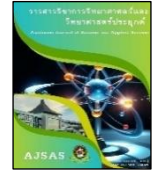




Faculty of Science and Technology, Uttaradit Rajabhat University

วารสารวิชาการวิทยาศาสตร์และเทคโนโลยีศาสตร์ประยุกต์

Journal homepage: <https://ph03.tci-thaijo.org/index.php/ajsas>

ระบบสำหรับการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยาย

Novel Similarity Analysis and Comparison System

พัชรพล กุศลทรัพย์¹, ภาวัต นาคเจือทอง¹, สุพารณ์ ชิมเจริญ¹Patcharaphon Koosomsarp¹, Pawat Narkjuathong¹, Supaporn Simcharoen¹¹ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปทุมธานี¹Department of Information Technology, Faculty of Industrial Technology and Management,

King Mongkut's University of Technology North Bangkok, Prachinburi Campus.

ข้อมูลบทความ

ประวัติบทความ

รับเมื่อ: 2 มิถุนายน 2569

แก้ไขเมื่อ: 20 สิงหาคม 2569

ตอบรับเมื่อ: 21 สิงหาคม 2569

เผยแพร่ออนไลน์:

5 กันยายน 2569

คำสำคัญ

การวิเคราะห์ความคล้ายคลึงข้อความ

เทคนิคการวิเคราะห์ความคล้ายคลึง

การแสดงผลเชิงกราฟ

*ผู้ประพันธ์บรรณกิจ

อีเมล:

[patcharaphon.koosomsarp](mailto:patcharaphon.koosomsarp@gmail.com)[@gmail.com](mailto:patcharaphon.koosomsarp@gmail.com)

(พัชรพล กุศลทรัพย์)

บทคัดย่อ

งานวิจัยนี้นำเสนอการประยุกต์ใช้เทคนิคการวิเคราะห์และเปรียบเทียบความคล้ายคลึงกัน (Similarity Analysis) ของนิยายทั้งที่เขียนด้วยภาษาไทยและภาษาอังกฤษ ระบบสามารถรองรับการนำเข้านิยายแบบไฟล์เดียวหรือหลายไฟล์พร้อมกันได้ในรูปแบบ TXT, DOCX และ PDF รวมถึงสามารถคัดลอกและวางข้อความได้โดยตรง โดยการทำงานของระบบ เริ่มจากการเตรียมและประมวลผลข้อมูลของนิยาย ทั้งการทำความสะอาดข้อความ การจัดรูปแบบ และการแบ่งหน่วยข้อความ เพื่อใช้ในการวิเคราะห์ความเหมือน ระบบสามารถวิเคราะห์ความคล้ายคลึงได้ตั้งแต่ระดับคำ ประโยค ย่อหน้า บท จนถึงทั้งเรื่องของนิยาย ผลลัพธ์จากการวิเคราะห์ที่ได้จะแสดงในรูปแบบคะแนนความเหมือน (Similarity Score) รวมทั้งแสดงในรูปแบบแผนภาพ เช่น Heatmap, Network Graph และ Timeline จากการประเมินประสิทธิภาพของระบบด้วยชุดข้อมูลทดสอบนิยายภาษาอังกฤษจำนวน 50 เรื่อง (150 ตอน) ครอบคลุม 5 หมวดหมู่ พบว่าที่ค่าเกณฑ์ความคล้ายคลึง (Threshold) 0.60 ระบบให้ค่าความแม่นยำ (Precision) ร้อยละ 84.8 ค่าความครอบคลุม (Recall) ร้อยละ 82.0 และค่า F1-Score ร้อยละ 83.4 ซึ่งแสดงให้เห็นว่าระบบมีประสิทธิภาพที่ดีในการจำแนกและเปรียบเทียบความคล้ายคลึงของเนื้อหา

ARTICLE INFO

Article History

Received: 2 June 2026

Revised: 20 August 2026

Accepted: 21 August 2026

Available online:

5 September 2026

ABSTRACT

This research presents the application of Similarity Analysis techniques for analyzing and comparing novels in both Thai and English. The system supports importing novels as single or multiple files in TXT, DOCX, and PDF formats, as well as direct copy-and-paste text input. The system's workflow begins with preprocessing and preparing the novel data, including text cleaning, formatting, and segmentation

Keywords:

Text Similarity Analysis
Similarity Analysis
Technique
Graph Visualization

Corresponding author*Email address:**

patcharaphon.koosomsarp@gmail.com
(Patcharaphon
Koosomsarp)

into analysis units. The system can perform similarity analysis at multiple levels, ranging from words to sentences, paragraphs, chapters, and entire novels. The results are presented in the form of similarity scores, and various visualizations, including heatmaps, network graphs, and timelines. Performance evaluation conducted on an English novel dataset comprising 50 titles (150 chapters) across 5 genres demonstrated that at an optimal similarity threshold of 0.60, the system achieved a precision of 84.8%, a recall of 82.0%, and an F1-score of 83.4%, indicating high effectiveness and reliability in text similarity detection and classification.

1. บทนำ

การนำเนื้อหาของนิยายมาวิเคราะห์เปรียบเทียบความคล้ายคลึงระหว่างเนื้อหาถือเป็นหนึ่งในงานที่ต้องมีความละเอียดและใช้เวลามาก เนื่องจากนักวิจัย นักเขียน และนักวิจารณ์วรรณกรรมจำเป็นต้องตรวจสอบและเปรียบเทียบเนื้อหาที่หลากหลายและมีจำนวนมาก เพื่อหาแนวคิด ลักษณะของเนื้อเรื่อง หรือความคล้ายคลึงระหว่างนิยายในแต่ละเรื่อง กระบวนการตรวจสอบดังกล่าวไม่เพียงแต่ใช้เวลานาน แต่ยังเสี่ยงที่จะมีข้อผิดพลาดเกิดขึ้นได้เช่นเดียวกัน ในกระบวนการนี้ การนำเทคโนโลยีเข้ามาช่วยโดยเฉพาะ NLP หรือ เทคโนโลยีการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ซึ่งจะช่วยวิเคราะห์ข้อความและข้อมูลที่อยู่ในรูปแบบภาษามนุษย์ได้ [1]

NLP เป็นหนึ่งในเทคโนโลยีที่มีประสิทธิภาพ ที่ช่วยให้ระบบคอมพิวเตอร์สามารถวิเคราะห์ข้อมูลจากภาษามนุษย์ได้ ช่วยให้ความซับซ้อนในการเปรียบเทียบเนื้อหาลดลง รวมทั้งช่วยให้การประมวลผลแม่นยำขึ้น ตัวอย่างการประยุกต์ใช้งาน NLP ที่พบได้ทั่วไป ได้แก่ การแยกคำ (Tokenization) การระบุเอนทิตีที่สำคัญในข้อความ (Named Entity Recognition: NER) และการวิเคราะห์ความคล้ายคลึงของข้อความ (Cosine Similarity) เพื่อเปรียบเทียบเนื้อหาหรือนิยาย [2]

ในมุมมองของทฤษฎีวรรณกรรมและวรรณกรรมวิจารณ์ การวิเคราะห์องค์ประกอบของนวนิยาย เช่น ทิมหลัก (Theme) โครงเรื่อง (Plot) และฉาก (Setting) มักถูกสะท้อนผ่านกลุ่มคำศัพท์เฉพาะ (Lexical Field) ที่นักเขียนเลือกใช้ ดังนั้น การนำเทคนิคการวิเคราะห์ความคล้ายคลึงเชิงหัวข้อ (Topic-level Similarity) และความถี่ของคำศัพท์มาประยุกต์ใช้ จึงมีความสอดคล้องกับแนวคิดเรื่องสัมพันธบท (Intertextuality) [2] ซึ่งอธิบายถึงความเชื่อมโยงระหว่างตัวบทวรรณกรรม การที่ระบบสามารถตรวจจับกลุ่มคำศัพท์ที่ซ้ำร่วมกันได้ จึงเป็นเครื่องมือเชิงประจักษ์ที่สะท้อนให้เห็นความคล้ายคลึงของแนวทางการเขียน (Genre) หรือสไตล์ของเรื่องได้อย่างสมเหตุสมผล

