ระบบแนะนำหนังสืออ่านเพิ่มเติม

Supplementary books recommendation system

เบญจมวรรณ ใจสูงเนิน

มหาวิทยาลัยบูรพา

2564

ระบบแนะนำหนังสืออ่านเพิ่มเติม

เบญจมวรรณ ใจสูงเนิน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการสารสนเทศ
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา
2564
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

Supplementary books recommendation system

BENCHAMAWAN CHAISOONGNOEN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR MASTER OF SCIENCE

IN INFORMATICS

FACULTY OF INFORMATICS

BURAPHA UNIVERSITY

2021

COPYRIGHT OF BURAPHA UNIVERSITY

คณะกรรมการควบคุมวิทยานิพนธ์และคณะกรรมการสอบวิทยานิพนธ์ได้พิจารณาวิทยานิพนธ์ของ เบญจมวรรณ ใจสูงเนิน ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศ ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมวิทยานิพนธ์

..... อาจารย์ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร. โกเมศ อัมพวัน)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธาน

(รองศาสตราจารย์ ดร. อนุชิต จิตพัฒนกุล)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. โกเมศ อัมพวัน)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. อุรีรัฐ สุขสวัสดิ์ชน)

คณะวิทยาการสารสนเทศอนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศ ของมหาวิทยาลัยบูรพา

..... คณบดีคณะวิทยาการสารสนเทศ

(ผู้ช่วยศาสตราจารย์ ดร. กฤษณะ ชินสาร)

วันที่.....เดือน.....พ.ศ.....

59910228: สาขาวิชา: วิทยาการสารสนเทศ; วท.ม. (วิทยาการสารสนเทศ)
คำสำคัญ: ระบบแนะนำหนังสือ, การสกัดคำสำคัญ

เบญจมวรรณ ใจสูงเนิน : ระบบแนะนำหนังสืออ่านเพิ่มเติม. (Supplementary books recommendation system) คณะกรรมการควบคุมวิทยานิพนธ์: โกเมศ อัมพวัน ปี พ.ศ. 2564.

ระบบค้นหาหนังสือและระบบแนะนำหนังสือถูกพัฒนาเพื่อช่วยให้ผู้อ่านมีความสะดวกในการเลือกหนังสือที่ตนเองสนใจ แต่อย่างไรก็ตาม ระบบที่ถูกพัฒนาขึ้นไม่ได้มุ่งเน้นที่การแนะนำหนังสือที่สอดคล้องกับการศึกษาในชั้นเรียน และไม่ได้มุ่งเน้นที่การแนะนำหนังสือในศาสตร์ที่มีคำศัพท์เฉพาะทางค่อนข้างมาก ดังนั้น งานวิจัยนี้จึงมีแนวคิดที่จะนำเสนอระบบแนะนำหนังสืออ่านเพิ่มเติม (เรียกโดยย่อว่า “ระบบเอสบีเอส”) เพื่อทำการแนะนำหนังสืออ่านเพิ่มเติมที่เกี่ยวเนื่องกับเนื้อหาในการเรียนวิชาหนึ่ง ๆ ในหลักสูตรวิทยาการคอมพิวเตอร์ โดยทำการพิจารณาชื่อรายวิชาและคำอธิบายรายวิชาของรายวิชาหนึ่ง ๆ ร่วมกับคำอธิบายหนังสือและสารบัญของหนังสือเล่มหนึ่ง ๆ ผ่านการดำเนินงาน 2 ขั้นตอนหลัก คือ 1) การเตรียมข้อมูลเบื้องต้น ซึ่งทำการรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี และการสร้างกฎทางภาษาศาสตร์เพื่อสกัดคำสำคัญ และ 2) การประมวลผลของระบบแนะนำ ซึ่งทำการสกัดคำสำคัญจากคำอธิบายรายวิชา, การค้นหาหนังสือที่เกี่ยวเนื่อง การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา และการสร้างรายการแนะนำหนังสือให้กับผู้ใช้งาน นอกจากนั้น งานวิจัยนี้ได้ทำการพัฒนาปรับปรุงระบบเอสบีเอส โดยนำเสนอระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ (เรียกโดยย่อว่า “ระบบเอสบีเอสพลัส”) โดยพัฒนาการดำเนินงานทั้งสองขั้นตอนข้างต้น เพื่อให้สามารถสกัดคำสำคัญและการคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชาได้แม่นยำมากขึ้น ท้ายสุด การประเมินประสิทธิภาพของระบบแนะนำที่นำเสนอดำเนินการกับ 640 คำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ จาก 13 มหาวิทยาลัยในประเทศไทย ประกอบด้วย 2 ส่วนได้แก่ 1) การประเมินประสิทธิภาพของการสกัดคำสำคัญที่ใช้ในระบบ และ 2) การประเมินประสิทธิภาพของระบบแนะนำ โดยทำการเปรียบเทียบกับระบบแนะนำที่สร้างด้วยขั้นตอนวิธีสกัดคำสำคัญ “เทอร์ไมน์” และ “เรค” ซึ่งเป็นขั้นตอนวิธีที่ได้รับความนิยมอย่างแพร่หลาย โดยจากผลการประเมินประสิทธิภาพสามารถสรุปได้ว่าระบบเอสบีเอสพลัสมีประสิทธิภาพมากกว่าระบบอื่น ๆ ที่นำมาเปรียบเทียบทุกแง่มุม

59910228: MAJOR: INFORMATICS; M.Sc. (INFORMATICS)
KEYWORDS: Book recommendation system, Keyword extraction

BENCHAMAWAN CHAISOONGNOEN : SUPPLEMENTARY BOOKS RECOMMENDATION SYSTEM. ADVISORY COMMITTEE: KOMATE AMPHAWAN, Ph.D. 2021.

Book searching and recommendation systems have been developed to facilitate readers to choose the books they interest. However, most of the systems do not focus on recommending books that are consistent with a course in a curriculum or a course that students have registered, and also do not focus on recommending books in the domain having a large number of terminologies. From these issues, this research aims to introduce a new system called “Supplementary Books Suggestion system (also called *SBS* system)” to recommend books related to a course in a computer science curriculum. *SBS* system regards the course name and course description of a course together with the book description and table of contents (ToC) of the book accordance to two main operations i.e. 1) Data preparation--including gathering terminology in computer technology and the establishment of linguistic rules to extract keywords and 2) Processing of the recommendation system--including keyword extraction, relevance book searching and similarity calculation, and creating the list of Top-N most interesting supplementary. Furthermore, a new improvement of *SBS* system called “an efficient Supplementary Book Suggestion system (*SBS*+ system)” is also developed. *SBS*+ improves the steps of extract keywords and calculate the similarity between the course and the book. Last, experiments were conducted to evaluate effectiveness and efficient of *SBS* and *SBS*+ on 640 computer science courses from 13 universities in Thailand. The experiment consists of two aspects i.e. 1) Effectiveness of keyword extraction and 2) Performance of the recommendation system. Moreover, a comparative study between *SBS* and *SBS*+ together with *Termine* and *RAKE* (the most popular keyword extraction algorithms). From the results, it can be seen and conclude that *SBS*+ system is more efficient than other systems in all aspects.

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงโดยความกรุณาจาก ผู้ช่วยศาสตราจารย์ ดร. โกเมศ อัมพวัน อาจารย์ที่ปรึกษา ที่กรุณาให้คำปรึกษา คำแนะนำ แนะแนวทางอันเป็นประโยชน์อย่างยิ่งสำหรับงานวิจัย อีกทั้งยังมอบประสบการณ์ในการดำเนินงานที่ดี ติดตามความก้าวหน้า ช่วยเหลือแก้ปัญหาข้อบกพร่องต่าง ๆ ที่เกิดขึ้นในงานวิจัยด้วยความละเอียดถี่ถ้วนและเอาใจใส่ เพื่อให้งานวิจัยฉบับนี้มีคุณภาพที่ดียิ่งขึ้น ตลอดจนคอยผลักดันและเป็นกำลังใจที่ดีเสมอมา ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของอาจารย์ที่ปรึกษาเป็นอย่างยิ่ง และขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณคุณแม่อุบล เทพบุตร และพี่สาว ที่คอยสนับสนุน ให้คำปรึกษา ให้กำลังใจ และเคียงข้างในยามท้อ เหนื่อย และหมดหวัง ทั้งคุณแม่และพี่สาวเป็นผู้สนับสนุนและกำลังใจที่ดีของผู้วิจัยเสมอมา

ขอขอบคุณ เพื่อน ๆ พี่ ๆ และ น้อง ๆ ทุกคนในห้องปฏิบัติการวิจัย Computational Innovation Laboratory (CIL) ที่มอบกำลังใจ คอยช่วยเหลือ ให้คำปรึกษา และมอบมิตรภาพที่ดีกับผู้วิจัย ตลอดจนคณะผู้บริหาร อาจารย์ และบุคลากรในคณะวิทยาการสารสนเทศที่คอยช่วยเหลือสนับสนุน แก้ปัญหาให้กับผู้วิจัยเสมอมา

เนื่องจากงานวิจัยฉบับนี้ได้รับทุนอุดหนุนวิทยานิพนธ์หรือดุษฎีนิพนธ์ สำหรับนิสิตระดับบัณฑิตศึกษาชาวไทย ปีงบประมาณ 2561 จากมหาวิทยาลัยบูรพา จึงขอขอบคุณมา ณ ที่นี้

สุดท้ายนี้ ขอบคุณตนเองที่มีจิตใจเข้มแข็ง พยายาม อดทน และไม่ย่อท้อต่ออุปสรรคต่าง ๆ จนสามารถดำเนินงานวิจัยฉบับนี้ได้สำเร็จลุล่วง

เบญจมวรรณ ใจสูงเนิน

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฎ

สารบัญภาพ ..... ฎ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 2

1.3 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 3

1.4 ขอบเขตของงานวิจัย 3

1.5 แผนการดำเนินงานวิจัย 4

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 6

2.1 ระบบแนะนำ (Recommender system) 6

2.1.1 วิธีการกรองเนื้อหา (Content-based filtering) 7

2.1.2 วิธีการกรองข้อมูลร่วม (Collaborative filtering) 8

2.1.3 วิธีการแบบผสมผสาน (Hybrid filtering) 9

2.2 การประมวลผลข้อความเบื้องต้น (Text preprocessing) 9

2.2.1 การตัดประโยค (Sentence tokenization) 10

2.2.2 การตัดคำ (Word tokenization) 10

2.2.3 การกำจัดคำหยุด (Stop word removal) 10

2.2.4 การวิเคราะห์หน้าที่ของคำ (Part of speech tagging) 10
2.2.5 การลดรูปของคำและการหารูปเดิมของคำ (Word stemming and lemmatization) 11
2.3 การสกัดคำสำคัญ (Keyword extraction) 11
2.3.1 วิธีการทางสถิติ (Statistical approach) 11
2.3.2 วิธีการทางภาษาศาสตร์ (Linguistic approach) 12
2.3.3 วิธีการทางการเรียนรู้ของเครื่องจักร (Machine learning approach) 12
2.3.4 วิธีการแบบผสมผสาน (Hybrid approach) 13
2.4 งานวิจัยที่เกี่ยวข้อง 13
2.4.1 งานวิจัยที่เกี่ยวข้องทางด้านการแนะนำหนังสือ (Recommender system) 13
2.4.1.1 การแนะนำหนังสือจากการพิจารณาเนื้อหาสาระในหนังสือ 13
2.4.1.2 การแนะนำหนังสือจากการให้คะแนนความนิยมของผู้อ่าน 14
2.4.1.3 การแนะนำหนังสือที่สืบเนื่องจากการแนะนำผู้เขียน 15
2.4.1.4 การแนะนำหนังสือจากประวัติการยืมหนังสือของผู้ใช้ห้องสมุด 17
2.4.1.5 การแนะนำหนังสือโดยพิจารณาจากระดับทักษะการอ่าน 19
2.4.2 งานวิจัยที่เกี่ยวข้องทางด้านการสกัดคำสำคัญ (Keyword extraction) 19
บทที่ 3 ระบบแนะนำหนังสืออ่านเพิ่มเติม 24
3.1 การเตรียมข้อมูลเบื้องต้น 25
3.1.1 การรวบรวมข้อมูลคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ 25
3.1.2 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี 25
3.1.3 การสร้างกฎทางภาษาศาสตร์ 27
3.2 การประมวลผลของระบบแนะนำ 28
3.2.1 การสกัดคำสำคัญ 28
3.2.1.1 การประมวลผลข้อความ 28

3.2.1.2 การระบุคำศัพท์เฉพาะ 32
3.2.1.3 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ 33
3.2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา 35
3.2.3 การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา และการจัดอันดับหนังสือ 38
บทที่ 4 ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ 42
4.1 การเตรียมข้อมูลเบื้องต้น 45
4.1.1 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี 45
4.1.2 การรวบรวมคำคุณศัพท์ 46
4.1.3 การปรับปรุงกฎทางภาษาศาสตร์ 46
4.2 การประมวลผลของระบบแนะนำ 47
4.2.1 การสกัดคำสำคัญจากคำอธิบายรายวิชา 47
4.2.1.1 การประมวลผลข้อความ (Text processing) 49
4.2.1.2 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี (Terminology identification) 52
4.2.1.3 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุง 54
4.2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา 57
4.2.3 การคำนวณความเกี่ยวเนื่องของหนังสือที่ครอบคลุมเนื้อหาในคำอธิบายรายวิชา 57
4.2.3.1 วิธีการเปรียบเทียบแบบตรงตัว (Exact matching) 58
4.2.3.2 วิธีการเปรียบเทียบแบบวลี (Paraphrase matching) 58
4.2.3.3 วิธีการเปรียบเทียบแบบพิจารณาคำย่อย (Sub-keyword matching) 58
4.3 การออกแบบการใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม 64
บทที่ 5 ผลการดำเนินงาน 68
5.1 การประเมินประสิทธิภาพของการสกัดคำสำคัญ 69

5.2 การประเมินการประมวลผลของระบบแนะนำ 74
5.3 การใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม 82
บทที่ 6 สรุปและอภิปรายผล 90
6.1 สรุปผลการดำเนินงาน 90
6.2 ข้อเสนอแนะ 94
บรรณานุกรม 94
ภาคผนวก 100
ภาคผนวก ก กฎทางภาษาศาสตร์สำหรับระบบ *SBS* 101
ภาคผนวก ข กฎทางภาษาศาสตร์สำหรับระบบ *SBS+* 131
ภาคผนวก ค เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์ 135
ภาคผนวก ง เอกสารเผยแพร่ผลงานวิจัย 137
ประวัติย่อของผู้วิจัย 152

สารบัญตาราง

หน้า

ตารางที่ 1 แผนการดำเนินงานวิจัย 4
ตารางที่ 2 รายการการปรับปรุงการเตรียมข้อมูลเบื้องต้นของระบบ SBS+ 44
ตารางที่ 3 รายการการปรับปรุงการประมวลผลของระบบแนะนำ SBS+ 48
ตารางที่ 4 จำนวนรายวิชาจากหลักสูตรวิทยาการคอมพิวเตอร์ของ 13 มหาวิทยาลัย 68
ตารางที่ 5 ตารางประเมินประสิทธิภาพความถูกต้องระหว่างข้อมูลคำสำคัญที่เกิดขึ้นจริงและข้อมูลคำสำคัญที่สกัดได้ 69
ตารางที่ 6 จำนวนคำสำคัญที่สกัดได้และจำนวนคำสำคัญที่มีความถูกต้องของ 4 ขั้นตอนวิธี 71
ตารางที่ 7 การเปรียบเทียบค่าร้อยละของความแม่นยำโดยรวม (Accuracy) 71
ตารางที่ 8 การเปรียบเทียบค่าร้อยละของความแม่นยำ (Precision) 73
ตารางที่ 9 การเปรียบเทียบค่าร้อยละของความถูกต้อง (Recall) 73
ตารางที่ 10 การเปรียบเทียบค่าร้อยละของประสิทธิภาพโดยรวม (F-measure) 74
ตารางที่ 11 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS 102
ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS 105
ตารางที่ 13 กฎทางภาษาศาสตร์สำหรับระบบ SBS+ 132

สารบัญภาพ

หน้า

ภาพที่ 1 ระบบแนะนำประกอบไปด้วยขั้นตอนวิธีต่าง ๆ 7
ภาพที่ 2 ขั้นตอนวิธีการกรองเนื้อหา 7
ภาพที่ 3 ขั้นตอนวิธีการกรองข้อมูลร่วม 8
ภาพที่ 4 ระบบการแนะนำหนังสือแบบไฮบริดรวมถึงการพิจารณาข้อมูลผู้อ่าน 15
ภาพที่ 5 ขั้นตอนการแนะนำแบบ top-n items 16
ภาพที่ 6 โครงสร้างข้อมูลต้นไม้ 3 ระดับ 17
ภาพที่ 7 การแบ่งปัญหาออกเป็นปัญหาย่อย 18
ภาพที่ 8 โครงสร้างระบบแนะนำหนังสืออ่านเพิ่มเติม 24
ภาพที่ 9 การจัดเก็บข้อมูลชื่อวิชาและคำอธิบายรายวิชาลงในฐานข้อมูล 25
ภาพที่ 10 ตัวอย่างคำศัพท์เฉพาะจากพจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด 26
ภาพที่ 11 ตัวอย่างคำศัพท์เฉพาะจากเว็บไซต์ Labautopedia 27
ภาพที่ 12 ตัวอย่างข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining” 29
ภาพที่ 13 การประมวลผลข้อความจากข้อมูลคำอธิบายรายวิชา “Data mining” 29
ภาพที่ 14 การระบุคำศัพท์เฉพาะจากข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining” 33
ภาพที่ 15 การระบุคำสำคัญในคำอธิบายรายวิชา “Data mining” โดยกฎทางภาษาศาสตร์ 35
ภาพที่ 16 ข้อมูลหนังสือที่ทำการเก็บรวมรวมไว้ในฐานข้อมูลของหนังสือ 37
ภาพที่ 17 การเปรียบเทียบคำสำคัญจากคำอธิบายรายวิชา “Data mining” และหนังสือ “Introduction to Data Mining” 39
ภาพที่ 18 การคำนวณคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวข้อง 40
ภาพที่ 19 การคำนวณคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวข้อง 40

ภาพที่ 20 โครงสร้างระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ 43
ภาพที่ 21 ข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining” 49
ภาพที่ 22 การประมวลผลข้อความของระบบ SBS+ จากข้อมูลคำอธิบายรายวิชา “Data mining” 50
ภาพที่ 23 การระบุคำศัพท์เฉพาะจากข้อมูลคำอธิบายรายวิชา “Data mining” 53
ภาพที่ 24 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุงกับคำอธิบายรายวิชา “Data mining” 56
ภาพที่ 25 ผลลัพธ์จากการสกัดคำสำคัญจากคำอธิบายรายวิชาของระบบ SBS+ 56
ภาพที่ 26 การเปรียบเทียบความคล้ายคลึงระหว่างคำสำคัญในคำอธิบายรายวิชาและคำสำคัญในหนังสือ “Data mining: Concepts and Techniques” และการคำนวณคะแนนความคล้ายคลึง 61
ภาพที่ 27 ออกแบบการแสดงผลเพื่อรับข้อมูลจากผู้ใช้ของระบบแนะนำหนังสืออ่านเพิ่มเติม 64
ภาพที่ 28 ออกแบบการแสดงผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความคล้ายคลึง 65
ภาพที่ 29 ออกแบบการแสดงผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความไม่คล้ายคลึง 66
ภาพที่ 30 การเปรียบเทียบค่าความครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติม 77
ภาพที่ 31 การเปรียบเทียบค่าความครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติมโดยใช้วิธีการเปรียบเทียบแบบตรงตัว, แบบวลี และแบบพิจารณาคำย่อย 78
ภาพที่ 32 การเปรียบเทียบค่าความครอบคลุมเนื้อหาสะสมของระบบแนะนำหนังสืออ่านเพิ่มเติม .. 79
ภาพที่ 33 การเปรียบเทียบค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติม 80
ภาพที่ 34 การเปรียบเทียบค่าความไม่ครอบคลุมเนื้อหาสะสมของระบบแนะนำหนังสืออ่านเพิ่มเติม 81
ภาพที่ 35 การแสดงผลเพื่อรับข้อมูลจากผู้ใช้งานของระบบแนะนำหนังสืออ่านเพิ่มเติม 83
ภาพที่ 36 ผู้ใช้งานกรอกข้อมูลคำค้นที่สนใจในหน้าต่างการรับข้อมูลของระบบแนะนำ 84
ภาพที่ 37 การแสดงผลเมื่อระบบทำการบันทึกข้อมูลของผู้ใช้งานเรียบร้อยแล้ว 84

ภาพที่ 38 รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง 85
ภาพที่ 39 รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง
..... 86
ภาพที่ 40 การตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมไปยังผู้ใช้งาน 88
ภาพที่ 41 หน้าต่างอีเมล์ของผู้ใช้งานที่ได้รับอีเมล์ตอบกลับจากระบบ SBS+..... 88
ภาพที่ 42 ผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่ผู้ใช้งานได้รับจากอีเมล์ตอบกลับ 89

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเป็นยุคของข้อมูลข่าวสาร เนื่องจากความก้าวหน้าทางเทคโนโลยีสารสนเทศได้พัฒนาไปอย่างรวดเร็ว ทำให้ในแต่ละวันมีการผลิตข้อมูลขึ้นมาจำนวนมาก ซึ่งข้อมูลเหล่านั้นอาจจะประกอบไปด้วยข้อมูลข่าวสาร สารสนเทศ หรือองค์ความรู้ โดยสามารถค้นหาข้อมูลได้จากเว็บไซต์ต่าง ๆ และจากสังคมออนไลน์ ในการรับรู้ข้อมูลข่าวสาร หรือความรู้ใหม่ ๆ แต่อย่างไรก็ตาม หนังสือก็ยังเป็นหนึ่งในแหล่งการเรียนรู้ที่ดี และถูกแนะนำให้ใช้เป็นแหล่งการเรียนรู้อย่างกว้างขวาง (Akabuike and Asika, 2012) อีกทั้งการอ่านหนังสือก็ยังเป็นปัจจัยสำคัญต่อการพัฒนาตนเองให้รอบรู้ ขยายคลังคำศัพท์ของผู้อ่าน รวมทั้งพัฒนาด้านกระบวนการทางความคิดได้ โดยในปัจจุบันรูปแบบของหนังสือได้เปลี่ยนแปลงไปอย่างมาก โดยมีทั้งรูปเล่มที่ถูกพิมพ์จากสำนักพิมพ์ และในรูปแบบอิเล็กทรอนิกส์ ซึ่งจะช่วยเพิ่มความสะดวกในการเพิ่มพูนความรู้ที่มากขึ้น จากความหลากหลายในรูปแบบของหนังสือจึงได้มีระบบค้นหาหนังสือและระบบแนะนำหนังสือต่าง ๆ มากมาย เพื่อช่วยให้ผู้อ่านมีความสะดวกในการค้นหาและเลือกหนังสือที่ตนเองสนใจได้ดียิ่งขึ้น อาทิเช่น

1. การค้นหาหนังสือจากความสอดคล้องของคำค้นหนึ่ง ๆ กับชื่อหนังสือ กับสารบัญหนังสือ (Garrido and Ilarri, 2014; Garrido, Pera and Ilarri, 2014; Ali, Khusro and Ullah, 2016)
2. การแนะนำหนังสือจากความนิยมของหนังสือ (Linden, Smith and York, 2003; Chandak, Girase and Mukhopadhyay, 2015)
3. การแนะนำหนังสือจากข้อคิดเห็นที่ได้รับจากผู้อ่าน (Lu and Zhang, 2015; Priyanka, Tewari and Barman, 2015)
4. การแนะนำหนังสือจากพฤติกรรมการยืมหนังสือในห้องสมุด (Saxena, Singh and Khan, 2014; Chen, Kuo and Liao, 2015; Teng *et al.*, 2015; Jomsri, 2018)

จากระบบข้างต้นสามารถช่วยให้ผู้อ่านหนังสือมีข้อมูลพื้นฐานในการเลือกอ่านหนังสือได้เหมาะสมมากยิ่งขึ้น นอกจากนั้นยังสามารถเพิ่มคุณภาพการให้บริการของห้องสมุดต่อผู้ใช้บริการ และการบริหารจัดการหนังสือในห้องสมุดได้อีกด้วย แต่อย่างไรก็ตาม แนวคิดทั้งหมดไม่ได้มุ่งเน้นที่การแนะนำหนังสือด้วยการพิจารณาความสอดคล้องของหนังสือที่เกี่ยวเนื่องกับการศึกษาในชั้นเรียน และไม่ได้มุ่งเน้นในการแนะนำหนังสือที่เกี่ยวข้องกับศาสตร์ที่มีคำศัพท์เฉพาะทางค่อนข้างมากกับคำค้น (คำค้นหนึ่ง ๆ

สามารถมีได้หลายรูปแบบ อาทิเช่น คำหนึ่ง ๆ , คำหลาย ๆ คำ, คำอธิบายรายวิชาที่ประกอบไปด้วยเนื้อหาต่าง ๆ และ/หรือประมวลผลรายวิชา) ที่ซึ่งช่วยให้ผู้อ่านได้รับการแนะนำที่สอดคล้องกับเนื้อหาที่ตนเองสนใจมากยิ่งขึ้น ตัวอย่างเช่น ถ้าผู้อ่านหนังสือต้องการค้นหาหนังสือเกี่ยวกับระบบแนะนำ (Recommendation system) ที่จะประกอบไปด้วย “นิยาม” “ขั้นตอนวิธีในการแนะนำ” “การวัดและประเมินประสิทธิภาพของการแนะนำ” “ตัวอย่าง/กรณีศึกษาของการประยุกต์ใช้ระบบแนะนำ” และอื่น ๆ หรืออาจเป็นการค้นหาหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาหนึ่ง ๆ อาทิเช่น คำอธิบายของรายวิชา การทำเหมืองข้อมูล (Data mining) ที่มีคำอธิบายรายวิชาเป็น “The fundamental concepts of data mining. types of data for data mining. the famous techniques for data mining. pattern mining and association rules mining. classification. clustering. outlier analysis. anomaly detection. data mining tools.” ซึ่งมีเนื้อหาค่อนข้างมาก ดังนั้นเนื้อหาในหนังสือเล่มหนึ่ง ๆ ควรที่จะต้องมีเนื้อหาที่สอดคล้องกับคำอธิบายค่อนข้างมากและอาจมีคำศัพท์เฉพาะค่อนข้างมากและหลากหลาย โดยจากความต้องการหนังสือที่มีความเฉพาะเจาะจงค่อนข้างสูงและมีเนื้อหาที่มีขอบเขตความรู้เฉพาะด้าน ระบบการแนะนำที่คิดค้นจากงานวิจัยข้างต้นอาจไม่สามารถแนะนำหนังสือที่ตรงกับความต้องการของผู้อ่านเท่าที่ควร

ด้วยเหตุนี้งานวิจัยนี้จึงมุ่งเน้นที่การแนะนำหนังสือโดยพิจารณาจากความสอดคล้องของหนังสือที่เพิ่มเติมจากการศึกษาในชั้นเรียน ที่ซึ่งเป็นหนังสือที่เกี่ยวเนื่องกับคำค้นที่เฉพาะเจาะจงเป็นหลัก โดยจะมุ่งเน้นที่การแนะนำหนังสืออ่านเพิ่มเติมจากการศึกษาในมหาวิทยาลัยผ่านการพิจารณาคำอธิบายรายวิชา และ/หรือประมวลรายวิชา โดยรูปแบบการแนะนำหนังสืออ่านเพิ่มเติมจะเป็นการสร้างรายการของหนังสือที่มีความเกี่ยวข้องกับคำค้นหรือกลุ่มของคำค้นที่ผู้อ่านกำหนด (คำอธิบายรายวิชา และ/หรือประมวลรายวิชา) โดยหนังสือในรายการที่ถูกแนะนำจะถูกเรียงลำดับตามความเกี่ยวข้อง (เรียงลำดับจากความเกี่ยวข้องกับคำค้นมากที่สุดไปน้อยที่สุด) ซึ่งรายการแนะนำหนังสือที่ได้จะช่วยให้ผู้อ่านสามารถตัดสินใจที่จะเลือกหนังสืออ่านเพิ่มเติมได้ง่ายและสะดวกยิ่งขึ้น เพื่อส่งเสริมการอ่าน และการเรียนรู้ด้วยตนเอง อันนำมาซึ่งการพัฒนาตนเอง พัฒนาการเรียนรู้อย่างยั่งยืน

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาเกี่ยวกับการแนะนำหนังสืออ่านเพิ่มเติม การประยุกต์ใช้การแนะนำหนังสือในแง่มุมต่าง ๆ และประโยชน์ที่ได้รับจากการประยุกต์ใช้การแนะนำหนังสือ โดยศึกษาจากงานวิจัยก่อนหน้าและเว็บไซต์ที่เกี่ยวข้อง

2. เพื่อศึกษาและค้นหาแนวทางการปรับปรุงพัฒนาขั้นตอนวิธีในการแนะนำหนังสืออ่านเพิ่มเติมทางด้านวิทยาการคอมพิวเตอร์ เพื่อเป็นต้นแบบในการสร้างระบบแนะนำหนังสืออ่านเพิ่มเติมจากการศึกษาในชั้นเรียน และการเรียนรู้ด้วยตนเอง
3. เพื่อสร้างระบบการแนะนำหนังสือเพื่ออ่านเพิ่มเติมทางด้านวิทยาการคอมพิวเตอร์แบบอัตโนมัติที่สามารถใช้ได้จริง ที่ซึ่งสามารถอำนวยความสะดวกให้นิสิต/นักศึกษา/บุคคลทั่วไปใช้ในการเลือกหนังสือเพื่ออ่านเพิ่มเติมได้
4. เพื่อสร้างต้นแบบงานวิจัยทางด้านการสร้างระบบแนะนำหนังสือที่จะสามารถเปิดโอกาสให้ผู้ที่สนใจสามารถนำความคิดที่นำเสนอไปศึกษาเพิ่มเติม พัฒนาเป็นผลผลิตเพื่อใช้ในองค์กรของตนเองและเพื่อประยุกต์ใช้ในงานวิจัยของตนเองต่อไป

1.3 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. ได้ขั้นตอนวิธีในการประมวลผลการแนะนำหนังสืออ่านเพิ่มเติมทางด้านวิทยาการคอมพิวเตอร์
2. ได้ระบบการแนะนำหนังสืออ่านเพิ่มเติมทางด้านวิทยาการคอมพิวเตอร์ ที่สามารถใช้งานได้จริง เพื่อแนะนำหนังสือสำหรับการอ่านเพิ่มเติมให้กับนิสิต/นักศึกษา
3. ได้ผลงานวิจัยที่จะสามารถตีพิมพ์ในงานประชุมวิชาการหรือวารสารวิชาการ ดังนี้
    i) 2018, 5th International Conference on Advanced Informatics: Concept Theory and Applications
    ii) 2020, 9th International Conference on Smart Media and Applications
4. สามารถนำแนวคิด ขั้นตอนวิธี และระบบการแนะนำหนังสืออ่านเพิ่มเติมไปพัฒนาต่อยอดเพื่อดำเนินงานวิจัยขั้นสูงในการส่งเสริมการอ่านและการเรียนรู้ด้วยตนเองของนักเรียนและนิสิต/นักศึกษาได้

1.4 ขอบเขตของงานวิจัย

1. ระบบแนะนำหนังสือทำเกี่ยวกับศาสตร์ทางด้านวิทยาการคอมพิวเตอร์ที่ซึ่งเป็นศาสตร์ที่มีคำศัพท์เฉพาะค่อนข้างมาก และเป็นศาสตร์ที่เกี่ยวข้องกับผู้วิจัย
2. การแนะนำหนังสือจะพิจารณาจากคำค้นของผู้ใช้งานที่เป็นภาษาอังกฤษเท่านั้น โดยคำค้นเป็นกลุ่มของคำสำคัญที่ได้จากคำอธิบายรายวิชาหนึ่ง ๆ หรือประมวลผลรายวิชาต่าง ๆ ในสาขาวิทยาการคอมพิวเตอร์

3. การพิจารณาความน่าสนใจของหนังสือที่จะแนะนำจะพิจารณาจากเนื้อหาในหนังสือที่เกี่ยวเนื่องกับคำค้นของผู้ใช้งาน
4. หนังสือเล่มหนึ่ง ๆ ที่จะถูกพิจารณาหาความน่าสนใจ และถูกแนะนำให้แก่ผู้ใช้งานจะเป็นหนังสือที่สามารถสืบค้นได้จากเว็บไซต์ www.barnesandnoble.com และเว็บไซต์ของสำนักพิมพ์ต่าง ๆ
5. ผลลัพธ์ที่ได้จากการแนะนำหนังสือจะถูกเรียงลำดับตามความน่าสนใจ โดยเรียงลำดับจากความน่าสนใจมากไปยังความน่าสนใจน้อย
6. การประเมินผลของการแนะนำหนังสือจะสามารถวัดได้จากประสิทธิภาพในการสร้างรายการแนะนำหนังสืออ่านเพิ่มเติม

1.5 แผนการดำเนินงานวิจัย

งานวิจัยนี้มีขั้นตอนการดำเนินงานภายใต้ระยะเวลาที่ระบุไว้ดังตารางที่ 1

ตารางที่ 1 แผนการดำเนินงานวิจัย

ปี	การดำเนินงาน	เดือน											
		1	2	3	4	5	6	7	8	9	10	11	12
2559	ศึกษางานวิจัยที่เกี่ยวข้องกับการแนะนำหนังสือและการสกัดคำสำคัญ								←	—	—	—	→
	กำหนดขอบเขตของระบบแนะนำ											←	→
2560	ออกแบบขั้นตอนวิธีของระบบแนะนำ	←	—	→									
	รวบรวมข้อมูลสำหรับการประมวลผลของระบบแนะนำ				←	—	→						
	เขียนโปรแกรมประมวลผลการสกัดคำสำคัญและทำการทดสอบ							←	—	—	—	—	→

ตารางที่ 1 แผนการดำเนินงานวิจัย (ต่อ)

ปี	การดำเนินงาน	เดือน											
		1	2	3	4	5	6	7	8	9	10	11	12
2560	เขียนโปรแกรมประมวลผลการแนะนำหนังสือและทำการทดสอบ										←	—	→
2561	สรุปการประมวลผลระบบแนะนำและจัดทำเอกสารงานวิจัยที่ 1	←	—	→									
	จัดทำเอกสารและสอบโครงร่างวิทยานิพนธ์				←	—	→						
	เผยแพร่งานวิจัยที่ 1							↔					
	ศึกษาการปรับปรุงประสิทธิภาพของระบบแนะนำ							←	—	—	—	—	→
2562	ปรับปรุงประมวลผลการแนะนำหนังสือและทำการทดสอบ	←	—	—	—	—	—	—	—	—	—	—	→
2563	สรุปการประมวลผลระบบแนะนำและจัดทำเอกสารงานวิจัยที่ 2	←	—	—	—	—	→						
	เผยแพร่งานวิจัยที่ 2								↔	↔			
	ปรับปรุงการประมวลผลการสกัดคำสำคัญและทำการทดสอบ							←	—	→			
	สรุปการประมวลผลระบบแนะนำ										←	—	→
2564	จัดทำเอกสารและสอบวิทยานิพนธ์	←	—	→									

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ได้อธิบายถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบแนะนำหนังสืออ่านเพิ่มเติม ผู้วิจัยได้ทำการศึกษาทฤษฎีต่าง ๆ ความเป็นมาและความสำคัญ ขั้นตอนวิธี รวมถึงงานวิจัยที่เกี่ยวข้องที่ซึ่งสามารถนำมาประยุกต์ใช้ในการพัฒนาระบบแนะนำ โดยมีหัวข้อดังนี้

2.1 ระบบแนะนำ (Recommender system)

2.2 การประมวลผลข้อความเบื้องต้น (Text preprocessing)

2.3 การสกัดคำสำคัญ (Keyword extraction)

2.4 งานวิจัยที่เกี่ยวข้อง

2.1 ระบบแนะนำ (Recommender system)

ระบบแนะนำ (Recommender system) เป็นระบบที่ถูกสร้างขึ้นมาเพื่อช่วยในการคัดกรองข้อมูลที่คาดว่าจะตรงตามความต้องการ/ความสนใจจากข้อมูลที่มีอยู่จำนวนมาก (Isinkaye, Folajimi and Ojokoh, 2015) มีเป้าหมายในการสร้างรายการแนะนำข้อมูลที่คาดเดาว่าผู้ใช้ชื่นชอบ หรือให้ความสนใจ ซึ่งช่วยให้ผู้ใช้สามารถเข้าถึงข้อมูลสินค้าหรือบริการที่ชื่นชอบได้อย่างมีประสิทธิภาพ และสามารถเพิ่มโอกาสทางธุรกิจ ผลกำไรและรักษาฐานลูกค้าให้กับระบบธุรกิจพาณิชย์อิเล็กทรอนิกส์ อาทิเช่น ระบบแนะนำหนังสือ[1] [2] , ระบบแนะนำภาพยนตร์[3] , ระบบแนะนำเพลง[4] [5] และระบบแนะนำวีดีโอ[6] เป็นต้น โดยวิธีการที่นิยมใช้ในการสร้างรายการแนะนำข้อมูลสามารถแบ่งได้ 3 วิธีการ (Thorat, Goudar and Barve, 2015) ได้แก่ วิธีการกรองเนื้อหา (Content-based filtering), วิธีการกรองข้อมูลร่วม (Collaborative filtering) และ วิธีการแบบผสมผสาน (Hybrid filtering) แสดงดังภาพที่ 1

---

[1] www.goodreads.com

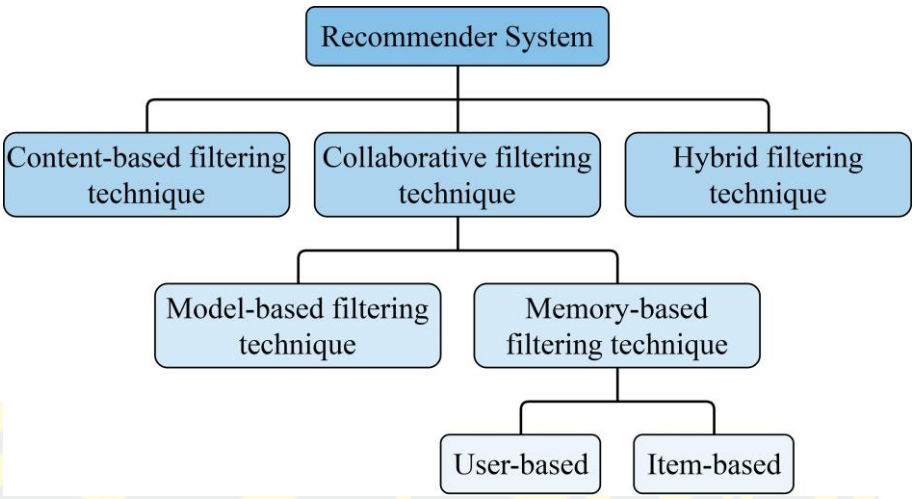
[2] www.amazon.com

[3] www.netflix.com

[4] www.lastfm.com

[5] www.pandora.com

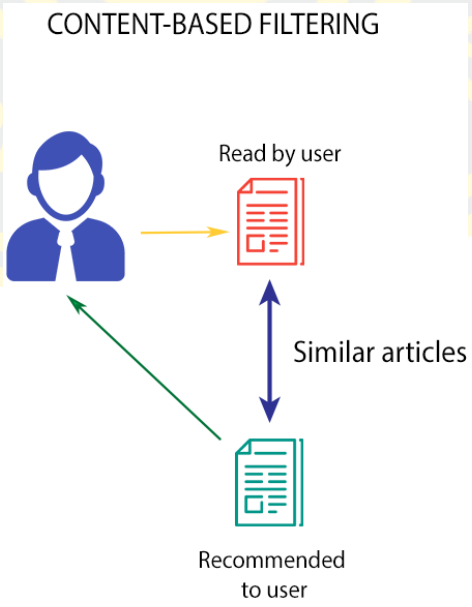
[6] www.youtube.com



ภาพที่ 1 ระบบแนะนำประกอบไปด้วยขั้นตอนวิธีต่าง ๆ

2.1.1 วิธีการกรองเนื้อหา (Content-based filtering)

วิธีการกรองเนื้อหาเป็นวิธีการสร้างรายการแนะนำข้อมูล โดยพิจารณาจากความคล้ายคลึงของเนื้อหาหรือคุณสมบัติต่าง ๆ ของข้อมูลที่คล้ายกันกับข้อมูลของผู้ใช้ในอดีต (Lops *et al.*, 2019) การทำงานของวิธีการนี้จะให้ความสนใจกับเนื้อหาของข้อมูลเป็นสำคัญ เช่น การสร้างรายการแนะนำหนังสือในวิธีการนี้จะทำการพิจารณาเนื้อหาของหนังสือ, ชื่อหนังสือ, ชื่อผู้แต่ง, หมวดหมู่ของหนังสือ และสำนักพิมพ์ที่มีความคล้ายกันกับหนังสือที่ผู้ใช้เคยอ่านในอดีต เป็นต้น แสดงดังภาพที่ 2

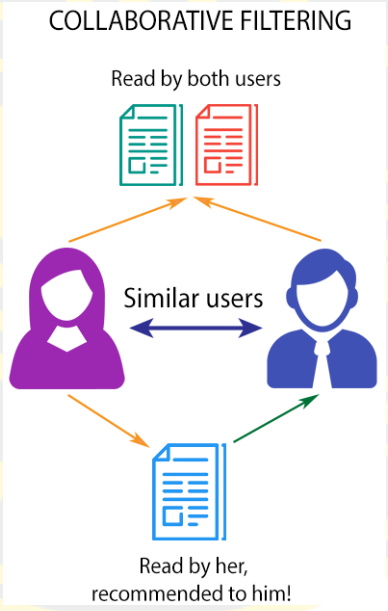


ภาพที่ 2 ขั้นตอนวิธีการกรองเนื้อหา

ข้อดีของวิธีการกรองเนื้อหา คือ การให้ความสำคัญกับเนื้อหาของข้อมูลเป็นสำคัญ เพื่อค้นหาข้อมูลที่ผู้ใช้สนใจ ซึ่งวิธีการนี้จะไม่ประสบกับปัญหาการให้คะแนนความชอบ (Rating) กับข้อมูลที่ไม่ทั่วถึง และปัญหาข้อมูลที่ยังไม่ได้ให้คะแนนความชอบ (Rating) จึงสามารถแนะนำข้อมูลที่ตรงกับความชื่นชอบของผู้ใช้ได้ อย่างไรก็ตามข้อเสียของวิธีการนี้คือข้อมูลที่แนะนำให้กับผู้ใช้มีความไม่หลากหลาย (Overspecialization)

2.1.2 วิธีการกรองข้อมูลร่วม (Collaborative filtering)

วิธีการกรองข้อมูลร่วมเป็นวิธีการสร้างรายการแนะนำข้อมูลที่นิยมมากที่สุด โดยการพิจารณาความคล้ายคลึงของข้อมูลหรือผู้ใช้ที่มีพฤติกรรมในการเลือกซื้อหรือบริโภคข้อมูลที่คล้ายกันจากคะแนนความชอบที่ผู้ใช้ให้กับข้อมูลโดยตรง (Rating) หรือประวัติในการเลือกซื้อหรือรับชมข้อมูลนั้น ๆ แสดงดังภาพที่ 3



ภาพที่ 3 ขั้นตอนวิธีการกรองข้อมูลร่วม

โดยวิธีการกรองข้อมูลร่วมแบ่งออกเป็น 2 ประเภท (Hameed, Al Jadaan and Ramachandram, 2012) คือ

1. วิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based filtering) เป็นวิธีการนำข้อมูลของผู้ใช้ทั้งหมดที่มีในฐานข้อมูล มาสร้างเป็นแบบจำลองสำหรับสร้างรายการแนะนำให้กับผู้ใช้ที่มีลักษณะคล้ายคลึงกัน

2. วิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based filtering) เป็นวิธีการนำข้อมูลของผู้ใช้แต่ละคนมาเปรียบเทียบว่ามีความคล้ายคลึงกับข้อมูลผู้ใช้คนอื่น ๆ ที่มีอยู่ในฐานข้อมูลหรือไม่ ซึ่งถ้าพิจารณาจากความคล้ายคลึงหรือใกล้เคียงกันระหว่างผู้ใช้ในฐานข้อมูล จะเรียกว่าวิธีการกรองข้อมูลร่วมโดยพิจารณาผู้ใช้เป็นหลัก (User-based Collaborative filtering) แต่ถ้าพิจารณาความคล้ายคลึงหรือใกล้เคียงกันระหว่างข้อมูลในฐานข้อมูล จะเรียกว่าวิธีการกรองข้อมูลร่วมโดยพิจารณาข้อมูลเป็นหลัก (Item-based Collaborative filtering)

ข้อดีของวิธีการกรองข้อมูลร่วม คือสามารถแนะนำข้อมูลให้กับผู้ใช้ได้หลากหลายมากขึ้น แต่อย่างไรก็ตามข้อเสียของวิธีการนี้คือ 1) เมื่อเกิดเหตุการณ์ที่มีผู้ใช้ใหม่หรือข้อมูลใหม่ ผู้ใช้ใหม่ที่เข้าสู่ระบบครั้งแรกจะยังไม่ได้รับการแนะนำและข้อมูลใหม่ก็จะไม่ถูกแนะนำเช่นกัน ปัญหานี้เรียกว่า “Cold start problem” 2) เมื่อเกิดเหตุการณ์ที่ข้อมูลในฐานข้อมูลได้รับการให้คะแนนความชอบจากผู้ใช้ (Rating) มีจำนวนไม่มากพอ ทำให้เกิดความเบาบางของข้อมูลขึ้น ปัญหานี้เรียกว่า “Data sparsity problem”

2.1.3 วิธีการแบบผสมผสาน (Hybrid filtering)

วิธีการแบบผสมผสานเป็นการรวมข้อดีของทั้งวิธีการกรองเนื้อหาและวิธีการกรองข้อมูลร่วมไว้ด้วยกัน ซึ่งสามารถช่วยในการหลีกเลี่ยงข้อจำกัดของทั้งสองวิธีการได้ (Çano and Morisio, 2017) ทำให้สามารถสร้างรายการแนะนำสินค้าที่คาดว่าผู้ใช้ชื่นชอบได้อย่างมีประสิทธิภาพ อย่างไรก็ตามวิธีการนี้ต้องใช้ข้อมูลของสินค้าจำนวนมากในการสร้างรายการแนะนำ อาจทำให้มีการใช้ทรัพยากรและมีความซับซ้อนในการแนะนำที่สูงขึ้น ซึ่งอาจไม่เหมาะสมกับระบบธุรกิจพาณิชย์อิเล็กทรอนิกส์ที่ไม่มีการจัดเก็บคุณสมบัติของข้อมูล

2.2 การประมวลผลข้อความเบื้องต้น (Text preprocessing)

การประมวลผลข้อความเป็นการแปลงรูปแบบข้อความให้อยู่ในรูปแบบที่ถูกต้อง หรือมีความพร้อมใช้งานในการประมวลผลเชิงลึกมากยิ่งขึ้น อาทิเช่น การสกัดคำสำคัญในเอกสาร, การสรุปใจความสำคัญในเอกสาร, การเปรียบเทียบความเหมือนกันระหว่างเอกสาร เป็นต้น โดยในงานวิจัยนี้ได้ประยุกย์ใช้วิธีการประมวลผลข้อความเบื้องต้นได้แก่ การตัดประโยค, การตัดคำ, การกำจัดคำหยุด, การวิเคราะห์หน้าที่ของคำ และการลดรูป/หารูปเดิมของคำ

2.2.1 การตัดประโยค (Sentence tokenization)

การแยก/ตัดประโยคเป็นขั้นตอนการแยกข้อความออกจากกัน ให้ประโยคอยู่ในรูปแบบประโยคความเดียว จะทำการวิเคราะห์ตามหลักไวยากรณ์ของประโยค และความสมบูรณ์ของความหมาย เช่น ในภาษาอังกฤษเมื่อพบ “.” (Full stop) หรือ “;” (Semi colon) ก็ดำเนินการแยก/ตัดกลุ่มของคำที่ปรากฏก่อนหน้าเป็นหนึ่งประโยค หรือหนึ่งหัวข้อ ซึ่งทำให้ข้อความถูกแบ่งออกเป็นหลายหัวข้อ

2.2.2 การตัดคำ (Word tokenization)

การตัดคำเป็นขั้นตอนการแบ่งองค์ประกอบของประโยคออกมาเป็นหน่วยคำ ซึ่งจะทำการวิเคราะห์ตามหลักไวยากรณ์ของประโยค เช่น ในภาษาอังกฤษเมื่อพบ “ ” (White space) ก็ดำเนินการแยก/ตัดคำออกมาเป็นหน่วยคำหนึ่ง ๆ

2.2.3 การกำจัดคำหยุด (Stop word removal)

เป็นขั้นตอนการกำจัดคำที่ไม่มีนัยยะสำคัญออกจากประโยค โดยที่ความหมายของประโยคหรือข้อความไม่ได้เปลี่ยนแปลง เช่น “a”, “an”, “the” เป็นต้น คำหยุดมักจะปรากฏบ่อยในทุก ๆ ข้อความหรือเอกสาร ถือได้ว่าคำหยุดเป็นคุณลักษณะที่ไม่เกี่ยวข้องในการประมวลผลข้อความ นอกจากนี้การกำจัดคำหยุดยังช่วยลดขนาดพื้นที่และเวลาในการประมวลผลข้อความอีกด้วย

2.2.4 การวิเคราะห์หน้าที่ของคำ (Part of speech tagging)

การวิเคราะห์หน้าที่ของคำคือ การระบุ/กำกับหน้าที่ของคำที่ประกอบอยู่ในประโยคว่าคำนั้น ๆ มีหน้าที่ชนิดใดในทางไวยากรณ์ ซึ่งหน้าที่ของคำในภาษาอังกฤษมีทั้งหมด 8 ชนิด ดังนี้

1. คำนาม (Noun) คือ คำที่ใช้เรียกแทนชื่อคน, สัตว์, สิ่งของ, สถานที่
2. คำสรรพนาม (Pronoun) คือ คำที่ใช้แทนคำนาม
3. คำกริยา (Verb) คือ คำที่แสดงอาการหรือการกระทำในประโยค
4. คำคุณศัพท์ (Adjective) คือ คำที่ทำหน้าที่ขยายคำนาม โดยตำแหน่งจะอยู่หน้าคำนามเสมอ
5. คำกริยาวิเศษณ์ (Adverb) คือ คำที่มีหน้าที่ขยายกริยา ขยายคุณศัพท์ และขยายกริยาวิเศษณ์ด้วยกันเอง

6. คำบุพบท (Preposition) คือ คำที่ใช้บอกตำแหน่ง วันเวลา ทิศทาง สถานที่ หรือแสดงความสัมพันธ์ระหว่างคำ กลุ่มคำ หรือประโยค
7. คำสันธาน (Conjunction) คือ คำที่ใช้เชื่อมคำกับคำ กลุ่มคำกับกลุ่มคำ หรือประโยคกับประโยค
8. คำอุทาน (Interjection) คือ คำที่ใช้บ่งบอกอารมณ์ ความรู้สึกที่เกิดขึ้นในตอนนั้น

2.2.5 การลดรูปของคำและการหารูปเดิมของคำ (Word stemming and lemmatization)

เป็นเทคนิคในการทำให้คำหนึ่ง ๆ เป็นมาตรฐาน และอยู่รูปแบบรากศัพท์ โดยการลดรูปของคำ (Stemming) จะพิจารณาจากกลุ่มของตัวอักษรในส่วนหน้า (prefix) หรือส่วนท้าย (suffix) ของคำ เช่น คำว่า “working” จะถูกพิจารณาส่วนท้ายของคำคือ “-ing” และกำจัดออก ผลลัพธ์จากการลดรูปของคำจะได้คำว่า “work” เป็นต้น ในส่วนของการหารูปเดิมของคำหรือการหารากศัพท์ (Lemmatization) จะทำการแปรคำ/ผันคำโดยใช้วิธีการรวบรวมคำศัพท์ที่มีความหมายคล้ายกันหรือมีรากศัพท์เดียวกันไว้เป็นรายการคำศัพท์ แล้วใช้ในการเปรียบเทียบหารากศัพท์ เช่น คำว่า “saw” จะถูกแปรคำเป็นคำว่า “see” เป็นต้น

2.3 การสกัดคำสำคัญ (Keyword extraction)

คำสำคัญสามารถเป็นตัวแทนของเอกสารหรือบทความหนึ่ง ๆ โดยย่อเหลือเพียงคำที่แสดงใจความสำคัญที่กระทัดรัด และชัดเจน ที่บ่งบอกว่าเอกสารหรือบทความนั้น ๆ เกี่ยวข้องกับเรื่องใด ซึ่งการระบุคำสำคัญของเอกสารหรือบทความจะช่วยลดระยะเวลาในการอ่าน และช่วยผู้อ่านพิจารณาได้ว่าเอกสารหรือบทความนั้นตรงตามความต้องการของผู้อ่านหรือไม่ โดยวิธีการสกัดคำสำคัญสามารถแบ่งได้ 4 วิธีการ (Zhang, 2008; Beliga, 2014) ได้แก่

2.3.1 วิธีการทางสถิติ (Statistical approach)

