วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์ ปีที่ 3 ฉบับที่ 3 พฤษภาคม – มิถุนายน 2568
Journal of Engineering and Industrial Technology, Kalasin University, Vol. 3 No. 3 May – June 2025

บทความวิจัย (Research Article)

การพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา

ภวิกา อ้วนละมัย[1], ธรรมนูญ ปัญญาทิพย์[1,*], ปนัดดา โพธินาม[1], อัจฉรา สุมังเกษตร[1], ณรงค์ฤทธิ์ มะสุใส[1], ทรงกรด พิมพิศาล[1] และ ไพฑูรย์ ทิพย์สันเทียะ[1]

[1] สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยีสุขภาพ มหาวิทยาลัยกาฬสินธุ์

* ผู้ประสานงานบทความต้นฉบับ: tammanoon@ksu.ac.th โทรศัพท์: 089-9373774
(รับบทความ: 17 มีนาคม 2568; แก้ไขบทความ: 28 เมษายน 2568; ตอบรับบทความ: 6 พฤษภาคม 2568)

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาอัลกอริทึมและพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา 2) พัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ข้อมูลที่ใช้ในการทดลองเป็นฐานข้อมูลหนังสือจาก Web Opac งานวิทยบริการ มหาวิทยาลัยกาฬสินธุ์ ประกอบด้วยหนังสือจำนวน 727 เรื่อง เครื่องมือที่ใช้ในการพัฒนา ได้แก่ ภาษาไพทอน (Python) และโคแลป (Colab) วิธีการดำเนินการวิจัยได้แก่ การตัดคำด้วยวิธีการประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) หาค่าความถี่ด้วยวิธี TF-IDF และแปลงค่าข้อความเป็นค่าตัวเลขด้วยเทคนิค Label Encoding และหาค่าความใกล้เคียง 2 วิธีการ คือ วิธีการหาความใกล้เคียงแบบโคไซต์ (Cosine Similarity) และวิธีการหาความใกล้เคียงแบบแอดจัสโคไซต์ (Adjusted Cosine) วัดประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) และค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) วิธีการแนะนำหนังสือด้วยวิธีการ Top-N ผลการวิจัยพบว่า การแนะนำหนังสือจำนวน 5 เรื่อง 1) Cosine Similarity มีความถูกต้องร้อยละ 92.97 ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย 0.0263 2) Adjusted Cosine มีความถูกต้องร้อยละ 88.72 ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย 0.0153 และแนะนำหนังสือจำนวน 10 เรื่อง 1) Cosine Similarity มีความถูกต้องร้อยละ 92.92 ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย 0.2819 2) Adjusted Cosine มีความถูกต้องร้อยละ 92.50 ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย 0.4960 และ Cosine Similarity มีประสิทธิภาพดีกว่าในด้านค่าความถูกต้องในทั้งกรณีแนะนำหนังสือ (Top-N) 5 และ 10 รายการ ขณะที่ Adjusted Cosine มีค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยต่ำกว่าในการแนะนำหนังสือ (Top-N) 5 รายการ แต่สูงกว่าสำหรับ 10 รายการ ผลการวิจัยนี้เน้นย้ำถึงประสิทธิภาพของวิธีการวัดความใกล้เคียงในการพัฒนาระบบแนะนำแบบอิงเนื้อหาอย่างมีประสิทธิภาพ

คำสำคัญ: การแนะนำแบบอิงเนื้อหา การหาความใกล้เคียง ค่าความถี่คำแบบ TF-IDF

การอ้างอิงบทความ: ภวิกา อ้วนละมัย, ธรรมนูญ ปัญญาทิพย์, ปนัดดา โพธินาม, อัจฉรา สุมังเกษตร, ณรงค์ฤทธิ์ มะสุใส, ทรงกรด พิมพิศาล และ ไพฑูรย์ ทิพย์สันเทียะ, "การพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ," *วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์*, ปีที่ 3, ฉบับที่ 3, หน้า 36-53, 2568.

วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์ ปีที่ 3 ฉบับที่ 3 พฤษภาคม – มิถุนายน 2568
Journal of Engineering and Industrial Technology, Kalasin University, Vol. 3 No. 3 May – June 2025

บทความวิจัย (Research Article)

Book Recommendation System using Content-Based Filtering Technique

Phawika Ouanlamai[1], Tammanoon Panyatip[1,*], Panatda Phothinam[1], Achara Sumungkaset[1], Narongrit Masusai[1], Songgrod Phimphisan[1] and Paitoon Thipsanthia[1]

[1] Department of Computer Science and Information Technology, Faculty of Science and Health Technology, Kalasin University

* Corresponding Author: tammanoon@ksu.ac.th, Tel: 089-9373774

(Received: March 17, 2025; Revised: April 28, 2025; Accepted: May 6, 2025)

Abstract

This research aims to: 1) develop an algorithm and a content-based book recommendation system, and 2) enhance the content-based book recommendation system. The dataset for the experiment consists of 727 books from the Web OPAC of the Library Services at Kalasin University. The tools used in the development process include Python programming language and Google Colab. The research methodology involves several steps: word segmentation using Natural Language Processing (NLP) techniques to tokenize text, TF-IDF calculation to determine the significance of words, and text-to-numeric conversion using Label Encoding. Two similarity measurement methods, Cosine Similarity and Adjusted Cosine Similarity, were employed. System performance was evaluated using Accuracy and Mean Absolute Error (MAE), and recommendations were generated for the top 5 and top 10 books. The findings indicate that for the top 5 book recommendations, the Cosine Similarity method produced an accuracy of 92.97% with an MAE of 0.0263, while the Adjusted Cosine Similarity method produced an accuracy of 88.72% with an MAE of 0.0153. For the top 10 book recommendations, Cosine Similarity produced an accuracy of 92.92% with an MAE of 0.2819, whereas Adjusted Cosine Similarity produced an accuracy of 92.50% with an MAE of 0.4960. The results reveal that Cosine Similarity consistently outperforms Adjusted Cosine Similarity in terms of accuracy for both the top 5 and top 10 recommendations. However, Adjusted Cosine Similarity demonstrated a lower MAE for the top 5 recommendations but a higher MAE for the top 10 recommendations. These findings underscore the effectiveness of these similarity measurement methods in developing efficient content-based recommendation systems.

Keywords: Content-Based Recommendation, Similarity, TF-IDF

Please cite this article as: P. Ouanlamai, T. Panyatip, P. Phothinam, A. Sumungkaset, N. Masusai, S. Phimphisan and P. Thipsanthia, “Book Recommendation System using Content-Based Filtering Technique,” *The Journal of Engineering and Industrial Technology, Kalasin University*, vol. 3, no. 3, pp. 36-53, 2025.

บทความวิจัย (Research Article)

1. บทนำ

ในปัจจุบันข้อมูลมีบทบาทสำคัญในชีวิตประจำวัน ระบบแนะนำ (Recommendation System) ได้กลายเป็นเครื่องมือที่ช่วยในการค้นหาและเลือกใช้ข้อมูลอย่างมีประสิทธิภาพ [1] และพัฒนาจนกลายเป็นเครื่องมือพื้นฐานในการตัดสินใจและการเลือกที่มีข้อมูลมากขึ้นในชีวิตประจำวัน [2] ระบบแนะนำได้รับการพัฒนาขึ้นในหลาย ๆ ด้าน เช่น เพลง ภาพยนตร์ ข่าว และผลิตภัณฑ์โดยทั่วไป ในยุคปัจจุบัน องค์กรส่วนใหญ่ใช้ระบบแนะนำเพื่อตอบสนองความต้องการของลูกค้า ตัวอย่างเช่น LinkedIn, Amazon และ Netflix LinkedIn [3] โดยเฉพาะด้านการแนะนำหนังสือ ซึ่งช่วยให้ผู้อ่านสามารถค้นพบหนังสือที่ตรงกับความสนใจและความต้องการของตนเองได้อย่างรวดเร็ว ระบบแนะนำหนังสือแบบอิงเนื้อหา (Content-Based Recommendation System) เหมาะสมกับลักษณะข้อมูลเชิงลักษณะของรายการ เช่น ประเภท หมวดหมู่ คำสำคัญ ใช้ได้กับข้อมูลผู้ใช้ไม่มาก ตัวอย่างประเภทข้อมูลที่เหมาะสม เช่น บทความ ข่าว ภาพยนตร์ หนังสือ เมื่อเปรียบเทียบกับวิธีการคัดกรองแบบใช้ความร่วมมือ (Collaborative Filtering) เหมาะกับลักษณะข้อมูลมีข้อมูลพฤติกรรมของผู้ใช้จำนวนมาก เช่น การให้คะแนน การคลิก หรือการซื้อสินค้า เหมาะสำหรับกรณีที่ไม่มีข้อมูลเนื้อหาเชิงลึกของรายการ ตัวอย่างประเภทข้อมูลที่เหมาะสม เช่น อีคอมเมิร์ซ ฟังเพลง การกดไลน์ กดแชร์ของโซเชียลมีเดีย ดังนั้น ระบบแนะนำหนังสือแบบอิงเนื้อหา เหมาะกับการวิเคราะห์เอกสาร หนังสือ โดยพิจารณาข้อมูลในอดีตและพยายามเรียนรู้แบบจำลองการตั้งค่าโดยอิงจากการแสดงเนื้อหาของรายการที่แนะนำตามคุณลักษณะ [4] ซึ่งใช้ข้อมูลคุณลักษณะของหนังสือ เช่น ชื่อเรื่อง ผู้แต่ง คำสำคัญ และประเภท เพื่อสร้างคำแนะนำที่มีความเกี่ยวข้องเฉพาะตัวกับผู้อ่าน การพัฒนาระบบนี้จึงมีความสำคัญอย่างยิ่งในยุคดิจิทัลและปัญญาประดิษฐ์ ที่การค้นหาข้อมูลมักเต็มไปด้วยความซับซ้อนและมีปริมาณข้อมูลที่มีขนาดใหญ่