แอปพลิเคชันสำหรับใช้ในวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยายที่พัฒนาขึ้น มีความสำคัญสำหรับช่วยในกระบวนการวิจัยและการวิจารณ์วรรณกรรมมีประสิทธิภาพเพิ่มมากขึ้น ข้อมูลสำคัญต่าง ๆ จากนิยาย เช่น ตัวละคร ทิมหลัก สถานการณ์ และโครงเรื่องจากนิยายเรื่องหนึ่ง นำมาใช้ในการเปรียบเทียบกับนิยายเรื่องอื่นๆ เพื่อทำการวิเคราะห์ว่ามีความคล้ายคลึง มีความแตกต่างกันเพียงใด การประยุกต์ใช้เทคโนโลยีอัตโนมัติในกระบวนการนี้ ในเชิงทฤษฎีคาดว่าจะสามารถช่วยลดภาระงาน ระยะเวลาในการตรวจสอบ และลดความเสี่ยงจากข้อผิดพลาด (Human Error) ที่เกิดจากการใช้มนุษย์อ่านเปรียบเทียบข้อมูลจำนวนมากได้ [3]

ที่ผ่านมามีการพัฒนาระบบตรวจสอบความคล้ายคลึงของข้อความและตรวจจับการคัดลอก (Plagiarism Detection) อย่างแพร่หลาย เช่น ระบบ iThenticate, Turnitin [4] หรือ Copyleaks [5] อย่างไรก็ตามระบบเหล่านั้นส่วนใหญ่มุ่งเน้นการตรวจสอบวรรณกรรมทางวิชาการและอาศัยการจับคู่ข้อความแบบตรงตัว (Exact Keyword Matching) ซึ่งอาจไม่ครอบคลุมถึงความยืดหยุ่นทางภาษาในทางวรรณกรรม ระบบที่พัฒนาขึ้นในงานวิจัยนี้จึงมีความแตกต่างและมีข้อได้เปรียบเหนือระบบทั่วไป โดยออกแบบมาเพื่อรองรับการวิเคราะห์นิยายโดยเฉพาะด้วยการผสานเทคนิค Instance-based Learning และ K-Nearest Neighbors เพื่อประเมินบริบทเชิงสถิติ แทนการใช้กฎเกณฑ์ตายตัว (Rule-based) รวมถึงมีการแสดงผลลัพธ์ในรูปแบบภาพ (Data Visualization) เช่น แผนภาพเครือข่ายความสัมพันธ์ (Network Graph) ที่ช่วยให้นักวิจารณ์วรรณกรรมสามารถมองเห็นภาพรวมความเชื่อมโยงของนิยายหลายเรื่องพร้อมกันได้อย่างเป็นรูปธรรม

จากที่กล่าวมาข้างต้น เทคนิคการประมวลผลภาษาธรรมชาติ (NLP) ถูกนำมาประยุกต์ใช้ร่วมกับเทคนิคการเรียนรู้แบบอิงตัวอย่าง (Instance-based Learning) ผ่านอัลกอริทึม K-Nearest Neighbors สำหรับการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยาย โดยมีวัตถุประสงค์ดังนี้

1) เพื่อพัฒนาระบบวิเคราะห์และเปรียบเทียบความคล้ายคลึงของเนื้อหา นิยาย โดยประยุกต์ใช้เทคนิค NLP และ Instance-based Learning

2) เพื่อให้ระบบสามารถรองรับการนำเข้าข้อมูลจากไฟล์ประเภทต่าง ๆ (.txt, .docx, .pdf) และรองรับการประมวลผลทั้งนิยายที่เขียนด้วยภาษาไทยและภาษาอังกฤษ

3) เพื่อพัฒนาระบบแสดงผลการวิเคราะห์ (Visualization Module) ที่สามารถนำเสนอระดับความคล้ายคลึงในรูปแบบเชิงภาพที่หลากหลาย เช่น แผนที่ความร้อน (Heatmap) และแผนภาพเครือข่ายความสัมพันธ์ (Network Graph)

4) เพื่อประเมินประสิทธิภาพของระบบในการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของเนื้อหา นิยาย

2. วิธีดำเนินการวิจัย

วิธีการดำเนินการวิจัยระบบวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยาย แสดงเป็นขั้นตอนดังภาพที่ 1 ซึ่งสรุปได้ดังนี้

2.1. Tool and Library Installation

ระบบพัฒนาด้วยภาษา Python โดยใช้ไลบรารีสำคัญได้ดังนี้

1) PyThaiNLP สำหรับรองรับการประมวลผลนิยายในกรณีที่เป็นภาษาไทย

2) sklearn และ gensim สำหรับใช้ในการคำนวณความคล้ายคลึง เช่น Topic-level Similarity และ Cosine Similarity

3) matplotlib, seaborn และ networkx สำหรับการสร้างภาพแสดงผลลัพธ์ของข้อมูล

2.2. Data Collection and Preparation

Data Collection and Preparation เป็นการรวบรวมและเตรียมข้อมูล ในงานวิจัยนี้ ผู้วิจัยได้ทำการรวบรวมชุดข้อมูลที่เป็นนิยาย เพื่อใช้เป็นชุดข้อมูลทดสอบประสิทธิภาพของระบบ โดยมีการจำแนกหมวดหมู่ออกเป็น 5 หมวดหมู่ ดังตารางที่ 1

ทั้งนี้ แม้ระบบจะถูกพัฒนาให้รองรับทั้งภาษาไทยและภาษาอังกฤษ แต่ในการทดสอบประสิทธิภาพเชิงปริมาณ (Quantitative Evaluation) ครั้งนี้ คณะผู้วิจัยได้เลือกใช้ชุดข้อมูลทดสอบมาตรฐานเป็นนิยายภาษาอังกฤษจำนวน 50 เรื่อง เพื่อให้ง่ายต่อการควบคุมตัวแปรและเปรียบเทียบกับระบบอื่นในระดับสากล

2.3. Text Preprocessing

กระบวนการประมวลผลข้อความเบื้องต้น (Text Preprocessing) ถูกออกแบบมาเพื่อรองรับการประมวลผลแบบสองภาษา (Bilingual Pipeline) เพื่อให้ข้อมูลพร้อมสำหรับการวิเคราะห์ โดยมีขั้นตอนดังนี้

1) Language Detection (การตรวจจำภาษา) ระบบจะทำการ

ตรวจสอบภาษาของไฟล์ข้อความนำเข้าโดยอัตโนมัติ เพื่อแยกสายการประมวลผล (Pipeline) ระหว่างภาษาไทยและภาษาอังกฤษ

2) Tokenization (การตัดคำ) และ Stopwords Removal (การลบคำหยุด)

- กรณีภาษาไทย จะเรียกใช้ไลบรารี PyThaiNLP เพื่อจัดการกับบริบทของภาษาไทยที่ไม่มีการเว้นวรรคคำ พร้อมทั้งตัดคำหยุด (Stopwords) ที่ไม่มีความสำคัญ เช่น "และ", "คือ", "ที่"

- กรณีภาษาอังกฤษ จะใช้เครื่องมือในการตัดคำ (Word Tokenization) และลบคำหยุดมาตรฐานของภาษาอังกฤษ

1) Text Normalization (การปรับรูปแบบข้อความ): ตัวอักษรและข้อความของทั้งสองภาษาจะถูกแปลงให้อยู่ในรูปแบบมาตรฐานเดียวกัน เช่น การแปลงตัวพิมพ์เล็ก-ใหญ่ในภาษาอังกฤษ และลบเครื่องหมายวรรคตอน (Punctuation) ที่ไม่จำเป็น



ภาพที่ 1 ขั้นตอนการดำเนินการวิจัย

ตารางที่ 1 ชุดข้อมูลที่ใช้ในการทดสอบ

Genre	Novel	Chapters	Language
Action	10 เรื่อง	30 ตอน	English
Adventure	10 เรื่อง	30 ตอน	English
Historical	10 เรื่อง	30 ตอน	English
Romance	10 เรื่อง	30 ตอน	English
Supernatural	10 เรื่อง	30 ตอน	English
รวมทั้งหมด	50 เรื่อง	150 ตอน	-

2) Text Normalization (การปรับรูปแบบข้อความ): ตัวอักษรและข้อความของทั้งสองภาษาจะถูกแปลงให้อยู่ในรูปแบบมาตรฐานเดียวกัน เช่น การแปลงตัวพิมพ์เล็ก-ใหญ่ในภาษาอังกฤษ และลบเครื่องหมายวรรคตอน (Punctuation) ที่ไม่จำเป็น

3) Vectorization (การแปลงข้อความเป็นเวกเตอร์): ข้อมูลที่ผ่านการทำความสะอาดแล้วจะถูกแปลงให้อยู่ในรูปแบบเวกเตอร์โดยใช้เทคนิค [ระบุชื่อเทคนิค เช่น TF-IDF หรือ Bag-of-Words] เพื่อใช้สำหรับการคำนวณความคล้ายคลึงเชิงหัวข้อ (Topic-level similarity)

4) Stopwords Removal หรือ การลบคำที่ไม่สำคัญ จะลบคำเชื่อมและคำที่ไม่ช่วยแยกความหมาย เช่น คำว่า "และ" "คือ" หรือ คำว่า "ที่"