วิธีการทางสถิติเป็นวิธีการอย่างง่ายที่สามารถใช้สถิติของข้อมูลคำในเอกสาร หรือบทความหนึ่ง ๆ ซึ่งอาจใช้การพิจารณาสถิติทางด้านความถี่หรือความน่าจะเป็น ในการระบุคำสำคัญของเอกสารต่าง ๆ ได้ โดยข้อดีที่สำคัญของวิธีการนี้คือการประยุกต์ใช้ได้อย่างอิสระ สามารถนำมาใช้งานได้หลากหหลายภาษา และเหมาะสมกับฐานข้อมูลที่มีขนาดใหญ่ ทฤษฎีที่นิยมนำมาใช้ อาทิเช่น

1. Word frequency เป็นค่าสถิติของการนับความถี่ของคำที่ปรากฏขึ้นในเอกสารหนึ่ง ๆ
2. TF-IDF เป็นค่าถ่วงน้ำหนักทางด้านสถิติที่ใช้ในการประเมินค่าความสำคัญของคำในเอกสาร

3. Word co-occurrences เป็นค่าทางด้านสถิติที่พิจารณาจากการเกิดขึ้นร่วมกันของคำสองคำหรือมากกว่าที่อยู่ติดกันในประโยคเดียวกันในเอกสาร

วิธีการทางสถิติเป็นวิธีที่ง่ายต่อการสกัดคำสำคัญในเอกสาร แต่อย่างไรก็ตามวิธีการนี้อาจจะตกหล่นคำสำคัญที่ปรากฏขึ้นในเอกสารไม่บ่อย (ค่าความถี่น้อย) จึงทำให้ระบุถึงคำสำคัญในเอกสารได้ไม่ครบถ้วนเท่าที่ควร

2.3.2 วิธีการทางภาษาศาสตร์ (Linguistic approach)

เป็นวิธีการที่ใช้องค์ความรู้ หรือคุณสมบัติทางด้านภาษาศาสตร์เข้ามาช่วยในการสกัดคำสำคัญ โดยการวิเคราะห์คำในประโยคหรือเอกสารต่าง ๆ โดยใช้หลักไวยกรณ์ของภาษา รวมถึงการวิเคราะห์คำศัพท์, วิเคราะห์ประโยค เป็นต้น ในส่วนของกฎที่ได้สร้างขึ้นนั้นจะถูกกำหนดไว้ตั้งแต่ก่อนเริ่มทำการสกัดคำและจะไม่มีการเปลี่ยนแปลงตามข้อมูลที่เข้ามาใหม่ ซึ่งวิธีการนี้อาจมีค่าความถูกต้องเป็นอย่างมาก แต่จะมีการคำนวณที่ละเอียดและจำเป็นจะต้องมีความเชี่ยวชาญในด้านภาษาศาสตร์ นอกจากนี้วิธีการทางภาษาศาสตร์สามารถนำไปใช้ร่วมกับวิธีการอื่น ๆ เพื่อเพิ่มประสิทธิภาพในการสกัดคำสำคัญมากยิ่งขึ้น

2.3.3 วิธีการทางการเรียนรู้ของเครื่องจักร (Machine learning approach)

เป็นวิธีการที่ใช้การเรียนรู้ของเครื่องจักรเข้ามาช่วย โดยเรียนรู้จากเอกสารตัวอย่างที่นำมาใช้ในการสอนและสร้างโมเดลสำหรับสกัดคำสำคัญจากเอกสารต่าง ๆ ซึ่งวิธีการทางการเรียนรู้ของเครื่องจักร สามารถแบ่งออกเป็น 2 ประเภท คือ

1. วิธีการเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นการเรียนรู้ที่สามารถจำแนกข้อมูลภายในชุดข้อมูลว่ามีผลลัพธ์ที่ผิดหรือถูกต้องได้ ที่ซึ่งในชุดข้อมูลจะประกอบไปด้วยข้อมูลฝึกสอนและข้อมูลเป้าหมาย สำหรับข้อมูลเหล่านี้จะถูกนำไปพยากรณ์ค่าข้อมูล โดยพื้นฐานการพยากรณ์จะพิจารณาจากข้อมูลทั้งหมดในชุดข้อมูล โดยเทคนิคที่ใช้วิธีการเรียนรู้แบบมีผู้สอน ได้แก่ Naïve Bayes, Support Vector Machine เป็นต้น
2. วิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นการเรียนรู้ที่ไม่มีการกำหนดข้อมูลเป้าหมาย จึงไม่มีการจำแนกข้อมูลว่ามีผลลัพธ์อย่างไร แต่เป็นการเรียนรู้ที่เกี่ยวข้องกับความสัมพันธ์ของข้อมูล จะนำไปใช้สำหรับจำแนกหรือแบ่งข้อมูลออกเป็นกลุ่ม ๆ

2.3.4 วิธีการแบบผสมผสาน (Hybrid approach)

วิธีการแบบผสมผสานเป็นวิธีการสกัดคำสำคัญโดยการนำวิธีการสกัดคำสำคัญต่าง ๆ ที่กล่าวไว้ข้างต้นมาดำเนินงานร่วมกัน เพื่อให้ได้มาซึ่งผลลัพธ์จากการสกัดคำสำคัญที่มีประสิทธิภาพมากยิ่งขึ้น นอกจากนี้ยังสามารถประยุกต์เทคนิคหรือข้อมูลของคำในการพิจารณาสกัดคำสำคัญ เช่น การนำคลังคำศัพท์ในศาสตร์ด้านใดด้านหนึ่งมาประยุกต์ใช้กับวิธีการต่าง ๆ ซึ่งจะใช้ประโยชน์ในด้านขององค์ความรู้เชิงลึกที่เกี่ยวข้องกับศาสตร์ด้านนั้น ๆ เป็นต้น

2.4 งานวิจัยที่เกี่ยวข้อง

ในส่วนนี้จะอธิบายถึงงานวิจัยที่เกี่ยวข้องทางด้านการแนะนำหนังสือ (Recommender system) โดยแบ่งการพิจารณาในการแนะนำออกเป็น 5 แง่มุม ได้แก่ การแนะนำหนังสือจากการพิจารณาเนื้อหาสาระในหนังสือ, การแนะนำหนังสือจากการให้คะแนนความนิยมของผู้อ่าน, การแนะนำหนังสือที่สืบเนื่องจากการแนะนำผู้เขียน, การแนะนำหนังสือจากประวัติการยืมหนังสือของผู้ใช้ห้องสมุด และการแนะนำหนังสือโดยพิจารณาจากระดับทักษะการอ่าน นอกจากนี้ยังอธิบายถึงงานวิจัยที่เกี่ยวข้องทางด้านการสกัดคำสำคัญ (Keyword extraction) ดังต่อไปนี้

2.4.1 งานวิจัยที่เกี่ยวข้องทางด้านการแนะนำหนังสือ (Recommender system)

2.4.1.1 การแนะนำหนังสือจากการพิจารณาเนื้อหาสาระในหนังสือ

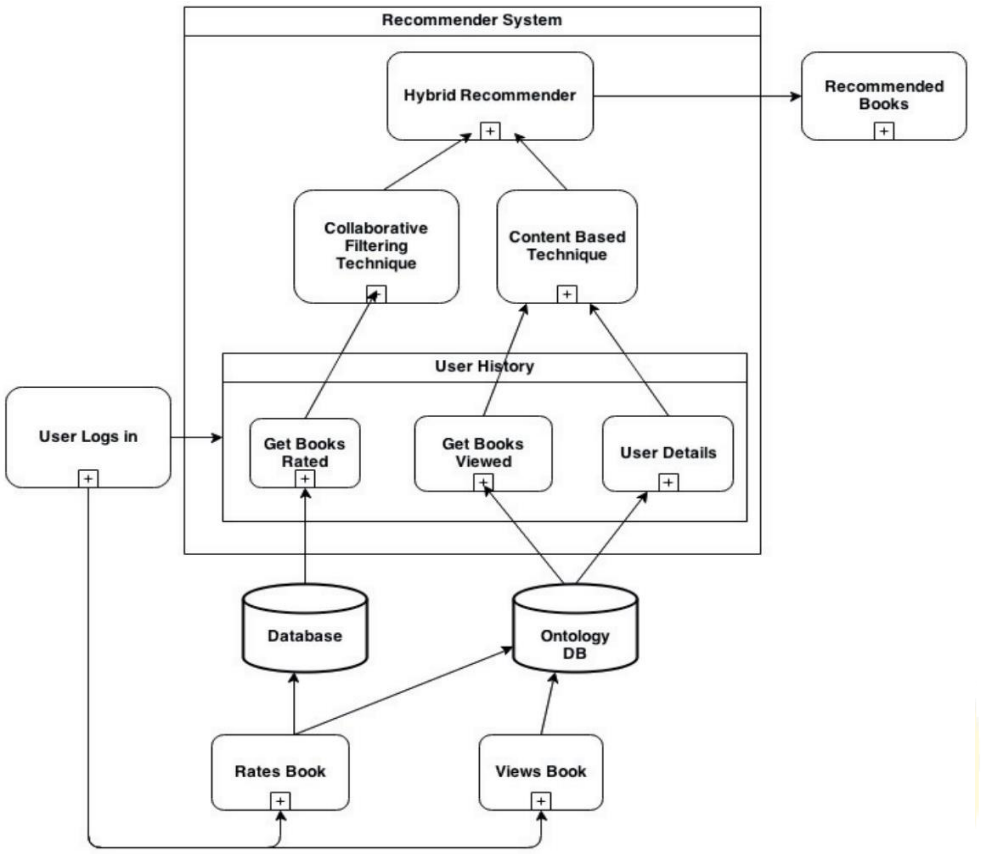
Garrido and Ilarri (2014) นำเสนอวิธีการสร้างรายการแนะนำหนังสือให้กับผู้อ่านแต่ละบุคคล จากสมมติฐานที่ว่าคำอธิบายหนังสือและบทความวิจารณ์ของผู้อ่านคนอื่นของหนังสือสามารถช่วยให้ผู้อ่านได้รับรายการหนังสือที่ตรงกับความชื่นชอบของผู้อ่าน จากแนวความคิดข้างต้น จึงได้นำเสนอวิธีการสร้างรายการแนะนำหนังสือที่เรียกว่า “Topic Map Recommender (TMR)” โดยการพิจารณาความสัมพันธ์ของหนังสือ (Topic map) ซึ่งสกัดจากคำอธิบายหนังสือโดยการใช้เครื่องมือที่เรียกว่า “TM-Gen” ที่จะพิจารณาคำสำคัญของคำอธิบายหนังสือแล้วทำการวิเคราะห์ความสัมพันธ์ของข้อมูล ร่วมกับการพิจารณาบทความวิจารณ์ของผู้อ่านคนอื่น โดยใช้วิธีการให้ค่าน้ำหนักของคำสำคัญ (TF-IDF) ในการวิเคราะห์บทความวิจารณ์ จากนั้นรายการแนะนำหนังสือของผู้อ่านจะถูกพิจารณาจากความคล้ายของความสัมพันธ์ของหนังสือทั้งในเชิงบวกและเชิงลบที่คาดว่าผู้อ่านน่าจะชื่นชอบ นอกเหนือจากระบบ TMR แล้วนั้น Garrido, Pera and Ilarri (2014) นำเสนอการสร้างระบบนำหนังสือที่มีชื่อว่า “SOLE-R” ที่ทำการพิจารณาความหมาย (Semantic) ออนโทโลยี (Ontology) และคุณลักษณะทางด้านภาษา (Linguistic) เพื่อให้สามารถแนะนำหนังสือได้อย่างมี

ประสิทธิภาพ นอกจากนั้นยังทำการพิจารณาความสัมพันธ์ของหนังสือ (Topic map) เพื่อทำการแนะนำหนังสือให้กับผู้อ่านหนังสือคนหนึ่ง ๆ

2.4.1.2 การแนะนำหนังสือจากการให้คะแนนความนิยมของผู้อ่าน

Linden, Smith and York (2003) นำเสนอการแนะนำหนังสือที่คาดว่าจะถูกซื้อร่วมกันของผู้ใช้เว็บไซต์ Amazon โดยงานวิจัยนี้ได้เสนอการแนะนำหนังสือด้วยวิธี “Item-to-Item Collaborative Filtering” ที่จะพิจารณาข้อมูลหนังสือที่ถูกซื้อ หรือถูกให้คะแนนความนิยมจากผู้ใช้คนหนึ่ง ๆ โดยพิจารณาร่วมกับข้อมูลหนังสือที่ได้รับคะแนนความนิยมคล้าย ๆ กัน (ข้อมูลหนังสือที่พิจารณาอาจประกอบไปด้วย ชื่อผู้เขียน คำสำคัญ ประเภทหรือหมวดหมู่ของหนังสือ และอื่น ๆ) ซึ่งจากการพิจารณาดังกล่าวจะทำให้สามารถสร้างตารางหรือเมทริกซ์ ที่แสดงถึงความเหมือนและความเกี่ยวเนื่องกันของหนังสือ ที่ซึ่งจะช่วยให้สามารถค้นหารายการหนังสือที่คาดว่าน่าจะถูกซื้อร่วมกันได้

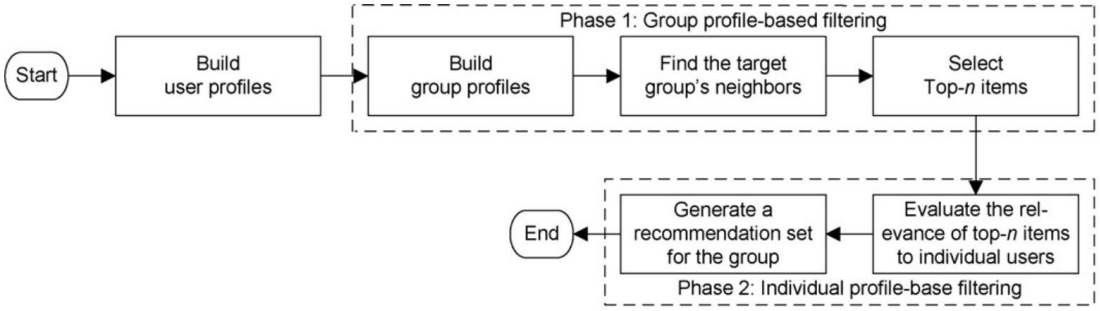
Chandak, Girase and Mukhopadhyay (2015) นำเสนอการแนะนำหนังสือให้แก่ผู้อ่านด้วยการพิจารณาข้อมูลอินพุต 3 ส่วนหลัก คือ 1) ข้อมูลผู้อ่าน (เช่น อายุ เพศ และอื่น ๆ) 2) ข้อมูลหนังสือ (เช่น คำสำคัญ ประเภทหนังสือ และอื่น ๆ) และ 3) คะแนนความนิยมของผู้อ่าน ตามลำดับ โดยในงานวิจัยนี้ได้นำเสนอระบบการแนะนำหนังสือแบบผสมผสาน (Hybrid book recommendation system) ที่จะผสมผสานการทำงานระหว่างขั้นตอนวิธี “Collaborative Filtering” และ “Content-based recommendation” รวมถึงการพิจารณาข้อมูลผู้อ่านเป็นส่วนประกอบหนึ่งของการแนะนำหนังสือ (ดังแสดงแผนผังการทำงานในภาพที่ 4) ที่จะสามารถเพิ่มมิติการพิจารณาและลดทอนข้อเสียของการประยุกต์ใช้วิธีการ “Collaborative Filtering” หรือ “Content-based recommendation” เพียงอย่างเดียว โดยจากอินพุตทั้ง 3 ส่วนข้างต้นจะถูกแยกการประมวลผลออกเป็นสองส่วนหลักและแยกเก็บไว้ในฐานข้อมูล และฐานข้อมูลออนโทโลยี เพื่อส่งต่อให้กับการคำนวณแบบ “Collaborative Filtering” และ “Content-based recommendation” โดยจากการทดลองเพื่อทดสอบประสิทธิภาพจะทำให้ทราบได้ว่าวิธีการดังกล่าวสามารถแนะนำหนังสือได้อย่างมีประสิทธิภาพ และลดปัญหาที่เกิดขึ้นจากการประยุกต์ใช้วิธีการคำนวณหนึ่ง ๆ ได้



ภาพที่ 4 ระบบการแนะนำหนังสือแบบไฮบริดรวมถึงการพิจารณาข้อมูลผู้อ่าน

2.4.1.3 การแนะนำหนังสือที่สืบเนื่องจากการแนะนำผู้เขียน

Kim *et al.* (2010) นำเสนอวิธีการแนะนำหนังสือให้กับกลุ่มผู้อ่านที่มีลักษณะความชอบหรือพฤติกรรมในการอ่านหนังสือที่คล้ายกัน จากสมมติฐานที่ว่าผู้อ่านหนังสือที่มีความชอบในการเลือกอ่านหนังสือที่คล้ายกันจะมีการแบ่งปันข้อมูลหนังสือที่มีความน่าสนใจกับผู้อ่านคนอื่น ๆ ภายในกลุ่ม ส่งผลให้ผู้อ่านในกลุ่มเดียวกันสามารถได้รับการแนะนำหนังสือที่ชื่นชอบโดยการพิจารณาการเลือกอ่านหนังสือของผู้อ่านคนอื่นภายในกลุ่ม ดังนั้น Kim *et al.* จึงได้นำเสนอวิธีการสร้างรายการแนะนำหนังสือด้วย 2 ขั้นตอนวิธีการแสดงดังภาพที่ 5 โดยขั้นตอนที่ 1) พิจารณาความชอบของกลุ่มผู้อ่านที่มีลักษณะความชอบในการอ่านหนังสือที่คล้ายกัน ด้วยวิธีการกรองข้อมูลร่วม (Collaborative Filtering) จากขั้นตอนนี้รายการหนังสือของกลุ่มผู้อ่านที่มีลักษณะความชอบคล้ายกันจะถูกแนะนำร่วมกับรายการหนังสือที่ถูกเลือกซื้อมากภายในกลุ่มผู้อ่าน และขั้นตอนที่ 2) วิธีการที่นำเสนอนี้ได้มีการพิจารณาความชอบของการอ่านหนังสือของผู้อ่านแต่ละบุคคล โดยหนังสือที่ผู้อ่านไม่ชื่นชอบจะไม่ถูกแนะนำ โดยการพิจารณาคีย์เวิร์ด (keyword) ของหนังสือจากประวัติการอ่านหนังสือของผู้อ่านแต่ละบุคคล

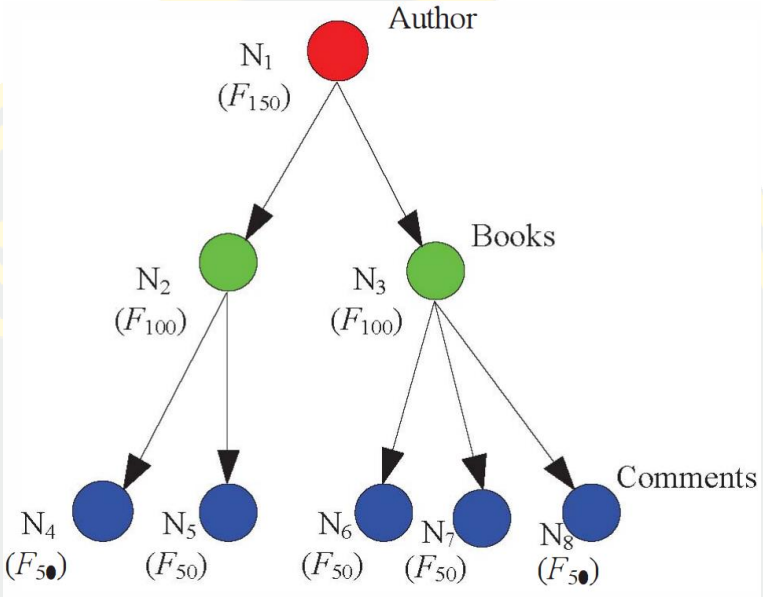


ภาพที่ 5 ขั้นตอนการแนะนำแบบ top-n items

Vaz, Martins de Matos and Martins (2012) นำเสนอวิธีการสร้างรายการแนะนำหนังสือให้กับผู้อ่าน เนื่องจากหนังสือภายในระบบมีจำนวนมากและมีความหลากหลายของเนื้อหาที่แตกต่างกัน ส่งผลให้ผู้อ่านมีความยุ่งยากในการเข้าถึงหนังสือที่ผู้อ่านชื่นชอบ ดังนั้น Vaz *et al.* ได้นำเสนอวิธีการสร้างรายการแนะนำหนังสือที่คาดว่าผู้อ่านชื่นชอบ จากสมมติฐานที่ว่าผู้อ่านน่าจะมีความชอบในการเลือกอ่านหนังสือจากสไตล์การเขียนหนังสือของนักเขียนที่ผู้อ่านเคยอ่านในอดีต วิธีการที่นำเสนอนี้จึงได้สร้างรายการหนังสือโดยการรวมค่าการทำนายคะแนนความชอบจาก 2 ขั้นตอนวิธีการ คือ 1) วิธีการกรองข้อมูลร่วมโดยพิจารณาความชอบของหนังสือเป็นหลัก (Item-based Collaborative Filtering) ซึ่งพิจารณาจากประวัติของผู้อ่านที่ให้คะแนนความชอบกับหนังสือในระบบ และ 2) วิธีการกรองเนื้อหา (Content-based filtering) ซึ่งพิจารณาสไตล์การเขียนหนังสือจากคำศัพท์ที่ใช้ในการเขียน, ขนาดของเนื้อหา, คำศัพท์ที่ใช้ร่วมกัน และคำศัพท์ที่ใช้บ่อยร่วมกับค่าคะแนนความชอบของหนังสือในเชิงบวกและเชิงลบจากคะแนนความชอบของหนังสือ โดยทั้ง 2 วิธีการจะถูกรวมเป็นค่าการทำนายความชอบของหนังสือโดยการใช้ค่าเฉลี่ยเพื่อสร้างรายการแนะนำหนังสือต่อไป

Lu and Zhang (2015) นำเสนอการแนะนำผู้เขียนหนังสือให้แก่ผู้อ่านหนังสือที่ซึ่งจะทำให้ผู้อ่านสามารถค้นหาหนังสือของผู้เขียนที่ถูกแนะนำได้ โดยงานวิจัยนี้มีการตั้งสมมติฐานที่ว่า “บุคคลหนึ่ง ๆ ที่ชอบหรือรักในการอ่านหนังสือมักจะชื่นชอบและติดตามหนังสือของผู้เขียนที่ตนเองชื่นชอบ บุคคลเหล่านี้มักจะมีความยินดีที่จะอ่านหนังสือจากผู้เขียนอื่น ๆ ที่มีแนวทางการเขียนหนังสือที่คล้ายคลึงกับนักเขียนที่พวกเขาชอบ” ซึ่งจากสมมติฐานดังกล่าว Lu and Zhang ได้เสนอการแนะนำผู้เขียนให้แก่ผู้อ่านหนังสือพร้อมกับหนังสือที่ถูกเขียนโดยผู้เขียนที่ถูกแนะนำ และข้อคิดเห็นของผู้อ่านอื่น ๆ ที่เคยอ่านหนังสือของผู้เขียนที่ถูกแนะนำ ด้วยการประยุกต์ใช้โครงสร้างข้อมูลแบบต้นไม้ 3 ระดับ (ดังแสดงตัวอย่างในภาพที่ 6 โดยระดับแรกใช้สำหรับจัดเก็บข้อมูลที่เกี่ยวเนื่องกับผู้เขียน ระดับที่สองใช้สำหรับจัดเก็บข้อมูลหนังสือที่ถูกเขียนโดยผู้เขียนนั้น ๆ และ ระดับที่ 3 ใช้สำหรับจัดเก็บข้อมูล

ข้อคิดเห็น (50 ข้อคิดเห็นแรก) จากผู้ที่ทำการอ่านหนังสือเล่มนั้น ๆ ของผู้เขียน ตามลำดับ) และทำการประยุกต์ใช้เทคนิค Multi-Layer Self-Organizing Map (ML-SOM) เพื่อทำการจัดกลุ่มความคล้ายคลึงกันระหว่างผู้เขียนหนังสือที่ซึ่งจะสามารถทำให้ทราบถึงผู้เขียนที่มีสไตล์ที่คล้ายคลึงกัน อันนำมาซึ่งการแนะนำหนังสือจากผู้เขียนที่มีความคล้ายคลึงกันได้



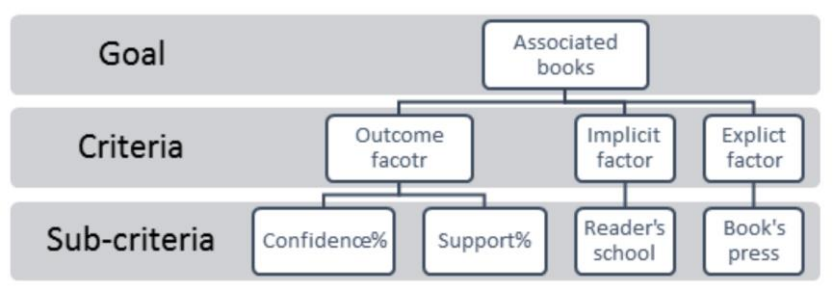
ภาพที่ 6 โครงสร้างข้อมูลต้นไม้ 3 ระดับ

2.4.1.4 การแนะนำหนังสือจากประวัติการยืมหนังสือของผู้ใช้ห้องสมุด

Saxena, Singh and Khan (2014) เสนอการแนะนำหนังสือให้แก่ผู้ใช้ห้องสมุดโดยตั้งสมมติฐานที่ว่าในการค้นหนังสือเล่มหนึ่ง ๆ ที่เกี่ยวกับเรื่องหนึ่ง ๆ ที่ผู้ใช้ห้องสมุดสนใจจะมีหนังสือหลายเล่มที่เกี่ยวข้องกับเรื่องนั้น ๆ โดยจากการที่มีหนังสือหลายเล่มทำให้ผู้ใช้ห้องสมุดไม่สามารถตัดสินใจได้ว่าจะเลือกยืมหนังสือเพื่ออ่านเพิ่มเติมเนื่องจากไม่ทราบข้อมูลเบื้องต้นว่าหนังสือแต่ละเล่มดีหรือไม่ดีอย่างไร ดังนั้น Saxena *et al.* ได้เสนอการประยุกต์ใช้แนวคิดของการค้นหารูปแบบปรากฏบ่อย (Frequent itemsets mining) เพื่อทำการค้นหาค่าความถี่ของจำนวนครั้งที่หนังสือเล่มหนึ่ง ๆ ถูกค้นหา หรือถูกยืมจากผู้ใช้ห้องสมุดทั้งหมด โดยผลลัพธ์ที่สามารถค้นคืนได้จะประกอบไปด้วยชื่อผู้แต่งหนังสือ ชื่อหนังสือ สำนักพิมพ์ และจำนวนครั้งที่หนังสือเล่มนั้น ๆ ถูกค้นหาหรือถูกยืมจากผู้ใช้ห้องสมุดทั้งหมด ซึ่งจากผลลัพธ์ดังกล่าวจะทำให้ผู้ใช้ห้องสมุดมีข้อมูลเบื้องต้นเกี่ยวกับความนิยมหนังสือ อันนำมาซึ่งการตัดสินใจในการเลือกยืมหนังสือที่ดีขึ้นได้

Teng *et al.* (2015) เสนอการแนะนำหนังสือให้แก่ผู้ใช้ห้องสมุดจากประวัติการยืมหนังสือในอดีตของผู้ใช้ห้องสมุดด้วยการประยุกต์ใช้แนวคิด Fuzzy Analytical Hierarchy Process (FAHP)

ร่วมกับอัลกอริทึม Apriori โดยขั้นตอนเริ่มแรก FAHP จะทำการแบ่งปัญหาออกเป็นปัญหาย่อย ๆ โดยแต่ละปัญหาย่อยสามารถถูกวิเคราะห์ได้โดยเป็นอิสระต่อกัน (ดังแสดงลำดับชั้นของการแบ่งปัญหาย่อยในภาพที่ 7) ซึ่งจากแนวคิดข้างต้น จะทำให้สามารถแบ่งส่วนของปัญหาออกเป็น 3 ส่วนย่อย คือ 1) Outcome factor จะเกี่ยวข้องกับการค้นหาค่าความถี่และค่าความเชื่อมั่นของหนังสือที่ถูกยืมร่วมกัน 2) Implicit factor จะเกี่ยวเนื่องกับข้อมูลพื้นฐานของผู้ใช้ห้องสมุด (อาทิเช่น ชื่อโรงเรียนของผู้ใช้ห้องสมุด โดยนักเรียนในโรงเรียนเดียวกันน่าจะมีความสนใจเหมือนกัน) และ 3) Explicit factor จะเกี่ยวเนื่องกับข้อมูลพื้นฐานของหนังสือ (อาทิเช่น สำนักพิมพ์ของหนังสือ) ตามลำดับ โดยจากการรวมกันของส่วนย่อยทั้งหมดและการกำหนดค่าความสำคัญของแต่ละส่วนย่อยโดยผู้ที่ต้องการวิเคราะห์ข้อมูลจะทำให้เราสามารถทราบถึงหนังสือที่มีความเกี่ยวเนื่องกันได้



ภาพที่ 7 การแบ่งปัญหาออกเป็นปัญหาย่อย

Chen, Kuo and Liao (2015) นำเสนอวิธีการแนะนำหนังสือให้กับผู้อ่านในห้องสมุดอิเล็กทรอนิกส์ ด้วยสมมติฐานที่ว่าผู้อ่านน่าจะมีความชอบในการเลือกอ่านหนังสือกับผู้อ่านคนอื่นที่มีพฤติกรรมการเลือกอ่านหนังสือที่คล้ายกัน แต่เนื่องจากหนังสือมีคุณลักษณะและคุณสมบัติที่หลากหลาย ทำให้การคำนวณค่าความคล้ายของลักษณะการอ่านหนังสือของผู้อ่านในระบบนั้นใช้เวลาในการคำนวณที่นาน เนื่องจากในระบบมีหนังสือจำนวนมากร่วมกับผู้อ่านมีการให้คะแนนกับหนังสือจำนวนน้อย ดังนั้น Chen *et al.* จึงได้นำเสนอการสร้างรายการแนะนำหนังสือ โดยใช้วิธีการกรองข้อมูลร่วมซึ่งพิจารณาความชอบในการอ่านหนังสือเพื่อแนะนำหนังสือให้กับผู้อ่านจากประวัติการยืมหนังสือของผู้อ่านที่มีพฤติกรรมการเลือกอ่านหนังสือที่คล้ายกัน ร่วมกับพิจารณาความสัมพันธ์ของคุณสมบัติของหนังสือของผู้อ่านโดยใช้วิธีการออนโทโลยี (Ontology) เพื่อสร้างรายการแนะนำหนังสือ โดยวิธีการที่นำเสนอมีการประมวลผลอยู่บนสถาปัตยกรรม Map-reduce โมเดลในเฟรมเวิร์คของ Hadoop ซึ่งสามารถใช้เวลาในการคำนวณที่รวดเร็วกว่าวิธีการแบบเดิม จากวิธีการที่นำเสนอนี้สามารถสร้างรายการแนะนำหนังสือได้อย่างมีประสิทธิภาพในด้านการประมวลผลข้อมูลและความถูกต้องของรายการแนะนำหนังสือ

2.4.1.5 การแนะนำหนังสือโดยพิจารณาจากระดับทักษะการอ่าน

Pera and Ng (2014) นำเสนอการสร้างรายการแนะนำหนังสือให้กับนักเรียน K-12 โดยการกำหนดเนื้อหาที่น่าสนใจของนักเรียนแต่ละกลุ่มจากครูผู้สอนผ่านข้อมูล 4 ข้อมูลคือ 1) กลุ่มของนักเรียน (range of grade levels) 2) เนื้อหาของหนังสือ (a brief content description) 3) หัวข้อที่สนใจ (a topical description) และ 4) องค์ประกอบอื่น ๆ เกี่ยวกับหนังสือ โดยข้อมูลทั้ง 4 ข้อมูลจะถูกกำหนดโดยครูผู้สอนนักเรียน K-12 จากนั้นระบบแนะนำหนังสือที่มีชื่อว่า “*BReT*” (an unsupervised *B*ook *Re*commender for K-12 *T*eachers) จะเปรียบเทียบข้อมูลที่ครูผู้สอนต้องการกับหนังสือด้วยขั้นตอนวิธี “Content-based recommendation” และสร้างรายการแนะนำให้กับนักเรียนตามข้อมูลที่ครูผู้สอนได้ระบุไว้ เพื่อช่วยครูผู้สอนในการค้นหาหนังสือที่ตรงตามองค์ความรู้ของนักเรียน

Ng (2016) ได้เสนอระบบ CBRec สำหรับแนะนำหนังสือให้กับเด็กที่ทำการศึกษาในระบบ K-12 ของสหรัฐอเมริกา โดยตั้งสมมติฐานที่ว่าเด็กยังไม่มีความรู้พื้นฐานในการเลือกหนังสือที่ชอบหรือที่สนใจ ซึ่งจากสมมติฐานดังกล่าวจะทำให้เด็กใช้เวลาเลือกค่อนข้างมากและอาจต้องพึ่งพาคำแนะนำของผู้ปกครอง หรือครู/อาจารย์ที่อาจมีความชอบไม่เหมือนกัน ด้วยเหตุนี้ ระบบ CBRec จึงถูกพัฒนาขึ้นเพื่อแก้ไขปัญหาข้างต้น CBRec จะประกอบด้วย 2 ขั้นตอนการทำงานหลัก คือ 1) การวิเคราะห์ระดับทักษะการอ่านของเด็ก ด้วยการวิเคราะห์ความยากง่ายของหนังสือที่เด็ก ๆ เคยอ่านไว้ และ 2) การระบุถึงหนังสือต่าง ๆ ที่ถูกจัดเก็บไว้ในเว็บไซต์ โดยหนังสือที่ถูกระบุจะมีความสอดคล้องกับระดับทักษะการอ่านของเด็กด้วยการผสมผสานระหว่างเทคนิค “User-based collaborative filtering” และ “Content-based filtering” ที่จะช่วยคัดกรองหนังสือที่จะแสดงให้เด็กแต่ละคนเห็นด้วยการพิจารณาจากการให้คะแนนความนิยมของเด็ก ๆ กับหนังสือ รวมถึงการวิเคราะห์คำอธิบาย (หรือเนื้อหา) ของหนังสือที่เด็กคนหนึ่ง ๆ เคยให้คะแนนความนิยมไว้ในอดีต ที่ซึ่งจะสามารถช่วยสร้างโปรไฟล์และสามารถทราบถึงความชื่นชอบของเด็กคนนั้น ๆ ได้ ซึ่งจากการดำเนินการดังกล่าวจะสามารถช่วยเพิ่มประสิทธิภาพของการแนะนำหนังสือได้

2.4.2 งานวิจัยที่เกี่ยวข้องทางด้านการสกัดคำสำคัญ (Keyword extraction)

Frantzi, Ananiadou and Mima (2000) นำเสนอวิธีการ C-value และ NC-value สำหรับสกัดคำสำคัญแบบอัตโนมัติ โดยวิธีการทั้งสองนี้ประยุกต์ใช้วิธีการทางภาษาศาสตร์ (Linguistic approach) และวิธีการทางสถิติ (Statistical approach) ในการสกัดกลุ่มคำสำคัญ ดังนี้ 1) วิธีการ C-value เป็นขั้นตอนสร้างรายการแคนดิเดท (Candidate) ของกลุ่มคำที่ถูกสกัดออกมาจากประโยค

โดยจะพิจารณาด้วยวิธีการทางภาษาศาสตร์ ซึ่งกลุ่มคำที่สนใจคือนามวลี (Noun phrase) จากนั้นจะถูกระบุเป็นคำสำคัญภายในรายการแคนดิเดท และนับความถี่ของคำสำคัญภายในรายการแคนดิเดท โดยใช้วิธีการทางสถิติ จากนั้นนำผู้เชี่ยวชาญเข้ามาประเมินผลลัพธ์ในรายการแคนดิเดทของกลุ่มคำสำคัญที่ถูกสกัด เมื่อผู้เชี่ยวชาญได้ระบุถึงคำสำคัญที่ถูกต้องแล้ว จะนำไปสู่วิธีการที่ 2) วิธีการ NC-value เป็นการพิจารณาคำอื่น ๆ ที่เกิดขึ้นร่วมกับคำสำคัญที่ผู้เชี่ยวชาญกำหนด รวมถึงความถี่ในการเกิดขึ้นร่วมกัน และระบุคำที่เกิดขึ้นร่วมกับคำสำคัญ เป็นกลุ่มคำสำคัญที่สามารถสกัดได้ซึ่งผลลัพธ์ในการสกัดกลุ่มคำสำคัญด้วยวิธีการ NC-value มีประสิทธิภาพมากกว่า C-value และ Frantzi *et al.* ยังทำการสร้างซอฟต์แวร์ชื่อว่า “*Termine*” จากวิธีการ C-value ดังกล่าวข้างต้น เพื่อให้บุคคลทั่วไปสามารถใช้ในการสกัดคำสำคัญได้อีกด้วย

Hulth (2003) ได้เสนอการสกัดคำสำคัญโดยอัตโนมัติจากบทคัดย่อภาษาอังกฤษ และนำวิธีการทางภาษาศาสตร์ (Linguistic approach) มาใช้ร่วมด้วยในการสกัดคำสำคัญ แทนที่จะใช้วิธีการทางสถิติเพียงอย่างเดียว โดยในการสกัดคำสำคัญแบ่งเป็น 3 วิธีการ ได้แก่ 1) วิธีการ *N*-gram โดยจะกรองข้อความที่ไม่ใช่ตัวอักษรและข้อความที่เป็นตัวเลขออกไป แต่จะสนใจเก็บคำนามเฉพาะ เช่น ชื่อผู้คน, ชื่อเรื่อง เป็นต้น 2) วิธีการ NP-chunks จะสกัดคำสำคัญโดยสนใจเฉพาะนามวลี (Noun phrase) 3) POS Tag Patterns จะสกัดคำสำคัญ โดยพิจารณาหน้าที่ของคำในประโยค ซึ่งจะสนใจเฉพาะคำนามที่เกิดขึ้นร่วมกัน และคำนามที่เกิดขึ้นร่วมกับคำคุณศัพท์เท่านั้น และผลลัพธ์ในการสกัดคำสำคัญโดยวิธีการ NP-chunks ให้ค่าความถูกต้อง (Precision) สูงที่สุด

Mihalcea and Tarau (2004) นำเสนอวิธีการ “*TextRank*” ที่ใช้วิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยทำการสร้าง Graph-based ranking ซึ่งเป็นโมเดลกราฟสำหรับการจัดลำดับคำสำคัญที่ได้จากข้อความ โดยจะนำคำทุกคำในเอกสารมาสร้างกราฟเชื่อมโยงคำที่อยู่ติดกันไปเรื่อย ๆ ต่อมาจะพิจารณาการเชื่อมโยงของคำที่เกิดขึ้นร่วมกัน โดยพิจารณาคำภายในโครงสร้างกราฟที่มีเส้นเชื่อมมากที่สุดแล้วจึงนำคำมารวมกัน ในการรวมกันระหว่างคำหนึ่ง ๆ กับคำอื่น ๆ จะใช้แนวความคิดของการ “voting” ระหว่างคำ เพื่อรวมคำเข้าด้วยกัน จากนั้นจะคำนวณค่าน้ำหนักของคำที่นำมารวมกัน และสร้างรายการคำสำคัญจากค่าน้ำหนัก สุดท้ายจะสกัดคำสำคัญจากการจัดอันดับค่าน้ำหนัก และทำการคัดกรองคำสำคัญที่ค่าน้ำหนักผ่านเกณฑ์

Wartena, Brussee and Slakhorst (2010) นำเสนอการสกัดคำสำคัญในเอกสารหนึ่ง ๆ โดยทำการคัดเลือกคำที่เหมาะสมที่สุดจากเนื้อหาภายในเอกสาร งานวิจัยนี้มีข้อมูลอินพุตคือบทคัดย่อจาก ACM abstracts และ BBC synopses และทำการสกัดคำสำคัญโดยใช้วิธีการทางสถิติ 4 วิธีการ คือ 1) TF-IDF 2) JSD 3) InfGain และ 4) Correlation จากนั้นนำค่าที่ได้ไปหาค่า Co-occurrence

distribution คือ การหาค่าน้ำหนักโดยพิจารณาคำสำคัญที่สกัดได้ว่ามีความสัมพันธ์กับข้อความภายในเอกสารนั้น ๆ หรือไม่ ซึ่งผลลัพธ์จากการสกัดคำสำคัญแต่ละวิธีการร่วมกับการหาค่า Co-occurrence distribution พบว่าวิธีการ TF-IDF ร่วมกับการหาค่า Co-occurrence distribution ให้ผลลัพธ์ที่ดีกว่าวิธีการอื่น ๆ

Rose *et al.* (2010) ได้นำเสนอวิธีการสกัดคำสำคัญโดยอัตโนมัติจากบทคัดย่อภาษาอังกฤษ และได้นำวิธีการทางภาษาศาสตร์ (Linguistic approach) เข้ามาใช้ในการพิจารณาคำสำคัญ วิธีการนี้เรียกว่า Rapid automatic keyword extraction หรือ “*RAKE*” โดยวิธีการมีดังนี้ 1) พิจารณาคำที่อยู่ติดกันไปเรื่อย ๆ จนเจอคำหยุด (Stop word) แล้วจึงนำคำเหล่านั้นมาสร้างรายการแคนดิเดทของคำสำคัญ 2) ทำการสร้างเมทริกซ์ (Matrix) ของแต่ละคำที่ปรากฏในรายการแคนดิเดทของคำสำคัญ เพื่อคำนวณคะแนนความถี่ของคำที่เกิดขึ้นร่วมกัน 3) รวมคะแนนของคำสำคัญที่ปรากฏในรายการแคนดิเดทของคำสำคัญ และทำการเรียงลำดับรายการแคนดิเดทของคำสำคัญโดยพิจารณาคะแนนของคำสำคัญแต่ละคำจากมากไปหาน้อย 4) พิจารณา top-T scoring โดยจะสนใจเพียง 1 ใน 3 คำสำคัญจากรายการแคนดิเดทของคำสำคัญ วิธีการ “*RAKE*” ได้ทำการเปรียบเทียบกับวิธีการ *N*-gram, NP-chunks และ POS Tag Patterns (Hulth, 2003) และวิธีการ “*TextRank*” (Mihalcea and Tarau, 2004) ซึ่งผลลัพธ์ในการสกัดคำสำคัญโดยวิธีการ “*RAKE*” ให้ค่าความถูกต้อง (Precision) มากที่สุด

Gemkow *et al.* (2018) นำเสนอวิธีการสกัดคำสำคัญแบบผสมผสาน (Hybrid approach) จากบทความหนึ่ง ๆ ของศาสตร์ทางด้านวิศวกรรม ซึ่งจะประยุกต์ใช้วิธีการ 1) วิธีการทางภาษาศาสตร์ (Linguistics approach) โดยเริ่มจากการนำบทความรับเข้าไปทำการตัดคำ (Word tokenization) เพื่อให้ได้มาซึ่งคำหนึ่ง ๆ จากนั้นนำไปวิเคราะห์หน้าที่ของคำ เพื่อนำไปใช้ในการสกัดคำสำคัญโดยการพิจารณาถึงนามวลี (Noun phrase) ที่เกิดขึ้น แล้วนำนามวลีเหล่านั้นมาสร้างเป็นรายการแคนดิเดตของคำสำคัญ สุดท้ายนำรายการแคนดิเดตของคำสำคัญที่ได้ไปทำการหารูปเดิมของคำหรือหารากศัพท์ (Lemmatization) เพื่อลดความซ้ำซ้อนของคำสำคัญในรายการแคนดิเดทและกำจัดคำที่เป็นคำหยุด (Stop word removal) และ 2) วิธีการทางสถิติ (Statistics approach) นำมาประยุกต์ใช้ในการลดขนาดของรายการแคนดิเดทของคำสำคัญที่สกัดได้ จากการพิจารณาจำนวนครั้งที่คำสำคัญหนึ่ง ๆ ปรากฏขึ้นในบทความ เพื่อนำไปวิเคราะห์ว่าคำสำคัญนั้น ๆ เกิดขึ้นมากกว่าเกณฑ์ที่กำหนดไว้หรือไม่ เพื่อให้ได้มาซึ่งรายการของคำสำคัญที่มีประสิทธิภาพมากยิ่งขึ้น

Yao, Pengzhou and Chi (2019) นำเสนอวิธีการสกัดคำสำคัญจากบทความของข่าว (ภาษาอังกฤษ) โดยวิธีการสกัดคำสำคัญแบบผสมผสาน (Hybrid approach) ระหว่างวิธีการทางสถิติ

(Statistics approach) และวิธีการทางการเรียนรู้ของเครื่องจักร (Machine Learning approach) โดยวิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) ซึ่งวิธีการสกัดคำสำคัญที่นำเสนอจะถูกแบ่งเป็น 3 ขั้นตอน ได้แก่ 1) วิธีการประมวลผลข้อความเบื้องต้น (Text preprocessing) เป็นการกำจัดคำหยุด, เครื่องหมายวรรคตอน (สัญลักษณ์พิเศษ) 2) การคัดกรองคำ โดยการคัดกรองเฉพาะคำที่ทำหน้าที่เป็นคำนาม (Noun) และคำกิริยา (Verb) เท่านั้น และ 3) การสกัดคำสำคัญ เริ่มจากการคำนวณหาค่าความถี่ของคำที่ถูกคัดกรองไว้จากขั้นตอนก่อนหน้า โดยวิธี TF-IDF จากนั้นใช้วิธีการของ “*TextRank*” เพื่อนำมาคำนวณหาค่าการปรากฏขึ้นร่วมกันระหว่างคำในบทความนั้น ๆ แล้วทำการจัดลำดับคะแนนของคำแต่ละคำ สุดท้ายทำการสกัดคำสำคัญ โดยผลลัพธ์ที่ได้แสดงให้เห็นว่าวิธีการสกัดคำสำคัญจากการประยุกต์ใช้วิธีการ TF-IDF และ “*TextRank*” มาดำเนินงานร่วมทำให้ได้มาซึ่งคำสำคัญที่มีประสิทธิภาพมากกว่าการคำสำคัญที่สกัดได้จากวิธีการทำงานของ TF-IDF หรือ “*TextRank*” เพียงวิธีการเดียว

Biswas (2019) นำเสนอวิธีการสกัดคำสำคัญจากข้อความบนทวิตเตอร์ในหัวข้อ Uri Attack และ American Election ซึ่งประยุกต์ใช้วิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยการสร้าง Graph-based model ซึ่งเรียกว่าวิธีการ “Keyword from Weighted Graph (KWG)” ซึ่งประกอบไปด้วย 3 ขั้นตอนคือ 1) การประมวลผลข้อความเบื้องต้น (Text preprocessing) โดยทำการตัดคำ (Word tokenization) และการกำจัดคำหยุด (Stop word removal) 2) ทำการเชื่อมโยงข้อมูลคำหนึ่ง ๆ และคำที่เกิดขึ้นร่วมกัน โดยกำหนดให้คำหนึ่ง ๆ คือโหนดหนึ่ง ๆ และเมื่อมีคำที่ปรากฏขึ้นร่วมกันจะทำการสร้างเส้นเชื่อมโยงระหว่างโหนด และ 3) การระบุคำสำคัญ โดยพิจารณาความเกี่ยวเนื่องระหว่างโหนดด้วยการหาค่าน้ำหนักจากวิธีการ “Node Edge (NE) rank centrality measure” ซึ่งคำนวณจากจำนวนเส้นเชื่อโยงและความใกล้ชิดของโหนดที่เกี่ยวเนื่องกัน จากนั้นทำการประเมินประสิทธิภาพโดยการเปรียบเทียบกับวิธีการการหาค่าน้ำหนักวิธีการ “*TextRank*” (คำนวณจากจำนวนเส้นเชื่อโยงระหว่างโหนด) จากผลการทดลองพบว่าประสิทธิภาพของการหาค่าน้ำหนักแบบวิธีการ KWG นั้นให้ผลลัพธ์ที่ดีกว่า

จากงานวิจัยที่เกี่ยวข้องทางด้านการแนะนำหนังสือที่ได้นำเสนอมาทั้ง 5 แง่มุมข้างต้น ได้แก่ การแนะนำหนังสือจากการพิจารณาเนื้อหาสาระในหนังสือ, การแนะนำหนังสือจากการให้คะแนนความนิยมของผู้อ่าน, การแนะนำหนังสือที่สืบเนื่องจากการแนะนำผู้เขียน, การแนะนำหนังสือจากประวัติการยืมหนังสือของผู้ใช้ห้องสมุด และการแนะนำหนังสือโดยพิจารณาจากระดับทักษะการอ่าน แต่อย่างไรก็ตาม แนวคิดทั้งหมดไม่ได้มุ่งเน้นที่การแนะนำหนังสือจากความสอดคล้องของหนังสือกับคำอธิบายรายวิชา และ/หรือประมวลผลรายวิชาที่ประกอบไปด้วยเนื้อหาต่าง ๆ ที่ซึ่งช่วยให้ผู้อ่านได้รับการแนะนำที่สอดคล้องกับเนื้อหาที่ตนเองสนใจมากยิ่งขึ้น

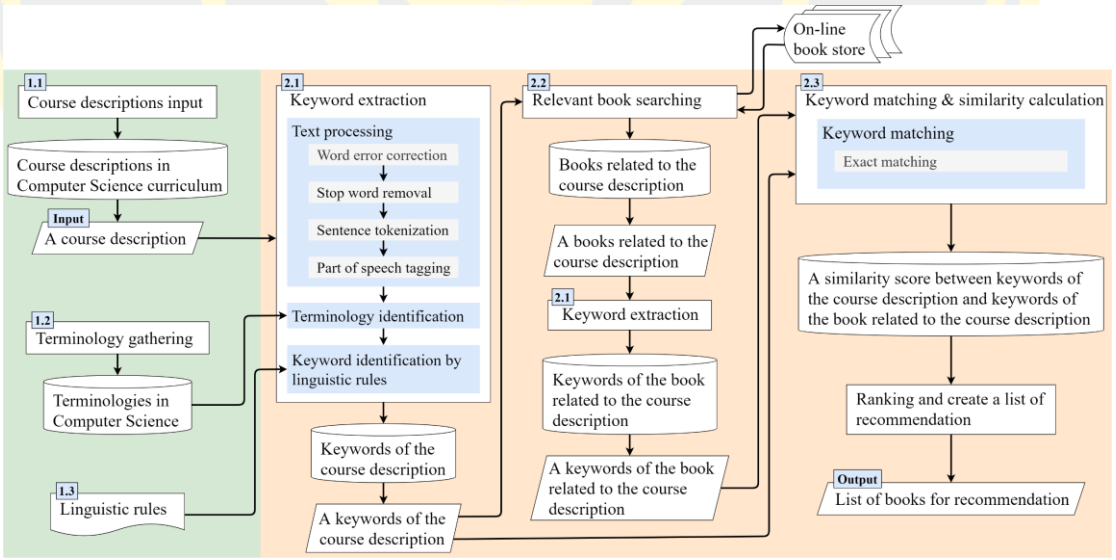
ด้วยเหตุนี้งานวิจัยนี้มุ่งเน้นที่การแนะนำหนังสืออ่านเพิ่มเติมโดยพิจารณาจากความสอดคล้องของหนังสือกับคำค้นที่เฉพาะเจาะจง (คำอธิบายรายวิชา และ/หรือประมวลผลรายวิชา) ในศาสตร์ทางด้านวิทยาการคอมพิวเตอร์เป็นหลัก เพื่อส่งเสริมการอ่านและการเรียนรู้ด้วยตนเอง อันนำมาซึ่งการพัฒนาตนเอง พัฒนาการเรียนรู้อย่างยั่งยืน ผู้วิจัยจึงประยุกต์ใช้วิธีการกรองเนื้อหา (Content-based filtering) ที่ทำการพิจารณาถึงคำสำคัญระหว่างคำอธิบายรายวิชาและหนังสือเป็นสำคัญ และใช้วิธีการหาค่าความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาเพื่อแสดงให้เห็นถึงคำสำคัญที่ครอบคลุมระหว่างในหนังสือเล่มหนึ่ง ๆ ต่อคำอธิบายรายวิชาหนึ่ง ๆ จากนั้นทำการสร้างรายการแนะนำหนังสืออ่านเพิ่มเติมให้กับผู้ใช้งาน

นอกจากนี้การแนะนำหนังสืออ่านเพิ่มเติมที่มีความสอดคล้องกับคำอธิบายรายวิชา และ/หรือประมวลรายวิชานั้น จะต้องทำการสกัดคำสำคัญจากคำอธิบายรายวิชา และหนังสืออ่านเพิ่มเติมที่ต้องการแนะนำ วิธีการสกัดคำสำคัญจึงมีบทบาทสำคัญในการประมวลผลของระบบในการแนะนำหนังสือ จากทฤษฎีและงานวิจัยที่เกี่ยวข้องทางด้านการสกัดคำสำคัญข้างต้นจะเห็นได้ว่าวิธีการทางภาษาศาสตร์นั้นมีส่วนช่วยให้ผลลัพธ์ในการสกัดคำสำคัญมีความถูกต้องสูง งานวิจัยนี้จึงประยุกต์ใช้วิธีการทางภาษาศาสตร์ร่วมกับการนำคลังคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี เพื่อสกัดคำสำคัญได้อย่างมีประสิทธิภาพ