ห้องสมุดหรือสำนักวิทยบริการ เป็นสถานที่รวบรวมหนังสือเพื่อให้บริการแก่นักศึกษา บุคลากร และบุคคลทั่วไป รวมทั้งเป็นแหล่งเรียนรู้ศึกษาเพิ่มเติมในการเรียนการสอน และส่งเสริมการเรียนรู้ด้วยตนเองตลอดชีวิต เมื่อเปรียบเทียบกับห้องสมุดประเภทอื่นแล้ว ห้องสมุดสถาบันอุดมศึกษาเป็นห้องสมุดขนาดใหญ่ มีหนังสือหลากหลายประเภทและมีจำนวนมาก จึงทำให้ห้องสมุดต้องเผชิญกับข้อมูลจำนวนมหาศาล ทั้งข้อมูลการยืม-คืนหนังสือ ข้อมูลระเบียนหนังสือ และข้อมูลของผู้ใช้บริการห้องสมุด ทั้งนี้ข้อมูลหนังสือมีจำนวนที่เพิ่มขึ้นเรื่อย ๆ จึงทำให้ผู้ใช้บริการห้องสมุดเกิดปัญหาในการค้นหาหนังสือหรือไม่สามารถค้นหาหนังสือของห้องสมุดได้ตรงกับความต้องการ และไม่สามารถใช้ทรัพยากรสารสนเทศที่มีในห้องสมุดได้อย่างเต็มประสิทธิภาพ [5] ปัจจุบันห้องสมุดมีระบบสืบค้นหนังสือ เพื่อช่วยให้ผู้ใช้บริการมีความสะดวกในการค้นหาและเลือกหนังสือที่ตนเองสนใจ เช่น การค้นหาหนังสือจากความสอดคล้องของคำค้นหนึ่ง ๆ กับชื่อหนังสือ [6] แต่ในบางครั้งผู้ใช้บริการไม่สามารถที่จะสืบค้นข้อมูลหนังสือที่ตรงกับความสนใจ และเสียเวลาในการสืบค้นข้อมูลหนังสือ ทุกวันนี้ได้มีระบบที่เข้ามาช่วยอำนวยความสะดวกสำหรับผู้ใช้บริการไม่ต้องเสียเวลาในการสืบค้น

บทความวิจัย (Research Article)

ข้อมูล และได้ข้อมูลตรงตามต้องการ นั่นคือระบบแนะนำข้อมูล [7] เป็นการประยุกต์ใช้จากประสบการณ์ของผู้ใช้ (User) เช่น ประสบการณ์การซื้อสินค้า การใช้บริการ การยืม-คืนหนังสือ เป็นต้น จากข้อมูลของผู้ใช้เหล่านั้นไปหาค่าความใกล้เคียง โดยใช้เทคนิคการเรียนรู้ของเครื่อง ประกอบการตัดสินใจให้กับผู้ใช้งาน เพื่อเพิ่มโอกาสให้กับผู้ใช้งานได้หนังสือที่ตรงตามความต้องการ โดยผู้ใช้ไม่ต้องเสียเวลาในการสืบค้นข้อมูล

ดังนั้น ผู้วิจัยจึงมีความสนใจในการพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา โดยใช้ฐานข้อมูลจาก Web Opac งานวิทยบริการ มหาวิทยาลัยกาฬสินธุ์ เป็นกรณีศึกษา เพื่ออำนวยความสะดวกแก่ผู้ใช้บริการ และส่งเสริมการยืมหนังสือภายในห้องสมุด

2. วัตถุประสงค์การวิจัย

1) เพื่อพัฒนาอัลกอริทึมระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา

2) เพื่อพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา

3. เอกสารและงานวิจัยที่เกี่ยวข้อง

ระบบแนะนำ (Recommender System : RS) เป็นเครื่องมือที่ช่วยให้ผู้ใช้ค้นหาเนื้อหาและเอาชนะข้อมูลขนาดใหญ่ โดยจะให้คำแนะนำตามรูปแบบรสนิยมของผู้ใช้และคาดการณ์ความสนใจของผู้ใช้ ปัจจุบัน ระบบแนะนำได้สร้างพื้นที่ให้กับตัวเองด้วยการมีอิทธิพลต่อการตัดสินใจของแต่ละบุคคลอย่างมาก เว็บไซต์สื่อสารโซเชียล เช่น Facebook, Google, Twitter, LinkedIn และ Netflix ล้วนประสบความสำเร็จอย่างมากจากระบบแนะนำ [8] ระบบแนะนำแบ่งประเภทได้ 3 ประเภท ได้แก่ ระบบการกรองแบบร่วมมือกัน (Collaborative Filtering) ระบบการกรองตามเนื้อหา (Content-Based Filtering) และระบบไฮบริด (Hybrid Systems) [3] ระบบเหล่านี้สร้างแบบจำลองจากพฤติกรรมในอดีตของผู้ใช้ระบบแนะนำคือการกรองตามเนื้อหา การกรองตามเนื้อหาจะวิเคราะห์พฤติกรรมที่ผ่านมาของผู้ใช้และแนะนำรายการที่คล้ายกันตามพารามิเตอร์ที่กำหนด [4] รายละเอียดดังนี้

1) ระบบการกรองแบบร่วมมือกัน (Collaborative Filtering : CF) เป็นวิธีที่อาศัยพฤติกรรมของผู้ใช้รายอื่นที่มีความสนใจคล้ายกันเพื่อแนะนำข้อมูล CF ใช้ข้อมูล User-Item เป็นอินพุตเพื่อสร้างคำแนะนำ โดยจะทำนายไอเท็มตามคะแนนที่ผ่านมาของผู้ใช้และคะแนนที่ผู้ใช้ที่มีโปรไฟล์เดียวกัน [9] ข้อดีของวิธีนี้คือ สามารถแนะนำรายการที่ผู้ใช้ไม่เคยมีประวัติการเข้าถึงมาก่อนได้ (Serendipity) และสามารถใช้ได้กับข้อมูลประเภทต่าง ๆ โดยไม่ต้องพึ่งพาคุณลักษณะเฉพาะของรายการ อย่างไรก็ตาม ข้อเสียคืออาจเกิดปัญหา Cold Start สำหรับผู้ใช้หรือรายการใหม่ ๆ ที่ไม่มีข้อมูลพฤติกรรมมาก่อน และอาจได้รับผลกระทบจากปัญหา Sparsity หากมีข้อมูลไม่เพียงพอ

2) ระบบการกรองตามเนื้อหา (Content-Based Filtering : CBF) เป็นระบบที่วิเคราะห์พฤติกรรมที่ผ่านมาของผู้ใช้และแนะนำรายการที่คล้ายกันตามพารามิเตอร์ที่กำหนด [4] ข้อดีของวิธีนี้คือสามารถ

บทความวิจัย (Research Article)

แนะนำข้อมูลโดยไม่ต้องพึ่งพาผู้ใช้อื่น ลดปัญหา Cold Start สำหรับผู้ใช้ที่มีข้อมูลประวัติการใช้งานอยู่แล้ว และสามารถอธิบายได้ว่าทำไมระบบจึงแนะนำรายการใดรายการหนึ่ง อย่างไรก็ตาม ระบบนี้มีข้อจำกัด เช่น ปัญหา Over-Specialization [10] ซึ่งอาจทำให้แนะนำเฉพาะรายการที่คล้ายกับสิ่งที่ผู้ใช้เคยมีประวัติการค้นหาเท่านั้น ส่งผลให้ขาดความหลากหลาย นอกจากนี้ยังต้องอาศัยการออกแบบคุณลักษณะของข้อมูล (Feature Engineering) ที่เหมาะสมเพื่อให้การแนะนำมีประสิทธิภาพ

