข. (2557). การวิเคราะห์การเผยแพร่ผลงานวิชาการและงานวิจัยซ้ำซ้อนในประเทศไทย (Analysis of Duplicate Publication in Thailand). สืบค้นเมื่อวันที่ 1 เมษายน 2569, จาก <https://www.gotoknow.org/posts/579456>
- บุญชม ศรีสะอาด. (2560). *การวิจัยเบื้องต้น* (พิมพ์ครั้งที่ 10 ฉบับปรับปรุงใหม่). กรุงเทพฯ: สุวีริยาสาส์น.
- ภวิกา อ้วนละมัย, ธรรมบุญ ปัญญาทิพย์, ปนัดดา โพธินาม, อัจฉรา สมังเกษตร, ณรงค์ฤทธิ์ มะสุใส, ทรงกรด พิมพ์ศาล, และ ไพฑูรย์ ทัพย์สันเทียะ. (2568). การพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา. *วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์*, 3(3), 36-53.
- พิรพัฒน์ หมั่นจิตร และฐิตาภรณ์ สีนจรรย์ศักดิ์. (2568). อิทธิพลของคุณภาพระบบสารสนเทศทางการบัญชี คุณภาพงานบริการบัญชีอิเล็กทรอนิกส์ และคุณภาพข้อมูลทางการบัญชีที่ส่งผลต่อประสิทธิภาพการปฏิบัติงานด้านบัญชีและความเจริญเติบโตของธุรกิจขนาดกลางและขนาดย่อมในประเทศไทย. *วารสารการวิจัยการบริหารการพัฒนา*, 15(2), 684-698.
- เมธิกา พ่วงแสง และวิสุตา วรรณห้วย. (2562). การพัฒนาระบบสารสนเทศสำหรับการจัดการข้อมูลงานวิจัยในยุคดิจิทัล มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. *วารสารเทคโนโลยีสารนิเทศ มทร.พระนคร*, 4(1), 8-17.
- วรพงศ์ บำรุงศรี. (2564). *การพัฒนาระบบประมวลผลภาษาธรรมชาติ เพื่อป้องกันการกลั่นแกล้งบนโซเชียลมีเดียในภาษาไทย* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). จุฬาลงกรณ์มหาวิทยาลัย.
- สุรพงษ์ วิริยะ, อุทัยวรรณ แก้วตะคุ, Nguyen, H. A., และกิตติพิเชษฐ์ ฐูปุข. (2567). การพัฒนาระบบสารสนเทศเพื่อสนับสนุนการบริหารจัดการงานวิจัย มหาวิทยาลัยเจ้าพระยา. *วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ*, 10(1), 22-35.
- เอนก รุ่งนาไร่. (2566). *การใช้เทคนิค Data Cleansing เพื่อปรับปรุงคุณภาพข้อมูล*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). มหาวิทยาลัยศิลปากร.

- เดชชนะ ศรีรัตนลัม และธิดารัตน์ สุขประภาภรณ์. (2568). การพัฒนาระบบสารสนเทศเพื่อการบริหารจัดการงานวิจัยของมหาวิทยาลัยราชภัฏเชียงราย. *วารสารการวิจัยภาษาละลองคำ มหาวิทยาลัยราชภัฏเชียงราย*, 19(1), 58-77.
- รัชพิชญ์ ชำนาญกิจ และฐิติรัตน์ ศิริบรรณกุล. (2565). การตรวจสอบข่าวปลอมภาษาไทยด้วยเทคนิคการประมวลผลภาษาธรรมชาติ. *วารสารวิจัยและพัฒนา มจร.*, 45(2), 275-287.
- โอบาส เอี่ยมสิริวงศ์. (2560). *ระบบสารสนเทศเพื่อการจัดการ*. กรุงเทพฯ: ซีเอ็ดดูเคชั่น.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. New York: Addison-Wesley.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. Retrieved from <https://arxiv.org/pdf/1301.3781>
- Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial Intelligence: Definition and Background. In *Mission AI: Research for Policy*. Cham: Springer.