บทที่ 3
ระบบแนะนำหนังสืออ่านเพิ่มเติม

จากที่กล่าวมาในข้างต้น เราทราบได้ว่ามีงานวิจัยที่นำเสนอการแนะนำหนังสือในแง่มุมที่หลากหลาย แต่อย่างไรก็ตามงานวิจัยดังกล่าวไม่ได้มุ่งเน้นที่การแนะนำหนังสือที่เกี่ยวเนื่องกับรายวิชาหนึ่ง ๆ ภายใต้การพิจารณาความสอดคล้องของเนื้อหาในหนังสือกับเนื้อหาในคำอธิบายรายวิชา จากการพิจารณาดังกล่าวจะทำให้ผู้อ่านได้รับการแนะนำหนังสือที่น่าสนใจและสอดคล้องกับเนื้อหาที่ตนเองต้องศึกษา

จากประเด็นข้างต้นงานวิจัยนี้จึงมีแนวคิดที่จะนำเสนอระบบแนะนำหนังสืออ่านเพิ่มเติม (Supplementary Book Suggestion system) หรือเรียกว่า “ระบบ *SBS*” โดยมุ่งเน้นที่การแนะนำหนังสือภายใต้การค้นหาความสอดคล้องระหว่างเนื้อหาในคำอธิบายรายวิชาหนึ่ง ๆ ในสาขาทางด้านคอมพิวเตอร์เทคโนโลยี กับ เนื้อหาในหนังสือเล่มหนึ่ง ๆ โดยการทำงานของระบบ *SBS* ประกอบด้วย 2 ขั้นตอนหลัก คือ 1) การเตรียมข้อมูลเบื้องต้น และ 2) การประมวลผลของระบบแนะนำ โดยรายละเอียดการทำงานทั้งสองขั้นตอนหลักแสดงได้ดังภาพที่ 8



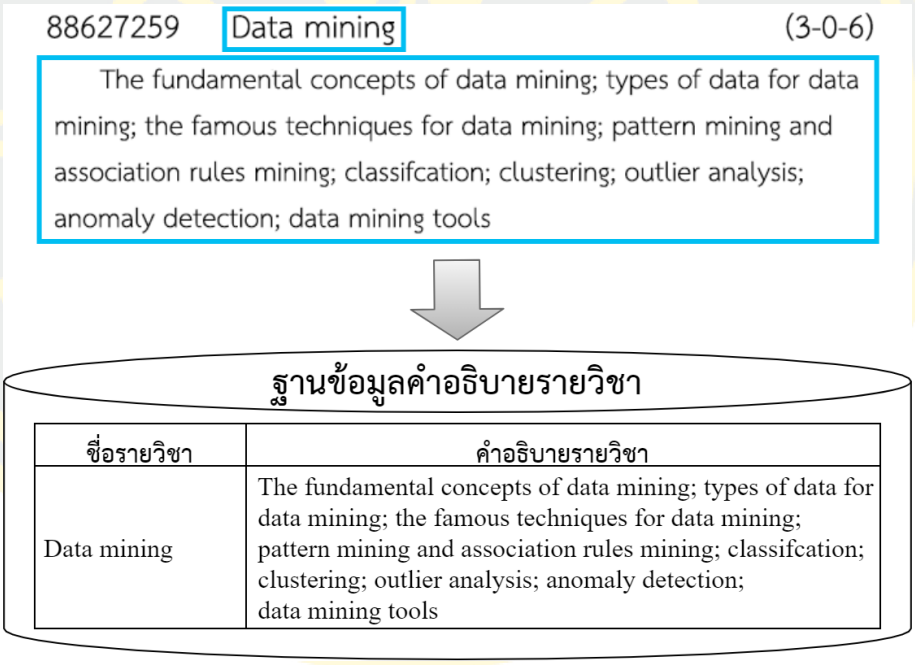
ภาพที่ 8 โครงสร้างระบบแนะนำหนังสืออ่านเพิ่มเติม

3.1 การเตรียมข้อมูลเบื้องต้น

ขั้นตอนแรกของระบบ *SBS* คือ การเตรียมข้อมูลเบื้องต้น ซึ่งประกอบไปด้วย 3 ส่วนย่อย คือ 1) การรวบรวมข้อมูลคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ 2) การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี และ 3) การสร้างกฎทางภาษาศาสตร์ ซึ่งมีรายละเอียดดังนี้

3.1.1 การรวบรวมข้อมูลคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์

ขั้นตอนนี้ได้ทำการรวบรวมคำอธิบายรายวิชาทางด้านวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา เพื่อใช้เป็นข้อมูลรับเข้าของระบบ *SBS* โดยชื่อรายวิชา และคำอธิบายรายวิชา (ในรูปแบบภาษาอังกฤษ) ของรายวิชาหนึ่ง ๆ จะถูกพิจารณาและจัดเก็บดังแสดง ภาพที่ 9 โดยส่วนที่อยู่ในกรอบสีฟ้าคือส่วนที่ต้องทำการจัดเก็บข้อมูลลงในฐานข้อมูล



ภาพที่ 9 การจัดเก็บข้อมูลชื่อวิชาและคำอธิบายรายวิชาลงในฐานข้อมูล

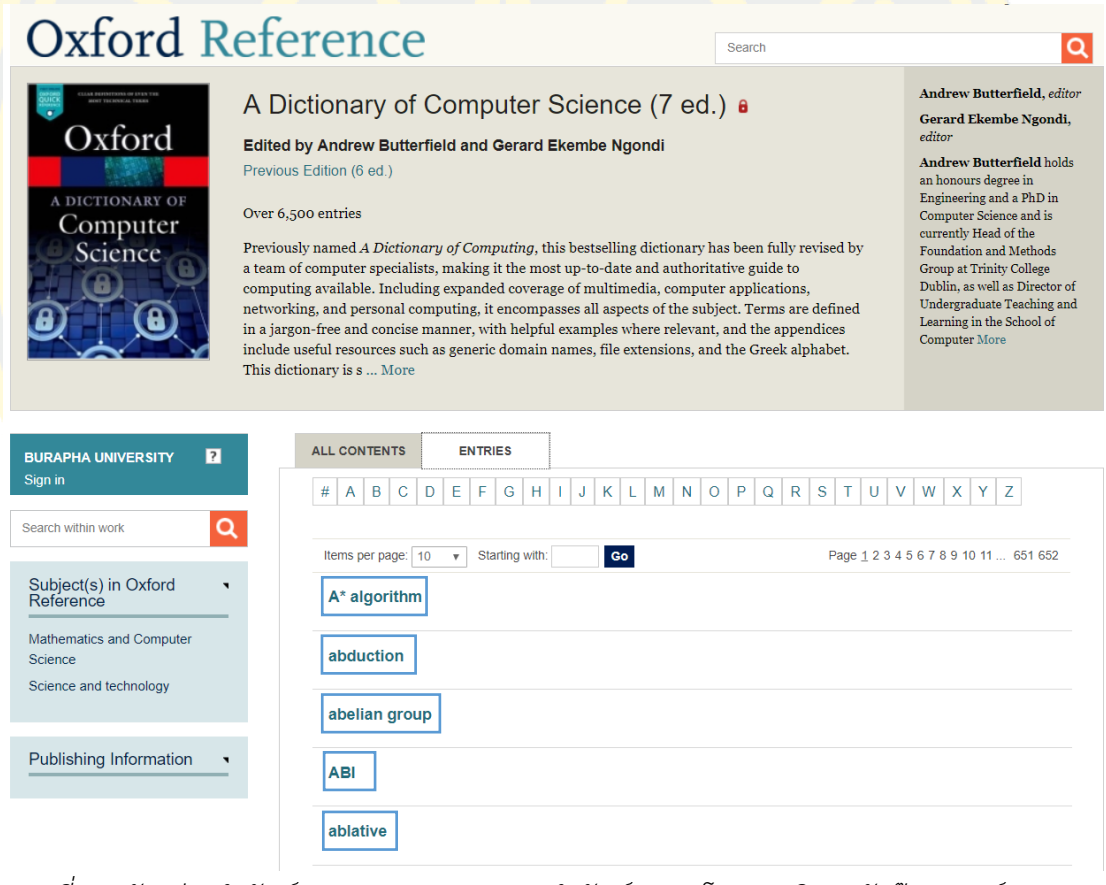
3.1.2 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี

ศาสตร์ทางด้านคอมพิวเตอร์เทคโนโลยีเป็นศาสตร์ที่มีคำศัพท์เฉพาะ (Terminology) ค่อนข้างมาก อาทิ เช่น “Natural Language Processing”, “Data Mining”, “Data Structure”, “Data Science”, “Big Data” เป็นต้น คำศัพท์เฉพาะสามารถช่วยในการระบุถึงเนื้อหาที่สำคัญและคำสำคัญที่เฉพาะเจาะจงได้ นอกจากนี้ การระบุถึงคำศัพท์เฉพาะในชื่อรายวิชา, คำอธิบายรายวิชา,

ชื่อหนังสือ และสารบัญของหนังสือ สามารถช่วยในการทำความเข้าใจเกี่ยวกับหัวข้อสำคัญที่ควรศึกษาได้

ด้วยเหตุนี้การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจึงเป็นสิ่งสำคัญประการหนึ่งในการสกัดเนื้อหาสำคัญและคำสำคัญ โดยในการรวบรวมคำศัพท์เฉพาะจะรวบรวมจากพจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด (Oxford University)[7] และ รายการคำศัพท์เฉพาะจากเว็บไซต์ Labautopedia[8] โดยใช้วิธีการสกัดข้อความบนเว็บไซต์ (Web Scraping)

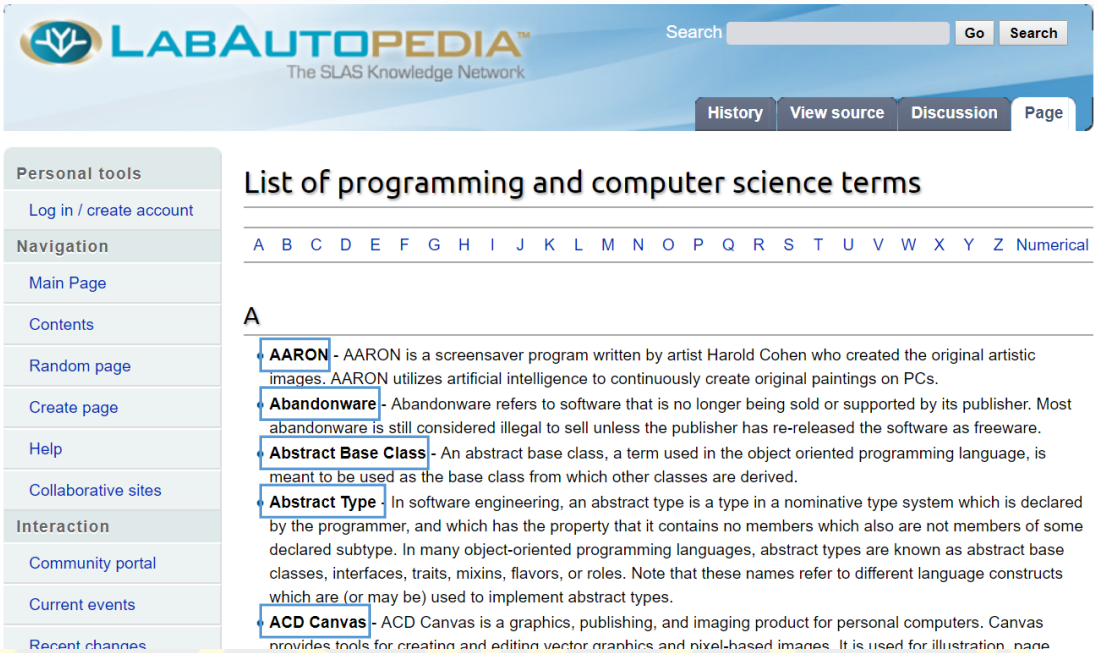
ภาพที่ 10 แสดงเว็บไซต์ที่มีข้อมูลพจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด (Oxford University) และภาพที่ 11 แสดงเว็บไซต์ที่มีข้อมูลรายการคำศัพท์จากเว็บไซต์ Labautopedia โดยกรอบสีฟ้าแสดงถึงคำศัพท์เฉพาะที่จะถูกเก็บรวบรวมไว้ในฐานข้อมูล



ภาพที่ 10 ตัวอย่างคำศัพท์เฉพาะจากพจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด

[7] http://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975

[8] http://www.labautopedia.org/mw/List_of_programming_and_computer_science_terms



ภาพที่ 11 ตัวอย่างคำศัพท์เฉพาะจากเว็บไซต์ Labautopedia

3.1.3 การสร้างกฎทางภาษาศาสตร์

กฎทางภาษาศาสตร์ถูกประยุกต์ใช้เพื่อระบุถึงเนื้อหาและคำสำคัญที่ซ่อนอยู่ในแต่ละประโยค โดยระบบ *SBS* ได้ทำการสร้างกฎทางภาษาศาสตร์ (แสดงในส่วนภาคผนวก ก) 2 ประเภท คือ

1. กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะ—กฎประเภทนี้จะพิจารณาคำศัพท์เฉพาะที่ปรากฏขึ้นในประโยคเป็นลำดับแรก จากนั้นทำการพิจารณาคำนาม (Noun) คำบุพบท (Preposition) หรือคำคุณศัพท์ (Adjective) ที่เกิดขึ้นในประโยคเดียวกันกับคำศัพท์เฉพาะ และทำการระบุคำสำคัญ ตามลำดับ ตัวอย่างเช่น กฎ : “Adjective (JJ) + Noun (NN) + Preposition (IN) + Terminology (TE)” เป็นกฎที่ประกอบด้วย 4 ส่วน คือ 1. คำคุณศัพท์ (Adjective) 2. คำนาม (Noun) 3. คำบุพบท (Preposition) และ 4. คำศัพท์เฉพาะ ตามลำดับ โดยคำทั้ง 4 ประเภทจะต้องปรากฏร่วมกันในประโยคแบบเรียงลำดับ
2. กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะ—กฎประเภทนี้จะประยุกต์ใช้กับประโยคที่ไม่มีคำศัพท์เฉพาะปรากฏ โดยจะทำการพิจารณานามวลี (Noun phrase) กล่าวคือ คำนาม (Noun) ที่ปรากฏขึ้น หรือคำนามที่ปรากฏร่วมกันกับคำคุณศัพท์ (Adjective) ในประโยค และทำการระบุคำสำคัญ ตัวอย่างเช่น กฎ : “Adjective (JJ) + Noun (NN)” เป็นกฎที่ประกอบด้วย 2 ส่วน คือ 1. คำคุณศัพท์ (Adjective) และ 2. คำนาม (Noun) โดยคำทั้ง 2 ประเภทจะต้องปรากฏร่วมกันในประโยคแบบเรียงลำดับ

หลังจากทำการเตรียมข้อมูลเบื้องต้นในข้อ 3.1.1 – 3.1.3 แล้วเสร็จ จะได้ฐานข้อมูลคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ ฐานข้อมูลคลังคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี และกฎทางภาษาศาสตร์ ตามลำดับ

3.2 การประมวลผลของระบบแนะนำ

เมื่อคำอธิบายรายวิชาหนึ่ง ๆ ถูกเพิ่มเข้ามา ระบบ *SBS* จะทำการประมวลผลกับคำอธิบายรายวิชานั้น โดยเริ่มจากการสกัดคำสำคัญที่ปรากฏอยู่ในคำอธิบายรายวิชา ทำการค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญที่สกัดก่อนหน้า ทำการสกัดคำสำคัญจากหนังสือแต่ละเล่ม และทำการเปรียบเทียบความเหมือนกันระหว่างคำสำคัญจากคำอธิบายรายวิชา และ คำสำคัญจากหนังสือแต่ละเล่ม ท้ายสุด ทำการจัดอันดับหนังสือที่มีความเกี่ยวเนื่องกับคำอธิบายรายวิชาสูง และแสดงผลให้แก่ผู้ใช้งาน ตามลำดับ

3.2.1 การสกัดคำสำคัญ

การสกัดคำสำคัญได้นำวิธีการทางภาษาศาสตร์ (Linguistic approach) ร่วมกับการนำคลังคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีที่ได้เก็บรวบรวมไว้ก่อนหน้ามาประยุกต์ใช้เพื่อสกัดคำสำคัญ โดยขั้นวิธีการสกัดคำสำคัญของระบบ *SBS* ประกอบไปด้วย 3 ขั้นตอนที่เชื่อมต่อกัน ดังนี้

3.2.1.1 การประมวลผลข้อความ

การประมวลผลข้อความมุ่งเน้นที่การตัดส่วนของคำและ/หรือเนื้อหาที่ไม่สำคัญออกจากคำอธิบายรายวิชา และทำให้ข้อมูลในคำอธิบายรายวิชาอยู่ในรูปแบบที่พร้อมสำหรับการประมวลผลในขั้นตอนต่อไป ประกอบไปด้วย 4 ขั้นตอน คือ 1) การกำจัดคำหยุด (Stop word removal) 2) การแก้ไขคำผิด (Word error correction) 3) การตัดประโยค (Sentence tokenization) และ 4) การวิเคราะห์หน้าที่ของคำในประโยค (Part of speech tagging) ตามลำดับ

ภาพที่ 12 แสดงตัวอย่างข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining” และภาพที่ 13 แสดงการประมวลผลข้อความจากข้อมูลคำอธิบายรายวิชาในภาพที่ 12 ที่ผ่านการประมวลผลข้อความจากทั้ง 4 ขั้นตอนที่เชื่อมต่อกัน โดยแต่ละขั้นตอนสามารถอธิบายประกอบตัวอย่างได้ดังนี้

ชื่อรายวิชา	Data mining
คำอธิบายรายวิชา	The fundamental concepts of data mining; types of data for data mining; the famous techniques for data mining; pattern mining and association rules mining; classification; clustering; outlier analysis; anomaly detection; data mining tools

ภาพที่ 12 ตัวอย่างข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining”

การประมวลผลข้อความ		
การกำจัดคำหยุด		~~The~~ fundamental concepts of data mining; types of data for data mining; ~~the~~ famous techniques for data mining; pattern mining and association rules mining; classifcation; clustering; outlier analysis; anomaly detection; data mining tools
การแก้ไขคำผิด	ก่อน	fundamental concepts of data mining; types of data for data mining; famous techniques for data mining; pattern mining and association rules mining; classifcation; clustering; outlier analysis; anomaly detection; data mining tools
	หลัง	fundamental concepts of data mining; types of data for data mining; famous techniques for data mining; pattern mining and association rules mining; classification; clustering; outlier analysis; anomaly detection; data mining tools
การตัดประโยค		fundamental concepts of data mining types of data for data mining famous techniques for data mining pattern mining and association rules mining classification clustering outlier analysis anomaly detection data mining tools
การวิเคราะห์หน้าที่ของคำ		('fundamental', 'JJ'), ('concepts', 'NNS'), ('of', 'IN'), ('data', 'NN'), ('mining', 'NN') ('types', 'NNS'), ('of', 'IN'), ('data', 'NN'), ('for', 'IN'), ('data', 'NN'), ('mining', 'NN') ('famous', 'JJ'), ('techniques', 'NNS'), ('for', 'IN'), ('data', 'NN'), ('mining', 'NN') ('pattern', 'NN'), ('mining', 'NN'), ('and', 'CC'), ('association', 'NN'), ('rules', 'NNS'), ('mining', 'NN') ('classification', 'NN') ('clustering', 'NN') ('outlier', 'JJR'), ('analysis', 'NN') ('anomaly', 'JJ'), ('detection', 'NN') ('data', 'NN'), ('mining', 'NN'), ('tools', 'NNS')

ภาพที่ 13 การประมวลผลข้อความจากข้อมูลคำอธิบายรายวิชา “Data mining”

1) การกำจัดคำหยุด

ขั้นตอนนี้จะเป็นการตัดคำหยุด (stop word) เช่น “a” “an” และ “the” ที่ปรากฏในประโยค โดยการตัดคำหยุดหนึ่ง ๆ จะไม่ทำให้ความหมายของประโยคเปลี่ยนแปลง แต่อย่างไรก็ตามจากการทดลองและทดสอบแนวคิด ผู้วิจัยได้สังเกตและทราบว่ามีคำหยุดบางคำที่มีส่วนช่วยในการสกัดคำสำคัญที่ซ่อนอยู่ในประโยคของคำอธิบายรายวิชา ด้วยเหตุนี้ คำหยุดประเภทคำบุพบท (Preposition) และคำสันธาน (Conjunction) จะไม่ถูกกำจัดออกไป ตัวอย่างเช่น ประโยค : “unstructured and structured data” สามารถสกัดคำสำคัญได้ 2 คำ คือ “unstructured data”

และ “structured data” ซึ่งจากประโยคอินพุตจะมีคำสันธาน “and” ซึ่งเชื่อมโยงคำที่ซ่อนอยู่ในประโยค เพื่อให้ได้เนื้อหาหรือข้อมูลที่แม่นยำมากขึ้น “and” ซึ่งเป็นคำหยุดจะไม่ถูกตัดออกจากประโยค

2) การแก้ไขคำผิด

คำอธิบายรายวิชาหนึ่ง ๆ อาจจะมีคำผิดปรากฏอยู่ ที่ซึ่งจะต้องได้รับการแก้ไขให้ถูกต้อง ขั้นตอนการแก้ไขคำผิดในคำอธิบายรายวิชาหนึ่ง ๆ งานวิจัยนี้ได้ประยุกต์ใช้วิธีการแก้ไขการสะกดผิด (Spelling Mistake Correction (*SMC*)) (Gupta, 2015) ซึ่งดำเนินการโดยทำการเปรียบเทียบคำหนึ่ง ๆ กับคำศัพท์ในพจนานุกรมภาษาอังกฤษ (Dictionary) ด้วยวิธีการ *N*-gram เพื่อทำการเปรียบเทียบอักขระที่ปรากฏในคำหนึ่ง ๆ กับคำศัพท์ในพจนานุกรม จากนั้นจึงดำเนินการแก้ไขคำผิด ตัวอย่างเช่นภาพที่ 13 ในส่วนของการแก้ไขคำผิด พบคำว่า “classifcation” มีการสะกดผิด ในขั้นตอนนี้จะทำการพิจารณาคำดังกล่าวด้วยวิธีการ *N*-gram เทียบกับคำศัพท์ในพจนานุกรมภาษาอังกฤษ และแก้ไขเป็นคำที่ถูกต้องคือ “classification”

3) การตัดประโยค

ขั้นตอนนี้จะทำการแยก/ตัดประโยคจากคำอธิบายรายวิชาให้อยู่ในรูปแบบประโยคความเดียว โดยจะทำการวิเคราะห์ตามหลักไวยากรณ์ของประโยค คือ เมื่อพบ “.” (Full stop) หรือ “;” (Semi colon) จะแยก/ตัดกลุ่มของคำที่ปรากฏก่อนหน้าเป็นหนึ่งประโยค หรือ หนึ่งหัวข้อ ซึ่งจะทำให้คำอธิบายรายวิชาถูกแบ่งออกเป็นหลายส่วน ตัวอย่างเช่นภาพที่ 13 แสดงคำอธิบายรายวิชา “Data mining” ที่ประกอบด้วยหัวข้อต่าง ๆ คั่นด้วย “;” ดังนั้น เมื่อพิจารณาการตัดประโยคด้วย “;” จะทำให้มีเนื้อหาเป็นจำนวนทั้งหมด 9 ประโยค คือ 1) “fundamental concepts of data mining” 2) “ types of data for data mining” 3) “ famous techniques for data mining” 4) “ pattern mining and association rules mining” 5) “ classification” 6) “ clustering” 7) “outlier analysis” 8) “anomaly detection” และ 9) “data mining tools” ตามลำดับ

4) การวิเคราะห์หน้าที่ของคำ

การวิเคราะห์หน้าที่ของคำจะทำการพิจารณาแต่ละคำในแต่ละประโยคของคำอธิบายรายวิชา โดยเริ่มจากการแบ่งประโยคออกมาเป็นหน่วยคำต่าง ๆ (Word tokenization) จากนั้นทำการระบุ/กำกับหน้าที่ของคำแต่ละชนิดไว้ในคำเหล่านั้น ด้วยการประยุกต์ Stanford part of speech tagger (Toutanova *et al.*, 2003) ในการระบุ/กำกับหน้าที่ของคำ โดยคำหนึ่ง ๆ จะถูกกำหนดหน้าที่อย่างใดอย่างหนึ่งจากกลุ่มดังต่อไปนี้

1. คำนาม (Noun) ประกอบด้วย
    i) “*NN*” (Noun, singular เช่น “desk”)
    ii) “*NNS*” (Noun, plural เช่น “desks”)
2. คำนามชี้เฉพาะ (Proper noun) ประกอบด้วย
    i) “*NNP*” (Proper noun, singular เช่น “Harrison”)
    ii) “*NNPS*” (Proper noun, plural เช่น “Americans”)
3. คำสรรพนาม (Pronoun) ประกอบด้วย
    i) “*PRP*” (Personal pronoun เช่น “he”)
    ii) “*PRP$*” (Possessive pronoun เช่น “his”)
4. คำกริยา (Verb) ประกอบด้วย
    i) “*VB*” (Verb, base form เช่น “take”)
    ii) “*VBD*” (Verb, past tense เช่น “took”)
    iii) “*VBG*”(Verb, present participle เช่น “taking”)
    iv) “*VBN*”(Verb, past participle เช่น “taken”)
    v) “*VBP*”(Verb, non-3rd person singular present เช่น “take”)
    vi) “*VBZ*” (Verb, 3rd person singular present เช่น “takes”)
5. คำคุณศัพท์ (Adjective) ประกอบด้วย
    i) “*JJ*” (Adjective เช่น “big”)
    ii) “*JJR*” (Adjective, comparative เช่น “bigger”)
    iii) “*JJS*” (Adjective, superlative เช่น “biggest”)
6. คำกริยาวิเศษณ์ (Adverb) ประกอบด้วย
    i) “*RB*” (Adverb เช่น “very”)
    ii) “*RBR*” (Adverb, comparative เช่น “better”)
    iii) “*RBS*” (Adverb, superlative เช่น “best”)
7. คำบุพบท (Preposition) ประกอบด้วย
    i) “*IN*” (Preposition เช่น “of”)
8. คำสันธาน (Conjunction) ประกอบด้วย
    i) “*CC*” (Coordinating conjunction เช่น “and”)
9. คำอุทาน (Interjection) ประกอบด้วย
    i) “*UH*” (Interjection เช่น “Oh”)

10. ตัวเลข (Digit) ประกอบไปด้วย
    i) “*CD*” (cardinal digit เช่น “3”)
11. คำกำกับคำนาม (Determiner) ประกอบไปด้วย
    i) “*DT*” (determiner เช่น “The”)

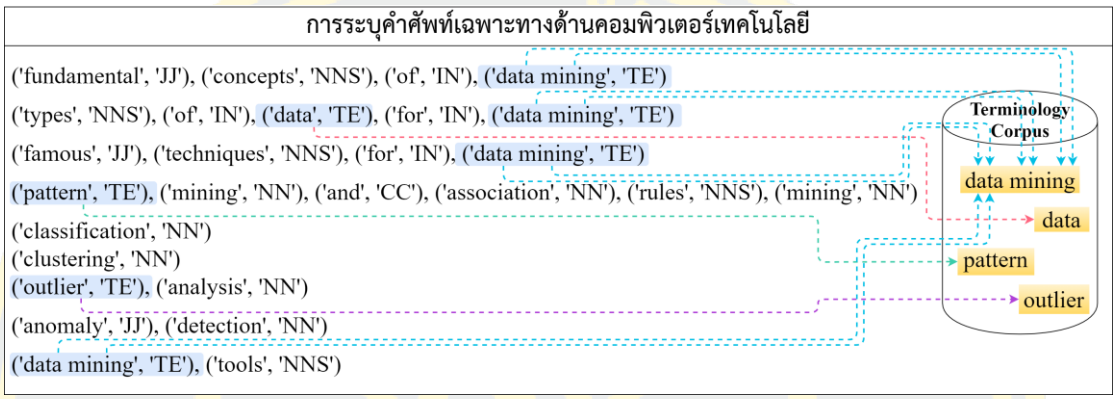
ตัวอย่างเช่น ประโยคที่ 1 ในส่วนของผลจากการตัดคำ ในภาพที่ 13 “fundamental concepts of data mining” ลำดับจะทำการแบ่งคำออกมาเป็นหน่วยคำจะได้ 1) (‘fundamental’) 2) (‘concepts’) 3) (‘of’) 4) (‘data’) และ 5) (‘mining’) ตามลำดับ จากนั้นทำการระบุหน้าที่ของคำโดยเก็บรายละเอียดอยู่ในรูปแบบ (คำ, หน้าที่ของคำ) จะได้เป็น 1) (‘fundamental’, ‘*JJ*’) 2) (‘concepts’, ‘*NNS*’) 3) (‘of’, ‘*IN*’) 4) (‘data’, ‘*NN*’) และ 5) (‘mining’, ‘*NN*’) โดยเมื่อพิจารณาทุกประโยคคำอธิบายรายวิชาจะได้รูปแบบประโยคดังต่อไปนี้

1. (‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’)
2. (‘types’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data’, ‘*NN*’), (‘for’, ‘*IN*’), (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’)
3. (‘famous’, ‘*JJ*’), (‘techniques’, ‘*NNS*’), (‘for’, ‘*IN*’), (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’)
4. (‘pattern’, ‘*NN*’), (‘mining’, ‘*NN*’), (‘and’, ‘*CC*’), (‘association’, ‘*NN*’), (‘rules’, ‘*NNS*’), (‘mining’, ‘*NN*’)
5. (‘classification’, ‘*NN*’)
6. (‘clustering’, ‘*NN*’)
7. (‘outlier’, ‘*JJR*’), (‘analysis’, ‘*NN*’)
8. (‘anomaly’, ‘*JJ*’), (‘detection’, ‘*NN*’)
9. (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’), (‘tools’, ‘*NNS*’)

3.2.1.2 การระบุคำศัพท์เฉพาะ

การค้นหาคำศัพท์เฉพาะในคำอธิบายรายวิชาจะประยุกต์ใช้วิธีการ *N*-gram (Lopez-Gazpio *et al.*, 2019) เพื่อเปรียบเทียบกลุ่มของคำในประโยคหนึ่ง ๆ ของคำอธิบายรายวิชากับคำศัพท์เฉพาะที่รวบรวมไว้ก่อนหน้า ตัวอย่างเช่นภาพที่ 13 (ส่วนการวิเคราะห์หน้าที่ของคำ) ประโยคแรกของคำอธิบายรายวิชา “Data mining” คือ (‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’) ซึ่งจากพิจารณากลุ่มของคำด้วยวิธีการ *N*-gram จะทำให้คำ (‘data’, ‘*NN*’) และ (‘mining’, ‘*NN*’) จะถูกรวมกลุ่มกันเป็น ‘data mining’ จากนั้นจะนำมา

เปรียบเทียบกับฐานข้อมูลคำศัพท์เฉพาะที่มีคำว่า ‘data mining’ ถูกบรรจุอยู่ ดังนั้น ‘data mining’ จึงถูกระบุว่าเป็นคำศัพท์เฉพาะ ถูกระบุหน้าที่ของคำเป็น “*TE*” (Terminology) ที่อยู่ในรูปแบบ (‘data mining’, ‘*TE*’) และนำไปแทนที่ (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’) ในประโยค ทำให้ประโยคประกอบด้วย 4 คำเรียงต่อกัน คือ (‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data mining’, ‘*TE*’) (ดังแสดงในภาพที่ 14)



ภาพที่ 14 การระบุคำศัพท์เฉพาะจากข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining”

3.2.1.3 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์

ขั้นตอนนี้ประยุกต์ใช้กฎทางภาษาศาสตร์ที่สร้างขึ้นก่อนหน้า ซึ่งประกอบไปด้วยกฎ 2 ประเภท คือ 1) กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะ และ 2) กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะ ตามลำดับ การระบุคำสำคัญจะเริ่มต้นจากการพิจารณาคำศัพท์เฉพาะที่ปรากฏขึ้นในประโยคหนึ่ง ๆ เป็นลำดับแรก ถ้าในประโยคมีคำศัพท์เฉพาะปรากฏ จะทำการประยุกต์ใช้กฎประเภท 1) แต่หากในประโยคไม่มีคำศัพท์เฉพาะปรากฏจะทำการประยุกต์ใช้กฎประเภท 2) ในการระบุคำสำคัญ ตัวอย่างเช่น พิจารณาประโยคที่ 1 คือ “(‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data mining’, ‘*TE*’)” ในภาพที่ 15 เมื่อพิจารณาประโยคดังกล่าวพบว่ามีคำศัพท์เฉพาะปรากฏในประโยค นั่นคือ “(‘data mining’, ‘*TE*’)” ดังนั้น ขั้นตอนการระบุคำสำคัญจะทำการประยุกต์ใช้กฎประเภทที่ 1 เป็นจำนวน 3 กฎ คือ

i) กฎ : “Adjective (*JJ*) + Noun (*NN*) + Preposition (*IN*) + Terminology (*TE*)” → “Terminology (*TE*)”,

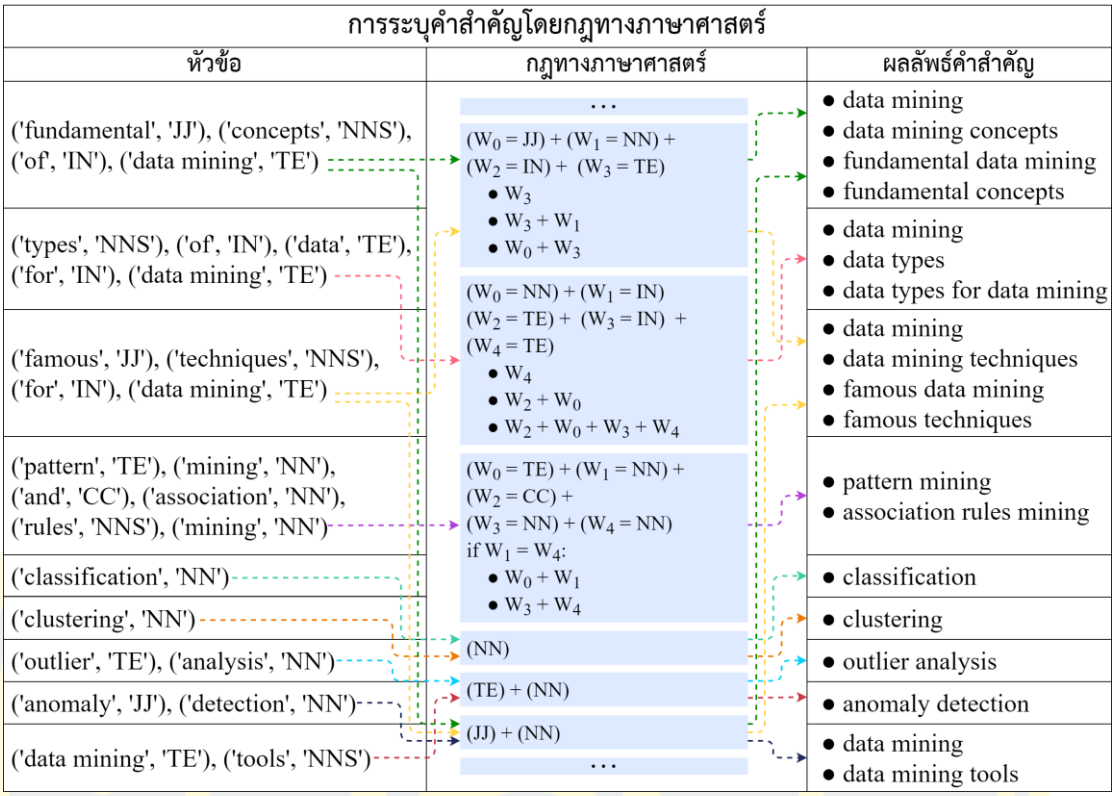
ii) กฎ : “Adjective (*JJ*) + Noun (*NN*) + Preposition (*IN*) + Terminology (*TE*)” → “Terminology (*TE*) + Noun (*NN*)”

iii) กฎ : “ Adjective (*JJ*) + Noun (*NN*) + Preposition (*IN*) + Terminology (*TE*)” → “Adjective (*JJ*) + Terminology (*TE*)”

โดยกฎ i) จะทำการแปลงข้อมูลจากประโยค “(‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data mining’, ‘*TE*’)” ให้กลายเป็น “(‘data mining’, ‘*TE*’)” กฎ ii) จะทำการแปลงข้อมูลอินพุตเดียวกัน ให้กลายเป็น “(‘data mining’, ‘*TE*’), (‘concepts’, ‘*NNS*’)” และ กฎ iii) จะทำการแปลงข้อมูลอินพุตเดียวกันให้กลายเป็น “(‘fundamental’, ‘*JJ*’), (‘data mining’, ‘*TE*’)” ตามลำดับ นอกเหนือจากข้างต้นแล้ว ขั้นตอนการระบุคำสำคัญจะทำการประยุกต์ใช้กฎประเภทที่ 2 เป็นจำนวน 1 กฎ คือ

กฎ : “ Adjective (*JJ*) + Noun (*NN*) + Preposition (*IN*) + Terminology (*TE*)” → “ Adjective (*JJ*) + Noun (*NN*)”

ซึ่งจะทำให้ได้ผลลัพธ์เป็น “(‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’)” โดยหลังจากทำการประยุกต์ใช้กฎทั้งสองประเภทแล้ว จะทำให้สามารถสกัดเนื้อหาสำคัญจากประโยคที่ 1 ได้ 4 ส่วน คือ 1) “(‘data mining’, ‘*TE*’)”, 2) “(‘data mining’, ‘*TE*’), (‘concepts’, ‘*NNS*’)” 3) “(‘fundamental’, ‘*JJ*’), (‘data mining’, ‘*TE*’)” และ 4) “(‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’)” ตามลำดับ การประมวลผลนี้จะประยุกต์ใช้กับทุกประโยคในคำอธิบายรายวิชา ซึ่งจะทำให้ได้ผลลัพธ์ดังภาพที่ 15



ภาพที่ 15 การระบุคำสำคัญในคำอธิบายรายวิชา "Data mining" โดยกฎทางภาษาศาสตร์

3.2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา

จากการสกัดคำสำคัญจากขั้นตอนก่อนหน้าจะทำให้ได้เซตของคำสำคัญของคำอธิบายรายวิชา cd ซึ่งสามารถเขียนอยู่ในรูปแบบ $K_{cd} = \{k_{cd,1}, k_{cd,2}, \ldots, k_{cd,m}\}$ จากนั้นทำการค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญแต่ละคำ $k_{cd,i} \in K_{cd}$ จากเว็บไซต์ Barnes & Noble[9] ซึ่งเป็นเว็บไซต์ขายหนังสือออนไลน์ ที่ก่อตั้งขึ้นตั้งแต่ปี พ.ศ. 2460 แต่มุ่งเน้นทางด้านหนังสือที่เกี่ยวข้องกับการศึกษาในปี ปี พ.ศ. 2556 โดยเว็บไซต์ดังกล่าวทำการรวบรวมหนังสือจากหลากหลายสำนักพิมพ์ เช่น McGraw-Hill Companies, Inc.[10], Addison-Wesley Publishing Company[11], Springer Publishing Company[12], John Wiley & Sons, Inc.[13], O'Reilly Media Inc.[14] เป็นต้น

[9] https://www.barnesandnoble.com

[10] https://www.mheducation.com

[11] https://www.pearson.com

[12] https://www.springer.com

[13] https://www.wiley.com

[14] https://www.oreilly.com

โดยเว็บไซต์ดังกล่าวมีการแสดงข้อมูลหนังสือที่หลากหลาย อาทิ เช่น ชื่อหนังสือ, ชื่อผู้แต่ง, ราคา, หมายเลข 13 หลักประจำหนังสือ (ISBN13), ชื่อสำนักพิมพ์, วันที่ตีพิมพ์หนังสือ, คำอธิบายหนังสือ และสารบัญหนังสือ และอื่น ๆ โดยข้อมูลที่แสดงผลข้างต้น สามารถนำมาวิเคราะห์เกี่ยวกับเนื้อหาสำคัญที่ปรากฏในหนังสือได้

เมื่อทำการพิจารณาคำสำคัญ $k_{cd,i} \in K_{cd}$ หนึ่ง ๆ และพบหนังสือเล่มหนึ่ง ๆ ที่เกี่ยวเนื่องกับคำสำคัญ $k_{cd,i}$ ระบบ *SBS* จะทำการจัดเก็บข้อมูลหนังสือเล่มนั้นโดยใช้วิธีการสกัดข้อความบนเว็บไซต์ (Web Scraping) ตัวอย่างเช่นภาพที่ 16 แสดงข้อมูลหน้าเว็บไซต์ของหนังสือ “Introduction to Data Mining” ที่ค้นพบจากการค้นหาด้วยคำสำคัญ “data mining” โดยข้อมูลสำคัญของหนังสือที่ต้องทำการจัดเก็บจะเป็นข้อมูลภายในกรอบสีฟ้าที่ปรากฏในภาพ ข้อมูลดังกล่าวจะถูกเก็บรวบรวมไว้ในฐานข้อมูลของหนังสือ ซึ่งสามารถเขียนในรูปแบบสัญลักษณ์ได้เป็น $B^{cd} = \{b_1^{cd}, b_2^{cd}, \cdots, b_x^{cd}\}$ ที่ซึ่งหนังสือ $b_i^{cd} =$ “*Introduction to Data Mining*” จะประกอบไปด้วยข้อมูลดังนี้

< ‘***Title***: *Introduction to Data Mining*’,
‘***Author***: *Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar*’,
‘***Price***: *108.50$*’,
‘***ISBN13***: *9780133128901*’,
‘***Publisher***: *Pearson Education*’,
‘***Publication date***: *01/04/2018*’,
‘***Description***: *Introducing the concepts and algorithms of data mining. Introduction to Data Mining, 2nd Edition, gives a comprehensive overview of the background and general themes of data mining and is designed to be useful to students, instructors, researchers, and professionals. Presented in a clear and accessible way, the book outlines fundamental concepts and algorithms for each topic, thus providing the reader with the necessary background for the application of data mining to real problems. The text helps readers understand the nuances of the subject, and includes important sections on classification, clustering, association analysis, and anomaly detection. This edition improves on the first iteration of the book, published over a decade ago, by addressing the significant changes in the industry as a result of advanced technology and data growth.*’,

‘***Table of Contents (ToC)***: *Introduction. Data. Classification: Basic Concepts and Techniques. Classification: Alternative Techniques. Association Analysis: Basic Concepts and Algorithms. Association Analysis: Advanced Concepts. Cluster Analysis: Basic Concepts and Algorithms. Cluster Analysis: Additional Issues and Algorithms. Pattern mining. Anomaly Detection. Avoiding False Discoveries’* >

ภาพที่ 16 ข้อมูลหนังสือที่ทำการเก็บรวมรวมไว้ในฐานข้อมูลของหนังสือ

หลังจากทำการรวบรวมหนังสือที่เกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชาครบถ้วนแล้ว ระบบ *SBS* จะพิจารณาข้อมูลส่วนข้อมูลคำอธิบายหนังสือ (*'Description'*) และสารบัญของหนังสือ (*'Table of Contents (ToC)'*) แต่ละเล่ม จากนั้นนำข้อมูลทั้งสองส่วนไปประมวลผลด้วยกระบวนการสกัดคำสำคัญ (หัวข้อ 3.2.1) เพื่อทราบถึงคำสำคัญที่แสดงถึงเนื้อหาในหนังสือ โดยผลลัพธ์ที่ได้เป็นเซตของคำสำคัญ $K^{b_i^{cd}} = \left\{k_1^{b_i^{cd}}, k_2^{b_i^{cd}}, \cdots, k_z^{b_i^{cd}}\right\}$ ของหนังสือ b_i^{cd}

3.2.3 การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา และการจัดอันดับหนังสือ

ขั้นตอนสุดท้ายของระบบ *SBS* คือ การคำนวณความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือแต่ละเล่ม จากนั้น ทำการจัดอันดับหนังสือตามคะแนนความคล้ายคลึงของคำสำคัญ (เรียงลำดับจากหนังสือที่มีความเกี่ยวเนื่องมากที่สุดก่อน)

การคำนวณความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือแต่ละเล่ม จะเริ่มจากการเปรียบเทียบกันระหว่างคำสำคัญหนึ่ง ๆ ที่สกัดจากคำอธิบายรายวิชา และคำสำคัญหนึ่ง ๆ ที่สกัดจากหนังสือแต่ละเล่ม โดยใช้วิธีการเปรียบเทียบแบบตรงตัว (Exact matching) กล่าวคือ เมื่อมีคำสำคัญหนึ่ง ๆ ของคำอธิบายรายวิชา ปรากฏอยู่ในหนังสือที่เกี่ยวเนื่องเล่มหนึ่ง ๆ จะให้คะแนนความคล้ายคลึงเท่ากับ 1 แต่ถ้าคำสำคัญของคำอธิบายรายวิชา ไม่ปรากฏอยู่ในหนังสือเลย จะให้คะแนนความคล้ายคลึงเท่ากับ 0 ตัวอย่างเช่นภาพที่ 17 แสดงการเปรียบเทียบคำสำคัญระหว่างเซตของคำสำคัญของคำอธิบายรายวิชาและเซตของคำสำคัญของหนังสือชื่อ “Introduction to Data Mining” จากรูปแสดงให้เห็นว่ามีคำสำคัญที่ถูกสกัดได้จากคำอธิบายรายวิชาคำว่า “data mining” ปรากฏในเนื้อหาสารบัญหนังสือดังกล่าว ดังนั้นคะแนนความคล้ายคลึงของคำสำคัญ “data mining” เท่ากับ 1 จะเห็นได้ว่ามี 5 คำสำคัญ ประกอบด้วย “data mining”, “data mining concepts”, “pattern mining”, “classification” และ “anomaly detection” ซึ่งทำให้คะแนนความคล้ายคลึงของทุกคำสำคัญดังกล่าวจะถูกกำหนดให้มีค่าเท่ากับ 1 แต่ในส่วนของคำสำคัญ “fundamental data mining”, “data types”, “data types for data mining”, “data mining techniques”, “association rules mining”, “clustering”, “outlier analysis” และ “data mining tools” ที่ปรากฏเฉพาะคำอธิบายรายวิชา แต่ไม่ปรากฏในหนังสือ “Introduction to Data Mining” คำสำคัญเหล่านี้จะถูกกำหนดให้มีค่าเท่ากับ 0 ตามลำดับ

	data mining	fundamental data mining	data mining concepts	data types	data types for data mining	data mining techniques	pattern mining	association rules mining	classification	clustering	outlier analysis	anomaly detection	data mining tools
คำอธิบายรายวิชา “Data mining”	1	1	1	1	1	1	1	1	1	1	1	1	1
ISNB13: 9780133128901 ชื่อหนังสือ: Introduction to Data Mining	1	0	1	0	0	0	1	0	1	0	0	1	0

ภาพที่ 17 การเปรียบเทียบคำสำคัญจากคำอธิบายรายวิชา “Data mining” และหนังสือ “Introduction to Data Mining”

จากข้างต้นการคำนวณคะแนนความคล้ายคลึง และคะแนนความไม่คล้ายคลึง ระหว่างหนังสือ b_i^{cd} และคำอธิบายรายวิชา cd สามารถคำนวณได้จากสมการที่ (1) และ (2) ดังนี้

$$SIM\left(cd, b_i^{cd}\right) = \frac{\left|K_{cd} \cap K^{b_i^{cd}}\right|}{|K_{cd}|} \quad (1)$$

$$disSIM\left(cd, b_i^{cd}\right) = 1 - SIM\left(cd, b_i^{cd}\right) \quad (2)$$

ตัวอย่างเช่นภาพที่ 18 แสดงการคำนวณคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา ที่คำนวณจากสมการ (1) และภาพที่ 19 แสดงการคำนวณคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา คำนวณจากสมการ (2)

หนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา “Data Mining”	**การคำนวณคะแนนความคล้ายคลึง** $SIM(cd, b_i^{cd}) = \frac{\left\|K_{cd} \cap K^{b_i^{cd}}\right\|}{\|K_{cd}\|}$	**คะแนนความคล้ายคลึง**
ISNB13: 9780133128901 ชื่อหนังสือ: Introduction to Data Mining	$SIM(cd, b_1^{cd}) = \frac{\|5\|}{\|13\|}$	0.385
ISNB13: 9780128014608 ชื่อหนังสือ: Predictive Analytics and Data Mining	$SIM(cd, b_2^{cd}) = \frac{\|6\|}{\|13\|}$	0.462
ISNB13: 9781118893173 ชื่อหนังสือ: Data Mining For Dummies	$SIM(cd, b_3^{cd}) = \frac{\|9\|}{\|13\|}$	0.692
ISNB13: 9780123814791 ชื่อหนังสือ: Data mining: Concepts and Techniques	$SIM(cd, b_4^{cd}) = \frac{\|6\|}{\|13\|}$	0.462
⋮	⋮	⋮

ภาพที่ 18 การคำนวณคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวข้อง

หนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา “Data Mining”	**การคำนวณคะแนนความไม่คล้ายคลึง** $disSIM(cd, b_i^{cd}) = 1 - SIM(cd, b_i^{cd})$	**คะแนนความไม่คล้ายคลึง**
ISNB13: 9780133128901 ชื่อหนังสือ: Introduction to Data Mining	$disSIM(cd, b_1^{cd}) = 1 - 0.385$	0.615
ISNB13: 9780128014608 ชื่อหนังสือ: Predictive Analytics and Data Mining	$disSIM(cd, b_2^{cd}) = 1 - 0.462$	0.538
ISNB13: 9781118893173 ชื่อหนังสือ: Data Mining For Dummies	$disSIM(cd, b_3^{cd}) = 1 - 0.692$	0.308
ISNB13: 9780123814791 ชื่อหนังสือ: Data mining: Concepts and Techniques	$disSIM(cd, b_4^{cd}) = 1 - 0.462$	0.538
⋮	⋮	⋮

ภาพที่ 19 การคำนวณคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชา “Data mining” และหนังสือที่เกี่ยวข้อง

หลังจากทำการคำนวณคะแนนความคล้ายคลึง และคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชา และหนังสือที่เกี่ยวข้องแล้ว เราสามารถทำการจัดอันดับหนังสือที่มีความคล้ายคลึงของเนื้อหากับเนื้อหาในคำอธิบายรายวิชา เพื่อแนะนำให้กับผู้ใช้งาน โดยรายการแนะนำหนังสือจะพิจารณาคะแนนความคล้ายคลึงที่มีค่ามากที่สุดไปยังน้อยที่สุด และพิจารณาการแนะนำหนังสือเป็นจำนวน N เล่ม โดยในการทดลองจะกำหนดค่า N เท่ากับ 10, 30, 50 และ 100 ตามลำดับ (Sohail, Siddiqui and Ali, 2013) ในทางกลับกัน เราสามารถสร้างรายการแนะนำหนังสือจากการจัดอันดับคะแนนความไม่คล้ายคลึงเพื่อให้ผู้ใช้ได้พิจารณาเกี่ยวกับหนังสือที่มีเนื้อหาที่ขยายหรือแตกต่างออกไปจากคำอธิบายรายวิชา

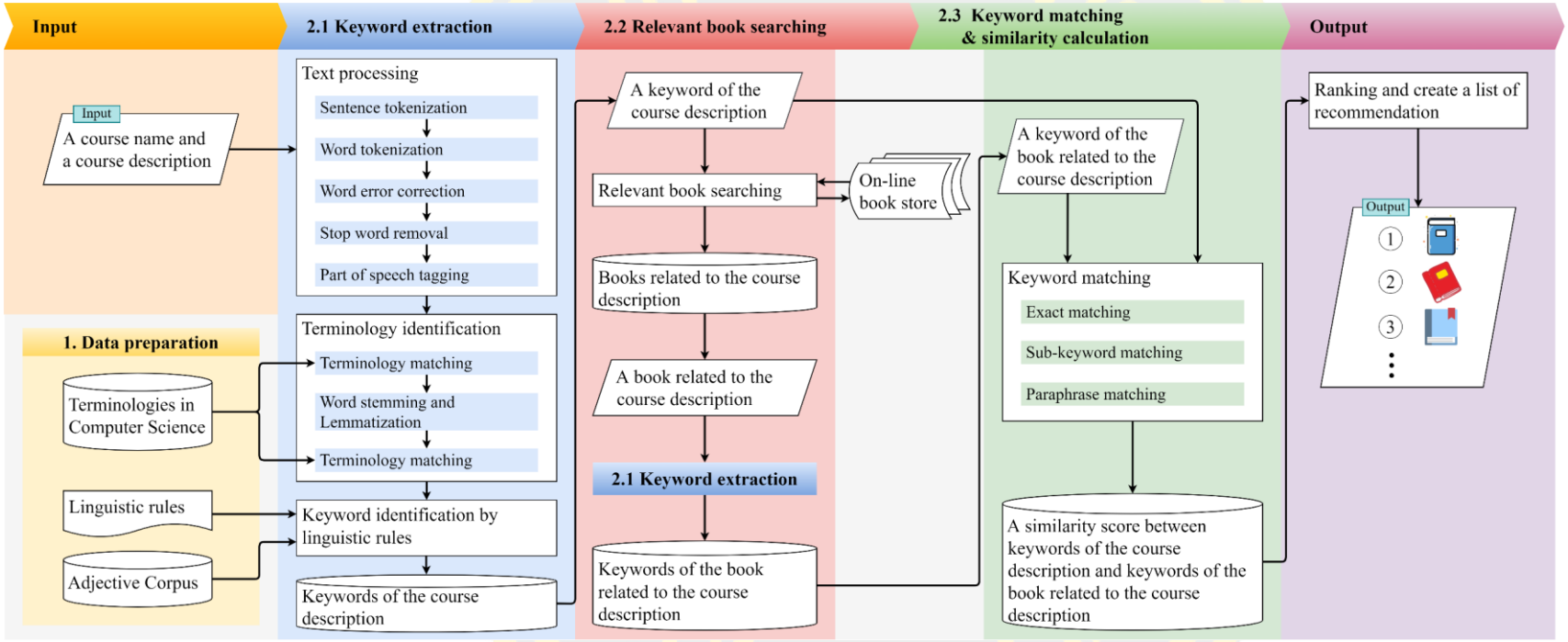
สรุปภาพรวมของระบบ *SBS* ได้ทำการเตรียมข้อมูลเบื้องต้นคือ 1) รวบรวมคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ มหาวิทยาลัยบูรพาเพื่อใช้เป็นข้อมูลรับเข้าจำนวน 68 วิชา 2) รวบรวมคำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์จำนวน 6,527 คำ และ 3) สร้างกฎทางภาษาศาสตร์เพื่อนำไปประยุกต์ใช้ในการสกัดคำสำคัญจากคำอธิบายรายวิชา (รวมทั้งการสกัดคำสำคัญจากหนังสือ) จากนั้น ทำการการสกัดคำสำคัญโดยประยุกต์ใช้กฎทางภาษาศาสตร์ที่สร้างขึ้นและนำคำสำคัญที่สกัดได้จากคำอธิบายรายวิชาไปค้นหาหนังสือที่เกี่ยวเนื่อง โดยจากการดำเนินการสามารถรวบรวมข้อมูลหนังสือทั้งหมดจำนวน 176,878 เล่ม ท้ายสุด ทำการคำนวณความคล้ายคลึงระหว่างคำสำคัญในคำอธิบายรายวิชาและคำสำคัญในหนังสือแต่ละเล่มที่เกี่ยวเนื่องกับคำอธิบายรายวิชาหนึ่ง ๆ ด้วยวิธีการเปรียบเทียบแบบตรงตัว (Exact matching) เพื่อสร้างรายการแนะนำหนังสืออ่านเพิ่มเติมสำหรับนิสิต/นักศึกษา และอาจารย์ผู้สอน ซึ่งสามารถลดเวลาในการเลือกหนังสืออ่านเพิ่มเติมจากในชั้นเรียนได้และสะดวก