3) ระบบไฮบริด (Hybrid Systems) เป็นการผสมผสานระหว่าง Collaborative Filtering และ Content-Based Filtering เพื่อใช้ประโยชน์จากข้อดีของทั้งสองระบบ โดยสามารถลดปัญหา Cold Start, Sparsity และ Overspecialization ได้ทำให้การแนะนำมีความแม่นยำและหลากหลายมากขึ้น อย่างไรก็ตาม ข้อเสียคือมีความซับซ้อนสูง ต้องใช้ทรัพยากรในการประมวลผลมากขึ้น และต้องมีการออกแบบโมเดลที่เหมาะสมเพื่อให้สามารถรวมข้อดีของทั้งสองวิธีได้อย่างมีประสิทธิภาพ [11]

ระบบการกรองตามเนื้อหา (Content-Based Filtering: CPF) เป็นวิธีการแนะนำที่คล้ายกับการกรองตามความรู้ความเข้าใจ ซึ่งเน้นการคาดการณ์และแนะนำรายการที่มีลักษณะคล้ายคลึงกับความสนใจหรือพฤติกรรมของผู้ใช้ [12] โดยการกรองตามเนื้อหาจะพิจารณาความคล้ายคลึงกันระหว่างบทความหรือรายการที่ผู้ใช้สนใจ [13,14] ใช้เทคนิคหาค่าความถี่ด้วย TF-IDF Vectorizer และคำนวณความคล้ายคลึงกัน [15] เพื่อเพิ่มประสิทธิภาพในการกรองตามเนื้อหา นักวิจัยหลายคนจึงคำสำคัญมาใช้ในการคำนวณความคล้ายคลึงกัน โดยหนึ่งในวิธีที่ได้รับความนิยมคือการใช้เมตริกความคล้ายคลึงของโคไซน์ (Cosine Similarity) และแอดจัสโคไซต์ (Adjusted Cosine) [16] ทั้งนี้มีนักวิจัยที่ให้ความสนใจการพัฒนาระบบแนะนำหนังสือ ได้แก่ การศึกษาของ Kommineni และคณะ [17] ได้ทำการพัฒนาระบบแนะนำหนังสือที่ใช้การเรียนรู้ของเครื่องและอัลกอริทึมการกรองแบบร่วมมือกันตามผู้ใช้ โดยเปรียบเทียบประสิทธิภาพของวิธีต่าง ๆ เช่น Pearson Coefficient Correlation (PCC), Constrained Pearson Coefficient (CPCC), Cosine Similarity และ Jaccard Similarity พบว่า CPCC เป็นวิธีที่มีประสิทธิภาพดีที่สุดในการแนะนำข้อมูล Top-10 ในงานศึกษาของ Sarma และคณะ [18] ได้พัฒนาระบบแนะนำหนังสือส่วนบุคคลด้วยการเรียนรู้ของเครื่องโดยใช้การจัดกลุ่ม K-means และการวัดความคล้ายคลึงของโคไซน์ ผลลัพธ์จากการทดลองแสดงให้เห็นว่าระบบมีความไวและความจำเพาะเฉลี่ยที่ 49.76% และ 56.74% ตามลำดับ และมีคะแนน F1 ที่ 52.84% จากการศึกษาเหล่านี้ สามารถสรุปได้ว่าการกรองตามเนื้อหาเป็นวิธีที่มีประสิทธิภาพในการแนะนำรายการที่เกี่ยวข้องกับความชอบของผู้ใช้ โดยเฉพาะในกรณีที่สามารถใช้ข้อมูลเชิงลึก เช่น คำอธิบาย หรือคุณสมบัติของรายการต่าง ๆ เพื่อเพิ่มความแม่นยำในการแนะนำ

ดังนั้น จากทฤษฎีและงานวิจัยที่เกี่ยวข้อง จะเห็นได้ว่าการเลือกใช้ระบบแนะนำหนังสือโดยใช้การกรองตามเนื้อหา เป็นแนวทางที่เหมาะสมสำหรับระบบแนะนำหนังสือ เนื่องจากสามารถวิเคราะห์คุณลักษณะของหนังสือ เช่น ประเภท ผู้แต่ง หรือคำ

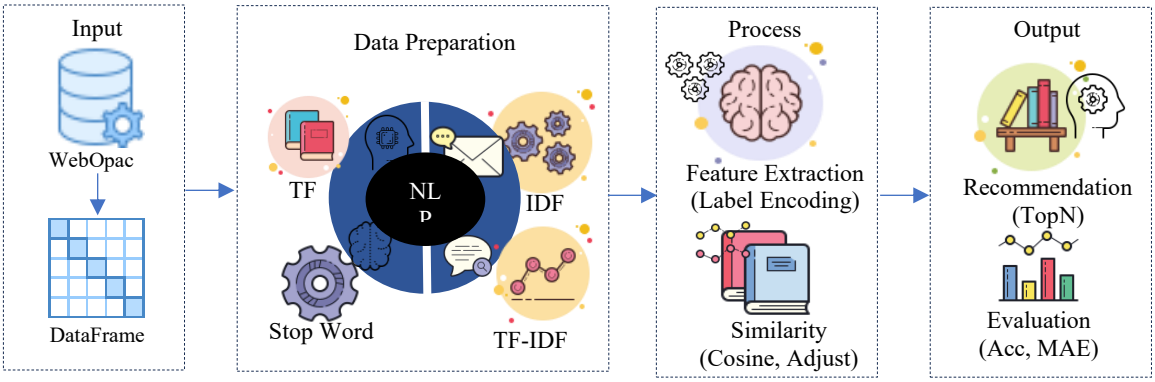
บทความวิจัย (Research Article)

สำคัญ แล้วแนะนำหนังสือที่คล้ายกับที่ผู้ใช้เคยอ่าน โดยไม่ต้องพึ่งพาข้อมูลจากผู้ใช้อื่น ทำให้สามารถสร้างคำแนะนำแบบเฉพาะบุคคลได้อย่างมีประสิทธิภาพ CBF ยังช่วยลดปัญหาผู้ใช้ใหม่ (Cold Start Problem) เพราะใช้ข้อมูลจากเนื้อหาหนังสือโดยตรง ไม่จำเป็นต้องมีประวัติการใช้งานของผู้ใช้ จึงเหมาะกับการแนะนำหนังสือเฉพาะกลุ่มที่มีข้อมูลพฤติกรรมจำกัด อีกจุดเด่นของ CBF คือ ความสามารถในการอธิบายคำแนะนำ เช่น "หนังสือเล่มนี้ถูกแนะนำเพราะมีเนื้อหาคล้ายกับที่คุณเคยอ่าน" ซึ่งช่วยเพิ่มความน่าเชื่อถือของระบบ ดังนั้น CBF จึงเป็นตัวเลือกที่เหมาะสม สำหรับระบบแนะนำหนังสือ โดยเฉพาะเมื่อต้องการคำแนะนำที่แม่นยำ เข้าใจง่าย และไม่ต้องอาศัยพฤติกรรมของผู้ใช้อื่น

4. วิธีดำเนินการวิจัย

ระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหาได้กลายเป็นเครื่องมือที่ช่วยส่งเสริมการเรียนรู้และการค้นคว้าอย่าง

มีประสิทธิภาพ งานวิจัยนี้มุ่งพัฒนาระบบแนะนำหนังสือโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ หาค่าความถี่และการวิเคราะห์ค่าความคล้ายคลึง เพื่อช่วยให้ผู้ใช้ค้นพบหนังสือที่ตรงกับความสนใจได้ง่ายและรวดเร็ว พร้อมประเมินประสิทธิภาพผ่านการทดลองที่ครอบคลุมและชัดเจน ซึ่งมีกระบวนการดำเนินการวิจัย ดังนี้



รูปที่ 1 กรอบระเบียบวิธีวิจัยระบบแนะนำหนังสือ

4.1 การเลือกข้อมูล (Data Collection)

ระบบแนะนำหนังสือเริ่มต้นด้วยการรวบรวมข้อมูลหนังสือทั้งหมด 727 เรื่องจากฐานข้อมูล Web Opac มหาวิทยาลัยกาฬสินธุ์ ซึ่งแบ่งออกเป็น 3 ประเภท ได้แก่ ประเภทคอมพิวเตอร์ จำนวน 389 เรื่อง กฎหมาย จำนวน 270 เรื่อง และบัญชี จำนวน 68 เรื่อง ข้อมูลแต่ละเล่มประกอบด้วย Features เช่น ชื่อเรื่อง (Title) ชื่อผู้แต่ง (Author) เมือง (City) ปีที่พิมพ์ (Year) โรงพิมพ์ (Published) คำสำคัญ (Keywords) และประเภทหนังสือ (Type) มีลักษณะข้อมูลดังรูปที่ 2