5) Text Normalization หรือ การปรับรูปแบบของข้อความ ตัวอักษร ข้อความ จะถูกแปลงให้เป็นรูปแบบเดียวกัน และลบเครื่องหมายที่ไม่จำเป็น

6) Vectorization หรือ การแปลงข้อความเป็นเวกเตอร์ เทคนิค Topic-level similarity ถูกใช้เพื่อแปลงข้อความเป็นเวกเตอร์สำหรับคำนวณความคล้ายคลึง

2.4. Analysis and Comparison of Similarity

ในการวิเคราะห์และเปรียบเทียบความคล้ายคลึงกัน เป็นกระบวนการที่ใช้หลักการทางคณิตศาสตร์โดยอาศัย การแทนค่าข้อมูลในรูปแบบเวกเตอร์ (Vector Representation) เพื่อวัดระยะห่างระหว่างข้อมูล กระบวนการนี้จัดเป็นเทคนิคแบบ Instance-based Learning ซึ่งระบบจะเรียนรู้ความคล้ายคลึงจากคุณลักษณะของข้อมูล (Data Features) โดยตรง วิธีการนี้มีความแตกต่างจากระบบแบบ Rule-based ทั่วไป เนื่องจากระบบไม่ได้พึ่งพากฎเกณฑ์หรือการจับคู่คีย์เวิร์ดที่ตายตัว (Exact Keyword Matching) แต่ใช้การคำนวณระยะห่างทางสถิติ ซึ่งทำให้มีความยืดหยุ่นและสามารถรองรับเนื้อหาใหม่ ๆ ได้อย่างมีประสิทธิภาพ

เพื่อสร้างมาตรฐานในการทดลองและให้สามารถทำซ้ำได้ (Reproducibility) คณะผู้วิจัยได้ประยุกต์ใช้ไลบรารี sklearn และ gensim ในกระบวนการแปลงเวกเตอร์และการคำนวณ โดยมีการกำหนดค่าพารามิเตอร์ (Hyperparameters) ที่สำคัญ ได้แก่ การใช้ฟังก์ชันเวกเตอร์โดยกำหนดค่าและใช้การวัดระยะห่างแบบ Cosine โดยมีขั้นตอนดังนี้

1) Similarity Calculation หรือ การคำนวณค่าความคล้ายคลึงระบบใช้การวัดระยะห่าง 2 รูปแบบ เพื่อวิเคราะห์ความคล้ายคลึงในระดับคำศัพท์ (Lexical Similarity) โดยมุ่งเน้นการเปรียบเทียบชุดคำสำคัญ (Keywords) ที่ปรากฏในเอกสาร ดังนี้

- Cosine Similarity ถูกใช้วัดทิศทางของเวกเตอร์เพื่อ

เปรียบเทียบโครงสร้างการกระจายตัวของคำศัพท์ โดยไม่ขึ้นกับความยาวของเอกสาร ซึ่งช่วยบ่งบอกความเหมือนของเอกสารผ่านความถี่ของคำที่ใช้ร่วมกัน

- Jaccard Similarity ถูกใช้วัดสัดส่วนของคำศัพท์ที่ปรากฏร่วมกัน เพื่อวัดความเหมือนของชุดคำสำคัญ (Keywords) ว่ามีการใช้คำศัพท์ชุดเดียวกันมากน้อยเพียงใด

2) Result Presentation ในการแสดงผลลัพธ์ระบบจะนำค่าระยะห่างที่คำนวณได้มาแปลงเป็นคะแนนความเหมือน (Similarity Score) ในรูปแบบเปอร์เซ็นต์ พร้อมทั้งทำการเน้นสี (Highlight) ข้อความที่มีความสัมพันธ์กัน เพื่อให้ผู้ใช้งานสามารถตีความผลลัพธ์เชิงปริมาณและเชิงคุณภาพได้โดยง่าย

2.5. Data Visualization

Data Visualization เป็นการนำข้อมูลมาแสดงด้วยกันเป็นภาพรวม ทำให้ข้อมูลเหล่านั้นถูกวิเคราะห์ได้ง่ายขึ้น โดยใช้เทคนิคดังนี้

- 1) Heatmap ใช้แสดงความคล้ายคลึงระหว่างเอกสารในรูปแบบตาราง สีเข้มหมายถึงความคล้ายสูง
- 2) Network Graph ใช้แสดงความสัมพันธ์ระหว่างนิยายด้วยโหนดและเส้นเชื่อม โหนดแทนเอกสาร เส้นเชื่อมแทนความคล้ายคลึง
- 3) Table and Matrix ใช้สำหรับตรวจสอบผลลัพธ์เชิงตัวเลขแบบละเอียด

2.6. Data Storage and Evaluation

การจัดเก็บและการประเมินผล ช่วยให้สามารถเรียกดูผลลัพธ์และวิเคราะห์ประสิทธิภาพของระบบได้ ข้อมูลและผลลัพธ์จะถูกเก็บในฐานข้อมูล และนำไปประเมินความแม่นยำของระบบในการจับคู่เอกสารได้ดังนี้

- 1) ผลลัพธ์การวิเคราะห์ถูกจัดเก็บในฐานข้อมูลสำหรับการอ้างอิงและประเมินผลในอนาคต
- 2) ประเมินความแม่นยำของระบบโดยการเปรียบเทียบ Similarity Score กับความเหมาะสมของ genre และเอกสารที่คล้ายกัน
- 3) ตัวอย่างผลลัพธ์ ดังนี้
 - Chapter 1 Outer Court Disciple.txt มีความคล้ายกับ Chapter 1 Cultivation Online.txt ที่คะแนน 0.281
 - Chapter 2 Beauty Elder.txt มีความคล้ายกับ Chapter 1 Rebirth.txt ที่คะแนน 0.210
 - Chapter 3 The Start of the Examination.txt มีความคล้ายกับ Chapter 1 Cultivation Online.txt ที่คะแนน 0.222

2.7. Result Summarization and Ranking

การสรุปและจัดอันดับช่วยให้ง่ายต่อการอ้างอิงและวิเคราะห์ ผลลัพธ์ที่แสดงจะเป็นรายการลำดับเอกสารในฐานข้อมูลที่ถูกจัดลำดับตามความ

คล้ายคลึงกับไฟล์นำเข้า ผลลัพธ์ที่แสดงนี้ช่วยให้ผู้ใช้เห็นเอกสารที่คล้ายกันมากที่สุด ซึ่งสามารถใช้ในการอ้างอิงหรือวิเคราะห์ต่อไป

3. ผลการวิจัยและอภิปรายผล

3.1. Novel Similarity Analysis

การวิเคราะห์ความคล้ายคลึงของนิยายที่ทดลองนำเข้าไฟล์ จำนวน 60 ไฟล์เพื่อนำไปเปรียบเทียบกับฐานข้อมูลที่มีอยู่ 150 ไฟล์ พบว่า ระบบสามารถระบุระดับความคล้ายคลึงและจัดอันดับเอกสารได้ โดยผลการวิเคราะห์ที่ได้แสดงให้เห็นว่า ไฟล์ที่นำเข้ามีความคล้ายคลึงในระดับต่ำถึงปานกลางกับเอกสารในฐานข้อมูล โดยมีคะแนนความคล้ายคลึงสูงสุดอยู่ที่ 0.281 ซึ่งเป็นระดับปกติสำหรับเนื้อหาที่ไม่ได้ลอกเลียนแบบโดยตรง ผลการจัดอันดับโดยรวมพบว่า Chapter 1 Cultivation Online.txt เป็นเอกสารที่มีคะแนนความคล้ายคลึงสูงสูดกับนิยายที่นำเข้าทั้งหมด

3.2. Performance Evaluation

การประเมินประสิทธิภาพของระบบ (Performance Evaluation) ดำเนินการโดยใช้วิธีการทดสอบเชิงสถิติ 2 รูปแบบ คือ Threshold Sensitivity Analysis และ Confusion Matrix เพื่อวัดความแม่นยำในการจำแนกประเภทและจับคู่ความคล้ายคลึง ได้ดังนี้

1) Threshold Sensitivity Analysis ทั้งนี้ผู้วิจัยได้ทำการทดสอบค่า Similarity Threshold ตั้งแต่ 0.4 ถึง 0.8 เพื่อหาจุดที่เหมาะสมที่สุด (Optimal Threshold) ที่ให้ค่าความแม่นยำสูงที่สุด ผลการทดสอบพบว่าที่ค่า Threshold = 0.60 ระบบสามารถสร้างสมดุลระหว่างค่า Precision และ Recall ได้ดีที่สุด ดังตารางที่ 2