บทที่ 4
ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ

ระบบ *SBS* ที่นำเสนอในบทที่ 3 สามารถแนะนำหนังสืออ่านเพิ่มเติมที่มีเนื้อหาคล้ายคลึงกับคำอธิบายรายวิชาให้แก่ผู้ใช้ได้ โดยเมื่อพิจารณาขั้นตอนการทำงานของระบบ *SBS* สังเกตได้ว่าการสกัดคำสำคัญจากคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้องเป็นขั้นตอนหลักของระบบ ซึ่งมีผลต่อประสิทธิภาพของระบบ ด้วยเหตุนี้ผู้วิจัยจึงทำการวิเคราะห์ประสิทธิภาพของการสกัดคำ และวิเคราะห์ผลลัพธ์ที่ได้จากระบบ *SBS* ซึ่งสังเกตได้ว่ายังสามารถพัฒนาประสิทธิภาพในการสกัดคำสำคัญของระบบ *SBS* เพิ่มเติมได้ เนื่องจาก 1) การระบุคำศัพท์เฉพาะที่ปรากฏขึ้นในประโยคหนึ่ง ๆ ของคำอธิบายรายวิชาหนึ่ง ๆ ยังทำได้ไม่ครอบคลุมทั้งหมด และ 2) คำศัพท์เฉพาะที่ระบุได้มีความซ้ำซ้อนและมีปริมาณมากเมื่อเทียบกับจำนวนเนื้อหาในคำอธิบายรายวิชาและในหนังสือ เป็นต้น

นอกจากข้างต้นผู้วิจัยได้ทำการวิเคราะห์เพิ่มเติมเกี่ยวกับวิธีการหาความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้องที่ประยุกต์ใช้วิธีการเปรียบเทียบแบบตรงตัว (Exact matching) โดยจากการวิเคราะห์ทำให้ทราบว่าวิธีการดังกล่าวให้ประสิทธิผลไม่ดีพอในการคำนวณคะแนนความคล้ายคลึง เนื่องจากวิธีการดังกล่าวไม่สามารถบ่งบอกถึงความคล้ายคลึงระหว่างบางส่วนของเนื้อหา (บางส่วนของคำสำคัญ) ที่ปรากฏในหนังสือกับเนื้อหา (หรือบางส่วนของเนื้อหา/บางส่วนของคำสำคัญ) ที่ปรากฏในคำอธิบายรายวิชาได้ นี้จึงนำมาซึ่งความไม่สมบูรณ์ของการหาความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง

จากทั้งสองประเด็นข้างต้น ผู้วิจัยจึงมีแนวคิดที่จะทำการปรับปรุงระบบ *SBS* เดิม โดยพัฒนาเพิ่มเติมในส่วนของการสกัดคำสำคัญและการหาความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง โดยเรียกระบบที่ปรับปรุงใหม่ว่า “ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ (An efficient Supplementary Book Suggestion system)” หรือเรียกโดยย่อว่า “ระบบ *SBS+*” โดยการดำเนินงานของระบบ *SBS+* ประกอบไปด้วย 2 ขั้นตอนหลักเหมือนกับระบบ *SBS* เดิม คือ 1) การเตรียมข้อมูลเบื้องต้น และ 2) การประมวลผลของระบบแนะนำ แต่รายละเอียดของการคำนวณแต่ละส่วนทั้งในส่วนที่ 1) และ 2) มีการพัฒนาที่ต่างจากเดิมค่อนข้างมาก ดังแสดงรายละเอียดการทำงานในภาพที่ 20



ภาพที่ 20 โครงสร้างระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ

จากภาพที่ 20 จะสังเกตได้ว่าระบบ *SBS+* มีการกำหนดอินพุตที่แตกต่างจากเดิม โดยกำหนดให้ผู้ใช้งานทำการป้อนชื่อรายวิชาและคำอธิบายรายวิชาเพื่อทำการค้นหารายการแนะนำหนังสือที่มีความเกี่ยวข้องกับเนื้อหาในคำอธิบายรายวิชามากที่สุด แต่อย่างไรก็ตามก่อนที่ผู้ใช้จะทำการป้อนข้อมูลอินพุต ระบบ *SBS+* ต้องทำการเตรียมข้อมูลเบื้องต้นที่ประกอบไปด้วย 1) การรวบรวมคำศัพท์เฉพาะ (Terminology) ทางด้านคอมพิวเตอร์เทคโนโลยี 2) การรวบรวมคำคุณศัพท์ (Adjective) ไว้เป็นคลังคำคุณศัพท์ (Adjective corpus) และ 3) การสร้างและปรับปรุงกฎทางภาษาศาสตร์ใหม่ ซึ่งการดำเนินการทั้งสามข้างต้นมีความแตกต่างจากระบบ *SBS* เดิม ดังแสดงรายละเอียดในตารางที่ 2

ตารางที่ 2 รายการการปรับปรุงการเตรียมข้อมูลเบื้องต้นของระบบ SBS+

ขั้นตอน	**ระบบ *SBS***	**ระบบ *SBS+***
การเตรียมข้อมูลเบื้องต้น	1. รวบรวมคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 68 รายวิชา จากมหาวิทยาลัยบูรพา 2. รวบรวมคำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์ จำนวน 6,527 คำ จาก 2 แหล่งข้อมูล 3. สร้างกฎทางภาษาศาสตร์	- 1. การรวบรวมคำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์ จำนวน 28,392 คำ จาก 8 แหล่งข้อมูล **2. การรวบรวมคำคุณศัพท์ จำนวน 4,998 คำ จาก 2 แหล่งข้อมูล** 3. การปรับปรุงและสร้างกฎทางภาษาศาสตร์ใหม่

4.1 การเตรียมข้อมูลเบื้องต้น

4.1.1 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี

จากหัวข้อ 3.1.2 เราทราบได้ว่าระบบ *SBS* ทำการรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจำนวน 6,527 คำ จาก 2 แหล่งข้อมูลคือ 1) พจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด (Oxford University)[15] และ 2) รายการคำศัพท์เฉพาะจากเว็บไซต์ Labautopedia[16] ซึ่งคำศัพท์เฉพาะดังกล่าวช่วยให้สามารถค้นหาคำศัพท์เฉพาะที่ปรากฏในคำอธิบายรายวิชาและคำอธิบายหนังสือได้ แต่อย่างไรก็ตามจากการวิเคราะห์ผลลัพธ์ ระบบ *SBS* ยังคงมีความผิดพลาดในการระบุถึงคำศัพท์เฉพาะ ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะเพิ่มการรวบรวมคำศัพท์เฉพาะให้มากขึ้นในระบบ *SBS+* โดยทำการรวบรวมคำศัพท์เฉพาะเพิ่มเติมจาก 6 แหล่งข้อมูลใหม่ ดังนี้

1. คำศัพท์เฉพาะที่เกี่ยวข้องกับเทคโนโลยีจากเว็บไซต์ TechTarget[17]
2. คำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจากเว็บไซต์ ComputerHope[18]
3. คำศัพท์เฉพาะที่เกี่ยวข้องกับคอมพิวเตอร์และอินเทอร์เน็ตจากเว็บไซต์ PC Glossary[19]
4. คำศัพท์เฉพาะที่เกี่ยวข้องทางด้านคอมพิวเตอร์จากเว็บไซต์ Glossary of Computer Related Terms[20]
5. คำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์ที่รวบรวมโดยวิกิพีเดีย (Wikipedia's Glossary of computer science)[21]
6. คำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจากเว็บไซต์ GCSE's Computer Science Glossary[22]

โดยจากการรวมคำศัพท์เฉพาะจาก 8 แหล่งข้อมูล จะทำให้สามารถรวบรวมคำศัพท์เฉพาะได้ทั้งสิ้น 28,392 คำ ซึ่งจากการเพิ่มขึ้นของการรวบรวมคำศัพท์เฉพาะจะช่วยให้สามารถระบุและสกัดคำสำคัญได้ดียิ่งขึ้น

---

[15] https://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975

[16] http://www.labautopedia.org/mw/List_of_programming_and_computer_science_terms

[17] https://whatis.techtarget.com/definitions/A

[18] https://www.computerhope.com/jargon.htm

[19] https://pc.net/glossary/

[20] http://www.math.utah.edu/~wisnia/glossary.html

[21] https://en.wikipedia.org/wiki/Glossary_of_computer_science

[22] https://www.computerscience.gcse.guru/glossary

4.1.2 การรวบรวมคำคุณศัพท์

คำคุณศัพท์ (Adjective) เป็นคำที่มักปรากฏขึ้นในเนื้อหาของคำอธิบายรายวิชา เช่น “fundamental”, “basic”, “famous”, “common” เป็นต้น ซึ่งในระบบ *SBS* จะได้ทำการรวมคำคุณศัพท์บางคำให้เป็นส่วนประกอบหนึ่งของคำสำคัญที่สกัดได้จากคำอธิบายรายวิชา แต่อย่างไรก็ตาม คำคุณศัพท์เหล่านั้นไม่ได้บ่งบอกใจความสำคัญของเนื้อหาจึงนำมาซึ่งความซ้ำซ้อน และความด้อยคุณภาพของคำสำคัญที่สกัดได้จากระบบ *SBS*

4.1.3 การปรับปรุงกฎทางภาษาศาสตร์

จากบทที่ 3 ระบบ *SBS* ได้ทำการสร้างกฎทางภาษาศาสตร์ 2 ประเภท คือ 1) กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะ และ 2) กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะ (หัวข้อ 3.1.3) เพื่อช่วยในการสกัดคำสำคัญจากคำอธิบายรายวิชาหนึ่งๆ และ/หรือหนังสือเล่มหนึ่ง ๆ (หัวข้อ 3.2.1.3) โดยจากตัวอย่างในภาพที่ 15 จะสังเกตได้ว่าคำสำคัญที่สกัดได้จากแต่ละประโยคที่ของคำอธิบายรายวิชา ยังคงมีความซ้ำซ้อน อาทิเช่น เมื่อพิจารณาประโยคที่ 1 คือ “(‘fundamental’, ‘*JJ*’), (‘concept’, ‘*NN*’), (‘of’, ‘*IN*’), (‘data mining’, ‘*TE*’)” ซึ่งสามารถสกัดคำสำคัญได้ 4 คำ คือ 1) “(‘data mining’, ‘*TE*’)” 2) “(‘data mining’, ‘*TE*’), (‘concept’, ‘*NN*’)” 3) “(‘fundamental’, ‘*JJ*’), (‘data mining’, ‘*TE*’)” และ 4) “(‘fundamental’, ‘*JJ*’), (‘concept’, ‘*NN*’)” ตามลำดับ โดยเมื่อพิจารณาและเปรียบเทียบคำสำคัญข้างต้น จะสังเกตได้ว่า

1. คำสำคัญที่ 1) เป็นส่วนประกอบของคำสำคัญที่ 2) และ 3) ที่ซึ่งสื่อถึงสิ่งเดียวกัน ด้วยเหตุนี้เราจึงสามารถลบคำสำคัญที่ 1) ออกจากการพิจารณาได้
2. คำสำคัญที่ 3) มีคำว่า “(‘fundamental’, ‘*JJ*’)” เป็นส่วนประกอบ โดยคำดังกล่าวไม่ได้บ่งบอกถึงใจความสำคัญอย่างมีนัยยะสำคัญ ดังนั้น จึงสามารถตัดคำดังกล่าวซึ่งเป็นคำคุณศัพท์ (โดยการเทียบเคียงกับคำคุณศัพท์ที่รวบรวมไว้ในหัวข้อ 4.1.2) ออกจากคำสำคัญได้ เมื่อตัดคำคุณศัพท์ดังกล่าวแล้ว จะทำให้คำสำคัญประกอบด้วย “(‘data mining’, ‘*TE*’)” เพียงคำเดียว ซึ่งเหมือนคำสำคัญที่ 1) ที่มีความซ้ำซ้อนกับคำสำคัญที่ 2) ดังนั้น จึงสามารถตัดคำสำคัญนี้ออกจากการพิจารณาได้
3. คำสำคัญที่ 4) “(‘fundamental’, ‘*JJ*’), (‘concept’, ‘*NN*’)” มีคำว่า “(‘fundamental’, ‘*JJ*’)” เป็นส่วนประกอบเช่นเดียวกับคำสำคัญที่ 3) ซึ่งสามารถตัดดังกล่าวออกจากคำสำคัญได้ โดยเมื่อตัดคำดังกล่าวแล้ว จะทำให้คำสำคัญที่ 4) มีความซ้ำซ้อนกับคำสำคัญที่ 2) ดังนั้น จึงสามารถตัดคำสำคัญนี้ออกจากการพิจารณาได้

จากทั้ง 3 ข้อสังเกตข้างต้น จะทำให้สามารถตัดคำสำคัญที่มีความซ้ำซ้อนและไม่ได้สื่อความหมายที่นอกเหนือจากคำสำคัญอื่น ๆ อย่างมีนัยยะสำคัญ โดยจะเหลือคำสำคัญเพียงแค่หนึ่งคำ คือ “(‘data mining’, ‘*TE*’), (‘concept’, ‘*NN*’)” ซึ่งบ่งบอกถึงแนวคิดของการทำเหมืองข้อมูล

จากข้างต้นผู้วิจัยจึงทำการปรับปรุงแก้ไขกฎทางภาษาศาสตร์เพื่อลดความซ้ำซ้อนในการสกัดคำสำคัญ โดยทำการปรับเปลี่ยนจากการพิจารณาคำศัพท์เฉพาะเป็นหลัก (เป็นอันดับแรก) จากนั้นทำการค้นหาคำข้างเคียงไปเป็นการพิจารณาการปรากฏขึ้นของคำสันธาน (Conjunction) และ/หรือคำบุพบท (Preposition) ในแต่ละหัวข้อของคำอธิบายรายวิชาหนึ่ง ๆ หรือ หนังสือเล่มหนึ่ง ๆ จากนั้นทำการแยกการพิจารณากลุ่มคำที่ปรากฏก่อนหน้าและปรากฏหลังจากคำคำสันธานและ/หรือคำบุพบทออกจากกัน รวมถึงการลบคำคุณศัพท์บางคำที่ไม่ได้บ่งบอกถึงใจความสำคัญออกจากการพิจารณา นั้นทำการรวมกลุ่มคำทั้งสองเข้าด้วยกันเพื่อระบุถึงคำสำคัญของหัวข้อนั้น ๆ โดยกฎทางภาษาศาสตร์ฉบับปรับปรุงจะถูกแสดงในส่วนภาคผนวก ข

4.2 การประมวลผลของระบบแนะนำ

เมื่อระบบ *SBS+* มีการรับข้อมูลรับคำอธิบายรายวิชาหนึ่ง ๆ จากผู้ใช้งาน ระบบ *SBS+* จะทำการประมวล 3 ขั้นตอน คือ 1) การสกัดคำสำคัญจากคำอธิบายรายวิชา 2) การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญหนึ่ง ๆ ในคำอธิบายรายวิชา (รวมทั้งการสกัดคำสำคัญจากหนังสือโดยใช้กระบวนการสกัดคำสำคัญเดียวกันกับขั้นตอนที่ 1) และ 3) การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา โดยทั้ง 3 ขั้นตอนนี้จะเป็นขั้นตอนที่เหมือนกับระบบ *SBS* เดิม แต่รายละเอียดการประมวลผลแต่ละขั้นตอนมีการพัฒนาเพิ่มเติม โดยแสดงรายการเปลี่ยนแปลงในภาพรวมดังตารางที่ 3

4.2.1 การสกัดคำสำคัญจากคำอธิบายรายวิชา

กระบวนการสกัดคำสำคัญจากคำอธิบายรายวิชาหนึ่ง ๆ ของระบบ *SBS+* ประกอบด้วย 3 ขั้นตอนต่อเนื่องกัน คือ 1) การประมวลผลข้อความ 2) การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี และ 3) การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุง (ดังแสดงในภาพที่ 24) โดยทั้ง 3 ขั้นตอนนี้จะเหมือนกับขั้นตอนในระบบ *SBS* เดิม แต่รายละเอียดการประมวลผลของแต่ละขั้นตอนมีการเปลี่ยนแปลงดังนี้

ตารางที่ 3 รายการการปรับปรุงการประมวลผลของระบบแนะนำ SBS+

ขั้นตอน	ระบบ *SBS*	ระบบ *SBS+*
1) การสกัดคำสำคัญจากคำอธิบายรายวิชา	1. การประมวลผลข้อความ 1.1 การกำจัดคำหยุด 1.2 การแก้ไขคำผิด 1.3 การตัดประโยค 1.4 การวิเคราะห์หน้าที่ของคำ 2. การระบุคำศัพท์เฉพาะ 2.1 การเปรียบเทียบหาคำศัพท์เฉพาะ 3. การระบุคำสำคัญโดยกฎทางภาษาศาสตร์	1. การประมวลผลข้อความ **1.1 การตัดประโยค** **1.2 การตัดคำ** **1.3 การแก้ไขคำผิด** **1.4 การกำจัดคำหยุด** **1.5 การวิเคราะห์หน้าที่ของคำ** 2. การระบุคำศัพท์เฉพาะ **2.1 การเปรียบเทียบหาคำศัพท์เฉพาะ (ครั้งที่ 1)** **2.2 การลดรูปของคำและการหารูปเดิมของคำ** **2.3 การเปรียบเทียบหาคำศัพท์เฉพาะ (ครั้งที่ 2)** **3. การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุง**
2) การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา	1. การค้นหาหนังสือและรวบรวมข้อมูลหนังสือบนเว็บไซต์ Barnes & Noble 2. การสกัดคำสำคัญจากหนังสือแต่ละเล่มโดยขั้นตอน 1)	1. การค้นหาหนังสือและรวบรวมข้อมูลหนังสือบนเว็บไซต์ Barnes & Noble 2. การสกัดคำสำคัญจากหนังสือแต่ละเล่มโดยขั้นตอน 1)
3) การคำนวณความเกี่ยวเนื่องของหนังสือที่ครอบคลุมเนื้อหาในคำอธิบายรายวิชา	1. การเปรียบเทียบแบบตรงตัว (Exact matching)	1. การเปรียบเทียบแบบตรงตัว (Exact matching) **2. การเปรียบเทียบแบบวลี (Paraphrase matching)** **3. การเปรียบเทียบแบบพิจารณาคำย่อย (Sub-keyword matching)**

4.2.1.1 การประมวลผลข้อความ (Text processing)

การประมวลผลข้อความของระบบ SBS+ ประกอบไปด้วย 5 ขั้นตอนต่อเนื่องกัน คือ 1) การตัดประโยค 2) การตัดคำ 3) การแก้ไขคำผิด 4) การกำจัดคำหยุด และ 5) การวิเคราะห์หน้าที่ของคำตามลำดับ ซึ่งจำนวนขั้นตอนการประมวลผลและลำดับของการประมวลผลมีการเปลี่ยนแปลงจากระบบ *SBS* เดิม (ดังแสดงในตารางที่ 3) การประมวลผลข้อความแบบใหม่จะช่วยให้การประมวลผลข้อความมีความสมบูรณ์มากขึ้น และสามารถลดความผิดพลาดในการประมวลข้อความได้ อาทิเช่น ระบบ *SBS* เดิมทำการกำจัดคำหยุดก่อนการแก้ไขคำผิด อาจทำให้กำจัดคำผิดที่ยังไม่ได้รับการแก้ไขออกไป ซึ่งอาจหมายถึงการกำจัดเนื้อหาที่สำคัญออกไปด้วยได้ นอกจากนี้การตัดประโยคควรจะอยู่ในการประมวลผลข้อความเป็นอันดับแรกเพราะข้อความจะถูกแยกส่วนออกมาเป็นแต่ละหัวข้อและตรงประเด็นมากขึ้น ช่วยลดความกำกวมของข้อความได้

ภาพที่ 21 แสดงตัวอย่างข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining” ซึ่งประกอบไปด้วยชื่อรายวิชา และคำอธิบายรายวิชา จากนั้นระบบ *SBS+* จะดำเนินการแยก/ตัดประโยค กล่าวคือ เมื่อพบ “.” (Full stop) หรือ “;” (Semi colon) จะทำการแยก/ตัดกลุ่มของคำที่ปรากฏในคำอธิบายรายวิชาออกเป็นส่วนย่อย ๆ ดังแสดงในส่วนของการตัดประโยคในภาพที่ 22 ซึ่งจะทำให้คำอธิบายรายวิชาถูกแบ่งออกเป็น 9 ส่วนย่อย

ชื่อรายวิชา	Data mining
คำอธิบายรายวิชา	The fundamental concepts of data mining; types of data for data mining; the famous techniques for data mining; pattern mining and association rules mining; classification; clustering; outlier analysis; anomaly detection; data mining tools

ภาพที่ 21 ข้อมูลรับเข้าคำอธิบายรายวิชา “Data mining”

ชื่อรายวิชา	: ‘‘Data mining’’
คำอธิบายรายวิชา	: ‘‘The fundamental concepts of data mining; types of data for data mining; the famous techniques for data mining; pattern mining and association rules mining; classifcation; clustering; outlier analysis; anomaly detection; data mining tools’’

การประมวลผลข้อความ

การตัดประโยค

tp_1 = {‘the fundamental concepts of data mining’}
tp_2 = {‘types of data for data mining’}
tp_3 = {‘the famous techniques for data mining’}
tp_4 = {‘pattern mining and association rules mining’}
tp_5 = {‘classifcation’}
tp_6 = {‘clustering’}
tp_7 = {‘outlier analysis’}
tp_8 = {‘anomaly detection’}
tp_9 = {‘data mining tools’}

การตัดคำ

tp_1 = {‘the’, ‘fundamental’, ‘concepts’, ‘of ’, ‘data’, ‘mining’}
tp_2 = {‘types’, ‘of’, ‘data’, ‘for’, ‘data’, ‘mining’}
tp_3 = {‘the’, ‘famous’, ‘techniques’, ‘for’, ‘data’, ‘mining’}
tp_4 = {‘pattern’, ‘mining’, ‘and’, ‘association’, ‘rules’, ‘mining’}
tp_5 = {‘classifcation’}
tp_6 = {‘clustering’}
tp_7 = {‘outlier’, ‘analysis’}
tp_8 = {‘anomaly’, ‘detection’}
tp_9 = {‘data’, ‘mining’, ‘tools’}

การแก้ไขคำผิด

tp_1 = {‘the’, ‘fundamental’, ‘concepts’, ‘of ’, ‘data’, ‘mining’}
tp_2 = {‘types’, ‘of’, ‘data’, ‘for’, ‘data’, ‘mining’}
tp_3 = {‘the’, ‘famous’, ‘techniques’, ‘for’, ‘data’, ‘mining’}
tp_4 = {‘pattern’, ‘mining’, ‘and’, ‘association’, ‘rules’, ‘mining’}
tp_5 = {‘classifcation’} ⇢ tp_5 = {‘classification’}
tp_6 = {‘clustering’}
tp_7 = {‘outlier’, ‘analysis’}
tp_8 = {‘anomaly’, ‘detection’}
tp_9 = {‘data’, ‘mining’, ‘tools’}

การวิเคราะห์หน้าที่ของคำ

tp_1 = {(‘fundamental’,‘JJ’), (‘concepts’,‘NNS’), (‘of ’,‘IN’), (‘data’,‘NN’), (‘mining’,‘NN’)}
tp_2 = {(‘types’,‘NNS’), (‘of’,‘IN’), (‘data’,‘NN’), (‘for’,‘IN’), (‘data’,‘NN’), (‘mining’,‘NN’)}
tp_3 = {(‘famous’,‘JJ’), (‘techniques’,‘NNS’), (‘for’,‘IN’), (‘data’,‘NN’), (‘mining’,‘NN’)}
tp_4 = {(‘pattern’,‘NN’), (‘mining’,‘NN’), (‘and’,‘CC’), (‘association’,‘NN’), (‘rules’,‘NNS’), (‘mining’,‘NN’)}
tp_5 = {(‘classification’,‘NN’)}
tp_6 = {(‘clustering’,‘NN’)}
tp_7 = {(‘outlier’,‘JJR’), (‘analysis’,‘NN’)}
tp_8 = {(‘anomaly’,‘JJ’), (‘detection’,‘NN’)}
tp_9 = {(‘data’,‘NN’), (‘mining’,‘NN’), (‘tools’,‘NNS’)}

การกำจัดคำหยุด

tp_1 = {‘~~the~~’, ‘fundamental’, ‘concepts’, ‘of ’, ‘data’, ‘mining’}
tp_2 = {‘types’, ‘of’, ‘data’, ‘for’, ‘data’, ‘mining’}
tp_3 = {‘~~the~~’, ‘famous’, ‘techniques’, ‘for’, ‘data’, ‘mining’}
tp_4 = {‘pattern’, ‘mining’, ‘and’, ‘association’, ‘rules’, ‘mining’}
tp_5 = {‘classification’}
tp_6 = {‘clustering’}
tp_7 = {‘outlier’, ‘analysis’}
tp_8 = {‘anomaly’, ‘detection’}
tp_9 = {‘data’, ‘mining’, ‘tools’}

ภาพที่ 22 การประมวลผลข้อความของระบบ SBS+ จากข้อมูลคำอธิบายรายวิชา “Data mining”

จากนั้นระบบ *SBS+* จะทำการแบ่งแต่ละประโยคออกเป็นส่วนย่อย โดยแบ่งออกเป็นหน่วยคำ กล่าวคือ เมื่อพบ “ ” (“White space) จะทำการแยก/ตัดคำออกมาเป็นหน่วยคำหนึ่ง ๆ ดังแสดงในส่วนการตัดคำในภาพที่ 22 เช่น ประโยค (หัวข้อ) ที่ 1 คือ tp_1 = “the fundamental concepts of data mining” ถูกตัดคำเป็น tp_1 = {‘the’, ‘fundamental’, ‘concepts’, ‘of’, ‘data’, ‘mining’} ต่อมาระบบ *SBS+* จะทำการแก้ไขการสะกดผิดโดยประยุกต์ใช้วิธีการ SMC (*S*pelling *M*istake *C*orrection) (Gupta, 2015) ดังแสดงในส่วนการสะกดคำผิดในภาพที่ 22 ซึ่งแสดงให้เห็นถึงการแก้คำผิดของคำในประโยคที่ 5 คือ $w_1^{tp_5}$ = “classifcation” ให้กลายเป็น $w_1^{tp_5}$ = {“classification”}

ต่อมาขั้นตอนการกำจัดคำหยุดถูกประยุกต์ใช้เพื่อกำจัดคำที่ไม่มีนัยสำคัญ โดยจะไม่กำจัดคำหยุด 2 ประเภท คือ คำบุพบท (Preposition) และ คำสันธาน (Conjunction) โดยในส่วนการกำจัดคำในภาพที่ 22 แสดงการกำจัดคำหยุด “the” ในหัวข้อที่ 1) และหัวข้อที่ 3) ตามลำดับ ท้ายสุดการระบุ/กำกับหน้าที่ของแต่คำประยุกต์ใช้ Stanford part of speech tagger (Toutanova, Klein, Manning & Singer, 2003) โดยในส่วนการวิเคราะห์หน้าที่ของคำในภาพที่ 22 แสดงการกำหนดหน้าที่ของแต่ละคำในแต่ละประโยค ตัวอย่างเช่น เมื่อทำการระบุหน้าที่ของคำในประโยคที่ 1 จะได้เป็น tp_1 = {(‘fundamental’, ‘*JJ*’), (‘concepts’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’)}

เมื่อคำอธิบายรายวิชาผ่านขั้นตอนการประมวลผลข้อความ จะได้ผลลัพธ์เป็นเซตของหัวข้อในคำอธิบายรายวิชา คือ $cd = \{tp_1, tp_2, \cdots, tp_n\}$ ซึ่งในแต่ละหัวข้อของคำอธิบายรายวิชา จะได้เซตของคำในหัวข้อนั้น ๆ คือ $tp_i = \{w_1^{tp_i}, w_2^{tp_i}, \cdots, w_m^{tp_i}\}$ ซึ่งแต่ละคำจะมีหน้าที่ของคำแต่ละชนิดที่กำกับไว้ ดังนั้น ส่วนประกอบของคำอธิบายรายวิชาหนึ่ง ๆ จะประกอบไปด้วย

$$cd = \begin{Bmatrix} \{(w_1^{tp_1}, tag), (w_2^{tp_1}, tag), \cdots, (w_p^{tp_1}, tag)\}, \\ \{(w_1^{tp_2}, tag), (w_2^{tp_2}, tag), \cdots, (w_q^{tp_2}, tag)\}, \cdots \\ \{(w_1^{tp_n}, tag), (w_2^{tp_n}, tag), \cdots, (w_r^{tp_n}, tag)\} \end{Bmatrix}$$

โดยจากคำอธิบายรายวิชา “Data mining” มีจำนวนทั้งหมด 9 หัวข้อ โดยแต่ละหัวข้อมีจำนวนคำ 6, 6, 6, 1, 1, 2, 2 และ 3 ตามลำดับ ดังแสดงในส่วนการวิเคราะห์หน้าที่ของคำในภาพที่ 22

4.2.1.2 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี (Terminology identification)

หลังจากการประมวลผลข้อความระบบ *SBS+* จะทำการค้นหา/ระบุคำศัพท์เฉพาะในคำอธิบายรายวิชาโดยใช้วิธีการ *N*-gram (Lopez-Gazpio *et al.*, 2019) เพื่อเปรียบเทียบคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชากับคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้า (ในหัวข้อที่ 4.1.2) ซึ่งจากระบบ *SBS* ได้ทำการพิจารณาคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชาระบุคำศัพท์และเปรียบเทียบคำคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้าเพียงครั้งเดียวเท่านั้น และไม่มีการประมวลผลอื่น ๆ เพิ่มเติม จึงส่งผลให้ไม่สามารถระบุคำศัพท์เฉพาะได้ครบถ้วน

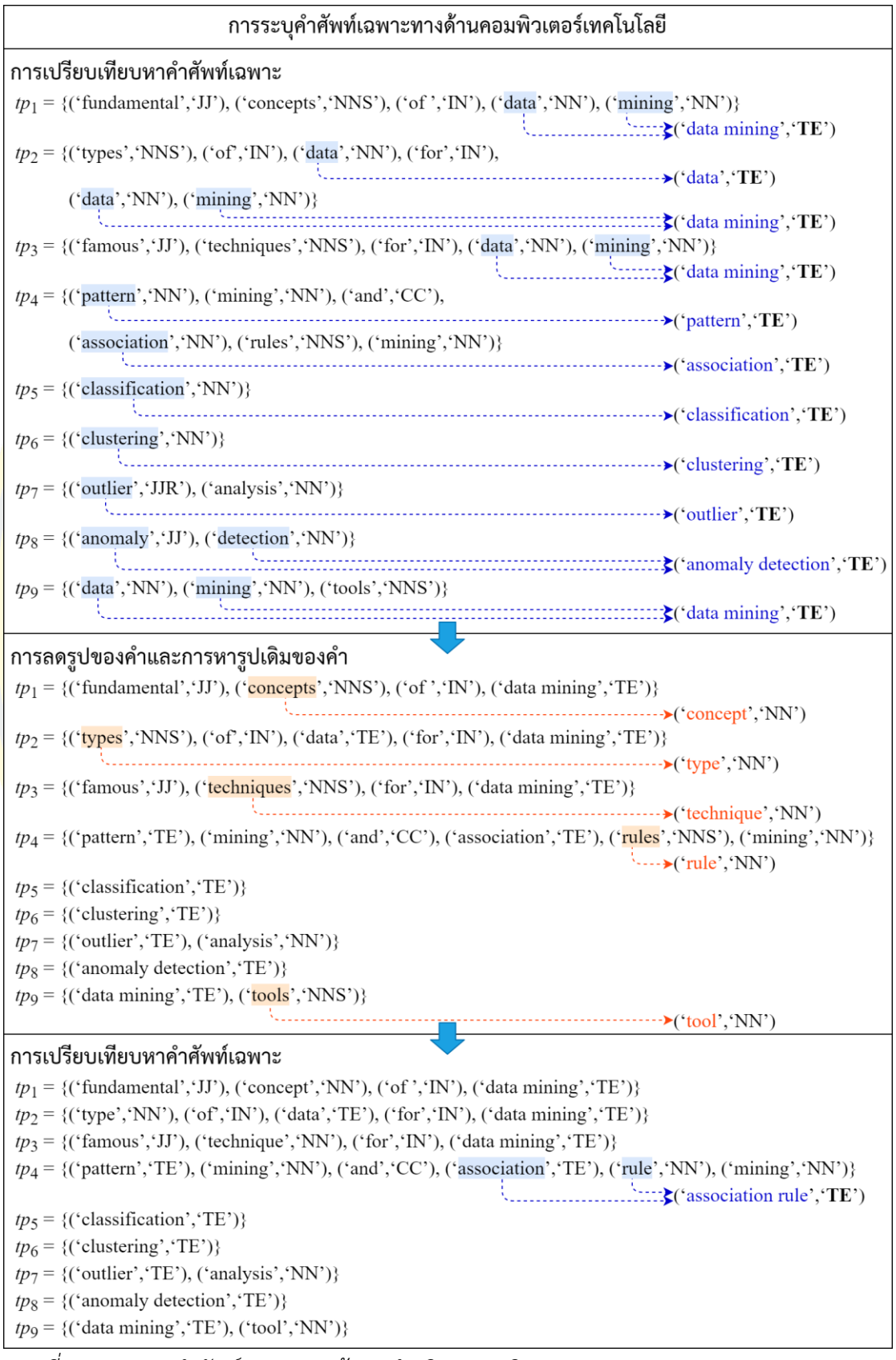
ด้วยเหตุนี้ ในระบบ *SBS+* จึงทำการปรับปรุงค้นหา/ระบุคำศัพท์เฉพาะ โดยปรับเปลี่ยนการประมวลเป็น 3 ขั้นตอน คือ 1) การเปรียบเทียบหาคำศัพท์เฉพาะระหว่างคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชากับคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้า 2) การลดรูปของคำและการหารูปเดิมของคำ (Word stemming and lemmatization) (Balakrishnan and Lloyd-Yemoh, 2014) และ 3) การเปรียบเทียบหาคำศัพท์เฉพาะระหว่างคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชากับคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้า ซึ่งจากขั้นตอนทั้ง 3 ข้างต้น จะสังเกตได้ว่าขั้นตอนที่ 1 และ 3 เป็นขั้นตอนเดียวกัน แต่ดำเนินการก่อนและหลังการการลดรูปของคำและการหารูปเดิมของคำ เพื่อให้สามารถเพิ่มการค้นหา/ระบุคำสำคัญได้

ภาพที่ 23 แสดงตัวอย่างการค้นหา/ระบุคำศัพท์เฉพาะจากคำอธิบายรายวิชา “Data mining” โดยในการเปรียบเทียบหาคำศัพท์เฉพาะระหว่างคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชากับคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้า จะสามารถระบุได้ว่ามีคำหนึ่ง ๆ อาทิเช่น

1. (‘data’, ‘*NN*’) ในหัวข้อที่ 2
2. (‘pattern’, ‘*NN*’) ในหัวข้อที่ 4
3. (‘association’, ‘*NN*’) ในหัวข้อที่ 4
4. (‘classification’, ‘*NN*’) ในหัวข้อที่ 5
5. (‘clustering’, ‘*NN*’) ในหัวข้อที่ 6
6. (‘outlier’, ‘*JJR*’) ในหัวข้อที่ 7

และกลุ่มของคำ คือ

1. (‘data’, ‘*NN*’), (‘mining’, ‘*NN*’) ในหัวข้อที่ 1, 2, 3 และ 9
2. (‘anomaly’, ‘*JJ*’), (‘detection’, ‘*NN*’) ในหัวข้อที่ 8



ภาพที่ 23 การระบุคำศัพท์เฉพาะจากข้อมูลคำอธิบายรายวิชา "Data mining"

ขั้นตอนถัดไปจะเป็นการลดรูป/หารูปเดิมของคำหนึ่ง ๆ เพื่อปรับคำนั้น ให้อยู่ในรูปแบบของรากศัพท์ ซึ่งในคำอธิบายรายวิชา “Data mining” มีคำที่จะต้องผ่านการลดรูป/หารูปเดิมของคำดังนี้

i) (‘*concept<u>s</u>*’, ‘NNS’) ในหัวข้อที่ 1 ถูกเปลี่ยนแปลงเป็น (‘*concept*’, ‘*NN*’)
ii) (‘*type<u>s</u>*’, ‘*NNS*’)” ในหัวข้อที่ 2 ถูกเปลี่ยนแปลงเป็น (‘*type*’, ‘*NN*’)
iii) (‘*technique<u>s</u>*’, ‘*NNS*’) ในหัวข้อที่ 3 ถูกเปลี่ยนแปลงเป็น (‘*technique*’, ‘*NN*’)
iv) (‘*rule<u>s</u>*’, ‘*NNS*’) ในหัวข้อที่ 4 ถูกเปลี่ยนแปลงเป็น (‘*rule*’, ‘*NN*’)
v) (‘*tool<u>s</u>*’, ‘*NNS*’) ในหัวข้อที่ 9 ถูกเปลี่ยนแปลงเป็น (‘*tool*’, ‘*NN*’)

จากการดำเนินการดังกล่าวจะช่วยให้สามารถทราบถึงรูปคำเดิมของคำ (แสดงในส่วนการการลดรูปของคำ/การหารูปเดิมของคำในภาพที่ 23) และช่วยให้ระบุถึงคำศัพท์เฉพาะได้แม่นยำมากขึ้น ด้วยการประมวลผลขั้นตอนการเปรียบเทียบหาคำศัพท์เฉพาะระหว่างคำหนึ่ง ๆ (หรือกลุ่มของคำหนึ่ง ๆ) ในคำอธิบายรายวิชากับคำศัพท์เฉพาะที่ได้รวบรวมไว้ก่อนหน้าถูกประยุกต์ใช้อีกครั้ง โดยจากการประมวลผลจะทำให้สามารถระบุได้ว่ามีหนึ่งกลุ่มคำ คือ (‘association’, ‘*TE*’), (‘rule’, ‘*NN*’) ในหัวข้อที่ 4 เป็นคำศัพท์เฉพาะและระบุหน้าที่ของคำเป็น ‘*TE*’ (แสดงในส่วนการเปรียบเทียบหาคำศัพท์เฉพาะด้านล่างสุดในภาพที่ 23)

จากการประมวลผลทั้ง 3 ขั้นตอนข้างต้น จะทำให้สังเกตได้ว่าการลดรูป/หารูปเดิมของคำและการเปรียบเทียบหาคำศัพท์ซ้ำ 2 ครั้ง จะช่วยเพิ่มความแม่นยำของการระบุคำศัพท์เฉพาะที่ปรากฏขึ้นในคำอธิบายรายวิชาได้

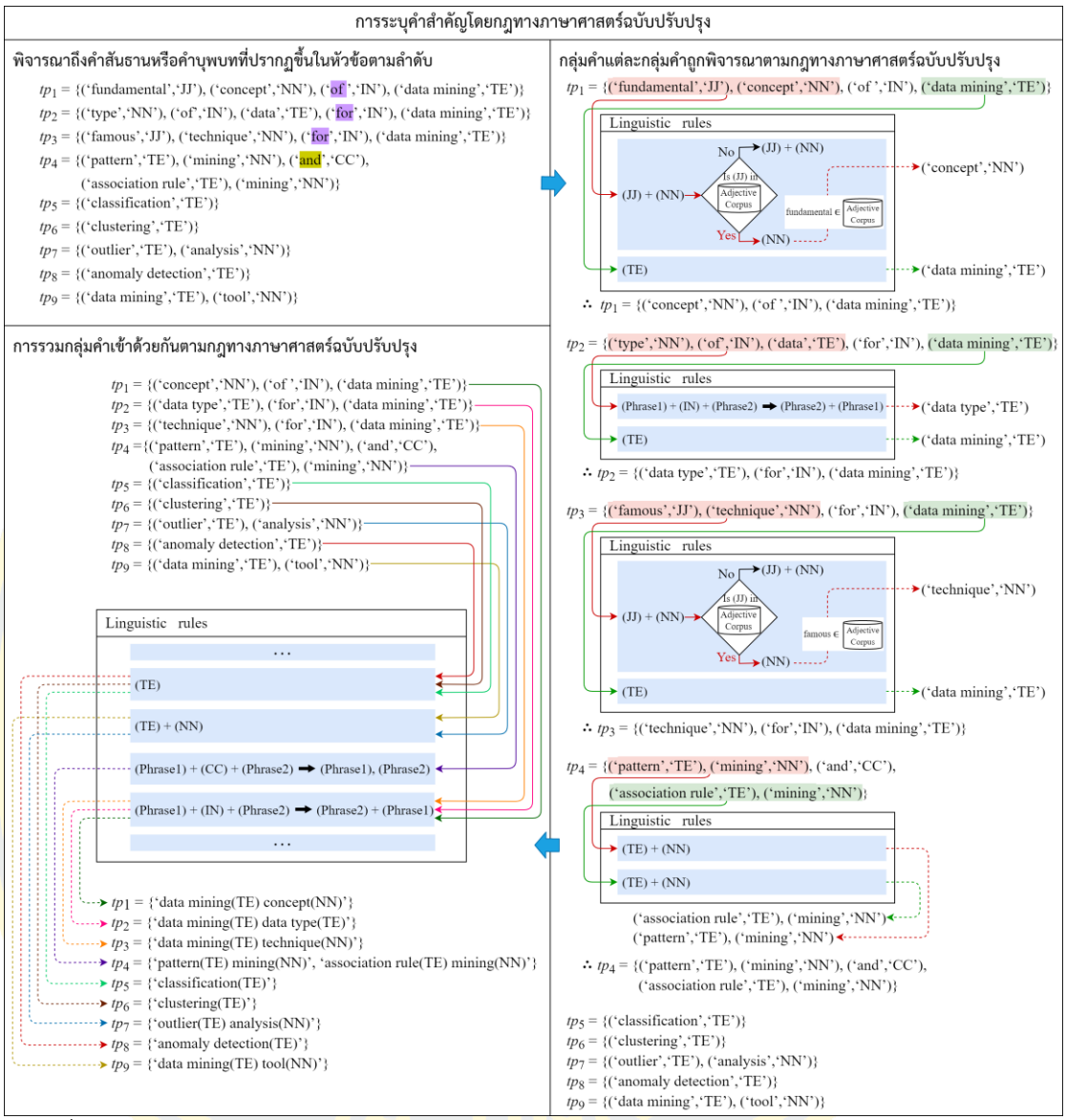
4.2.1.3 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุง

ในขั้นตอนนี้จะทำการระบุถึงคำสำคัญซึ่งอาจจะเป็นคำศัพท์เฉพาะที่ระบุจากขั้นตอนก่อนหน้า กลุ่มคำ หรือ นามวลี โดยคำเหล่านี้สามารถชี้เฉพาะเจาะจงถึงเนื้อหาสำคัญได้ กฎทางภาษาศาสตร์ฉบับปรับปรุง (หัวข้อ 4.1.4) จะถูกนำมาประยุกต์ใช้ในการระบุคำสำคัญของ*ระบบ SBS+*

ภาพที่ 24 แสดงการประยุกต์ใช้กฎทางภาษาศาสตร์ฉบับปรับปรุงกับคำอธิบายรายวิชา “Data mining” ตัวอย่างเช่น การประยุกต์ใช้กฎทางภาษาศาสตร์กับคำในหัวข้อที่ 1 ที่ประกอบด้วย {(‘fundamental’, ‘*JJ*’), (‘concept’, ‘*NNS*’), (‘of’, ‘*IN*’), (‘data mining’, ‘*TE*’)} โดยเริ่มจากการค้นหาคำบุพบท (Preposition) ที่ปรากฏขึ้นในหัวข้อ โดยจากการค้นหาพบว่ามี (‘of’, ‘*IN*’) ซึ่งเป็นคำบุพบท จากนั้นจะทำการแบ่งคำในหัวข้อที่ 1 ออกเป็น 2 กลุ่ม คือ 1) กลุ่มคำที่ปรากฏหน้าคำบุพบทคือ {(‘fundamental’, ‘*JJ*’), *(*‘concept’, ‘*NNS*’)} และ 2) กลุ่มคำที่ที่ปรากฏหลังคำบุพบท

คือ {(‘data mining’, ‘*TE*’)} ตามลำดับ จากนั้นจะทำการประยุกต์ใช้กฎทางภาษาศาสตร์ฉบับปรับปรุงกับแต่ละกลุ่มคำ โดยทำการประยุกต์ใช้ กฎทางภาษาศาสตร์ “Adjective (JJ) + Noun (NN) → Noun (NN)” กับกลุ่มคำที่ 1 เนื่องจาก (‘fundamental’, ‘*JJ*’) ซึ่งเป็นคำคุณศัพท์ที่ไม่บ่งบอกถึงนัยยะสำคัญ (กล่าวคือคำคุณศัพท์ดังกล่าวปรากฏในคลังคุณศัพท์ที่รวบรวมไว้ก่อนหน้า—ในหัวข้อ 4.1.2) ดังนั้นจะระบุคำสำคัญจากกลุ่มคำที่ 1 ได้เป็น {(‘concept’, ‘*NN*’)} จากนั้นทำการพิจารณากลุ่มคำที่ 2) แต่เนื่องด้วยกลุ่มคำดังกล่าวมีเพียงคำศัพท์เฉพาะปรากฏเพียงคำเดียวจึงไม่ต้องพิจารณาตามกฎทางภาษาศาสตร์ และคงรูปคำเดิมไว้ จากนั้นทำรวมกลุ่มคำทั้งสองเข้าด้วยกันด้วยการประยุกต์ใช้กฎ “(Phrase1) + Preposition (*IN*) + (Phrase2)” → “(Phrase2) (Phrase1)” จะทำให้ได้ “data mining concept” ซึ่งจัดเก็บอยู่ในรูปแบบ “(‘data mining’, ‘*TE*’) (‘concept’, ‘*NN*’)” เป็นคำสำคัญของหัวข้อที่ 1 ในคำอธิบายรายวิชา

ในส่วนของหัวข้อที่ 2 ที่ประกอบด้วย {(‘type’, ‘*NN*’), (‘of’, ‘*IN*’), (‘data’, ‘*TE*’), (‘for’, ‘*IN*’), (‘data mining’, ‘*TE*’)} พบ (‘for’, ‘*IN*’) เป็นคำบุพบทที่ปรากฏขึ้นจึงทำการแบ่งคำออกเป็น 2 กลุ่มคำ คือ 1) {(‘type’, ‘*NN*’), (‘of’, ‘*IN*’), (‘data’, ‘*TE*’)} และ 2) {(‘data mining’, ‘*TE*’)} ตามลำดับ จากนั้นทำการพิจารณากลุ่มคำที่ 1 ปรากฏว่ายังคงพบ (‘of’, ‘*IN*’) เป็นคำบุพบทปรากฏจึงทำการแบ่งคำในกลุ่มคำออกเป็น 2 กลุ่มคำที่ 1 ออกเป็น 2 กลุ่มคำย่อย คือ 1.1) {(‘type’, ‘*NN*’)} และ 1.2) {(‘data’, ‘*TE*’)} เมื่อไม่พบคำบุพบทในทั้งสองกลุ่มคำย่อยจะทำรวมกลุ่มคำย่อยทั้งสองเข้าด้วยกัน ด้วยการประยุกต์ใช้กฎ “(Phrase1) + Preposition (*IN*) + (Phrase2)” → “(Phrase2) (Phrase1)” จะทำให้ได้ “data type” ซึ่งจัดเก็บอยู่ในรูปแบบ “(‘data type’, ‘*TE*’)” เป็นคำสำคัญของกลุ่มคำที่ 1 จากนั้นจะทำการรวมกลุ่มคำทั้งสองเข้าด้วยกัน ด้วยการประยุกต์ใช้กฎ “(Phrase1) + Preposition (*IN*) + (Phrase2)” → “(Phrase2) (Phrase1)” จะทำให้ได้ “data mining data type” ซึ่งจัดเก็บอยู่ในรูปแบบ “(‘data mining’, ‘*TE*’) (‘data type’, ‘*TE*’)” เป็นคำสำคัญของหัวข้อที่ 2 สำหรับส่วนของหัวข้อที่ 3 ถึง 9 จะถูกประมวลผลด้วยวิธีการเดียวกันซึ่งจะทำให้ผลลัพธ์ดังส่วนล่างซ้ายของภาพที่ 24 โดยในหัวข้อที่ 4 จะมีคำสำคัญ 2 คำ เนื่องจากมีคำ (‘and’, ‘*CC*’) เชื่อมระหว่างคำสำคัญทั้งสอง โดยผลลัพธ์ที่ได้จากการสกัดคำสำคัญจากคำอธิบายรายวิชา “Data mining” จะได้คำสำคัญทั้งสิ้น 10 คำดังแสดงในภาพที่ 25 ที่ซึ่งสามารถเขียนอธิบายเกี่ยวกับคำอธิบายรายวิชาหนึ่ง ๆ ในเชิงคณิตศาสตร์ได้เป็น $cd = \{tp_1, tp_2, \cdots, tp_n\}$ โดยในแต่ละหัวข้อหนึ่ง ๆ จะมีเซตของคำสำคัญที่เกี่ยวเนื่องกับหัวข้อนั้น ๆ คือ $K^{tp_i} = \left\{k_{tp_{i,1}}^{cd}, k_{tp_{i,2}}^{cd}, \cdots, k_{tp_{i,m}}^{cd}\right\}$



ภาพที่ 24 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์ฉบับปรับปรุงกับคำอธิบายรายวิชา "Data mining"

ผลลัพธ์จากการสกัดคำสำคัญจากคำอธิบายรายวิชา

tp_1 = {'data mining(TE) concept(NN)'}
tp_2 = {'data mining(TE) data type(TE)'}
tp_3 = {'data mining(TE) technique(NN)'}
tp_4 = {'pattern(TE) mining(NN)', 'association rule(TE) mining(NN)'}
tp_5 = {'classification(TE)'}
tp_6 = {'clustering(TE)'}
tp_7 = {'outlier(TE) analysis(NN)'}
tp_8 = {'anomaly detection(TE)'}
tp_9 = {'data mining(TE) tool(NN)'}

ภาพที่ 25 ผลลัพธ์จากการสกัดคำสำคัญจากคำอธิบายรายวิชาของระบบ SBS+

4.2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา

การค้นหาหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาจะดำเนินการหลังจากทำการสกัดคำสำคัญแล้ว (หัวข้อที่ 4.2.1) โดยขั้นตอนนี้จะมีการประมวลผลเช่นเดียวกับระบบ *SBS* เดิม (หัวข้อ 3.2.2) โดยจากการพิจารณา 10 คำสำคัญในภาพที่ 25 และใช้คำดังกล่าวเป็นคำค้นหาจะทำให้ได้หนังสือที่มีความเกี่ยวเนื่องเป็นจำนวนทั้งสิ้น 384 เล่ม จากนั้นทำการสกัดคำสำคัญจากข้อมูลคำอธิบายหนังสือและสารบัญหนังสือแต่ละเล่ม เพื่อทราบถึงเนื้อหาสำคัญที่อธิบายในหนังสือแต่ละเล่ม