บทความวิจัย (Research Article)

code	title	author	city	pulished	year	Keyword	Type
3781	ผ่า ! ...คอมพิวเตอร์	อนิรุทธิ์ รัชตะวราห์	กรุงเทพมหานคร	โปรวิชั่น	2543	คอมพิวเตอร์	คอมพิวเตอร์
4741	สงครามคอมพิวเตอร์	Wilson,Cronville	กรุงเทพมหานคร	กอไผ่	2526	คอมพิวเตอร์	คอมพิวเตอร์
4767	โปรแกรมคอมพิวเตอร์	ยืน ภู่สุวรรณ	กรุงเทพมหานคร	ซีเอ็ดยูเคชั่น	2535	คอมพิวเตอร์	คอมพิวเตอร์
4898	1001 ปัญหาคอมพิวเตอร์	Komando , Kim	กรุงเทพมหานคร	ซีเอ็ดยูเคชั่น	2540	ปัญหา คอมพิวเตอร์	คอมพิวเตอร์
5829	คอมพิวเตอร์เบื้องต้น	นฤชิต แววศรีผ่อง	กรุงเทพมหานคร	ซีเอ็นยูเคชั่น	2548	คอมพิวเตอร์	คอมพิวเตอร์
5987	เครือข่ายคอมพิวเตอร์	จารุวรรณ ทุลธรรม	ขอนแก่น	วิทยาเขตขอนแก่น	2548	เครือข่ายคอมพิวเตอร์	คอมพิวเตอร์
6537	คอมพิวเตอร์กับงานบัญชี	ฉวีวรรณ โสภาจารีย์	กรุงเทพมหานคร	พระจอมเกล้าพระนครเ	2547	คอมพิวเตอร์ บัญชี	คอมพิวเตอร์
6569	คอมพิวเตอร์ทำงานอย่างไร	วิโรจน์ อัศวรังสี	กรุงเทพมหานคร	ซีเอ็ดยูเคชั่น	2538	คอมพิวเตอร์ ทำงาน	คอมพิวเตอร์
6573	คู่มือช่างคอมพิวเตอร์ 2006	อนิรุจน์ รัชตะวราห์	กรุงเทพมหานคร	โปรวิชั่น	2548	คู่มือ คอมพิวเตอร์	คอมพิวเตอร์
6584	ถอดรหัสไวรัสคอมพิวเตอร์	ดลเดช ศุภวิจิตรกุล	กรุงเทพมหานคร	ซีเอ็ดยูเคชั่น	2545	ไวรัสคอมพิวเตอร์	คอมพิวเตอร์

รูปที่ 2 ลักษณะของข้อมูล

สามารถแสดงกลุ่มคำ (Word Cloud) ที่สร้างจากชื่อเรื่องหนังสือ (Title) ของชุดข้อมูล ดังรูปที่ 3 (1) และแสดงกลุ่มคำที่สร้างจากคำสำคัญ (Keywords) ดังรูปที่ 3 (2) ซึ่งชุดข้อมูลเหล่านี้ได้รับการทำความสะอาดก่อนที่จะถูกแปลงเป็นเวกเตอร์

(1) (2)

รูปที่ 3 แสดงกลุ่มคำของชุดข้อมูล

4.2 การเตรียมข้อมูล (Data Preparation)

ในกระบวนการเตรียมข้อมูล เป็นขั้นตอนสำคัญในการพัฒนาระบบแนะนำหนังสือแบบอิงเนื้อหา โดยข้อมูลดิบจากฐานข้อมูลจะถูกจัดเตรียมให้เหมาะสมต่อการประมวลผลด้วยเทคนิคต่าง ๆ เช่น การทำความสะอาดข้อมูล การแยกข้อความ และการหาค่าความถี่ของคำ เพื่อให้ระบบสามารถวิเคราะห์และสร้างคำแนะนำได้อย่างแม่นยำ มีขั้นตอนดังนี้

4.2.1 การทำความสะอาดข้อมูล (Data Cleaning)

กระบวนการตรวจสอบ แก้ไข และลบข้อมูลที่ไม่ถูกต้อง หรือไม่สมบูรณ์ในชุดข้อมูล เพื่อปรับปรุงคุณภาพของข้อมูลให้เหมาะสมสำหรับการวิเคราะห์และการสร้างโมเดล จากข้อมูลที่ดึงมาจาก Web Opac เป็นลักษณะ Text ไฟล์ ประกอบด้วยข้อมูล Bib, Leader, ISBN, Dewey Call, Title, Edition, Published/Created, Note, Bibliography,

บทความวิจัย (Research Article)

Sebject, Added Author, Keyword ผู้วิจัยได้ทำแปลงข้อมูลเป็นลักษณะตาราง ไฟล์ csv และ Features ดังนี้ Title, Author, City, Year, Published, Keywords, Type

4.2.2 การรวมข้อมูลและแยกข้อความ (Word Segmentation)

ในขั้นตอนนี้ ข้อมูลข้อความดิบจะถูกแปลงเป็นหน่วยย่อยที่เรียกว่า "คำ" โดยใช้กระบวนการ Word Segmentation ซึ่งเป็นการแยกคำในภาษาที่ไม่มีตัวเว้นวรรค เช่น ภาษาไทย ออกจากกันอย่างถูกต้อง เพื่อให้สามารถนำไปประมวลผลต่อในงานด้าน การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) [19] สำหรับการตัดคำในภาษาไทย PyThaiNLP เป็นหนึ่งในเครื่องมือที่ได้รับความนิยม ซึ่งใช้วิธีการตัดคำ Newmm ที่อาศัยกฎไวยากรณ์และข้อมูลสถิติช่วยให้การตัดคำมีความแม่นยำและยืดหยุ่นมากขึ้น ตัวอย่างเช่น คอลัมน์ชื่อหนังสือ (Title) เช่น ชื่อเรื่องหนังสือ “การเขียนโปรแกรม Python สำหรับนักศึกษา” เมื่อทำการตัดคำจะได้ กลุ่มคำ คือ [“การเขียนโปรแกรม”, “Python”, “สำหรับ”, “นักศึกษา”]

4.2.3 การลบคำหยุด (Stop-word Removal)

คือกระบวนการในการกำจัด คำที่มีความหมายทั่วไป หรือ คำที่ไม่มีคุณค่าทางข้อมูล ในการประมวลผลภาษาธรรมชาติ (NLP) เช่น คำที่ใช้บ่อยในภาษา แต่ไม่ช่วยในการทำความเข้าใจเนื้อหาหรือบริบทของข้อความมากนัก เพื่อลดคำที่ไม่มีความสำคัญ เช่น "และ", "หรือ", "ของ" ซึ่งถูกกำจัดออกจากชุดข้อมูล เช่น จากกลุ่มคำ [“การเขียนโปรแกรม”, “Python”, “สำหรับ”, “นักศึกษา”] จะลบคำว่า “สำหรับ” จะได้กลุ่มคำใหม่คือ [“การเขียนโปรแกรม”, “Python”, “นักศึกษา”]

4.2.4 การคำนวณค่าความถี่คำ

Term Frequency (TF) [8] ใช้วัดว่าคำใดปรากฏบ่อยในข้อความหนังสือแต่ละเล่ม โดยมีสูตรสมการคือ

$$\mathrm{TF(t,d)} = \frac{\text{จำนวนครั้งที่คำ } \boldsymbol{t} \text{ ปรากฏในเอกสาร } \boldsymbol{d}}{\text{จำนวนคำทั้งหมดในเอกสาร } \boldsymbol{d}} \quad (1)$$

เมื่อ t คือคำที่สนใจ และ d คือข้อความของหนังสือเล่มหนึ่ง

4.2.5 การคำนวณค่าความสำคัญของคำ

Inverse Document Frequency (IDF) ใช้ระบุว่าคำนั้นเป็นเอกลักษณ์เพียงใดในชุดข้อมูลทั้งหมด โดยพิจารณาจำนวนเอกสารที่คำปรากฏอยู่ โดยมีสูตรสมการคือ [8]

$$IDF(t) = log\frac{N}{1+DF(t)} \quad (2)$$

เมื่อ N คือจำนวนหนังสือทั้งหมดในฐานข้อมูล และ DF(t) คือ จำนวนหนังสือที่มีคำ t ปรากฏ

4.2.6 การรวมค่าความถี่คำและค่าความสำคัญ (TF-IDF)