2) Confusion Matrix จากการทดสอบที่ค่า Threshold 0.60 กับชุดข้อมูล 150 เอกสาร เพื่อประเมินความสามารถในการระบุประเภทนิยาย (Classification Accuracy) ผลลัพธ์แสดงดังตารางที่ 3

3.2.1 ข้อจำกัดของการประเมินประสิทธิภาพ

แม้ว่าสถาปัตยกรรมของระบบจะถูกออกแบบมาให้รองรับการทำงานแบบสองภาษาโดยบูรณาการไลบรารี PyThaiNLP เข้ามาในขั้นตอนการเตรียมข้อมูล อย่างไรก็ตาม เนื่องจากข้อจำกัดในการรวบรวมคลังข้อมูลจำลอง (Benchmark Dataset) สำหรับนิยายภาษาไทยที่เป็นมาตรฐาน การทดสอบประสิทธิภาพเชิงปริมาณ (Quantitative Evaluation) ในงานวิจัยฉบับนี้จึงจำกัดขอบเขตอยู่ที่ชุดข้อมูลภาษาอังกฤษจำนวน 50 เรื่องเป็นหลัก เพื่อให้สามารถควบคุมตัวแปรได้อย่างรัดกุม

นอกจากนี้ การประเมินประสิทธิภาพในปัจจุบันยังมุ่งเน้นไปที่การวิเคราะห์ความสอดคล้องภายในโมเดลผ่าน Threshold Sensitivity Analysis เป็นหลัก การเปรียบเทียบคะแนนประสิทธิภาพกับระบบตรวจจับมาตรฐานอื่น (Baseline Systems) หรือการใช้ผู้เชี่ยวชาญด้านวรรณกรรม (Human Annotators) มาร่วมประเมินความสอดคล้อง ถือเป็นข้อจำกัดของงานวิจัยในระยะนี้ ซึ่งคณะผู้วิจัยได้กำหนดให้เป็น

แนวทางสำคัญสำหรับการทดลองและพัฒนาในอนาคต (Future Work) เพื่อยืนยันความน่าเชื่อถือของระบบในระดับสากลต่อไป

3.3. Data Visualization

การแสดงผลลัพธ์ในรูปแบบภาพ (Data Visualization) เพื่อให้เข้าใจง่ายและเห็นความสัมพันธ์ระหว่างเอกสารอย่างชัดเจนได้ดังนี้

- 1) Heatmap สำหรับใช้แสดงค่าความคล้ายคลึงระหว่างประเภทของนิยาย ช่องที่ตัดกันระหว่าง ACTION NOVELS กับ ACTION NOVELS มีสีเข้มที่สุด แสดงถึงความแม่นยำสูงในการจัดกลุ่มประเภท
- 2) Network Graph สำหรับใช้แสดงความสัมพันธ์ระหว่างนิยายที่มีความคล้ายคลึงกัน เช่น Chapter 1 Outer Court Disciple.txt, Chapter 3 The Start of the Examination .txt, และ Chapter 1 Cultivation Online.txt ซึ่งสอดคล้องกับผลการวิเคราะห์เชิงตัวเลข

3.4. Visual Reporting

ระบบสามารถนำเสนอผลลัพธ์เป็นรายงานในรูปแบบต่างๆเพื่อช่วยให้การตรวจสอบ และการศึกษาความเหมือนหรือความแตกต่างของนิยายแต่ละเรื่องเป็นไปได้ง่ายมีประสิทธิภาพและสะดวกขึ้น การดูข้อมูลรายงานสามารถดูได้จากการแสดงผลที่หน้าเว็บแอปพลิเคชันและสามารถ Export ออกมาในรูปแบบของไฟล์ PDF ได้ นอกจากนี้ ยังสามารถนำข้อมูลออกมาในรูปแบบไฟล์ CSV และ JSON สำหรับการนำข้อมูลไปใช้ในการวิเคราะห์ต่อยอดโดยนักวิจัยหรือนักพัฒนาต่อไป

ตารางที่ 2 แสดงค่า Precision Recall และ F1-Score

Threshold	Precision	Recall	F1-Score
0.40	0.65	0.95	0.77
0.50	0.78	0.88	0.82
0.60	0.85	0.82	0.83
0.70	0.92	0.65	0.86
0.80	0.96	0.45	0.61

ตารางที่ 3 Confusion Matrix ของการจำแนกประเภทนิยาย

	ทำนายว่าเหมือน (Predicted Positive)	ทำนายว่าไม่เหมือน (Predicted Negative)
เหมือน		
(Actual Positive)	123	27
ไม่เหมือน		
(Actual Negative)	22	128

3.5. Keyword-Based Semantic Analysis

นอกจากการวัดความคล้ายคลึงของคำศัพท์ ระบบยังสามารถนำผลลัพธ์มาขยายผลเพื่อวิเคราะห์เนื้อหาโดยอาศัยชุดคำสำคัญ (Keyword-Based Analysis) ได้ดังนี้

- 1) วิเคราะห์แนวทางของนิยายจากชุดคำศัพท์ ใช้การตรวจสอบกลุ่มคำที่พบบ่อย เพื่อระบุแนวทางของเรื่อง เช่น หากพบชุดคำที่เกี่ยวข้องกับ "ดาบ เวทมนตร์ หรือ มังกร" สามารถจัดกลุ่มเป็นแนวแฟนตาซีได้ ซึ่งช่วยให้เข้าใจความสัมพันธ์ของนิยายแต่ละเล่มผ่านรูปแบบภาษาที่ใช้
- 2) ตรวจสอบความสัมพันธ์ระหว่างกลุ่มนิยาย ข้อมูลความคล้ายคลึงของคำศัพท์ (Lexical Similarity) สามารถนำไปใช้จัดกลุ่มเอกสารที่มีเนื้อหาใกล้เคียงกัน หรือศึกษาความเชื่อมโยงของงานเขียนที่มีรูปแบบการใช้คำซ้ำกัน ซึ่งเป็นประโยชน์ต่อการสืบค้นและการวิจัยความสัมพันธ์ด้านวรรณกรรม

3.6. อภิปรายผล (Discussion)

จากผลการวิจัยและทดสอบประสิทธิภาพของระบบ สามารถนำมาอภิปรายผลได้ดังนี้

- 1) ประสิทธิภาพการวิเคราะห์และจัดอันดับ การที่ระบบสามารถระบุคะแนนความคล้ายคลึงสูงสุดได้ที่ 0.281 สำหรับเอกสารที่ไม่ได้มีการคัดลอกโดยตรง แสดงให้เห็นว่าการใช้เทคนิค Cosine Similarity ร่วมกับ K-Nearest Neighbors มีความไว (Sensitivity) เพียงพอที่จะตรวจสอบความเชื่อมโยงของเนื้อหาและคำศัพท์ที่ใช้ร่วมกันในหมวดหมู่เดียวกันได้ โดยไม่เกิดผลบวกสูง (False Positive) ในระดับที่สูงเกินไป
- 2) ความเหมาะสมของเกณฑ์ (Threshold) จากการทดสอบ Threshold Sensitivity Analysis พบว่าค่า 0.60 เป็นจุดที่เหมาะสมที่สุด (Optimal Threshold) ที่ให้ค่า F1-Score สูงถึงร้อยละ 83.4 สะท้อนให้เห็นว่าในบริบทของการวิเคราะห์งานวรรณกรรม ซึ่งมีความยืดหยุ่นของการใช้ภาษาและคำพ้องความหมายสูง การตั้งค่าเกณฑ์ที่เข้มงวดเกินไป (เช่น 0.80) จะทำให้ค่า Recall ลดลงอย่างรวดเร็ว (เหลือร้อยละ 45) เนื่องจากการตัดเอกสารที่มีความคล้ายคลึงเชิงบริบทออกไป
- 3) การแสดงผลเชิงภาพ (Data Visualization) การแปลงผลลัพธ์เชิงตัวเลขให้อยู่ในรูปแบบ Heatmap และ Network Graph ช่วยแก้ปัญหาความซับซ้อนในการตีความข้อมูลจำนวนมาก สอดคล้องกับแนวคิดที่ว่า การแสดงผลเชิงสัมพันธ์แบบกราฟ (Graph-based Relationship) ช่วยให้นักวิจัยวรรณกรรมสามารถมองเห็นกลุ่มก้อน (Clustering) และความเชื่อมโยงข้ามเนื้อหา (Cross-document Linkage) ได้อย่างเป็นรูปธรรมมากกว่าการพิจารณาเป็นรายเอกสาร

ตัวอย่างการแสดงผลระบบการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยาย

การทำงานในหน้าหลักของเว็บแอปพลิเคชันในงานวิจัยนี้ ประกอบด้วย ส่วนของการอัปโหลด ผู้ใช้สามารถเลือกไฟล์นิยายหรือ