4.2.3 การคำนวณความเกี่ยวเนื่องของหนังสือที่ครอบคลุมเนื้อหาในคำอธิบายรายวิชา

ขั้นตอนสุดท้ายของระบบ *SBS+* คือ การคำนวณความเกี่ยวเนื่องของหนังสือเล่มหนึ่ง ๆ กับคำอธิบายรายวิชาหนึ่ง ๆ โดยพิจารณาจากคะแนนความคล้ายคลึงของคำสำคัญที่ปรากฏในคำอธิบายรายวิชาและคำสำคัญที่ปรากฏในหนังสือ จากนั้นทำการจัดอันดับหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาตามคะแนนความคล้ายคลึง โดยเรียงลำดับจากหนังสือที่มีความเกี่ยวเนื่องมากที่สุดไปยังหนังสือที่มีความเกี่ยวเนื่องน้อยที่สุด

จากหัวข้อที่ 4.2.1 และ 4.2.2 แต่ละคำอธิบายรายวิชา cd ประกอบด้วยเซตของหัวข้อของเนื้อหาซึ่งสามารถอธิบายได้ในรูปแบบ $cd = \{tp_1, tp_2, \cdots, tp_n\}$ ซึ่งแต่ละ $tp_i \in cd$ จะถูกนำไปสกัดคำสำคัญจะได้เซตของคำสำคัญ $K^{tp_i} = \{k_{tp_{i,1}}^{cd}, k_{tp_{i,2}}^{cd}, \cdots, k_{tp_{i,m}}^{cd}\}$ จากนั้นนำคำสำคัญหนึ่ง ๆ ในคำอธิบายรายวิชาไปค้นหาหนังสือที่เกี่ยวเนื่อง จะได้เซตของหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาทั้งหมดเป็น $B^{cd} = \{b_1^{cd}, b_2^{cd}, \cdots, b_x^{cd}\}$ ซึ่งหนังสือแต่ละเล่มจะถูกนำไปสกัดคำสำคัญจะได้เซตคำสำคัญ $K^{b_i^{cd}} = \left\{k_1^{b_i^{cd}}, k_2^{b_i^{cd}}, \cdots, k_z^{b_i^{cd}}\right\}$

ดังนั้น การคำนวณคะแนนความคล้ายคลึงจะพิจารณาจากการเปรียบเทียบความคล้ายคลึงระหว่างคำสำคัญ $k_{tp_{p,q}}^{cd}$ หนึ่ง ๆ ในเซตของคำสำคัญ K^{tp_p} ของคำอธิบายรายวิชา cd กับคำสำคัญ $k_j^{b_i^{cd}}$ หนึ่ง ๆ ในเซตคำสำคัญ $K^{b_i^{cd}}$ ที่เกี่ยวเนื่องกับคำอธิบายรายวิชาโดยใช้วิธีการเปรียบเทียบแบบตรงตัว (Exact matching), การเปรียบเทียบแบบวลี (Paraphrase matching) และการเปรียบเทียบแบบพิจารณาคำย่อย (Sub-keyword matching) (Gomaa and Fahmy, 2013) โดยสามารถอธิบายได้ดังนี้

4.2.3.1 วิธีการเปรียบเทียบแบบตรงตัว (Exact matching)

วิธีการเปรียบเทียบแบบตรงตัวเป็นการเปรียบเทียบในระดับของความเหมือนกันของตัวอักขระและลำดับการปรากฏของอักขระ ซึ่งคะแนนที่ได้จากการเปรียบเทียบจะมีค่าเป็น 0 หรือ 1 โดยคะแนนจะเป็น 1 ก็ต่อเมื่อคำสำคัญทั้งสองคำเหมือนกันทั้งอักขระและลำดับของการปรากฏของอักขระ และ คะแนนเป็น 0 ก็ต่อเมื่อ คำสำคัญทั้งสองคำแตกต่างกัน

4.2.3.2 วิธีการเปรียบเทียบแบบวลี (Paraphrase matching)

การเปรียบเทียบแบบวลีเป็นการเปรียบเทียบคำสำคัญ 2 คำ ที่ซึ่งแต่ละคำสำคัญจะมีมากกว่า 2 พยางค์ โดยพิจารณาการสลับตำแหน่งคำที่นำมาเปรียบเทียบ ซึ่งเริ่มต้นจากการวนสลับตำแหน่งในทุก ๆ ตำแหน่งของคำสำคัญที่นำมาเปรียบเทียบจนครบ แล้วนำรายการของคำสำคัญที่ได้จากการสลับตำแหน่งมาเปรียบเทียบกับคำสำคัญตั้งต้น เช่น การเปรียบเทียบระหว่างคำว่า “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” (คำสำคัญตั้งต้น) และ “(‘mining’, ‘*NN*’) (‘association rule’, ‘*TE*’)” (คำสำคัญเปรียบเทียบ) ในวิธีการนี้จะนำคำสำคัญเปรียบเทียบมาทำการสลับตำแหน่งของคำ ซึ่งได้ผลลัพธ์ดังนี้ “(‘mining’, ‘*NN*’) (‘association rule’, ‘*TE*’)” และ “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” ซึ่งจะเห็นว่า “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” (คำสำคัญตั้งต้น) ตรงกับคำที่อยู่ในรายการของคำสำคัญที่ได้จากการสลับตำแหน่งคือคำว่า “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” ดังนั้น จึงสรุปได้ว่าการเปรียบเทียบระหว่างคำสำคัญ “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” (คำสำคัญตั้งต้น) และ “(‘mining’, ‘*NN*’) (‘association rule’, ‘*TE*’)” (คำสำคัญเปรียบเทียบ) ทั้งสองคำเหมือนกัน ซึ่งคะแนนที่ได้จากการเปรียบเทียบจะมีค่าเป็น 0 หรือ 1 โดยคะแนนจะเป็น 1 ก็ต่อเมื่อคำสำคัญทั้งสองคำเหมือนกัน และคะแนนเป็น 0 ก็ต่อเมื่อคำสำคัญทั้ง 2 คำแตกต่างกัน

4.2.3.3 วิธีการเปรียบเทียบแบบพิจารณาคำย่อย (Sub-keyword matching)

วิธีการเปรียบเทียบแบบพิจารณาคำย่อยเป็นการเปรียบเทียบคำสำคัญที่เหมือนกันบางส่วน ซึ่งแบ่งออกเป็น 3 กรณีคือ

1. หากคำสำคัญในคำอธิบายรายวิชาปรากฏเป็นบางส่วนของคำสำคัญในหนังสือที่เกี่ยวเนื่อง จะให้คะแนนเท่ากับ 1
2. หากบางส่วนของคำสำคัญในคำอธิบายรายวิชาเป็นบางส่วนของคำสำคัญหรือเหมือนกับคำสำคัญในหนังสือที่เกี่ยวเนื่อง จะให้คะแนนเท่ากับ 0.5

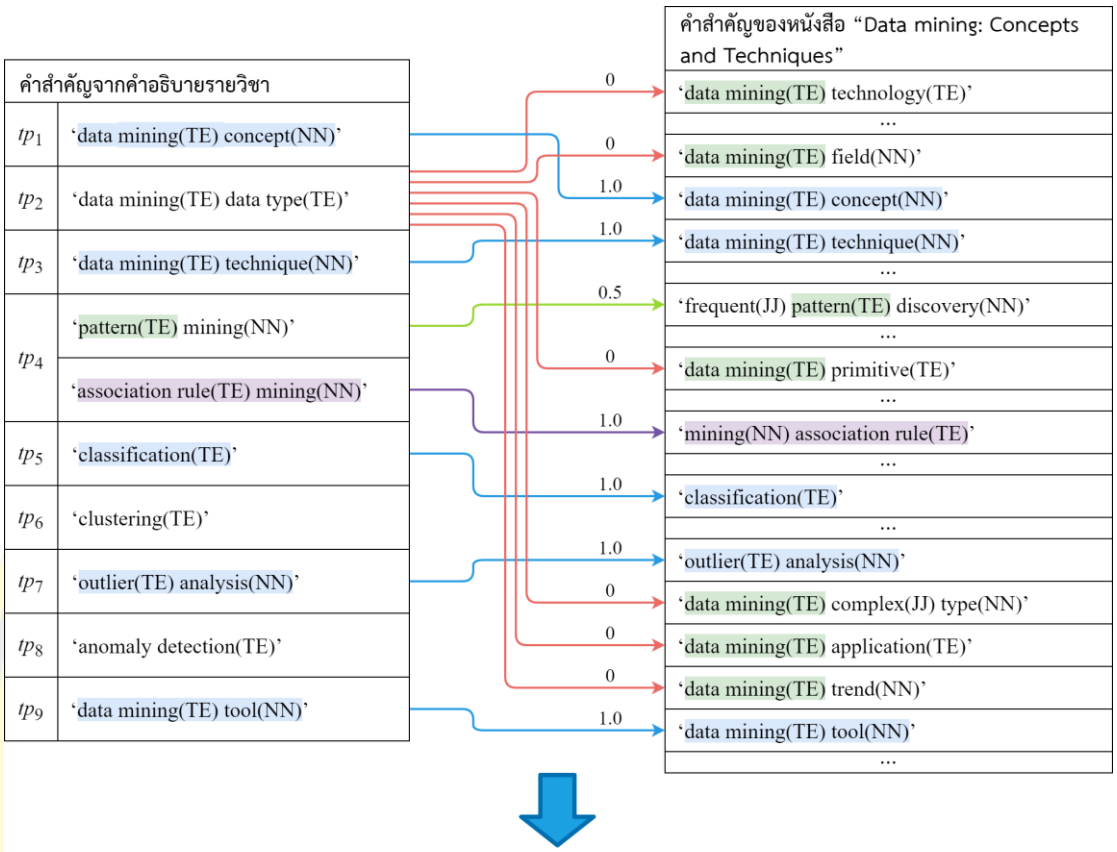
3. หากบางส่วนของคำสำคัญในคำอธิบายรายวิชาเหมือนกันกับบางส่วนของคำสำคัญในหนังสือที่เกี่ยวเนื่อง จะให้คะแนนเท่ากับ 0.5

ซึ่งจากทั้ง 3 กรณีข้างต้นสามารถอธิบายเชิงคณิตศาสตร์ได้ดังสมการ (3) ดังนี้

$$sim\left(k_{tp_{p,q}}^{cd}, k_j^{b_i^{cd}}\right) = \begin{cases} 1, & k_{tp_{p,q}}^{cd} \subset k_j^{b_i^{cd}} \\ 0.5, & u \subseteq k_j^{b_i^{cd}} \; where \; u \subset k_{tp_{p,q}}^{cd} \\ 0.5, & u = v \; where \; u \subset k_{tp_{p,q}}^{cd}, v \subset k_j^{b_i^{cd}} \end{cases} \quad (3)$$

เมื่อทำการเปรียบเทียบความคล้ายคลึงระหว่างแต่ละคำสำคัญในคำอธิบายรายวิชาและแต่ละคำสำคัญในหนังสือที่เกี่ยวข้อง โดย 3 วิธีข้างต้นแล้ว จะทำการรวมผลคะแนนความคล้ายคลึงระหว่างของทุกคำสำคัญในคำอธิบายรายวิชาและทำการหารด้วยจำนวนหัวข้อของเนื้อหาในคำอธิบายรายวิชาเพื่อบ่งบอกถึงร้อยละคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง ตัวอย่างเช่น ดังภาพที่ 26 แสดงตัวอย่างการเปรียบเทียบความคล้ายคลึงระหว่างคำสำคัญในคำอธิบายรายวิชา “Data mining” และคำสำคัญในหนังสือ “Data mining: Concepts and Techniques” โดยเริ่มจากการพิจารณาคำสำคัญที่ 1 ของคำอธิบายรายวิชา คือ “(‘data mining’, ‘*TE*’) (‘concept’, ‘*NN*’)” เพื่อเปรียบเทียบกับคำสำคัญในหนังสือเมื่อเปรียบเทียบแล้วจึงพบว่า คำสำคัญที่ 1 เหมือนกับคำสำคัญในหนังสือแบบตรงตัว ดังนั้น คะแนนความคล้ายคลึงของคำสำคัญที่ 1 (ปรากฏในหัวข้อที่ 1) จะมีค่าเท่ากับ 1 และเมื่อคะแนนความคล้ายคลึงถูกกำหนดให้มีค่าเท่ากับ 1 แล้ว จะไม่ทำการเปรียบเทียบด้วยวิธีการเปรียบเทียบแบบบวลีและวิธีการเปรียบเทียบแบบพิจารณาคำย่อยแล้ว ต่อมาเป็นการพิจารณาคำสำคัญที่ 2 คือ “(‘data mining’, ‘*TE*’) (‘data type’, ‘*TE*’)” เพื่อเปรียบเทียบกับคำสำคัญในหนังสือด้วยวิธีการแบบตรงตัว และ แบบบวลี ปรากฏว่าไม่พบคำสำคัญใด ๆ ในหนังสือที่เหมือนกัน แต่สำหรับวิธีการเปรียบเทียบแบบพิจารณาคำย่อย พบว่า “(‘data mining’, ‘*TE*’)” ซึ่งเป็นคำย่อย เหมือนกันกับ หลายคำย่อยของคำสำคัญในหนังสือ (กรณีที่ 3 ในสมการ (3)) แต่อย่างไรก็ตาม “(‘data mining’, ‘*TE*’)” เหมือนกับชื่อรายวิชา ซึ่งไม่ได้บ่งบอกถึงความหมายเพิ่มเติมที่มีนัยยะสำคัญ ดังนั้นคะแนนความคล้ายคลึงของคำสำคัญที่ 2 จะถูกกำหนดให้มีค่าเท่ากับ 0 และ คำสำคัญคำที่ 3 คือ “(‘data mining’, ‘*TE*’) (‘technique’, ‘*NN*’)” เหมือนกับคำสำคัญในหนังสือแบบตรงตัว ดังนั้นคะแนนความคล้ายคลึงของหัวข้อที่ 3 จะถูกกำหนดให้มีค่าเท่ากับ 1

ในส่วนของหัวข้อที่ 4 มี 2 คำสำคัญปรากฏ คือ คำสำคัญที่ 4 “(‘pattern’, ‘*TE*’) (‘mining’, ‘*NN*’)” และ คำสำคัญที่ 5 “(‘association rule’, ‘*TE*’) (‘mining’, ‘*NN*’)” ซึ่งคำสำคัญที่ 4 มีบางส่วนของคำสำคัญคือ “(‘pattern’, ‘*TE*’)” ที่เหมือนกับคำสำคัญในหนังสือคือ “(‘frequent’, ‘*JJ*’) (‘pattern’, ‘*TE*’) (‘discovery’, ‘*NN*’)” ด้วยวิธีการเปรียบเทียบแบบพิจารณาคำย่อย คะแนนความคล้ายคลึงจะถูกกำหนดให้มีเท่ากับ 0.5 และคำสำคัญที่ 5 เหมือนกับคำสำคัญในหนังสือคือ “(‘mining’, ‘*NN*’) (‘association rule’, ‘*TE*’)” ด้วยวิธีการเปรียบเทียบแบบบวลี คะแนนความคล้ายคลึงจะถูกกำหนดให้มีค่าเท่ากับ 1 แต่เนื่องด้วยหัวข้อที่ 4 มีคำสำคัญ 2 คำสำคัญปรากฏ ดังนั้นคะแนนความคล้ายคลึงของหัวข้อที่ 4 จะคำนวณจากค่าเฉลี่ยของคะแนนความคล้ายคลึงของคำสำคัญที่ 4 และ คำสำคัญที่ 5 ซึ่งสามารถคำนวณได้เป็น (0.5+1)/2 = 0.75 ในส่วนของคำสำคัญในหัวข้อที่ 5 ถึง 9 จะถูกพิจารณาด้วยวิธีเดียวกัน ดังแสดงตัวอย่างในภาพที่ 26



หัวข้อ	คำสำคัญจากคำอธิบายรายวิชา	การเปรียบเทียบแบบตรงตัว	การเปรียบเทียบวลี	การปรียบเทียบแบบพิจารณาคำย่อย	คะแนนความคล้ายคลึง
tp_1	'data mining(TE) concept(NN)'	1.0	0	0	1.0
tp_2	'data mining(TE) data type(TE)'	0	0	0	0
tp_3	'data mining(TE) technique(NN)'	1.0	0	0	1.0
tp_4	$tp_{4,1}$ = 'pattern(TE) mining(NN)'	0	0	0.5	0.75 = ((0.5+1.0)/2)
	$tp_{4,2}$ = 'association rule(TE) mining(NN)'	0	1.0	0	
tp_5	'classification(TE)'	1.0	0	0	1.0
tp_6	'clustering(TE)'	0	0	0	0
tp_7	'outlier(TE) analysis(NN)'	1.0	0	0	1.0
tp_8	'anomaly detection(TE)'	0	0	0	0
tp_9	'data mining(TE) tool(NN)'	1.0	0	0	1.0
					0.64

ภาพที่ 26 การเปรียบเทียบความคล้ายคลึงระหว่างคำสำคัญในคำอธิบายรายวิชาและคำสำคัญในหนังสือ "Data mining: Concepts and Techniques" และการคำนวณคะแนนความคล้ายคลึง

เมื่อทำการคำนวณคะแนนความคล้ายคลึงของทุกหัวข้อในคำอธิบายรายวิชาแล้ว ระบบ *SBS+* จะทำการคำนวณคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชากับหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา โดยคำนวณจากผลรวมคะแนนความคล้ายคลึงของหัวข้อในคำอธิบายรายวิชาหารด้วยจำนวนหัวข้อในคำอธิบายรายวิชา โดยจากภาพที่ 26 ผลรวมคะแนนความคล้ายคลึงทุก ๆ หัวข้อในคำอธิบายรายวิชา "Data mining" เท่ากับ 5.75 และจำนวนหัวข้อในคำอธิบายรายวิชาเท่ากับ 9

ดังนั้นคะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชากับหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา 5.75/9 ≈ 0.64 ซึ่งสามารถเขียนเป็นสมการได้ดังสมาการที่ (4)

$$SIM\left(cd, b_i^{cd}\right) = \frac{\sum_{i=1}^{n} sim_score_{tp_i}}{n} \quad (4)$$

เมื่อ cd คือ คำอธิบายรายวิชา

b_i^{cd} คือ หนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา

tp_i คือ หัวข้อหนึ่ง ๆ ที่อยู่ในคำอธิบายรายวิชา

$sim_score_{tp_i}$ คือ คะแนนความคล้ายคลึงในหัวข้อหนึ่ง ๆ

n คือ จำนวนหัวข้อที่อยู่ในคำอธิบายรายวิชา

นอกจากนี้ยังสามารถคำนวณคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชากับหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชาได้โดยสมการที่ (5)

$$disSIM = 1 - SIM\left(cd, b_i^{cd}\right) \quad (5)$$

เมื่อ $SIM\left(cd, b_i^{cd}\right)$ คือ คะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชา cd และหนังสือ b_i^{cd} ที่เกี่ยวเนื่องกับคำอธิบายรายวิชา

จากนั้นหนังสือที่เกี่ยวเนื่องกับคำอธิบารายวิชาทุก ๆ เล่มจะต้องถูกนำมาจัดอันดับตามคะแนนความคล้ายคลึงและคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชากับหนังสือที่เกี่ยวข้อง ซึ่งจะทำให้ได้รายการแนะนำหนังสือที่มีค่ามากที่สุดไปยังคะแนนความคล้ายคลึงที่มีค่าน้อยที่สุด และทำการแสดงผลการแนะนำหนังสือจากคะแนนความคล้ายคลึงที่มีค่ามากที่สุด N เล่ม โดยพิจารณาค่า N ที่น่าสนใจที่ 10, 30, 50 และ 100 ตามลำดับ ในทางกลับกันรายการแนะนำหนังสือสามารถสร้างขึ้นได้จากการจัดอันดับคะแนนความไม่คล้ายคลึงได้เช่นกัน โดยสร้างจากการพิจารณาคะแนนความไม่คล้ายคลึง เพื่อให้ผู้ใช้ได้พิจารณาหนังสือที่มีคำสำคัญที่ขยาย หรือแตกต่างออกไปจากคำสำคัญของคำอธิบายรายวิชา

จะเห็นได้ว่าในบทนี้ระบบ *SBS+* ได้พัฒนาปรับปรุงข้อบกพร่องจากระบบ *SBS* เดิม ตั้งแต่ส่วนการเตรียมข้อมูลเบื้องต้นคือ

1. รวบรวมคำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์จำนวน 28,392 คำศัพท์ จากเดิม 6,527 คำศัพท์
2. รวมรวมคำคุณศัพท์จำนวน 4,998 คำ จากเดิมไม่ได้รวบรวม
3. ปรับปรุงกฎทางภาษาศาสตร์ เพื่อลดความซ้ำซ้อนและลดเนื้อหาที่ไม่มีนัยยะสำคัญในการสกัดคำสำคัญ

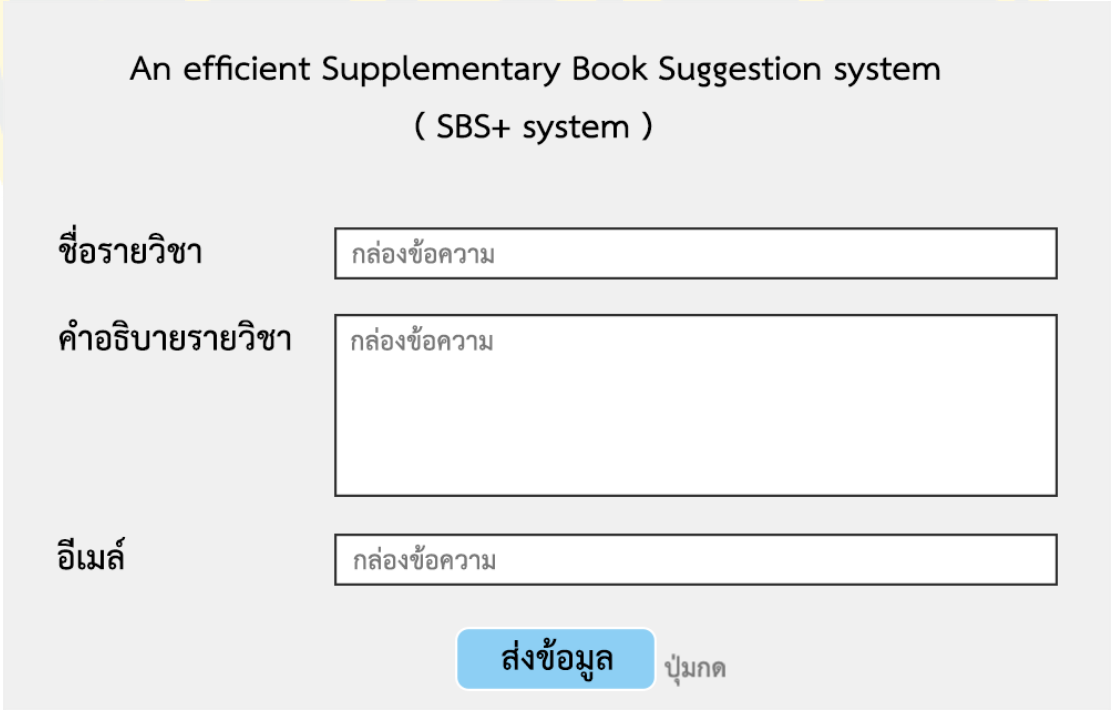
และพัฒนาเพิ่มเติมในส่วนการประมวลผลของระบบแนะนำ คือ

1. การสกัดคำสำคัญโดยประยุกต์ใช้กฎทางภาษาศาสตร์ฉบับรับปรุง ที่ซึ่งในการสกัดคำสำคัญนั้นประกอบไปด้วย 3 ขั้นตอน คือ
    i) การประมวลผลข้อความ โดยทำการเรียงลำดับขั้นตอนในการประมวลผลข้อความใหม่ เพื่อลดความผิดพลาดและช่วยให้ข้อความมีความสมบูรณ์/มีความพร้อมใช้งานในขั้นตอนต่อไปมากยิ่งขึ้น
    ii) การระบุคำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์ ได้ประยุกต์ใช้การลดรูปของคำและการหารูปเดิมของคำ และเพิ่มการเปรียบเทียบหาคำศัพท์เฉพาะทั้งหมด 2 รอบ (รอบที่ 1 เปรียบเทียบหาคำศัพท์เฉพาะก่อนการลดรูปของคำและการหารูปเดิมของคำ และรอบที่ 2 เปรียบเทียบหาคำศัพท์เฉพาะหลังการลดรูปของคำและการหารูปเดิมของคำ) ซึ่งช่วยให้ลดการตกหล่น/ไม่ถูกระบุคำศัพท์เฉพาะในหัวข้อหนึ่ง ๆ จากเดิมการเปรียบเทียบหาคำศัพท์เฉพาะจะทำเพียงครั้งเดียว
    iii) การระบุถึงคำสำคัญในหัวข้อหนึ่ง ๆ โดยประยุกต์ใช้กฎทางภาษาศาสตร์ฉบับปรับปรุง ทำให้ได้คำสำคัญที่ไม่ซ้ำซ้อนและกำจัดเนื้อหาที่ไม่บ่งบอกถึงใจความสำคัญออกไปได้
2. การคำนวณความเกี่ยวเนื่องของหนังสือที่ครอบคลุมเนื้อหาในคำอธิบายรายวิชา ได้ทำการเพิ่มวิธีการเปรียบเทียบคำสำคัญระหว่างคำสำคัญในคำอธิบายรายวิชาและคำสำคัญในหนังสือที่เกี่ยวเนื่องกับคำอธิบายรายวิชา โดยเพิ่ม 2 วิธีการคือ 1) การเปรียบเทียบแบบวลี (Paraphrase matching) และ 2) การเปรียบเทียบแบบพิจารณาคำย่อย (Sub-keyword matching) จากเดิมใช้วิธีการเปรียบเทียบแบบตรงตัว (Exact matching) เพียงอย่างเดียว

จากการพัฒนาปรับปรุงขั้นตอนต่าง ๆ ที่ได้อธิบายข้างต้นทำให้ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพสามารถสกัดคำสำคัญได้อย่างถูกต้องและตรงประเด็นมากขึ้น ส่งผลให้คำสำคัญถูกนำไปค้นหาหนังสือได้ตรงประเด็นและเฉพาะเจาะจง และสามารถเปรียบเทียบหาความเกี่ยวเนื่องกับคำอธิบายรายวิชาสามารถพบคำสำคัญ/เนื้อหาที่มีความหมายเดียวกัน/เนื้อหาบางส่วนเพื่อคำนวณความคล้ายคลึง และสร้างรายการแนะนำหนังสืออ่านเพิ่มเติมอย่างมีประสิทธิภาพสำหรับนิสิต/นักศึกษา และอาจารย์ผู้สอนใช้ในการเลือกหนังสืออ่านเพิ่มเติมจากในชั้นเรียนได้ง่าย สะดวกและตรงตามความสนใจมากขึ้น เพื่อส่งเสริมการอ่านและการเรียนรู้และศึกษาด้วยตนเองนอกชั้นเรียน

4.3 การออกแบบการใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม

หลังจากได้ทำการเตรียมข้อมูลเบื้องต้น และพัฒนาการประมวลผลของระบบแนะนำ ต่อมาจะเป็นการออกแบบพัฒนาระบบแนะนำหนังสืออ่านเพิ่มเติมในส่วนติดต่อผู้ใช้งาน โดยการออกแบบจะต้องคำนึงถึงความสะดวกในการใช้งาน และสามารถทำความเข้าใจได้ง่าย จึงทำการออกแบบการแสดงผลเพื่อรับข้อมูลนำเข้าจากผู้ใช้งาน ดังภาพที่ 27



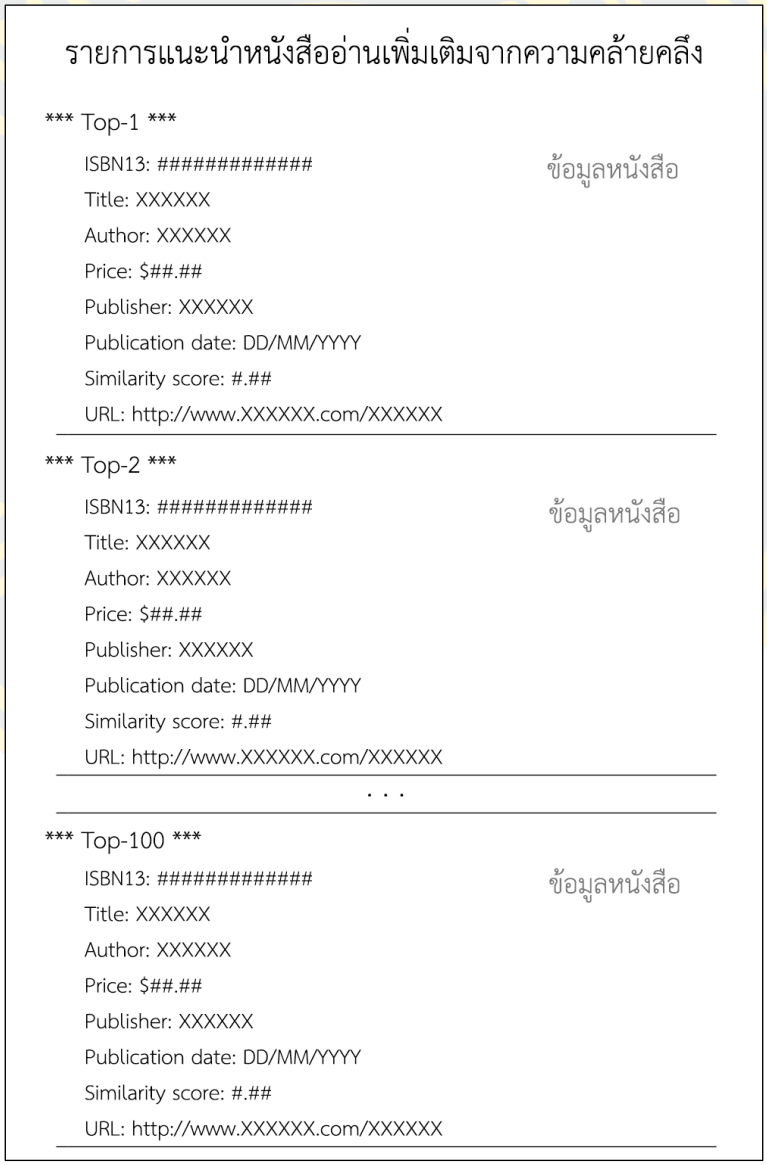
ภาพที่ 27 ออกแบบการแสดงผลเพื่อรับข้อมูลจากผู้ใช้ของระบบแนะนำหนังสืออ่านเพิ่มเติม

ในภาพที่ 27 แสดงผลหน้าต่างเพื่อรับข้อมูลนำเข้าจากผู้ใช้งาน โดยมีข้อมูลที่จำเป็นต้องกรอก 3 ส่วนคือ ชื่อรายวิชา, คำอธิบายรายวิชา และอีเมล์สำหรับตอบกลับรายการแนะนำหนังสือ

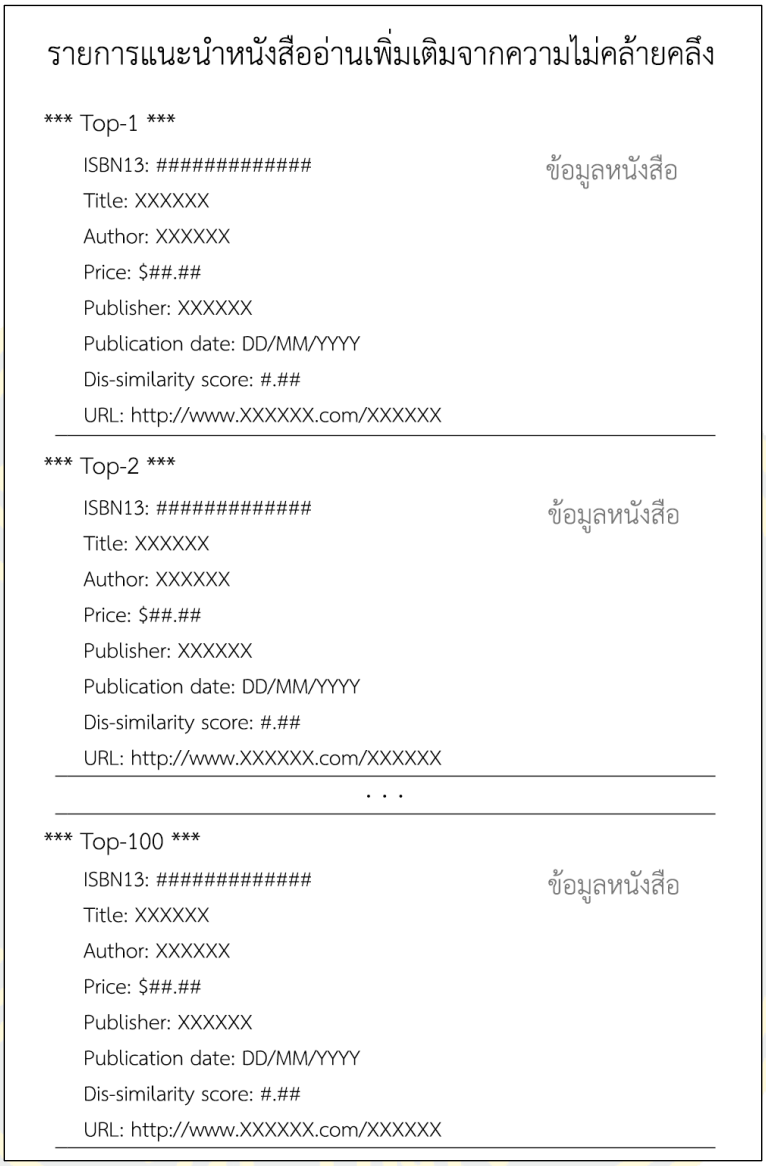
อ่านเพิ่มเติม และปุ่ม “ส่งข้อมูล” เพื่อทำการจัดเก็บข้อมูลรับเข้าจากผู้ใช้งาน และนำไปประมวลผลเพื่อสร้างรายการแนะนำหนังสืออ่านเพิ่มเติมสำหรับรายวิชาที่ผู้ใช้สนใจ เมื่อได้ผลลัพธ์รายการแนะนำเรียบร้อยแล้วจะทำการส่งอีเมล์กลับไปยังผู้ใช้งาน โดยผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมจะอยู่ในรูปแบบไฟล์เอกสารข้อมูล (.pdf) ซึ่งจะมีทั้งหมด 2 รายการคือ

1. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความคล้ายคลึง
2. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความไม่คล้ายคลึง

แสดงดังภาพที่ 28 และภาพที่ 29 ตามลำดับ



ภาพที่ 28 ออกแบบการแสดงผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความคล้ายคลึง



ภาพที่ 29 ออกแบบการแสดงผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับโดยพิจารณาจากคะแนนความไม่คล้ายคลึง

จากภาพที่ 28 และภาพที่ 29 ผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่อยู่ในรูปแบบไฟล์เอกสารข้อมูล (.pdf) นั้นจะประกอบไปด้วยชื่อรายการแนะนำหนังสืออ่านเพิ่มเติม และข้อมูลหนังสือที่ถูกแนะนำแต่ละเล่มจำนวนทั้งหมด 100 เล่ม เนื่องจากงานวิจัยนี้พิจารณาค่า N ที่น่าสนใจที่ 100 เป็นค่ามากที่สุด โดยข้อมูลหนังสือจะประกอบไปด้วย

1. อับดับของหนังสือที่ถูกแนะนำ
2. หมายเลข 13 หลักประจำหนังสือ (ISBN13)
3. ชื่อหนังสือ
4. ชื่อผู้แต่ง
5. ราคา
6. ชื่อสำนักพิมพ์
7. วันที่ตีพิมพ์หนังสือ
8. คะแนนความคล้ายคลึง หรือคะแนนความไม่คล้ายคลึง
9. ลิงก์เชื่อมโยงไปยังเว็บไซต์ของหนังสือเล่มนั้น ๆ

บทที่ 5
ผลการดำเนินงาน

ในบทนี้จะอธิบายถึงผลการดำเนินงานของระบบแนะนำหนังสืออ่านเพิ่มเติม (ระบบ *SBS*) และ ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ (ระบบ *SBS+*) ที่ได้นำเสนอไว้ในบทก่อนหน้า โดยในการทดลอง ผู้วิจัยได้ทำการรวบรวม 640 คำอธิบายรายวิชา (เฉพาะภาษาอังกฤษเท่านั้น) จากหลักสูตรวิทยาการคอมพิวเตอร์ของ 13 มหาวิทยาลัยไทยที่ได้รับการจัดอันดับจากเว็บไซต์ QS University Ranking of Asian[23] โดยที่แต่ละหลักสูตรของแต่ละมหาวิทยาลัยจะมีจำนวนวิชาที่ไม่เท่ากันแสดงรายละเอียดดังตารางที่ 4

ตารางที่ 4 จำนวนรายวิชาจากหลักสูตรวิทยาการคอมพิวเตอร์ของ 13 มหาวิทยาลัย

ชื่อมหาวิทยาลัย	จำนวนรายวิชา	จำนวนหัวข้อ	จำนวนคำสำคัญที่ถูกต้อง
Chulalongkorn University (CU)	46	332	414
Mahidol University (MU)	35	433	554
Chiang Mai University (CMU)	57	515	641
Thammasat University (TU)	84	789	985
Kasetsart University (KU)	69	609	759
Prince of Songkla University (PSU)	61	583	756
King Mongkut's University of Technology Thonburi (KMUTT)	40	520	627
King Mongkut's Institute of Technology Ladkrabang (KMITL)	87	891	1,103
Naresuan University (NU)	47	497	629
Silpakorn University (SU)	50	511	660
Mae Fah Luang University (MFU)	39	357	449
Mahasarakham University (MSU)	57	556	735
Burapha University (BUU)	68	715	967
จำนวนทั้งหมด	640	7,308	9,279

[23] https://www.topuniversities.com/university-rankings/asian-university-rankings/2021

การประเมินระบบ *SBS* และ *SBS+* แบ่งออกเป็น 2 ส่วน ได้แก่ 1) การประเมินประสิทธิภาพของการสกัดคำสำคัญ (เนื่องจากการสกัดคำสำคัญเป็นขั้นตอนที่สำคัญและส่งผลต่อประสิทธิภาพของระบบ) และ 2) การประเมินความครอบคลุมของเนื้อหาของหนังสือที่แนะนำจำนวน N เล่ม (Top-N recommendation) โดยกำหนดค่า N เท่ากับ 10, 30, 50 และ 100 ตามลำดับ แสดงรายละเอียดดังต่อไปนี้

5.1 การประเมินประสิทธิภาพของการสกัดคำสำคัญ

จากบทก่อนหน้า เราทราบได้ว่าการสกัดคำสำคัญทั้งจากคำอธิบายรายวิชาและหนังสือเล่มหนึ่งเป็นขั้นตอนหลักของระบบที่นำเสนอ ซึ่งขั้นตอนดังกล่าวส่งผลต่อประสิทธิภาพและประสิทธิผลของการสร้างรายการแนะนำหนังสือ ดังนั้นผู้วิจัยจึงทำการประยุต์ใช้การประเมินค่าความแม่นยำโดยรวม ($Accuracy$), ค่าความแม่นยำ ($Precision$), ค่าความถูกต้อง ($Recall$) และ ค่าประสิทธิภาพโดยรวม ($F\text{-}measure$) (Tharwat, 2020) เพื่อวัดประสิทธิภาพของการสกัดคำสำคัญ โดยแต่ละค่าสามารถคำนวณได้จากข้อมูลภายในตารางประเมินประสิทธิภาพความถูกต้อง (Confusion matrix) ซึ่งเป็นตารางแสดงข้อมูลระหว่างข้อมูลคำสำคัญที่เกิดขึ้นจริงและข้อมูลคำสำคัญที่สกัดได้ โดยข้อมูลแกนตั้งคือข้อมูลคำสำคัญที่เกิดขึ้นจริง ส่วนข้อมูลแกนนอนคือข้อมูลคำสำคัญที่สกัดได้ แสดงดังตารางที่ 5

ตารางที่ 5 ตารางประเมินประสิทธิภาพความถูกต้องระหว่างข้อมูลคำสำคัญที่เกิดขึ้นจริงและข้อมูลคำสำคัญที่สกัดได้

		ข้อมูลคำสำคัญที่เกิดขึ้นจริง	
		คำสำคัญที่ถูกต้อง	**คำสำคัญที่ไม่ถูกต้อง**
ข้อมูลคำสำคัญที่สกัดได้	**คำสำคัญที่ถูกสกัด**	TP	FP
	คำสำคัญที่ไม่ถูกสกัด	FN	TN

โดยแต่ละค่าที่ปรากฏอยู่ในตาราง สามารถอธิบายได้ดังนี้

1. TP (True Positive) คือ คำสำคัญที่ถูกสกัดได้ และเป็นคำสำคัญที่ถูกต้อง
2. TN (True Negative) คือ คำสำคัญที่ไม่ถูกสกัด และเป็นคำสำคัญที่ไม่ถูกต้อง
3. FP (False Positive) คือ คำสำคัญที่ถูกสกัดได้ และเป็นคำสำคัญที่ไม่ถูกต้อง
4. FN (False Negative) คือ คำสำคัญที่ไม่ถูกสกัด และเป็นคำสำคัญที่ถูกต้อง

จากข้อมูลในตารางประเมินประสิทธิภาพความถูกต้อง สามารถคำนวณค่า $Accuracy$, $Precision$, $Recall$ และ $F\text{-}measure$ ได้จากสมการที่ 6, 7, 8 และ 9 ตามลำดับดังนี้

$$Accuracy = \frac{TP}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F\text{-}measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (9)$$

การทดสอบประสิทธิภาพของการสกัดคำสำคัญได้ทำการเปรียบเทียบวิธีการสกัดคำสำคัญกับขั้นตอนวิธี "*Termine*" (Frantzi, Ananiadou and Mima, 2000) และขั้นตอนวิธี "*RAKE*" (Rose *et al.*, 2010) ซึ่งเป็นวิธีการสกัดคำสำคัญที่ได้รับความนิยม โดยขั้นตอนวิธีดังกล่าวสามารถอธิบายได้ดังนี้

1. "*Termine*"—เป็นขั้นตอนวิธีในการสกัดกลุ่มคำจากข้อความหนึ่ง ๆ ซึ่งกลุ่มคำสำคัญที่สนใจคือ นามวลี (Noun phrase) กล่าวคือ กลุ่มของคำนามที่อยู่ติดกัน หรือกลุ่มของคำคุณศัพท์และคำนามที่อยู่ติดกัน
2. "*RAKE*"— เป็นขั้นตอนวิธีที่ในการพิจารณาคำสำคัญ โดย 1) ทำการพิจารณาคำที่อยู่ติดกันไปเรื่อย ๆ จนเจอคำหยุด (Stop word) แล้วจึงนำกลุ่มคำที่ปรากฏก่อนหน้าคำหยุดมาสร้างรายการแคนดิเดทของคำสำคัญ 2) ทำการสร้างเมทริกซ์ (Matrix) ของแต่ละคำที่ปรากฏในรายการแคนดิเดท เพื่อคำนวณคะแนนความถี่ของคำที่เกิดขึ้นร่วมกัน 3) ทำการรวมคะแนนของคำสำคัญที่ปรากฏในรายการแคนดิเดท และจัดอันดับคำสำคัญโดยพิจารณาคะแนนของคำสำคัญแต่ละคำจากมากไปหาน้อย 4) ทำการพิจารณาคะแนนของแต่ละคำสำคัญ โดยจะสนใจคำสำคัญที่มีคะแนนมากที่สุดเพียง 1 ใน 3 ของคำสำคัญทั้งหมดในรายการแคนดิเดท

ด้วยเหตุนี้การทดสอบประสิทธิภาพของการสกัดคำจะทดสอบกับ 4 ขั้นตอนวิธี คือ *SBS*, *SBS+*, *Termine* และ *RAKE* โดยทำการคำนวณหาค่า $Accuracy$, $Precision$, $Recall$ และ $F\text{-}measure$ ตารางที่ 6 แสดงจำนวนคำสำคัญที่สกัดได้และจำนวนคำสำคัญที่สกัดได้อย่างถูกต้อง และตารางที่ 7 แสดงการเปรียบเทียบค่าความแม่นยำโดยรวม ($Accuracy$) ของการสกัดคำสำคัญทั้ง 4 ขั้นตอนวิธีดังกล่าว

ตารางที่ 6 จำนวนคำสำคัญที่สกัดได้และจำนวนคำสำคัญที่มีความถูกต้องของ 4 ขั้นตอนวิธี

ชื่อมหาวิทยาลัย (คำสำคัญรวมในคำอธิบาย)	จำนวนคำสำคัญที่สกัดได้				จำนวนคำสำคัญที่สกัดได้ถูกต้อง			
	Termine	*RAKE*	*SBS*	*SBS+*	*Termine*	*RAKE*	*SBS*	*SBS+*
CU (414)	262	474	568	415	179	283	345	407
MU (554)	242	614	678	554	184	409	456	547
CMU (641)	390	732	820	641	286	447	525	639
TU (985)	578	1,195	1,386	985	344	638	756	975
KU (759)	443	901	1,001	758	311	529	613	756
PSU (756)	480	867	1,000	750	348	524	616	747
KMUTT (627)	353	734	808	623	250	433	513	621
KMITL (1,103)	681	1,313	1,487	1,103	429	726	887	1,096
NU (629)	376	711	830	624	244	395	507	616
SU (660)	388	789	912	657	259	420	549	656
MFU (449)	250	538	622	445	164	296	355	444
MSU (735)	478	882	1,082	722	285	448	530	717
BUU (967)	525	1,142	1,297	967	338	632	774	965
จำนวน (9,279)	5,446	10,892	12,491	9,244	3,621	6,180	7,426	9,186

ตารางที่ 7 การเปรียบเทียบค่าร้อยละของความแม่นยำโดยรวม (Accuracy)

ชื่อมหาวิทยาลัย (คำสำคัญรวมในคำอธิบาย)	ค่าความแม่นยำโดยรวม $(Accuracy)$			
	Termine	*RAKE*	*SBS*	*SBS+*
CU (414)	36.02	46.78	54.16	96.45
MU (554)	30.07	53.89	58.76	97.50
CMU (641)	38.39	48.27	56.09	99.38
TU (985)	28.22	41.37	46.81	97.99
KU (759)	34.90	46.77	53.44	99.34
PSU (756)	39.19	47.68	54.04	98.42
KMUTT (627)	34.25	46.66	55.64	98.73
KMITL (1,103)	31.66	42.96	52.08	98.74
NU (629)	32.06	41.80	53.26	96.70
SU (660)	32.83	40.82	53.67	99.24
MFU (449)	30.65	42.84	49.58	98.67
MSU (735)	30.71	38.32	41.18	96.89
BUU (967)	29.29	42.79	51.95	99.59
จำนวน (9,279)	32.61	44.17	51.77	98.38

จากตารางที่ 6 เห็นได้ว่าแต่ละขั้นตอนวิธีที่สามารถสกัดคำสำคัญได้เป็น *Termine* (5,446 คำ), *RAKE* (10,892 คำ), *SBS* (12,491 คำ) และ *SBS+* (9,244 คำ) โดยจะสังเกตได้ว่า *RAKE* และ *SBS* สามารถสกัดคำได้มากกว่าคำสำคัญที่ปรากฏขึ้นจริงในคำอธิบายรายวิชา เนื่องจากทั้งสองขั้นตอนวิธีสกัดคำสำคัญซ้ำซ้อน ส่วนขั้นตอนวิธี *Termine* สามารถสกัดคำสำคัญได้น้อยที่สุดเนื่องจากพิจารณาเฉพาะนามวลี (Noun phrase) ที่ปรากฏหัวข้อเท่านั้น แต่สำหรับขั้นตอนวิธี *SBS+* สามารถสกัดคำมีจำนวนใกล้เคียงกับจำนวนคำสำคัญที่ปรากฏขึ้นจริงในคำอธิบายรายวิชา และสามารถสกัดคำสำคัญที่ถูกต้องออกมาได้มากที่สุด คือ *SBS+* (9,186), *SBS* (7,426), *RAKE* (6,180) และ *Termine* (3,621) ตามลำดับ ซึ่งจากการพัฒนาในส่วนต่าง ๆ ดังอธิบายในบทที่ 4 ส่งผลให้ค่า $\boldsymbol{Accuracy}$ โดยเฉลี่ยของขั้นตอนวิธี *SBS+* (98.38%) มีค่ามากกว่าขั้นตอนวิธี *SBS* (51.77%), *RAKE* (44.17%) และ *Termine* (32.61%) ดังแสดงในตารางที่ 7

ตารางที่ 8 และตารางที่ 9 เป็นการเปรียบเทียบประสิทธิภาพในการสกัดคำสำคัญ โดยค่า $\boldsymbol{Precision}$ และ $\boldsymbol{Recall}$ ซึ่งจะเห็นได้ว่าขั้นตอนวิธี *SBS+* (99.33%) มีค่า $\boldsymbol{Precision}$ สูงที่สุดเมื่อเทียบกับขั้นตอนวิธี *Termine* (67.31%), *SBS* (59.97%) และ *RAKE* (57.24%) และขั้นตอนวิธี *SBS+* (98.93%) ยังคงมีค่า $\boldsymbol{Recall}$ มากที่สุดเมื่อเทียบกับขั้นตอนวิธี *SBS* (80.29%), *RAKE* (66.87%) และ *Termine* (39.23%) ตามลำดับ โดยจากค่าทั้งสองสามารถบ่งบอกได้ว่าขั้นตอนวิธี *SBS+* มีความสามารถในการสกัดคำสำคัญจากคำอธิบายรายวิชาที่ค่อนข้างสูง นอกจากนั้นตารางที่ 10 แสดงว่าขั้นตอนวิธี *SBS+* (99.13%) มีค่า $\boldsymbol{F\text{-}measure}$ มากที่สุดเมื่อเทียบกับขั้นตอนวิธี *SBS* (68.62%), *RAKE* (61.67%) และ *Termine* (49.47%) ตามลำดับ ที่ซึ่งแสดงถึงประสิทธิภาพในการสกัดคำสำคัญ

ตารางที่ 8 การเปรียบเทียบค่าร้อยละของความแม่นยำ (Precision)

ชื่อมหาวิทยาลัย (คำสำคัญรวมในคำอธิบาย)	ค่าความแม่นยำ ($Precision$)			
	Termine	*RAKE*	*SBS*	*SBS+*
CU (414)	68.32	59.70	60.74	98.07
MU (554)	76.03	66.61	67.26	98.74
CMU (641)	73.33	61.07	64.02	99.69
TU (985)	59.52	53.39	54.55	98.98
KU (759)	70.20	58.71	61.24	99.74
PSU (756)	72.50	60.44	61.60	99.60
KMUTT (627)	70.82	58.99	63.49	99.68
KMITL (1,103)	63.00	55.29	59.65	99.37
NU (629)	64.89	55.56	61.08	98.72
SU (660)	66.75	53.23	60.20	99.85
MFU (449)	65.60	55.02	57.07	99.78
MSU (735)	59.62	50.79	48.98	99.31
BUU (967)	64.38	55.34	59.68	99.79
จำนวน (9,279)	67.31	57.24	59.97	99.33

ตารางที่ 9 การเปรียบเทียบค่าร้อยละของความถูกต้อง (Recall)

ชื่อมหาวิทยาลัย (คำสำคัญรวมในคำอธิบาย)	ค่าความถูกต้อง ($Recall$)			
	Termine	*RAKE*	*SBS*	*SBS+*
CU (414)	43.24	68.36	83.33	98.31
MU (554)	33.21	73.83	82.31	98.74
CMU (641)	44.62	69.73	81.90	99.69
TU (985)	34.92	64.77	76.75	98.98
KU (759)	40.97	69.70	80.76	99.60
PSU (756)	46.03	69.31	81.48	98.81
KMUTT (627)	39.87	69.06	81.82	99.04
KMITL (1,103)	38.89	65.82	80.42	99.37
NU (629)	38.79	62.80	80.60	97.93
SU (660)	39.24	63.64	83.18	99.39
MFU (449)	36.53	65.92	79.06	98.89
MSU (735)	38.78	60.95	72.11	97.55
BUU (967)	34.95	65.36	80.04	99.79
จำนวน (9,279)	39.23	66.87	80.29	98.93

ตารางที่ 10 การเปรียบเทียบค่าร้อยละของประสิทธิภาพโดยรวม (F-measure)

ชื่อมหาวิทยาลัย (คำสำคัญที่ถูกต้อง)	ค่าประสิทธิภาพโดยรวม ($F\text{-}measure$)			
	Termine	*RAKE*	*SBS*	*SBS+*
CU (414)	52.96	63.74	70.26	98.19
MU (554)	46.23	70.03	74.03	98.74
CMU (641)	55.48	65.11	71.87	99.69
TU (985)	44.02	58.53	63.77	98.98
KU (759)	51.75	63.73	69.66	99.67
PSU (756)	56.31	54.57	70.16	99.20
KMUTT (627)	51.02	63.63	71.50	99.36
KMITL (1,103)	48.09	60.10	68.49	99.37
NU (629)	48.56	58.96	69.50	98.32
SU (660)	49.43	57.97	69.85	99.62
MFU (449)	46.92	59.98	66.29	99.33
MSU (735)	46.99	55.41	58.34	98.42
BUU (967)	45.31	59.93	68.37	99.79
จำนวน (9,279)	49.47	61.67	68.62	99.13