ประกอบด้วย TF และ IDF ซึ่งก็คือความถี่ของคำและความถี่ของเอกสารผกผันตามลำดับ TF แสดงถึงความถี่ที่คำปรากฏในเอกสาร แนวคิดหลักของ IDF คือ หากคำปรากฏในเอกสารอื่นหลายครั้ง คำนั้นจะมี

บทความวิจัย (Research Article)

ความสำคัญน้อยลงค่าคะแนน TF-IDFคำนวณจากผลคูณของ TF และ IDF มีสูตรสมการดังนี้ [8]

$$\mathrm{TF-IDF(t,d)} = TF(t,d) \times IDF(t) \qquad (3)$$

สำหรับกระบวนการเตรียมข้อมูลช่วยทำให้ชุดข้อมูลของหนังสือทั้ง 727 เรื่องถูกจัดการให้อยู่ในรูปแบบที่เหมาะสมสำหรับการคำนวณในขั้นต่อไป โดยเน้นการประมวลผลข้อความและการแปลงข้อมูลเป็นค่าตัวเลขเพื่อลดความซับซ้อนและเพิ่มความแม่นยำในระบบแนะนำหนังสือ

4.3 การแปลงข้อมูลเป็นเชิงตัวเลข (Feature Extraction)

ในกระบวนการนี้ ข้อมูลจากหนังสือถูกแปลงให้อยู่ในรูปแบบเชิงตัวเลขที่เหมาะสำหรับการคำนวณ โดยใช้เทคนิคที่เรียกว่า TF-IDF (Term Frequency-Inverse Document Frequency) [8] จาก Features Title (0) นอกจากนี้ ยังใช้ Label Encoding เพื่อแปลงข้อความเหล่านี้ให้เป็นตัวเลขจาก Features ดังนี้ Author, City, Year, Published, Keywords, Type ดังรูปที่ 4

	code	author	city	pulished	year	Keyword	Type	0
0	3781	529.0	0.0	230.0	17.0	270.0	0	[[0.7071067811865475, 0.7071067811865475]]
1	4741	17.0	0.0	32.0	2.0	270.0	0	[[0.5773502691896258, 0.5773502691896258, 0.57...
2	4767	327.0	0.0	50.0	8.0	270.0	0	[[0.7071067811865475, 0.7071067811865475]]
3	4898	16.0	0.0	50.0	13.0	414.0	0	[[0.5, 0.5, 0.5, 0.5]]
4	5829	190.0	0.0	51.0	22.0	270.0	0	[[0.5, 0.5, 0.5, 0.5]]

รูปที่ 4 ลักษณะการแปลงข้อมูล

4.4 การคำนวณค่าความใกล้เคียง (Similarity Calculation)

กระบวนการนี้มีเป้าหมายเพื่อวัดความคล้ายคลึงกันระหว่างหนังสือสองเล่ม โดยใช้เวกเตอร์ที่สร้างจากค่าคะแนน TF-IDF ของคำและ Label Encoding มาทำการคำนวณค่าความใกล้เคียงทำได้ด้วยวิธี Cosine Similarity [17, 20] คือผลคูณจุดของจุดข้อมูลสองจุดวัดมุมโคไซน์ระหว่างวัตถุ หากมุมระหว่างวัตถุทั้งสองเท่ากับ 0 องศา ความคล้ายคลึงจะเป็น 1 หากมุมเท่ากับ 90 องศา คะแนนความคล้ายคลึงจะเป็น 0 นับเป็นเมทริกซ์

ที่นิยมใช้กันมากในการหาความคล้ายคลึงเนื่องจากประเมินได้ง่าย โดยเฉพาะอย่างยิ่งสำหรับเวกเตอร์แบบเบาบาง และวิธี Adjusted Cosine Similarity [1, 20] ซึ่งสามารถบ่งบอกถึงระดับความเหมือนหรือความแตกต่างของหนังสือในเชิงเนื้อหาโดยรายละเอียดดังนี้

4.4.1 Cosine Similarity มีสูตรสมการ [17] ดังนี้

$$S_{cosine}(A,B) = \frac{A.B}{\|A\| \, \|B\|} \qquad (4)$$

เมื่อ *A* และ *B* คือเวกเตอร์ TF-IDF ของหนังสือสองเล่ม, *A.B* คือผลคูณของเวกเตอร์ *A* และ *B*, $\|A\|$ คือ ความยาวหรือขนาดของเวกเตอร์ *A*, $\|B\|$ คือ ความยาวหรือขนาดของเวกเตอร์ *B*, S_{cosine} คืออยู่ในช่วง [0,1] โดย 1 หมายถึงเหมือนกันมากที่สุด

4.4.2 Adjusted Cosine Similarity [20] มีสูตรสมการ ดังนี้

$$S_{adjusted}(A,B) = \frac{\sum_{i=1}^{n}(A_i-\bar{A})(B_i-\bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i-\bar{A})^2}\sqrt{\sum_{i=1}^{n}(B_i-\bar{B})^2}} \qquad (5)$$

เมื่อ A_i, B_i คือค่าของคำที่ตำแหน่ง *i* ในเวกเตอร์ *A* และ *B*, $\bar{A}, \bar{B}$ คือ ค่าเฉลี่ยของเวกเตอร์ *A* และ *B*

การคำนวณความใกล้เคียงด้วยวิธีการ Cosine Similarity และ Adjusted Cosine Similarity ช่วยประเมินความคล้ายคลึงของหนังสือในเชิงเนื้อหา โดย Cosine Similarity ใช้สำหรับการวัดทั่วไป และ Adjusted Cosine Similarity ใช้ในกรณีที่ต้องปรับค่าตามระดับความนิยมของคำในหนังสือ

4.5 การแนะนำหนังสือ (Book Recommendation)

การทดลองวิจัยได้มีการแบ่งชุดข้อมูลเป็นชุดฝึก (Training Set) และชุดทดสอบ (Test Set) คือ 95:5, 90:10 และ 80:20 สำหรับการทดลองโดยใช้ข้อมูลทดสอบในสัดส่วนต่าง ๆ เช่น 5%, 10%, และ 20% การทดลองนี้เพื่อศึกษาผลกระทบของขนาดข้อมูลทดสอบที่แตกต่างกันในการประเมินประสิทธิภาพของโมเดล โดยช่วยให้เราสามารถเข้าใจได้ว่า ขนาดข้อมูลทดสอบที่มากขึ้นหรือความหลากหลายของข้อมูลทดสอบจะส่งผลต่อความแม่นยำของโมเดลในลักษณะใด และช่วยให้เราสามารถปรับปรุงโมเดลได้อย่างมีประสิทธิภาพมากขึ้น

กระบวนการแนะนำหนังสือใช้ค่าความใกล้เคียง (Similarity) ที่คำนวณได้ในขั้นตอนก่อนหน้า เพื่อสร้างรายการแนะนำหนังสือ โดยวิธีการ Top-N Recommendation จะคัดเลือกหนังสือที่มีค่าความใกล้เคียงกับหนังสือที่ผู้ใช้สนใจมากที่สุด N รายการ กระบวนการนี้ช่วยให้ผู้ใช้ได้รับคำแนะนำที่ตรงกับความต้องการหรือความสนใจ ซึ่งผู้วิจัยได้เลือกการแนะนำหนังสือจำนวน N = 5 เรื่อง [21] และ N = 10 เรื่อง [1, 17] สูตรสมการสำหรับการจัดอันดับ (Ranking) จากการวัดค่าความใกล้เคียง $S_{similarity}$ *(A,B)* ของหนังสือเป้าหมาย *A* และหนังสืออื่น ๆ *B1, B2, ..., Bm* จะถูกจัดเรียงในลำดับจากมากไปน้อย เพื่อเลือกหนังสือ N เล่มที่คล้ายกับหนังสือเป้าหมายมากที่สุด สำหรับกระบวนการแนะนำหนังสือใช้ค่าความใกล้เคียงเพื่อจัดอันดับและเลือกหนังสือ *N* รายการที่มีเนื้อหาคล้ายกับหนังสือเป้าหมายมากที่สุด วิธีการนี้ช่วยสร้างคำแนะนำที่สอดคล้องกับความสนใจของผู้ใช้อย่างมีประสิทธิภาพ

บทความวิจัย (Research Article)

5. การประเมินผล (Evaluation)

การวัดประเมินผลระบบแนะนำหนังสือแบบอิงเนื้อหา มีวิธีการวัดด้วยวิธีวัดค่าความถูกต้อง (Accuracy) [22] และวัดค่าความคลาดเคลื่อน (Mean Absolute Error (MAE) [1] ดังนี้