เอกสารที่ต้องการนำมาวิเคราะห์ โดยระบบสามารถรองรับไฟล์ .txt .docx .pdf และอัปโหลดไฟล์ได้สูงสุด 60 ไฟล์ เมื่ออัปโหลดไฟล์แล้ว ระบบจะแสดงรายการไฟล์ที่พร้อมใช้งาน และสามารถตั้งชื่อนิยายที่อัปโหลด เช่น Martial God Asura เพื่อให้ง่ายต่อการจัดการและอ้างอิง ดังภาพที่ 2

จากภาพที่ 3 ส่วนของการตั้งค่าต่าง ๆ ได้แก่

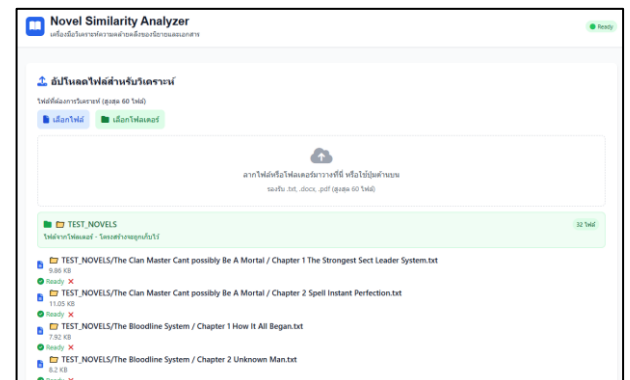
- 1) K-Neighbors เป็นการกำหนดค่าจำนวนเอกสารที่คล้ายกัน คือ การกำหนดค่าสัมประสิทธิ์ K สำหรับอัลกอริทึม K-Nearest Neighbors เพื่อใช้ในการค้นหา K-Neighbors ที่ใกล้เคียงที่สุด เช่น

- Training Set จะเป็นไฟล์ฐานข้อมูลนิยาย (.zip) ที่ผู้ใช้นำเข้าสู่ระบบ

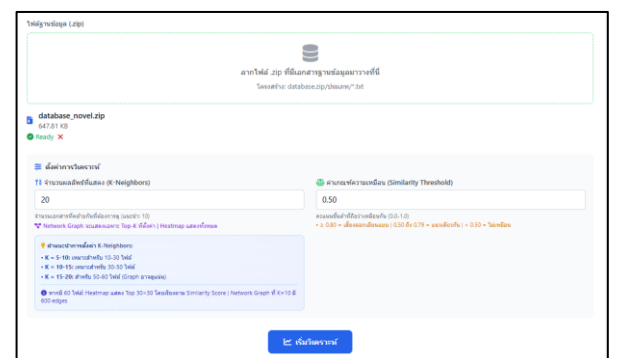
- Label เป็นข้อมูลประเภทของนิยาย (Genre) และชื่อเรื่องที่กำกับอยู่กับเอกสารในฐานข้อมูลนิยาย

โดยกระบวนการทำงาน เริ่มต้นจะคำนวณระยะห่างเพื่อจัดลำดับความคล้ายคลึง (Similarity Ranking) และดึงเอกสารจำนวน K เรื่องที่มีความเหมือนสูงสุดออกมา เพื่อใช้ในการวิเคราะห์แนวโน้มของประเภทนิยายจากรูปแบบนิยายของเอกสารในลำดับดังกล่าว

- 1) Target Accuracy คือ ค่าความน่าจะเป็นขั้นต่ำ (0.0-1.0) ที่ระบบใช้เป็นเกณฑ์ในการตัดสินใจเพื่อจำแนกประเภทนิยาย หากค่าความ



ภาพที่ 2 ตัวอย่างหน้าแรกของเว็บแอปพลิเคชันสำหรับวิเคราะห์ความคล้ายคลึงของนิยายจากการอัปโหลดไฟล์



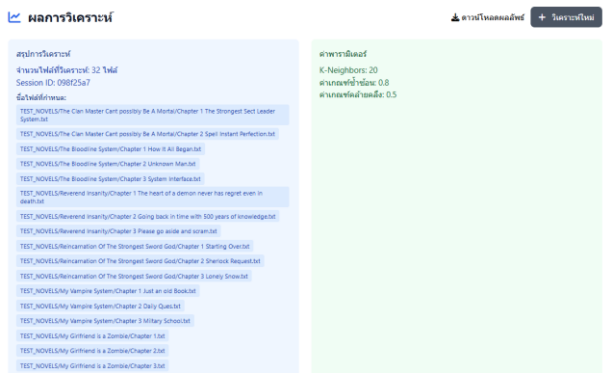
ภาพที่ 3 ตัวอย่างการเลือกเพิ่มไฟล์ฐานข้อมูลและตั้งค่าเกณฑ์ซ้ำซ้อน

เชื่อมั่นของผลลัพธ์ต่ำกว่าเกณฑ์ที่กำหนด ระบบจะไม่นำผลลัพธ์นั้นมาสรุปเป็นคำตอบ

2) Similarity Threshold โดยค่าคะแนนความเหมือน (0.0-1.0) ที่ใช้สำหรับกรองผลลัพธ์การค้นหา โดยระบบจะแสดงผลเฉพาะเอกสารที่มีค่า Cosine Similarity Score สูงกว่าค่าที่กำหนดเท่านั้น เพื่อลด Noise จากเอกสารที่ไม่เกี่ยวข้อง

จากภาพที่ 4 แสดงข้อมูลจำนวนไฟล์ที่ถูกวิเคราะห์ทั้งหมด 32 ไฟล์ ค่า Session ID และรายการชื่อไฟล์ ซึ่งในส่วนของค่าพารามิเตอร์จะแสดงข้อมูล ค่า K-Neighbors ค่าเกณฑ์ซ้ำซ้อน และค่าคล้ายคลึง โดยที่ค่าเกณฑ์ซ้ำซ้อน (Duplicate Threshold) ที่ได้คือ 0.8 และค่าเกณฑ์คล้ายคลึง (Similarity Threshold) คือ 0.5 ซึ่งผลการตรวจสอบพบว่าความคล้ายคลึงกันของนิยายอยู่ในเกณฑ์ 'ปานกลางค่อนข้างต่ำ' ทั้งนี้การแปลผลดังกล่าวอ้างอิงจากเกณฑ์การวัดระดับความคล้ายคลึงของระบบ (อ้างอิงตามค่า Heatmap) โดยกำหนดให้ค่าที่ใกล้เคียง 1.0 หมายถึงความคล้ายคลึงสูง ค่าระหว่าง 0.5-0.7 หมายถึงความคล้ายคลึงระดับปานกลาง (คล้ายคลึงบางส่วน) และค่าที่ใกล้เคียง 0.0 หมายถึงมีความคล้ายคลึงต่ำหรือแตกต่างกันมาก

จากภาพที่ 5 แสดงผลของการเปรียบเทียบความคล้ายคลึงของนิยายแต่ละเรื่องที่ต้องการตรวจสอบกับนิยายต่าง ๆ ที่มีอยู่ในฐานข้อมูล โดย



ภาพที่ 4 ตัวอย่างผลการวิเคราะห์ความคล้ายคลึงกันของนิยาย

Analysis by Input File		
Input: Earth's Greatest Magus Chapter 1: Magus Academy		
TITLE / DATABASE FILE	GENRE	SIMILARITY
Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	30.08%
Rebirth Of A Genius Creator and Destroyer - Chapter 2 New World.txt Rebirth Of A Genius Creator and Destroyer	Novel	28.95%
Death And Me - Chapter 1 I Don't Wanna.txt Death And Me	Novel	28.86%
I Am Loaded with Passive Skills - Chapter 2 Corpse Collectors.txt I Am Loaded with Passive Skills	Novel	24.3%
Walker Of The Worlds - Chapter 3 The Mysterious Ring.txt Walker Of The Worlds	Novel	23.73%
Cultivation Online - Chapter 3 Becoming a Cultivator.txt Cultivation Online	Novel	22.88%
I Am Loaded with Passive Skills - Chapter 1 Black Wheel.txt I Am Loaded with Passive Skills	Novel	22.83%
Rebirth Of A Genius Creator and Destroyer - Chapter 1 Sam.txt Rebirth Of A Genius Creator and Destroyer	Novel	22.8%
Walker Of The Worlds - Chapter 1 Lin Mu.txt Walker Of The Worlds	Novel	21.63%
Parallel Memory - Chapter 2 Me And Lul.txt Parallel Memory	Novel	21.49%
Nine Star Hegemon Body Arts - Chapter 2 Despicable Scum.txt Nine Star Hegemon Body Arts	Novel	21.08%

ภาพที่ 5 ตัวอย่างการวิเคราะห์ตามไฟล์อินพุต

ระบบสามารถจัดลำดับความคล้ายคลึงสูงสุดไปต่ำสุด และแสดงชื่อนิยาย หรือชื่อไฟล์ได้

ในส่วนของการวิเคราะห์เพื่อจัดลำดับความคล้ายคลึงจากสูงสุดไปยังต่ำสุดของแต่ละ Chapter ของนิยายที่เปรียบเทียบแสดงได้ดังภาพที่ 6 - ภาพที่ 8 และตัวอย่างของรายงานสรุป แสดงได้ดังภาพที่ 9

Input: Chapter 1 Awakening After Rebirth		
TITLE / DATABASE FILE	GENRE	SIMILARITY
Cultivation Online - Chapter 2 Mysterious Little Girl.txt Cultivation Online	Novel	23.44%
Cultivation Online - Chapter 3 Becoming a Cultivator.txt Cultivation Online	Novel	23.36%
Nine Star Hegemon Body Arts - Chapter 2 Despicable Scum.txt Nine Star Hegemon Body Arts	Novel	19.73%
I Am Loaded with Passive Skills - Chapter 1 Black Wheel.txt I Am Loaded with Passive Skills	Novel	19.1%
Soul Land IV (Douluo Dalu) Ultimate Fighting - Chapter 3 A distressed start.txt Soul Land IV (Douluo Dalu) Ultimate Fighting	Novel	18.99%
Nine Star Hegemon Body Arts - Chapter 1 Memories of a Pill God.txt Nine Star Hegemon Body Arts	Novel	18.43%
Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	18%
Rebirth Of A Genius Creator and Destroyer - Chapter 2 New World.txt Rebirth Of A Genius Creator and Destroyer	Novel	17.63%
Nine Star Hegemon Body Arts - Chapter 3 Revealing his Sharp Side.txt Nine Star Hegemon Body Arts	Novel	17.49%
I Am Loaded with Passive Skills - Chapter 2 Corpse Collectors.txt I Am Loaded with Passive Skills	Novel	17.39%
Walker Of The Worlds - Chapter 3 The Mysterious Ring.txt Walker Of The Worlds	Novel	16.92%
The Supreme Martial King Shocking All Realms - Chapter 1 Fancious Older Sister_1.txt The Supreme Martial King Shocking All Realms	Novel	16.59%

ภาพที่ 6 ตัวอย่างไฟล์ที่วิเคราะห์ของ Chapter 1

Input: Chapter 2 Camp Of The Seven Sages		
TITLE / DATABASE FILE	GENRE	SIMILARITY
I Am Loaded with Passive Skills - Chapter 2 Corpse Collectors.txt I Am Loaded with Passive Skills	Novel	26.26%
Alchemy Emperor of the Divine Dao - Chapter 3 You're Sick I Can Heal You.txt Alchemy Emperor of the Divine Dao	Novel	19.99%
Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	18.86%
Alchemy Emperor of the Divine Dao - Chapter 2 Might.txt Alchemy Emperor of the Divine Dao	Novel	16.65%
Rebirth Of A Genius Creator and Destroyer - Chapter 2 New World.txt Rebirth Of A Genius Creator and Destroyer	Novel	16.28%
Walker Of The Worlds - Chapter 3 The Mysterious Ring.txt Walker Of The Worlds	Novel	16.12%
Cultivation Online - Chapter 3 Becoming a Cultivator.txt Cultivation Online	Novel	16.03%
Nine Star Hegemon Body Arts - Chapter 2 Despicable Scum.txt Nine Star Hegemon Body Arts	Novel	15.8%
Soul Land IV (Douluo Dalu) Ultimate Fighting - Chapter 3 A distressed start.txt Soul Land IV (Douluo Dalu) Ultimate Fighting	Novel	14.94%
Cultivation Online - Chapter 2 Mysterious Little Girl.txt Cultivation Online	Novel	14.68%
I Am Loaded with Passive Skills - Chapter 1 Black Wheel.txt I Am Loaded with Passive Skills	Novel	14.57%
Rebirth Of A Genius Creator and Destroyer - Chapter 1 Sam.txt Rebirth Of A Genius Creator and Destroyer	Novel	14.31%

ภาพที่ 7 ตัวอย่างไฟล์ที่วิเคราะห์ของ Chapter 2

Input: Chapter 3 Seventh Level Of The Qi Refinement Realm, Too Charming		
TITLE / DATABASE FILE	GENRE	SIMILARITY
Alchemy Emperor of the Divine Dao - Chapter 3 You're Sick I Can Heal You.txt Alchemy Emperor of the Divine Dao	Novel	41.96%
Alchemy Emperor of the Divine Dao - Chapter 2 Might.txt Alchemy Emperor of the Divine Dao	Novel	39.69%
Alchemy Emperor of the Divine Dao - Chapter 1 Rebirth.txt Alchemy Emperor of the Divine Dao	Novel	37.92%
Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	20.21%
Cultivation Online - Chapter 3 Becoming a Cultivator.txt Cultivation Online	Novel	18.7%
I Am Loaded with Passive Skills - Chapter 1 Black Wheel.txt I Am Loaded with Passive Skills	Novel	18.38%
I Am Loaded with Passive Skills - Chapter 2 Corpse Collectors.txt I Am Loaded with Passive Skills	Novel	18.14%
Rebirth Of A Genius Creator and Destroyer - Chapter 2 New World.txt Rebirth Of A Genius Creator and Destroyer	Novel	18.12%
Nine Star Hegemon Body Arts - Chapter 2 Despicable Scum.txt Nine Star Hegemon Body Arts	Novel	17.73%
Walker Of The Worlds - Chapter 3 The Mysterious Ring.txt Walker Of The Worlds	Novel	17.37%
Cultivation Online - Chapter 2 Mysterious Little Girl.txt Cultivation Online	Novel	16.52%
Nine Star Hegemon Body Arts - Chapter 3 Revealing his Sharp Side.txt Nine Star Hegemon Body Arts	Novel	15.18%

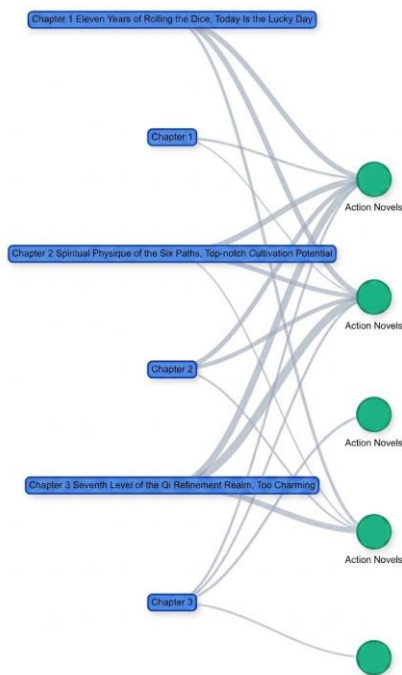
ภาพที่ 8 ตัวอย่างไฟล์ที่วิเคราะห์ของ Chapter 3

INPUT FILE	BEST MATCH	GENRE	SCORE	STATUS	TOP GENRES
Chapter 1 - Miscellaneous Info.txt	Parallel Memory - Chapter 1 Prologue.txt Parallel Memory	Novel	15.0%	Difficult	Novelman0.15,mean0.04
Chapter 1 Awakening after Rebirth.txt	Cultivation Online - Chapter 2 Mysterious Little Girl.txt Cultivation Online	Novel	23.0%	Difficult	Novelman0.23,mean0.12
Chapter 1 Eleven Years of Rolling the Dice. Today is the Lucky Day.txt	Alchemy Emperor of the Divine Dao - Chapter 2 Might.txt Alchemy Emperor of the Divine Dao	Novel	37.0%	Difficult	Novelman0.37,mean0.13
Chapter 1 End Of Misery.txt	Infinite Mana In The Apocalypse - Chapter 1 Awaken.txt Infinite Mana In The Apocalypse	Novel	35.0%	Difficult	Novelman0.35,mean0.14
Chapter 1 How It All Began.txt	Walker Of The Worlds - Chapter 3 The Mysterious Ring.txt Walker Of The Worlds	Novel	28.0%	Difficult	Novelman0.28,mean0.17
Chapter 1 Just an old Book.txt	I Am Loaded with Passive Skills - Chapter 2 Corpse Collector.txt I Am Loaded with Passive Skills	Novel	27.0%	Difficult	Novelman0.27,mean0.17
Chapter 1 Lu Ya.txt	Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	19.0%	Difficult	Novelman0.19,mean0.17
Chapter 1 Mage Academy.txt	Cultivation Online - Chapter 1 Cultivation Online.txt Cultivation Online	Novel	30.0%	Difficult	Novelman0.30,mean0.14
Chapter 1 Outer Court Disciple.txt	Reincarnation of the Strongest Healer - Chapter 2 Preparation.txt Reincarnation of the Strongest Healer	Novel	41.0%	Difficult	Novelman0.41,mean0.13
Chapter 1 Prologue.txt	Death And Me - Chapter 1 Dont Wanna.txt Death And Me	Novel	38.0%	Difficult	Novelman0.38,mean0.18
Chapter 1 Reborn.txt	Infinite Mana In The Apocalypse - Chapter 1 Awaken.txt Infinite Mana In The Apocalypse	Novel	43.0%	Difficult	Novelman0.43,mean0.15

ภาพที่ 9 ตัวอย่างรายงานสรุป

ผลการจับคู่โดยระบบ		
	Novel	45% จำนวนการจับคู่
ค่าเฉลี่ย: 13%		
Top 20 รายการโดยอันดับ		
อันดับ	ชื่อเรื่อง / Selected	คะแนน
1	Reincarnation of the Strongest Healer - Chapter 2 Preparation.txt Reincarnation of the Strongest Healer	45%
2	Infinite Mana In The Apocalypse - Chapter 1 Awaken.txt Infinite Mana In The Apocalypse	43%
3	Global Game AFK In The Zombie Apocalypse Game - Chapter 1.txt Global Game AFK In The Zombie Apocalypse Game	42%
#4	Alchemy Emperor of the Divine Dao - Chapter 3 You're Sick I Can Heal You.txt Alchemy Emperor of the Divine Dao	42%
#5	Reincarnation of the Strongest Healer - Chapter 3 Login.txt Reincarnation of the Strongest Healer	40%
#6	Alchemy Emperor of the Divine Dao - Chapter 2 Might.txt Alchemy Emperor of the Divine Dao	40%
#7	Parallel Memory - Chapter 2 Me And I.txt Parallel Memory	39%
#8	Death And Me - Chapter 1 Dont Wanna.txt Death And Me	38%
#9	Infinite Mana In The Apocalypse - Chapter 3 Apes Paradise.txt Infinite Mana In The Apocalypse	38%

ภาพที่ 10 ตัวอย่างการจัดอันดับ 10 อันดับ เอกสารที่คล้ายคลึงที่สุด



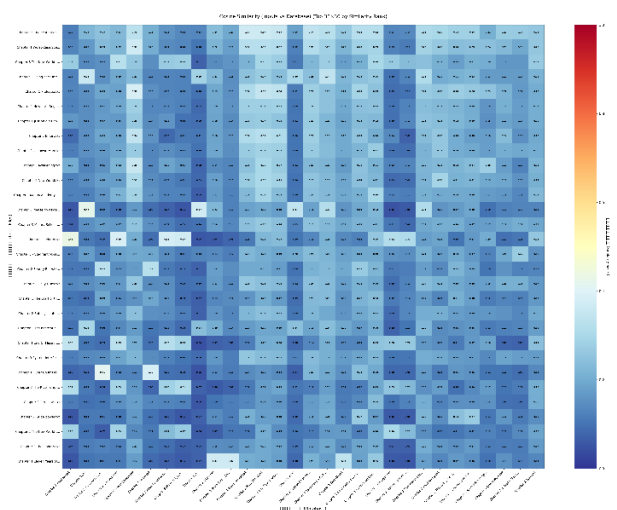
ภาพที่ 11 กราฟเครือข่ายของความสัมพันธ์ระหว่างไฟล์นิยายที่ตรวจสอบและนิยายที่อยู่ในฐานข้อมูล

ตารางเปรียบเทียบ (Comparison Table)									
* ข้อมูล: ตารางที่แสดงการเปรียบเทียบโครงเรื่องภายในเอกสารโดยอิงตาม Similarity (top_similarity) และการเชื่อมโยงโครงเรื่อง (top_genre) และความสัมพันธ์ (top_similarity) และความสัมพันธ์ (top_similarity).									
Input disc	Input_similarity	top_disc	top_genre	top_genre_title	top_folder_name	top_chapter_name	top_display_name	top_similarity	rel
Chapter 1 The heart of a desert never has regret even in death.txt	50 matches	Chapter 1.txt	Novel	Adventure Novels	Adventure Novels	Chapter 1	Adventure Novels - Chapter 1	29.4%	+
Chapter 2 Going back in time 500 years of knowledge.txt	50 matches	Chapter 1.txt	Novel	Adventure Novels	Adventure Novels	Chapter 1	Adventure Novels - Chapter 1	35.6%	+
Chapter 3 Please go with me and return.txt	50 matches	Chapter 1.txt	Novel	Adventure Novels	Adventure Novels	Chapter 1	Adventure Novels - Chapter 1	42.0%	+

ภาพที่ 12 ตัวอย่างตารางเปรียบเทียบความคล้าย

	Chapter 1 Rebirth.txt	Chapter 1 Might.txt	Chapter 3 You're Sick I Can Heal You.txt	Chapter 1 Cultivation Online.txt	Chapter 2 Mysterious Little Girl.txt	Chapter 3 Becoming a Cultivator.txt	Chapter 1 Reborn as a Wolf.txt	Chapter 2 Automating Growth.txt	Chapter 3 Going Out Hunting.txt	Chapter 1 Black Wheat.txt	Chapter 1 Corpse Collector
Chapter 1 - Miscellaneous Info.txt	7.8%	5.0%	6.3%	4.6%	1.9%	3.5%	4.2%	4.0%	2.6%	2.8%	4.5%
Chapter 1 Awakening after Rebirth.txt	12.1%	14.1%	11.7%	10.2%	23.4%	23.4%	11.2%	7.2%	6.1%	18.1%	17.4%
Chapter 1 Eleven Years of Rolling the Dice, Today is the Lucky Day.txt	33.7%	37.1%	37.0%	26.5%	17.8%	17.8%	9.7%	5.9%	5.5%	20.3%	21.1%
Chapter 1 End Of Misery.txt	13.1%	7.0%	7.8%	23.0%	16.2%	20.7%	9.4%	4.2%	3.7%	11.8%	13.5%
Chapter 1 How It All Began.txt	19.4%	19.1%	15.6%	24.8%	24.0%	24.9%	14.4%	7.5%	8.7%	27.6%	27.9%
Chapter 1 Just an old Book.txt	18.7%	18.4%	14.3%	22.9%	24.8%	22.4%	13.6%	8.2%	8.6%	26.2%	27.1%
Chapter 1 Lu Ya.txt	11.1%	12.9%	15.9%	15.2%	15.4%	16.2%	8.4%	5.6%	6.9%	14.6%	18.9%
Chapter 1 Mage Academy.txt	19.2%	16.9%	12.5%	30.1%	20.8%	22.9%	12.0%	5.8%	8.0%	22.8%	24.3%

ภาพที่ 13 ตัวอย่างตารางเมทริกซ์ความคล้ายคลึง



ภาพที่ 14 ตัวอย่างการแสดงความคล้ายคลึงด้วย Heatmap

จากภาพที่ 9 แสดงรายงานสรุปผลการวิเคราะห์ความคล้ายคลึงของนิยายที่นำเข้า (Input) โดยแจกแจงรายละเอียดค่าความคล้ายคลึงในแต่ละตอน (Chapter) พร้อมทั้งแสดงค่าเฉลี่ย (Mean) นอกจากนี้ ระบบยังนำเสนอสรุปผลรวมของการจัดอันดับนิยายในฐานข้อมูลที่มีความสอดคล้องและจับคู่ได้ดีที่สุด (Best Match) กับนิยายที่นำมาตรวจสอบ

จากภาพที่ 10 แสดงการจัดอันดับ 10 อันดับ เอกสารที่คล้ายคลึงที่สุด ได้แก่ ลำดับ 1-10 ชื่อเรื่อง/เอกสาร หมวดหมู่ รวมทั้งความคล้าย

ในการแสดงกราฟเครือข่ายของความสัมพันธ์ระหว่างโพลีนิยายที่ตรวจสอบและนิยายที่อยู่ในฐานข้อมูลแสดงได้ดังภาพที่ 11 ได้ มีรายละเอียดดังนี้

- 1) รูปสี่เหลี่ยม แสดงถึง โพลีนิยายที่นำมาวิเคราะห์
- 2) รูปวงกลม แสดงถึง เอกสารในฐานข้อมูลที่มีความคล้ายคลึงสูง
- 3) เส้นเชื่อม ความหนา และความยาวของเส้นเชื่อมแสดงระดับ

ความคล้ายมาก หรือน้อย

จากภาพที่ 12 แสดงตารางการเปรียบเทียบความคล้าย โดยแสดงรายละเอียดของแต่ละโพลีนิยายที่นำเข้า รวมถึงนิยายที่คล้ายคลึงที่สุดที่อยู่ในฐานข้อมูล คะแนนความคล้ายคลึง และการจัดประเภทความสัมพันธ์

การแสดงความคะแนนความคล้ายคลึงระหว่างโพลีนิยายกับเอกสารทั้งหมดในฐานข้อมูลในรูปแบบตารางเมทริกซ์ สามารถแสดงได้ดังภาพที่ 13 โดย

- ค่าใกล้ 1 คือ คล้ายคลึงมาก
- ค่าใกล้ 0 คือ แตกต่างมาก

ผลลัพธ์ที่ได้จากตัวอย่างนี้ถือว่ามีความคล้ายคลึงกันไม่มาก

นอกจากนี้ ระบบยังมีในส่วนของการแสดงแผนที่ความร้อนความคล้ายคลึงดังภาพที่ 14 ที่มีรายละเอียดดังนี้

- 1) สีแดงเข้ม มีค่าความคล้ายสูง (ใกล้ 1.0) แสดงว่า เนื้อหาที่มีความคล้ายคลึงมาก
- 2) สีเหลือง มีค่าความคล้ายปานกลาง (0.5-0.7) แสดงว่า มีเนื้อหาคล้ายบางส่วน
- 3) สีน้ำเงิน มีค่าความคล้ายต่ำ (ใกล้ 0.0) แสดงว่า เนื้อหาที่แตกต่างกัน

4. สรุปผล

ผลการวิจัยจากการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยายสามารถให้ข้อมูลที่สำคัญเชิงลึกเกี่ยวกับความสัมพันธ์ของเนื้อหาหลายเรื่องได้อย่างมีประสิทธิภาพ จากการวิเคราะห์นิยายที่นำเข้า ระบบสามารถคำนวณคะแนนความคล้ายคลึงที่แสดงถึงระดับการซ้ำซ้อนของเนื้อหาได้ และสามารถระบุได้ว่าไฟล์ใดเป็นเอกสารที่มีคะแนนความคล้ายคลึงสูงที่สุดกับไฟล์ที่นำเข้า ซึ่งบ่งชี้ว่าระบบมีความสามารถในการจัดอันดับความสัมพันธ์ของเนื้อหาได้เช่นเดียวกัน

การนำเทคนิคประมวลผลภาษาธรรมชาติที่เหมาะสมกับภาษาไทยมาใช้ เช่น การนำไลบรารี PyThaiNLP มาใช้จัดการการตัดคำ และลบคำหยุด ซึ่งเป็นปัญหาเฉพาะของภาษาไทย ทำให้ระบบสามารถวิเคราะห์เนื้อหาได้อย่างแม่นยำขึ้นในส่วนของการจับคู่ความเหมือนระหว่างนิยาย การนำเสนอผลลัพธ์ ระบบ

ได้แสดงผลเป็นภาพ ซึ่งช่วยให้ผู้ใช้เข้าใจข้อมูลที่ซับซ้อนได้อย่างชัดเจน โดยใช้แผนที่ความร้อน (Heatmap) จะแสดงถึงความสัมพันธ์ระหว่างประเภทของนิยาย ซึ่งช่องที่ตัดกันระหว่าง ACTION NOVELS กับ ACTION NOVELS จะมีสีเข้มที่สุด ในขณะที่แผนภาพเครือข่ายความสัมพันธ์ (Network Graph) จะแสดงถึงความเชื่อมโยงของนิยายที่มีความคล้ายคลึงกัน โดยมีเส้นเชื่อมโยงระหว่างโพลีที่มีคะแนนความคล้ายคลึงสูงสุด

การวิจัยนี้ ในอนาคตสามารถนำผลที่ได้รับไปปรับใช้ได้ เช่น ใช้ในการตรวจสอบการคัดลอกผลงานสำหรับสำนักพิมพ์ นักเขียน หรือใช้เป็นเครื่องมือสำหรับนักวิจัยด้านวรรณกรรมในการศึกษาความเชื่อมโยงระหว่างงานเขียนต่าง ๆ นอกจากนี้ยังสามารถต่อยอดเพื่อเพิ่มความสามารถในการวิเคราะห์ความเหมือนเชิงโครงสร้าง (Plot Structure) และแนวคิดหลัก (Theme) ในระดับที่ลึกขึ้น โดยอาจประยุกต์ใช้โมเดลภาษาขนาดใหญ่ (Large Language Models: LLMs) เข้ามาช่วยสกัดบริบทเชิงความหมายที่ซับซ้อน หรือใช้เทคโนโลยีโครงข่ายประสาทแบบกราฟ (Graph Neural Networks: GNN) เพื่อวิเคราะห์ความสัมพันธ์ของตัวละครและลำดับเหตุการณ์ ซึ่งจะเป็ประโยชน์อย่างยิ่งต่อการพัฒนาการวรรณกรรมต่อไป อย่างไรก็ตาม เพื่อให้มีหลักฐานเชิงประจักษ์ที่ครอบคลุมมากยิ่งขึ้น งานวิจัยในอนาคตควรมีการเก็บสถิติระยะเวลาการประมวลผล (Execution Time) ของระบบ เปรียบเทียบกับระยะเวลาเฉลี่ยที่มนุษย์ใช้ในการอ่านวิเคราะห์ รวมถึงการสร้างคลังข้อมูลทดสอบสำหรับนิยายภาษาไทยโดยเฉพาะ เพื่อนำมาใช้ประเมินประสิทธิภาพ และเปรียบเทียบกับเกณฑ์มาตรฐาน (Baseline) ต่อไป

กิตติกรรมประกาศ

โครงการวิจัยเรื่องระบบสำหรับการวิเคราะห์และเปรียบเทียบความคล้ายคลึงของนิยาย สามารถสำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับการสนับสนุนทางด้านความรู้ คำแนะนำ และการตรวจทานแก้ไขข้อบกพร่องต่าง ๆ จากบุคคลที่ผู้จัดทำจะกล่าวถึงดังต่อไปนี้ ขอขอบพระคุณพระคุณอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.สุพารณ์ ชุ่มเจริญ ที่ได้กรุณาให้คำปรึกษาให้ความรู้ ตลอดจนสละเวลาคอยช่วยเหลือและให้คำแนะนำในการศึกษาค้นคว้าตั้งแต่เริ่มต้นจนสำเร็จลุล่วงไปด้วยดี ขอกราบขอบพระคุณคณาจารย์ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ให้วิชาความรู้และสถานที่ในการจัดทำโครงการวิจัยขึ้นนี้

เอกสารอ้างอิง

- [1] บุชรินทร์ เจียเจริญ. (2565). การพิจารณาความคล้ายคลึงของนวนิยาย เพื่อวิเคราะห์การละเมิดลิขสิทธิ์จากการทำซ้ำหรือดัดแปลงงานวรรณกรรม [วิทยานิพนธ์นิติศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์].
- [2] ชมพูนุท เหลืองสมบุญรณ์. (2561). เปรียบเทียบการเล่าเรื่อง สัมพันธบท และบริบทของนวนิยาย ละครโทรทัศน์และภาพยนตร์

- กรณีศึกษา: ทวิภพ [วิทยานิพนธ์นิเทศศาสตรมหาบัณฑิต, มหาวิทยาลัยกรุงเทพ].
- [3] พรเทพ ตีสม. (2563). การวัดความคล้ายคลึงข้อความโดยใช้การวิเคราะห์แบบ Word Embedding [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง].
- [4] รัตติมา จีนาหงษา และ ชัยสิทธิ์ อังคะปัญญาเดช. (2568). การเปรียบเทียบประสิทธิภาพการตรวจสอบการคัดลอกระหว่างโปรแกรม iThenticate และ Turnitin. สำนักวิทยบริการมหาวิทยาลัยมหาสารคามและหน่วยงานห้องสมุดมหาวิทยาลัยส่วนภูมิภาค.
- [5] รัตติมา จีนาหงษา และ ชัยสิทธิ์ อังคะปัญญาเดช. (2567). การศึกษาประสิทธิภาพในการตรวจสอบการคัดลอกของโปรแกรม TurnItIn และ Copyleaks. ศูนย์บรรณสารและสื่อการศึกษา มหาวิทยาลัยแม่ฟ้าหลวงและหน่วยงานห้องสมุดมหาวิทยาลัยส่วนภูมิภาค.
- [6] จิรา สุวันทโรจน์ และ วีรยุทธ เลิศนที. (2568). การประยุกต์ใช้อัลกอริทึมระยะทางเลเวนชเตย์นเพื่อตรวจหาข้อความที่มองคล้ายกัน. มหาวิทยาลัยศิลปากร.