5.2 การประเมินการประมวลผลของระบบแนะนำ

การคำนวณคะแนนความคล้ายคลึงของหนังสือเล่มหนึ่ง ๆ ที่เกี่ยวเนื่องกับคำอธิบายรายวิชา บ่งบอกได้ว่าหนังสือเล่มนั้น ๆ มีความครอบคลุมเนื้อหา (คำสำคัญ) ในคำอธิบายรายวิชามากน้อยเพียงใด ดังนั้นการประเมินประสิทธิภาพของระบบแนะนำพิจารณาได้ดังนี้

1. ค่าความครอบคลุมเนื้อหาโดยเฉลี่ย (Average Percentage of Coverage of Contents) หรือเรียกว่าค่า “$APCoC_N$” คำนวณได้จากสมการที่ 10

$$APCoC_N = \frac{\sum_{j=0}^{N} SIM\left(cd, b_i^{cd}\right)}{N} \quad (10)$$

เมื่อ $SIM\left(cd, b_i^{cd}\right)$ คือ คะแนนความคล้ายคลึงระหว่างคำอธิบายรายวิชา cd และหนังสือ b_i^{cd} ที่เกี่ยวเนื่องกับคำอธิบายรายวิชา

N คือ จำนวนหนังสือที่ถูกแนะนำได้แก่ 10, 30, 50 หรือ 100

โดยค่าความครอบคลุมเนื้อหาโดยเฉลี่ยจะแสดงถึงค่าเฉลี่ยของคะแนนความคล้ายคลึงระหว่างคำอธิบายและหนังสือที่เกี่ยวข้องหนึ่ง ๆ (หรือจำนวนเนื้อหาที่มีความเหมือนกันระหว่างเนื้อหาในคำอธิบายรายวิชาและเนื้อหาในหนังสือที่เกี่ยวข้อง) ในรายการแนะนำหนังสือ N เล่ม หรือพิจารณาอีกแบบหนึ่งได้ว่าเป็นค่าเฉลี่ยของความคล้ายคลึงของเนื้อหา

2. ค่าความครอบคลุมเนื้อหาสะสม (Cumulative Percentage of Coverage of Contents) หรือเรียกว่าค่า "$CPCoC_N$" คำนวณได้จากสมการที่ 11

$$CPCoC_N = \frac{\sum_{j=0}^{n}\begin{pmatrix} max\left(sim_score_{tp_1}(cd, b_1^{cd}), sim_score_{tp_1}(cd, b_1^{cd}), \cdots, sim_score_{tp_1}(cd, b_N^{cd})\right), \\ max\left(sim_score_{tp_2}(cd, b_1^{cd}), sim_score_{tp_2}(cd, b_1^{cd}), \cdots, sim_score_{tp_2}(cd, b_N^{cd})\right), \cdots, \\ max\left(sim_score_{tp_n}(cd, b_1^{cd}), sim_score_{tp_n}(cd, b_1^{cd}), \cdots, sim_score_{tp_n}(cd, b_N^{cd})\right) \end{pmatrix}}{n} \quad (11)$$

เมื่อ $sim_score_{tp_j}(cd, b_i^{cd})$ คือ คะแนนความคล้ายคลึงของหัวข้อ tp_j ระหว่างคำอธิบายรายวิชา cd และหนังสือ b_i^{cd} ที่เกี่ยวเนื่องกับคำอธิบายรายวิชา

N คือ จำนวนหนังสือที่ถูกแนะนำได้แก่ 10, 30, 50 หรือ 100

n คือ จำนวนหัวข้อที่อยู่ในคำอธิบายรายวิชา

โดยค่าความครอบคลุมเนื้อหาสะสมบ่งบอกถึงค่าความคล้ายคลึงของคำอธิบายรายวิชาหนึ่ง ๆ กับรายการแนะนำหนังสือ N เล่ม โดยพิจารณาจากคะแนนความคล้ายคลึงของแต่ละหัวข้อที่มีค่ามากที่สุดระหว่างเนื้อหาคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง

3. ค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ย (Average Percentage of Uncoverage of Contents) หรือเรียกว่าค่า "$APUoC_N$" คำนวณได้จากสมการที่ 12

$$APUoC_N = \frac{\sum_{j=0}^{N} disSIM(cd, b_i^{cd})}{N} \quad (12)$$

เมื่อ $disSIM(cd, b_i^{cd})$ คือ คะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชา cd และหนังสือ b_i^{cd} ที่เกี่ยวเนื่องกับคำอธิบายรายวิชา

N คือ จำนวนหนังสือที่ถูกแนะนำได้แก่ 10, 30, 50 หรือ 100

โดยค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ยแสดงถึงค่าเฉลี่ยของคะแนนความไม่คล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้องเล่มหนึ่ง ๆ ในรายการแนะนำหนังสือ N เล่ม (หนังสือเล่มหนึ่ง ๆ จะถูกคำนวณคะแนนความคล้ายคลึง และคะแนนความไม่คล้ายคลึงคือค่าตรงกันข้ามกับคะแนนความคล้ายคลึง)

4. ค่าความไม่ครอบคลุมเนื้อหาสะสม (Cumulative Percentage of Uncoverage of Contents) หรือเรียกว่าค่า "$CPUoC_N$" คำนวณได้จากสมการที่ 13

$$CPUoC_N = 1 - CPCoC_N \quad (13)$$

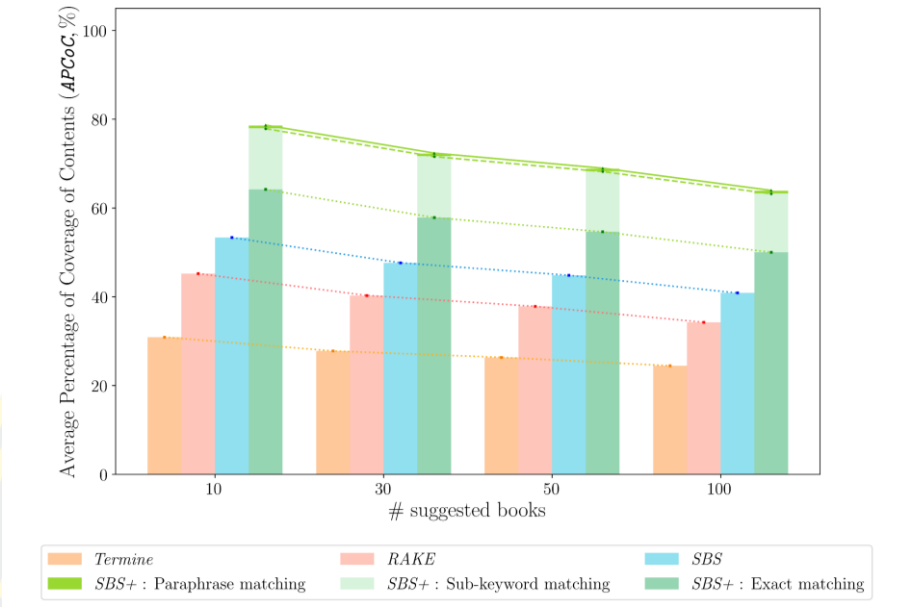
เมื่อ $CPCoC_N$ คือ ค่าความครอบคลุมเนื้อหาสะสม

N คือ จำนวนหนังสือที่ถูกแนะนำได้แก่ 10, 30, 50 หรือ 100

โดยค่าความไม่ครอบคลุมเนื้อหาสะสมบ่งบอกถึงค่าความไม่คล้ายคลึงของคำอธิบายรายวิชาหนึ่ง ๆ กับรายการแนะนำหนังสือ N เล่ม กล่าวคือค่าความไม่ครอบคลุมเนื้อหาสะสมคือค่าตรงกันข้ามกับค่าความครอบคลุมเนื้อหาสะสมในรายการแนะนำหนังสือ N เล่ม

การประเมินประสิทธิภาพการประมวลผลของระบบแนะนำจะทำการเปรียบเทียบค่า $APCoC_N$, $CPCoC_N$, $APUoC_N$ และ $CPUoC_N$ ระหว่างระบบ *SBS+*, ระบบ *SBS*, ระบบแนะนำที่ใช้ขั้นตอนวิธี *RAKE* ในการสกัดคำสำคัญ โดยจะเรียกว่า "ระบบ *RAKE*" และ ระบบแนะนำที่ใช้ขั้นตอนวิธี *Termine* ในการสกัดคำสำคัญ โดยจะเรียกว่า "ระบบ *Termine*"

ภาพที่ 30 แสดงการเปรียบเทียบค่า $APCoC_{10}$, $APCoC_{30}$, $APCoC_{50}$ และ $APCoC_{100}$ ของทั้ง 4 ระบบแนะนำ โดยในบทที่ 3 ได้อธิบายไว้ว่าระบบ *SBS* ใช้วิธีการเปรียบเทียบแบบตรงตัวเพียงวิธีการเดียวในการคำนวณคะแนนความคล้ายคลึง และในบทที่ 4 ระบบ *SBS+* ใช้วิธีการเปรียบเทียบแบบตรงตัว, แบบวลี และแบบพิจารณาคำย่อยในการคำนวณคะแนนความคล้ายคลึง ดังนั้นภาพที่ 30 จึงแสดงการเปรียบเทียบค่า $APCoC_N$ โดยระบบ *SBS*, *RAKE* และ *Termine* ที่ประยุกต์ใช้วิธีการเปรียบเทียบแบบตรงตัวเพียงอย่างเดียว และระบบ *SBS+* ใช้วิธีการเปรียบเทียบแบบตรงตัว, แบบวลี และแบบพิจารณาคำย่อย



ภาพที่ 30 การเปรียบเทียบค่าความครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติม

จากภาพที่ 30 สังเกตได้ว่า เมื่อจำนวนในการแนะนำหนังสือเพิ่มมากขึ้น (ค่า N เพิ่มขึ้น) ค่า $APCoC_N$ จะลดลง กล่าวคือ เมื่อทำการแนะนำหนังสือมากขึ้นจะทำให้มีโอกาสที่จะได้หนังสือที่เกี่ยวข้องกับคำอธิบายรายวิชาน้อยลง (หนังสือที่แนะนำทั้งหมดมีการจัดอันดับความเกี่ยวเนื่องกับคำอธิบายรายวิชาโดยเรียงลำดับจากหนังสือที่มีความเกี่ยวเนื่องมากที่สุดไปยังน้อยที่สุด) ดังนั้นสามารถบอกได้ว่าหนังสือที่ถูกจัดอับดับอยู่ในอันดับต้น ๆ นั้นมีความเกี่ยวเนื่องกับคำอธิบายรายวิชามากที่สุด จากภาพที่ 30 การเปรียบเทียบค่า $APCoC_N$ ของแต่ละระบบแนะนำมีผลลัพธ์ (โดยเรียงลำดับตามค่า N คือ 10, 30, 50 และ 100) ดังนี้

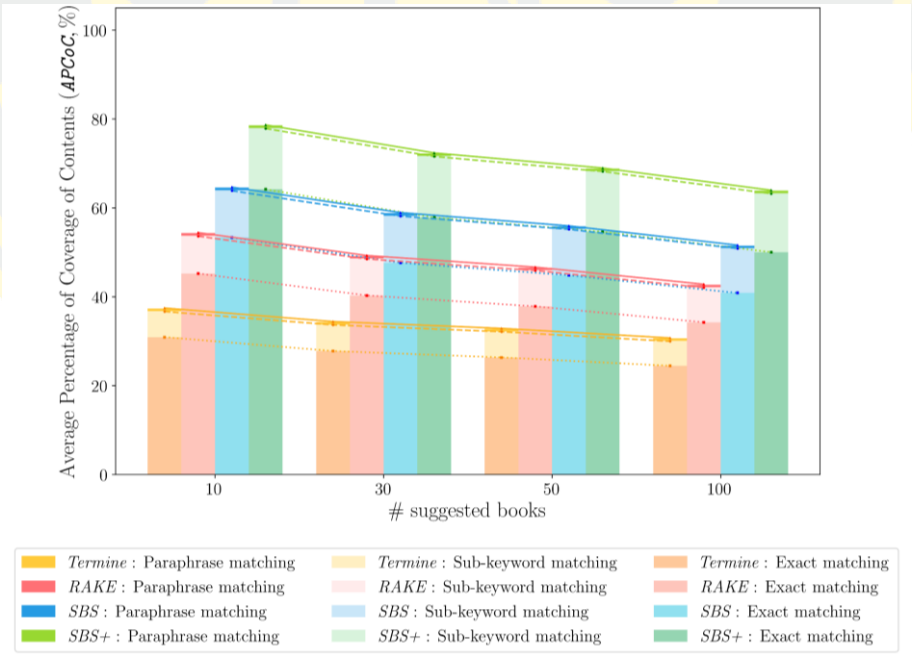
1. ระบบ *Termine* มีค่า 30.9% , 27.8%, 26.3% และ 24.4%
2. ระบบ *RAKE* มีค่า 45.2%, 40.3%, 37.8% และ 34.2%
3. ระบบ *SBS* มีค่า 53.3%, 47.6%, 44.8% และ 40.8%
4. ระบบ SBS+ มีค่า 78.0%, 71.7%, 68.3% และ 63.3%

โดยจะสามารถคำนวณเปรียบเทียบประสิทธิภาพได้ว่าระบบ *SBS+* มีค่า $APCoC_N$ มากกว่าระบบ *Termine* ≈ 43.0%, *RAKE* ≈ 30.9% และ *SBS* ≈ 23.7% และมากกว่าทั้ง 3 ระบบโดยเฉลี่ย ≈ 32.5% เหตุผลคือ ระบบ *SBS+* ได้เพิ่มการเปรียบเทียบแบบวลี และการเปรียบเทียบแบบพิจารณาคำย่อยในการคำนวณคะแนนความคล้ายคลึง ซึ่งส่งผลให้ค่า $APCoC_N$ ของระบบ *SBS+* เพิ่มขึ้นโดยเฉลี่ย ≈ 13.7%

ภาพที่ 31 แสดงการเปรียบเทียบค่า $APCoC_{10}$, $APCoC_{30}$, $APCoC_{50}$ และ $APCoC_{100}$ ของทั้ง 4 ระบบแนะนำโดยทำการเพิ่มการเปรียบเทียบแบบวลี และแบบพิจารณาคำย่อยในการคำนวณคะแนนความคล้ายคลึงให้กับระบบ *SBS*, *RAKE* และ *Termine* (เดิมจากภาพที่ 30 ระบบ *SBS*, *RAKE* และ *Termine* ประยุกต์ใช้วิธีการเปรียบเทียบแบบตรงตัวเพียงวิธีการเดียวในการคำนวณคะแนนความคล้ายคลึง) โดยจากการทดลองและการเปรียบเทียบค่า $APCoC_N$ สังเกตได้ว่าค่า $APCoC_N$ ของระบบ *SBS*, *RAKE* และ *Termine* เพิ่มขึ้น (โดยเรียงลำดับตามค่า N คือ 10, 30, 50 และ 100) ดังนี้

1. ระบบ *Termine* มีค่า 36.8%, 33.8%, 32.3% และ 30.1% เพิ่มขึ้นจากเดิมโดยเฉลี่ย 5.9%
2. ระบบ *RAKE* มีค่า 53.8%, 48.7%, 46.0% และ 42.2% เพิ่มขึ้นจากเดิมโดยเฉลี่ย 8.9%
3. ระบบ *SBS* มีค่า 64.0%, 58.3%, 55.3% และ 51.0% เพิ่มขึ้นจากเดิมโดยเฉลี่ย 10.5%

แต่อย่างไรก็ตาม จากการสังเกตในภาพที่ 31 ระบบ *SBS+* ยังคงมีค่า $APCoC_N$ มากกว่าระบบ *Termine* ≈ 37.1%, *RAKE* ≈ 22.6% และ *SBS* ≈ 13.2% และมากกว่าทั้ง 3 ระบบโดยเฉลี่ย ≈ 24.3%

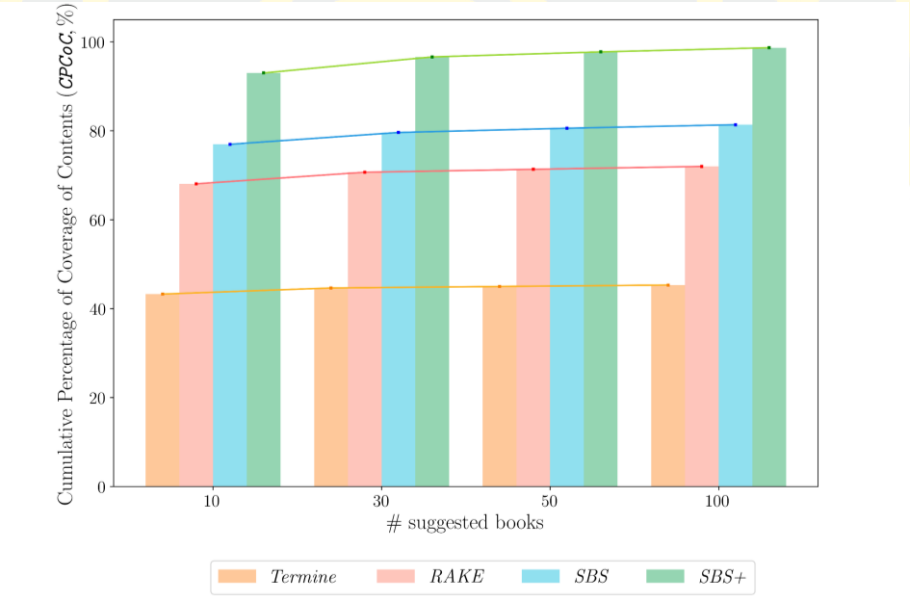


ภาพที่ 31 การเปรียบเทียบค่าความครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติมโดยใช้วิธีการเปรียบเทียบแบบตรงตัว, แบบวลี และแบบพิจารณาคำย่อย

จากที่ได้อธิบายไว้ข้างต้นว่าเมื่อจำนวนในการแนะนำหนังสือเพิ่มมากขึ้น (ค่า N เพิ่มมากขึ้น) ค่า $APCoC_N$ จะลดลง ในทางตรงกันข้ามค่า $CPCoC_N$ จะเพิ่มมากขึ้นเมื่อจำนวนหนังสือในการแนะนำเพิ่มมากขึ้น กล่าวคือ เมื่อจำนวนหนังสือที่แนะนำเพิ่มมากยิ่งขึ้นก็จะมีความเหมือนของคำสำคัญระหว่างหนังสือและคำอธิบายรายวิชาปรากฏมากขึ้นไปด้วย ภาพที่ 32 แสดงการเปรียบเทียบค่า $CPCoC_{10}$, $CPCoC_{30}$, $CPCoC_{50}$ และ $CPCoC_{100}$ ของระบบแนะนำ ซึ่งมีผลลัพธ์ (โดยเรียงลำดับตามค่า N คือ 10, 30, 50 และ 100) ดังนี้

1. ระบบ *Termine* มีค่า 42.7%, 44.1%, 44.5% และ 44.8%
2. ระบบ *RAKE* มีค่า 67.5%, 70.1%, 70.8% และ 71.5%
3. ระบบ *SBS* มีค่า 76.4%, 79.1%, 80.0% และ 80.8%
4. ระบบ *SBS+* มีค่า 92.5%, 96.1%, 97.2% และ 98.2%

จากข้างต้นจะสามารถคำนวณได้ว่าระบบ *SBS+* มีค่า $CPCoC_N$ มากกว่าระบบ *Termine* ≈ 51.9%, *RAKE* ≈ 26.0% และ *SBS* ≈ 16.9% และ มากกว่าทั้ง 3 ระบบโดยเฉลี่ย ≈ 31.6%

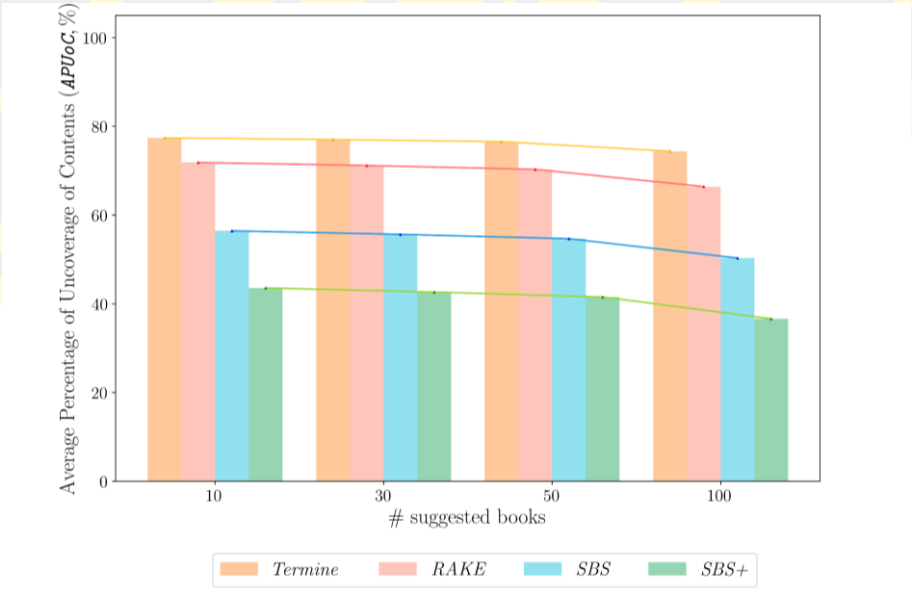


ภาพที่ 32 การเปรียบเทียบค่าความครอบคลุมเนื้อหาสะสมของระบบแนะนำหนังสืออ่านเพิ่มเติม

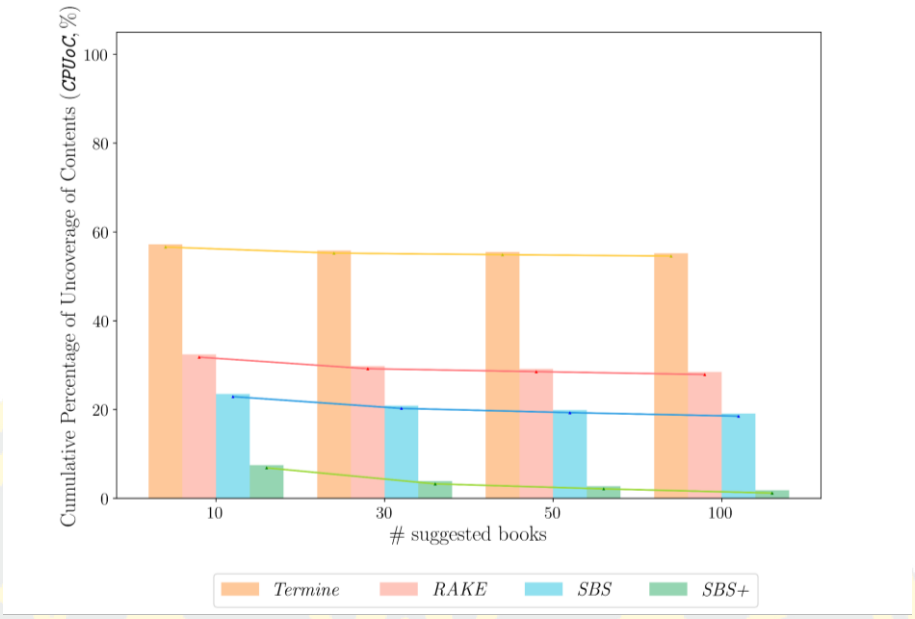
จากข้างต้นสามารถอธิบายได้ว่าความเหมือนของคำสำคัญระหว่างหนังสือและคำอธิบายรายวิชาจะปรากฏมากขึ้นเมื่อจำนวนหนังสือในการแนะนำเพิ่มมากขึ้น บ่งบอกว่าหนังสือแต่ละเล่มจะมีความเกี่ยวเนื่องกับคำอธิบายรายวิชาในเนื้อหาที่แตกต่างกันออกไป เมื่อเลือกอ่านหนังสือเพิ่มมากขึ้นจะสามารถเพิ่มพูนความรู้ความเข้าใจในเนื้อหาต่าง ๆ ที่เกี่ยวข้องกับคำอธิบายรายวิชาได้มากยิ่งขึ้น

และพบว่าระบบ *SBS+* มีค่า $CPCoC_N$ สูงกว่าระบบ *SBS*, *RAKE* และ *Termine* หมายความว่าระบบ *SBS+* สามารถสร้างรายการแนะนำหนังสือได้ครอบคลุมเนื้อหาในคำอธิบายรายวิชาได้มากกว่าทั้ง 3 ระบบแนะนำที่นำมาเปรียบเทียบ

นอกจากการสร้างรายการแนะนำหนังสือโดยพิจารณาจากคะแนนความคล้ายคลึงแล้วนั้น คะแนนความไม่คล้ายคลึงของหนังสือ เราสามารถสร้างรายการแนะนำหนังสือโดยทำการเรียงลำดับจากหนังสือที่มีความไม่เกี่ยวเนื่องมากที่สุดไปยังน้อยที่สุดได้เช่นเดียวกัน โดยการดำเนินการดังกล่าวสามารถแนะนำหนังสือที่มีคำสำคัญที่เพิ่มมากขึ้นหรือแตกต่างออกไปจากคำสำคัญในเนื้อหาคำอธิบายรายวิชา ซึ่งสามารถประเมินประสิทธิภาพของระบบแนะนำได้จากค่า $APUoC_N$ โดยแสดงการเปรียบเทียบค่า $APUoC_{10}$, $APUoC_{30}$, $APUoC_{50}$ และ $APUoC_{100}$ ของระบบแนะนำดังภาพที่ 33 และค่า $CPUoC_N$ (ค่าที่บ่งบอกถึงความแตกต่างระหว่างเนื้อหาคำอธิบายรายวิชาหนึ่ง ๆ และรายการหนังสือที่เกี่ยวข้อง N เล่ม เมื่อหนังสือที่แนะนำมีจำนวนมากขึ้นความแตกต่างระหว่างเนื้อหาคำอธิบายรายวิชาหนึ่ง ๆ และรายการหนังสือที่เกี่ยวข้องก็จะลดลง หรือเรียกได้ว่าเป็นค่าตรงกันข้ามกับค่า $CPCoC_N$) ในภาพที่ 34 แสดงการเปรียบเทียบค่า $CPUoC_{10}$, $CPUoC_{30}$, $CPUoC_{50}$ และ $CPUoC_{100}$ ของระบบแนะนำ



ภาพที่ 33 การเปรียบเทียบค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ยของระบบแนะนำหนังสืออ่านเพิ่มเติม



ภาพที่ 34 การเปรียบเทียบค่าความไม่ครอบคลุมเนื้อหาสะสมของระบบแนะนำหนังสืออ่านเพิ่มเติม

จากภาพที่ 33 สังเกตว่าเมื่อจำนวนในการแนะนำหนังสือเพิ่มมากขึ้น ค่า $APUoC_N$ จะลดลง ซึ่งบ่งบอกว่าหนังสือที่ถูกจัดอันดับอยู่ในอันดับต้น ๆ ของรายการแนะนำคือหนังสือที่มีความแตกต่างกับเนื้อหาคำอธิบายรายวิชามากที่สุด เมื่อทำการเปรียบเทียบค่า $APUoC_N$ ของแต่ละระบบแนะนำมีผลลัพธ์ (โดยเรียงลำดับตามค่า N คือ 10, 30, 50 และ 100) ดังนี้

1. ระบบ *Termine* มีค่า 77.3%, 77.0%, 76.5% และ 74.3%
2. ระบบ *RAKE* มีค่า 71.8%, 71.1%, 70.3% และ 66.4%
3. ระบบ *SBS* มีค่า 56.4%, 55.6%, 54.6% และ 50.3%
4. ระบบ *SBS+* มีค่า 43.5%, 42.6%, 41.5% และ 36.6%

จากข้างต้น พบว่าระบบ *SBS+* มีค่า $APUoC_N$ น้อยกว่าระบบ *Termine* ≈ 35.2%, *RAKE* ≈ 28.8% และ *SBS* ≈ 13.1% และน้อยกว่าทั้ง 3 ระบบโดยเฉลี่ย ≈ 25.7% ซึ่งรายการแนะนำหนังสือที่สร้างขึ้นจากคะแนนความไม่คล้ายคลึง จะทำให้นิสิต/นักศึกษาสามารถเลือกอ่านหนังสือที่มีเนื้อหาเพิ่มเติมมากขึ้นจากในชั้นเรียน (พิจารณาจากคำอธิบายรายวิชา) ได้เรียนรู้เนื้อหาใหม่ ๆ เพื่อเพิ่มพูนทักษะและองค์ความรู้ของตนเอง โดยพบว่าระบบ *SBS+* มีค่า $APUoC_N$ น้อยกว่าระบบอื่น เนื่องจากระบบ *SBS+* สามารถสกัดคำสำคัญได้ถูกต้องและครอบคลุมเนื้อหาของคำอธิบายรายวิชา ส่งผลให้สามารถค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำอธิบายรายวิชาค่อนข้างสูง ดังนั้นรายการแนะนำหนังสือจึงมีหนังสือที่มีความไม่เกี่ยวเนื่องน้อยกว่าทั้ง 3 ระบบแนะนำที่นำมาเปรียบเทียบ

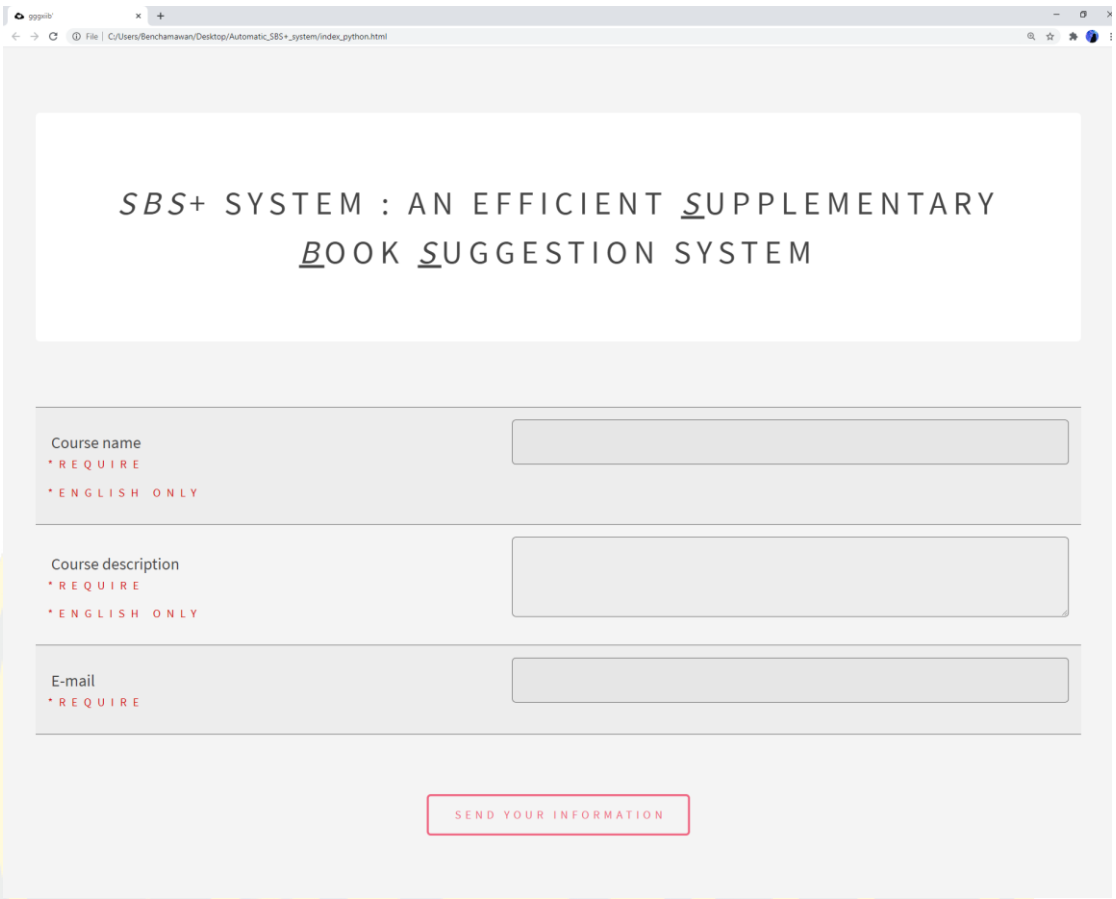
ท้ายสุด จากภาพที่ 34 สามารถสังเกตว่าค่า $CPUoC_N$ จะลดลงเมื่อจำนวนการแนะนำหนังสือมากขึ้น เนื่องจากค่า $CPUoC_N$ เป็นค่าตรงกันข้ามกับค่า $CPCoC_N$ ที่ซึ่งเมื่อจำนวนหนังสือที่แนะนำเพิ่มมากยิ่งขึ้นความแตกต่างของคำสำคัญระหว่างหนังสือและคำอธิบายรายวิชาจะลดลงไปด้วย เมื่อทำการเปรียบเทียบค่า $CPUoC_N$ ของแต่ละระบบแนะนำมีผลลัพธ์ (โดยเรียงลำดับตามค่า N คือ 10, 30, 50 และ 100) ดังนี้

1. ระบบ *Termine* มีค่า 57.3%, 55.9%, 55.5% และ 55.2%
2. ระบบ *RAKE* มีค่า 32.5%, 29.9%, 29.2% และ 28.5%
3. ระบบ *SBS* มีค่า 23.6%, 20.9%, 20.0% และ 19.2%
4. ระบบ *SBS+* มีค่า 7.5%, 3.9%, 2.8% และ 1.8%

จากข้างต้น เห็นได้ว่าระบบ *SBS+* มีค่า $CPUoC_N$ ที่น้อยกว่าระบบ *Termine* ≈ 51.9%, *RAKE* ≈ 26.0% และ *SBS* ≈ 16.9% และน้อยกว่าทั้ง 3 ระบบโดยเฉลี่ย ≈ 31.6% เนื่องจากระบบ *SBS+* สามารถสร้างรายการแนะนำหนังสือได้ครอบคลุมเนื้อหาของคำอธิบายรายวิชา ดังนั้นหนังสือที่ถูกแนะนำจึงมีความแตกต่างกับเนื้อหาของคำอธิบายรายวิชาน้อยกว่าทั้ง 3 ระบบแนะนำที่นำมาเปรียบเทียบ

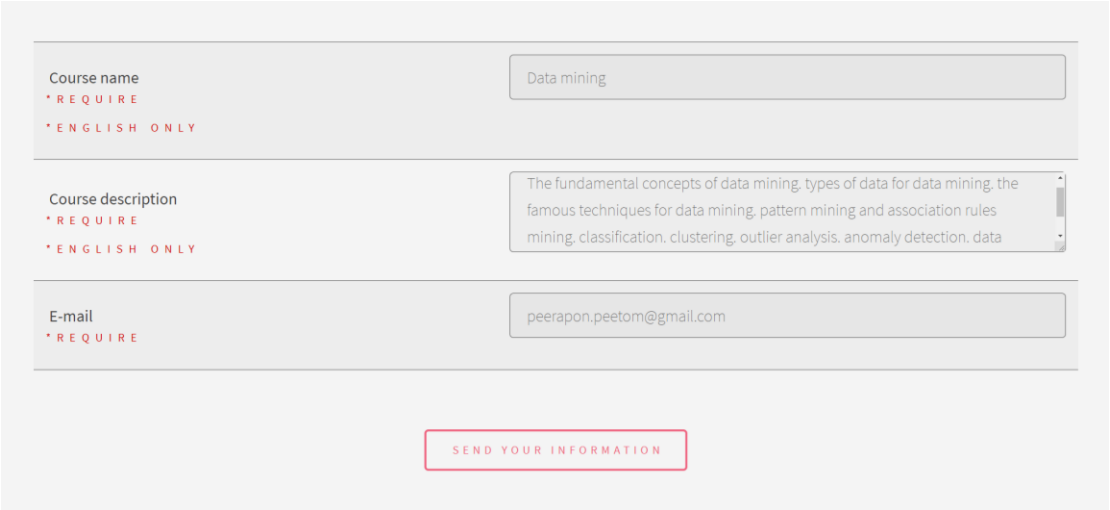
5.3 การใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม

ในส่วนนี้อธิบายถึงระบบแนะนำหนังสืออ่านเพิ่มเติมทางด้านวิทยาการคอมพิวเตอร์ที่ผู้วิจัยได้พัฒนาขึ้นเพื่อสร้างรายการแนะนำหนังสือให้กับนิสิต/นักศึกษา อาจารย์ผู้สอนรายวิชา หรือผู้ที่สนใจค้นหาหนังสืออ่านเพิ่มเติมนอกชั้นเรียน โดยได้ออกแบบส่วนติดต่อผู้ใช้งานของระบบ *SBS+* เพื่อให้ผู้ใช้งานสามารถกรอกข้อมูลคำค้นที่ต้องการ ซึ่งในหน้าต่างการรับข้อมูล ประกอบไปด้วย 1) ชื่อรายวิชา 2) คำอธิบายรายวิชา และ 3) อีเมล์ของผู้ใช้งานสำหรับตอบกลับรายการแนะนำหนังสืออ่านเพิ่มเติม แสดงดังภาพที่ 35



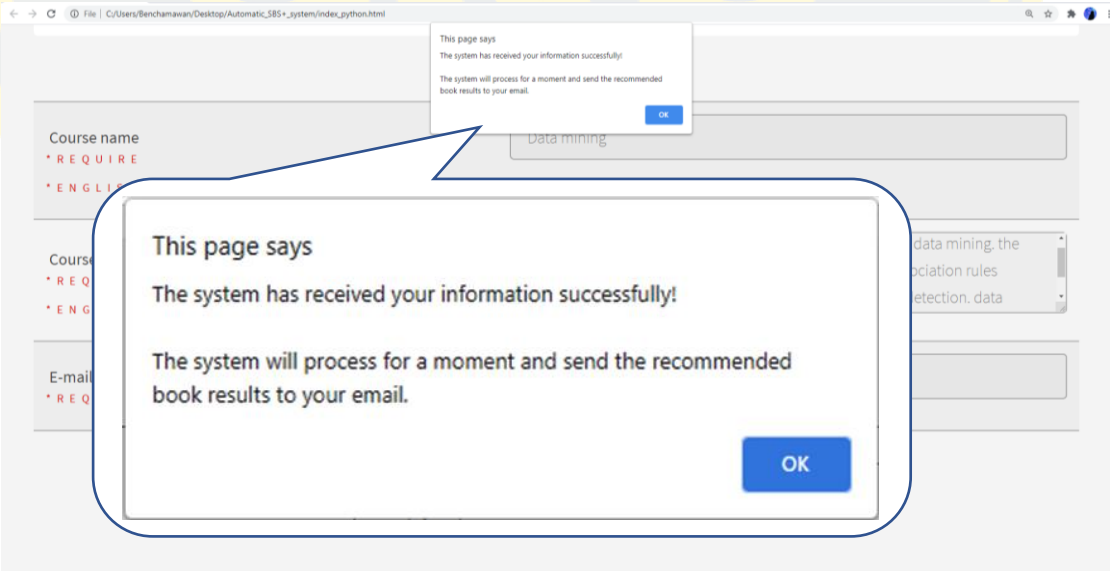
ภาพที่ 35 การแสดงผลเพื่อรับข้อมูลจากผู้ใช้งานของระบบแนะนำหนังสืออ่านเพิ่มเติม

เมื่อผู้ใช้งานต้องการค้นหาหนังสือที่เกี่ยวข้องกับรายวิชาที่สนใจ สามารถกรอกข้อมูลคำค้นเพื่อเป็นข้อมูลรับเข้าของระบบ *SBS+* โดยมีข้อมูลที่ต้องกรอกมี 3 ส่วนคือ ชื่อรายวิชา (ภาษาอังกฤษ), คำอธิบายรายวิชา (ภาษาอังกฤษ) และอีเมล์สำหรับตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมให้ครบถ้วน จากนั้นกดปุ่ม “SEND YOUR INFORMATION” ระบบจะทำการบันทึกข้อมูลรับเข้าจากผู้ใช้งาน ตัวอย่างดังภาพที่ 36 ผู้ใช้งานทำการกรอกข้อมูลชื่อรายวิชา คือ “Data mining”, คำอธิบายรายวิชาคือ “The fundamental concepts of data mining. types of data for data mining. the famous techniques for data mining. pattern mining and association rules mining. classification. clustering. outlier analysis. anomaly detection. data mining tools.” และอีเมล์สำหรับตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมคือ “peerapon.peetom@gmail.com”



ภาพที่ 36 ผู้ใช้งานกรอกข้อมูลคำค้นที่สนใจในหน้าต่างการรับข้อมูลของระบบแนะนำ

เมื่อผู้ใช้งานกดปุ่ม “SEND YOUR INFORMATION” ระบบจะทำการบันทึกข้อมูลรับเข้า เมื่อบันทึกสำเร็จจะแสดงหน้าต่างดังภาพที่ 37 เพื่อให้ผู้ใช้งานทราบว่าระบบได้ทำการบันทึกข้อมูลรับเข้า และแจ้งผู้ใช้งานว่าระบบจะทำการประมวลผลและส่งผลลัพธ์ตอบกลับไปยังอีเมล์ของผู้ใช้งาน



ภาพที่ 37 การแสดงผลเมื่อระบบทำการบันทึกข้อมูลของผู้ใช้งานเรียบร้อยแล้ว

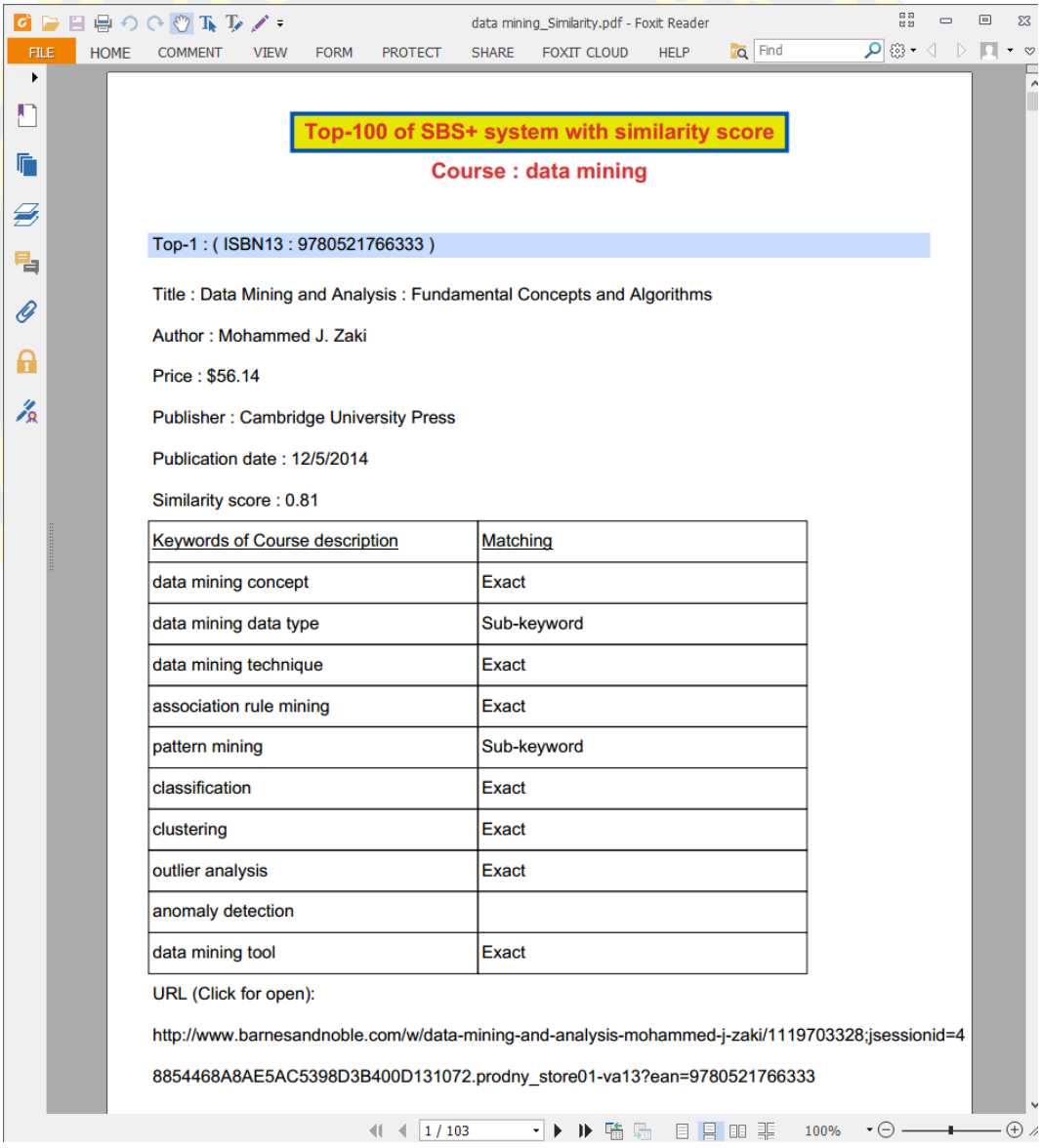
จากนั้นข้อมูลรับเข้าดังภาพที่ 36 จะถูกนำไปทำการประมวลผลโดยอัตโนมัติในระบบ *SBS+* ซึ่งข้อมูลที่จะนำไปประมวลผลคือ ชื่อรายวิชาและคำอธิบายรายวิชา และทำการประมวลผลตามขั้นตอนที่ได้กล่าวไว้ในหัวข้อที่ 4.2 คือ การสกัดคำสำคัญจากคำอธิบายรายวิชา, การค้นหาหนังสือที่

เกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา และการคำนวณความเกี่ยวเนื่องของหนังสือที่ครอบคลุมเนื้อหาในคำอธิบายรายวิชา

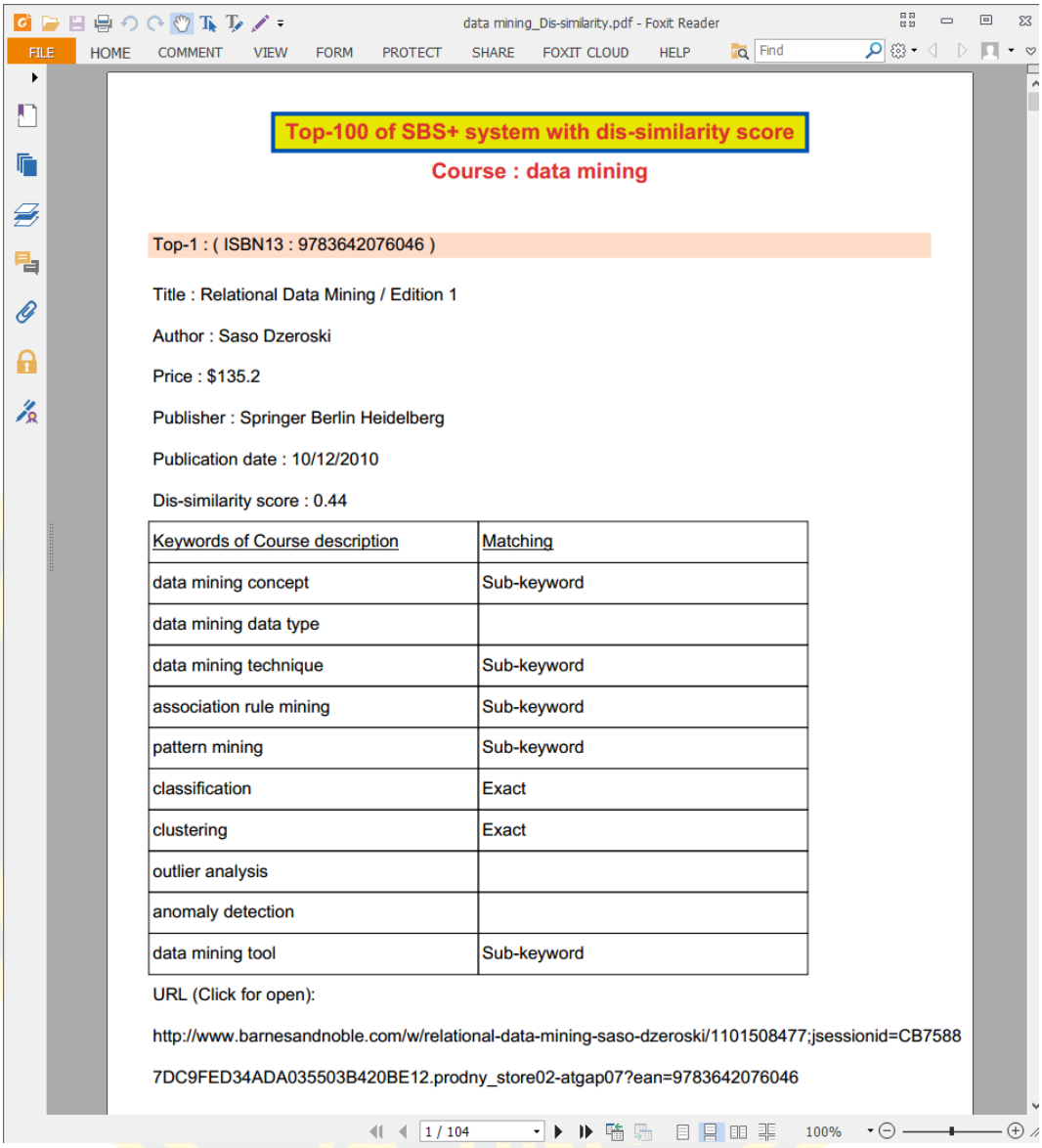
หลังจากทำการประมวลผลแล้วเสร็จจะได้รายการแนะนำหนังสือ 2 รายการที่อยู่ในรูปแบบไฟล์เอกสารข้อมูล (.pdf) คือ

3. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง
4. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง

แสดงดังภาพที่ 38 และภาพที่ 39 ตามลำดับ



ภาพที่ 38 รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง



ภาพที่ 39 รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง

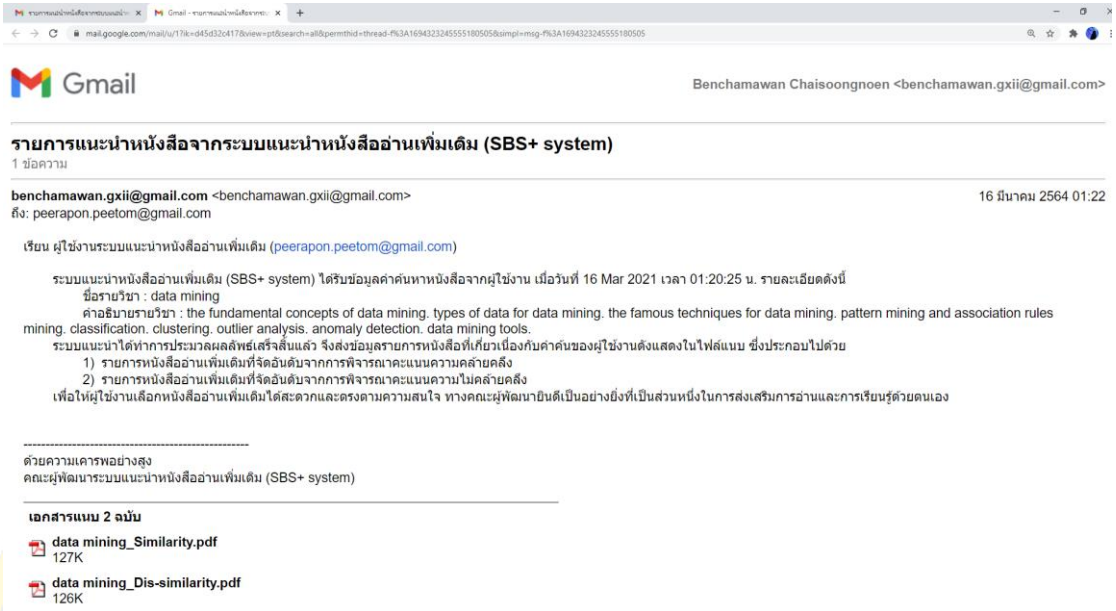
จากภาพที่ 38 และภาพที่ 39 แสดงผลลัพธ์รายการหนังสืออ่านเพิ่มเติมที่ถูกแนะนำให้กับผู้ใช้งานอยู่ในรูปแบบไฟล์เอกสารข้อมูล (.pdf) ประกอบไปด้วยข้อมูลดังนี้

1. ชื่อรายการหนังสือ (กรอบสีน้ำเงิน) จะบอกได้ว่ารายการหนังสือที่ผู้ใช้งานกำลังพิจารณาอยู่นั้น เป็นรายการหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึงหรือคะแนนความไม่คล้ายคลึง
2. ชื่อรายวิชา (ตัวอักษรสีแดง) บอกถึงรายวิชาที่ผู้ใช้งานกรอกข้อมูลเข้ามาเป็นรายวิชาใด

3. ข้อมูลหนังสือที่ทำการแนะนำแต่ละเล่ม ซึ่งประกอบไปด้วย
    i) อันดับของหนังสือที่ถูกแนะนำ
    ii) หมายเลข 13 หลักประจำหนังสือ (ISBN13)
    iii) ชื่อหนังสือ
    iv) ชื่อผู้แต่ง
    v) ราคา
    vi) ชื่อสำนักพิมพ์
    vii) วันที่ตีพิมพ์หนังสือ
    viii) คะแนนความคล้ายคลึง หรือคะแนนความไม่คล้ายคลึง
    ix) ตารางรายการคำสำคัญของคำอธิบายรายวิชาที่ปรากฏในหนังสือเล่มหนึ่ง ๆ
    x) ลิงก์เชื่อมโยงไปยังเว็บไซต์ของหนังสือเล่มนั้น ๆ