5.1 ค่าความถูกต้อง

$$\text{Accuracy} = \frac{\text{จำนวนคำแนะนำที่ถูกต้อง}}{\text{จำนวนคำแนะนำทั้งหมด}} \quad (7)$$

เมื่อจำนวนคำแนะนำที่ถูกต้อง คือจำนวนคำแนะนำที่ตรงกับประเภทที่เราต้องการ หรือประเภทที่ถูกต้อง และจำนวนคำแนะนำที่เลือก คือจำนวนคำแนะนำทั้งหมดที่เราตัดสินใจเลือกมา

ยกตัวอย่าง สมมุติว่าเราได้ทำการแนะนำหนังสือ 5 รายการ (Top 5) เพื่อคำนวณค่าความถูกต้อง (Accuracy) ซึ่งในที่นี้การแนะนำที่ถูกต้องจะต้องเป็นหนังสือประเภทคอมพิวเตอร์ที่เราคาดหวัง โดยจาก 5 รายการแรกที่เลือกมา พบว่า ทุกคำแนะนำใน 5 รายการนี้มีประเภทที่ตรงกับประเภทคอมพิวเตอร์จำนวน 4 รายการ ดังนั้น จำนวนคำแนะนำที่ถูกต้องเมื่อใช้สูตรคำนวณค่าความถูกต้อง จะได้ Accuracy=4/5=0.8 หรือ 80%

5.2 ค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์

$$\text{MAE} = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \quad (8)$$

เมื่อ y_i คือค่าความจริง (Actual Value) ของตัวอย่างที่ *i*, $\hat{y}_i$ คือค่าที่คาดการณ์ (Predicted Value) ของตัวอย่างที่ *i*, *n* คือจำนวนตัวอย่างทั้งหมดในชุดข้อมูล และ $|\hat{y}_i - y_i|$ คือค่าสัมบูรณ์ของความแตกต่างระหว่างค่าความจริงกับค่าที่คาดการณ์

6. ผลการวิจัย

6.1 ผลการพัฒนาอัลกอริทึมระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา

ผู้วิจัยได้ดำเนินการพัฒนากระบวนการระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ซึ่งมีกระบวนการพัฒนาตามวิธีการดำเนินการวิจัยแต่ขั้นตอน สามารถสรุปเป็นอัลกอริทึมระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ดังนี้

ผู้วิจัยได้ดำเนินการพัฒนากระบวนการระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ซึ่งมีกระบวนการพัฒนาตามวิธีการดำเนินการวิจัยแต่ละขั้นตอน สามารถสรุปเป็นอัลกอริทึมระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ดังนี้

1. การเลือกข้อมูลและการเตรียมข้อมูล

1.1 อินพุต ชุดข้อมูล D ซึ่งประกอบด้วยหนังสือ D = {B1, B2, ..., Bn} พร้อมฟีเจอร์ F = {ชื่อเรื่อง, ผู้แต่ง, เมือง, ปีที่พิมพ์, โรงพิมพ์, คำสำคัญ, ประเภทหนังสือ}

1.2 แยกข้อความ $Tokens(d_i) \leftarrow Tokenize\big(Features(d_i)\big), \forall d_i \in D$

1.3 ทำความสะอาดข้อมูล ลบคำฟุ่มเฟือยหรือคำที่ไม่มีความสำคัญ ด้วย Stop words

2. การแปลงข้อมูลเป็นเชิงตัวเลข

2.1 คำนวณ TF-IDF สำหรับหนังสือทุกเล่ม

$$TF-IDF(t,d_i) \leftarrow \frac{Freq(t,d_i)}{Total\ Term\ in\ d_i} \times log\frac{Total\ Documents}{Documents\ Containing\ t}, \forall t \in tokens(d_{i)})$$

2.2 แปลงค่าข้อมูลแต่ละ Features

$$Encoded\ Label(d_i) \leftarrow LabelEncoder(Category(d_i))$$

3. คำนวณค่าความใกล้เคียง

3.1 สร้างเวกเตอร์ TF-IDF สำหรับหนังสือเป้าหมาย q

$$Vector(q) \leftarrow TF-IDF(q)$$

3.2 คำนวณ Cosine Similarity ระหว่าง q และทุก d_i

$$S_{cosine}(q,d_i) = \frac{Vector(q)\cdot Vector(d_i)}{||Vector(q)||\,||Vector(d_i)||}$$

3.3 คำนวณ Adjusted Cosine Similarity

$$S_{adjusted}(q,d_i) = \frac{\sum_k (q_{k-\bar{q}})(d_{i,k}-\bar{d}_i)}{\sqrt{\sum_k (q_{k-\bar{q}})^2}\sqrt{(d_{i,k}-\bar{d}_i)^2}}$$

4. การเลือกแนะนำหนังสือ

4.1 เรียงลำดับค่าความใกล้เคียง

$$R \leftarrow Sort(\{S(q,d_1,S(q,d_2),\dots,S(q,d_n)\},Descending)$$

4.2 เลือก Top-N

$$R = \{r_1,r_2,\dots,r_N\}, where\ r_i\ is\ Top-N\ By\ Similarity$$

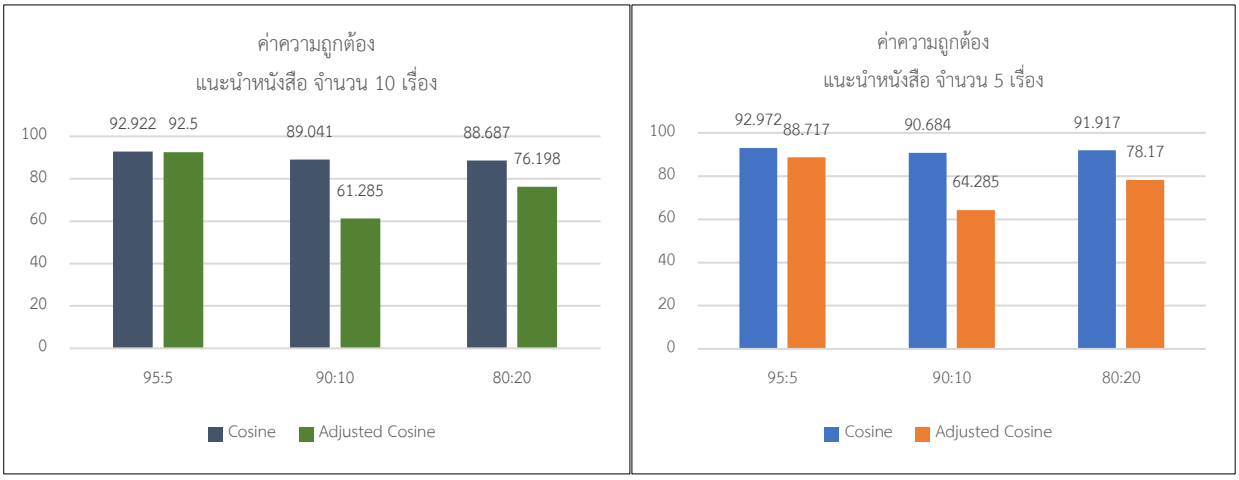
6.2 ผลการพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา

ผู้วิจัยได้ดำเนินการพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหาช่วยเพิ่มประสิทธิภาพในการเลือกหนังสือที่ตรงกับความสนใจของผู้ใช้ โดยใช้ข้อมูลคุณลักษณะของหนังสือเป็นพื้นฐาน ระบบสามารถสร้างคำแนะนำที่มีความถูกต้องและเหมาะสมกับความต้องการของผู้ใช้งานได้อย่างมีประสิทธิภาพ สามารถสรุปผลการวัดความใกล้เคียงด้วยวิธีการโคไซต์ (Cosine) ดังตารางที่ 1 และรูปที่ 5 และสรุปผลการวัดความใกล้เคียงด้วยวิธีการแอดจัสโคไซต์ (Adjusted Cosine) ดังตารางที่ 2 และรูปที่ 6

บทความวิจัย (Research Article)

ตารางที่ 1 ผลการเปรียบเทียบวัดความใกล้เคียงด้วยวิธีการโคไซต์ (Cosine) และวิธีการแอดจัสโคไซต์ (Adjusted Cosine)

ข้อมูลทดสอบ	ค่าความถูกต้อง (Accuracy) (%)			
	วิธีการโคไซต์ (Cosine)		วิธีการแอดจัสโคไซต์ (Adjusted Cosine)	
	แนะนำ 10 เรื่อง	แนะนำ 5 เรื่อง	แนะนำ10 เรื่อง	แนะนำ 5 เรื่อง
5%	92.92	**92.97**	92.50	88.71
10%	89.04	90.68	61.28	64.28
20%	88.68	91.91	76.19	78.17



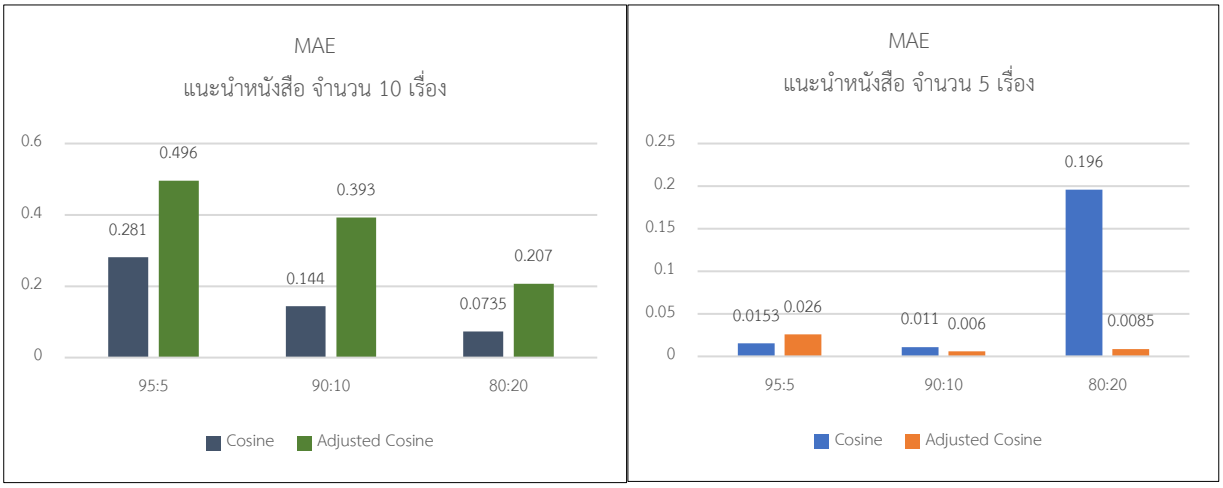
รูปที่ 5 เปรียบเทียบค่าความถูกต้องแนะนำหนังสือ จำนวน 10 เรื่อง และ 5 เรื่อง

จากตารางที่ 1 และรูปที่ 5 พบว่า การวัดความใกล้เคียงด้วยวิธีการโคไซต์และวิธีการแอดจัสโคไซต์ระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ด้วยการแยกข้อมูลเป็นชุดทดสอบ จำนวน 5%, 10%, 20% และแนะนำหนังสือ จำนวน 10 เรื่อง และ 5 เรื่อง ตามลำดับ ผลปรากฏว่า วิธีการแบบโคไซต์ มีค่าความถูกต้อง (Accuracy) มากที่สุดเท่ากับ 92.97% สำหรับแนะนำหนังสือ (TopN) จำนวน 5 เรื่อง จากข้อมูลทดสอบ 5% รองลงมาวิธีการแบบโคไซต์ มีค่าความถูกต้อง (Accuracy) เท่ากับ 92.92% สำหรับแนะนำหนังสือ (TopN) จำนวน 10 เรื่อง จากข้อมูลทดสอบ 5% และค่าความถูกต้อง (Accuracy) น้อยที่สุด คือวิธีการแบบแอดจัสโคไซต์ มีค่าความถูกต้อง (Accuracy) เท่ากับ 61.28% สำหรับแนะนำหนังสือ (TopN) จำนวน 10 เรื่อง จากข้อมูลทดสอบ 10%

บทความวิจัย (Research Article)

ตารางที่ 2 ผลการเปรียบเทียบค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) ด้วยวิธีการโคไซต์ (Cosine) และวิธีการแอดจัสโคไซต์ (Adjusted Cosine)

ข้อมูลทดสอบ	ค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE)			
	วิธีการโคไซต์ (Cosine)		วิธีการแอดจัสโคไซต์ (Adjusted Cosine)	
	แนะนำ 10 เรื่อง	แนะนำ 5 เรื่อง	แนะนำ 10 เรื่อง	แนะนำ 5 เรื่อง
5%	0.2819	0.0153	0.4960	0.0260
10%	0.1440	0.0110	0.3936	**0.0060**
20%	0.0741	0.1960	0.2073	0.0085



รูปที่ 6 เปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์ แนะนำหนังสือ จำนวน 10 เรื่อง และ 5 เรื่อง

จากตารางที่ 2 และรูปที่ 6 พบว่า ค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) ด้วยวิธีการโคไซต์และวิธีการแอดจัสโคไซต์ ระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา ด้วยการแยกข้อมูลเป็นชุดทดสอบ จำนวน 5%, 10%, 20% และแนะนำหนังสือ จำนวน 10 เรื่อง และ 5 เรื่อง ตามลำดับ ผลปรากฏว่า วิธีการแบบแอดจัสโคไซต์ มีค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) น้อยที่สุดเท่ากับ 0.0060 สำหรับแนะนำหนังสือ (TopN) จำนวน 5 เรื่อง จากข้อมูลทดสอบ 10% รองลงมาวิธีการแบบแอดจัสโคไซต์ มีค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) เท่ากับ 0.0085 สำหรับแนะนำหนังสือ (TopN) จำนวน 5 เรื่อง จากข้อมูลทดสอบ 20% และค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) มากที่สุด คือวิธีการแบบแอดจัสโคไซต์ มีค่าคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAE) เท่ากับ เท่ากับ 0.4960 สำหรับ

บทความวิจัย (Research Article)

แนะนำหนังสือ (TopN) จำนวน 10 เรื่อง จากข้อมูลทดสอบ 5%

7. อภิปรายผล

จากผลการพัฒนาอัลกอริทึมสำหรับระบบแนะนำหนังสือแบบอิงเนื้อหาและทดลองแสดงให้เห็นถึงประสิทธิภาพของอัลกอริทึมที่ใช้เทคนิค Cosine Similarity มีประสิทธิภาพสูงกว่า Adjusted Cosine Similarity ในแง่ของค่าความแม่นยำ (Accuracy) ทั้งในกรณีแนะนำหนังสือจำนวน 5 เรื่อง และ 10 เรื่อง โดย Cosine Similarity มีค่าความแม่นยำสูงสุดอยู่ที่ 92.97% สำหรับการแนะนำ จำนวน 5 เรื่อง และความแม่นยำ 92.92% สำหรับการแนะนำ จำนวน 10 เรื่อง ซึ่งแสดงถึงความสามารถของอัลกอริทึมในการจับความคล้ายคลึงของเนื้อหาหนังสือได้อย่างมีประสิทธิภาพ ในขณะที่ Adjusted Cosine Similarity มีข้อได้เปรียบในด้านค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAE) ที่ต่ำกว่าในกรณีแนะนำหนังสือ จำนวน 5 เรื่อง แต่มี MAE ที่สูงกว่าเมื่อเพิ่มจำนวนรายการเป็น 10 เรื่อง ที่เป็นเช่นนี้อาจเป็นเพราะการจัดการค่าถ่วงน้ำหนักที่ซับซ้อนมากขึ้นในชุดข้อมูลขนาดใหญ่

จากผลการทดลอง Cosine Similarity จึงเหมาะสมกับระบบแนะนำหนังสือที่เน้นความแม่นยำสูง โดยสามารถทำงานได้ดีแม้ในกรณีที่มีข้อมูลจำนวนมาก นอกจากนี้ยังสอดคล้องกับงานวิจัยของ Sarma และคณะ [18] ที่ใช้ Cosine Similarity ร่วมกับการจัดกลุ่ม K-means เพื่อพัฒนาระบบแนะนำหนังสือส่วนบุคคล และการศึกษาของ Kommineni และคณะ [17] ที่ศึกษาอัลกอริทึมการกรองแบบร่วมมือกันตามผู้ใช้ ซึ่งพบว่าการใช้ Cosine Similarity ให้ผลลัพธ์ที่มีความแม่นยำสูงและสามารถแนะนำหนังสือได้อย่างมีประสิทธิภาพ ดังนั้นจะเห็นได้ว่า Cosine Similarity เป็นวิธีที่เหมาะสมสำหรับกรณีที่ต้องการวัดความคล้ายคลึงของเนื้อหาหนังสือในระดับที่ดี โดยไม่ต้องพึ่งพาคะแนนจากผู้ใช้ ขณะที่ Adjusted Cosine Similarity จะมีความซับซ้อนในการคำนวณในกรณีที่ข้อมูลมีจำนวนมากขึ้น ซึ่งอาจส่งผลให้มีการลดประสิทธิภาพในบางกรณีที่มีรายการแนะนำจำนวนมาก นอกจากนี้ การเปรียบเทียบผลในกรณีแนะนำหนังสือ จำนวน 5 และ 10 เรื่อง ยังชี้ให้เห็นว่าความแม่นยำลดลงเล็กน้อยเมื่อจำนวนรายการหนังสือเพิ่มขึ้น หรือปริมาณข้อมูลเพิ่มขึ้น ซึ่งเป็นข้อท้าทายในการออกแบบระบบแนะนำแบบอิงเนื้อหา

8. สรุปผล

การพัฒนาระบบแนะนำหนังสือด้วยวิธีการแบบอิงเนื้อหา เป็นการพัฒนาอัลกอริทึมด้วยวิธีการวัดความใกล้เคียงแบบ Cosine Similarity และ Adjusted Cosine Similarity สามารถตอบสนองความต้องการในการแนะนำหนังสือแบบอิงเนื้อหาได้ดีในระดับหนึ่ง โดย Cosine Similarity มีความโดดเด่นในด้านความแม่นยำ เหมาะสมสำหรับการประยุกต์ใช้ในสถานการณ์ที่ต้องการผลลัพธ์ที่ครอบคลุมและแม่นยำสูง ในขณะที่ Adjusted Cosine Similarity อาจเหมาะสมในบริบทที่ต้องการลดข้อผิดพลาดในกรณีข้อมูลที่มีความละเอียดอ่อน อย่างไรก็ตาม ข้อจำกัดบางประการ เช่น ความหลากหลายของข้อมูลและข้อจำกัดในเชิงคำนวณ ยังเปิดโอกาสสำหรับการพัฒนาเพิ่มเติม เช่น การรวมเทคนิคแบบไฮบริด (Hybrid) การเพิ่มข้อมูลความคิดเห็นผู้ใช้

(User Feedback) หรือการใช้โมเดลเชิงลึก (Deep Learning) ที่สามารถจัดการความสัมพันธ์เชิงซับซ้อนได้ดียิ่งขึ้น ผลการวิจัยนี้ชี้ให้เห็นถึงความสำคัญของการเลือกเทคนิคที่เหมาะสมกับลักษณะของข้อมูลและความต้องการในระบบแนะนำ นอกจากนี้ การพัฒนาระบบในอนาคตควรพิจารณาการเพิ่มข้อมูลผู้ใช้หรือการผสานโมเดลแบบไฮบริด เพื่อเพิ่มความแม่นยำและตอบสนองความต้องการของผู้ใช้งานได้ดียิ่งขึ้น

9. ข้อเสนอแนะ

การพัฒนาระบบแนะนำหนังสือแบบอิงเนื้อหาเป็นแนวทางที่สำคัญสำหรับการปรับปรุงประสบการณ์ของผู้ใช้งานในระบบสารสนเทศ โดยเฉพาะในบริบทที่มีชุดข้อมูลขนาดใหญ่และความหลากหลายของเนื้อหา จากการทดลองวิจัยดังกล่าว มีข้อเสนอแนะดังนี้

1) ควรเพิ่มจำนวนและความหลากหลายของข้อมูลหนังสือในฐานข้อมูลเพื่อให้ระบบแนะนำสามารถรองรับประเภทหนังสือที่หลากหลายมากขึ้นและเพิ่มความแม่นยำของอัลกอริทึมในบริบทที่กว้างขึ้น

2) การเพิ่มข้อมูลจากความคิดเห็นของผู้ใช้ (User Feedback) เช่น การให้คะแนนหรือรีวิวหนังสือสามารถช่วยเพิ่มประสิทธิภาพของระบบแนะนำโดยเฉพาะเมื่อรวมกับโมเดลที่เน้นพฤติกรรมของผู้ใช้ (Collaborative Filtering)

3) ควรพัฒนาอัลกอริทึมที่สามารถผสานข้อมูลจากพฤติกรรมผู้ใช้และเนื้อหาหนังสือ เพื่อเพิ่มประสิทธิภาพในการแนะนำรายการที่สอดคล้องกับความต้องการของผู้ใช้มากขึ้น การทดลองใช้โมเดลการเรียนรู้เชิงลึก เช่น Deep Neural Networks จะช่วยเพิ่มศักยภาพของระบบให้สามารถประมวลผลข้อมูลเชิงซับซ้อนในอนาคตได้ดียิ่งขึ้น

10. กิตติกรรมประกาศ

ขอขอบคุณคณาจารย์ทุกท่าน สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยีสุขภาพ มหาวิทยาลัยกาฬสินธุ์ ที่ช่วยแนะนำ ให้ข้อเสนอแนะ

11. เอกสารอ้างอิง

[1] A. Jadon and A. Patil, "A comprehensive survey of evaluation techniques for recommendation systems," in *Proceedings of the International Conference on Computation of Artificial Intelligence & Machine Learning*, Jan. 2024, pp. 281–304.

[2] S. Wang *et al.*, "A survey on session-based recommender systems," *ACM Computing Surveys*, vol. 54, no. 7, pp. 1–38, 2021.

[3] S. R. S. Reddy *et al.*, "Content-based movie recommendation system using genre correlation," in *Smart Intelligent Computing and Applications: Proceedings of the Second International*

Conference on SCI 2018, vol. 2, pp. 391–397, 2019.

[4] P. Lops *et al.*, "Trends in content-based recommendation: Preface to the special issue on recommender systems based on rich item descriptions," *User Modeling and User-Adapted Interaction*, vol. 29, pp. 239–249, 2019.

[5] ธนพร เฟื่องขจร, "การพัฒนาบริการแนะนำหนังสือโดยประยุกต์ใช้เทคนิคเหมืองข้อมูลและเทคโนโลยีความเป็นจริงเสริม," วิทยานิพนธ์, มหาวิทยาลัยมหาสารคาม, 2564.

[6] โกเมส อัมพวัน *และคณะ*, "ระบบแนะนำหนังสืออ่านเพิ่มเติมเพื่อส่งเสริมการอ่านและการศึกษาด้วยตนเอง," มหาวิทยาลัยบูรพา, 2561.

[7] ธรรมนูญ ปัญญาทิพย์, "ระบบให้คำแนะนำภาพยนตร์ด้วยวิธีการผสมผสาน," วิทยานิพนธ์, มหาวิทยาลัยมหาสารคาม, 2562.

[8] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Impact of feature selection on content-based recommendation system," in *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, Apr. 2019, pp. 1–6.

[9] S. Sharma, V. Rana, and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm," *Education and Information Technologies*, vol. 27, no. 2, pp. 1523–1538, 2022.

[10] O. Stitini, S. Kaloun, and O. Bencharef, "An improved recommender system solution to mitigate the over-specialization problem using genetic algorithms," *Electronics*, vol. 11, no. 2, p. 242, 2022.

[11] P. Jomsri *et al.*, "Hybrid recommender system model for digital library from multiple online publishers," *F1000Research*, vol. 12, pp. 1–23, 2024.

[12] S. Amara and R. R. Subramanian, "Collaborating personalized recommender system and content-based recommender system using TextCorpus," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020, pp. 105–109.

[13] Y. H. Lee *et al.*, "Small clues tell: A collaborative expansion approach for effective content-based recommendations," *Journal of Organizational Computing and Electronic Commerce*, vol. 30, no. 2, pp. 111–128, 2020.

[14] A. Gholami and Y. Forghani, "Improving multi-class co-clustering-based collaborative recommendation using item tags," *Revue d'Intelligence Artificielle*, vol. 34, no. 1, 2020.

[15] S. S. E. Alqallaf, W. M. Medhat, and T. A. El-Shishtawy, "A hybrid recommender framework for selecting a course reference books," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 4, pp. 1004–1013, 2022.

[16] A. A. Amer, H. I. Abdalla, and L. Nguyen, "Enhancing recommendation systems performance using highly-effective similarity measures," *Knowledge-Based Systems*, vol. 217, p. 106842, 2021.

[17] M. Kommineni *et al.*, "Machine learning based efficient recommendation system for book selection using user based collaborative filtering algorithm," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, Jan. 2020, pp. 66–71.

[18] D. Sarma, T. Mittra, and M. S. Hossain, "Personalized book recommendation system using machine learning algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[19] A. K. Sharma *et al.*, "An efficient approach of product recommendation system using NLP technique," *Materials Today: Proceedings*, vol. 80, pp. 3730–3743, 2023.

[20] Y. C. Chen, L. Hui, and T. Thaipisutikul, "A collaborative filtering recommendation system with dynamic time decay," *The Journal of Supercomputing*, vol. 77, pp. 244–262, 2021.

[21] N. Ifada, T. F. Rahman, and M. K. Sophan, "Comparing collaborative filtering and hybrid based approaches for movie recommendation," in *2020 6th Information Technology International Seminar (ITIS)*, Oct. 2020, pp. 219–223.

[22] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: Recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, p. 141, 2022