จากนั้น ระบบ *SBS+* จะทำการตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมไปยังผู้ใช้งานโดยอัตโนมัติ ซึ่งประกอบไปด้วยข้อความที่บอกให้ผู้ใช้งานทราบถึงข้อมูลที่ถูกรับเข้ามาในระบบ คือ 1) ชื่อรายวิชา 2) คำอธิบายรายวิชา 3) วันที่และเวลาที่ผู้ใช้งานได้กรอกข้อมูลเข้าระบบ และ 4) ผลลัพธ์รายรายการแนะนำหนังสืออ่านเพิ่มเติมในรูปแบบไฟล์เอกสารข้อมูล (.pdf) ทั้ง 2 รายการที่ได้กล่าวไปข้างต้น แสดงดังภาพที่ 40

เมื่อผู้ใช้งานได้รับอีเมล์ตอบกลับจากระบบเรียบร้อยแล้ว แสดงดังภาพที่ 41 สามารถเปิดดูผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมจากระบบ *SBS+* และพิจารณาเลือกอ่านหนังสืออ่านเพิ่มเติม โดยไฟล์เอกสารข้อมูล (.pdf) ที่ผู้วิจัยตอบกลับไปนั้นสามารถคลิกที่ลิงก์เชื่อมโยงไปยังเว็บไซต์ของหนังสือเล่มหนึ่ง ๆ ที่ผู้ใช้งานสนใจจากภายในไฟล์ได้ทันที แสดงดังภาพที่ 42



Gmail

Benchamawan Chaisoongnoen <benchamawan.gxii@gmail.com>

รายการแนะนำหนังสือจากระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system)

1 ข้อความ

benchamawan.gxii@gmail.com <benchamawan.gxii@gmail.com> 16 มีนาคม 2564 01:22
ถึง: peerapon.peetom@gmail.com

เรียน ผู้ใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม (peerapon.peetom@gmail.com)

ระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system) ได้รับข้อมูลคำค้นหาหนังสือจากผู้ใช้งาน เมื่อวันที่ 16 Mar 2021 เวลา 01:20:25 น. รายละเอียดดังนี้
ชื่อรายวิชา : data mining
คำอธิบายรายวิชา : the fundamental concepts of data mining. types of data for data mining. the famous techniques for data mining. pattern mining and association rules mining. classification. clustering. outlier analysis. anomaly detection. data mining tools.
ระบบแนะนำได้ทำการประมวลผลลัพธ์เสร็จสิ้นแล้ว จึงส่งข้อมูลรายการหนังสือที่เกี่ยวเนื่องกับคำค้นของผู้ใช้งานดังแสดงในไฟล์แนบ ซึ่งประกอบไปด้วย
1) รายการหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง
2) รายการหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง
เพื่อให้ผู้ใช้งานเลือกหนังสืออ่านเพิ่มเติมได้สะดวกและตรงตามความสนใจ ทางคณะผู้พัฒนายินดีเป็นอย่างยิ่งที่เป็นส่วนหนึ่งในการส่งเสริมการอ่านและการเรียนรู้ด้วยตนเอง

ด้วยความเคารพอย่างสูง
คณะผู้พัฒนาระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system)

เอกสารแนบ 2 ฉบับ

data mining_Similarity.pdf
127K

data mining_Dis-similarity.pdf
126K

ภาพที่ 40 การตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมไปยังผู้ใช้งาน

รายการแนะนำหนังสือจากระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system) กล่องจดหมาย ×

benchamawan.gxii@gmail.com 01:22 (8 นาทีที่ผ่านมา)
ถึง ฉัน

เรียน ผู้ใช้งานระบบแนะนำหนังสืออ่านเพิ่มเติม (peerapon.peetom@gmail.com)

ระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system)
ได้รับข้อมูลคำค้นหาหนังสือจากผู้ใช้งาน เมื่อวันที่ 16 Mar 2021 เวลา
01:20:25 น. รายละเอียดดังนี้
ชื่อรายวิชา : data mining
คำอธิบายรายวิชา : the fundamental concepts of data mining. types of data for data mining. the famous techniques for data mining. pattern mining and association rules mining. classification. clustering. outlier analysis. anomaly detection. data mining tools.
ระบบแนะนำได้ทำการประมวลผลลัพธ์เสร็จสิ้นแล้ว
จึงส่งข้อมูลรายการหนังสือที่เกี่ยวเนื่องกับคำค้นของผู้ใช้งานดังแสดงในไฟล์แนบ
ซึ่งประกอบไปด้วย
1) รายการหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง
2) รายการหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง
เพื่อให้ผู้ใช้งานเลือกหนังสืออ่านเพิ่มเติมได้สะดวกและตรงตามความสนใจ
ทางคณะผู้พัฒนายินดีเป็นอย่างยิ่งที่เป็นส่วนหนึ่งในการส่งเสริมการอ่านและการเรียนรู้ด้วยตนเอง

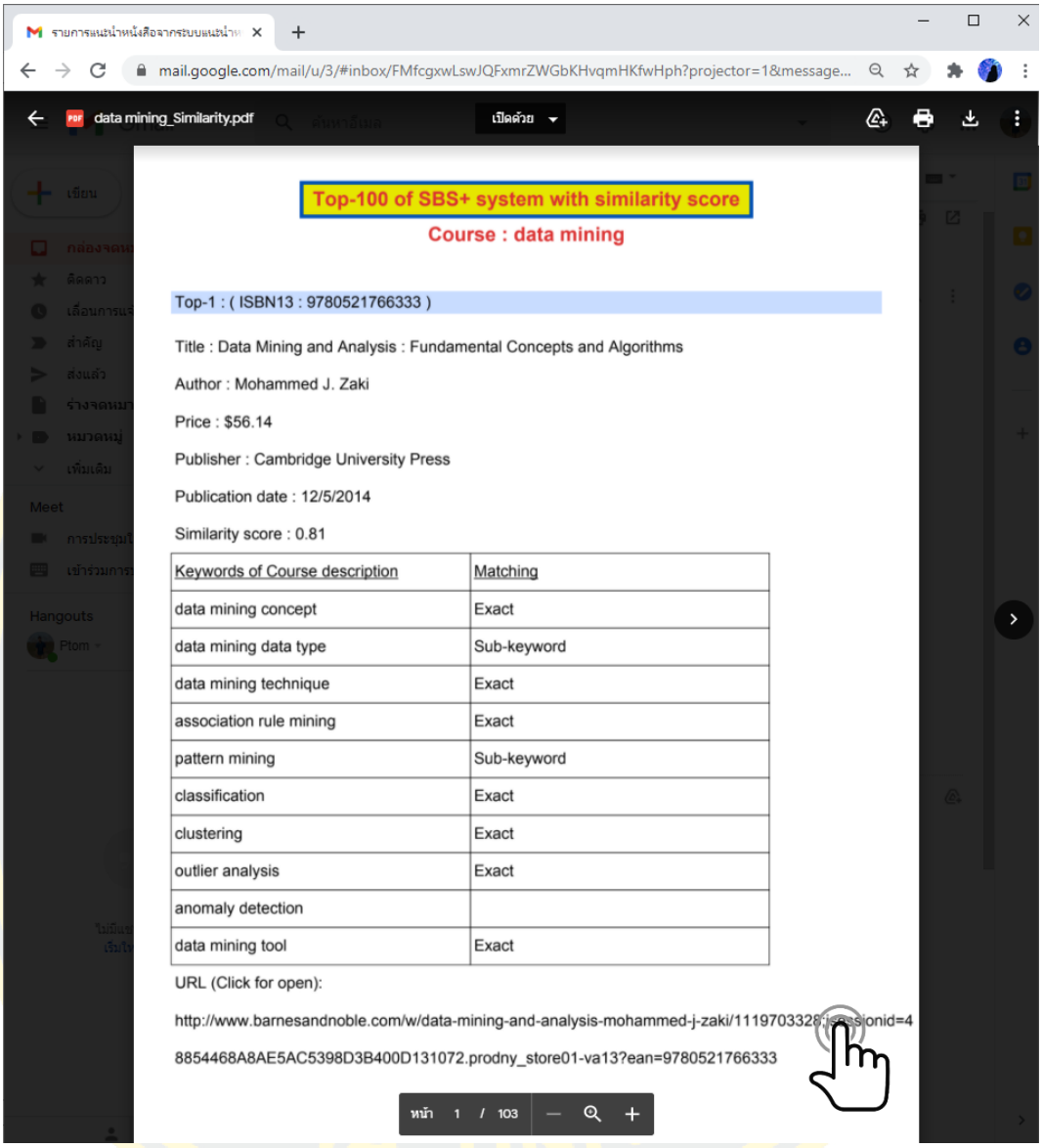
ด้วยความเคารพอย่างสูง
คณะผู้พัฒนาระบบแนะนำหนังสืออ่านเพิ่มเติม (SBS+ system)

2 ไฟล์แนบ

data mining_Simila...

data mining_Dis-si...

ภาพที่ 41 หน้าต่างอีเมล์ของผู้ใช้งานที่ได้รับอีเมล์ตอบกลับจากระบบ SBS+



ภาพที่ 42 ผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมที่ผู้ใช้งานได้รับจากอีเมล์ตอบกลับ

บทที่ 6
สรุปและอภิปรายผล

6.1 สรุปผลการดำเนินงาน

ปัจจุบันมีระบบค้นหาหนังสือและระบบแนะนำหนังสือต่าง ๆ ที่ช่วยเพิ่มความสะดวกให้แก่ให้ผู้อ่านในการค้นหาหนังสือได้ดียิ่งขึ้น อาทิเช่น การค้นหาหนังสือจากความสอดคล้องของคำค้นหนึ่ง ๆ กับชื่อหนังสือหรือสารบัญหนังสือ การแนะนำหนังสือจากความนิยมของหนังสือ การแนะนำหนังสือจากข้อคิดเห็นที่ได้รับจากผู้อ่าน การแนะนำหนังสือจากพฤติกรรมการยืมหนังสือในห้องสมุด เป็นต้น แต่อย่างไรก็ตามแนวคิดทั้งหมดไม่ได้มุ่งเน้นที่การแนะนำหนังสือจากความสอดคล้องของหนังสือที่เกี่ยวเนื่องกับการศึกษาในชั้นเรียน และไม่ได้มุ่งเน้นในการแนะนำหนังสือที่เกี่ยวข้องกับศาสตร์ที่มีคำศัพท์เฉพาะทางค่อนข้างมากกับคำค้น ที่ซึ่งช่วยให้ผู้อ่านได้รับการแนะนำที่สอดคล้องกับเนื้อหาที่ตนเองสนใจมากยิ่งขึ้น

ดังนั้นงานวิจัยนี้จึงนำเสนอระบบแนะนำหนังสืออ่านเพิ่มเติม (Supplementary Book Suggestion system) หรือเรียกโดยย่อว่า “ระบบ *SBS*” ระบบที่นำเสนอมุ่งเน้นที่การแนะนำหนังสือโดยพิจารณาจากความสอดคล้องของหนังสือที่ควรอ่านเพิ่มเติมจากการศึกษาในชั้นเรียนด้วยการพิจารณาเนื้อหาของคำอธิบายรายวิชา และ/หรือประมวลรายวิชา โดยขั้นตอนหลักของการประมวลผลประกอบด้วย 2 ขั้นตอน คือ

1. การเตรียมข้อมูลเบื้องต้น
    1.1 การรวบรวมข้อมูลคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์
    1.2 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี
    1.3 การสร้างกฎทางภาษาศาสตร์
2. การประมวลผลของระบบแนะนำ
    2.1 การสกัดคำสำคัญโดยประยุกต์ ประกอบด้วย 3 ขั้นตอน คือ
        2.1.1 การประมวลผลข้อความ ประกอบด้วย 4 ขั้นตอน คือ
            i) การกำจัดคำหยุด (Stop word removal)
            ii) การแก้ไขคำผิด (Word error correction)
            iii) การตัดประโยค (Sentence tokenization)
            iv) การวิเคราะห์หน้าที่ของคำในประโยค (Part of speech tagging)

2.1.2 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี

2.1.3 การระบุคำสำคัญโดยกฎทางภาษาศาสตร์

2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำอธิบายรายวิชา

2.3 การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา ประกอบด้วย 3 ขั้นตอน คือ

2.3.1 การเปรียบเทียบคำสำคัญระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้องโดยวิธีการเปรียบเทียบแบบตรงตัว

2.3.2 การจัดอันดับหนังสือ โดยเรียงลำดับจากความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชามากที่สุดไปน้อยที่สุด

2.3.3 สร้างรายการแนะนำหนังสือจำนวน N เล่ม (ค่า N ที่สนใจคือ 10, 30, 50 และ 100 ตามลำดับ)

จากข้างต้น ระบบ *SBS* สามารถแนะนำหนังสืออ่านเพิ่มเติมที่มีเนื้อหาคล้ายคลึงกับคำอธิบายรายวิชาให้แก่ผู้ใช้งานได้ โดยเมื่อพิจารณาขั้นตอนการทำงานของระบบ *SBS* จะสังเกตได้ว่ายังสามารถพัฒนาประสิทธิภาพในการสกัดคำสำคัญของระบบ *SBS* ได้ นอกจากนี้การหาความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง ที่ประยุกต์ใช้วิธีการเปรียบเทียบแบบตรงตัว (Exact matching) เพียงอย่างเดียวทำให้ประสิทธิผลไม่เพียงพอในการคำนวณคะแนนความคล้ายคลึง เนื่องจากไม่สามารถบ่งบอกถึงความคล้ายคลึงระหว่างบางส่วนของเนื้อหาที่ปรากฏในหนังสือกับเนื้อหา (หรือบางส่วนของเนื้อหา) ที่ปรากฏในคำอธิบายรายวิชาได้

ด้วยเหตุนี้ผู้วิจัยจึงได้ทำการปรับปรุงระบบ *SBS* เดิม โดยพัฒนาเพิ่มเติมในส่วนของการสกัดคำสำคัญและการหาความคล้ายคลึงระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง โดยเรียกระบบที่พัฒนาขึ้นใหม่ว่า “ระบบแนะนำหนังสืออ่านเพิ่มเติมที่มีประสิทธิภาพ (An efficient Supplementary Book Suggestion system)” หรือ “ระบบ *SBS+*” การประมวลผลของระบบ *SBS+* ประกอบไปด้วย 2 ขั้นตอนหลักเช่นเดียวกับระบบ *SBS* แต่รายละเอียดของการคำนวณแต่ละขั้นตอนมีการพัฒนาที่ต่างจากเดิมดังนี้

1. การเตรียมข้อมูลเบื้องต้น

1.1 การรวบรวมคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี (จำนวน 28,392 คำ เพิ่มขึ้นจากเดิมที่รวบรวมไว้เพียง 6,527 คำ)

1.2 การรวบรวมคำคุณศัพท์ (จำนวน 4,998 คำ เป็นขั้นตอนที่เพิ่มขึ้น ซึ่งเดิมไม่ได้รวบรวม)

1.3 การการปรับปรุงกฎทางภาษาศาสตร์ เพื่อลดความซ้ำซ้อนในการสกัดคำสำคัญ

2. การประมวลผลของระบบแนะนำ

2.1 การสกัดคำสำคัญ ประกอบไปด้วย 3 ขั้นตอน คือ

2.1.1 การประมวลผลข้อความ ได้ทำการเรียงลำดับขั้นตอนในการประมวลผลเพิ่มขั้นตอนใหม่เพื่อลดความผิดพลาดและช่วยให้ข้อความมีความสมบูรณ์/มีความพร้อมใช้งานในขั้นตอนต่อไปมากยิ่งขึ้น ประกอบไปด้วย 5 ขั้นตอน คือ

i) การตัดประโยค (Sentence tokenization)

ii) การตัดคำ (Word tokenization)

iii) การแก้ไขคำผิด (Word error correction)

iv) การกำจัดคำหยุด (Stop word removal)

v) การวิเคราะห์หน้าที่ของคำ (Part of speech tagging)

2.1.2 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี ได้ประยุกต์ใช้การลดรูปของคำและการหารูปเดิมของคำ (Word stemming and lemmatization) และ เพิ่มการเปรียบเทียบคำศัพท์เฉพาะซ้ำ 2 รอบ (รอบที่ 1 เปรียบเทียบหาคำศัพท์เฉพาะก่อนการลดรูปของคำและการหารูปเดิมของคำ และรอบที่ 2 เปรียบเทียบหาคำศัพท์เฉพาะหลังการลดรูปของคำและการหารูปเดิมของคำ)

2.1.3 การระบุคำสำคัญโดยประยุกต์ใช้กฎทางภาษาศาสตร์ฉบับปรับปรุง

2.2 การค้นหาหนังสือที่มีความเกี่ยวเนื่องกับคำสำคัญในคำอธิบายรายวิชา

2.3 การคำนวณความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชา ประกอบด้วย 3 ขั้นตอน คือ

2.3.1 การเปรียบเทียบคำสำคัญระหว่างคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้อง ประกอบไปด้วย 3 วิธีการ คือ

i) วิธีการเปรียบเทียบแบบตรงตัว

ii) วิธีการเปรียบเทียบแบบวลี

iii) วิธีการเปรียบเทียบแบบพิจารณาคำย่อย

2.3.2 การจัดอันดับหนังสือ โดยเรียงลำดับจากความเกี่ยวเนื่องของหนังสือกับคำอธิบายรายวิชามากที่สุดไปน้อยที่สุด

2.3.3 สร้างรายการแนะนำหนังสือจำนวน N เล่ม (ค่า N ที่สนใจคือ 10, 30, 50 และ 100 ตามลำดับ)

งานวิจัยนี้ได้ทำการรวบรวมคำอธิบายรายวิชาในหลักสูตรวิทยาการคอมพิวเตอร์ทั้งหมด จำนวน 640 วิชา จาก 13 มหาวิทยาลัย (มีคำสำคัญที่ถูกต้องทั้งหมด 9,279 คำ) เพื่อใช้เป็นข้อมูลทดสอบของระบบแนะนำหนังสืออ่านเพิ่มเติม และเนื่องจากการสกัดคำสำคัญจากคำอธิบายรายวิชาและหนังสือที่เกี่ยวข้องเป็นขั้นตอนหลักของระบบ จึงทำการเปรียบเทียบวิธีการสกัดคำสำคัญกับขั้นตอนวิธี "*Termine*" และขั้นตอนวิธี "*RAKE*" ซึ่งเป็นวิธีการสกัดคำสำคัญที่ได้รับความนิยม ดังนั้นการประเมินประสิทธิภาพของระบบแนะนำหนังสืออ่านเพิ่มเติมแบ่งออกเป็น 2 ส่วน ได้แก่

1. การประเมินประสิทธิภาพของการสกัดคำสำคัญที่ใช้ในระบบ โดยประยุต์ใช้การประเมินค่าความแม่นยำโดยรวม ($\boldsymbol{Accuracy}$), ค่าความแม่นยำ ($\boldsymbol{Precision}$), ค่าความถูกต้อง ($\boldsymbol{Recall}$) และ ค่าประสิทธิภาพโดยรวม ($\boldsymbol{F\text{-}measure}$) โดยจากผลลัพธ์การทดสอบประสิทธิภาพของการสกัดคำสำคัญจะเห็นได้ว่าขั้นตอนวิธีการสกัดคำสำคัญของระบบ *SBS+* มีประสิทธิภาพมากที่สุดเนื่องจากสามารถสกัดคำสำคัญที่ไม่ซ้ำซ้อนและกำจัดคำสำคัญที่ไม่ได้บ่งบอกถึงใจความสำคัญออกไป ดังนั้นค่า $\boldsymbol{Accuracy}$ ของขั้นตอนวิธีการสกัดคำสำคัญของระบบ *SBS+* จึงมีค่ามากกว่าขั้นตอนวิธีอื่น ๆ นอกจากนั้นระบบ *SBS+* มีค่า $\boldsymbol{Precision}$ และมีค่า $\boldsymbol{Recall}$ มากที่สุด เนื่องจากสามารถสกัดคำสำคัญที่ถูกต้องได้ค่อนข้างมากจากคำสำคัญที่สกัดได้ทั้งหมด และสามารถสกัดคำสำคัญได้ครอบคลุมเนื้อหาของคำอธิบายรายวิชา ซึ่งจากความสามารถทั้งสองส่งผลให้ค่า $\boldsymbol{F\text{-}measure}$ ที่บ่งบอกถึงประสิทธิภาพโดยรวมของระบบ *SBS+* มีค่ามากกว่าขั้นตอนวิธีอื่น ๆ
2. การประเมินการประมวลผลของระบบแนะนำ โดยพิจารณาจากค่าความครอบคลุมเนื้อหาโดยเฉลี่ย ค่าความครอบคลุมเนื้อหาสะสม ค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ย และ ค่าความไม่ครอบคลุมเนื้อหาสะสม ตามลำดับ โดยผลลัพธ์แสดงให้เห็นว่า i) ระบบ *SBS+* มีค่าความครอบคลุมเนื้อหาโดยเฉลี่ยที่สูงกว่าระบบแนะนำอื่น ๆ ซึ่งทำให้สามารถสร้างรายการแนะนำหนังสือที่มีความเกี่ยวเนื่องกับคำอธิบายรายวิชาได้มากกว่าทั้ง 3 ระบบ ii) ค่าความครอบคลุมเนื้อหาสะสมของระบบ *SBS+* มีค่าสูงกว่าระบบแนะนำอื่น ๆ เช่นเดียวกัน บ่งบอกได้ว่าระบบ *SBS+* สามารถสร้างรายการแนะนำหนังสือได้ครอบคลุมเนื้อหาในคำอธิบายรายวิชาได้มากกว่า กลับกัน iii) ระบบ *SBS+* มีค่าความไม่ครอบคลุมเนื้อหาโดยเฉลี่ยน้อยกว่าระบบแนะนำอื่น ๆ ซึ่งทำให้รายการแนะนำหนังสือที่สร้างขึ้นจากคะแนนความไม่คล้ายคลึงจึงมีหนังสือที่มีความไม่เกี่ยวเนื่องน้อยกว่าระบบอื่น ๆ และ iv) ระบบ *SBS+* มีค่าความไม่ครอบคลุมเนื้อหาสะสมน้อยกว่าระบบแนะนำอื่น ๆ ซึ่งบ่งบอกได้ว่าระบบ *SBS+* สามารถสร้างรายการแนะนำหนังสือที่มีความแตกต่างกับเนื้อหาของคำอธิบายรายวิชาได้น้อยกว่าระบบอื่น ๆ

นอกเหนือจากแบบจำลองระบบการประมวลผลข้างต้น ผู้วิจัยได้พัฒนาเว็บไซต์สำหรับรับข้อมูลเข้าให้กับระบบแนะนำหนังสืออ่านเพิ่มเติม จากนั้นสร้างรายการแนะนำหนังสือให้กับนิสิต/นักศึกษา อาจารย์ผู้สอนรายวิชา หรือผู้ที่สนใจค้นหาหนังสืออ่านเพิ่มเติมนอกชั้นเรียน โดยผู้ใช้งานสามารถกรอกข้อมูลคำค้นที่สนใจ (คำอธิบายรายวิชา และ/หรือประมวลรายวิชา) เข้ามาในระบบ ประกอบไปด้วย 1) ชื่อรายวิชา 2) คำอธิบายรายวิชา และ 3) อีเมล์ของผู้ใช้งานสำหรับตอบกลับรายการหนังสืออ่านเพิ่มเติม จากนั้นเว็บไซต์จะนำข้อมูลจากผู้ใช้งานไปประมวลผลตามขั้นตอนของระบบ *SBS+* ที่ได้กล่าวไว้ข้างต้น ผลลัพธ์จากการประมวลผลการแนะนำจะได้รายการแนะนำหนังสือ 2 รายการที่อยู่ในรูปแบบไฟล์เอกสารข้อมูล (.pdf) คือ

1. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความคล้ายคลึง
2. รายการแนะนำหนังสืออ่านเพิ่มเติมที่จัดอันดับจากการพิจารณาคะแนนความไม่คล้ายคลึง

จากนั้นผู้วิจัยจะทำการตอบกลับผลลัพธ์รายการแนะนำหนังสืออ่านเพิ่มเติมไปยังผู้ใช้งาน เพื่อพิจารณาเลือกอ่านหนังสือได้ตรงตามความสนใจ

6.2 ข้อเสนอแนะ

1. ระบบแนะนำหนังสืออ่านเพิ่มเติมมุ่งเน้นในการแนะนำหนังสือเกี่ยวกับศาสตร์ทางด้านวิทยาการคอมพิวเตอร์ ในอนาคตสามารถนำไปประยุกต์ใช้ในหลักสูตร/สาขาวิชาอื่น ๆ ได้
2. การเปรียบเทียบความคล้ายคลึงที่ใช้ในระบบแนะนำ โดยใช้วิธีการเปรียบเทียบแบบตรงตัว, แบบวลี และแบบพิจารณาคำย่อย เป็นการเปรียบเทียบความคล้ายคลึงเชิงอักขระ (String matching) สามารถพัฒนาการเปรียบเทียบความคล้ายคลึงเชิงความหมาย (Semantic matching) ระหว่างเนื้อหาคำอธิบายรายวิชาและเนื้อหาของหนังสือที่เกี่ยวข้องได้ ซึ่งจะทำให้ระบบแนะนำหนังสือมีความแม่นยำ และแนะนำหนังสือได้ตรงกับความสนใจของผู้ใช้งานมากขึ้น

บรรณานุกรม

Akabuike, I. G. and Asika, I. E. (2012) 'Reading habits of undergraduates and their academic performances: Issues and perspectives', *African Research Review*, 6(2), pp. 246–257.

Ali, Z., Khusro, S. and Ullah, I. (2016) 'A hybrid book recommender system based on table of contents (toc) and association rule mining', in *Proceedings of the 10th International Conference on Informatics and Systems*, pp. 68–74.

Balakrishnan, V. and Lloyd-Yemoh, E. (2014) 'Stemming and lemmatization: A comparison of retrieval performances'.

Beliga, S. (2014) 'Keyword extraction: a review of methods and approaches', *University of Rijeka, Department of Informatics, Rijeka*, pp. 1–9.

Biswas, S. K. (2019) 'Keyword extraction from tweets using weighted graph', in *Cognitive Informatics and Soft Computing*. Springer, pp. 475–483.

Çano, E. and Morisio, M. (2017) 'Hybrid recommender systems: A systematic literature review', *Intelligent Data Analysis*, 21(6), pp. 1487–1524.

Chandak, M., Girase, S. and Mukhopadhyay, D. (2015) 'Introducing hybrid technique for optimization of book recommender system', *Procedia Computer Science*, 45, pp. 23–31.

Chen, L.-C., Kuo, P.-J. and Liao, I.-E. (2015) 'Ontology-based library recommender system using MapReduce', *Cluster Computing*, 18(1), pp. 113–121.

Frantzi, K., Ananiadou, S. and Mima, H. (2000) 'Automatic recognition of multi-word terms:. the c-value/nc-value method', *International journal on digital libraries*, 3(2), pp. 115–130.

Garrido, A. L. and Ilarri, S. (2014) 'TMR: a semantic recommender system using topic maps on the items' descriptions', in *European Semantic Web Conference*. Springer, pp. 213–217.

Garrido, A. L., Pera, M. S. and Ilarri, S. (2014) 'SOLE-R: A semantic and linguistic approach for book recommendations', in *2014 IEEE 14th International Conference on Advanced Learning Technologies*. IEEE, pp. 524–528.

Gemkow, T. *et al.* (2018) 'Automatic glossary term extraction from large-scale requirements specifications', in *2018 IEEE 26th International Requirements Engineering Conference (RE)*. IEEE, pp. 412–417.

Gomaa, W. H. and Fahmy, A. A. (2013) 'A survey of text similarity approaches', *International Journal of Computer Applications*, 68(13), pp. 13–18.

Gupta, S. (2015) 'A correction model for real-word errors', *Procedia Computer Science*, 70, pp. 99–106.

Hameed, M. A., Al Jadaan, O. and Ramachandram, S. (2012) 'Collaborative filtering based recommendation system: A survey', *International Journal on Computer Science and Engineering*, 4(5), p. 859.

Hulth, A. (2003) 'Improved automatic keyword extraction given more linguistic knowledge', in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pp. 216–223.

Isinkaye, F. O., Folajimi, Y. O. and Ojokoh, B. A. (2015) 'Recommendation systems: Principles, methods and evaluation', *Egyptian Informatics Journal*, 16(3), pp. 261–273.

Jomsri, P. (2018) 'FUCL mining technique for book recommender system in library service', *Procedia Manufacturing*, 22, pp. 550–557.

Kim, J. K. *et al.* (2010) 'A group recommendation system for online communities', *International journal of information management*, 30(3), pp. 212–219.

Linden, G., Smith, B. and York, J. (2003) 'Amazon. com recommendations: Item-to-item collaborative filtering', *IEEE Internet computing*, 7(1), pp. 76–80.

Lopez-Gazpio, I. *et al.* (2019) 'Word n-gram attention models for sentence similarity and inference', *Expert Systems with Applications*, 132, pp. 1–11.

Lops, P. *et al.* (2019) 'Trends in content-based recommendation', *User Modeling and User-Adapted Interaction*, 29(2), pp. 239–249.

Lu, L. and Zhang, H. (2015) 'A tree-structured representation for book author and its recommendation using multilayer SOM', in *2015 International Joint Conference on Neural Networks (IJCNN)*.

Mihalcea, R. and Tarau, P. (2004) 'Textrank: Bringing order into text', in *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 404–411.

Ng, Y.-K. (2016) 'Recommending books for children based on the collaborative and content-based filtering approaches', in *International Conference on Computational Science and Its Applications*. Springer, pp. 302–317.

Pera, M. S. and Ng, Y. K. (2014) 'How can we help our k-12 teachers?: Using a recommender to make personalized book suggestions', in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE, pp. 335–342.

Priyanka, K., Tewari, A. S. and Barman, A. G. (2015) 'Personalised book recommendation system based on opinion mining technique', in *2015 global conference on communication technologies (GCCT)*. IEEE, pp. 285–289.

Rose, S. *et al.* (2010) 'Automatic keyword extraction from individual documents', *Text mining: applications and theory*, 1, pp. 1–20.

Saxena, S., Singh, B. and Khan, Z. (2014) 'Frequent pattern mining technique for improving book lending recommendation service', *Journal of Basic and Applied Engineering Research*, 1(10), pp. 38–41.

Sohail, S. S., Siddiqui, J. and Ali, R. (2013) 'Book recommendation system using opinion mining technique', in *2013 international conference on advances in computing, communications and informatics (ICACCI)*. IEEE, pp. 1609–1614.

Teng, Y. *et al.* (2015) 'A Novel FAHP Based Book Recommendation Method by Fusing Apriori Rule Mining', in *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. IEEE, pp. 237–243.

Tharwat, A. (2020) 'Classification assessment methods', *Applied Computing and Informatics*.

Thorat, P. B., Goudar, R. M. and Barve, S. (2015) 'Survey on collaborative filtering, content-based filtering and hybrid recommendation system', *International Journal of Computer Applications*, 110(4), pp. 31–36.

Toutanova, K. *et al.* (2003) 'Feature-rich part-of-speech tagging with a cyclic dependency network', in *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 252–259.

Vaz, P. C., Martins de Matos, D. and Martins, B. (2012) 'Stylometric relevance-feedback towards a hybrid book recommendation algorithm', in *Proceedings of the fifth ACM workshop on Research advances in large digital book repositories and complementary media*, pp. 13–16.

Wartena, C., Brussee, R. and Slakhorst, W. (2010) 'Keyword extraction using word co-occurrence', in *2010 Workshops on Database and Expert Systems Applications*. IEEE, pp. 54–58.

Yao, L., Pengzhou, Z. and Chi, Z. (2019) 'Research on news keyword extraction technology based on TF-IDF and TextRank', in *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*. IEEE, pp. 452–455.

Zhang, C. (2008) 'Automatic keyword extraction from documents using conditional random fields', *Journal of Computational Information Systems*, 4(3), pp. 1169– 1180.

ภาคผนวก

ภาคผนวก ก

กฎทางภาษาศาสตร์สำหรับระบบ *SBS*

ตารางที่ 11 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
1	(W_0 = 'NN' or 'NNS' or 'NNP' or 'NNPS'), ..., (W_n = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• W_0, ..., W_n
2	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP' or 'NNPS'), ..., (W_n = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• $W_0 + W_1$, ..., W_n
3	(W_0 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_3 = 'IN') + (W_4 = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• $W_0 + W_1 + W_2$ • $W_2 + W_3 + W_4$
4	(W_0 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_1 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• $W_0 + W_1$ • $W_0 + W_3$
5	(W_0 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_3 = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• $W_0 + W_3$ • $W_2 + W_3$
6	(W_0 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• W_0 • W_2
7	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP' or 'NNPS') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP' or 'NNPS')	if $W_1 = W_4$: • $W_0 + W_1$ • $W_3 + W_4$
8	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$

ตารางที่ 11 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
9	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
10	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$
11	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'JJ' or 'JJR' or 'JJS') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3 + W_4 + W_5$ • $W_2 + W_3 + W_4 + W_5$
12	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$
13	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
14	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_2 + W_3$
15	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_0 + W_3$

ตารางที่ 11 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
16	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
17	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$
18	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_2 + W_3$
19	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'JJ' or 'JJR' or 'JJS') + (W_5 = 'NN' or 'NNS' or 'NNP') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_4 + W_5 + W_6$ • $W_0 + W_1 + W_2 + W_3 + W_4 + W_5 + W_6$
20	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3 + W_4$
21	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS')	• W_0 • W_2
22	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$

ตารางที่ 11 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
23	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_0 + W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
1	(W_0, ..., W_n = 'RW')	• W_0, ..., W_n
2	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$
3	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP'), ..., (W_n = 'NN' or 'NNS' or 'NNP' or 'NNPS')	• $W_0 + W_1$, ..., W_n
4	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$
5	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_2 + W_3$
6	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'RW')	• $W_1 + W_2$ • $W_0 + W_1 + W_2$
7	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_1 + W_2$ • $W_0 + W_1 + W_2$
8	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3$ • $W_0 + W_1 + W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
9	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'RW')	• $W_0 + W_1$
10	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW')	• $W_0 + W_1$
11	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_0 + W_1 + W_2$
12	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW')	• W_3 • $W_0 + W_3$ • $W_3 + W_1$
13	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'IN') + (W_4 = 'RW')	• $W_2 + W_0$ • $W_2 + W_0 + W_1 + W_4$
14	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_3 + W_4$ • $W_0 + W_3 + W_4$ • $W_3 + W_4 + W_1$
15	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_3 + W_4$ • $W_3 + W_4 + W_0$ • $W_3 + W_4 + W_1$
16	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_2 + W_3 + W_4$ • $W_2 + W_3 + W_4 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
17	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_1 + W_2 + W_3$
18	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_2 + W_0$
19	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_3 + W_4 + W_5$ • $W_3 + W_4 + W_5 + W_1$ • $W_0 + W_3 + W_4 + W_5$
20	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • W_3
21	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_0 + W_3$
22	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	if $W_1 = W_4$: • $W_0 + W_1$ • $W_3 + W_4$
23	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	if $W_0 = W_2$: • W_0 • $W_2 + W_3$
24	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_0 + W_3$ • $W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
25	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW')	• W_0 • W_2
26	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3$ • $W_2 + W_3$
27	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW')	• W_0 • W_2
28	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3$
29	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
30	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP')	• W_0 • $W_2 + W_3$
31	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW')	• $W_0 + W_3$ • $W_2 + W_3$
32	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	5. ผลลัพธ์
33	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_0 + W_3$
34	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• W_2 • $W_2 + W_0$ • $W_2 + W_4$
35	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'JJ' or 'JJR' or 'JJS') + (W_6 = 'RW') + (W_7 = 'NN' or 'NNS' or 'NNP')	• $W_5 + W_6 + W_7$ • $W_2 + W_3$ • $W_5 + W_6 + W_7 + W_0$ • $W_2 + W_3 + W_0$
36	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'NN' or 'NNS' or 'NNP') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3$ • $W_5 + W_6$ • $W_2 + W_3 + W_0$ • $W_5 + W_6 + W_0$
37	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'RW')	• $W_2 + W_0$ • $W_4 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	6. ผลลัพธ์
38	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'IN') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_4 + W_5 + W_6$ • $W_3 + W_1 + W_4 + W_5 + W_6$
39	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'RW')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
40	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$
41	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• W_0 • $W_2 + W_3$
42	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'RW')	• W_0 • $W_2 + W_3$
43	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
44	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4$
45	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_2 + W_0$
46	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW')	• $W_0 + W_3$ • $W_2 + W_3$
47	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'IN') + (W_5 = 'JJ' or 'JJR' or 'JJS') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$ • $W_2 + W_3 + W_0 + W_4 + W_5 + W_6$
48	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4$
49	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_2 + W_3 + W_4$ • $W_0 + W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
50	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$ • $W_0 + W_1 + W_2 + W_3 + W_4$
51	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_2 + W_3 + W_0$
52	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'RW') + (W_5 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4 + W_5$
53	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$
54	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_4 + W_5$
55	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_4$ • $W_0 + W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
56	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$
57	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'IN') + (W_5 = 'RW')	• $W_0 + W_4 + W_5$ • $W_2 + W_3 + W_4 + W_5$
58	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'IN') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_0$ • $W_2 + W_0 + W_3 + W_4 + W_5$
59	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'CC') + (W_5 = 'RW') + (W_6 = 'RW')	• $W_2 + W_3 + W_0$ • $W_5 + W_6 + W_0$
60	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4$
61	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
62	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	● $W_2 + W_3 + W_0$
63	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW')	● $W_2 + W_0$
64	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'RW')	● $W_0 + W_1 + W_2$
65	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP')	● $W_0 + W_1 + W_2 + W_3$
66	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	● $W_2 + W_3 + W_4 + W_0$
67	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	● $W_0 + W_1 + W_2$
68	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	● $W_0 + W_1 + W_2$
69	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW')	● $W_0 + W_1 + W_2$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
70	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW') + (W_4 = 'CC') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$ • $W_5 + W_0$
71	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'RW')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
72	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_0 + W_1 + W_2 + W_3$
73	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'RW') + (W_5 = 'IN') + (W_6 = 'RW')	• $W_0 + W_1 + W_5 + W_6$ • $W_3 + W_4 + W_5 + W_6$
74	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW')	• $W_0 + W_1 + W_2 + W_3$
75	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3$ • $W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
76	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3$
77	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_4$ • $W_0 + W_3 + W_4$
78	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'JJ' or 'JJR' or 'JJS') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_0$ • $W_4 + W_5 + W_6 + W_0$
79	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW')	• W_0 • W_2
80	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW')	• W_0 • W_2
81	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_0 + W_1 + W_2 + W_3$
82	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	if $W_0 = W_2$: • W_0 • $W_2 + W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
83	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$ • $W_0 + W_1 + W_2 + W_5$
84	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_1 + W_2 + W_3 + W_4$
85	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'RW')	• $W_2 + W_0$ • $W_4 + W_0$
86	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW')	• $W_3 + W_1$ • $W_0 + W_3$
87	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$
88	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'JJ' or 'JJR' or 'JJS') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_0$ • $W_4 + W_5 + W_6 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
89	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'RW')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
90	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$
91	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'IN') + (W_3 = 'RW')	• $W_0 + W_1 + W_2 + W_3$
92	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$
93	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'RW')	• $W_0 + W_1 + W_2 + W_3 + W_4$
94	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'RW')	• $W_0 + W_1 + W_2 + W_3 + W_4$
95	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
96	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_0 + W_3$
97	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$
98	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
99	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3$
100	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3 + W_4$
101	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
102	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'IN') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_4 + W_5 + W_6$ • $W_3 + W_4 + W_5 + W_6$
103	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'RW')	• $W_0 + W_3$ • $W_2 + W_3$
104	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3$
105	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$
106	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'RW')	• W_0 • $W_2 + W_3$
107	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP')	• W_0 • W_2
108	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
109	(W_0 = 'RW') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'RW') + (W_4 = 'IN') + (W_5 = 'JJ' or 'JJR' or 'JJS')	• $W_1 + W_2 + W_3$ • $W_0 + W_1 + W_2 + W_3 + W_4 + W_5$
110	(W_0 = 'RW') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'RW') + (W_4 = 'RW') + (W_5 = 'IN') + (W_6 = 'RW')	• $W_1 + W_2 + W_3 + W_4$ • $W_0 + W_1 + W_2 + W_3 + W_4 + W_5 + W_6$
111	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_0$ • $W_5 + W_6 + W_0$
112	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP')	• W_0 • $W_2 + W_3$
113	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$
114	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW')	• W_0 • W_2

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
115	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $+ W_4 + W_5$
116	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$
117	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'IN') + (W_5 = 'RW')	• $W_5 + W_3$ • $W_5 + W_1$ • $W_0 + W_5$
118	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_4$ • $W_2 + W_3 + W_4$
119	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
120	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
121	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• W_2 • $W_2 + W_0$ • $W_2 + W_4$
122	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_2 + W_3$
123	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_3 + W_4$ • $W_0 + W_3 + W_4$ • $W_3 + W_4 + W_1$
124	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_0 + W_1 + W_2$
125	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'IN') + (W_4 = 'RW')	• $W_2 + W_0$ • $W_2 + W_0 + W_3 + W_4$
126	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = ''CC) + (W_4 = 'NN' or 'NNS' or 'NNP') + (W_5 = 'NN' or 'NNS' or 'NNP')	if $W_2 = W_5$: • $W_0 + W_1 + W_2$ • $W_4 + W_5$
127	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW')	• $W_0 + W_1$ • W_3

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
128	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_2 + W_3 + W_4$ • $W_2 + W_3 + W_4 + W_0$
129	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'RW')	• $W_0 + W_1 + W_2$
130	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'RW')	• $W_0 + W_1 + W_2 + W_3$
131	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4$
132	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'JJ' or 'JJR' or 'JJS') + (W_6 = 'RW') + (W_7 = 'NN' or 'NNS' or 'NNP')	• $W_5 + W_6 + W_7$ • $W_2 + W_3$ • $W_5 + W_6 + W_7 + W_0$ • $W_2 + W_3 + W_0$
133	(W_0 = 'RW') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'RW') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3$ • $W_5 + W_6$ • $W_2 + W_3 + W_0$ • $W_5 + W_6 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
134	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'CC') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_0 + W_3$
135	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'RW')	• $W_0 + W_1 + W_2$
136	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = ''CC) + (W_4 = 'NN' or 'NNS' or 'NNP')	if $W_2 = W_4$: • $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
137	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1$ • $W_3 + W_4 + W_5$
138	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP') + (W_5 = 'CC') + (W_6 = 'RW')	• $W_0 + W_1$ • $W_3 + W_4$ • $W_0 + W_1 + W_6$ • $W_3 + W_4 + W_6$
139	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'CC') + (W_5 = 'RW')	• $W_2 + W_3 + W_0$ • $W_4 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
140	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP') + (W_5 = 'IN') + (W_6 = 'RW') + (W_7 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_5 + W_6 + W_7$ • $W_3 + W_4 + W_5 + W_6 + W_7$
141	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'RW')	• $W_0 + W_4$ • $W_2 + W_3 + W_4$
142	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'RW')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
143	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP') + (W_6 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_4 + W_5 + W_6$
144	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'CC') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'IN') + (W_4 = 'JJ' or 'JJR' or 'JJS') + (W_5 = 'RW')	• $W_0 + W_3 + W_4 + W_5$ • $W_2 + W_3 + W_4 + W_5$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
145	(W_0 = ‘JJ’ or ‘JJR’ or ‘JJS’) + (W_1 = ‘RW’)	• $W_0 + W_1$
146	(W_0 = ‘RW’) + (W_1 = ‘RW’) + (W_2 = ‘NN’ or ‘NNS’ or ‘NNP’) + (W_3 = ‘CC’) + (W_4 = ‘NN’ or ‘NNS’ or ‘NNP’) + (W_5 = ‘NN’ or ‘NNS’ or ‘NNP’)	• $W_0 + W_1 + W_2$ • $W_4 + W_5$
147	(W_0 = ‘JJ’ or ‘JJR’ or ‘JJS’) + (W_1 = ‘RW’) + (W_2 = ‘RW’) + (W_3 = ‘CC’) + (W_4 = ‘RW’)	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
148	(W_0 = ‘NN’ or ‘NNS’ or ‘NNP’) + (W_1 = ‘IN’) + (W_2 = ‘NN’ or ‘NNS’ or ‘NNP’) + (W_3 = ‘RW’)	• $W_2 + W_3 + W_0$
149	(W_0 = ‘RW’) + (W_1 = ‘RW’) + (W_2 = ‘CC’) + (W_3 = ‘RW’) + (W_4 = ‘RW’)	• $W_0 + W_1$ • $W_3 + W_4$
150	(W_0 = ‘RW’) + (W_1 = ‘RW’) + (W_2 = ‘RW’)	• $W_0 + W_1 + W_2$
151	(W_0 = ‘JJ’ or ‘JJR’ or ‘JJS’) + (W_1 = ‘IN’) + (W_2 = ‘RW’) + (W_3 = ‘RW’)	• $W_0 + W_2 + W_3$
152	(W_0 = ‘NN’ or ‘NNS’ or ‘NNP’) + (W_1 = ‘IN’) + (W_2 = ‘JJ’ or ‘JJR’ or ‘JJS’) + (W_3 = ‘RW’)	• $W_2 + W_3 + W_0$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
153	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'RW')	• $W_0 + W_1 + W_2$
154	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2 + W_3$
155	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_2 + W_0$
156	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_2 + W_3$ • $W_0 + W_1 + W_2 + W_3$
157	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_4 + W_0$
158	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
159	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'RW')	• $W_0 + W_1$ • $W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
160	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW') + (W_4 = 'RW')	• $W_2 + W_3 + W_4 + W_0$
161	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'IN') + (W_2 = 'RW')	• $W_0 + W_2$
162	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'JJ' or 'JJR' or 'JJS') + (W_3 = 'RW')	• W_0 • $W_2 + W_3$
163	(W_0 = 'RW') + (W_1 = 'CC') + (W_2 = 'RW') + (W_3 = 'NN' or 'NNS' or 'NNP') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_3 + W_4$ • $W_2 + W_3 + W_4$
164	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'IN') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_3 + W_4 + W_1$ • $W_0 + W_3 + W_4$
165	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'JJ' or 'JJR' or 'JJS') + (W_2 = 'RW') + (W_3 = 'RW') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_2 + W_3 + W_4$ • $W_0 + W_1 + W_2 + W_3 + W_4$
166	(W_0 = 'JJ' or 'JJR' or 'JJS') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_2 + W_3$ • $W_0 + W_1 + W_2 + W_3$

ตารางที่ 12 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะสำหรับระบบ SBS (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
167	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'RW') + (W_5 = 'RW')	• $W_0 + W_1 + W_2$ • $W_4 + W_5$
168	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'NN' or 'NNS' or 'NNP') + (W_3 = 'CC') + (W_4 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_1 + W_4$
169	(W_0 = 'RW') + (W_1 = 'RW') + (W_2 = 'RW') + (W_3 = 'CC') + (W_4 = 'RW') + (W_5 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_2$ • $W_0 + W_4 + W_5$
170	(W_0 = 'RW') + (W_1 = 'NN' or 'NNS' or 'NNP') + (W_2 = 'CC') + (W_3 = 'JJ' or 'JJR' or 'JJS') + (W_4 = 'NN' or 'NNS' or 'NNP') + (W_5 = 'IN') + (W_6 = 'RW') + (W_7 = 'NN' or 'NNS' or 'NNP')	• $W_0 + W_1 + W_5 + W_6 + W_7$ • $W_3 + W_4 + W_5 + W_6 + W_7$
171	(W_0 = 'NN' or 'NNS' or 'NNP') + (W_1 = 'IN') + (W_2 = 'RW') + (W_3 = 'RW')	• $W_2 + W_3 + W_0$

ภาคผนวก ข

กฎทางภาษาศาสตร์สำหรับระบบ *SBS+*

ตารางที่ 13 กฎทางภาษาศาสตร์สำหรับระบบ SBS+

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
1	$JJ_1, ..., JJ_p$	● $JJ_i, ..., JJ_k \subseteq JJ_1, ..., JJ_p$ When $JJ_u \in JJ_i, ..., JJ_k \cap JJ_u \notin$ Adjective corpus
2	$JJ_1, ..., JJ_p$ + $NN_1, ..., NN_q$	● $JJ_i, ..., JJ_k \subseteq JJ_1, ..., JJ_p + NN_1, ..., NN_q$ When $JJ_u \in JJ_i, ..., JJ_k \cap JJ_u \notin$ Adjective corpus
3	$JJ_1, ..., JJ_p$ + $TE_1, ..., TE_q$	● $JJ_i, ..., JJ_k \subseteq JJ_1, ..., JJ_p + TE_1, ..., TE_q$ When $JJ_u \in JJ_i, ..., JJ_k \cap JJ_u \notin$ Adjective corpus
4	$NN_1, ..., NN_p$	● $NN_1, ..., NN_p$
5	$NN_1, ..., NN_p$ + $TE_1, ..., TE_q$	● $NN_1, ..., NN_p + TE_1, ..., TE_q$
6	$TE_1, ..., TE_p$	● $TE_1, ..., TE_p$
7	$TE_1, ..., TE_p$ + $NN_1, ..., NN_q$	● $TE_1, ..., TE_p + NN_1, ..., NN_q$
8	$JJ_1, ..., JJ_p$ + $NN_1, ..., NN_q$ + $TE_1, ..., TE_r$	● $JJ_i, ..., JJ_k \subseteq JJ_1, ..., JJ_p + NN_1, ..., NN_q +$ $TE_1, ..., TE_r$ When $JJ_u \in JJ_i, ..., JJ_k \cap JJ_u \notin$ Adjective corpus
9	$JJ_1, ..., JJ_p$ + $TE_1, ..., TE_q$ + $NN_1, ..., NN_r$	● $JJ_i, ..., JJ_k \subseteq JJ_1, ..., JJ_p + TE_1, ..., TE_q +$ $NN_1, ..., NN_r$ When $JJ_u \in JJ_i, ..., JJ_k \cap JJ_u \notin$ Adjective corpus
10	$NN_1, ..., NN_p$ + $TE_1, ..., TE_q$ + $NN_1, ..., NN_r$	● $NN_1, ..., NN_p + TE_1, ..., TE_q + NN_1, ..., NN_r$
11	$TE_1, ..., TE_p$ + $NN_1, ..., NN_q$ + $TE_1, ..., TE_r$	● $TE_1, ..., TE_p + NN_1, ..., NN_q + TE_1, ..., TE_r$

ตารางที่ 13 กฎทางภาษาศาสตร์สำหรับระบบ SBS+ (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
12	($Phrase_1$) + IN + ($Phrase_2$)	• $Phrase_2$ + $Phrase_1$
13	($Phrase_1$) + IN + ($Phrase_2$) + IN + ($Phrase_3$)	• $Phrase_3$ + $Phrase_2$ + $Phrase_1$
14	($Phrase_1$) + CC + ($Phrase_2$)	• $Phrase_1$ • $Phrase_2$
15	($Phrase_1$) + CC + ($Phrase_2$) *if ($Phrase_2$) has one word	• $Phrase_1$ • $Phrase_1$(without last word) + $Phrase_2$
16	($Phrase_1$) + CC + ($Phrase_2$) * if ($Phrase_1$) has one word	• $Phrase_1$ + $Phrase_2$(without first word) • $Phrase_2$
17	($Phrase_1$) + CC + ($Phrase_2$) * if last word of ($Phrase_1$) = last word of ($Phrase_2$)	• $Phrase_1$ • $Phrase_2$
18	($Phrase_1$) + CC + ($Phrase_2$) * if first word of ($Phrase_1$) = first word of ($Phrase_2$)	• $Phrase_1$ • $Phrase_2$

ตารางที่ 13 กฎทางภาษาศาสตร์สำหรับระบบ SBS+ (ต่อ)

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
19	($Phrase_1$) + CC + ($Phrase_2$) * if ($Phrase_1$) has one word and ($Phrase_1$) = first word of ($Phrase_2$)	• $Phrase_2$
20	($Phrase_1$) + IN + ($Phrase_2$) + CC + ($Phrase_3$)	• $Phrase_2$ + $Phrase_1$ • $Phrase_3$ + $Phrase_1$
21	($Phrase_1$) + CC + ($Phrase_2$) + IN + ($Phrase_3$)	• $Phrase_3$ + $Phrase_1$ • $Phrase_3$ + $Phrase_2$

ภาคผนวก ค

เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์

สำนักงานบริหารกิจการ
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา
เลขที่ 813
๒๗ ส.ค. ๒๕๖๑
วันที่
เวลา 13.30

1384
๒๔ ส.ค. ๒๕๖๑
15.30

บันทึกข้อความ

ส่วนงาน สำนักงานอธิการบดี กองบริหารการวิจัยและนวัตกรรม โทร. ๒๕๖๑ - ๒๕๖๒
ที่ ศธ ๖๒๐๐/ ๐๙๙๒๒ **วันที่** ๒๓ สิงหาคม พ.ศ. ๒๕๖๑
เรื่อง ขอแจ้งรับรองโครงการวิจัยที่ส่งมาขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา

เรียน นางสาวเบญจมวรรณ ใจสูงเนิน (นิสิตระดับบัณฑิตศึกษา คณะวิทยาการสารสนเทศ)

ตามที่ท่านได้ส่งเอกสารโครงการวิจัยเพื่อขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ ในหัวข้อโครงการวิจัยเรื่อง ระบบแนะนำหนังสืออ่านเพิ่มเติม นั้น

บัดนี้ คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาตามวิธีดำเนินการมาตรฐาน (Standard Operating Procedures, SOP) ฉบับที่ ๑ พ.ศ. ๒๕๖๐ ที่ได้ประกาศใช้เมื่อวันที่ ๙ มกราคม พ.ศ. ๒๕๖๐ แล้วว่า โครงการวิจัยดังกล่าวไม่ได้ทำการศึกษาวิจัยในมนุษย์ จึงเห็นสมควรให้ดำเนินการวิจัยได้

จึงเรียนมาเพื่อโปรดทราบ

(ผู้ช่วยศาสตราจารย์ ดร.วิทวัส แจ้งเอี่ยม)
ประธานคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์
มหาวิทยาลัยบูรพา

ภาคผนวก ง
เอกสารเผยแพร่ผลงานวิจัย

2018
ICAICTA
International Conference on Advanced Informatics:
Concepts, Theory and Applications
Conference : 14-17 August, 2018 | Krabi, Thailand
ICAICTA2018
@Krabi, Thailand
August 14 - 17, 2018
Call for Papers
ICAICTA 2018 - International Conference on Advanced Informatics: Concept Theory and Applications
Prospective authors are invited to submit full papers of no more than six (6) pages including results, according to template provided below. Papers will be subject to double blind peer review in which the reviewers do not know the author's identity. In order to make blind reviewing possible, authors must omit their names and affiliations from the paper.
Conference website : http://services.informatics.buu.ac.th/icaicta2018/
Scope of Conference
Computational Science and Engineering
High Performance Computing
Distributed Systems and Pervasive Computing
Numerical Methods and Simulations
Game Theory and Optimizations
Cryptography and Data Hiding
Media Technology and Information Systems
Computer Vision, Graphics, and Image Processing
Natural Language and Speech Processing
Human Computer Interactions
Multimedia Services and Applications
Audit, Security and Governance
Intelligent Robotics
Intelligent Systems and Data Sciences
Information Retrieval and Data Mining
Database and Programming
Artificial Intelligence and Machine Learning
Expert and Recommendation Systems
Learning Systems and Knowledge Modeling
Computation for Science
Computational Chemistry
Computational Physics
Computational Biology
Computational Social Science
Dec 1, 2017
CALL FOR PAPERS
May 20, 2018 (EXTENDED)
SUBMISSION DEADLINE
Jul 1, 2018
NOTIFICATION OF ACCEPTANCE
Jul 15, 2018
EARLY-BIRD REGISTRATION
Jul 15, 2018
CAMERA-READY SUBMISSION
Aug 1, 2018
REGULAR REGISTRATION
Aug 14 -17, 2018
CONFERENCE DAY
TOYOHASHI
UNIVERSITI SAINS MALAYSIA
IEEE
IEEE Xplore Digital Library
IEEE THAILAND SECTION

Supplementary book suggestion for computer science courses

Benchamawan Chaisoongnoen[§], Komate Amphawan[†]
Computational Innovation Laboratory,
Faculty of Informatics, Burapha University
Chonburi, 20131, Thailand
Email: benchamawan.gxii@gmail.com[§], komate@gmail.com[†]

Aekapop Bunpeng
Faculty of Informatics, Burapha University
Chonburi, 20131, Thailand
Email: aekapop@buu.ac.th

***Abstract*—On-line selling website is currently famous and popular. There are several websites selling products and/or services including on-line book-selling websites. At present, the current book-selling websites usually apply recommender systems to recommend a book or a set of books to customers. However, the recommender systems mostly focus on recommending books that users usually view or buy together and also on books having high review rates. This may cause failure to recommend books that cover most required contents, for example, books related to a course description of a course students have registered. To address on this issue, we here introduce an alternative recommender system called *Supplementary Books Suggestion system* (*SBS system*) to create a list of supplementary books related/relevance to a course description of a course in computer science domain by regarding relevance between a book and a course description. This can help students easily find supplementary books to read and also may help to encourage the students doing self-learning. Experiments on real course descriptions were conducted to investigate the effectiveness of the *SBS system* in the terms of precision, recall, F-measure and average (also total) coverage/uncoverage of contents between a list of supplementary books and a course description.**

***Keywords*—Book recommender system, Keyword extraction, Supplementary books, Course description;**

I. Introduction

Currently, it is the era of 4.0 based on the growth of Internet and technology. People use more and more Internet in daily life in order to assists their living and to serve their needs such as finding goods/products, services, travel, places, friends and so on. Meanwhile, in business aspect, it is the era that focuses on digital transformation and innovation. Many businesses starts to improve their product or service by applying AI and/or IoT. Moreover, to improve marketing, understanding individual (personalize) needs is really important. Based on this, recommendation systems have become essential tools to assists users to find what they want and it is currently applied in a wide range of businesses such as music, movie, clothes, and so on.

In addition, recommendation systems are also widely applied in on-line book-selling industry. Amazon, O'Reilly, Goodreads, Barns & Noble are well-known websites applying recommendation systems to suggest books to customers based on relevance, clicks, buying behaviors, reviews and so on. By this, several techniques are proposed to improve recommendation performance, for example, (*i*) Book recommendation system based on combine features of content based filtering, collaborative filtering and association mining [1], (*ii*) Book recommendation system based on considering both table of contents (ToC) and association rule mining [2], (*iii*)Book Recommender System Using Fuzzy Linguistic Quantifiers [3] and so on.

Besides, there are several approaches on book recommendation using different criteria such as (*i*) Recommendation model for personalized book lists [4] aiming to recommend books to groups of people having common interests and then evaluate the accuracy satisfaction of the model by using questionnaires, (*ii*) Book recommendation system for digital library [5] trying to identify relationship between books that users interest through their loaning histories, (*iii*) Recommendation system using hybrid filtering (collaborative and content-based filtering) incorporate with best price identification [6], [7] proposing to recommend books that meet users need and to inform about on-line book-selling website with lowest price, etc. However, previous book recommendation systems (mentioned above) face with several issues to provide relevant book to user needs since most of them do not take into account book contents at deep level and regard only the mere information of books provided on web pages along with meta-data and rating. This may cause failure to recommend books that cover most required contents, for example, books related to a course description of a course students have registered.

To cope with this issue, we here introduce an alternative approach on book recommender system called *Supplementary Books Suggestion system* (also called *SBS system* for short) in order to create a list of supplementary books related/relevance to a course description of a course in computer science domain by regarding relevance between a book and a course description. This list of supplementary books can help students easily find supplementary books (or should read books) to read and also may help to encourage the students doing self-learning. Experiments on real course descriptions were conducted to investigate effectiveness of the proposed *SBS system* in the terms of precision, recall, F-measure and average (also total) coverage/uncoverage of contents between a list of supplementary books relevant to a course description.

978-1-5386-4804-9/18/$31.00 @2018 IEEE

II. Supplementary book suggestion system

In this section, we here describe details of our proposed supplementary book suggestion system (also called *SBS system* in short). The system aims to assist students to easily see a list of "should-read books" related to a computer science course they have registered. Moreover, this system can help to provide add-on information and ability to perform self-learning and also help to extend body of knowledge of students. The *SBS system* consists of five computational steps (as shown in Fig. 1) : 1) Terminology gathering, 2) Course descriptions input, 3) Keyword extraction, 4) Relevance books searching 5) Coverage calculation and ranking, respectively. Each of the above steps can be described in details as follows.

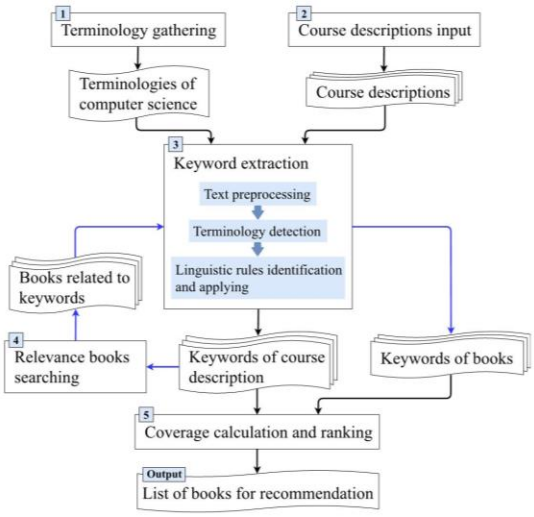


Fig. 1: Framework of the supplementary book suggestion system

A. Terminology gathering

In computer science domain, there is a large amount of reserved words and terminologies with specific meaning, for example, "natural language processing", "data mining", "data structure", "data science", "big data" and so on. Most of these are well-known in the domain which can help to identify specific objects or actions. To study in this domain, students and/or anyone interested should have a background or know about meaning of reserved words and terminologies since course titles, course descriptions, book title and table of contents of books contain numerous of these terminologies with extension words, for instance, "data science concepts" composing from 1) "data science"–a terminology and 2) "concepts"–an extension word, respectively. In addition, identifying terminologies in the course titles, course descriptions, book title and table of contents of books can help to know about important topics to study in each course. Thus, to automatically extraction of the topics to study, terminologies in computer science domain should be prior gathered.

Based on our *SBS system*, terminologies from the lists of computer science terms of *Oxford university*[1] and *Labautopedia*[2] are gathered into our corpus. Last, by regarding and removing redundant terminologies from both sites, our corpus contains $6,527$ terminologies.

B. Course descriptions input

The second step of *SBS system* is to collect course descriptions in order to know main topics related to each course. To do that, course descriptions in the Computer Science curriculum of Faculty of Informatics, Burapha University [8] are considered and used as an example input. Then, the title and details of course description (in English) of a course are collected into database (as highlighted in blue in Fig. 2).

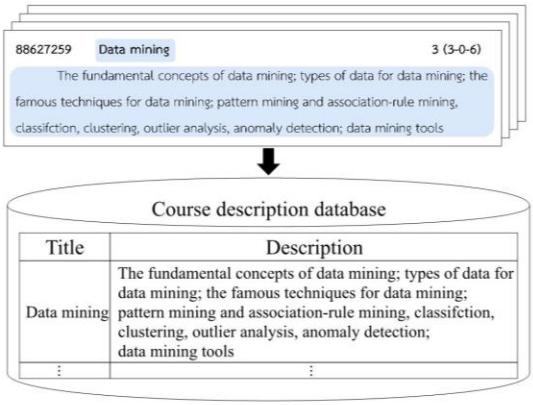


Fig. 2: Example of course descriptions input

C. Keywords extraction

Whenever a course description is input (as shown in Fig. 2), the process of keyword extraction is performed to determine important objects or actions hidden in the course description. This task consists of three consecutive steps as follows (see Fig. 3):

1) *Text preprocessing*–performs on (*i*) *Stop word removal*–removes stop words except prepositions and conjunctions since we have to keep these, for example, the word "and" is a conjunction which can help to split the noun phrase "unstructured and structured data" to be "unstructured data" and "structured data", (*ii*) *Word error correction*–corrects words from typos or input errors by spelling mistake correction (SMC) method [9], (*iii*) *Sentence tokenization*–splits sentences or phrases in the course description to be individual sentences, and (*iv*) *Part of speech tagging*–identifies type of words by using Standford corpus [10] (Noted "JJ", "JJR", "JJS" refers to Adjective, "NN", "NNS", "NNP", "NNPS" refers to Noun, "IN" refers to preposition and "CC" refers to conjunction, respectively).

[1] http://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975

[2] http://www.labautopedia.org/mw/List_of_programming_and_computer_science_terms

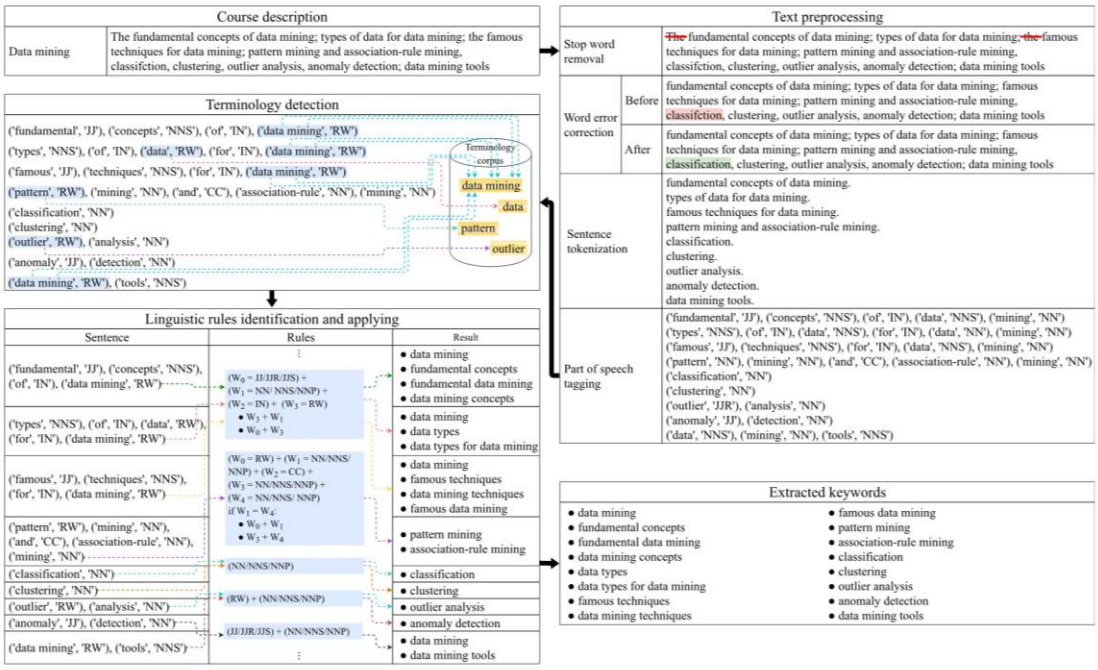


Fig. 3: Example of keyword extraction from an input course description

2) *Terminology detection*–finds terminologies in the course description by comparing n-grams of words in each sentence of the course description and terminologies gathered from the first step. For example, the first sentence of the course description contains ('fundamental', 'JJ'), ('concepts', 'NNS'), ('of', 'IN'), ('data', 'NNS'), ('mining','NN'). From the sentence, the words 'data' and 'mining' are grouped together to be 'data mining' by the concept of 2-grams. Then, it is compared with terminologies contained in the terminology corpus and then identified as 'reserved word', *i.e.* 'RW' (see Fig. 3).

3) *Linguistic rules identification and applying*–creates linguistic rules by applying the concept of domain-specific term extraction of [11], [12] and then applies these rules to identify keywords hidden in each sentence of the course description. To create rules, there are two cases to be considered: (*i*) rules with terminology and (*ii*) rules without terminology, respectively. For the first type of rules, each terminology in the considered sentence is firstly regarded. It is thus merged with its surrounding noun and adjective and then identified as keywords. On the other hand, rules without terminology try to identify noun phrases being in the form of combination between nouns, adjective and/or preposition, respectively. For example, there is a created rule sequentially containing "adjective (JJ) + noun (NN or NNS or NNP) + preposition (IN) + terminology (RW)". It can converse the above sequence of words to be keywords of (*i*) "terminology (RW) + noun (NN or NNS or NNP)" and (*ii*) "adjective (JJ) + terminology (RW)", respectively. For the first sentence from the previous step *i.e.* "fundamental (JJ) + concepts (NNS) + of (IN) + data mining (RW)", it can be extracted to identify keywords by the above rule as (*i*) "data mining concepts" (RW + NNS) and (*ii*) "fundamental data mining" (JJ + RW), respectively. Noted that all of the rules we have created and used in the *SBS system* can be viewed and downloaded from http://staff.informatics.buu.ac.th/~komate/SBSsystem.

D. Relevance books searching

After extracting keywords $K_{c_i} = \{k_{c_i,1}, k_{c_i,2}, \ldots, k_{c_i,n}\}$ from a course description c_i, the searching relevance books is performed. Each keyword $k_{c_i,x} \in K_{c_i}$ is thus considered and used to search for related books from https://www.barnesandnoble.com (*i.e.* on-line book selling website focusing on academic books). By this, a list of books in which the important information each books *e.g.* title, descriptions, table of contents (TOC) and other essential information (see Fig. 4) is collected into our database and set to be relevant to the course description. For example, as in Fig. 4, the keyword "data mining" is considered and used to searching for related books. From the figure, the related book is "Introduction to Data mining" is collected with *title* = *"Introduction to Data Mining"*, *author* = *"Ping-Ning Tan, Michael Steinbacj, Anuj Kapatne, Vipin Kumar"*,

price = "108.50", isbn13 = "9780133128901", publisher = "Pearson", publicationdate = "01/18/2018" edition = "2", bookdescription = "Introducing the fundamental concepts and algorithms of data mining. Introduction to Data Mining, 2nd Edition , gives a comprehensive overview of the background and general themes of data mining and is designed to be useful to students, instructors, researchers, and professionals." etc. and *toc = "Introduction. Data. Classification: Basic Concepts and Techniques. Classification: Alternatives Techniques. Association Analysis: Basic Concepts and Algorithms. Association Analysis: Advanced Concepts. Cluster Analysis: Basic Concepts and Algorithms. Cluster Analysis: Additional Issues and Algorithms. Anomaly Detection. Avoiding False Discoveries"*, respectively.

From these information, *bookdescription* and *toc* are then considered and transferred to *Keyword extraction* step (Section II-C) in order to get keywords. Then, the keywords from *bookdescription* and *toc* are merged together and collected as the set of keywords of the related book. For example, the first sentence from *book-description* is "Introducing the fundamental concepts and algorithms of data mining". It then transferred to *Keyword extraction* where we can get "fundamental data mining", "data mining concepts", and "data mining algorithm" as keywords for the book, respectively.

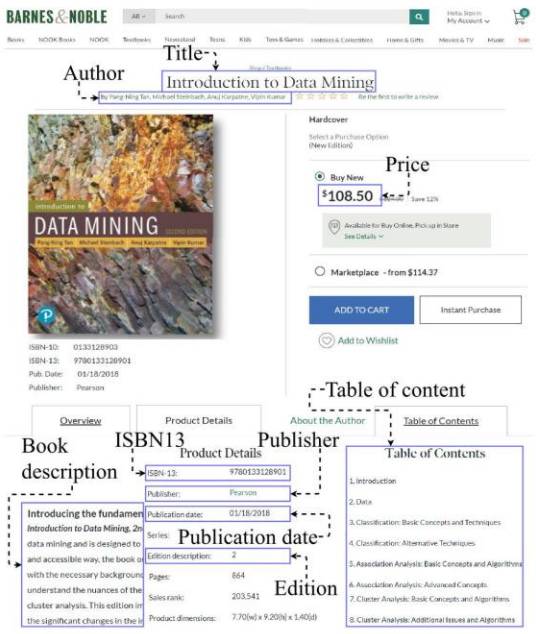


Fig. 4: Example of collecting of a book related to the keyword "Data mining" from the course description of Data mining course

E. Coverage calculation and ranking

The last step of the *SBS system* is the calculation of coverage (similarities) between the keywords extracted from a course description and the keywords from each of its related books and then ranking related books based on their coverages. To calculate similarity between a course description c_i with keywords $K_{c_i} = \{k_{c_i,1}, k_{c_i,2}, \ldots, k_{c_i,n}\}$ and a book b_j related to c_i with keywords $K_{b_j} = \{k_{b_j,1}, k_{b_j,2}, \ldots, k_{b_j,m}\}$, the percentage of *coverage* and *uncoverage* [13] between the book b_j and the course description c_i is applied as follows:

$$C(c_i, b_j) = \frac{|K_{c_i} \cap K_{b_j}|}{|K_{c_i}|}$$

$$U(c_i, b_j) = 1 - C(c_i, b_j)$$

Therefore, as described in Section II-D, each course description c_i has a list of related books $B_{c_i} = \{b_j, \ldots, b_k\}$ in which each book b_j associates with the set of keywords $K_{b_j} = \{k_{b_j,1}, k_{b_j,2}, \ldots, k_{b_j,m}\}$. Therefore, by using the two formulas as above, there will be a list of coverage values $C_{c_i} = \{C(c_i, b_j), \ldots, C(c_i, b_k)\}$ and a list of uncoverage values $U_{c_i} = \{U(c_i, b_j), \ldots, U(c_i, b_k)\}$ expressing similarities and dissimilarities between the course description c_i and the list of related books B_{c_i}, respectively. Thus, the list of recommended books can be generated by ranking coverage values in descending order and filtered out only the top-*N* books with highest coverage (Noted that this paper specifies the value of N to be 10, 30, 50 and 100, respectively). On the other hand, the list of recommended booked can also be created in the same manner as above by using uncoverage value. Thus can help to gain a list of books having details (keywords) extending from details (keywords) in the course description.

III. Experimental Study

In this section, we here describe experimental studies conducted to evaluate our proposed *SBS system*. Seventy-two course descriptions from the Computer Science curriculum of Faculty of Informatics, Burapha University [8] are considered and collected. To create a list of supplementary books related to a course description, the number of recommended books is set to be 10, 30, 50 and 100, respectively. Noted that these numbers are usually used for top-*N* recommendations [14], [15] and mining top-*k* (*N*-most) interesting patterns [16], [17].

As detailed in Section II, we can see that keyword extraction is the major computation of the *SBS system* affecting to its effectiveness. Thus, we then also observe the performance of the keyword extraction used in the *SBS system* and make a comparative study with *Termine* [11] and *RAKE* [18] (two famous keyword extraction techniques). Three popular metrics, recall, precision and F-measure, are applied to observe accuracy of the three keyword extraction techniques in which each metric can be defined as follows.

$$precision = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ items\ retrieved}$$

$$recall = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ relevant\ items}$$

$$F\text{-}measure = 2 \times \frac{precision^{*}recall}{precision + recall}$$

As shown in Fig. 5, the performance of our keyword extraction in comparison with *Termine* and *RAKE* are shown. Anyhow, to evaluate whether an extracted keyword is a correct keyword or not, we have labeled keywords on course descriptions based on the help from expertises (lecturers on the courses). Based on this, the number of extracted keywords, the number of correct extracted keywords, recall, precision, and F-measure of the three techniques can be calculated. From the figure, we can see that our technique outperforms both *Termine* and *RAKE* techniques in all metrics. Based on *Termine*, noun and preposition phrases are first considered to create a list of candidate keywords and then sent the list to expertises in order to identify keywords. However, *Termine* does not take into account domain terminologies and other linguistic rules (such as combination of words, combination of words with a terminology) which can help to improve its performance. Meanwhile, *RAKE* mainly focuses on considering "word co-occurrence" by applying the n-grams technique to identify keywords. It then regards on frequency of occurrence of co-occurrence words, ranks co-occurrence words by descending order of their frequency and selects one-third highest frequency co-occurrence words to be candidate keywords. Last, the keywords are thus identified from the list of candidates by expertises. However, for our keyword extraction technique, it applies several linguistic rules including noun phrases, adjective phrases, preposition phrases, term phrases (*i.e.* terminology + noun, terminology + adj, and terminology + preposition) and coordination of phrases (*i.e.* generating a keyword based on combinations between nouns, adjective and/or preposition occurring surrounding of the conjunction "*and*" or "*or*"). These can help to increase accuracy of keyword extraction. Hence, the precision of our technique is 91.3% and the recall is 95.3% causing the F-measure is quite high as 93.1%. Thanks to the terminologies prior collected in the database and the created rules that mostly cover all the case of keywords in the course descriptions.

Approach	Extracted keywords	Correct keywords	Precision	Recall	F-measure
Our keyword extraction	1447	1253	**91.3**	**95.3**	**93.1**
Termine	1365	523	91.1	40.6	55.1
RAKE	1695	837	70.1	62.6	66.0

Fig. 5: Evaluation of keyword extraction used in *SBS system*

In addition, to investigate quality of recommendation results, the average and total coverage (also for uncoverage) of all supplementary books from the *SBS system* are calculated and scrutinized. The average coverage (also for uncoverage) is calculated by averaging coverage (uncoverage) values of all supplementary books related to course descriptions. Meanwhile, the total coverage (uncoverage) is computed by considering a list of supplementary books related to a course description. Then, all of matched keywords (unmatched keywords for uncoverage) between keywords of the supplementary books and keywords of the course description is accumulated. Last, the number of matched keywords (unmatched keywords) of all course description is counted and averaged. Furthermore, we thus also observe and compare the coverage and uncoverage of our *SBS system* (in the terms of average and total) with that of the *SBS system* using *Termine* or *RAKE* in stead of using our keyword extraction technique (denoted these systems as *SBS system*^{T} and *SBS system*^{R}).

Figure 6 shows the average and total coverage of the three systems. From the figure, it can be observed that as the number of recommended books increases, the average coverage decreases. Obviously, with higher number of required recommended books, there are more chance to get less relevance books related to a course description (Noted that all of recommended books are ranked by descending order of their courage and then selected by the top-*N* rank). Conversely, the total coverage of the three systems increases as the number of recommended books increases. It is because higher number of recommended books leads to have more topics related to course descriptions. Last, it can be seen that our *SBS system* can give higher average and total coverage than the other two systems (Thanks to our keyword extraction technique).

The average and total uncoverage of the three systems are also observed and illustrated in Fig. 7 to see expansion of topics extended from course descriptions. By this, a (another) list of books related to each course description having several additional contents is created. This list can help students to get/know more topics related to the course to learn about. The average and total uncoverages are calculated in the contrary of the average and total coverage. By this, all books in the list of recommended books related to a course description are ranked by descending order of uncoverage value (the coverage value is not used on this basis). From the figure, it can be seen that the average uncoverage of the three systems decreases as the number of required recommended books increases. As mentioned above, all of recommended books are ranked by descending order of uncoverage value. Thus, higher number of books returned causes lower value of average coverage. Moreover, the total uncoverage of the three systems also decreases as the number of required recommended books increases. Higher number of books suggested makes the systems take more and more books with low uncoverage to be results. Last, we can also observe that our *SBS system* gives lower average and total coverage than the other two systems. Based on *SBS system*, it usually extracts correct keywords and then retrieves relevance books based on the correct keywords. This leads the *SBS system* to generates low uncoverage lists of recommended books in comparison with the other two systems.

IV. Conclusion

In this paper, we have introduced a new book recommender system called *Supplementary Book Suggestion system* (*SBS system*). The system can suggest a list of should read books related to a course description of a course in computer science domain. This can help students easily find supplementary books to read and also may help to encourage the students doing self-learning. Based on a course, its course description is regarded to extract important keywords in order to understand/know about objects and actions concerned to the course. These keywords is thus used to match between the course and its relevant books. Experiments on real courses were done to investigate precision, recall and F-measure of our keyword extraction technique in a comparison to *RAKE* and *Termine* since performance of keyword extraction affect

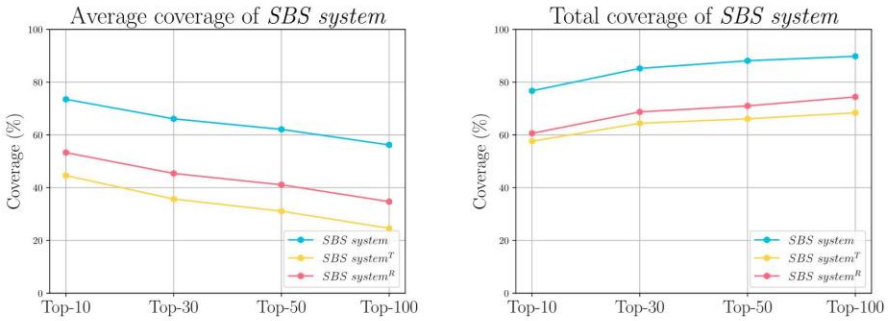


Fig. 6: Average and total coverage of *SBS system*, *SBS system*^{T} and *SBS system*^{R}

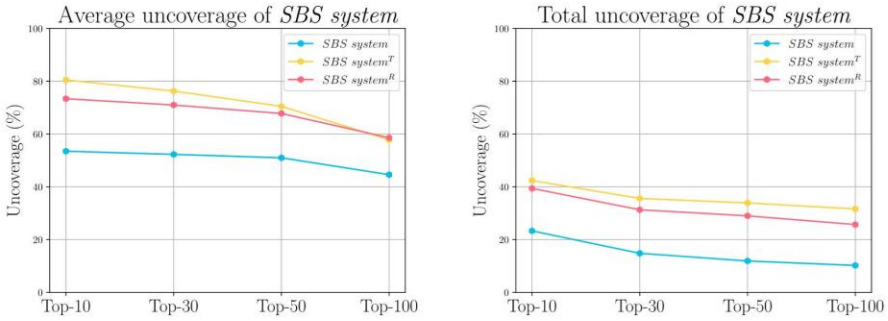


Fig. 7: Average and total uncoverage of *SBS system*, *SBS system*^{T} and *SBS system*^{R}

to performance of *SBS system*. In addition, the average and total coverages (also for uncoverage) of the list of books to the course description are also observed. From all of experiment, it can be seen that our keyword extraction technique and *SBS system* can effectively suggest supplementary books related to a course.

Acknowledgement

This work was financially supported by a research grant of Burapha University through the National Research Council of Thailand (Grant No. 16/2561). In addition, the authors would like to give a special thanks to all lecturers of Faculty of Informatics, Burapha University on help and supports.

References

[1] A. S. Tewari, A. Kumar, and A. G. Barman, "Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining," in *2014 IEEE International Advance Computing Conference (IACC)*, 2014, pp. 500–503.

[2] Z. Ali, S. Khusro, and I. Ullah, "A hybrid book recommender system based on table of contents (toc) and association rule mining," in *Proceedings of the 10th International Conference on Informatics and Systems*, 2016, pp. 68–74.

[3] S. S. Sohail, J. Siddiqui, and R. Ali, *Book Recommender System Using Fuzzy Linguistic Quantifiers*. Singapore: Springer Singapore, 2017, pp. 47–60.

[4] S. Maneewongvatana and S. Maneewongvatana, "A recommendation model for personalized book lists," in *2010 10th International Symposium on Communications and Information Technologies*, 2010, pp. 389–394.

[5] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in *Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014)*, 2014, pp. 130–134.

[6] A. S. Tewari and K. Priyanka, "Book recommendation system based on collaborative filtering and association rule mining for college students," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 2014, pp. 135–138.

[7] A. Goel, D. Khandelwal, J. Mundhra, and R. Tiwari, "Intelligent and integrated book recommendation and best price identifier system using machine learning," in *Intelligent Engineering Informatics*, 2018, pp. 397–412.

[8] B. U. Faculty of Informatics, "Curriculum of computer science," https://www.informatics.buu.ac.th/2016/wp-content/uploads/2014/11/4.6.2.37-BSc-CS-ProgramSpec-cs.pdf, 2016.

[9] S. Sharma and S. Gupta, "A correction model for real-word errors," *Procedia Computer Science*, vol. 70, pp. 99 – 106, 2015.

[10] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, 2003, pp. 173–180.

[11] K. Frantzi, S. Ananiadou, and H. Mima, "Automatic recognition of multi-word terms:. the c-value/nc-value method," *International Journal on Digital Libraries*, vol. 3, no. 2, pp. 115–130, 2000.

[12] K. Balachandran and S. Ranathunga, "Domain-specific term extraction for concept identification in ontology construction," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2016, pp. 34–41.

[13] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: Evaluating recommender systems by coverage and serendipity," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, 2010, pp. 257–260.

[14] H. K. Kim, H. Y. Oh, J. C. Gu, and J. K. Kim, "Commenders: A recommendation procedure for online book communities," *Electronic Commerce Research and Applications*, vol. 10, no. 5, pp. 501–509, 2011.

[15] X. Ning and G. Karypis, "Slim: Sparse linear methods for top-n recommender systems," in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 497–506.

[16] A. Wai-chee Fu, R. Wang-wai Kwong, and J. Tang, "Mining n-most interesting itemsets," in *Foundations of Intelligent Systems*, 2010, pp. 59–67.

[17] K. Amphawan and P. Lenca, "Mining top-k frequent-regular closed patterns," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7882 – 7894, 2015.

[18] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," in *Text Mining: Applications and Theory*, 2010, pp. 1 – 20.

View publication stats

Gmail **Benchamawan Chaisoongnoen <benchamawan.gxii@gmail.com>**

Fwd: ICAICTA-2018 notification for paper 59

1 ข้อความ

Komate Amphawan <komate@gmail.com> 12 กรกฎาคม 2561 17:02
ถึง: เบญจมวรรณ ใจสูงเนิน <benchamawan.gxii@gmail.com>

Komate Amphawan

Computational Innovation Laboratory (CIL)
Faculty of Informatics, Burapha University
Chonburi, 20131, Thailand

---------- Forwarded message ----------
From: **ICAICTA-2018** <icaicta2018@easychair.org>
Date: Thu, Jul 12, 2018 at 4:29 PM
Subject: ICAICTA-2018 notification for paper 59
To: Komate Amphawan <komate@gmail.com>

Dear Komate Amphawan

We are pleased to inform you that your paper with paper id 59 entitled Supplementary books suggestion for computer science courses has been accepted with Minor Correction for “Oral Presentation” at the 2018 – 5th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA-2018) which will be held on August 14-17, 2018 at Beyond Hotels & Resorts, Krabi, Thailand. Therefore, we cordially invite you to attend ICAICTA-2018 for presenting your paper.

Reviewers' comments are listed at the end of this message. It is strongly recommended to the author to revise the paper following reviewer's comments in camera-ready submission.

You MUST strictly adhere to the following requirements:
1. At least one author of each paper must register as Regular no later than the final camera-ready paper submission due date. The registration rates are as follows: http://services.informatics.buu.ac.th/icaicta2018/registration.php
2. Your paper may have been accepted with a condition that the paper must be revised according to the reviewer comments. Failing to do so may exclude your paper from the program.
3. The final camera ready paper is due on July 25, 2018. This is a firm deadline for the production of the proceedings. Please submit your paper using your EasyChair author account via Submission Page. The page limit is 6 page(s) and is strict. You can refer to the Final Manuscript format as follows: http://services.informatics.buu.ac.th/icaicta2018/format
Please note that ICAICTA-2018 conference is technical co-sponsorship by IEEE Thailand Section. Therefore, all accepted and presented papers should have the similarity score lower than 25% to be submitted for inclusion in IEEEXplore® Digital Library. Please also be aware of self-plagiarism.

Best regards
ICAICTA 2018 Organizing Committee

---------- REVIEW 1 ----------
PAPER: 59
TITLE: Supplementary books suggestion for computer science courses
AUTHORS: Benchamawan Chaisoongnoen, Komate Amphawan and Aekapop Bunpeng

Relevance to the conference: 4 (Very good)
Novelty and originality: 2 (Limited / Minor variations on a well investigated subject.)
Technical content and scientific rigour: 3 (Good / Valid work but limited contribution.)
Quality of presentation: 3 (Good / Readable, but revision is needed in some parts.)

---------- Detailed comments ----------
The paper presents a supplementary book suggestion system that recommends books relevant to a course description. It has been evaluated showing that the proposed method is better than 'termine' and 'rake'. The paper should provides detail description of 'termine' and 'rake', as well as highlight their similarities and differences from the

Smart Media
KOREAN INSTITUTE OF SMART MEDIA

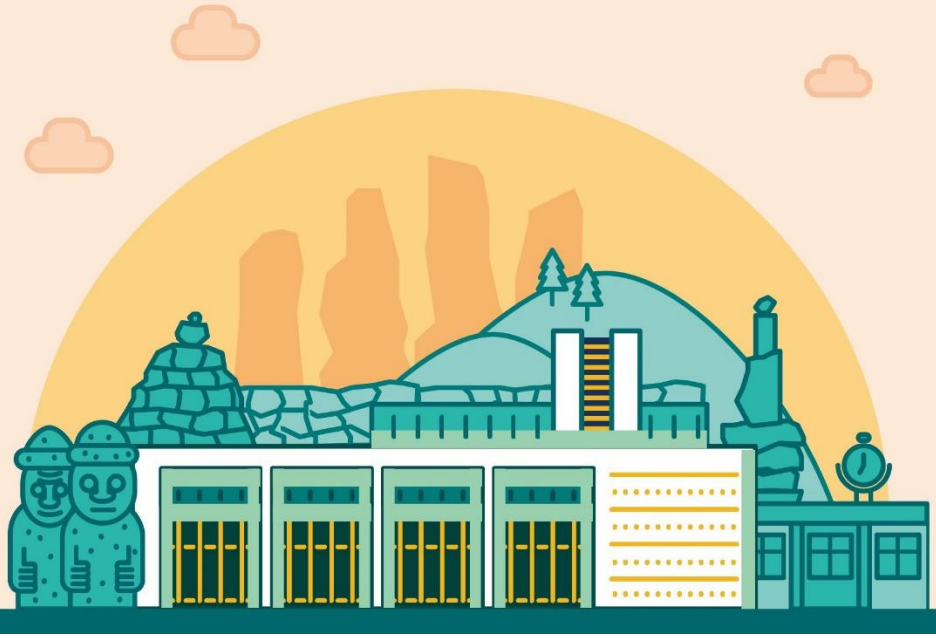
In-Cooperation with acm SIGAPP

SMA 2020

The 9th International Conference on Smart Media & Applications

September 17-19, 2020 | Ramada Plaza Jeju, KOREA

PROGRAM BOOK



An improvement of supplementary book suggestion system

Benchamawan Chaisoongnoen
Computational Innovation Laboratory,
Faculty of Informatics, Burapha University,
Chonburi, 20130, Thailand
benchamawan.gxii@gmail.com

Komate Amphawan
Computational Innovation Laboratory,
Faculty of Informatics, Burapha University,
Chonburi, 20130, Thailand
komate@gmail.com

ABSTRACT

In the era of data-driven, the fields of mathematics, statistics, computer science, information technology and other computer-related fields have become more popular for students and people. Several universities, institutes and companies start to set up new courses across these fields causing a new course may cover several topics to study. By this, it is a challenge for teachers, students and learners to find the right supplementary books that mostly relate to the topics of the new course. To address this issue, an efficient Supplementary Book Suggestion system, also called *eSBS* system, is introduced to efficiently discover books with contents mostly relevant to that of the course. With *eSBS* system, it is more convenient, less time-consuming to find the right book and may help to encourage people for further reading. Experiments were done on CS' course descriptions of Thai universities and show that the *eSBS* system outperforms previous systems in terms of accuracy, precision, recall, F-measure and average & total percentages of coverage on contents.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Computer Science, Supplementary Book, Book recommendation, Course description analysis;

Benchamawan Chaisoongnoen and Komate Amphawan. 2020. An improvement of supplementary book suggestion system. In *Proceedings of The 9th International Conference on Smart Media and Applications, September 17-19, 2020 (SMA 2020).* ACM, New York, NY, USA, Article 4, 3 pages.

1 INTRODUCTION

Currently, technologies in AI, Data Science, Machine learning, Big data, and Robotics are widely studied, developed and applied in several organizations. This causes the need for employees or people having skills in these topics. To respond to this need, universities, institutes and educational companies set up and create new courses across computer-related fields causing a new course may cover several topics to study. By this, it is a challenge for teachers, students and learners to find the right supplementary books that mostly relate to the topics of the new course.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SMA 2020, September 17-19, Jeju, Republic of Korea,
© 2020 Copyright held by the owner/author(s).

To solve this challenge, several book recommender systems are recently proposed based on different aspects of consideration such as user opinion and ToC of books [1], user profile and genres of book [4, 5], user feedback and rating [7], accessing Wikipedia's articles [9], and so on. As above, it can be observed that most of recent book recommendation systems aims to recommend books based on users' action and users' profile. In addition, most of them do not focus on suggestion of books related to a course description belong to a university, institute, and company. With these issues, a supplement book suggestion system named *SBS* system [2] is introduced to create a list of supplementary books relevant to a course in CS without the needs of users' information. However, after releasing to the *SBS* system and deeply analyzing its results, there are some issues to be concerned *i.e. i*) missing on identification of reserved words and keywords, *ii*) redundancy of generated keywords, *iii*) missing on similarity matching, and *iv*) having bias, inequality and unreliability on similarity calculation, respectively. Thus, it is challenging to improve computational performance of the *SBS* system to obtain better quality of results.

To address all the issues, an improvement of CS's supplementary book suggestion system, called *eSBS* (*efficient Supplementary Book Suggestion*) system, is introduced. Based on the *eSBS* system, several computations are greatly revised and modified. The *eSBS* system collects more course descriptions and terminologies, adds new procedures, rearranges the order of computations, reapplies computational procedure, and designs new techniques & calculation in order to improve computational performance and quality of results. Experiments and two comparative studies were conducted on CS's course descriptions of Thai universities. Four metrics (*i.e.* accuracy precision, recall and F-measure) are applied to keyword extraction to evaluate its performance. Besides, two additional metrics (*i.e.* average and total percentages of coverage contents) are used for investigating the quality of books suggested from the *eSBS* system.

2 THE PROPOSED METHOD

The *eSBS* system detailed in Fig. 1 consists of 4 main steps: *i*) Input gathering, *ii*) Keyword Extraction, *iii*) Relevance book searching and keyword extraction, and *iv*) Similarity matching and calculation, respectively. These main steps are similar to that of the *SBS* system [2] but computations in each main step are greatly modified and improved.

A list of improvements is shown in Table 1. First, *eSBS* gathers more CS' course descriptions and more CS's terminologies by exploring new sources, *i.e.* 740 course descriptions from 13 universities instead of 68 course descriptions of 1 university and 28, 392 terminologies from 8 sources increasing from 6, 527 terminologies of 2 sources. Secondly, the text processing in Keyword extraction

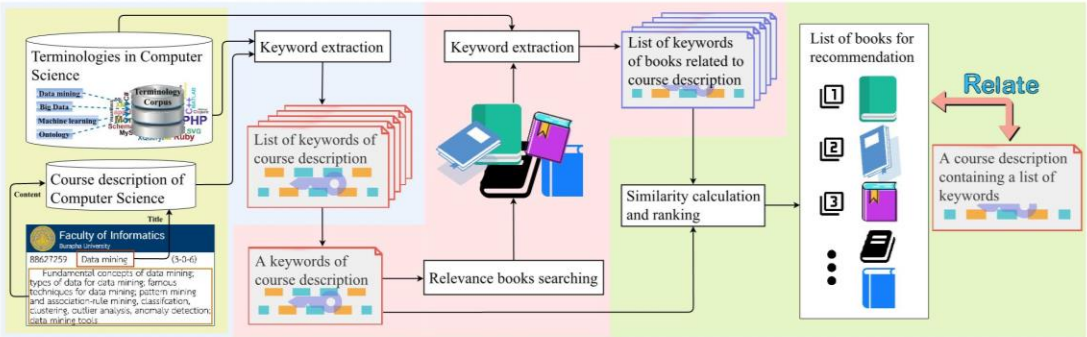


Figure 1: The Framework of the *eSBS* system

Table 1: Improvements of the proposed *eSBS* against the *SBS* system

Step	*SBS* system	*eSBS* system
i) Input gathering	1. 68 course descriptions gathered from 1 university 2. 6, 527 terminologies gathered 2 sources	1. 740 course descriptions gathered from 13 universities 2. 28, 392 gathered from 8 sources
ii) Keyword extraction	1. Text preprocessing 1.1 Stop word removal 1.2 Word error correction 1.3 Sentence Tokenization 1.4 Part of speech tagging 2. Reserved word identification 2.1 Terminology matching 3. Keyword identification by linguistic rules	1. Text preprocessing 1.1 Sentence Tokenization 1.2 Word Tokenization 1.3 Word error correction 1.4 Stop word removal 1.5 Part of speech tagging 2. Reserved word identification 2.1 Terminology matching 2.2 Word stemming & Lemmatization 2.3 Terminology matching 3. Keyword identification by (updated) linguistic rules 4. Adjective & Sub-keyword removal
iii) Relevant book searching & keyword extraction	1. Searching on on-line book store 2. Keyword extraction (similar as above)	1. Searching on on-line book store 2. Keyword extraction (similar as above)
iv) Similarity matching & calculation	1. Similarity matching 1.1 Exact matching 2. Similarity calculation calculated by the number of extracted keywords	1. Similarity matching 1.1 Exact matching 1.2 Paraphrase matching 1.3 Subset matching 1.4 Sub-keyword matching 2. Similarity calculation calculated by the number of topics

procedure is revised. Word tokenization is applied and the order of computation is rearranged to be *i*) Sentence tokenization, *ii*) Word tokenization, *iii*) Word error correction, *iv*) Stop word removal, and *v*) POS tagging, respectively. Besides, word stemming & lemmatization are added to the reserved word identification procedure and terminology matching is applied twice (one before and one after the new added computation). Next, linguistic rules[1] are updated and modified to better identify keywords from words and reserved words. Two new procedures, adjective & sub-keyword removal [6], are applied. Third, three new matching techniques, paraphrase, subset and sub-keyword matching, are designed and utilized. Last, the calculation of similarity percentage is modified by considering the number of topics contained in a course description instead of the number of keywords extracted from a course description.

[1]http://staff.informatics.buu.ac.th/~komate/eSBS

With all of these modifications and improvements, unessential words are precisely removed, more reserved words are discovered, keywords are correctly extracted, redundant keywords are eliminated, similarity matching is more precise and similarity percentage is more concise, reasonable and accurate, respectively. These lead the *eSBS* system to efficiently discover supplementary books related to a course description.

3 EXPERIMENTAL RESULTS

Experiments were conducted on 740 CS' course descriptions gathered from 13 Thai universities ranked in QS Asia University Ranking 2020. Then, two comparative studies were done by considering the computational performance of the *eSBS* system in comparison with the *SBS* system [2] and the two well-known keyword extraction techniques, *i.e. Termine* [3] and *RAKE* [8]. To do that, accuracy, precision, recall, and F-measure on keyword extraction is regarded.

Besides, the quality of suggested books identified by each technique is considered through the average and total percentage of their contents (keywords) cover to the contents of their related course descriptions.

In Table 2, it can be seen that the accuracy, precision, recall, and F-measure on *eSBS*'s keyword extraction significantly outperforms that of other techniques. The accuracy of the *eSBS*'s keyword extraction increases around 28% from *SBS* system and up to 47-60% against *Termine* and *RAKE*. Similarly, precision, recall and F-measure on the *eSBS*'s keyword extraction are greater than *SBS* system around 15-31% and between 25-47% in comparison with to *Termine* and *RAKE*. Thanks to the new collected terminologies and the three additional text processing techniques, *i.e.* terminology detection, adjective removal and sub-word removal, which help the *eSBS* system to extract reserved words and keywords more accurately.

Subsequently, Fig. 2 firstly shows that in average, each suggested book from the *eSBS* system can cover 60-75% of contents belonging to a course description. This is greater than that of other techniques by 8-39%. Moreover, it can be observed that the average percentage of coverage contents of all techniques decrease as the number of suggested books increases. It is because with a low number of suggested books, only high coverage books are selected to suggest. Meanwhile, with a high number of suggested books, some books with low coverage contents are output.

In addition, it is also shown that all suggested books from the *eSBS* system can cover 91-98% of course descriptions' contents. Meanwhile, books from the other techniques can cover 46-89% which is 8-48% less than that of the *eSBS* system. Last, it can be seen that the total coverage of all techniques slightly increases as the number of suggested books increases. With high number of books, more books with different contents are selected causing increasing of contents cover to course descriptions.

Table 2: Efficiency of keyword extraction

	Termine	*RAKE*	*SBS*	*eSBS*
Accuracy	35.00	47.22	56.64	**95.14**
Precision	70.96	61.13	65.40	**96.64**
Recall	40.97	67.67	82.61	**98.71**
F-measure	50.51	64.04	72.45	**97.58**

4 CONCLUSION

In this paper, an efficient supplementary book suggestion system, called *eSBS* (*efficient Supplementary Book Suggestion*) system, is introduced to improve efficiency of the current system on discovering books relevant to a course description of a course. *eSBS* collects more course descriptions and terminologies, revises text processing, adds two new procedures: *i*) adjective removal and *ii*) subset removal. Besides, some linguistic rules are updated. These three changes can help to correctly identify important contents (*i.e.* reserved word and keywords) contained in a course description (also in a book). Last, three new similarity matching techniques and a new similarity calculation are designed and applied to increase accuracy of identifying books relevant to a course description. Based on the experiments, the results show that the new proposed *eSBS* system outperforms the existing *SBS* system and the other two well-know keyword extraction algorithms in the terms of accuracy, recall, precision and F-measure on keyword extraction. Moreover, the percentages that suggested books from the *eSBS* system can cover course contents are also improved.

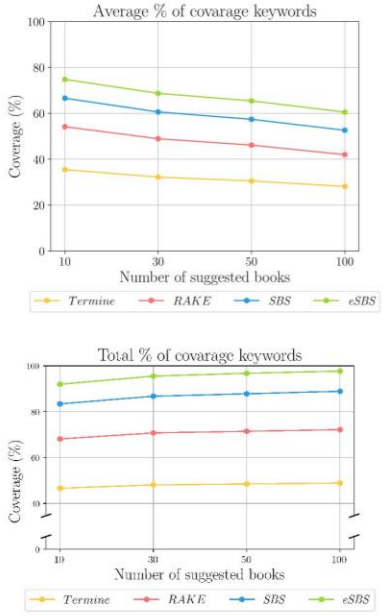


Figure 2: The average and total percentages of keywords from suggested books cover to contents in course descriptions

REFERENCES

[1] Zafar Ali, Shah Khusro, and Irfan Ullah. 2016. A Hybrid Book Recommender System Based on Table of Contents (ToC) and Association Rule Mining. In *Proceedings of the 10th International Conference on Informatics and Systems*. 68–74.

[2] B. Chaisoongnoen, K. Amphawan, and A. Bunpeng. 2018. Supplementary Book Suggestion for Computer Science Courses. In *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*. 84–90.

[3] Katerina Frantzi, Sophia Ananiadou, and Hideki Mima. 2000. Automatic recognition of multi-word terms:. the C-value/NC-value method. *International Journal on Digital Libraries* 3, 2 (2000), 115–130.

[4] ". I. Hariadi and D. Nurjanah. 2017. Hybrid attribute and personality based recommender system for book recommendation. In *2017 International Conference on Data and Software Engineering (ICoDSE)*. 1–5.

[5] Pijitra Jomsri. 2018. FUCL mining technique for book recommender system in library service. *Procedia Manufacturing* 22 (2018), 550 – 557.

[6] Peerapon Kamlangpuech and Komate Amphawan. 2020. A new system for analyzing contents of Computer Science courses. In *7th International Conference on Advanced Informatics: Concepts, Theory and Applicationsy (to be appeared)*.

[7] Gautam Ramakrishnan, V. Saicharan, K. Chandrasekaran, M. V. Rathnamma, and V. Venkata Ramana. 2020. Collaborative Filtering for Book Recommendation System. In *Soft Computing for Problem Solving*. 325–338.

[8] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic Keyword Extraction from Individual Documents. In *Text Mining: Applications and Theory*. 1 – 20.

[9] K. Tsuji. 2019. Book Recommender System for Wikipedia Article Readers in a University Library. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*. 121–126.

Gmail **Benchamawan Chaisoongnoen <benchamawan.gxii@gmail.com>**

Author presentation information and others for SMA 2020

1 ข้อความ

SMA-2020 <sma2020-0@easychair.org> 11 กันยายน 2563 21:34
ถึง: Benchamawan Chaisoongnoen <benchamawan.gxii@gmail.com>

Dear Benchamawan Chaisoongnoen, authors of papers accepted for SMA 2020.

Congraturation again! Our conference dates fast approaching!

Please, prepare your oral/poster presentations with the following guidelines:

1. All accepted papers can be found and the presentation type (oral or poster) is as well at https://sma-2020.github.io/accepted_papers/

2. We have selected best papers based on the review scores plus deep paper analysis. Big congratulation! The result is also available at https://sma-2020.github.io/accepted_papers/ The best paper award ceremony will be at the closing session on September 19. All winners of the best paper selections should come and get your certificate.

3. We have completed the conference programe, which can be accessed at https://sma-2020.github.io/program/

4. For oral presentation, all presentations are scheduled on September 18. Each oral presentation time is 10 minutes with 3-4 minutes of QnA. Please prepare your presentation slides for your oral presentation. In case that you cannot attend in-person due to COVID-19, then please send an email to kism1122@kism.or.kr with your paper ID and title. For those who cannot attend due to COVID-19, then your video presentation file (10 minutes) is required to be sent to kism1122@kism.or.kr until September 15; otherwise your paper will not be included in the conference proceedings.

5. For poster presentation, all presentations are video streaming and scheduled on September 19. Each poster presentation time is 5-6 minutes. Please prepare your video presentation file (5-6 minutes) and send the file to kism1122@kism.or.kr until September 15; otherwise your poster paper will not be included in the conference proceedings. When you send your video presentation file, you must provide your paper ID and title also.

6. For poster presentation's QnA, it will be done in online from 10:00 - 11:00 on September 19. All poster authors must be online and try to best for answering to questions or comments on your poster. The online link for the poster QnA session will be announced at https://sma-2020.github.io/program/

7. Online video streaming links for oral and poster presentations will be announced at https://sma-2020.github.io/program/

We are looking forward to seeing you at the conference or online!

Jong-Hyouk Lee, Program Chair of SMA 2020.

ประวัติย่อของผู้วิจัย

ชื่อ-สกุล	เบญจมวรรณ ใจสูงเนิน
วัน เดือน ปี เกิด	30 กรกฎาคม 2536
สถานที่เกิด	32 หมู่ที่ 10 ตำบล พญาเย็น อำเภอ ปากช่อง จังหวัด นครราชสีมา 30320
สถานที่อยู่ปัจจุบัน	32 หมู่ที่ 10 ตำบล พญาเย็น อำเภอ ปากช่อง จังหวัด นครราชสีมา 30320
ประวัติการศึกษา	มัธยมศึกษาตอนต้น โรงเรียนสระบุรีวิทยาคม
	มัธยมศึกษาตอนปลาย โรงเรียนสระบุรีวิทยาคม
	ปริญญาตรี วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา