



การสกัดรูปแบบคำยืมสำหรับการประเมินระดับความยากง่ายของข้อความในภาษาไทย

Pattern-based Loanword Extraction for Text Readability Assessment in Thai

บุรฮาน หวังเหล็ม

Burhan Wanglem

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา

วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Information Technology

Prince of Songkla University

2560

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์



การสกัดรูปแบบคำยืมสำหรับการประเมินระดับความยากง่ายของข้อความในภาษาไทย

Pattern-based Loanword Extraction for Text Readability Assessment in Thai

บุรฮาน หวังเหล็ม

Burhan Wanglem

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา

วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Information Technology

Prince of Songkla University

2560

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ การศึกษารูปแบบคำยืมสำหรับการประเมินระดับความยากง่ายของข้อความใน
ภาษาไทย

ผู้เขียน นายบุรุษาน หวังเหล็ม

สาขาวิชา เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(ดร.ณัฐพงศ์ ทองเทพ)

คณะกรรมการสอบ

.....ประธานกรรมการ
(ดร.กาญจนา ทองกลิ่น)

.....กรรมการ
(ดร.ณัฐพงศ์ ทองเทพ)

.....กรรมการ
(ดร.พุทธิพร ธนธรรมเมธี)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้บัณฑิตวิทยาลัยรับนี้เป็น
ส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยี
สารสนเทศ

.....
(รองศาสตราจารย์ ดร.ดำรงศักดิ์ ฟ้ารุ่งสว่าง)

คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ.....

(ดร.ณัฐพงศ์ ทองเทพ)

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....

(นายบูรฮาน หวังเหล็ม)

นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และ
ไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....

(นายบูรฮาน หวังเหล็ม)

นักศึกษา

ชื่อวิทยานิพนธ์	การสกัดรูปแบบคำยืมสำหรับการประเมินระดับความยากง่ายของข้อความในภาษาไทย
ผู้เขียน	นายบุรุษาน หวังเหล็ม
สาขาวิชา	เทคโนโลยีสารสนเทศ
ปีการศึกษา	2560

บทคัดย่อ

ในแต่ละวันมนุษย์ต้องอ่านเอกสารเป็นจำนวนมาก และการเลือกเอกสารให้ตรงกับความต้องการได้อย่างรวดเร็วเป็นสิ่งสำคัญสำหรับผู้อ่าน ปัจจัยที่ส่งผลต่อการเลือกเอกสาร ได้แก่ เนื้อหาภายในเอกสารและวิธีการเขียนที่มีความยากง่ายของข้อความ เช่น ความยากง่ายของคำ วลี ไวยากรณ์ อย่างไรก็ตามการวิจัยเกี่ยวกับความยากง่ายของข้อความในเอกสารภาษาไทยอยู่ในขั้นเริ่มต้น ดังนั้นงานวิจัยนี้ได้นำเสนอวิธีสำหรับการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยรูปแบบใหม่ ประกอบด้วยคุณสมบัติของระดับความยากง่ายในข้อความและเทคนิคการประเมินระดับความยากง่าย ซึ่งงานวิจัยนี้ใช้คำยืม 7 ประเภท เป็นคุณสมบัติสำหรับการจัดระดับความยากง่ายของข้อความซึ่งเกิดจากการสังเกตรูปแบบของคำยืมจากชุดเอกสารภาษาไทย ได้แก่ 1) บาลี (Pali: P) 2) สันสกฤต (Sanskrit: S) 3) ตัวการันต์ (Orthography: O) 4) บาลีและสันสกฤต (Pali & Sanskrit: PS) 5) บาลีและตัวการันต์ (Pali & Orthography: PO) 6) สันสกฤตและตัวการันต์ (Sanskrit & Orthography: SO) 7) บาลี สันสกฤต และตัวการันต์ (Pali, Sanskrit, Orthography: PSO) และงานวิจัยนี้ได้นำเสนอเทคนิคการประเมินระดับความยากง่ายของข้อความรูปแบบใหม่ 3 เทคนิค คือ 1) การจัดกลุ่มเอกสารที่คำนวณจากสัดส่วนของจำนวนเอกสารทั้งหมด (Document Clustering A: DoC-A) 2) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุด (Document Clustering B: DoC-B) และ 3) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุดและน้อยที่สุดที่เกิดขึ้นจริง (Document Clustering C: DoC-C) ซึ่งเทคนิคการจัดกลุ่มทั้ง 3 เทคนิคนี้จะจัดระดับความยากง่าย ออกเป็น 3 ระดับ คือ 1) ง่าย 2) ปานกลาง และ 3) ยาก โดยอ้างอิงจากจำนวนคำยืมในเอกสาร และนำผลการประเมินของเทคนิคทั้ง 3 เทคนิค เปรียบเทียบกับผลการประเมินของกลุ่มผู้อ่าน ผลการเปรียบเทียบแสดงให้เห็นว่าการประเมินระดับความยากง่ายด้วยเทคนิค DoC-A มีความใกล้เคียงกับผลการประเมินระดับความยากง่ายของกลุ่มผู้อ่านมากที่สุดถึงร้อยละ 75

คำสำคัญ : การประเมินระดับความยากง่ายของข้อความ, ระบบการเขียน, ภาษาไทย, คำยืม, ภาษาบาลี, ภาษาสันสกฤต, การจัดกลุ่มเอกสาร, การสกัดคำยืม

Thesis Title	Pattern-based Loanword Extraction for Text Readability Assessment in Thai
Author	Mr.Burhan Wanglem
Major Program	Information Technology
Academic	2017

ABSTRACT

We read many documents every day, and it is desirable to be able to choose which documents to read more easily and quickly. In due course, not only what is written but also how it is written are important factors for document choice. There are a variety of characteristics in how a text is written, and one important idea is text readability, which is characterized by difficulty levels of words, phrases, grammar, etc. that are employed. Text readability for Thai documents, however, have not been extensively investigated. This research proposes a new method for text readability assessment for Thai documents which consists of new text readability features and a new readability assessment technique. Based on human observations of a large Thai document set, this research focuses specifically on the use of seven different types of loanwords in Thai: 1) Pali word (P), 2) Sanskrit word (S), 3) Orthography (O), 4) Pali and Sanskrit word (PS), 5) Pali word and Orthography (PO), 6) Sanskrit word and Orthography (SO), and 7) Pali, Sanskrit word and Orthography (PSO). Employing features of these loanwords, we propose three new Thai text readability assessment techniques and compared them with human assessment. These three techniques count the frequency of those loanwords to cluster Thai documents into three levels: Easy, Medium, and Hard. Each technique is based on different clustering methods: 1) Document clustering using the proportion of the total number of documents (DoC-A) 2) Document clustering using class interval calculated from the actual maximum value (DoC-B) 3) Document clustering using class interval calculated from the actual maximum and minimum values (DoC-C). Our comparative experiment among these three techniques shows that the DoC-A technique is most approximate to human assessment at 75% of accuracy.

Keywords: Text Readability Assessment, Writing Systems, Thai Language, Loanwords, Pali Language, Sanskrit Language, Text Clustering, Loanword Extraction

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือของ ดร.ณัฐพงศ์ ทองเทพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยช่วยปรับแก้เนื้อหาวิทยานิพนธ์และแนะนำแนวทางการจัดทำวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

นอกจากนี้วิทยานิพนธ์ฉบับนี้ก็จะสำเร็จมิได้ หากขาดคณะกรรมการสอบวิทยานิพนธ์ ที่คอยแนะนำและปรับแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์มากขึ้น คือ ดร.กาญจนา ทองกลิ่น และ ดร.พุทธิพร ชนธรรมเมธี รวมถึงคณาจารย์และบุคลากรประจำวิทยาลัยการคอมพิวเตอร์ทุกท่าน ที่คอยแนะนำ ตักเตือน อบรมสั่งสอน อำนวยความสะดวก และถ่ายทอดวิชาความรู้ที่เป็นประโยชน์แก่การจัดทำวิทยานิพนธ์ฉบับนี้

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณบิดา มารดา พี่ น้อง และญาติพี่น้องของข้าพเจ้า ที่คอยสนับสนุนและเป็นกำลังใจเสมอมาสำหรับการจัดทำวิทยานิพนธ์ฉบับนี้

บูรฮาน หวังเหล็ม

สารบัญ

	หน้า
บทคัดย่อ	(5)
บทคัดย่อ (ภาษาอังกฤษ)	(6)
กิตติกรรมประกาศ	(7)
สารบัญ	(8)
สารบัญตาราง	(10)
สารบัญภาพประกอบ	(12)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาของปัญหาและปัญหา	1
1.2 คำถามวิจัย	3
1.3 วัตถุประสงค์	3
1.4 สมมติฐาน	4
1.5 ความสำคัญและประโยชน์ของการวิจัย	4
1.6 ขอบเขตของการวิจัย	4
บทที่ 2 การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง	5
2.1 ระบบการเขียน	5
2.1 หลักภาษาไทย	12
2.3 การยืมภาษา	32
2.4 ภาษาสันสกฤต	34
2.5 ภาษาบาลี	34
2.6 หลักการสังเกตคำยืมในภาษาไทย	35
2.7 การประเมินระดับความยากง่ายของข้อความ	41
2.8 การประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย	44
2.9 การจัดกลุ่มเอกสาร	46

สารบัญ (ต่อ)

	หน้า
บทที่ 3 วิธีดำเนินการวิจัย	51
3.1 การสกัดคำยืม	51
3.2 การจัดกลุ่มเอกสาร	67
3.3 การประเมินความยากง่ายของข้อความแบบอัตโนมัติ	78
บทที่ 4 ผลการทดลองและการอภิปรายผล	80
4.1 ผลการทดลองการสกัดคำยืมและการอภิปรายผล	80
4.2 ผลการทดลองการจัดกลุ่มเอกสารและการอภิปรายผล	87
บทที่ 5 สรุปและข้อเสนอแนะ	95
5.1 สรุปผลการทดลอง	95
5.2 ปัญหาและอุปสรรค	96
5.3 ข้อเสนอแนะ	96
5.4 ลักษณะการนำไปใช้ประโยชน์	97
เอกสารอ้างอิง	98
ภาคผนวก ก เอกสารหัวเรื่อง Data mining	106
ภาคผนวก ข เอกสารหัวเรื่อง iOS operating system	120
ภาคผนวก ค แนะนำการใช้งานระบบการตรวจจับตำแหน่งคำยืม	131
ภาคผนวก ง แนะนำการใช้งานระบบการประเมินผลการจัดระดับความยากง่ายของเอกสาร	138
ภาคผนวก จ รายการคำยืมที่จำแนกโดยผู้เชี่ยวชาญ	145
ภาคผนวก ฉ รายการคำยืมที่จำแนกโดยแบบจำลองตรวจจับรูปแบบคำยืม	148
ประวัติผู้เขียน	152

สารบัญตาราง

	หน้า
ตารางที่ 2.1 ตัวอย่างคำที่แทนด้วยเสียง ช (sh) ของแต่ละภาษา	6
ตารางที่ 2.2 มาตรฐานสะกด	16
ตารางที่ 2.3 อักษรไตรยางค์	17
ตารางที่ 2.4 รูปสระ	19
ตารางที่ 2.5 การจำแนกสระตามการออกเสียง	20
ตารางที่ 2.6 ประเภทของกลุ่มอักษรไทย	23
ตารางที่ 2.7 พยัญชนะและสระที่ใช้เขียนภาษาไทย บาลี และสันสกฤต	36
ตารางที่ 2.8 วรรณคดีและแถวของพยัญชนะสำหรับภาษาบาลี	38
ตารางที่ 2.9 คุณสมบัติของเทคนิคการจัดกลุ่มแต่ละเทคนิค	49
ตารางที่ 3.1 รายละเอียดของเอกสารที่จัดเก็บ	55
ตารางที่ 3.2 รูปแบบสำหรับสกัดคำยืม	61
ตารางที่ 3.3 ตัวอย่างการแปลงรูปแบบที่สังเกตได้เป็นรูปแบบเงื่อนไข	62
ตารางที่ 3.4 ประเภทของคำยืมและไม่ใช่คำยืมสำหรับการจำแนก	64
ตารางที่ 3.5 การจำแนกประเภทคำโดยผู้เชี่ยวชาญ	65
ตารางที่ 3.6 Confusion matrix	66

สารบัญตาราง (ต่อ)

	หน้า
ตารางที่ 4.1 ค่าเฉลี่ย Precision ของการสกัดคำยืม	81
ตารางที่ 4.2 ค่าเฉลี่ย Recall ของการสกัดคำยืม	82
ตารางที่ 4.3 ค่าเฉลี่ย F-measure ของการสกัดคำยืม	83
ตารางที่ 4.4 ค่าเฉลี่ยความถูกต้องของการสกัดคำยืม	84
ตารางที่ 4.5 ผลการประเมินความใกล้เคียงในการจัดกลุ่มของกลุ่มตัวอย่าง	87
ตารางที่ 4.6 สรุปการประเมินผลและการคำนวณผลความใกล้เคียงในการจัดกลุ่ม ของหัวข้อเรื่อง Data mining	88
ตารางที่ 4.7 สรุปการประเมินผลและการคำนวณผลความใกล้เคียงในการจัดกลุ่ม ของหัวข้อเรื่อง iOS operating system	89
ตารางที่ 4.8 การเปรียบเทียบงานวิจัยด้านการประเมินระดับความยากง่ายของข้อความ ในเอกสารภาษาไทย	93

สารบัญภาพประกอบ

	หน้า
รูปที่ 2.1 การใช้สระและไม่ใช้สระของอักษรอาหรับและฮีบรู	6
รูปที่ 2.2 การแทนสัญลักษณ์ของอักษรในแต่ละพยางค์ระหว่างอักษรฮิระกานะ (Hiragana) และอักษรเอสกีโม (Inuktitut)	7
รูปที่ 2.3 Old European	9
รูปที่ 2.4 Harappa script	9
รูปที่ 2.5 Proto-Elamite	10
รูปที่ 2.6 Linear A	10
รูปที่ 2.7 Phaistos Disk	11
รูปที่ 2.8 อักษรเทวนาครี (Devanagari) ที่ประกอบด้วยพยัญชนะ สระ พยัญชนะ ที่กำกับสระ และกลุ่มพยัญชนะ	12
รูปที่ 2.9 โครงสร้างลำดับชั้นของหน่วยภาษาไทย	13
รูปที่ 2.10 ที่มาของอักษรไทย	14
รูปที่ 2.11 ตัวอย่างข้อความภาษาไทย	22
รูปที่ 2.12 ลักษณะการจัดกลุ่มแบบ Agglomerative และ Divisive	48
รูปที่ 3.1 วิธีการดำเนินการวิจัย	52
รูปที่ 3.2 การดำเนินการวิจัยการสกัดคำยืม	53
รูปที่ 3.3 ตัวอย่างการค้นหาเอกสารด้วยโปรแกรมค้นหาของกูเกิล (Google)	54
รูปที่ 3.4 ตัวอย่างส่วนต่อประสานกับผู้ใช้ของเว็บไซต์เอกสารดิจิทัล	57
รูปที่ 3.5 ตัวอย่างเอกสารข้อความดิจิทัล	58
รูปที่ 3.6 ตัวอย่างชุดเอกสารข้อความดิจิทัลที่ผ่านการจำกัดจำนวนอักขระ	59
รูปที่ 3.7 ตัวอย่างข้อความที่ผ่านการตัดคำ	60
รูปที่ 3.8 ตัวอย่างการแปลงรูปแบบที่สามารถสังเกตกลายเป็นรูปแบบเงื่อนไจ	61
รูปที่ 3.9 ตัวอย่างการตรวจจับตำแหน่งคำยืมในข้อความ	63
รูปที่ 3.10 ตัวอย่างการจำแนกคำยืมในข้อความ	63

สารบัญภาพประกอบ (ต่อ)

	หน้า
รูปที่ 3.11 การดำเนินการวิจัยการจัดกลุ่มเอกสาร	68
รูปที่ 3.12 ตัวอย่างการจัดกลุ่มเอกสารที่มีคะแนนเท่ากัน	69
รูปที่ 3.13 เทคนิคการจัดกลุ่มเอกสาร DoC-A	70
รูปที่ 3.14 ตัวอย่างในกรณีค่า N เป็นจำนวนจริงของ DoC-A	70
รูปที่ 3.15 เทคนิคการจัดกลุ่มเอกสาร DoC-B	71
รูปที่ 3.16 เทคนิคการจัดกลุ่มเอกสาร DoC-C	72
รูปที่ 3.17 ขั้นตอนการใช้งานระบบประเมินผลการจัดกลุ่มเอกสารของแต่ละเทคนิค	74
รูปที่ 3.18 ส่วนต่อประสานกับผู้ใช้ในการเข้าใช้งานระบบประเมิน	76
รูปที่ 3.19 ส่วนต่อประสานกับผู้ใช้ของผลการจัดกลุ่มเอกสารแต่ละเทคนิคในระบบประเมิน	76
รูปที่ 3.20 ส่วนต่อประสานกับผู้ใช้ของเนื้อหาในเอกสารในระบบประเมิน	77
รูปที่ 3.21 ส่วนต่อประสานกับผู้ใช้ของการประเมินผลในระบบประเมิน	77
รูปที่ 3.22 ขั้นตอนการใช้งานระบบการประเมินระดับความยากง่ายของ ข้อความในเอกสารภาษาไทยแบบอัตโนมัติ	79
รูปที่ 4.1 การสรุปผลของการประเมินความถูกต้องในการสกัดคำยืมแต่ละประเภท	85
รูปที่ 4.2 ร้อยละจำนวนกลุ่มตัวอย่างในการประเมินความใกล้เคียงสำหรับการจัดกลุ่ม เอกสารของเทคนิคในหัวเรื่อง Data mining	90
รูปที่ 4.3 ร้อยละจำนวนกลุ่มตัวอย่างในการประเมินความใกล้เคียงสำหรับการจัดกลุ่ม เอกสารของเทคนิคในหัวเรื่อง iOS operating system	91
รูปที่ 4.4 ค่าเฉลี่ยระดับความใกล้เคียงในการจัดกลุ่มเอกสารของเทคนิคในแต่ละหัวเรื่อง	92

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหาและปัญหา

จากสถิติการอ่านของประชากรในประเทศไทย ปี พ.ศ. 2558 ประชากรส่วนใหญ่ ร้อยละ 41.9 อ่านเอกสารเพื่อเพิ่มพูนความรู้ (สำนักงานสถิติแห่งชาติ, 2555) โดยเอกสารประเภทให้ความรู้ที่มีหัวเรื่องเดียวกันจะมีลักษณะการเขียนแตกต่างกัน เป็นเพราะวัตถุประสงค์ของผู้แต่งเอกสารแต่ละคนแตกต่างกัน ตัวอย่างเช่น นักเขียน 2 คน ต้องการเขียนหนังสือการให้ความรู้เกี่ยวกับต้นไม้ โดยนักเขียนคนแรกต้องการเขียนหนังสือสำหรับกลุ่มวัยเด็ก ส่งผลให้ลักษณะเนื้อหาภายในหนังสือเน้นการสื่อความหมายโดยใช้รูปภาพมากกว่าข้อความและการเขียนข้อความที่ไม่ซับซ้อน ตรงกันข้ามสำหรับนักเขียนคนต่อมาต้องการเขียนหนังสือสำหรับกลุ่มวัยเยาวชน ส่งผลให้ลักษณะของเนื้อหาในหนังสือเน้นการเขียนอธิบายโดยใช้ข้อความมากกว่ารูปภาพและใช้คำศัพท์เฉพาะทาง เพื่อเพิ่มพูนความรู้ตามกลุ่มวัย จากตัวอย่างดังกล่าวหากมีผู้อ่านต้องการเลือกเอกสารเพื่ออ่านหนึ่งฉบับที่เหมาะสมกับตนเองในหัวเรื่องใดหัวเรื่องหนึ่ง ในขั้นตอนเบื้องต้นผู้อ่านจะต้องเปิดอ่านเอกสารแต่ละฉบับเพื่อค้นหาเอกสารที่เหมาะสมกับตนเอง หากเอกสารที่กำลังเลือกในขณะนั้นมีจำนวนมากส่งผลให้ผู้อ่านใช้ระยะเวลาในการเลือกเอกสารเพิ่มขึ้นตามจำนวนเอกสาร เนื่องจากโดยปกติระยะเวลาในการเลือกเอกสารขึ้นกับจำนวนเอกสารที่กำลังเลือกในขณะนั้น

จากปัญหาในข้างต้น หากเอกสารที่ผู้อ่านกำลังเลือกในขณะนั้นถูกจัดกลุ่มและจัดลำดับตามระดับความยากของเนื้อหาในเอกสาร ส่งผลให้ผู้อ่านสามารถเลือกเอกสารให้ตรงตามระดับความสามารถในการอ่านของตนเองได้รวดเร็วยิ่งขึ้น เป็นเหตุผลที่ผู้วิจัยต้องการศึกษาหา

เทคนิคในการประเมินระดับความยากง่ายของเอกสาร โดยประเมินจากข้อความในเอกสาร เนื่องจากระดับความยากในการอ่านและระดับความเข้าใจของผู้อ่านขึ้นกับคุณลักษณะทางภาษาของข้อความในเอกสาร ส่งผลให้งานวิจัยส่วนใหญ่ที่ผ่านมามุ่งเน้นการประเมินระดับความยากง่ายของข้อความในเอกสาร (Collins-Thompson, 2014)

สำหรับวิธีการประเมินระดับความยากง่ายของข้อความที่ใช้ในปัจจุบันมีหลากหลาย เช่น Flesh (Flesch, 1948) Flesh-Kincaid Score (Kincaid, *et al.*, 1975) ซึ่งใช้วิธีการพิจารณาคำ และความซับซ้อนของไวยากรณ์ เช่น การหาค่าเฉลี่ยของจำนวนพยางค์ต่อคำ การหาค่าเฉลี่ยของจำนวนคำต่อประโยค อย่างไรก็ตามวิธีการดังกล่าวเหมาะสำหรับใช้ในภาษาอังกฤษ แต่ยังมีข้อจำกัดในการนำไปประยุกต์ใช้กับภาษาอื่น ๆ เช่น ภาษาไทย เนื่องจากภาษาไทยเป็นระบบการเขียนแบบอักษรพยางค์ (Syllabic Alphabets) (Ager, 2017) เป็นระบบภาษาที่ยากต่อกระบวนการประมวลผลข้อความ เช่น การหาจุดจบของคำหรือประโยค โดยงานวิจัยที่เกี่ยวข้องกับการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยจำเป็นต้องดำเนินการประมวลผลข้อความในขั้นตอนการเตรียมข้อมูลก่อนนำข้อความเข้าสู่กระบวนการประเมินระดับความยากง่ายเป็นกระบวนการถัดไป ส่งผลให้งานวิจัยในด้านนี้มีจำนวนน้อย เช่น ผลงานของ Daowadung and Chen, (2011) Daowadung and Chen, (2012) Tongtep, *et al.*, (2014) และ Chen and Daowadung, (2015) ผู้วิจัยต้องการพัฒนาเทคนิคการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย

เทคนิคที่ผู้วิจัยใช้สำหรับประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย คือ การสกัดและนับจำนวนคำยืมในเอกสาร เนื่องจากคำยืมของภาษาไทยมีลักษณะที่ตรงกันข้ามกับคำไทยแท้และถูกนำมาใช้เฉพาะทาง เช่น คำราชาศัพท์ คำศัพท์ทางศาสนา คำศัพท์ในวรรณคดี (ร้านา พลอยขาว, 2549) ส่งผลให้มีความเฉพาะเจาะจงในข้อความ ซึ่งสามารถนำมาเป็นคุณสมบัติในการประเมินความยากง่ายของข้อความในเอกสารภาษาไทยได้ โดยคำยืมที่ผู้วิจัยสกัดและนับจำนวนดังที่กล่าวในข้างต้น คือ คำยืมที่มาจากภาษาบาลี สันสกฤต และภาษาอื่น ๆ ที่มีเครื่องหมายที่ขนาดเท่ากับด้านบนพยัญชนะ (ตัวการ์นต์) หลังจากได้จำนวนคำยืมของแต่ละเอกสาร แล้วจึงนำเอกสารเหล่านั้นจัดกลุ่มตามเทคนิคการจัดกลุ่มแต่ละเทคนิคที่ผู้วิจัยได้พัฒนาขึ้นจำนวน 3 เทคนิค ดังนี้ 1) การจัดกลุ่มเอกสารที่คำนวณจากสัดส่วนของจำนวนเอกสารทั้งหมด

(Document Clustering A: DoC-A) 2) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุด (Document Clustering B: DoC-B) 3) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุดและน้อยที่สุดที่เกิดขึ้นจริง (Document Clustering C: DoC-C) โดยการจัดกลุ่มแต่ละเทคนิคเป็นการจัดกลุ่มตามระดับความยากง่ายที่ได้กำหนด คือ ระดับง่าย (Easy) ระดับปานกลาง (Medium) ระดับยาก (Hard) เมื่อเอกสารถูกจัดระดับความยากง่ายออกเป็นกลุ่ม ผู้อ่านสามารถค้นหาเอกสารที่เหมาะสมกับตนเอง โดยที่ผู้อ่านไม่จำเป็นต้องเปิดอ่านเอกสารเพื่อดูเนื้อหาทุกฉบับ ส่งผลให้ผู้อ่านสามารถค้นหาเอกสารที่เหมาะสมกับตนเองรวดเร็วยิ่งขึ้น

1.2 คำถามวิจัย

- 1) คำชี้แจงสามารถใช้เป็นคุณสมบัติสำหรับการประเมินระดับความยากง่ายของข้อความได้หรือไม่ ?
- 2) เทคนิคการจัดกลุ่มเอกสารที่พัฒนาขึ้นสามารถประเมินความยากง่ายของเอกสารได้หรือไม่ ?

1.3 วัตถุประสงค์

- 1) เพื่อศึกษาคุณสมบัติสำหรับการประเมินความยากง่ายของข้อความในเอกสารดิจิทัลที่เป็นภาษาไทย
- 2) เพื่อจัดกลุ่มเอกสารสำหรับการประเมินความยากง่ายของข้อความในเอกสารดิจิทัลที่เป็นภาษาไทย

1.4 สมมติฐาน

- 1) คำชี้แจงสามารถใช้เป็นคุณสมบัติสำหรับการประเมินระดับความยากง่ายของข้อความ
- 2) เทคนิคการจัดกลุ่มเอกสารที่พัฒนาขึ้นสามารถประเมินความยากง่ายของเอกสาร

1.5 ความสำคัญและประโยชน์ของการวิจัย

- 1) ผู้อ่านสามารถเลือกอ่านเอกสารให้เหมาะสมกับตนเอง
- 2) ผู้อ่านสามารถค้นหาเอกสารที่เหมาะสมกับตนเองได้รวดเร็วยิ่งขึ้น
- 3) เป็นงานวิจัยที่สามารถนำไปต่อยอดในงานวิจัยทางการประเมินระดับความยากง่ายในภาษาไทยได้

1.6 ขอบเขตของการวิจัย

- 1) ประเมินระดับความยากง่ายของเอกสารตามความยากง่ายของข้อความ
- 2) ดำเนินการโดยใช้เอกสารดิจิทัลภาษาไทยที่ได้จากโปรแกรมค้นหา (Search engine) เป็นกรณีศึกษา
- 3) ดำเนินการโดยใช้เอกสารหัวเรื่องเดียวกันและจำนวนเอกสารในแต่ละหัวเรื่องต้องไม่น้อยกว่า 2 ฉบับ

บทที่ 2

การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง

2.1 ระบบการเขียน

ระบบการเขียน (Writing System) คือ ชุดของสัญลักษณ์หรือตัวอักษรสำหรับการสื่อสารแทนการพูดในภาษาต่าง ๆ อย่างเป็นระบบ โดยบุคคลที่ใช้ระบบเดียวกันสามารถอ่านและเข้าใจในระบบการเขียนนั้นได้ตรงกันโดยไม่จำเป็นต้องมีความรู้เฉพาะทางสำหรับการทำความเข้าใจของสัญลักษณ์หรืออักษรนั้น ๆ ซึ่งแตกต่างจากสัญลักษณ์ในแผ่นป้าย สัญลักษณ์ทางคณิตศาสตร์ สัญลักษณ์ภาพวาด และอื่น ๆ ชนิดของระบบการเขียนแบ่งออกเป็น 6 ชนิด ดังนี้

2.1.1 Abjad

Abjad คือ ระบบการเขียนที่โดยปกติแล้วจะมีเพียงแค่พยัญชนะเท่านั้น แต่จะมีการกำกับสระในบางบริบทเท่านั้น เช่น บริบทสำหรับเด็ก บริบททางศาสนา ดังรูปที่ 2.1 แสดงถึงการใช้สระและไม่ใช้สระของภาษาอาหรับ (Arabic) และภาษาฮีบรู (Hebrew) โดยสัญลักษณ์ที่ล้อมด้วยเส้นประ หมายถึง สระ นอกจากนี้ระบบการเขียน Abjad จะมีทิศทางการเขียนจากขวาไปซ้าย ยกเว้นอักษรยูการิติก (Ugaritic) แต่ในปัจจุบันเลิกใช้งานไปแล้ว ซึ่งอักษรที่กำลังใช้ในปัจจุบันมีดังนี้

- 1) ฮีบรู
- 2) อาหรับ
- 3) ซาแมเรีย (Samaritan) และ
- 4) ซีเรีย (Syriac) (Ager, 2017)



รูปที่ 2.1 การใช้สระและไม่ใช้สระของอักษรอาหรับและฮีบรู (Ager, 2017)

2.1.2 Alphabet

Alphabet คือ ระบบการเขียนซึ่งประกอบด้วยพยัญชนะและสระ ในบางภาษาสำหรับแต่ละอักษร หรือการรวมกันของอักษรจะแทนเพียงแค่หนึ่งเสียง เช่น ภาษาเช็ก (Czech) แต่ในทางกลับกันในบางภาษาจะแทนได้หลายเสียง เช่น ภาษาอังกฤษ (English) แสดงดังตารางที่ 2.1 โดยอักษรที่ขีดเส้นใต้ หมายถึง อักษรที่แทนเสียง ช (sh) ของแต่ละภาษา

ตารางที่ 2.1 ตัวอย่างคำที่แทนด้วยเสียง ช (sh) ของแต่ละภาษา (Ager, 2017)

ภาษา	ตัวอย่างคำ
ภาษาแอฟริกา (Afrikaans)	<u>sj</u> ampoe (shampoo)
ภาษาเช็ก (Czech)	<u>ř</u> plouchat (to splash)
ภาษาอังกฤษ (English)	<u>sh</u> ip, elect <u>i</u> on, mach <u>in</u> e, miss <u>ion</u> , press <u>ur</u> e, <u>sch</u> edule
ภาษาฝรั่งเศส (French)	<u>ch</u> at (cat)
ภาษาเยอรมัน (German)	<u>sch</u> ön (beautiful)
ภาษาไอริช (Irish)	<u>mi</u> se (me)
ภาษาเวลส์ (Welsh)	<u>si</u> arad (to speak)
ภาษาอาร์มีเนีย (Armenian)	<u>շ</u> աբաթոց (Saturday)
ภาษาจอร์เจีย (Georgian)	ბოლო <u>ბ</u> ო (sorry)

ตัวอย่างอักษรในระบบ Alphabet ที่กำลังใช้ในปัจจุบัน เช่น อาร์เมเนีย (Armenian) รัสเซีย (Cyrillic) จอร์เจีย (Georgian) กรีก (Greek) เกาหลี (Korean) ละติน โรมัน (Latin/Roman) มองโกเลีย (Mongolian) (Ager, 2017)

2.1.3 Syllabaries

Syllabaries คือ ระบบการเขียนที่มีความเกี่ยวข้องกับการออกเสียง โดย Syllabaries จะใช้สัญลักษณ์ที่แสดงถึงพยางค์ ซึ่งพยางค์ใน Syllabaries โดยส่วนใหญ่จะประกอบด้วยพยัญชนะ และสระ หรือมีเพียงแค่สระอย่างเดียว

Hiragana		Inuktitut	
ka	か	ki	p
ki	き	ku	d
ku	く	ka	b
ke	け	-k	b
ko	こ		

รูปที่ 2.2 การแทนสัญลักษณ์ของอักษรในแต่ละพยางค์ระหว่างอักษรฮิระกานะ (Hiragana) และ อักษรเอสกีโม (Inuktitut) (Ager, 2017)

จากรูปที่ 2.2 แสดงถึง ตัวอย่างการเปรียบเทียบการแทนสัญลักษณ์ของอักษรในแต่ละพยางค์ระหว่างอักษร Hiragana และ Inuktitut ในอักษร Hiragana จะแทนเสียงของแต่ละพยางค์ด้วยสัญลักษณ์ที่แตกต่างกัน ส่วนอักษร Inuktitut จะแทนเสียงของแต่ละพยางค์ด้วยความแตกต่างของการวางแนวของสัญลักษณ์ และสัญลักษณ์ขนาดเล็กจะใช้เขียนแทนพยัญชนะสะกด

ตัวอย่างอักษรในระบบ Syllabaries ที่ใช้ในปัจจุบัน เช่น เซอโรกี (Cherokee) เอสคายา (Eskayan) ฮิระกานะ คาทาคานะ (Katakana) (Ager, 2017)

2.1.4 Samanto-phonetic

Samanto-phonetic คือ ระบบการเขียนที่ใช้สัญลักษณ์แทนเสียงและความหมาย โดยทั่วไปบทเขียนในระบบการเขียนนี้จะมีการใช้สัญลักษณ์จำนวนมาก โดยเริ่มตั้งแต่หลักร้อย จนถึงหลักพัน บทเขียนในระบบการเขียนนี้ถูกเรียกว่า Logophonetic Morphonemic Logographic หรือ Logosyllabic สัญลักษณ์ของระบบการเขียนแบบ Samanto-phonetic แบ่งออกเป็น 3 ประเภท ดังนี้

2.1.4.1 Pictograms และ Logograms โดยที่ Pictograms คือ สัญลักษณ์ที่มีความคล้ายคลึงกับรูปลักษณะของสิ่งที่มีหมายถึง และ Logogram คือ สัญลักษณ์แทนส่วนหนึ่ง หรือ ทั้งหมดของคำ

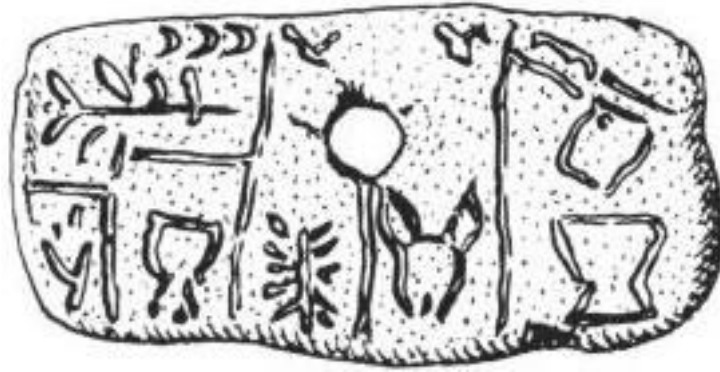
2.1.4.2 Ideograms คือ สัญลักษณ์ที่แทนความคิดที่เป็นนามธรรม

2.1.4.3 Compound characters คือ สัญลักษณ์ที่ประกอบด้วย 2 ส่วน คือ 1) องค์ประกอบเชิงความหมาย ซึ่งจะแสดงถึงความหมาย และ 2) องค์ประกอบเกี่ยวกับเสียง ซึ่งจะแสดงถึงการออกเสียง

ในปัจจุบันอักษรที่อยู่ในระบบการเขียน Samanto-phonetic มีดังนี้ จีน (Chinese) ญี่ปุ่น (Japanese) (Ager, 2017)

2.1.5 Undeciphered script

Undeciphered script คือ ระบบการเขียนที่ยังไม่ได้รับการถอดรหัส หรืออาจจะได้รับการถอดรหัสแค่บางส่วนเท่านั้น (Ager, 2017) ซึ่งได้แสดงตัวอย่างของระบบการเขียน Undeciphered script ดังภาพต่อไปนี้



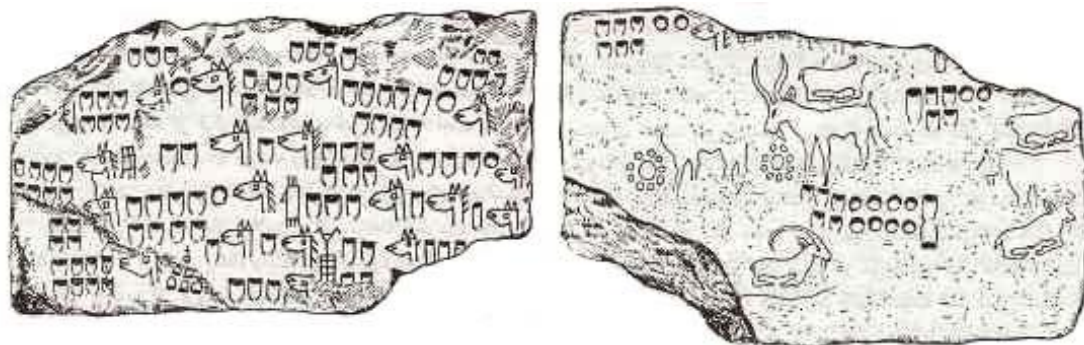
รูปที่ 2.3 Old European (Ager, 2017)

จากรูปที่ 2.3 คือ ตัวอย่างสัญลักษณ์ที่ถูกประดิษฐ์ในช่วง 6,000 ถึง 4,500 ปี ก่อนคริสตศักราช ซึ่งได้ถูกค้นพบในพื้นที่ทางตะวันออกเฉียงใต้ของทวีปยุโรป (Ager, 2017)



รูปที่ 2.4 Harappa script (Ager, 2017)

จากรูปที่ 2.4 คือ ตัวอย่างสัญลักษณ์ที่ถูกใช้ในอารยธรรมลุ่มแม่น้ำสินธุของอินเดีย ในช่วง 3,500 ถึง 2,000 ปี ก่อนคริสตศักราช บางคนเชื่อว่านี่ไม่ใช่สัญลักษณ์ของภาษา ในขณะที่บางคนเชื่อว่าสัญลักษณ์นี้เป็นภาษาของชาวพื้นเมืองดั้งเดิมของอินเดีย (Dravidian) (Ager, 2017)



รูปที่ 2.5 Proto-Elamite (Ager, 2017)

จากรูปที่ 2.5 คือ ตัวอย่างสัญลักษณ์ที่ปรากฏขึ้นในประมาณ 2,900 ปี ก่อนคริสต์ศักราช ในเมืองซูซา (Susa) ซึ่งเป็นเมืองหลวงของจักรวรรดิเอลาไมท์ (Elamite) ปัจจุบันสถานที่แห่งนี้ คือ ประเทศอิหร่าน (Ager, 2017)



รูปที่ 2.6 Linear A (Ager, 2017)

จากรูปที่ 2.6 คือ ตัวอย่างสัญลักษณ์ที่ใช้ในช่วง 1,800 ถึง 1,450 ปี ก่อนคริสต์ศักราช เป็นแผ่นจารึกที่บันทึกลงบนดินเหนียว ถูกค้นพบที่พระราชวังคนอสซอส (Knossos) บนเกาะครีต (Crete) ประเทศกรีซ (Ager, 2017)



รูปที่ 2.7 Phaistos Disk (Ager, 2017)

จากรูปที่ 2.7 คือ ตัวอย่างสัญลักษณ์ที่ใช้ในประมาณ 2,000 ปี ก่อนคริสต์ศักราช เป็นแผ่นจารึกลงบนดินเหนียว ถูกค้นพบที่พระราชวังไมนวน (Minoan Palace) ในเมืองไฟตอส (Phaistos) บนเกาะครีต ประเทศกรีซ (Ager, 2017)

2.1.6 Syllabic alphabets

Syllabic alphabets หรือ Abugidas คือ ระบบการเขียนที่มีส่วนประกอบหลัก คือ พยางค์ โดยพยางค์สร้างขึ้นจากพยัญชนะ ซึ่งพยัญชนะทุกอักษรจะมีสระที่ไม่ปรากฏรูปกำกับอยู่ เมื่อมีการกำกับเครื่องหมายหรือรูปสระเข้าไปจะสามารถเปลี่ยเสียงหรือเจียบเสียงได้

Consonants					Consonants with vowel diacritics				
क	ख	ग	घ	ङ	का	खि	गी	घु	ङु
ka	kha	ga	gha	ṅa	kā	khi	gī	ghu	ṅū
Independent vowel letters					Conjunct consonants				
अ	आ	इ	ई	उ	क्क	क्ख	ग्घ	घङ	ङ्क
a	ā	i	ī	u	kka	khka	ggha	ghṅa	ṅka

รูปที่ 2.8 อักษรเทวนาครี (Devanagari) ที่ประกอบด้วยพยัญชนะ สระ พยัญชนะที่กำกับสระ และกลุ่มพยัญชนะ (Ager, 2017)

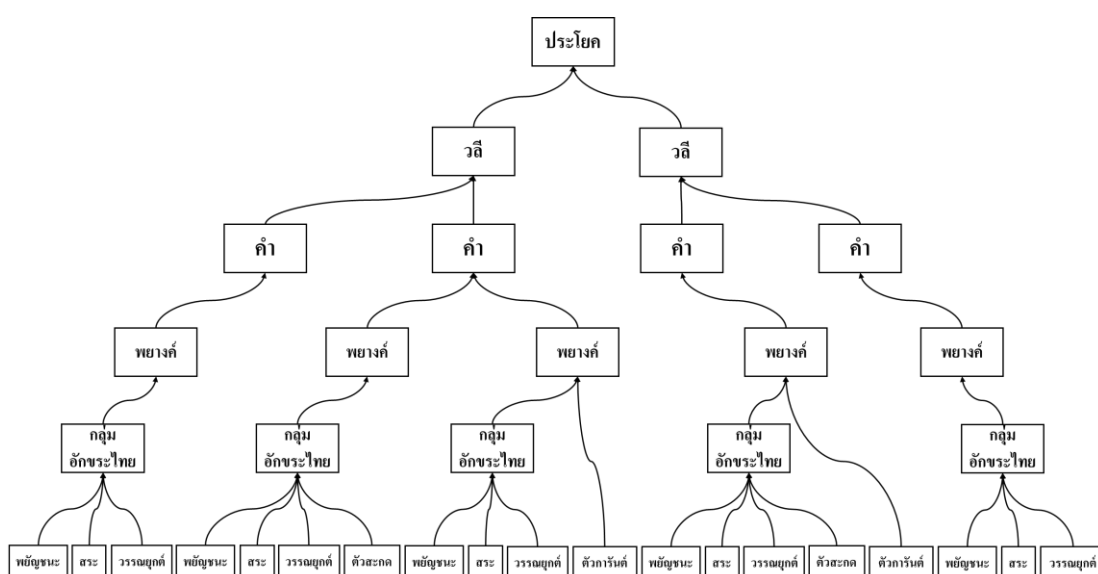
จากรูปที่ 2.8 แสดงให้เห็นถึง อักษรเทวนาครีที่ประกอบด้วยพยัญชนะ สระ พยัญชนะที่กำกับสระ และกลุ่มพยัญชนะที่เชื่อมกัน โดยส่วนที่ล้อมด้วยเส้นประ คือ รูปสระที่เขียนกำกับในพยัญชนะ ซึ่งอักษรเทวนาครีใช้เขียนภาษาฮินดู (Hindi) เนปาล (Nepali) บาลี (Pali) สันสกฤต (Sanskrit) และอื่น ๆ

ตัวอย่างอักษรในระบบ Syllabic alphabets ที่ใช้ในปัจจุบัน เช่น บังกลาเทศ (Bengali) พม่า (Burmese) เทวนาครี เฟรเซอร์ (Fraser) เอธิโอเปีย (Ethiopic) ปัญจาบ (Punjabi) ม้ง (Hmong) เอสกีโม (Inuktitut) เขมร (Khmer) ลานนา (Lanna) ลาว (Lao) มาลายาลัม (Malayalam) ฉาน (Shan) ทมิฬ (Tamil) ทิเบต (Tibetan) ไทย (Thai) (Ager, 2017)

2.1 หลักภาษาไทย

ภาษาไทยมีระบบการเขียนแบบ Syllabic Alphabet (สามารถดูรายละเอียดได้ที่ ข้อ 2.1.6) ที่เป็นพยางค์ ซึ่งแต่ละพยางค์ของภาษาไทย อย่างน้อยต้องประกอบด้วย 3 ส่วน คือ พยัญชนะ สระ และวรรณยุกต์ เช่น ตา ไป นา บางพยางค์ประกอบด้วย 4 ส่วน คือ พยัญชนะ สระ

วรรณยุกต์ และตัวสะกด เช่น คน กิน ข้าว หรือพยัญชนะ สระ วรรณยุกต์ และการันต์ เช่น โหล่ เล่ห์ ฟ่าห์ และบางพยางค์ประกอบด้วย 5 ส่วน คือ พยัญชนะ สระ วรรณยุกต์ ตัวสะกด และการันต์ เช่น แพทย์ รัศมี สิทธิ โดยพยางค์จะเป็นหน่วยเสียงของคำ ซึ่งในหนึ่งคำอาจจะประกอบไปด้วยหนึ่งพยางค์เป็นต้นไป เมื่อนำคำมาประกอบกันอย่างน้อยสองคำขึ้นไปกลายเป็นกลุ่มคำหรือวลี เมื่อนำวลีที่เกี่ยวข้องกันมาเรียงต่อกันเป็นระเบียบจนมีเนื้อหาที่ครบสมบูรณ์โดยปกติจะต้องมีบทประธานและบทกริยา เรียกว่า ประโยค (คำชัย ทองหล่อ, 2550) แสดงดังรูปที่ 2.9

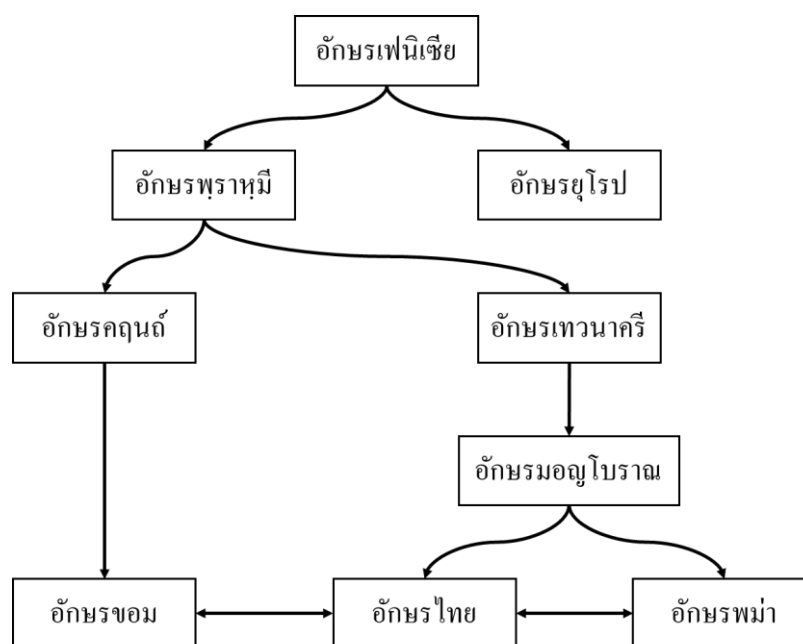


รูปที่ 2.9 โครงสร้างลำดับชั้นของหน่วยภาษาไทย

2.2.1 อักษรไทย

เป็นที่ทราบกันว่าอักษรไทยนั้นมาจากการประดิษฐ์คิดค้น โดยพ่อขุนรามคำแหงมหาราช โดยอักษรไทยที่พ่อขุนรามคำแหงประดิษฐ์ขึ้นมีต้นเค้ามาจากอักษรขอม อักษรมอญ อักษรพม่า เนื่องจากมีลักษณะตัวอักษรบางตัวที่คล้ายคลึงกัน อักษรมอญและพม่าได้ดัดแปลงมาจากอักษรเทวนาครี ซึ่งเป็นอักษรของอินเดียฝ่ายเหนือ ใช้เขียนหลายภาษา เช่น เนปาล บาลี สันสกฤต และมีระบบการเขียนที่เหมือนกับภาษาไทย คือ Syllabic Alphabets (สามารถดูรายละเอียดได้ที่ข้อ 2.1.6) ส่วนอักษรขอมได้ดัดแปลงอักษรคฤนต์ หรือ ครนต์ ของอินเดียฝ่ายใต้ โดยทั้งอักษรคฤนต์และอักษรเทวนาครีมีต้นกำเนิดมาจากอักษรเดียวกัน คือ อักษรพราหฺมิของพราหมณ์ ซึ่งอักษรนี้ได้

ดัดแปลงมาจากอักษรเฟนิเชีย ซึ่งสรุปได้ว่าอักษรไทยได้รับการถ่ายทอดลักษณะมาจากอักษรเฟนิเชีย ซึ่งเป็นอักษรที่เก่าแก่ และเป็นอักษรที่เป็นต้นกำเนิดอักษรหลายๆ ภาษาในทวีปยุโรปและเอเชีย แสดงดังรูปที่ 2.10 (สุธิวงศ์ พงศ์ไพบูลย์, 2532)



รูปที่ 2.10 ที่มาของอักษรไทย (สุธิวงศ์ พงศ์ไพบูลย์, 2532)

อักษรไทยในปัจจุบันแบ่งออกเป็น 4 ประเภท คือ พยัญชนะ สระ วรรณยุกต์ และเลข (กำชัย ทองหล่อ, 2550) ซึ่งสามารถอธิบายได้ดังนี้

2.2.1.1 พยัญชนะ คือ สัญลักษณ์หรืออักษรสำหรับประสมกับสระ เพื่อใช้สร้างคำหรือประโยคสำหรับจดบันทึกแทนภาษาพูดหรือบันทึกภาษาพูดไว้ไม่ให้สูญหาย พยัญชนะไทย มี 44 รูป คือ (กำชัย ทองหล่อ, 2550)

ก	ข	ฃ	ค	ฅ	ฉ	ง	จ
ฉ	ช	ฌ	ฌ	ญ	ฎ	ฏ	ฐ
ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ

น	บ	ป	ผ	ฝ	พ	ฟ	ภ
ม	ย	ร	ล	ว	ศ	ษ	ส
ห	ฬ	อ	ฮ				

โดยพยัญชนะไทย 44 รูป จำแนกออกเป็น 3 ประเภท คือ (กำชัย ทองหล่อ, 2550)

1) พยัญชนะกลาง คือ พยัญชนะสำหรับเขียนคำทั่ว ๆ ไป ทั้งคำไทย คำบาลี คำสันสกฤต และคำจากภาษาอื่น ๆ ที่ยืมมาใช้ในภาษาไทย เช่น เขมร อังกฤษ จีน ชาวประเภอบด้วยทั้งหมด 21 รูป คือ

ก	ข	ค	ง	จ	ฉ	ช
ต	ถ	ท	น	ป	ผ	พ
ม	ย	ร	ล	ว	ศ	ห

2) พยัญชนะเฉพาะ คือ พยัญชนะสำหรับเขียนในขอบเขตจำกัด เฉพาะคำที่มาจากภาษาบาลี สันสกฤต และภาษาทวีปยุโรปบางคำ ประกอบด้วยทั้งหมด 13 รูป คือ

ฌ	ฌ	ญ	ฎ	ฐ	ฑ	ฒ
ณ	ธ	ภ	ศ	ษ	ฬ	

3) พยัญชนะเดิม คือ พยัญชนะที่ภาษาไทยกำหนดเพิ่มขึ้นเพื่อให้เพียงพอแก่การเขียนสำเนียงในภาษาไทย เนื่องจากพยัญชนะที่มีสำหรับภาษาบาลีและสันสกฤต ไม่สามารถเขียนสำเนียงในภาษาไทยได้ครบทุกเสียง จึงจำเป็นต้องบัญญัติเพิ่มขึ้น ประกอบด้วยทั้งหมด 10 รูป คือ

ช ค ซ ฎ ค
บ ฝ ฟ อ ฮ

พยัญชนะไทย 44 รูป หากจำแนกตามวิธีใช้ มีดังนี้ คือ มาตราสะกด คำเป็นคำตาย ไตรยางค์ อักษรคู่ อักษรเดี่ยว อักษรควบ อักษรนำ ตัวหัน และตัวการ์นต์ สามารถแสดงรายละเอียดได้ดังต่อไปนี้ (กำชัย ทองหล่อ, 2550)

1) มาตราตัวสะกด คือ การจำแนกพยัญชนะท้ายคำหรือพยางค์ (ตัวสะกด) ตามการออกเสียงสำหรับสะกดในคำหรือพยางค์นั้น ๆ จำแนกเป็น 8 มาตรา แสดงดังตารางที่ 2.2

ตารางที่ 2.2 มาตราสะกด

มาตรา	พยัญชนะสะกด	ตัวอย่างคำ
กก	ก ข ค ฅ	นก เลข พรรค เมฆ
กบ	บ ป ฟ ภ ฝ	ลบ บาบ ศัพท์ ลาก กราฟ
กค	ค ต จ ซ ฎ ฏ ฐ ฒ ส ศ ษ ท ธ ช ฑ	มด จิต กิจ ประโยชน์ ฎฎ ปราบฎ อุฐ วุฒิ รหัส ประเทศ พิเศษ บาท โกรธ ก้าช ครุฑ
กม	ม	สะสม ออม คม ขม ลม สาม
กน	น ณ ญ ร ล ฬ	หิน คุณ บุญ กร นิล ทมิฬ
กง	ง	รณรงค์ ชง สำเนียง เสียง กลาง
เกย	ย	มหาลัย เนย เสวย วินัย ทราย เลย
เกอว	ว	ข้าว ดาว สาว ลิว แก้ว เกี้ยว

2) คำเป็น คือ พยางค์ที่ไม่มีตัวสะกดที่ประสมด้วยสระเสียงยาว (ทีฆสระ) พยางค์ที่ประสมด้วยสระ อำ ไอ โอ เอา และพยางค์ที่มีตัวสะกดในมาตรา กง กม กน เกย เกอว เช่น คุณย่ากินลำไย ตาของฉันทันมองไม่เห็น

3) คำตาย คือ พยางค์ที่ไม่มีตัวสะกดที่ประสมด้วยสระเสียงสั้น (รัสสระ) ยกเว้น อำ ไอ โอ เอา และพยางค์ที่มีตัวสะกดในมาตรา กก กค กข เช่น มดกัดซากกบ

4) ไตรยางค์ คือ การจำแนกพยัญชนะต้นออกเป็น 3 หมู่ตามพื้นฐานการออกเสียงโดยที่ไม่ได้มีการผันวรรณยุกต์ คือ อักษรสูง อักษรกลาง อักษรต่ำ โดยที่อักษรต่ำสามารถแบ่งออกเป็น 2 ประเภท คือ อักษรเสียงต่ำที่มีคู่เสียงอักษรเสียงสูง เรียกว่า “อักษรต่ำคู่” หรือ “อักษรคู่” และอักษรเสียงต่ำที่ไม่มีคู่เสียงกับอักษรเสียงสูง เรียกว่า “อักษรต่ำเดี่ยว” หรือ “อักษรเดี่ยว” แสดงดังตารางที่ 2.3

ตารางที่ 2.3 อักษรไตรยางค์

หมู่พยัญชนะ		พยัญชนะ	ตัวอย่างคำ
อักษรสูง		ข ข ฉ ฐ ถ ผ ฝ ศ ช ส ห	ฝั้ว ฝัก ถุง ข้าว สาร ให้ ฉันทัน
อักษรกลาง		ก จ ฉ ฎ ฏ ด ต บ ป อ	ไก่ จิก เด็ก ตาย บน ปาก โอง
อักษรต่ำ	อักษรคู่	ค ต ฉ ม ช ฌ ฉ ฆ ฒ ท ฑ พ ฟ ภ ฮ	พ่อ คำ ฟัน ทอง ชื้อ ช้าง ฮ่อ
	อักษรเดี่ยว	ง ญ ฌ น ษ ร วม พล	งู ใหญ่ นอน อยู่ ณ ริม วัค โม พี โลก

5) อักษรควบ คือ พยัญชนะ 2 อักษร ควบหรือกล้ำและใช้สระเดียวกัน ประกอบด้วย 4 พยัญชนะ คือ ร ล ว ห โดยในการควบกล้ำมี 2 รูปแบบ คือ 1) อักษรควบแท้ คือ พยัญชนะ ร ล ว ควบกล้ำโดยใช้สระตัวเดียวกันกับพยัญชนะที่ควบอยู่ เมื่อควบกล้ำแล้วจะต้องออกเสียงพร้อมกัน เป็นตัวสะกดหรือการันต์ ถือว่าเป็นอักษรควบแท้ เช่น ครู กลอง บุตรี นิทรา แพกตร์ ศาสตร์ และ 2) อักษรควบไม่แท้ คือ พยัญชนะ 2 อักษร ควบกล้ำกันโดยใช้สระเดียวกัน โดยเมื่อควบกล้ำกันแล้วจะออกเสียงเฉพาะตัวหน้า เช่น จริง ศรีทธา เสรี ออกเสียงแปรไปเป็นเสียงตัวอื่น

เช่น อินทรีย์ ทรวดทรง พุทรา หรือกลายเป็นตัวสะกดร่วมกัน เช่น ธรรม มรรค บรรณ พรหม
พราหมณ์ พรหมจารี

6) อักษรนำ คือ พยัญชนะ 2 รูป ใช้สระเดียวกัน ซึ่งจะออกเสียง
ร่วมกันสนิท เช่น หมี หน่อ หนี อยู่ อวยาก หรือออกเสียงคล้ายกับเป็น 2 พยางค์ เนื่องจากประสมกัน
ไม่สนิท เช่น ผนัง แผนก ผนวก จรัส ฉลาด ขนม

7) ตัวหัน คือ พยัญชนะที่แปลงรูปจากวิสรรชนีย์มาเป็นพยัญชนะที่
รูปเหมือนตัวสะกด แต่ทำหน้าที่แทนหันอากาศ และต้องเรียงไว้หน้าตัวสะกดที่มีรูปเหมือนกัน ซึ่ง
พยัญชนะทุกอักษรที่เป็นตัวสะกดสามารถเป็นตัวหันได้ทั้งหมด แต่ในปัจจุบันนิยมใช้ ร หันเท่านั้น
เช่น ธรรม กรรม สรร

8) ตัวการ์นต์ คือ พยัญชนะที่ไม่ออกเสียง โดยจะมีทัศนธาตุเป็น
เครื่องหมายกำกับบนพยัญชนะ โดยปกติการรับคำจากภาษาอื่นที่มีพยางค์มาก เช่น ภาษาบาลี
สันสกฤต และอังกฤษ มาใช้ในภาษาไทยที่มีลักษณะเป็นคำโดดที่มีพยางค์น้อย จึงต้องหาวิธีที่จะนำ
คำจากภาษาอื่นมาปรับใช้ให้เข้ากับภาษาไทย ด้วยวิธีการลดพยางค์ ทั้งนี้เพื่อให้คงรูปศัพท์เดิมไว้
ด้วยจึงเลือกใช้ตัวการ์นต์ในการลดพยางค์และคงรูปศัพท์ เช่น พักตร์ กาญจน์ สิงห์ อินทร์ พันธุ์
สรรพ ธรรม์ ลักษณะ

2.2.1.2 สระ คือ เครื่องหมายหรือสัญลักษณ์สำหรับแทนเสียงที่เปล่งออกมา
สามารถออกเสียงตามลำพังได้ ประกอบด้วยรูปสระที่ไม่ซ้ำกันทั้งหมดมี 21 รูป (กำชัย ทองหล่อ,
2550) แสดงดังตารางที่ 2.4

ตารางที่ 2.4 รูปสระ

สระ	ชื่อ	ตัวอย่างการใช้
ะ	วิสรรชนีย์	อะ เอะ แอะ โอะ เอาะ เอียะ เอือะ อัวะ
ั	หันอากาศหรือไม้ฝัด	อัวะ อัว
็	ไม้ไต่คู้	อึ
า	ลากข้าง	อา อ้า เอา เอาะ
ิ	พินทุ พินทุอิ	อิ อี อึ อือ เอียะ เอีย เอือะ เอือ
ุ	ฝนทอง	อู อู เอือะ เอีย
"	พินทุ	อึ เอือะ เอือ
อ	หยาดน้ำค้าง	อ้า อึ
ุ	ตีนเหยียด	อุ
ู	ตีนคู้	อู
เ	ไม้หน้า	เอ แอ เอะ แอะ เอาะ เอออะ เอียะ เอีย เอือะ เอือ เอา
ไ	ไม้ม้วน	ไอ
ใ	ไม้มลาย	ไอ
โ	ไม้โอ	โอ โอะ
ฤ	รี	กฤษณา
ฤา	รือ	ฤาษี
ฃ	ลี	คงฃก
ฃา	ลือ	ฃาชา
ย	ตัวยอ	เอียะ เอีย
ว	ตัววอ	อัวะ อัว
อ	ตัวออ	ออ อือ ออ เอออะ เอือ

และเพื่อให้เพียงพอต่อการใช้งาน จึงมีการประสมสระ ประกอบด้วยทั้งหมด 32 รูป (กำชัย ทองหล่อ, 2550) คือ

อะ	อา	อิ	อี	อึ	อื	อุ	อู
เอะ	เอ	แอะ	แเอ	โอะ	โ	เอาะ	ออ
เออะ	เออ	เอียะ	เอีย	เอือะ	เอือ	อ้าวะ	อ้าว
ฤ	ฤา	ฃ	ฃา	อำ	ไอ	ไอ	เอา

และสระทั้ง 32 รูปนี้สามารถจำแนกออกเป็น 4 ประเภท ตามการออกเสียงของสระ (กำชัย ทองหล่อ, 2550) แสดงดังตารางที่ 2.5

ตารางที่ 2.5 การจำแนกสระตามการออกเสียง

เสียงสั้น (รัสสระ)	เสียงยาว (ทิมสระ)	เสียงเบา (ลหุ)	เสียงหนัก (ครุ)
อะ	อา	อะ	อา
อิ	อี	อิ	อี
อึ	อื	อึ	อื
อุ	อู	อุ	อู
เอะ	เอ	เอะ	เอ
แอะ	แเอ	แอะ	แเอ
โอะ	โ	โอะ	โ
เออะ	ออ	เออะ	ออ
เออะ	เออ	เออะ	เออ
เอียะ	เอีย	เอียะ	เอีย
เอือะ	เอือ	เอือะ	เอือ
อ้าวะ	อ้าว	อ้าวะ	อ้าว
ฤ	ฤา	ฤ	ฤา

ตารางที่ 2.5 การจำแนกสระตามการออกเสียง (ต่อ)

เสียงสั้น (รัศสระ)	เสียงยาว (ทิมสระ)	เสียงเบา (ลหุ)	เสียงหนัก (ครุ)
ฤ	รฺ	ฤ	รฺ
อํ			อํ
โ			โ
ไ			ไ
เ			เ
			สระสั้น + ตัวสะกด

2.2.1.3 วรรณยุกต์ คือ ระดับเสียงสูงหรือต่ำของคำในภาษาไทยประกอบด้วย 5 เสียง คือ เสียงสามัญ เสียงเอก เสียงโท เสียงตรี และเสียงจัตวา แต่วรรณยุกต์มีเพียง 4 รูป ดังนี้

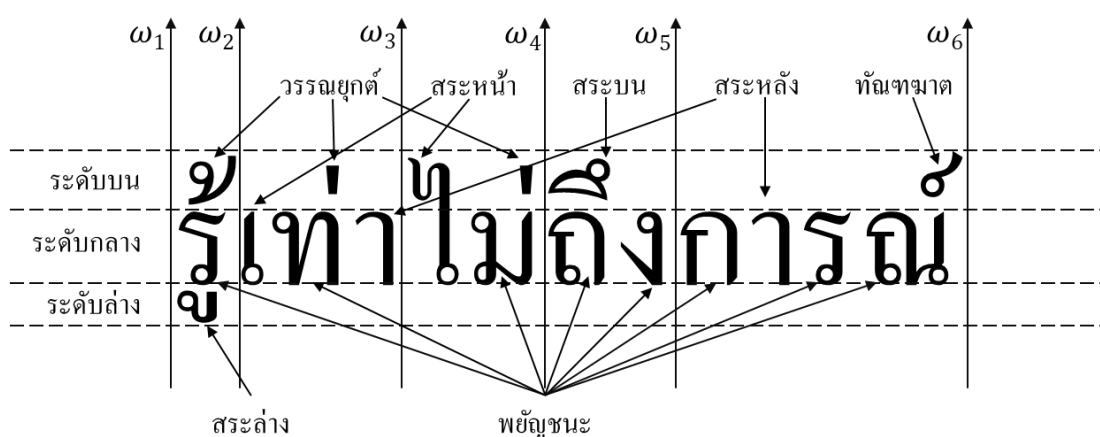
- ◌̄ เรียกว่า ไม้เอก
- ◌̄̄ เรียกว่า ไม้โท
- ◌̄̄̄ เรียกว่า ไม้ตรี
- ◌̄̄̄̄ เรียกว่า ไม้จัตวา

2.2.1.4 ตัวเลข มีทั้งหมด 10 รูป คือ

- | | |
|---------------------|---------------------|
| ๑ เรียกว่า เลขหนึ่ง | ๒ เรียกว่า เลขสอง |
| ๓ เรียกว่า เลขสาม | ๔ เรียกว่า เลขสี่ |
| ๕ เรียกว่า เลขห้า | ๖ เรียกว่า เลขหก |
| ๗ เรียกว่า เลขเจ็ด | ๘ เรียกว่า เลขแปด |
| ๙ เรียกว่า เลขเก้า | ๐ เรียกว่า เลขศูนย์ |

2.2.2 กลุ่มอักขระในภาษาไทย

กลุ่มอักขระ คือ การจำแนกประเภทอักขระตามวิธีการใช้งานของอักขระในระบบการเขียนของภาษาไทยเป็นหลัก ซึ่งอาจจะเล็กกว่าหรือเท่ากับหน่วยคำ อาจจะใหญ่กว่าหรือเท่ากับระดับหน่วยอักขระ อาจจะมีความหมายหรือไม่มีความหมาย อาจจะเล็กกว่าหรือเท่ากับหน่วยพยางค์ นั้นหมายความว่า การจำแนกในลักษณะนี้จะไม่คำนึงถึงสิ่งอื่นโดยเว้นหน้าที่การใช้งานของอักขระเท่านั้น เช่น ความหมาย การอ่านออกเสียง โดยประเภทของกลุ่มอักขระที่ได้จำแนกแสดงดังตารางที่ 2.6 (Tongtep and Theeramunkong, 2011) สำหรับงานวิจัยนี้ได้มีการปรับเปลี่ยนบางส่วนของตารางเพื่อให้ข้อมูลตรงกับทฤษฎีของภาษาไทยที่ใช้อ้างอิง คือ เรื่องตัวการ์นต์ เนื่องจากตัวการ์นต์เป็นพยัญชนะที่ไม่ออกเสียง ซึ่งประกอบด้วยกัน 2 ส่วน คือ พยัญชนะและเครื่องหมายทัณฑฆาต (สามารถดูรายละเอียดได้ที่ข้อ 2.6.1) จากรูปที่ 2.11 แสดงถึงตำแหน่งของอักขระและขอบเขตของคำในภาษาไทยของข้อความ “รู้เท่าไม่ถึงการณ์” ซึ่ง ω_1 คือ ขอบเขตของคำ และเมื่อนำอักขระมาจำแนกกลุ่มจะมีผลลัพธ์ ดังนี้ รู้เท่าไม่ถึง|การ|ณ์ ซึ่งเมื่อนำกลุ่มอักขระเหล่านี้มาแปลงให้อยู่ในรูปแบบของประเภทของกลุ่มอักขระแต่ละประเภทดังตารางที่ 2.6 ได้ดังนี้ T4T|T6T|T1T|T3|C|T2|C|TK



รูปที่ 2.11 ตัวอย่างข้อความภาษาไทย

ตารางที่ 2.6 ประเภทของกลุ่มอักขระไทย

ประเภท	ความหมาย	ตัวอย่าง	ประเภท	ความหมาย	ตัวอย่าง
ประเภทกลุ่มอักขระแบบทั่วไป					
S	ช่องว่าง	<ช่องว่าง>	V	สระ	แ ไ โ ะ
D	ตัวเลข	๐-๙, 0-1	P	เครื่องหมาย วรรคตอน	/'-๑๑,,:=!: ?; ; ()#@...
C	พยัญชนะไทย	ก-ฮ	T	วรรณยุกต์	่ ้ ๊ ๋
E	อักษรอังกฤษ	a-z, A-Z	O	อื่น ๆ	
ประเภทกลุ่มอักขระแบบซับซ้อน					
T1	C+VF (สระหน้า)	ไฟ ไกล เข แก โอ	T1T	T1+T	ไฟ ไข่ เข้ แก่ โอ้ ใจ ไกล
T2	C+VB (สระหลัง)	จะ คำ วา	T2T	T2+T	จะ ล้า ว่า
T3	C+VU (สระบน)	ก็ อี ตี อี	T3T	T3+T	ตี้ กี้ อี้
T4	C+VL (สระล่าง)	คู ดู	T4T	T4+T	คู้ ดู๋
TT	C+T	บ่ ลิ่น กิ่ง ตัง	TK	C+ทัณฑฆาต	อ๋
T6	C+VF+VB	และ โละ เรา	T6T	T6+T	โละ เต่า เสร์่า
T7	C+VF+VU	เต็ม เป็ย	T7T	T7+T	เต็ม เล็ยง
T8	C+VB+VU	บ๊วะ ผัวะ	T8T	T8+T	บ๊วะ อัวะ
T9	C+VF+VB+VU	เผ็ยะ เสือะ	T9T	T9+T	เผ็ยะ เสือะ

2.2.3 คำไทย

คำ คือ หน่วยพยางค์ตั้งแต่ 1 พยางค์ขึ้นไป ประกอบกันก่อให้เกิดความหมาย คำไทยสามารถจำแนกออกเป็น 7 ประเภท (กำชัย ทองหล่อ, 2550) คือ

2.2.3.1 คำนาม คือ คำที่เป็นชื่อคน สัตว์ สิ่งของ สถานที่ สิ่งมีชีวิตต่าง ๆ สิ่งที่ไม่มีชีวิต สิ่งที่เป็นรูปธรรมและนามธรรม คำนามจำแนกออกเป็น 5 ประเภท ดังนี้

1) सामानยนาม คือ คำนามที่เป็นชื่อทั่วไปไม่ชี้เฉพาะของคน สัตว์ สถานที่ สิ่งของ เช่น คน เด็ก มหาวิทยาลัย แมว ภูเขา บ้าน ดินสอ กระเป๋า วิชา ปัญญา

2) วิสามานยนาม คือ คำนามที่เป็นชื่อเฉพาะของคน สัตว์ สิ่งของ สถานที่ เช่น สมหมาย แดง คำ เอรಾವัน ดาวดิงส์ มหาวิทยาลัยสงขลานครินทร์

3) ลักษณะนาม คือ คำนามที่ทำหน้าที่ประกอบคำนามอื่น ๆ เพื่อแสดงรูปลักษณะ ขนาด หรือปริมาณ ของนามนั้น ๆ ให้ชัดเจนยิ่งขึ้น เช่น ปากกา 1 **ด้าม** ช้อน 2 **คัน** ช้างป่า 1 **ตัว** ลากซุง 3 **ท่อน**

4) สมุหนาม คือ คำนามที่บอกหมวดหมู่ของนามทั่วไปและนามเฉพาะ เพื่อบอกถึงลักษณะการรวมกันเป็นกลุ่ม หมู่ หรือพวก เช่น **โขลง** ช้างเดินผ่านทุ่ง **ฝูง** นกกำลังบินกลับรัง **มหาวิทยาลัย** ต้องการนักศึกษาที่มีคุณธรรม

5) อาการนาม คือ คำนามที่เกิดจากคำกริยาหรือคำวิเศษณ์ที่มีคำว่า “การ” หรือ “ความ” นำหน้า เช่น การวิ่ง การเดิน การนั่ง ความรู้ ความดี ความชั่ว

2.2.3.2 คำสรรพนาม คือ คำสำหรับแทนคำนามที่กล่าวมาแล้ว แต่ไม่ต้องการกล่าวคำนามนั้นซ้ำอีกครั้ง จำแนกออกเป็น 6 ประเภท ดังนี้

1) บุรุษสรรพนาม คือ คำสรรพนามสำหรับแทนบุคคลในการสนทนา สามารถจำแนกเป็น 3 ประเภท คือ แทนชื่อผู้พูด (บุรุษที่ 1) เช่น ฉัน ผม ข้าพเจ้า อีฉัน แทนชื่อผู้ที่กำลังสนทนา (บุรุษที่ 2) เช่น เธอ ท่าน คุณ ได้เท่า ฝ่าพระบาท แทนชื่อผู้ที่ถูกอ้างถึง (บุรุษที่ 3) เช่น เขา มัน ผู้ใด พระองค์

2) ประพันธสรรพนาม คือ คำสรรพนามสำหรับแทนคำนามหรือคำสรรพนามที่อยู่ติดกันด้านหน้า เช่น ผู้ที่ อัน ผู้ซึ่ง ดังตัวอย่าง เธอคือครู**ผู้**เสียสละ คน**ที่**เป็นครูต้องมีความอดทน เขา**ผู้**ซึ่งชอบการพนันจะต้องพินาศ

3) วิภาคสรรพนาม คือ คำสรรพนามสำหรับแทนคำนามหรือคำสรรพนามที่แยกออก หรือ รวมกลุ่มเข้ากัน ได้แก่ บ้าง ต่าง กัน ดังตัวอย่าง พนักงาน**ต่าง**ทำหน้าที่ของตน เพื่อนช่วยช่วยเหลือซึ่ง**กัน**และ**กัน** นักเรียนทั้งหลาย**บ้าง**ก็เล่น **บ้าง**ก็อ่านหนังสือ **บ้าง**ก็คุย

4) นิยมสรรพนาม คือ คำสรรพนามสำหรับแทนคำนามหรือข้อความที่กล่าวมาแล้วสำหรับความเฉพาะเจาะจงเพื่อความชัดเจน เช่น นี้ นี้ นั่น โน่น โน้น เช่น นี้ เช่น นั่น อย่าง นี้ ดังตัวอย่าง ดอกไม้ที่ นี้ สวยกว่าที่ โน่น นี้ พนักงานใหม่ นั่น พนักงานเก่า หนังสือเล่ม นี้ อ่านยากกว่าหนังสือเล่ม นั่น

5) อนิยมสรรพนาม คือ คำสรรพนามสำหรับแทนคำนามทั่วไป ไม่เฉพาะเจาะจง เช่น ใคร อะไร ไหน ผู้ใด ผู้อื่น ผู้หนึ่งผู้ใด ดังตัวอย่าง ใคร ๆ ก็ชอบมัน เขาทาน อะไร ก็ได้ ห้าม ผู้ใด เข้าก่อน ได้รับอนุญาต

6) ปรากฏสรรพนาม คือ คำสรรพนามสำหรับแทนคำนามที่มีเนื้อหาเป็นคำถาม เช่น ใคร อะไร ไหน ผู้ใด ดังตัวอย่าง ใคร อยู่ในห้อง ? อะไร อยู่ในกล่อง ? บ้านของคุณอยู่ที่ ไหน ?

2.2.3.3 คำกริยา คือ คำที่แสดงถึงอาการของคำนามและสรรพนาม หรือแสดงถึงการกระทำของประธาน จำแนกออกเป็น 5 ประเภท ดังนี้

1) สกรรมกริยา คือ คำกริยาที่ต้องมีกรรมมารับจึงจะมีความหมายสมบูรณ์ เช่น ทหาร ถือ ปืน พ่อค้า ขาย ของ คน ตัด ต้นไม้

2) อกรรมกริยา คือ คำกริยาที่มีความหมายสมบูรณ์ในตัว ถึงแม้ไม่มีกรรมมารับความหมายยังคงสมบูรณ์ เช่น ฉัน เดิน ฉัน นั่ง เขา ยืน เขา วิ่ง ในห้อง

3) วิกตรกริยา คือ คำกริยาที่ไม่สำเร็จความหมายในตัว และไม่สามารถใช้เป็นกริยาของประธานตามลำพังได้ จะต้องมีคำนาม คำสรรพนามหรือคำวิเศษณ์มาขยาย เช่น เป็น เหมือน คล้าย เสมือน เท่า คู่จ คือ ราวกับ ดังตัวอย่าง ผม เป็น ครู ต้นไม้ คือ สิ่งมีชีวิตชนิดหนึ่ง เขา คล้าย นั้น

4) กริยานุเคราะห์ คือ คำกริยาสำหรับช่วยให้คำกริยาอื่นแสดงความหมายออกมาเป็นกาล มาลา หรือวากจก ดังตัวอย่าง เขา กำลัง ทำงาน เขา คง มาที่นี่ ผู้ร้าย ถูก จับกุม

5) กริยาสภาวามาลา คือ คำกริยาที่ทำหน้าที่คล้ายกับคำนาม อาจเป็นประธาน เป็นกรรม หรือ เป็นบทขยายส่วนใดส่วนหนึ่งของประโยค เช่น นอนมีประโยชน์กว่าอิริยาบถอื่น (นอน เป็นประธานของกริยา มี) ฉันชอบดูภาพยนตร์ (ดู เป็นกรรมของกริยา ชอบ) เขาออมทรัพย์ไว้เพื่อซื้อรองเท้า (ซื้อ เป็นบทขยายของกริยา ออม)

2.2.3.4 คำวิเศษณ์ คือ คำที่ทำหน้าที่สำหรับประกอบคำนาม คำสรรพนาม คำกริยา และคำวิเศษณ์ด้วยกันให้ได้ความชัดเจนยิ่งขึ้น สามารถจำแนกออกเป็น 10 ประเภท ดังนี้

1) ลักษณะวิเศษณ์ คือ คำวิเศษณ์ที่ประกอบบอกลักษณะต่าง ๆ เช่น ชนิด อากาณ ขนาด สี กลิ่น รส ดังตัวอย่าง คนอ้วนเป็นเพื่อนกับคนผอม เขาวิ่งเร็ว

2) กาลวิเศษณ์ คือ คำวิเศษณ์สำหรับบอกกาลเวลา เช่น เช้า สาย บ่าย เย็น อดีต ปัจจุบัน อนาคต ดังตัวอย่าง เขาตื่นตอนเช้าเพื่อออกกำลังกาย

3) สถานที่วิเศษณ์ คือ คำวิเศษณ์สำหรับบอกสถานที่ ระยะเวลา เช่น บน ล่าง เหนือ ใต้ ไกล ใกล้ บก น้ำ ดังตัวอย่าง ผมอยู่ไกล คุณอยู่ใกล้

4) ปริมาณวิเศษณ์ คือ คำวิเศษณ์สำหรับบอกจำนวนนับและปริมาณ เช่น หนึ่ง สอง สาม มาก น้อย ทั้งหลาย จู ทั้งหมด คนละ บาง บ้าง ดังตัวอย่าง บางคนก็ทำบ้าง บ้าง

5) นิยมวิเศษณ์ คือ คำวิเศษณ์สำหรับการชี้เฉพาะหรือจำกัดลงไปว่าเป็นเช่นนั้น เช่น นี้ นั่น โนนี่ นี้นั้น โนนี่ ทั้งนี้ ทั้งนั้น อย่างนี้ อย่างนั้น ดังนี้ ดังนั้น แท้จริงเอง เฉพาะ ดังตัวอย่าง คนคนนี้พูดจริงทำจริงและมีใจมั่นคงแน่นอน

6) อนิยมวิเศษณ์ คือ คำวิเศษณ์สำหรับแสดงถึงความไม่ชี้เฉพาะหรือไม่ได้จำกัดลงไปว่าเป็นเช่นนั้น เช่น ทำไม ใด อะไร ไหน ไร ก็ ฉันทใด เช่นไร ดังตัวอย่าง เธอทำอะไรฉันก็กินได้ทั้งนั้น คนใดคนหนึ่งในวันนี้เป็นขโมย

7) ปฤจฉาวิเศษณ์ คือ คำวิเศษณ์สำหรับบอกเนื้อความเป็นคำถามหรือความสงสัย เช่น ใด ไร ก็ อะไร ฉันทใด เช่นไร ไหน ไหม อันใด อย่างไร เท่าไร ไย ดังตัวอย่าง อะไรอยู่ในตู้คุณรู้ไหม ? แมวตัวนั้นหายไปไหน ?

8) ประดิษฐานวิเศษณ์ คือ คำวิเศษณ์สำหรับบอกเสียงร้องเรียกและเสียงขานรับเพื่อแสดงความสละสลวยของภาษาและแสดงความเป็นกันเองระหว่างผู้พูด เช่น จำ จ๊ะ ขอรับ โว้ย ครับ ดังตัวอย่าง แม่จำ หนูหิวข้าว ได้ครับ ได้ขอรับ

9) ประดิษฐานวิเศษณ์ คือ คำวิเศษณ์สำหรับบอกความปฏิเสธหรือไม่ยอมรับ เช่น ไม่ ไม่ใช่ มิใช่ บ่ มิได้ ไม่ได้ ดังตัวอย่าง เขาไม่ยอมกินข้าว เพราะเขาไม่หิวข้าว วันนี้มาไม่ได้มาโรงเรียน ฉันไม่ใช่ขโมย

10) ประพันธวิเศษณ์ คือ คำวิเศษณ์สำหรับประกอบคำกริยาและคำวิเศษณ์ เพื่อเชื่อมประโยคให้มีข้อความเกี่ยวข้อง เช่น ที่ ซึ่ง อย่างที่ ว่า ให้ ว่า ที่ว่า ชนิดที่ คือ เพื่อ ดัง ตัวอย่าง มาจะเป็นนักเรียนที่มีคุณธรรม **ซึ่ง**เป็นสิ่งสำคัญของการเป็นนักเรียนดีเด่น

2.2.3.5 คำบุพบท คือ คำที่แสดงความสัมพันธ์ระหว่างคำหรือประโยค เพื่อให้ทราบว่าคำหรือประโยคที่ถัดจากคำบุพบทนั้นมีหน้าที่เกี่ยวข้องกับคำหรือประโยคที่อยู่ด้านหน้าอย่างไร เช่น บุตร**ของ**นางสมศรีขยันเรียนมาก ฉันท่าน**เพื่อ**หาเงิน คำบุพบทจำแนกออกเป็น 5 ประเภท ดังนี้

1) คำบุพบทสำหรับแสดงความสัมพันธ์ระหว่างนามและนาม นามและสรรพนาม หรือนามและกริยา เช่น ฉันทชอบนาฬิกา**ของ**เธอ โด๊ะ**ใน**ห้องสำนักงานมีไม้พ้อ เอกสาร**สำหรับ**จัดระดับความยากง่าย จำนวน 30 ฉบับ

2) คำบุพบทสำหรับแสดงความสัมพันธ์ระหว่างคำสรรพนามและนาม คำสรรพนามและคำสรรพนาม หรือคำสรรพนามและคำกริยา เช่น เธอคุยกับใคร**ใน**ห้องเรียน

3) คำบุพบทสำหรับแสดงความสัมพันธ์ระหว่างคำกริยาและคำนาม คำกริยาและคำสรรพนาม คำกริยาและคำกริยา หรือคำกริยาและคำวิเศษณ์ เช่น เธอมา**กับ**ฉัน เขาทำ**เพื่อ**หาผลประโยชน์ ผมนั่ง**ใกล้**เขา

4) คำบุพบทสำหรับแสดงความสัมพันธ์ระหว่างประโยคและประโยค เช่น เธอเก็บอาหารไว้**สำหรับ**คนเก็บขยะหน้าบ้าน เขาออกกำลังกายทุกวัน**หลังจาก**เขาไปหาหมอ

5) คำบุพบทสำหรับไม่แสดงความสัมพันธ์กับบทอื่น เช่น ดูกร ดูก่อน ดูณะ ดูรา ข้าแต่ ดังแสดงตัวอย่าง **ข้าแต่** ท่านทั้งหลาย **ดูก่อน** ฟังน้องทั้งหลาย

2.2.3.6 คำสันธาน คือ คำที่ทำหน้าที่เชื่อมระหว่างคำกับคำ ประโยคกับประโยค ข้อความกับข้อความ หรือเชื่อมความให้สละสลวย สามารถจำแนกออกเป็น 3 ประเภท ดังนี้

1) คำสันธานที่เชื่อมให้เป็นอนุกรมประโยค จำแนกออกเป็น 4 ประเภท คือ เชื่อมความที่คล้ายตามกัน เช่น และ ก็ กับ จึง เช่น ว่า ให้ คือ ทั้ง ครั้น...ก็ เมื่อ...ก็ พอ...ก็ ก็ดี ก็ได้ เชื่อมความขัดแย้งกัน เช่น แต่ แต่ว่า แต่ทว่า กว่า...ก็ ถึง...ก็ เชื่อมความเป็นเหตุเป็นผลกัน

เช่น เพราะ ด้วย จึง ฉะนั้น ฉะนั้น เหตุว่า ฉะนั้น...จึง เชื่อมความที่เลือกเอาอย่างใดอย่างหนึ่ง เช่น หรือ ไม่ก็ ไม่เช่นนั้น มิฉะนั้น

2) คำสันธานที่เชื่อมให้เป็นสังกรประโยค จำแนกออกเป็น 6 ประเภท คือ เชื่อมความแสดงลักษณะอาการ เช่น เขาพูดว่าเธอเป็นขโมย เขาให้ผมเขียนหนังสือ เชื่อมความแสดงประมาณ เช่น เธอพูดมากจนเพื่อนของเธอหลับ เชื่อมความแสดงเวลา เช่น เขาเก็บอาหารไว้จนหนึ่งปีผ่านไป เชื่อมความแสดงเหตุ เช่น เธอไม่สบายเพราะตากฝน เชื่อมความแสดงผลลัพธ์ เช่น ฝนตกจนน้ำท่วม เชื่อมความแสดงการเปรียบเทียบ เช่น เธอสวยราวกับนางฟ้า

3) คำสันธานที่เชื่อมให้เนื้อหาความเด่น จำแนกออกเป็น 2 ประเภท คือ เชื่อมความคนละตอนให้ประสานกัน เช่น นายแดงและนายดำเป็นเพื่อนกัน วันหนึ่งนายแดงยกของ ฝ้ายนายดำเห็นเช่นนั้นก็ไปช่วยนายแดงยกของ เชื่อมความให้สละสลวย เช่น อย่างไรก็ตาม ผมจะไปช่วยเธอให้ได้ อย่างไรก็ตาม ธรรมชาติย่อมชนะธรรมชาติ

2.2.3.7 คำอุทาน คือ คำสำหรับแสดงถึงเสียงที่เปล่งออกมาในเวลาจิตใจ เสียใจ ตกใจ หรือประหลาดใจ หรือเป็นคำที่ใช้ต่อถ้อยเสริมบทให้สมบูรณ์ยิ่งขึ้น สามารถจำแนกออกเป็น 2 ประเภท ดังนี้

1) คำอุทานบอกอาการสำหรับแสดงความรู้สึกต่าง ๆ ในการพูดและสำหรับเป็นคำขึ้นต้นคำประพันธ์ เช่น ชี ! ชี ๆ ! วะ ! วา ! แหม ! เอ๊ะ ! เฮ้ย ! เฮ้อ ! โอ๊ย ! อ้า ! โอ้ ! โอ้ว !

2) คำอุทานเสริมบท คือ คำอุทานเสริมบทสำหรับคำสร้อย สำหรับคำแทรก สำหรับคำเสริม เพื่อให้มีความครบถ้วนตามที่ต้องการ หรือให้มีความกระชับ หรือให้สละสลวยขึ้น จำแนกออกเป็น 3 ประเภท คือ อุทานเสริมบทสำหรับคำสร้อย เช่น โลกแล เขียนนา หน้อยรา นี้เอ๋ย อุทานเสริมบทสำหรับคำแทรก เช่น เวียนมาสิเวียนไป สนุกนิเราเศร้าสิ้น สัตว์อะไร เอ๋ย มีขา 4 ขา เด็กเอ๋ยเด็กน้อย อุทานเสริมบทสำหรับคำเสริม เช่น วัดวาอาราม ลูกเต้าเหล่าใคร พิธีรีตอง อายน้ำอาบท่า รดรา เข้าอกเข้าใจ ไม่รู้ไม่ซี หนังสือหน้าหา

2.2.4 วลีไทย

วลีหรือกลุ่มคำ คือ การนำคำอย่างน้อยสองคำมาเรียงต่อกัน ส่งผลให้เกิดความหมายเพิ่มขึ้น ซึ่งสามารถเข้าใจได้แต่ยังไม่สมบูรณ์ วลีจะใช้ตามลำพังไม่ได้ จะต้องใช้เป็น

ส่วนใดส่วนหนึ่งของประโยค คือ บทประธาน กริยา กรรม หรือบทขยายอื่น ๆ วลีสามารถจำแนกออกเป็น 7 ประเภท (กำชัย ทองหล่อ, 2550) ดังนี้

2.2.4.1 นามวลี คือ กลุ่มคำที่มีคำนามนำหน้า เช่น นาฬิกาข้อมือเรือนทอง อยู่บนโต๊ะ หนังสือหลักภาษาไทย อยู่ในห้องสมุด

2.2.4.2 สรรพนามวลี คือ กลุ่มคำที่มีคำสรรพนามนำหน้า เช่น กระผม นาย กล้า แกร่ง ศึก มีความรักชาติอย่างแรงกล้า ท่านอาจารย์ จะไปไหน ?

2.2.4.3 กริยวลี คือ กลุ่มคำที่มีคำกริยานำหน้า เช่น ผม กำลังพยายาม ทบทวน บทเรียน ต้อง อด สำห้ ปาก เพียร เรียน หนังสือ

2.2.4.4 วิเศษณ์วลี คือ กลุ่มคำที่มีคำวิเศษณ์นำหน้า เช่น เขาร้องเพลง ไพเราะ อะไร เช่นนี้ มันเป็นพื้นที่ที่ ใหญ่ โต โหด โหด มาก

2.2.4.5 บุพบทวลี คือ กลุ่มคำที่มีคำบุพบทนำหน้า เช่น จดหมาย ของ ฉัน ตั้ง อยู่ บน โต๊ะ ทำงาน

2.2.4.6 สันธานวลี คือ กลุ่มคำสำหรับใช้เป็นคำสันธาน จะมีคำสันธานนำหน้าหรือไม่ก็ได้ เช่น ถึง อย่างไร ก็ ตาม เขา เป็น เพื่อน ของ ฉัน เพราะ ฉะนั้น เขา จึง ต้อง ไป หา หมอ

2.2.4.7 อุทานวลี คือ กลุ่มคำสำหรับใช้เป็นคำอุทาน จะมีคำอุทานนำหน้าหรือไม่ก็ได้ เช่น เราควร เข้า อก เข้า ใจ ผู้อื่น เขาร้องขึ้นว่า โธ ตาย แล้ว !

2.2.5 ประโยคไทย

ประโยค คือ กลุ่มคำที่เรียงเรียงขึ้นเพื่อแสดงความคิดหรือเรื่องราวที่ครบถ้วนสมบูรณ์โดยปกติแล้วประโยคจะต้องมีบทประธานและบทกริยาเป็นหลักสำคัญ (กำชัย ทองหล่อ, 2550)

2.2.5.1 บท คือ คำ วลี หรือประโยค บทต่าง ๆ ที่เป็นองค์ประกอบของประโยคถูกจำแนกออกเป็น 7 ประเภท (กำชัย ทองหล่อ, 2550) ดังนี้

1) บทประธาน คือ บทที่ทำหน้าที่เป็นหัวหน้าของประโยค และคุมบทกริยาให้มีลักษณะเป็นประโยคต่าง ๆ โดยบทประธานจะใช้คำทั้ง 7 ประเภท วลี หรือ ประโยค ก็ได้ เช่น คน แต่ง หนังสือ เขา มี พี่ สาว สอง คน กิน มาก เกิน ไป เป็นภัยต่อร่างกาย

2) บทขยายประธาน คือ บทที่ทำหน้าที่สำหรับประกอบบทประธาน ให้ได้ความยิ่งขึ้น โดยบทขยายประธานจะใช้คำ วลี หรือ ประโยค ก็ได้ เช่น คน อ้วนมีเหงื่อมาก นักเรียนในห้องนี้เรียนเก่ง คนที่เป็นสมาชิกมีสิทธิพิเศษ

3) บทกริยา คือ บทสำหรับแสดงความเป็นไปของบทประธาน เพื่อให้ทราบว่าบทประธานเป็นผู้ นำ ผู้ถูกกระทำ หรือผู้ถูกใช้ จะเป็นคำหรือวลีก็ได้ เช่น เขาเป็นเด็กดี เขากำลังเรียนวิชาภาษาอังกฤษ

4) บทขยายกริยา คือ บทสำหรับประกอบบทกริยาให้ได้ความยิ่งขึ้น จะใช้คำวลี หรือประโยคก็ได้ เช่น เขาเป็นเด็ก เขารับประทานอาหารมากเกินไป เธอเป็นคนเรียนเก่งมาก

5) บทกรรม คือ บทที่ทำหน้าที่เป็นผู้ถูกกระทำ จะใช้คำ วลี หรือประโยคก็ได้ เช่น นายพรานยิงกระต่าย เขาต้องการเงิน ผมเห็นเด็กนักเรียนชาย ฉันเห็นรถยนต์คันบนถนน

6) บทขยายกรรม คือ บทสำหรับประกอบบทกรรมให้ได้ความยิ่งขึ้น จะใช้คำ วลี หรือประโยคก็ได้ เช่น เขาต้องการเด็กดี ผมเห็นเด็กนักเรียนชายคนหนึ่ง ฉันเห็นรถยนต์ที่เล่นบนถนนใหญ่

7) บทเชื่อม คือ คำหรือวลีที่ไม่ใช่บททั้ง 6 ประเภทที่กล่าวมาในข้างต้น แต่เป็นบทที่ทำหน้าที่เป็นบทเชื่อมระหว่างบทกับ ประโยคกับประโยค หรือทำหน้าที่เสริมความให้เด่นชัด สามารถจำแนกออกเป็น 4 ประเภท คือ ทำหน้าที่เป็นบทบุพบท เช่น เขาถูกข่มเหง ในเมื่อเขาตกอับ ทำหน้าที่เป็นบทสันธาน เช่น เขาเดินมาแต่ฉันเดินไป ทำหน้าที่เป็นบทอุทาน เช่น ไอ้โฮ! เครื่องบินลำนี้ใหญ่จริง ทำหน้าที่เป็นบทอาลปนะ (คำเรียกหรือทักทาย) เช่น ท่านทั้งหลาย ผมขอเชิญทุกท่านฟังทางนี้

2.2.5.2 ประเภทของประโยค

1) ประเภทของประโยค ที่จำแนกตามเนื้อความในประโยค มี 3 ประเภท (กำชัย ทองหล่อ, 2550) ดังนี้

1.1) เอกรรตประโยชน์ (ประโยชน์ความเดียว) คือ ประโยคที่มีใจความสำคัญเพียงแต่ใจความเดียวโดยมีความหมายครบถ้วนสมบูรณ์ เช่น โอโห ! ฝนตกหนักมาก นกตัวใหญ่บินอยู่บนท้องฟ้า นักศึกษากำลังไปห้องสมุด

1.2) อเนกรรตประโยชน์ (ประโยชน์ความรวม) คือ ประโยคที่มีใจความสำคัญอย่างน้อย 2 ใจความรวมกันและระหว่างประโยคจะต้องมีบทสันธานเชื่อม หรือละสันธานไว้ในฐานเข้าใจ เช่น ฉันเป็นพ่อค้าแต่ เขาเป็นเกษตรกร พ่อเป็นชาวน และแม่ก็เป็นชาวน เพราะเขายัน เขาจึงเป็นคนโปรดของหัวหน้า

1.3) สังกประโยชน์ (ประโยชน์ความซ้อน) คือ ประโยคที่ประกอบด้วยประโยคอย่างน้อย 2 ประโยคขึ้นไปรวมกัน แต่มีประโยคหลักที่มีใจความสำคัญเพียงประโยคเดียว ส่วนประโยคอื่นนอกเหนือจากนี้ทำหน้าที่เป็นประโยคประกอบให้กับประโยคหลักโดยมีการเชื่อมกันระหว่างประโยค เช่น เธอเป็นผู้หญิงที่เรียนเก่งมาก ฉันเห็นแมวที่คาบปลา แมวตายเพราะมันถูกรถทับ เขาเป็นคนดีที่สังคมต้องการ

2) ประเภทของประโยค ที่จำแนกตามเจตนาของผู้ส่งสาร มี 5 ประเภท (ราเนา พลอยขาว, 2549) ดังนี้

2.1) ประโยคบอกเล่า คือ ประโยคที่มีใจความบอกเล่าเรื่องราวทั่วไป โดยที่ไม่เป็นคำถาม ไม่เป็นปฏิเสธ ไม่เป็นคำสั่ง หรือคำร้อง เช่น เมื่อวานผมหุงข้าว ผมชอบการออกกำลังกาย

2.2) ประโยคคำถาม คือ ประโยคที่มีใจความเป็นคำถามและต้องการคำตอบ เช่น เมื่อวานใครหุงข้าว ? คุณกำลังทำอะไร ?

2.3) ประโยคปฏิเสธ คือ ประโยคที่มีใจความปฏิเสธหรือไม่ตอบรับ ซึ่งจะมีคำวิเศษณ์ประกอบด้วยเพื่อแสดงความปฏิเสธ เช่น ฉันไม่หุงข้าว ผมไม่ชอบออกกำลังกาย

2.4) ประโยคคำสั่ง คือ ประโยคที่ส่งสารเพื่อสั่งให้ผู้รับสารปฏิบัติตาม หรือมิให้ปฏิบัติตาม เช่น ผมบอกให้เธอไปหุงข้าว หมอให้เธอหมั่นออกกำลังกาย

2.5) ประโยคขอร้อง คือ ประโยคที่ส่งสารเพื่อขอร้องวิงวอนหรือชักชวนให้ผู้รับสารกระทำอย่างใดอย่างหนึ่ง เช่น คุณช่วยหุงข้าวให้ผมหน่อย กรุณาถอดรองเท้าก่อนเข้าห้อง

2.3 การยืมภาษา

ภาษาทุกภาษาย่อมมีคำจากภาษาอื่นเข้ามาปะปน เพื่อให้ภาษานั้นมีคำใช้เพิ่มมากขึ้น สะดวกในการสื่อสาร และทำให้ภาษานั้นเจริญงอกงามขึ้น ทั้งนี้ต้องเป็นไปตามธรรมชาติของมัน (สุริวงค์ พงศ์ไพบูลย์, 2532) หากเปรียบเทียบการยืมยืมคำจากภาษาอื่น ๆ มาใช้กับการรับประทานอาหารของมนุษย์ สามารถแบ่งได้เป็น 3 ระยะ ดังนี้ (สุริวงค์ พงศ์ไพบูลย์, 2532)

ระยะที่ 1 รับประทานเพื่อประทังชีวิตเพื่อความอยู่รอด คือ มีความจำเป็นต้องรับคำจากภาษาอื่น ๆ มาใช้ เพื่อความสะดวกในการสื่อสาร เช่น ภาษาไทยรับคำว่า ฟุตบอล ของภาษาอังกฤษมาใช้เพื่อทำให้สื่อสารได้สะดวกขึ้น เนื่องจากในภาษาไทยยังไม่มีคำใดที่รัดกุมเท่ากับคำต้นฉบับที่ยืมมา

ระยะที่ 2 รับประทานเพื่อเสริมสร้างความสมบูรณ์ คือ ยืมคำมาเพื่อช่วยให้ภาษานั้นมีความสละสลวย ไพเราะ เสนาะหูมากยิ่งขึ้น แม้ว่าในภาษานั้นจะมีคำนั้นอยู่แล้วก็ตาม เช่น ในภาษาไทยมีการยืมคำว่า บิคา มาใช้แทนคำว่า พ่อ ในบางสถานการณ์ที่แตกต่างกัน

ระยะที่ 3 รับประทานเกินควร คือ การรับคำจากภาษาอื่นมากไปจนเกินความจำเป็น ส่งผลให้ภาษานั้น ๆ เสื่อมเสีย

ดังที่กล่าวไปแล้วว่า ทุกภาษาย่อมมีคำจากภาษาอื่น ๆ เข้ามาปะปน และในภาษาไทยก็มีการยืมคำจากภาษาอื่น ๆ มาใช้เช่นกัน เช่น ภาษาบาลี สันสกฤต จีน เขมร อังกฤษ ซึ่งสาเหตุที่ทำให้คำจากภาษาอื่น ๆ เข้ามาปะปนในภาษาไทย มีดังนี้ (สุริวงค์ พงศ์ไพบูลย์, 2532)

- เจ้าของภาษาทั้งสองมีความสัมพันธ์ทางเชื้อชาติ สัญชาติ หรือที่อยู่อาศัยตามสภาพภูมิศาสตร์ เช่น ดินแดนใกล้ชิดกัน ตัวอย่างเช่น ภาคเหนือของประเทศไทยมีภูมิประเทศที่ติด

กับประเทศพม่าและลาว ภาคตะวันออกเฉียงเหนือของไทยมีภูมิประเทศติดกับประเทศเขมร หรือภาคใต้ของประเทศไทยมีภูมิประเทศติดกับประเทศมาเลเซีย

- ความสัมพันธ์ทางด้านประวัติศาสตร์ เช่น การอพยพถิ่นฐาน ที่ทำให้ภาษาของเจ้าของถิ่นเดิมตกค้างอยู่ หรือผู้โยกย้ายเข้ามาใหม่ นำภาษาใหม่เข้ามาใช้

- ความสัมพันธ์ทางการค้า เนื่องจากจะต้องมีการสื่อสารกันในระหว่างการค้าขาย ส่งผลให้เกิดการหยิบยืมภาษาซึ่งกันและกัน

- การนับถือศาสนา ในการเรียนการสอนเรื่องของศาสนาจำเป็นต้องใช้ภาษาและคำราชา ส่งผลให้ผู้เรียนรับเอาภาษาที่ผู้สอนใช้สอนในเรื่องศาสนาไปด้วย

- การนำเครื่องมือของใช้ ตลอดจนความรู้ทางวิชาการเฉพาะทางของชนชาติที่เจริญแล้วเข้ามา

- ความสัมพันธ์ในเรื่องการศึกษา เช่น การศึกษาต่างประเทศ การศึกษาภาษานั้น ๆ โดยตรง เพื่อแลกเปลี่ยนเรียนรู้ระหว่างภาษาซึ่งกันและกัน

โดยทั่วไปคำในภาษาไทยที่ยืมจากภาษาเขมรมักเป็นคำราชาศัพท์และคำทั่ว ๆ ไป คำที่ยืมจากภาษาจีนมักใช้เกี่ยวกับชื่อวัตถุ สิ่งของ หรืออาหาร เช่น เต้าหู้ บะหมี่ ก๋วยเตี๋ยว ปุ้งกี๋ อั้งโล่ ฯลฯ คำที่ยืมจากภาษาอังกฤษมักใช้เกี่ยวกับชื่อวิชาการ ชื่อวัตถุเครื่องใช้ และชื่อยารักษาโรคต่าง ๆ คำที่ยืมจากภาษาบาลีและสันสกฤตมักเป็นคำศัพท์ทั่วไปหรือคำที่เกี่ยวกับชื่อคน ชื่อสถานที่ ชื่อศิลปวิทยาการต่าง ๆ และคำที่เกี่ยวกับลัทธิศาสนา (กำชัย ทองหล่อ, 2550) และนอกจากความสัมพันธ์ทางการยืมคำแล้ว ภาษาบาลีและสันสกฤตยังมีความสัมพันธ์ทางด้านระบบการเขียนของทั้ง 3 ภาษา คือ อักษรไทยซึ่งใช้สำหรับเขียนภาษาไทย บาลี และสันสกฤต และ อักษรเทวนาครีซึ่งใช้สำหรับเขียนภาษาบาลีและสันสกฤต มีระบบการเขียนที่เหมือนกัน คือ Syllabic alphabets (สามารถดูรายละเอียดที่ข้อ 2.1.6) อีกทั้งอักษรไทยยังได้รับการสืบทอดลักษณะอักษรมาจากอักษรเทวนาครีอีกด้วยแสดงดังรูปที่ 2.10 ด้วยเหตุนี้ ภาษาบาลี และสันสกฤตจึงมีความสัมพันธ์ร่วมกับภาษาไทยยิ่งขึ้น ส่งผลให้มีการหยิบยืมภาษาได้สะดวกมากยิ่งขึ้น

2.4 ภาษาสันสกฤต

สันสกฤต หมายถึง ขัดเกลามาแล้ว ภาษาสันสกฤตถือกำเนิดมาจากภาษาพระเวท ในช่วง 1,500 ปีก่อนคริสตกาล เนื่องด้วยมีกฎเกณฑ์รัดกุมตายตัวไม่สามารถเปลี่ยนแปลงได้ ภาษาสันสกฤตจึงกลายเป็นภาษาของชนชั้นสูงสำหรับผู้มีการศึกษา และเป็นภาษาวรรณคดี ซึ่งไม่ใช่ภาษาในชีวิตประจำวัน ในช่วงประมาณยุคกลางของอินเดีย ภาษาสันสกฤตถูกนำมาใช้เขียนคัมภีร์พระเวทเป็นลายลักษณ์อักษร ภาษาสันสกฤตจึงกลายเป็นภาษาศักดิ์สิทธิ์และภาษาศาสนา ส่งผลให้มีผู้ศึกษาและผู้รู้ภาษาลดน้อยลงยิ่งขึ้น จึงกลายเป็นภาษาตายในที่สุด ซึ่งไม่ได้ใช้สำหรับพูดหรือเขียนอีก (สุภาพร มากแจ้ง, 2535) แต่ปัจจุบันภาษาสันสกฤตส่วนใหญ่ถูกใช้สำหรับพิธีกรรมทางศาสนาฮินดู คือ เป็นภาษาสำหรับบทสวดมนต์ นอกจากนี้ยังมีบางหมู่บ้านในรัฐकर्णाฏกะ (Karnataka) ของประเทศอินเดียที่พยายามฟื้นฟูภาษาสันสกฤตให้กลายเป็นภาษาพูดในชีวิตประจำวัน (Ager, 2017) และในขณะเดียวกันไทยเองยังคงศึกษาและใช้งานภาษาสันสกฤตเช่นเดียวกัน

2.5 ภาษาบาลี

ภาษาบาลี คือ ภาษาปรากฤต ซึ่งภาษาปรากฤตเป็นภาษาที่วิวัฒนาการมาจากภาษาพระเวทในยุคกลางของอินเดีย ภาษาปรากฤตนี้แบ่งออกเป็น 3 สมัย คือ ภาษาปรากฤตสมัยเก่าหรือภาษาบาลี ภาษาปรากฤตสมัยกลาง ภาษาอปภรมุส ซึ่งนี่เป็นระยะสุดท้ายของภาษาในยุคกลางของอินเดีย ภาษาบาลีถูกนำมาใช้สำหรับบันทึกพุทธวจนะเป็นลายลักษณ์อักษร เมื่อประมาณพุทธศตวรรษที่ 3 ปรากฏหลักฐานครั้งแรกในจารึกพระเจ้าอโศกมหาราช ต่อมาภาษาบาลีก็มีสภาพเฉกเช่นภาษาสันสกฤต กลายเป็นภาษาที่ไม่มีการเจริญเติบโต ไม่มีการใช้สำหรับพูดหรือเขียนในชีวิตประจำวัน เป็นภาษาที่สร้างขึ้นด้วยวัตถุประสงค์เพียงสิ่งเดียว คือ ใช้สำหรับบันทึกพุทธวจนะ

(สุภาพร มากแจ้ง, 2535) ปัจจุบันยังคงมีบางพื้นที่ที่ศึกษาภาษาบาลี เช่น ประเทศพม่า ศรีลังกา ไทย ลาว และกัมพูชา (Ager, 2017)

2.6 หลักการสังเกตคำยืมในภาษาไทย

โดยธรรมชาติของคำที่ถูกยืมมามักจะถูกเปลี่ยนรูปคำ เสียง และความหมายใน ภาษาใหม่ เพื่อความสะดวกในการออกเสียง และเป็นไปตามลักษณะสำคัญของภาษาของผู้ยืม (สุทธิ วงศ์ พงศ์ไพบุลย์, 2532) ส่งผลให้คำยืมเหล่านั้นมีร่องรอยบ่งบอกว่าเป็นคำที่หยิบยืมมา แต่ในบาง กรณีก็เป็นการลบร่องรอยจากภาษาเดิมด้วยเช่นกัน อย่างไรก็ตามคำที่ถูกยืมมาบางคำมักจะมี ร่องรอยของภาษาเดิมอยู่ ซึ่งเป็นรูปแบบของภาษาเดิมนั้น โดยเฉพาะ ดังนั้นสามารถที่จะใช้ร่องรอย เหล่านั้นสืบค้นได้ว่าคำนั้น ๆ มีที่มาจากภาษาใด หลักการสังเกตเบื้องต้นเป็นการศึกษาที่มาที่ไปของ อักษรของคำยืม

2.6.1 หลักการสังเกตคำยืมทั่วไป

เป็นที่ทราบกันว่าเครื่องหมายทัศนมาตร คือ เครื่องหมายที่ใช้บังคับห้ามออกเสียง บนตัวพยัญชนะ ซึ่งเรียกว่า ตัวการันต์ การใช้ตัวการันต์เป็นวิธีที่ใช้ลดพยางค์ของคำที่ยืมมา เช่น ภาษาบาลี สันสกฤต อังกฤษ (กำชัย ทองหล่อ, 2550) แม้กระทั่งคำทับศัพท์บางคำจำเป็นต้องใช้ การันต์ด้วยเช่นกัน เช่น เปอร์เซนต์ (Percent) เซลล์ (Cell) อิเล็กทรอนิกส์ (Electronics) การ์ตูน (Cartoon) ตัวอย่างคำยืม เช่น ภาพยนตร์ วงศ์ ทูทซ์ สุทธิ โดยทั่วไปคำไทยจะไม่มีตัวการันต์ เพราะ เป็นภาษาคำโดดจึงมีพยางค์น้อย ดังนั้นหลักการสังเกตเบื้องต้นของคำที่ถูกหยิบยืมมา สามารถ สังเกตได้จากการมีหรือไม่มีของตัวการันต์ได้

2.6.2 หลักการสังเกตคำยืมจากภาษาบาลี

เนื่องจากภาษาบาลีเป็นภาษาที่ไม่มีอักษรเป็นของตนเอง (สุภาพร มากแจ้ง, 2535) หากภาษาใดที่ยืมคำบาลีมาใช้งานจะต้องใช้พยัญชนะของภาษานั้นในการเขียนคำบาลีขึ้น ดังนั้นคำ

ตารางที่ 2.7 พยัญชนะและสระที่ใช้เขียนภาษาไทย บาลี และสันสกฤต (ต่อ)

ภาษา	สระ				พยัญชนะ					จำนวน (รูป)
สันสกฤต	อะ	อา	อิ	อี	ก	ข	ค	ฌ	ง	สระ = 14
	อุ	อู	เอ	โอ	จ	ฉ	ช	ฌ	ญ	พยัญชนะ = 35
	ฤ	ฤ๑	ฌ	ฌ๑	ฎ	ฐ	ฑ	ฒ	ณ	รวม = 49
	โ	เอา			ด	ถ	ท	ธ	น	
					ป	ผ	พ	ภ	ม	
					ย	ร	ล	ว	ศ	
					ศ	ษ	ห	ฬ	อ	

2.6.2.1 การสังเกตการใช้พยัญชนะและสระเบื้องต้น ซึ่งหลักเกณฑ์ในการสังเกตการใช้พยัญชนะและสระของภาษาบาลีในภาษาไทยมีดังนี้

- 1) คำที่มีสระ โ และพยัญชนะตัว ย สะกด เช่น เวไนย ไวยากรณ์
ภาลินย ไททาน สาลัย
- 2) คำที่ใช้พยัญชนะตัว พ เช่น จุพา เวฬุริย กิฬา กักขพะ เขพะ
- 3) คำที่มีการตัดพยัญชนะตัวสะกด เมื่อพยัญชนะตัวสะกดและ
พยัญชนะตัวตามเป็นพยัญชนะเดียวกัน เช่น

<u>ภาษาบาลี</u>	<u>ภาษาไทย</u>
จิตต	จิต
ทิจฺฐิ	ทิจู
วฺจฺฉิ	วจุฉิ
อจฺฉ	อจฉ
รฺจฺฉ	รจฉ
เขตต	เขต

4) คำที่มีพยัญชนะ ห เป็นพยัญชนะตามหลังพยัญชนะสะกด เช่น
อุณหภูมิ ตัณหา กัณหา ปัญหา สายัณห์

ตารางที่ 2.8 วรรคและแถวของพยัญชนะสำหรับภาษาบาลี

วรรค	แถวที่ 1	แถวที่ 2	แถวที่ 3	แถวที่ 4	แถวที่ 5	เศษวรรค
วรรค กะ	ก	ข	ค	ฌ	ง	ย ร ล ว ส ห พ
วรรค จะ	จ	ฉ	ช	ฌ	ญ	
วรรค ฎะ	ฎ	ฐ	ฑ	ฒ	ณ	
วรรค ตะ	ต	ถ	ท	ธ	น	
วรรค ปะ	ป	ผ	พ	ภ	ม	

2.6.2.2 การสังเกตการใช้พยัญชนะและสระตามทฤษฎีพยัญชนะสะกดและพยัญชนะตาม จากตารางที่ 2.8 แสดงถึงพยัญชนะไทยที่ใช้เขียนคำบาลีที่ลูกแบ่งออกเป็น 6 วรรค และ 5 แถว ซึ่งสามารถอธิบายทฤษฎีพยัญชนะสะกดและพยัญชนะตามได้ ดังนี้

1) เมื่อพยัญชนะในแถวที่ 1 ของแต่ละวรรคเป็นตัวสะกด พยัญชนะในแถวที่ 1 และ 2 ต้องเป็นพยัญชนะตามหลังพยัญชนะสะกดในวรรคนั้น ๆ เช่น คำว่า ทุกข์ พยัญชนะ ก ทำหน้าที่เป็นตัวสะกด และพยัญชนะ ข ทำหน้าที่เป็นตัวตาม

2) เมื่อพยัญชนะในแถวที่ 3 ของแต่ละวรรคเป็นตัวสะกด พยัญชนะในแถวที่ 3 และ 4 ต้องเป็นพยัญชนะตามหลังพยัญชนะสะกดในวรรคนั้น ๆ เช่น คำว่า อัคคี พยัญชนะ ค ทำหน้าที่เป็นตัวสะกด และพยัญชนะ ค ทำหน้าที่เป็นตัวตาม

3) เมื่อพยัญชนะในแถวที่ 5 ของแต่ละวรรคเป็นตัวสะกด พยัญชนะในแถวที่ 1 ถึง 5 ต้องเป็นพยัญชนะตามหลังพยัญชนะสะกดในวรรคนั้น ๆ ยกเว้น พยัญชนะ ง จะ

เป็นตัวตามของพยัญชนะ ง เองไม่ได้ เช่น คำว่า คงคา พยัญชนะ ง ทำหน้าที่เป็นตัวสะกด และ พยัญชนะ ค ทำหน้าที่เป็นตัวตาม

4) สำหรับพยัญชนะในเศววรรค เมื่อทำหน้าที่เป็นตัวกล้ำจะมีหลักเกณฑ์ดังนี้

- 4.1) พยัญชนะ ย สามารถกล้ำกับพยัญชนะของแถวที่ 1 ถึง 5
- 4.2) พยัญชนะ ร สามารถกล้ำกับพยัญชนะของแถวที่ 1 และ 3
- 4.3) พยัญชนะ ล สามารถกล้ำกับพยัญชนะของแถวที่ 1
- 4.4) พยัญชนะ ว สามารถกล้ำกับพยัญชนะของแถวที่ 1 ถึง 5
- 4.5) พยัญชนะ ห สามารถกล้ำกับพยัญชนะของแถวที่ 5

2.6.3 หลักการสังเกตคำยืมจากภาษาสันสกฤต

ภาษาสันสกฤตเป็นภาษาหนึ่งที่ไม่มีอักษรเป็นของตนเอง (สุภาพร มากแจ้ง, 2535) เช่นเดียวกับภาษาบาลี ดังนั้นหลักการในการสังเกตคำสันสกฤตจะคล้ายกับหลักการสังเกตคำบาลี ในภาษาไทย ซึ่งจะแตกต่างกันตรงที่การใช้พยัญชนะและสระบางตัวที่เพิ่มเข้ามา คือ ศ ษ ฎ ฏ ฐ ฑ ฒ ไอ และ เอา ดังที่แสดงในตารางที่ 2.7 ซึ่งในภาษาสันสกฤตเองก็มีพยัญชนะสะกดและพยัญชนะตาม แต่มีรูปแบบไม่เป็นที่ชัดเจนเช่นเดียวกับภาษาบาลี

2.6.3.1 การสังเกตการใช้พยัญชนะและสระเบื้องต้น ซึ่งหลักเกณฑ์ในการสังเกตการใช้พยัญชนะและสระของภาษาสันสกฤตในภาษาไทยมีดังนี้

- 1) คำที่มีสระ ฤ ไอ และ เอา เช่น ฤษี ฤค ฤชฌา ไกลาส ไอศวรรย์ เอารส เมาลี
- 2) คำที่ใช้สระ แอ แล้วมีพยัญชนะสะกด เนื่องจากคำเดิมใช้สระ ไอ แล้วมีพยัญชนะสะกด ซึ่งภาษาไทยไม่สามารถอ่านเสียงในลักษณะนี้ได้ จึงเปลี่ยนมาใช้สระ แอ แทนสระ ไอ เช่น

<u>ภาษาสันสกฤต</u>	<u>ภาษาไทย</u>
ไวศย	แพศย์
ไวทย	แพตย์
ไวทย	แพทย์
ไวศฺร	แจศร
ไวสนยา	แสนยา

3) คำที่ใช้พยัญชนะ ศ ษ เช่น ลักษณะ ศีล สัน โดย ทฤษฏี ศรี ศาลา

ภาษา

4) คำที่ใช้พยัญชนะ รร เนื่องจากคำเดิมเป็น ร แต่ภาษาไทยนำมา

แปลงเป็น รร เช่น

<u>ภาษาสันสกฤต</u>	<u>ภาษาไทย</u>
ภารฺยา	ภรรยา
ครฺรภ	ครรรภ์
อาศฺจฺรฺย	อัศจรรย์
วฺรฺณ	วรรณ
ทฺรฺม	ทรรม

5) คำที่ใช้พยัญชนะ ร เป็นตัวสะกด เช่น

<u>ภาษาสันสกฤต</u>	<u>ภาษาไทย</u>
อาจฺรฺย	อาจารย์
สุรฺย	สุรย์

6) คำที่ใช้พยัญชนะ ร เป็นตัวควบ เช่น

<u>ภาษาสันสกฤต</u>	<u>ภาษาไทย</u>
จनुทฺรา	จันทรา
มหินฺทฺร	มหินทร์
วกฺตฺร	พกตฺร
ปฺตฺร	บุตร
มิตฺร	มิตร

7) คำที่มีพยัญชนะซ้อนต้นคำ เช่น

<u>ภาษาสันสกฤต</u>	<u>ภาษาไทย</u>
เกษฺตฺร	เกษตร
กษฺย	กษัย
สฺตฺรี	สตรี
กษฺตฺริย	กษัตริย์
สฺวสฺติ	สวัสดี

2.7 การประเมินระดับความยากง่ายของข้อความ

การประเมินระดับความยากง่ายของข้อความในเอกสาร หมายถึง การวัดระดับความยากหรือง่ายของเอกสาร โดยใช้คุณสมบัติทางข้อความเป็นตัวบ่งชี้ถึงระดับความยากง่าย ซึ่งระดับความยากง่ายของเอกสาร หมายถึง ความยากง่ายในการอ่านและความเข้าใจในเนื้อหาของเอกสาร (Patteera, 2009) ระดับความยากของเนื้อหาประกอบด้วยหลายปัจจัย เช่น องค์ประกอบของเนื้อหา ขนาดตัวอักษร รูปแบบการจัดวางองค์ประกอบของเนื้อหา สีของเนื้อหา ความรู้ภูมิหลังของผู้อ่าน วรรคตอนของข้อความ ระดับความยากง่ายของเนื้อหาข้อความสิ่งเหล่านี้จะส่งผลต่อการอ่าน

และระดับความเข้าใจในเนื้อหา ซึ่งปัจจัยที่มีอิทธิพลต่อความยากง่ายในการอ่านและเข้าใจของเนื้อหามากที่สุด ซึ่งสามารถวิเคราะห์ได้โดยง่าย คือ ปัจจัยในระดับความยากง่ายของข้อความในเอกสาร ดังนั้นงานวิจัยที่ผ่านมาได้มุ่งเน้นการประเมินระดับความยากง่ายของข้อความในเอกสารเป็นส่วนใหญ่ (Collins-Thompson, 2014) และคุณสมบัติสำคัญของข้อความสำหรับประเมินความยากง่าย มีดังนี้ 1) คุณสมบัติระดับคำ (Lexical features) เช่น คำที่ไม่คุ้นเคย คำศัพท์เทคนิค คำกำกวม 2) คุณสมบัติระดับไวยากรณ์ (Syntactic features) เช่น ความซับซ้อนของโครงสร้างประโยค และ 3) คุณสมบัติระดับบริบท (Discourse features) เช่น การวิเคราะห์ความสัมพันธ์ระหว่างประโยค การวิเคราะห์ทิศทางของข้อความ (Zhang, *et al.*, 2013)

ความยากง่ายของข้อความที่เขียนขึ้นมานั้น สามารถประเมินได้โดยการใช้สมการทางคณิตศาสตร์ โดยปกติแล้วการประเมินระดับความยากง่ายของข้อความนั้น จะพิจารณาจากคุณสมบัติเฉพาะ เช่น จำนวนของคำในประโยค จำนวนอักขระในคำ หรือแม้กระทั่งจำนวนพยางค์ของคำนั้น ๆ ทฤษฎีเกี่ยวกับการคำนวณหรือสมการของการประเมินระดับความยากง่ายของข้อความได้มีการศึกษามาตั้งแต่ปี ค.ศ. 1920 (Kitson, 1921) ซึ่งสมการการประเมินมีจำนวนมาก เช่น Flesch Reading Ease (Flesch, 1948) Flesch-Kincaid Grade Level (Kincaid, *et al.*, 1975) Coleman-Liau Index (CLI) (Coleman and Liau, 1975) Automated Readability Index (ARI) (Senter and Smith, 1967) ส่วนสมการทางคณิตศาสตร์สำหรับการประเมินระดับความยากง่ายของข้อความที่เป็นที่นิยมใช้ในข้อความภาษาอังกฤษและภาษาอื่น ๆ ที่น่าสนใจมี 3 สมการ คือ 1) Flesch-Kincaid Grade Level 2) CLI และ 3) ARI โดยสมการ Flesch-Kincaid Grade Level ถูกพัฒนามาจากสมการ Flesch Reading Ease เพื่อให้เหมาะสมสำหรับการใช้งานในกองทัพเรือของสหรัฐอเมริกา (Dave and Ruth, 2017) ซึ่งจะประเมินระดับความยากง่ายเป็นระดับเกรด (Grade level) โดยการคำนวณวัดระดับความยากง่ายทำได้โดยการหาค่าเฉลี่ยของคำต่อประโยคและค่าเฉลี่ยของพยางค์ต่อคำ แสดงดังสมการที่ (2.1)

$$RG_{FK} = 0.39 \left(\frac{\text{Total words}}{\text{Total sentences}} \right) + 11.8 \left(\frac{\text{Total syllables}}{\text{Total words}} \right) - 15.59 \quad (2.1)$$

โดย Total words คือ จำนวนคำทั้งหมด

Total sentences คือ จำนวนประโยคทั้งหมด

Total syllables คือ จำนวนพยางค์ทั้งหมด

สมการที่ (2.1) ถูกนำไปประยุกต์ใช้เป็นคุณสมบัติของโปรแกรมประมวลผลคำ (Word processing software) เช่น Microsoft Word (Mitchell, 1987) ในขณะที่สมการ CLI จะแบ่งระดับความยากง่ายของข้อความเป็นระดับเกรด โดยใช้ค่าความยาวเฉลี่ยของคำและประโยค แสดงดังสมการที่ (2.2)

$$CLI = 0.0588L - 0.296S - 15.8 \quad (2.2)$$

โดย L คือ ค่าเฉลี่ยของจำนวนอักขระต่อจำนวน 100 คำ

S คือ ค่าเฉลี่ยของจำนวนประโยคต่อจำนวนคำ 100 คำ

ส่วนสมการ ARI จะให้ค่าประมาณการของระดับความยากหรือง่ายของข้อความ โดยคำนวณจากจำนวนอักขระต่อจำนวนคำหนึ่งคำ และจำนวนคำต่อจำนวนประโยคหนึ่งประโยค โดยผลลัพธ์ที่ได้จะแบ่งเป็นระดับอายุของผู้อ่านที่ควรที่จะเข้าใจในการอ่าน แสดงดังสมการที่ (2.3)

$$ARI = 4.71 \left(\frac{C}{W} \right) + 0.5 \left(\frac{W}{S} \right) - 21.43 \quad (2.3)$$

โดย C คือ ผลรวมของจำนวนอักขระ ตัวเลข และเครื่องหมายวรรคตอน

W คือ จำนวนของช่องว่าง

S คือ จำนวนของประโยค

เทคนิค CLI และ ARI นี้มีข้อดี คือ การนับจำนวนอักขระ ค่อนข้างถูกต้องและสะดวกกว่าการนับจำนวนพยางค์ที่มีใช้ในบางสมการของการประเมินระดับความยากง่ายของข้อความ เช่น Flesch-Kincaid Grade Level โดยสูตรคำนวณแบบ CLI และ ARI นี้ สามารถนำมาเป็นเกณฑ์พื้นฐานในการประเมินระดับความยากหรือง่ายของข้อความในภาษาแบบ Syllabic Alphabets ได้ (ซึ่งภาษาไทยก็จัดอยู่ในกลุ่มภาษาดังกล่าว) อย่างไรก็ตามสมการส่วนใหญ่ที่ได้อ้างอิงไว้ได้ปรับ

ให้เหมาะสมสำหรับข้อความภาษาอังกฤษเป็นสำคัญ แต่ไม่ได้มีการปรับค่าให้เหมาะสมสำหรับภาษาอื่น รวมถึงภาษาไทยในการนำไปใช้งาน (Daowadung and Chen, 2011) ส่งผลให้ค่าความถูกต้องค่อนข้างน้อยเมื่อเปรียบเทียบกับเทคนิคที่สร้างเพื่อประเมินความยากง่ายสำหรับภาษาไทย โดยเฉพาะ (Tongtep, *et al.*, 2014)

2.8 การประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย

ในปัจจุบัน ได้มีการวิจัยที่เกี่ยวข้องกับการประเมินระดับความยากง่ายในภาษาต่าง ๆ เพิ่มมากขึ้น เช่น ภาษาอังกฤษ (Collins-Thompson and Callan, 2004; Collins-Thompson and Callan, 2005; Kidwell, *et al.*, 2011; Graesser, *et al.*, 2004 และ Pitler and Nenkova, 2008) ภาษาฝรั่งเศส (Todirascu, *et al.*, 2013; Dascalu, 2014 และ François and Fairon, 2012) ภาษาจีน (Sung, *et al.*, 2015; Pang, 2006 และ Chen, *et al.*, 2013) ภาษาอาหรับ (Al-Khalifa and Al-Ajlan, 2010 และ Saddiki, *et al.*, 2015) ภาษาเยอรมัน (Vor Der Brück and Hartrumpf, 2007) ภาษาสวีเดน (Sjöholm, 2012 และ Pilán, *et al.*, 2014) ภาษาญี่ปุ่น (Sato, *et al.*, 2008) และได้นำวิธีการประเมินความยากง่ายไปประยุกต์ใช้ในด้านต่าง ๆ เช่น สนับสนุนด้านการศึกษา (Heilman, *et al.*, 2010; Beinborn, *et al.*, 2012 และ Dascalu, 2014) สนับสนุนโปรแกรมค้นหา (Collins-Thompson, *et al.*, 2011 และ Tan, *et al.*, 2012) สนับสนุนผู้อ่านที่พิการ (Abedi, *et al.*, 2011; Sitbon and Bellot, 2008 และ Feng, *et al.*, 2009) สนับสนุนด้านสุขภาพ (Wang, 2006) และสำหรับงานวิจัยการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยยังเป็นขั้นเริ่มต้น (Daowadung and Chen, 2011) ผลงานวิจัยทางด้านนี้ในภาษาไทยมีน้อย และงานวิจัยที่ผ่านมาที่ดำเนินการวิจัยเกี่ยวกับการประเมินระดับความง่ายของข้อความในเอกสารภาษาไทย มีดังนี้

- Daowadung and Chen (2011) ได้ประเมินระดับความยากง่ายของข้อความภาษาไทยโดยการจำแนกประเภทเอกสารการเรียนสำหรับระดับชั้นประถมศึกษาที่ 1 ถึง 6 ของประเทศไทย ซึ่งดำเนินการจำแนกโดยใช้เทคนิคการเรียนรู้ด้วยเครื่องคอมพิวเตอร์ (Machine

Learning) คือ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และคุณสมบัตินี้ (Features) ในการประเมินความยากง่าย คือ Term Frequency Inverse Document Frequency (TF-IDF) เป็นวิธีการให้น้ำหนักสำหรับคำที่มีความสำคัญ ดำเนินการทดลองโดยการวัดความแม่นยำของการจำแนกโดยพิจารณาแยกทีละประเภท (Precision) วัดความถูกต้องของการจำแนกโดยพิจารณาแยกทีละประเภท (Recall) และการวัด Precision และ Recall ของการจำแนกพร้อมกันโดยพิจารณาแยกทีละประเภท (F-measure) ในการทดลองนี้ได้แบ่งออกเป็น 2 แบบ คือ 1) การจำแนกเอกสารโดยกำหนดให้เอกสารของระดับชั้นประถมศึกษาปีที่ 1 ถึง 2 คือ ตัวแทนของชุดข้อมูลที่เป็นบวก (Positive) และกำหนดให้เอกสารของระดับชั้นประถมศึกษาปีที่ 3 ถึง 6 คือ ตัวแทนของชุดข้อมูลที่เป็นลบ (Negative) ซึ่งในการทดลองการจำแนกนี้เรียกว่า “Lower grades” และ 2) การจำแนกเอกสารโดยกำหนดให้เอกสารของระดับชั้นประถมศึกษาปีที่ 1 ถึง 4 คือ ตัวแทนของชุดข้อมูลที่เป็นบวก (Positive) และกำหนดให้เอกสารของระดับชั้นประถมศึกษาปีที่ 5 ถึง 6 คือ ตัวแทนของชุดข้อมูลที่เป็นลบ (Negative) ซึ่งในการทดลองการจำแนกนี้เรียกว่า “Middle grades” ผลในการดำเนินการในครั้งนี้สรุปได้ว่า การจำแนกเอกสารให้เหมาะสมสำหรับระดับ Middle grades ได้รับค่า F-measure สูงถึงร้อยละ 83.00

- Daowadung and Chen (2012) ได้ประเมินความยากง่ายของข้อความภาษาไทยโดยการจำแนกประเภทเอกสารการเรียนสำหรับระดับชั้นประถมศึกษาปีที่ 1 ถึง 6 ของประเทศไทย ซึ่งการวิจัยนี้เป็นการวิจัยต่อยอดจากงานวิจัยก่อนหน้าของ Daowadung and Chen (2011) โดยการเพิ่มการกำจัดคำหยุด (Stop Word) ในเอกสาร เพื่อต้องการทราบถึงอิทธิพลของคำหยุดที่มีต่อการประเมินระดับความยากง่ายของภาษาไทย ผลในการดำเนินการนี้สรุปได้ว่า การจำแนกเอกสารให้เหมาะสมสำหรับระดับ Middle grades ได้รับค่า F-measure สูงถึงร้อยละ 87.00 ซึ่งค่า F-measure สูงกว่างานวิจัยก่อนหน้า

- Tongtep, *et al.* (2014) ได้ประเมินระดับความยากง่ายของข้อความภาษาไทยโดยการจำแนกเอกสารสำหรับนักเรียนระดับชั้นมัธยมตอนต้น มัธยมตอนปลาย และปริญญาตรี โดยได้นำเสนอเทคนิคสำหรับการประเมินระดับความยากง่ายทั้งหมด 7 เทคนิค ดังนี้ 1) Word Ranking Graph 2) Part of Speech (POS) Ranking 3) Word and POS Ranking 4) Number of Characters Ranking Graph 5) Character Per Word (CPW) Document Ranking 6) Words Per

Paragraph (CPW) Document Ranking 7) Hybrid ranking ซึ่งนำเทคนิคทั้งหมดนี้มาเปรียบเทียบ เพื่อหาเทคนิคสำหรับประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยได้ถูกต้องที่สุด ผลในการดำเนินการนี้สรุปได้ว่า เทคนิค CWP Document Ranking ให้ค่าเฉลี่ยของ F-measure สูงที่สุดถึงร้อยละ 46.53

- Chen and Daowadung (2015) ได้ประเมินความยากง่ายของข้อความภาษาไทยโดยการจำแนกประเภทเอกสารการเรียนสำหรับระดับชั้นประถมศึกษาที่ 1 ถึง 6 ของประเทศไทย ซึ่งการวิจัยนี้เป็นการวิจัยต่อยอดจากงานวิจัยก่อนหน้าของ Daowadung and Chen (2011, 2012) โดยได้นำเสนอคุณสมบัติในการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยทั้งหมด 3 ชุดคุณสมบัติ ได้แก่ 1) Shallow (SL) 2) Language model (LM) 3) Term frequency (TF) ซึ่งได้นำชุดคุณสมบัติทั้ง 3 นี้มาเปรียบเทียบ เพื่อหาชุดคุณสมบัติสำหรับประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยมีค่าความถูกต้องของการจำแนกโดยพิจารณารวมทุกประเภท (Accuracy) สูงที่สุด ผลในการดำเนินการในครั้งนี้สรุปได้ว่า ชุดคุณสมบัติที่มีการผสมผสานกันของทั้ง 3 ชุดคุณสมบัติ (TF+SL+LM) ได้ให้ค่า Accuracy สูงที่สุดถึงร้อยละ 50.97

2.9 การจัดกลุ่มเอกสาร

การจัดกลุ่มเอกสาร (Text clustering หรือ Document clustering) เป็นเทคนิคที่ประยุกต์มาจากเทคนิคการวิเคราะห์กลุ่ม (Cluster analysis) ซึ่งเป็นเทคนิคที่ใช้จัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันให้อยู่ภายใต้กลุ่มเดียวกัน โดยข้อมูลดังกล่าวเป็นข้อมูลที่ไม่ทราบกลุ่มของข้อมูลและจำนวนกลุ่มที่แน่ชัด ส่งผลให้ได้กลุ่มข้อมูลที่เป็นธรรมชาติ (Svadas and Jha, 2014) สำหรับเทคนิคในการจัดกลุ่มข้อมูลโดยพื้นฐานประกอบด้วย 4 เทคนิค ดังนี้

2.9.1 วิธีแบบพาร์ทิชัน

วิธีแบบพาร์ทิชัน (Partitioning method) คือ เทคนิคการจัดกลุ่มข้อมูลโดยการแบ่งส่วน (Partition) ข้อมูลตามความคล้ายคลึงของข้อมูล ซึ่งจะแทนข้อมูลแต่ละข้อมูลด้วยวัตถุ โดยจะ

มีลักษณะในการจัดกลุ่มเป็นรอบจนกว่าจะได้กลุ่มข้อมูลที่ดีที่สุด (Han, *et al.*, 2011) โดยมีเงื่อนไข ดังนี้

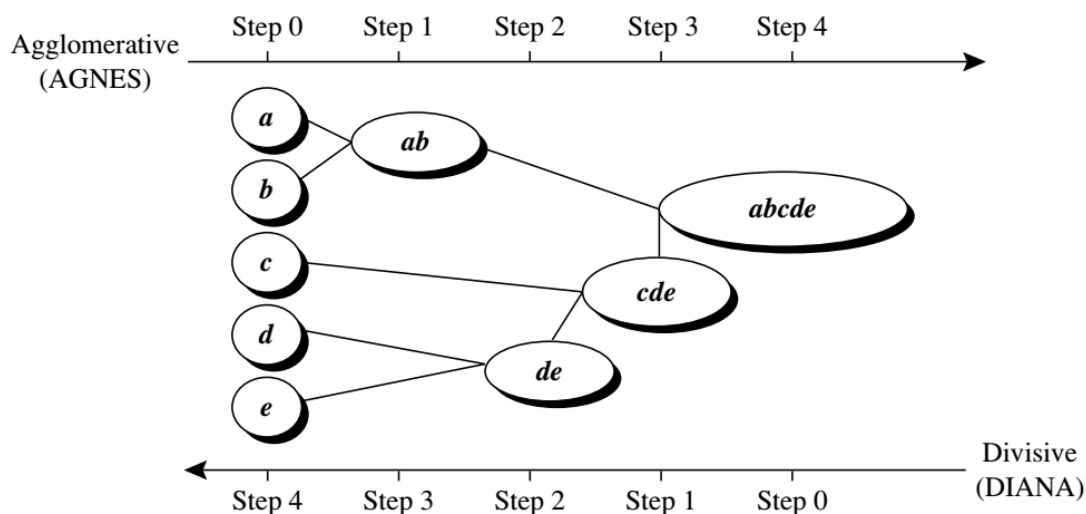
- ต้องกำหนดจำนวนกลุ่มของข้อมูลไว้ก่อนหน้า
- จำนวนกลุ่มที่กำหนดจะต้องไม่น้อยกว่าจำนวนวัตถุที่ต้องการการจัดกลุ่ม
- แต่ละกลุ่มจะต้องมีวัตถุอย่างน้อย 1 วัตถุ
- สำหรับผลลัพธ์ของการจัดกลุ่มในแต่ละวัตถุจะต้องสังกัดภายใต้กลุ่มใดกลุ่มหนึ่ง

เท่านั้น

เทคนิคนี้เหมาะสำหรับค้นหากลุ่มข้อมูลที่มีรูปร่างเป็นทรงกลม เทคนิคที่ได้รับ ความนิยมและนำไปประยุกต์ใช้กับการจัดกลุ่มเอกสาร ได้แก่ เค-มีนส์ (K-means) (Calvillo, *et al.*, 2013)

2.9.2 วิธีแบบไฮราร์คี

วิธีแบบไฮราร์คี (Hierarchical method) คือ เทคนิคการจัดกลุ่มข้อมูลโดยการแบ่ง ข้อมูลเป็นลำดับชั้น (Hierarchy) ตามความคล้ายคลึงของข้อมูล โดยมีลักษณะการจัดกลุ่ม 2 รูปแบบ คือ 1) การรวม คือ การเริ่มต้นการจัดกลุ่มจากกลุ่มเล็ก ๆ หลายกลุ่ม นำมารวมกันกลายเป็นกลุ่มใหญ่ เรียกว่า Agglomerative 2) การแยก คือ การแยกกลุ่มจากกลุ่มใหญ่ให้กลายเป็นกลุ่มเล็ก ๆ หลายกลุ่ม เรียกว่า Divisive แสดงดังรูปที่ 2.12 โดยใช้เดนโดแกรม (Dendrogram) สำหรับแสดงผลลัพธ์จาก การแบ่งกลุ่มในเทคนิคนี้ เทคนิคนี้เหมาะสำหรับค้นหากลุ่มข้อมูลที่มีรูปร่างเป็นทรงกลม (Han, *et al.*, 2011)



รูปที่ 2.12 ลักษณะการจัดกลุ่มแบบ Agglomerative และ Divisive (Han, *et al.*, 2011)

2.9.3 วิธีแบบเดนซิตีเบส

วิธีแบบเดนซิตีเบส (Density-based method) คือ เทคนิคการจัดกลุ่มข้อมูลโดยการวิเคราะห์ตำแหน่งพื้นที่ความหนาแน่น (Density) ของข้อมูล โดยการวัดระยะห่างระหว่างข้อมูล ซึ่งจะแทนข้อมูลแต่ละข้อมูลด้วยวัตถุ เทคนิคนี้สามารถจัดการกับข้อมูลที่มีรูปร่างเฉพาะได้โดยง่าย และสามารถกรองข้อมูลที่ผิดปกติออกจากกลุ่ม เทคนิคนี้เป็นเทคนิคพื้นฐานสำหรับการพัฒนาเทคนิคใหม่ขึ้น เช่น DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Han, *et al.*, 2011)

2.9.4 วิธีแบบกริดเบส

วิธีแบบกริดเบส (Grid-based method) คือ เทคนิคการจัดกลุ่มข้อมูลโดยการคัดเลือกข้อมูลใส่ในพื้นที่ที่มีลักษณะเป็นช่อง (Cell) ในรูปแบบโครงสร้างตารางที่มีหลายมิติ โดยต้องเตรียมช่องสำหรับใส่ข้อมูลไว้ก่อนหน้า ข้อดีของเทคนิคนี้ คือ มีความรวดเร็วในการประมวลผลในการจัดกลุ่มข้อมูลตามจำนวนช่องที่เตรียม และ STING (Statistical Information Grid) เป็นหนึ่งในวิธีการจัดกลุ่มที่มีการจัดกลุ่มด้วยเทคนิค Grid-based (Han, *et al.*, 2011)

ซึ่งสามารถสรุปลักษณะสำคัญของเทคนิคการจัดกลุ่มข้อมูลแต่ละเทคนิค แสดงดังตารางที่ 2.9

ตารางที่ 2.9 คุณสมบัติของเทคนิคการจัดกลุ่มแต่ละเทคนิค

คุณสมบัติ	Partitioning	Hierarchical	Density-based	Grid-based
จัดกลุ่มข้อมูลแบบอัตโนมัติ		X		
สามารถกรอค่าผิดปกติทิ้ง รบกวนออกได้		X	X	
สามารถปรับใช้กับข้อมูลข้อมูล แบบพลวัต	X		X	
เหมาะสำหรับชุดข้อมูลขนาด เล็กและกลาง	X			
สามารถค้นหากลุ่มข้อมูลที่มี รูปร่างเฉพาะ			X	
สามารถปรับใช้กับเทคนิค อื่น ๆ ได้				X
สามารถค้นหากลุ่มข้อมูลที่มี รูปร่างทรงกลม	X	X		
ความเร็วในการประมวล ผลการจัดกลุ่มข้อมูล				X

เทคนิคในข้างต้นเป็นเพียงเทคนิคพื้นฐานสำหรับจัดกลุ่มข้อมูล ปัจจุบันมีการคิดค้นเทคนิคการจัดกลุ่มเพิ่มมากขึ้นตามการนำไปประยุกต์ใช้ในด้านต่าง ๆ สำหรับการนำเทคนิคการจัดกลุ่มมาประยุกต์ใช้ในด้านการจัดกลุ่มเอกสารมีหลากหลายอัลกอริทึม (Algorithm) เช่น เค-มีนส์,

Fuzzy Clustering, Expectation Minimization, Self Organizing Maps โดยมีตัวอย่างงานวิจัยที่นำเทคนิคการจัดกลุ่มเอกสารมาประยุกต์ใช้ในการจัดกลุ่มเอกสาร ดังนี้

- Ma, *et al.* (2012) ได้ใช้เทคนิคการจัดกลุ่มเอกสาร โดยใช้อัลกอริทึม Self Organizing Maps เพื่อจัดประเภทของเอกสารโครงร่างงานวิจัยที่เป็นภาษาจีนและอังกฤษ

- Calvillo, *et al.* (2013) ได้จัดกลุ่มบทความงานวิจัย เพื่อเพิ่มประสิทธิภาพของสารสนเทศและค้นหาได้รวดเร็วยิ่งขึ้น โดยใช้อัลกอริทึมเค-มีนส์

- สุพัฒชยา ปัญโญ และ ทศนัย ชุ่มวัฒนะ (2559) ได้จัดกลุ่มคำอธิบายรายวิชาสำหรับการเทียบโอนผลการเรียนของมหาวิทยาลัยรังสิต โดยใช้เทคนิคโครงข่ายประสาทเทียม (Neural Networks)

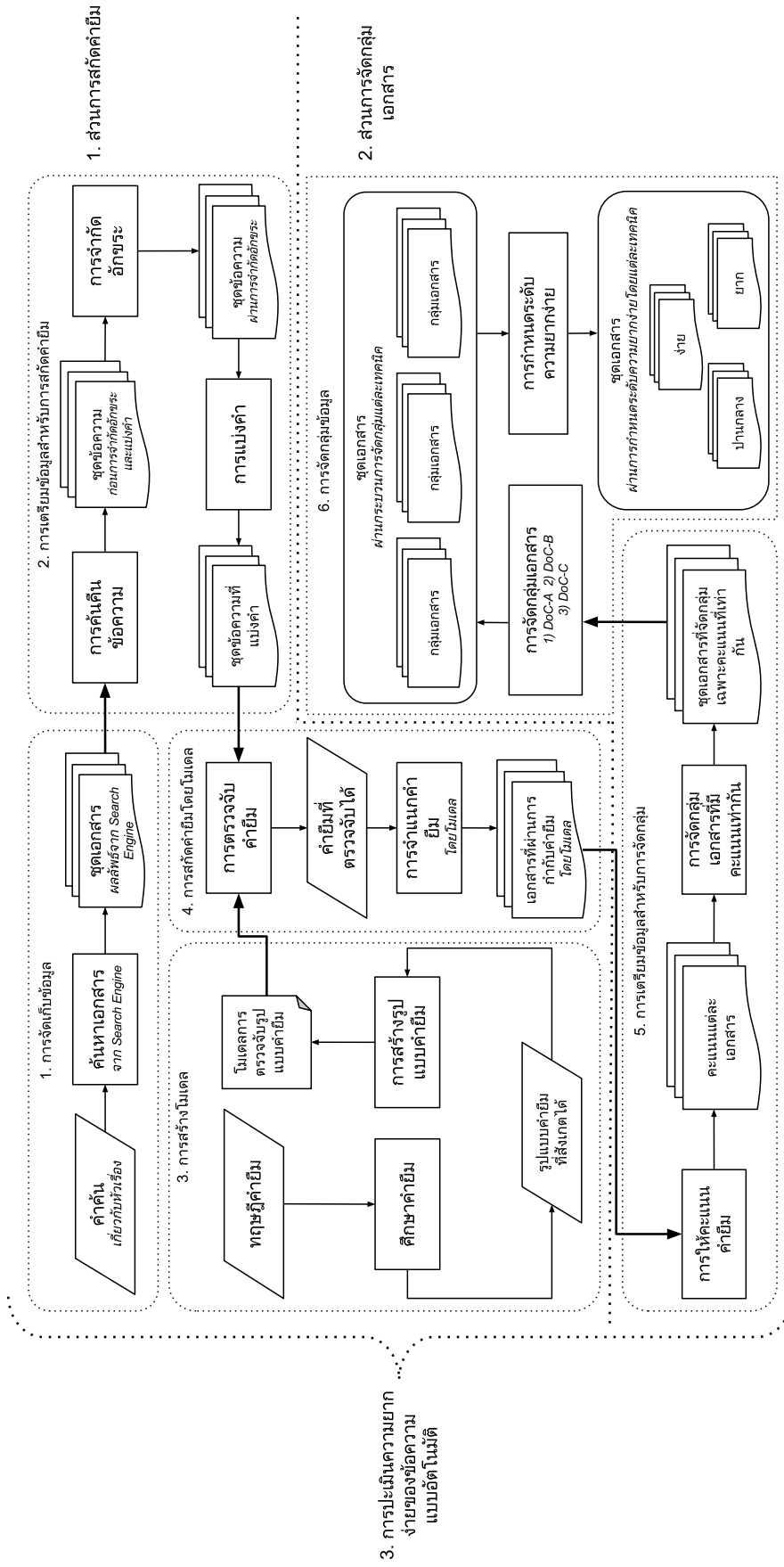
บทที่ 3

วิธีดำเนินการวิจัย

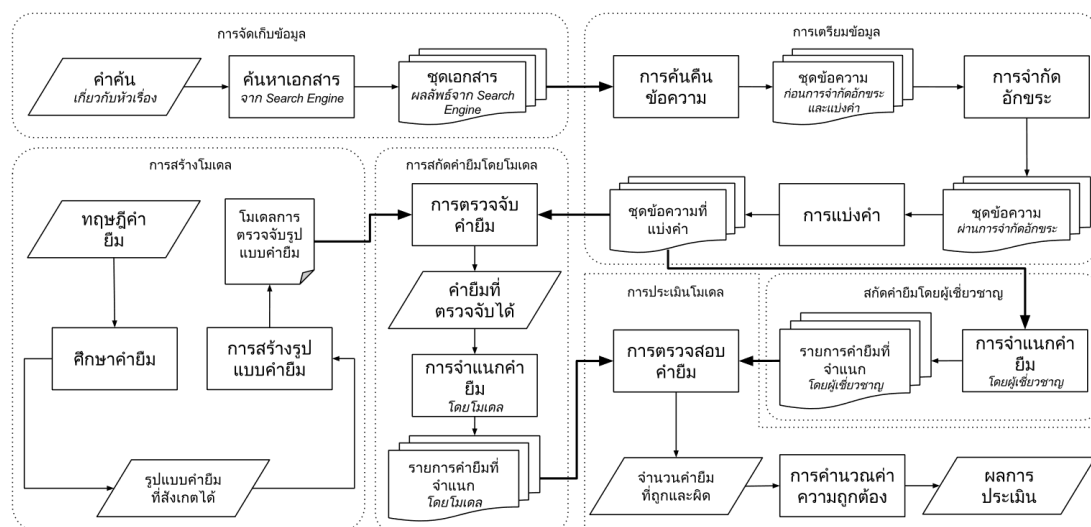
ในบทนี้กล่าวถึงขั้นตอนในการดำเนินงานวิจัย งานวิจัยนี้เป็นงานวิจัยที่ดำเนินการศึกษาและค้นหาวิธีการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย โดยใช้คำยืมในภาษาไทยเป็นคุณสมบัติในการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย และประยุกต์ใช้เทคนิคการจัดกลุ่มเอกสารในการจัดกลุ่มเอกสารตามระดับความยากง่ายของข้อความในเอกสาร ผู้วิจัยได้แบ่งส่วนวิธีดำเนินการวิจัยออกเป็น 3 ส่วน คือ 1) การสกัดคำยืม (Loanword extraction) 2) การจัดกลุ่มเอกสาร (Text clustering) 3) การประเมินระดับความยากง่ายของข้อความแบบอัตโนมัติ (Automated text readability assessment) ดังที่แสดงในรูปที่ 3.1 และสามารถอธิบายได้ดังนี้

3.1 การสกัดคำยืม

เนื่องจากระดับความยากง่ายของเอกสารขึ้นกับจำนวนคำยืมในเอกสาร ดังนั้นงานวิจัยนี้จึงใช้คำยืมเป็นคุณสมบัติในการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย และเพื่อให้ได้มาซึ่งตำแหน่งของคำยืมในเอกสารดิจิทัลภาษาไทย ดังนั้นงานวิจัยนี้ได้ศึกษาหาวิธีในการสกัดตำแหน่งของคำยืม ซึ่งในส่วนนี้จะอธิบายเกี่ยวกับกระบวนการในการสกัดตำแหน่งคำยืมและกระบวนการประเมินผลความถูกต้องของการสกัดตำแหน่งคำยืม แสดงดังรูปที่ 3.2 ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้



รูปที่ 3.1 วิธีการดำเนินการวิจัย



รูปที่ 3.2 การดำเนินการวิจัยการสกัดคำยืม

3.1.1 การจับเก็บข้อมูล

ขั้นตอนนี้ คือ การค้นหาเอกสารเพื่อจัดเก็บและนำไปใช้สำหรับการสกัดคำยืม โดยเริ่มจากการกำหนดคำสำคัญในการค้นหาเอกสาร (Keywords หรือ คำค้น) เพื่อใช้สำหรับค้นหาเอกสารที่มีเนื้อหาเกี่ยวข้องกับโปรแกรมค้นหา เนื่องจากการประเมินความยากง่ายจำเป็นต้องใช้เอกสารที่มีเนื้อหาที่เกี่ยวข้องกัน ผลลัพธ์ที่ได้จากขั้นตอนนี้ คือ เอกสารที่มีความเกี่ยวข้องกัน ซึ่งอยู่ในรูปแบบการเชื่อมโยงหรือลิงค์ (Link) ไปยังเนื้อหาในหน้าเว็บไซต์หลัก แสดงดังรูปที่ 3.3

Google data mining คือ

ทั้งหมด ค้นรูป วิดีโอ ข่าวสาร แผนที่ เพิ่มเติม การตั้งค่า เครื่องมือ

ผลการค้นหาประมาณ 82,200 รายการ (0.51 วินาที)

บทความทางการศึกษาเกี่ยวกับ data mining คือ
Introduction to Data Mining - Ch - อ้างโดย3

Data Mining - compcenter
compcenter.bu.ac.th/news-information/data-mining
ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล Data Mining เป็นขั้นตอนการ
ค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่. - Pattern Evaluation ...
คุณไปที่หน้าเว็บนี้เมื่อ 13/6/2017

เหมืองข้อมูล (data mining) – { Automated Software } ;
www.autosoft.in.th/data-warehouse/เหมืองข้อมูล-data-mining/
6 ก.พ. 2559 - Data Mining คือ การวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (big data) ... การทำเหมืองข้อมูล (data
mining) ขอสรุป Algorithm of Data mining ...
คุณไปที่หน้าเว็บนี้เมื่อ 13/6/2017

การทำเหมืองข้อมูล - วิกิพีเดีย
https://th.wikipedia.org/wiki/การทำเหมืองข้อมูล
... data mining) หรืออาจจะเรียกว่า การค้นหาความรู้ในฐานข้อมูล (อังกฤษ: Knowledge ... การทำเหมืองข้อมูล คือ
กระบวนการที่กระทำกับข้อมูล(โดยส่วนใหญ่จะมีจำนวนมาก) ...
ขั้นตอนการทำเหมืองข้อมูล · ประโยชน์จากการทำเหมืองข้อมูล · ดูเพิ่ม · แหล่งข้อมูลอื่น
คุณไปที่หน้าเว็บนี้เมื่อ 13/6/2017

Data Mining: Data Mining (เหมืองข้อมูล)
sajeegm301.blogspot.com/2015/11/data-mining.html
3 พ.ย. 2558 - Data Mining (เหมืองข้อมูล) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและ
ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ...
คุณไปที่หน้าเว็บนี้เมื่อ 13/6/2017

รูปที่ 3.3 ตัวอย่างการค้นหาเอกสารด้วยโปรแกรมค้นหาของกูเกิล (Google)

จากรูปที่ 3.3 แสดงถึงตัวอย่างส่วนต่อประสานกับผู้ใช้ (User interface) ของโปรแกรมค้นหาของกูเกิล ประกอบด้วย 3 ส่วน ดังนี้

ส่วนที่ 1 ช่องสำหรับรับข้อมูล keywords

ส่วนที่ 2 ผลลัพธ์จากการค้นหาด้วย keywords ซึ่งแต่ละผลลัพธ์ สามารถเชื่อมโยง

ไปยังหน้าเว็บไซต์หลักโดยตรง

ส่วนที่ 3 ลิงค์สำหรับนำไปยังหน้าเว็บไซต์หลักของเอกสารดิจิทัล

ในงานวิจัยนี้ได้กำหนดหัวเรื่องสำหรับการกำหนดคำค้นในการค้นหาเอกสาร จำนวน 2 หัวเรื่อง คือ 1) Data mining 2) iOS operating system และมีจำนวนเอกสารที่ค้นหาด้วย โปรแกรมค้นหาของกูเกิล โดยการคัดลอกเนื้อหาข้อความในเอกสารดิจิทัลด้วยตนเอง (Manual) จำนวน 30 ฉบับ แบ่งเป็นเอกสารในหัวเรื่อง Data mining จำนวน 11 ฉบับ และหัวเรื่อง iOS operating system จำนวน 19 ฉบับ โดยสาเหตุของการเลือกหัวเรื่องของเอกสารในงานวิจัยนี้ เนื่องจากหัวเรื่อง Data mining และ iOS operating system มีเนื้อหาเชิงให้ความรู้ ซึ่งเหมาะสำหรับการทดลองในการประเมินระดับความยากง่าย และทั้ง 2 หัวเรื่องเป็นหัวเรื่องที่ถูกกลุ่มตัวอย่างกำลังศึกษาในขณะนั้น สำหรับหลักเกณฑ์ในการค้นหาเอกสารและเลือกเอกสารจากโปรแกรมค้นหาของกูเกิล เริ่มจากการค้นหาเอกสารด้วยคำค้นที่แสดงดังตารางที่ 3.1 โดยผลลัพธ์การค้นหาจากโปรแกรมค้นหาดังกล่าวจะเรียงลำดับเอกสารจากเอกสารที่มีความสัมพันธ์กับคำค้นมากที่สุดไปยังเอกสารที่มีความสัมพันธ์กับคำค้นน้อยที่สุด ดังนั้นงานวิจัยนี้จะพิจารณาเอกสารดิจิทัล 30 ฉบับแรกที่ได้จากผลลัพธ์จากการค้นหา หลังจากนั้นจะคัดเลือกเฉพาะเอกสารที่ไม่ซ้ำกัน เพื่อนำมาใช้ในขั้นตอนถัดไป คือ การเตรียมข้อมูล

ตารางที่ 3.1 รายละเอียดของเอกสารที่จัดเก็บ

หัวเรื่อง	คำค้น	จำนวนเอกสาร (ฉบับ)	จำนวนอักขระของเอกสารทั้งหมด	จำนวนอักขระเอกสาร
Data mining	Datamining คือ data mining คือ การทำเหมืองข้อมูล คำดำไม่นิ่ง คืออะไร	11	27,852	2,532
iOS operating system	iOS คือ ระบบปฏิบัติการ ios	19	10,336	544

3.1.2 การเตรียมข้อมูล

ขั้นตอนนี้ คือ การเตรียมเอกสารให้อยู่ในรูปแบบที่พร้อมสำหรับการสกัดคำยืม โดยปกติแล้วเว็บไซต์เอกสารดิจิทัลจะประกอบด้วยเนื้อหาหลายรูปแบบ เช่น ข้อความ ภาพ วิดีทัศน์ แสดงดังรูปที่ 3.4 ซึ่งงานวิจัยนี้มุ่งเน้นไปยังเนื้อหาในรูปแบบข้อความดิจิทัลเท่านั้น โดยการเตรียมข้อมูลเริ่มจากกระบวนการดึงเนื้อหาในเอกสารเฉพาะข้อความจากผลลัพธ์ในขั้นตอนการจัดเก็บข้อมูล ซึ่งดำเนินการโดยการกำจัดเนื้อหาที่เป็นสิ่งรบกวนต่อการประมวลผล เช่น ภาพ วิดีทัศน์ คำสั่งเชิงภาษาโปรแกรม ผลลัพธ์ที่ได้ คือ ชุดเอกสารข้อความ แสดงดังรูปที่ 3.5 กระบวนการลำดับถัดมา คือ การจำกัดอักขระในชุดเอกสารข้อความแต่ละฉบับเพื่อให้เอกสารทั้งชุดมีจำนวนอักขระที่เท่ากัน ผลลัพธ์ คือ เอกสารทั้งหมดในชุดเดียวกันมีจำนวนอักขระเท่ากัน แสดงดังรูปที่ 3.6 กระบวนการลำดับถัดมา คือ การตัดคำ หรือ แบ่งคำ ในชุดเอกสารที่ผ่านกระบวนการจำกัดอักขระแต่ละฉบับ ผลลัพธ์ที่ได้ คือ ชุดเอกสารข้อความที่ถูกแบ่งขอบเขตของแต่ละคำในเอกสาร แสดงดังรูปที่ 3.7

Home Projects Software Development Web Design About me

{ Automated Software ;}

Home Dashboard Design Data Warehouse Data Science Business Intelligence Programming ห้องทดลอง1 ห้องทดลอง2

Green DATA.

ชื่อเรื่อง

เนื้อหา

เหมืองข้อมูล (data mining)

Business Intelligence and Platform

รูปภาพ

ข้อความ

เรื่องล่าสุด

ความเห็นล่าสุด

ส่วนที่ 1 Data Source เป็นข้อมูลจากแหล่งต่างๆที่เราต้องการเก็บรวมไว้บน data warehouse ของเราไม่ว่าจะเป็น customer, product, Order และอื่นๆซึ่งเป็นข้อมูลภายในองค์กรและข้อมูลภายนอกที่ต้องนำมาใช้ประโยชน์ เช่น location base, wiki, social, etc.

ส่วนที่ 2 Data Management Platform เน้นอนว่าข้อมูลที่เรารวเก็บและทำการ Transformation มาจากแหล่งต่างๆ(data source) จะถูกนำไปใช้จำนวนคือเช่น ทำ report, data analytic, OLAP, data mining มันจะต้องอยู่ในรูปที่สามารถนำไปใช้งานได้โดยสะดวกเร็ว(speed)และมีคุณภาพ(data quality) เพราะฉะนั้นในการออกแบบ data warehouse เพื่อเก็บข้อมูลในส่วนนี้จึงถือว่ามีค่าสำคัญ

ส่วนที่ 3 BI Platform เป็นส่วนที่เรานำข้อมูลมาทำการวิเคราะห์(data analytics) ซึ่งอาจจะใช้เทคนิคที่เรียกว่า OLAP(Online Analytical Processing) เพื่อให้การวิเคราะห์ข้อมูลในมุมมองต่างๆ สามารถทำได้อย่างมีประสิทธิภาพ นอกจากนี้เราสามารถจะใช้เทคนิคขั้นสูงเพื่อจำแนกข้อมูล และหาความน่าจะเป็นโดยใช้ data mining ซึ่งจะมาแชร์ให้ฟังต่อไปครับ

Data Mining คืออะไร?

กำลังส่งหารูป : <http://bdcservicesgroup.com>

Data Mining คือ การวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (big data) เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้นเพื่อ

ไม่ได้มีความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่างๆ เช่น ตลาดหลักทรัพย์, หางสรรพสินค้า, ทางด้านการแพทย์, ยุทธศาสตร์ทหาร เป็นต้น

ห้องทดลอง1 ห้องทดลอง2

สร้าง Date Table/Dimension Table(DAX) Calendar function

เพิ่มจัดการมาเอา Excel Dashboard บน

การสร้าง Measure

การเขียน data source type จาก excel file เป็น sql driver หรือ data source ขึ้นวิธีการอย่างไร

วิธีการสร้าง Line chart เพื่อแสดงผลการเพิ่ม Sales Forecast จากข้อมูลที่เราคำนวณบน Microsoft Power BI

Mr. Automated Software

244 likes

Like Page

Share

Be the first of your friends to like this

Mr. Automated Software shared โปรแกรมเมอร์ไทย Thai programmer's post June 11 at 3:20pm

โปรแกรมเมอร์ไทย Thai programmer June 11 at 12:27pm

++++ ภาษาโปรแกรมมิ่งในเมืองไทยใช่++++

Java 53.7%

PHP 45.9%

C# 30.2%

ส่วนภาษา R และ Swift ยังมีน้อย

... See More

รูปที่ 3.4 ตัวอย่างส่วนต่อประสานกับผู้ใช้ของเว็บไซต์เอกสารดิจิทัล (<http://www.autosoft.in.th/data-warehouse/เหมืองข้อมูล-data-mining/>, 26 มกราคม 2559)

iOS ถือได้ว่าเป็นระบบปฏิบัติการบนสมาร์ทโฟนและแท็บเล็ตที่ได้รับความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปัจจุบันและยังเป็นคู่แข่งคนสำคัญในตลาดสมาร์ทโฟนและแท็บเล็ตโลกของ Android OS อีกด้วย ซึ่งหลายคนอาจกำลังใช้งานผ่าน iPhone, iPad, iPod อยู่ แต่ก็อาจมีอีกหลายคนที่จะยังไม่รู้จักหรือรู้จักแบบคร่าวๆ ดังนั้นเรามารู้จัก iOS ระบบปฏิบัติการจาก Apple กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน iOS (iPhoneOS) คืออะไร? iOS หรือเดิมชื่อ iPhone OS เป็นระบบปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาขึ้นมาโดยบริษัท Apple Inc. (Apple Computer Inc.) โดยมีพื้นฐานในการพัฒนาจาก Mac OS X และ Darwin (OS แบบเปิดพัฒนาโดย Apple Inc.) ภายใต้แนวความคิด "Direct Manipulation" (การควบคุมโดยตรง) ซึ่งเน้นการควบคุมและใช้งานผ่านระบบสัมผัสด้วยนิ้ว (Multi-touch), ปุ่มและสวิตช์ โดยจะมีเซ็นเซอร์ตรวจจับการเคลื่อนไหวโดยเฉพาะใส่ไว้ในตัว Device (Built in) ซึ่งเป็นตัวกลางในการควบคุม และในปัจจุบันระบบปฏิบัติการ iOS ได้ถูกนำไปใช้งานผ่าน Device พกพาของ Apple เอง ได้แก่ iPhone, iPad, iPod และในอนาคตอาจได้เห็นบนอุปกรณ์สวมใส่เช่น Smartwatch ด้วยเช่นกัน Version ของ iOS iOS ถูกนำมาใช้งานเป็นครั้งแรกในปี 2007 ซึ่งเปิดตัวพร้อมสมาร์ทโฟน iPhone รุ่นแรกและถูกเรียกว่า iPhone OS ก่อนจะมาเปลี่ยนชื่อเป็น iOS ในภายหลัง ซึ่งนับจากปี 2007-ปัจจุบัน Apple ได้ปล่อย iOS ออกมาให้ใช้งานกันทั้งหมดถึง 7 เวอร์ชันหลักแล้ว โดยมีรายละเอียดดังนี้ iOS X หรือ iPhone OS (First Generation) iOS เวอร์ชันแรกนี้เปิดตัวภายในงาน MacWorld Expo ซึ่งมาพร้อมกับสมาร์ทโฟนเครื่องแรกของบริษัท Apple และสร้างความฮือฮาเป็นอย่างมาก เพราะเป็นสมาร์ทโฟนที่นำปฏิวัติวงการมือถือ ณ ตอนนั้นเลยทีเดียว ได้ไม่ว่าจะเป็นการมาพร้อมระบบหน้าจอสัมผัสแบบ Capacitive Multi-touch ที่สามารถใช้การควบคุมผ่านนิ้วด้วยการแสดงท่าทางของนิ้วเช่น การใช้นิ้วจิบเข้า-ออกเพื่อใช้ฟังก์ชัน Zoom ภาพ เป็นต้น และนอกจากนี้ยังสามารถใช้งานแป้นพิมพ์คีย์บอร์ดด้วยการสัมผัสและเก็บได้เองเมื่อไม่ใช้งานอีกด้วย ซึ่ง ณ ตอนนั้น สตีฟ จ๊อบส์ได้กล่าวว่า iPhone OS ของ Apple เป็น OS บนอุปกรณ์พกพาที่ก้าวนำสมาร์ทโฟนอื่นๆ ในสมัยนั้นไปข้างหน้าถึง 5 ปีเลยทีเดียว iOS 2.0 หรือ iPhone OS 2.0 iOS เวอร์ชันที่ 2 นี้เปิดตัวพร้อมกับการมาของ iPhone 3G ในปี 2008 และมีการเพิ่ม App Store เข้ามาใช้งานด้วย ซึ่งนับเป็นการปฏิวัติวงการสมาร์ทโฟนอีกครั้งด้วยการเปิดให้ผู้ใช้สามารถดาวน์โหลดแอปพลิเคชันจากผู้พัฒนาต่างๆ ลงบน iPhone ได้ผ่านทาง App Store ที่ได้รับการคัดกรองจากทาง Apple ว่าปลอดภัยและไม่ต้องไปหาโหลดผ่านเว็บของผู้ผลิตหรือที่อื่นๆ เอง ซึ่งอาจเกิดปัญหาในเรื่องการคัดข้อมูลของผู้ใช้งานได้อีกด้วยและทาง Apple ก็ยังเปิดให้ผู้ที่สนใจสามารถเขียนแอปพลิเคชันส่งขึ้น App Store ได้ด้วย iOS 3.0 หรือ iPhone OS 3.0 iOS เวอร์ชันที่ 3 นี้ก็ยังคงเรียกว่า iPhone OS อยู่เหมือนเดิม ซึ่งในเวอร์ชันนี้ทาง Apple เปิดให้ใช้งานได้ในปี 2009 และมาพร้อมกับการเปิดตัวของ iPhone 3GS ด้วย โดยในเวอร์ชันนี้ได้มีการเพิ่มฟังก์ชันการใช้งานที่ขาดหายไปก่อนหน้านี้เช่น การเพิ่มการ Copy-Paste ข้อความ, เพิ่มความสามารถในการส่ง MMS, เพิ่มฟังก์ชันเขย่าแล้ว undo ข้อความ, เพิ่มประสิทธิภาพในการใช้งานแอปพลิเคชันในแอนดรอยด์ ฯลฯ เป็นต้น นอกจากนี้ในเวอร์ชันนี้ยังรองรับการใช้งานบน iPod และ iPad ที่จะทำการเปิดตัวในช่วงต้นปี 2010 และนับเป็นการปฏิวัติวงการ Tablet

รูปที่ 3.5 ตัวอย่างเอกสารข้อความดิจิทัล (<http://news.siamphone.com/news-16304.html>, 27 มกราคม 2559)

IT อีพเพทข่าวล่าสุดกับ ป้าเอก TechXcite ยังไม่ทันจะถึงงาน WWDC 2014 ที่ Apple จะจัดขึ้นในเดือนมิถุนายน โน่นเลยแต่ตอนนี้ปรากฏว่าเริ่มมีข้อมูลเกี่ยวกับ iOS 8 และ OS X 10.10 มาให้ได้ติดตามกันบ้างแล้วว่าจะมีอะไรใหม่ๆเพิ่มเข้ามาหรือปรับปรุงให้ดีขึ้นในระบบปฏิบัติการเวอร์ชันต่อไปของ Apple ทั้งบน iPhone, iPad, iPod Touch และผลิตภัณฑ์ตระกูล Mac ทั้งหมดครับ ส่วนจะจริงหรือไม่ไปรอลุ้นกันเดือนมิถุนายนกันได้นะจ๊ะ D ข้อมูลแรกระบบปฏิบัติการ iOS 8 iOS 8 สำหรับ iPhone, iPad และ iPod Touch จะใช้ชื่อโค้ดเนมว่า Okemo iOS 8 Healthbook iOS 8 จะรองรับแ

iOS ถือได้ว่าเป็นระบบปฏิบัติการบนสมาร์ตโฟนและแท็บเล็ตที่ได้รับความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปัจจุบันและยังเป็นคู่แข่งคนสำคัญในตลาดสมาร์ตโฟนและแท็บเล็ตโลกของ Android OS อีกด้วย ซึ่งหลายคนอาจกำลังใช้งานผ่าน iPhone, iPad, iPod อยู่ แต่ก็อาจมีอีกหลายคนที่ยังไม่รู้จักรู้จักหรือรู้จักแบบคร่าวๆ ดังนั้นเรามารู้จัก iOS ระบบปฏิบัติการจาก Apple กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน iOS (iPhoneOS) คืออะไร? iOS หรือเดิมชื่อ iPhone OS เป็นระบบปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาขึ้นมาโดยบริษัท Apple Inc. (Apple Computer Inc.) โดยมีพื้นฐานในการพัฒนา

iOS คือ ระบบปฏิบัติการ ที่เป็นของบริษัท Apple ในตอนแรกถูกพัฒนาขึ้นมาเพื่อ iPhone เปิดตัวครั้งแรก 29 มิถุนายน 2550 และมีการพัฒนาต่อไปที่ iPod Touch และ iPad ระบบปฏิบัติการนี้ สามารถเชื่อมต่อกับ itunes store เพื่อสามารถดาวน์โหลด Application ได้ ในเวอร์ชัน ios 2.0 เป็นครั้งแรก ในตอนนี้เป็นที่ฮือฮาสำหรับ ระบบปฏิบัติการนี้ จนตอนนี้มีทั้งหมด ถึง ios 7 แล้ว Z9dCA7D [sws_divider_basic] iOS Device คือ อุปกรณ์ที่สามารถใช้ระบบปฏิบัติการ iOS ได้ นั่นก็คือ iPhone , iPod , iPad โดยเราสามารถเรียก อุปกรณ์ทั้งหมดโดยรวมที่เรียกว่า iOS Device เช่น ล

แอปเปิลเปิดตัว iOS 8 ในครั้งนี้ แม้จะไม่ได้เปลี่ยนดีไซน์ครั้งใหญ่เหมือนครั้ง iOS 7 แต่เป็นการปรับปรุงและพัฒนาการใช้งาน UX (User eXperience) ให้ดียิ่งขึ้น อ่านเพิ่ม : รวมทุกเรื่องต้องรู้และวิธีย่อเทพ iOS 8 สำหรับ iPhone, iPad, iPod Touch ios-8 การรีวิวนี้จะเป็นการเล่าประสบการณ์การใช้งานผ่านมุมมองที่ใช้ iPhone แทบจะติดมือและทำหลายอย่างมากในชีวิตประจำวัน เพื่อให้เห็นการใช้งานจริงมากที่สุด โดยเวอร์ชันที่รีวิวจะเป็น iOS 8 Gold Master คือ ถ้าไม่มีปัญหาอะไร เวอร์ชันนี้จะเป็น iOS 8 ตัวเต็มนั่นเอง First Look-02 1. ดีไซน์ อย่างแร

รูปที่ 3.6 ตัวอย่างชุดเอกสารข้อความดิจิทัลที่ผ่านการจำกัดจำนวนอักขระ

iOS | ถือได้ว่าเป็นระบบปฏิบัติการบนสมาร์ทโฟนและแท็บเล็ตที่ได้รับ
 ความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปัจจุบันและยังเป็นคู่แข่ง
 คนสำคัญในตลาดสมาร์ทโฟนและแท็บเล็ตโลกของ | Android | OS | อีกด้วย
 ซึ่งหลายคนอาจกำลังใช้งานผ่าน | iPhone, | iPad, | iPod | อยู่ แต่ก็อาจมี
 อีกหลายคนที่ยังไม่รู้จักหรือรู้จักแบบคร่าวๆ | ดังนั้นเรามารู้จัก
 iOS | ระบบปฏิบัติการจาก | Apple | กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน
 iOS | (iPhoneOS) | คืออะไร? | iOS | หรือเดิมชื่อ | iPhone | OS | เป็นระบบ
 ปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาขึ้นมาโดยบริษัท | Apple | Inc. |
 (Apple Computer Inc.) | โดยมีพื้นฐานในการพัฒนาจ

รูปที่ 3.7 ตัวอย่างข้อความที่ผ่านการตัดคำ

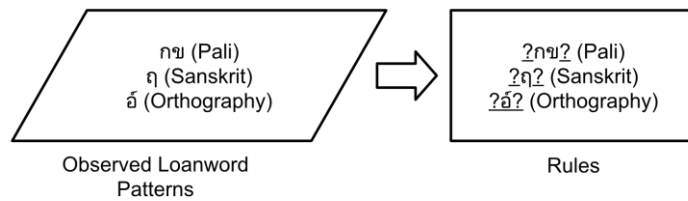
ในงานวิจัยนี้ได้สกัดข้อความออกจากเอกสาร สำหรับการจำกัดอักขระ ดำเนินการ
 โดยการเขียนชุดคำสั่งภาษาโปรแกรม PHP เพื่อต้องการให้จำนวนอักขระในแต่ละเอกสารเท่ากัน
 และสุดท้ายสำหรับการตัดคำภาษาไทยดำเนินการโดยใช้โปรแกรมตัดคำภาษาไทยเล็กซ์โต (LexTo)
 โดย LexTo คือ โปรแกรมตัดคำสำหรับข้อความภาษาไทย โดยใช้เทคนิคตัดคำแบบยาวที่สุด
 (Longest word matching) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
 (National Electronics and Computer Technology Center: NECTEC) (ศูนย์เทคโนโลยีอิเล็กทรอนิกส์
 และคอมพิวเตอร์แห่งชาติ, 2559) โดยงานวิจัยนี้ได้ปรับปรุงคลังคำศัพท์โดยการเพิ่มคำศัพท์เพื่อให้
 การตัดคำมีความครอบคลุมเพิ่มขึ้น

3.1.3 การสร้างแบบจำลอง

ขั้นตอนนี้ คือ การสร้างแบบจำลองสำหรับตรวจจับตำแหน่งของคำยืมในเอกสาร
 โดยเริ่มต้นด้วยกระบวนการศึกษาและค้นคว้าทฤษฎีที่เกี่ยวข้องกับคำยืมในภาษาไทย เพื่อหา
 รูปแบบ (Pattern) ของคำยืมในภาษาไทย ซึ่งรูปแบบดังกล่าว คือ รูปแบบที่มนุษย์สามารถสังเกตเห็น
 ได้ด้วยตา ดังที่ได้อธิบายในหัวข้อที่ 2.6 ผลลัพธ์ที่ได้ คือ รูปแบบที่มนุษย์สามารถสังเกตเห็นได้ แสดง
 ดังตารางที่ 3.2 และกระบวนการสุดท้าย คือ แปลงรูปแบบดังกล่าวแต่ละรูปแบบของแต่ละคำยืม

กลายเป็นรูปแบบที่เป็นเงื่อนไข หรือกฎสำหรับค้นหาตำแหน่งของคำยืม แสดงดังรูปที่ 3.8 ผลลัพธ์คือ แบบจำลองซึ่งสามารถตรวจจับตำแหน่งของคำยืมในเอกสารภาษาไทย

ในงานวิจัยนี้ได้ศึกษาและค้นคว้าทฤษฎีที่เกี่ยวกับคำยืมจากภาษาบาลี สันสกฤต และภาษาอื่น ๆ ที่มีเครื่องหมายทัศนมาตรกำกับด้านบนของพยัญชนะ (ตัวการ์นต์) โดยรูปแบบของภาษาบาลีและสันสกฤตของงานวิจัยนี้ได้อ้างอิงทฤษฎีหลักตัวสะกดตัวตาม และหลักการใช้พยัญชนะ ซึ่งได้อธิบายในหัวข้อที่ 2.6.2 และ 2.6.3 เช่น ก ฏ ศ แสดงดังตารางที่ 3.2 และในการสร้างแบบจำลองดำเนินการโดยการเขียนชุดคำสั่งภาษาโปรแกรม PHP



รูปที่ 3.8 ตัวอย่างการแปลงรูปแบบที่สามารถสังเกตกลายเป็นรูปแบบเงื่อนไข

ตารางที่ 3.2 รูปแบบสำหรับสกัดคำยืม

คำยืม	จำนวนรูปแบบ	รายละเอียดรูปแบบ
บาลี (Pali)	47	กก กข คค คฌ กง ขง งค งฌ จจ จฉ ชช ชฌ ฉจ ฉช ฉฌ ตต ตถ ทท ทธ นต นถ นท นธ นน ปป ฝฝ ฟฝ ภภ มป มฝ มพ มภ มม ยย ลล สส
สันสกฤต (Sanskrit)	25	ฤ ฤ ศษ รร รฌ กร รค คร คย ทย นย ฌย ฌย สก รด ตร ปต ตว ทร ปร ลป รป รฌ มท
ตัวการ์นต์ (Orthography)	1	อี

ตารางที่ 3.3 ตัวอย่างการแปลงรูปแบบที่สังเกตได้เป็นรูปแบบเงื่อนไข

คำยืม	ตัวอย่างรูปแบบที่สังเกตได้	ตัวอย่างรูปแบบเงื่อนไข	ตัวอย่างคำ
บาลี	กข	?กข?	นิรทุภข์ รักษา นักขัต
สันสกฤต	ชย	?ชย?	มัธยม มัธยมฐาน อยุชยา สนชยา
ตัวการ์นต์	อ์	อ์	แอร์ บาร์เลย์ ซับสไครป์

จากตารางที่ 3.3 แสดงให้เห็นถึงตัวอย่างการแปลงรูปแบบที่สามารถสังเกตเห็น กลายเป็นเงื่อนไขสำหรับค้นหาตำแหน่งคำยืม ตัวอย่างเช่น คำว่า “ทูกข์” รูปแบบที่สามารถสังเกตเห็น คือ กข หมายถึง ก คือ พยัญชนะสะกด และ ข คือ พยัญชนะตามหลังพยัญชนะสะกด โดยแปลงให้กลายเป็นรูปแบบเงื่อนไข คือ ?กข? หมายถึง ด้านหน้าพยัญชนะ ก เป็นอักษรใด ๆ ส่วน ก จะต้องเป็นพยัญชนะสะกด พยัญชนะถัดจาก ก จะต้องเป็น ข เท่านั้น และถัดจาก ข เป็นอักษรใด ๆ

3.1.4 การสกัดคำยืมด้วยแบบจำลอง

ขั้นตอนนี้ คือ การสกัดตำแหน่งของคำยืมจากชุดเอกสารข้อความที่ผ่านกระบวนการจำแนกคำโดยใช้แบบจำลองในการตรวจจับตำแหน่งของคำยืม โดยเริ่มจากการตรวจจับตำแหน่งของคำยืมตามรูปแบบของแบบจำลองในแต่ละคำยืมที่พัฒนาขึ้น ผลลัพธ์ คือ ตำแหน่งของคำยืม แสดงดังรูปที่ 3.9 โดยเส้นใต้ตัวอักษร หมายถึง รูปแบบของคำยืมที่ตรวจจับได้ และสุดท้าย คือ กระบวนการจำแนกประเภทของคำยืมตามรูปแบบที่ได้ระบุด้วยแบบจำลอง กล่าวคือ หนึ่งคำสามารถถูกระบุได้มากกว่าหนึ่งรูปแบบ ผลลัพธ์ที่ได้ คือ รายการคำยืมที่ได้จำแนกประเภทของคำยืมด้วยแบบจำลอง แสดงดังรูปที่ 3.10 ซึ่งแสดงถึง ตำแหน่งของคำยืมที่ผ่านการจำแนกประเภท โดยหนึ่งประเภทของคำยืมจะพิจารณาจากรูปแบบของคำยืมที่ตรวจจับได้ในหนึ่งคำ

iOS | ถือได้ว่าเป็นระบบปฏิบัติการบนสมาร์ทโฟนและแท็บเล็ตที่ได้รับความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปัจจุบันและยังเป็นคู่แข่งคนสำคัญในตลาดสมาร์ทโฟนและแท็บเล็ตโลกของ |Android| OS| อีกด้วย |ซึ่งหลายคนอาจกำลังใช้งานผ่าน |iPhone|, |iPad|, |iPod| |อยู่| แต่ก็อาจมีอีกหลายคนที่ยังไม่รู้จักหรือรู้จักแบบคร่าวๆ |ดังนั้นเรามารู้จัก |iOS| ระบบปฏิบัติการจาก |Apple| กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน |iOS| (|iPhoneOS|) |คืออะไร?| |iOS| หรือเดิมชื่อ |iPhone| OS| เป็นระบบปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาขึ้นมาโดยบริษัท |Apple| Inc. |(|Apple| Computer| Inc.)| |โดยมีพื้นฐานในการพัฒนาจ

รูปที่ 3.9 ตัวอย่างการตรวจจับตำแหน่งคำยืมในข้อความ

iOS | ถือได้ว่าเป็นระบบปฏิบัติการบนสมา<ORTHOGRAPHY>ร</ORTHOGRAPHY>ทโฟนและแท็บเล็ตที่ได้รับความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปี<PALI>จ</PALI>นและยังเป็นคู่แข่งคนสำคัญในตลาดสมา<ORTHOGRAPHY>ร</ORTHOGRAPHY>ทโฟนและแท็บเล็ตโลกของ |Android| OS| อีกด้วย |ซึ่งหลายคนอาจกำลังใช้งานผ่าน |iPhone|, |iPad|, |iPod| |อยู่| แต่ก็อาจมีอีกหลายคนที่ยังไม่รู้จักหรือรู้จักแบบคร่าวๆ |ดังนั้นเรามารู้จัก |iOS| ระบบปฏิบัติการจาก |Apple| กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน |iOS| (|iPhoneOS|) |คืออะไร?| |iOS| หรือเดิมชื่อ |iPhone| OS| เป็นระบบปฏิบัติการบนอุป<SANSKRIT>กร</SANSKRIT><ORTHOGRAPHY>ณ</ORTHOGRAPHY>พกพาที่ถูกพัฒนาขึ้นมาโดยบริ<SANSKRIT>ษ</SANSKRIT>ท |Apple| Inc. |(|Apple| Computer| Inc.)| |โดยมีพื้นฐานในการพัฒนาจ

รูปที่ 3.10 ตัวอย่างการจำแนกคำยืมในข้อความ

ในงานวิจัยนี้ได้ดำเนินการระบุตำแหน่งและจำแนกประเภทคำยืมด้วยการเขียนชุดคำสั่งภาษาโปรแกรม PHP โดยประเภทของคำยืมที่ใช้จำแนกได้กำหนดขึ้นจากประเภทของคำยืมที่เป็นไปได้ แสดงดังตารางที่ 3.4

ตารางที่ 3.4 ประเภทของคำยืมและไม่ใช่คำยืมสำหรับการจำแนก

ประเภทคำยืม	นิยาม	ตัวอย่าง
บาลี (P)	คำที่ยืมจากภาษาบาลี	ปัจจุบัน
สันสกฤต (S)	คำที่ยืมจากภาษาสันสกฤต	กัญชายน
ตัวการ์นต์ (O)	คำที่มีตัวการ์นต์	ออนไลน์
บาลีและสันสกฤต (PS)	คำที่ยืมจากภาษาบาลีและสันสกฤต	ประสิทธิภาพ
บาลีและตัวการ์นต์ (PO)	คำที่ยืมจากภาษาบาลีและมีตัวการ์นต์	กลยุทธ์
สันสกฤตและตัวการ์นต์ (SO)	คำที่ยืมจากภาษาสันสกฤตและมีตัวการ์นต์	วิเคราะห์
บาลี สันสกฤต และตัวการ์นต์ (PSO)	คำที่ยืมจากภาษาบาลี สันสกฤต และมีตัวการ์นต์	ประสิทธิ์
คำอื่น ๆ (Other)	คำในภาษาไทยที่ไม่ใช่คำยืมจากภาษาบาลี สันสกฤต และคำที่ไม่มีตัวการ์นต์	ข้อมูล

จากตารางที่ 3.4 แสดงให้เห็นถึงประเภทของคำยืมที่ใช้สำหรับจำแนกและประเภทคำในภาษาไทยที่ไม่ใช่คำยืมจากภาษาบาลี สันสกฤต และคำที่ไม่มีตัวการ์นต์ โดยมีความหมายและตัวอย่างคำของแต่ละประเภท ยกตัวอย่างประเภทของคำยืมจากบาลี สันสกฤต และตัวการ์นต์ หมายถึง คำไทยที่มีตัวการ์นต์ โดยยืมคำมาจากภาษาบาลีและสันสกฤต ตัวอย่างคำ เช่น ประสิทธิ์ (ปร คือ รูปแบบของคำสันสกฤต ทธ คือ รูปแบบของคำบาลี และ ธิ คือ รูปแบบของคำยืมจากภาษาอื่น ๆ)

3.1.5 การสกัดคำยืมด้วยผู้เชี่ยวชาญ

ขั้นตอนนี้ คือ การจำแนกประเภทคำยืมด้วยผู้เชี่ยวชาญเพื่อเตรียมชุดของรายการคำยืมที่ถูกต้องสำหรับการประเมินผลแบบจำลองที่พัฒนาขึ้น โดยเริ่มจากผู้เชี่ยวชาญจำแนกประเภทคำยืมจากเอกสารข้อความที่ผ่านการจำแนกคำ ผลลัพธ์ คือ รายการคำยืมที่ถูกต้อง

งานวิจัยนี้ได้กำหนดเกณฑ์สำหรับจำแนกประเภทของผู้เชี่ยวชาญดังตารางที่ 3.4 ซึ่งผลสำหรับการจำแนกประเภทของคำแสดงดังตารางที่ 3.5

ตารางที่ 3.5 การจำแนกประเภทคำโดยผู้เชี่ยวชาญ

ประเภทของคำ	หัวเรื่องของเอกสาร (คำ)	
	iOS operating system	Data mining
บาลี (P)	1	4
สันสกฤต (S)	11	18
การันต์ (O)	14	4
บาลีและสันสกฤต (PS)	1	1
บาลีและการันต์ (PO)	0	5
สันสกฤตและการันต์ (SO)	6	24
บาลี สันสกฤต และการันต์ (PSO)	0	0
คำอื่น ๆ (Other)	645	1,905

3.1.6 การประเมินผลแบบจำลอง

ขั้นตอนนี้ คือ การประเมินผลการจำแนกประเภทคำยืมของแบบจำลอง โดยเริ่มจากการเปรียบเทียบการจำแนกประเภทระหว่างแบบจำลองและผู้เชี่ยวชาญที่ได้จำแนกในตารางที่ 3.5 เพื่อหาจำนวนของคำยืมที่จำแนกถูกต้องและไม่ถูกต้องของแบบจำลอง และสุดท้ายดำเนินการคำนวณผลเพื่อหาค่าความถูกต้องในการจำแนกประเภทคำยืมของแบบจำลองที่พัฒนาขึ้น ผลลัพธ์ คือ ผลของความถูกต้องในการจำแนกประเภทคำยืมของแบบจำลอง

ตารางที่ 3.6 Confusion matrix (Han, *et al.*, 2011)

Actual class / Predict class	C1	¬ C1	Total
C1	True Positives (TP)	False Negatives (FN)	P
¬ C1	False Positives (FP)	True Negatives (TN)	N
Total	P'	N'	P + N

งานวิจัยนี้ได้คำนวณหาค่าความถูกต้องโดยใช้ตาราง Confusion matrix ดังที่ได้แสดงในตารางที่ 3.6 โดยที่

Positives (P) คือ จำนวนคำยืมทั้งหมดที่จำแนกโดยผู้เชี่ยวชาญ

Negatives (N) คือ จำนวนคำทั้งหมดที่ไม่ใช่คำยืมที่จำแนกโดยผู้เชี่ยวชาญ

True positives (TP) คือ จำนวนคำยืมที่จำแนกถูกต้องด้วยแบบจำลองที่พัฒนาขึ้นเมื่อเปรียบเทียบกับ P

False positives (FP) คือ จำนวนคำยืมที่จำแนกไม่ถูกต้องด้วยแบบจำลองที่พัฒนาขึ้นเมื่อเปรียบเทียบกับ P

True negatives (TN) คือ จำนวนคำที่ไม่ใช่คำยืมที่จำแนกถูกต้องด้วยแบบจำลองที่พัฒนาขึ้นเมื่อเปรียบเทียบกับ N

False negatives (FN) คือ จำนวนคำที่ไม่ใช่คำยืมที่จำแนกไม่ถูกต้องด้วยแบบจำลองที่พัฒนาขึ้นเมื่อเปรียบเทียบกับ N

$$Accuracy = \frac{TP+TN}{P+N} \quad (3.1)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3.2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.3)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.4)$$

Accuracy คือ การวัดความถูกต้องของแบบจำลอง โดยพิจารณาจากคำชี้มทุกประเภท คำนวณโดยสมการ (3.1)

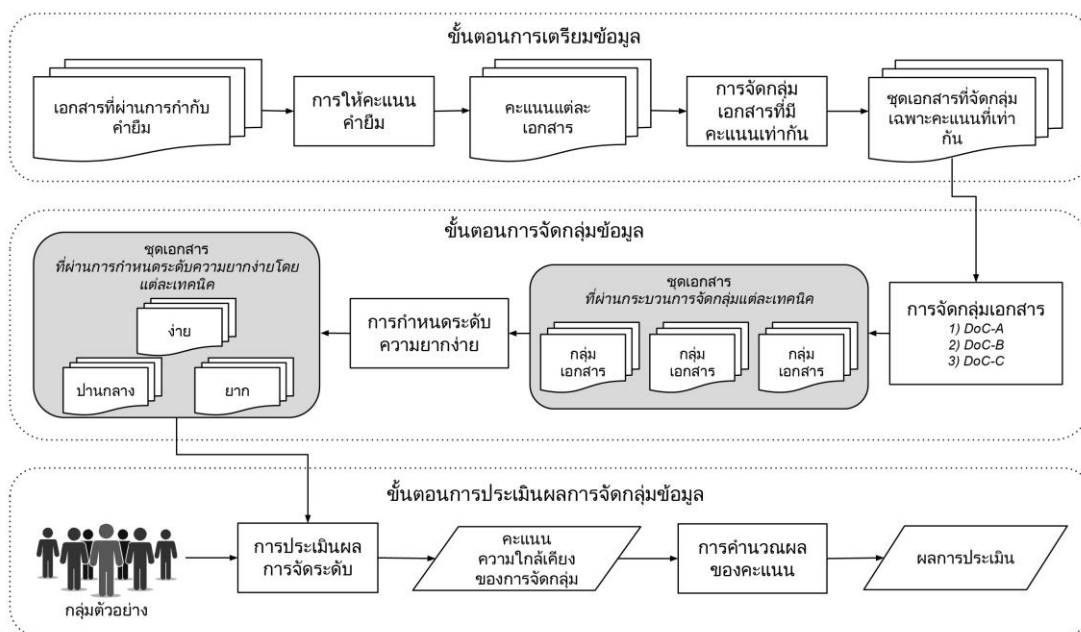
F – measure คือ การวัดค่า Precision และ Recall พร้อมกันของแบบจำลอง โดยพิจารณาจากคำชี้มแต่ละประเภท คำนวณโดยสมการ (3.2)

Precision คือ การวัดความแม่นยำของแบบจำลอง โดยพิจารณาจากคำชี้มแต่ละประเภท คำนวณโดยสมการ (3.3)

Recall คือ การวัดความถูกต้องของแบบจำลอง โดยพิจารณาจากคำชี้มแต่ละประเภท คำนวณโดยสมการ (3.4)

3.2 การจัดกลุ่มเอกสาร

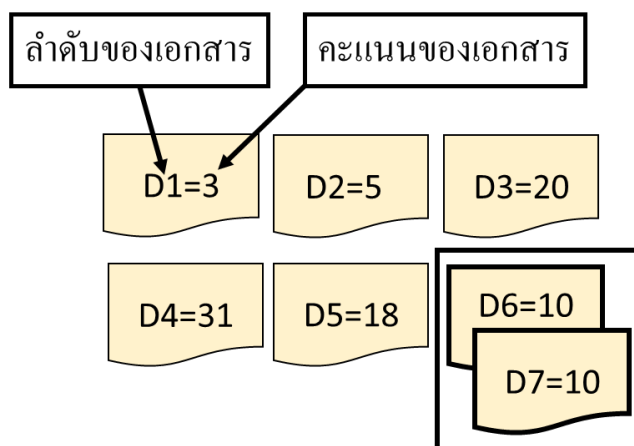
เมื่อได้ชุดเอกสารที่มีการกำกับคำชี้มจากส่วนการสกัดคำชี้ม ในส่วนนี้จะดำเนินการประเมินระดับความยากง่ายของข้อความในเอกสาร คือ การจัดเรียงหรือจัดกลุ่มเอกสารตามระดับความยากง่ายของข้อความที่เกิดขึ้นในเอกสาร งานวิจัยนี้ได้นำเทคนิคการจัดกลุ่มเอกสาร (Text clustering) มาประยุกต์ใช้ในด้านการจัดระดับความยากง่ายสำหรับเอกสาร ซึ่งในส่วนนี้จะอธิบายเกี่ยวกับขั้นตอนในการจัดกลุ่มเอกสารเพื่อประเมินระดับความยากง่ายและขั้นตอนการประเมินผลความใกล้เคียงในการกลุ่มของเทคนิคการจัดกลุ่มเอกสาร แสดงดังรูปที่ 3.11 ซึ่งประกอบด้วย 3 ขั้นตอน ดังนี้



รูปที่ 3.11 การดำเนินการวิจัยการจัดกลุ่มเอกสาร

3.2.1 การเตรียมข้อมูล

ขั้นตอนนี้ คือ การนำข้อมูลเข้าเพื่อเตรียมความพร้อมก่อนเข้าสู่ขั้นตอนการจัดกลุ่ม โดยกระบวนการเริ่มแรก คือ การนำเข้าเอกสารที่มีการระบุคำยืมเพื่อตรวจหาตำแหน่งของคำยืม โดย 1 ตำแหน่งที่ตรวจจับได้มีค่าเท่ากับ 1 คะแนน ทำให้ได้คะแนนของคำยืมในแต่ละเอกสาร ผลลัพธ์ คือ คะแนนคำยืมของแต่ละเอกสาร และกระบวนการถัดมาเป็นการหาคะแนนที่ซ้ำกันหรือเท่ากันของแต่ละเอกสาร เนื่องจากต้องการจัดกลุ่มเอกสารที่มีคะแนนเท่ากันให้อยู่ในกลุ่มเดียวกัน ผลลัพธ์ที่ได้ คือ กลุ่มของเอกสารที่มีคะแนนเท่ากัน แสดงดังรูปที่ 3.12



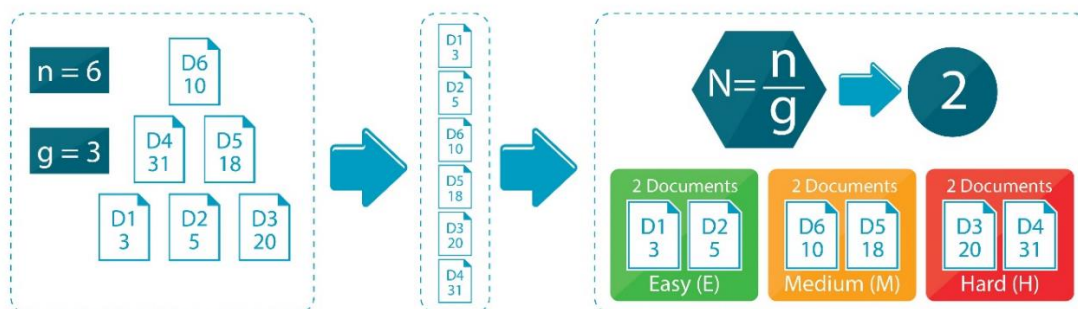
รูปที่ 3.12 ตัวอย่างการจัดกลุ่มเอกสารที่มีคะแนนเท่ากัน

งานวิจัยนี้ได้ใช้เอกสารจากชุดเอกสารกำกับคำยืมที่ได้จากส่วนการสกัดคำยืมจำนวน 20 ฉบับ แบ่งเป็นเอกสารในหัวเรื่อง Data mining จำนวน 10 ฉบับ และหัวเรื่อง iOS operating system จำนวน 10 ฉบับ และดำเนินการให้คะแนนคำยืมและจัดกลุ่มเอกสารเฉพาะเอกสารที่มีคะแนนเท่ากันด้วยการเขียนชุดคำสั่งภาษาโปรแกรม PHP

3.2.2 การจัดกลุ่มข้อมูล

ขั้นตอนนี้ คือ การจัดกลุ่มให้กับเอกสาร โดยเริ่มจากการนำเอกสารที่ผ่านขั้นตอนการเตรียมข้อมูลเข้าสู่กระบวนการจัดกลุ่มเอกสารในแต่ละเทคนิค ซึ่งเทคนิคการจัดกลุ่มเอกสารที่งานวิจัยนี้ได้พัฒนาและนำเสนอประกอบด้วย 3 เทคนิค ดังนี้

3.2.2.1 การจัดกลุ่มเอกสารที่คำนวณจากสัดส่วนของจำนวนเอกสารทั้งหมด (Document Clustering A: DoC-A) คือ การจัดกลุ่มเอกสารด้วยเทคนิค DoC-A ดำเนินการโดยการหาเศษส่วนเพื่อกำหนดจำนวนสมาชิก (เอกสาร) ในแต่ละกลุ่มเพื่อให้ได้จำนวนเอกสารในแต่ละกลุ่มใกล้เคียงกันมากที่สุด แสดงดังรูปที่ 3.13 โดยมีขั้นตอน และเงื่อนไขดังนี้



รูปที่ 3.13 เทคนิคการจัดกลุ่มเอกสาร DoC-A

- 1) จัดเรียงเอกสารตามคะแนนจากน้อยไปมาก
- 2) หาจำนวนสมาชิก (เอกสาร) ของแต่ละกลุ่มโดยใช้สมการ (3.5)

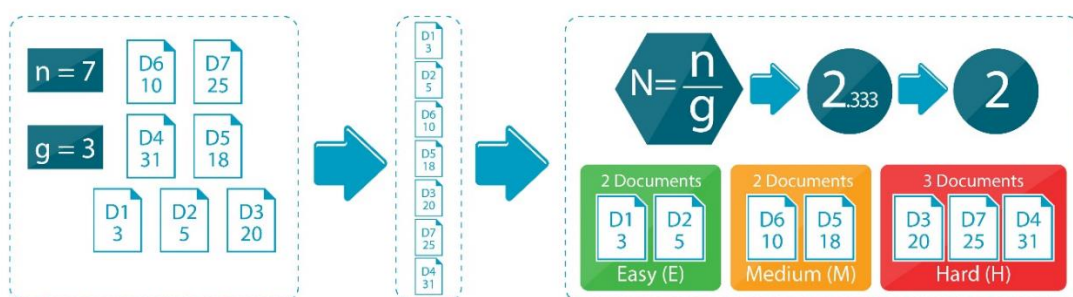
$$N = \frac{n}{g} \quad (3.5)$$

โดยที่ N คือ จำนวนสมาชิกที่จะต้องกำหนดให้ในแต่ละกลุ่ม
 n คือ จำนวนของเอกสารทั้งหมด

g คือ จำนวนกลุ่มที่ต้องการแบ่งเอกสาร และในงานวิจัยนี้

กำหนดให้ g เท่ากับ 3 กลุ่ม คือ ระดับง่าย (Easy) ระดับปานกลาง (Medium) และระดับยาก (Hard)

- 3) นำค่า N ที่ได้จากสมการ (3.5) มาเป็นเกณฑ์ในการกำหนดจำนวนสมาชิกของแต่ละกลุ่ม ซึ่งมีเกณฑ์ ดังนี้

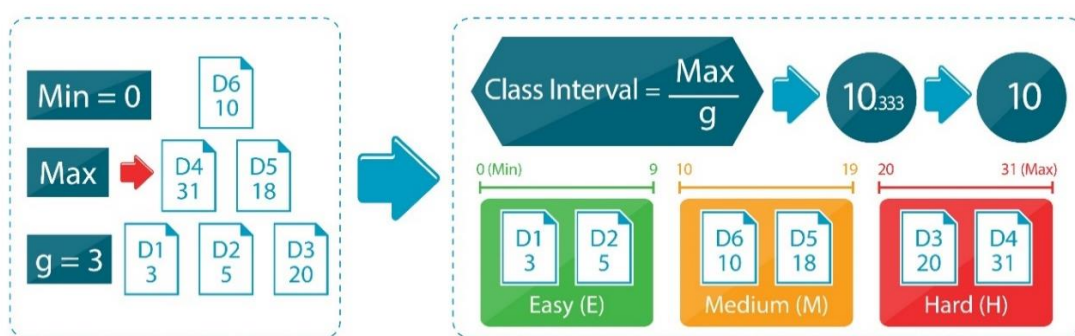


รูปที่ 3.14 ตัวอย่างในกรณีค่า N เป็นจำนวนจริงของ DoC-A

3.1) หากค่า N เป็นจำนวนจริงที่น้อยกว่า 0.5 ให้ลดค่า N เป็นจำนวนเต็ม และหากมากกว่าหรือเท่ากับ 0.5 ให้เพิ่มค่า N เป็นจำนวนเต็ม ดังที่ได้แสดงในรูปที่ 3.14

3.2) ดำเนินการแบ่งจำนวนสมาชิกให้กับกลุ่ม Easy และ Medium ก่อนตามลำดับ โดยแบ่งตามค่า N และสุดท้ายแบ่งให้กับ Hard ตามจำนวนสมาชิกที่คงเหลือ ดังที่ได้แสดงในรูปที่ 3.14

3.2.2.2 การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุด (Document Clustering B: DoC-B) คือ การจัดกลุ่มเอกสารด้วยเทคนิค DoC-B ดำเนินการโดยการหาช่วงของอันตรภาคชั้นในแต่ละกลุ่มจากค่าคะแนนที่มากและน้อยที่สุดของเอกสารที่มีโอกาสเกิดขึ้นจริง แสดงดังรูปที่ 3.15 โดยมีขั้นตอนและเงื่อนไขดังนี้



รูปที่ 3.15 เทคนิคการจัดกลุ่มเอกสาร DoC-B

1) หาช่วงของอันตรภาคชั้นของแต่ละกลุ่มโดยใช้สมการ (3.6)

$$Class\ Interval = \frac{Max}{g} \quad (3.6)$$

โดยที่ Class Interval คือ ความกว้างของอันตรภาคชั้น

Max คือ คะแนนของเอกสารที่มีค่ามากที่สุด

g คือ จำนวนกลุ่ม หรือจำนวนชั้นที่ต้องการแบ่ง

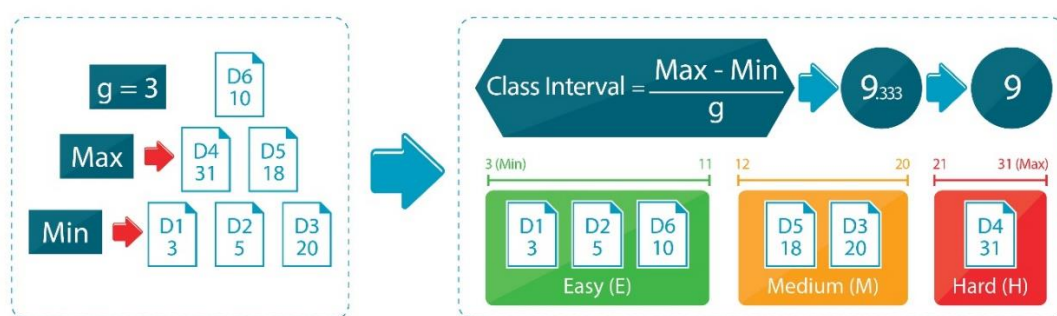
เอกสาร และในงานวิจัยนี้กำหนดให้ g เท่ากับ 3 กลุ่ม คือ Easy Medium และ Hard

2) หากค่า Range เป็นจำนวนจริงที่น้อยกว่า 0.5 ให้ลดค่า Range เป็นจำนวนเต็ม และหากมากกว่าหรือเท่ากับ 0.5 ให้เพิ่มค่า Range เป็นจำนวนเต็ม

3) นำค่า Range แบ่งช่วงของอันตรายภคชั้นในแต่ละกลุ่ม ซึ่งเขตข้อมูลจะเริ่มต้นที่ 0 ถึง Max โดยเริ่มแบ่งช่วงให้กับ Easy และ Medium ให้เท่ากันตามลำดับ และสุดท้ายแบ่งช่วงให้กับ Hard ตามจำนวนช่วงที่เหลือ

4) แบ่งเอกสารที่มีคะแนนของคำยืมตามช่วงของอันตรายภคชั้นที่สร้างขึ้น

3.2.2.3 การจัดกลุ่มเอกสารด้วยอันตรายภคชั้นที่คำนวณจากค่าที่มากที่สุดและน้อยที่สุดที่เกิดขึ้นจริง (Document Clustering C: DoC-C) คือ การจัดกลุ่มเอกสารด้วยเทคนิค DoC-C ดำเนินการ โดยการสร้างอันตรายภคชั้นสำหรับแต่ละกลุ่มจากค่าคะแนนคำยืมที่มากที่สุดและน้อยที่สุดของเอกสารที่เกิดขึ้นจริง แสดงดังรูปที่ 3.16 โดยมีขั้นตอนและเงื่อนไขดังนี้



รูปที่ 3.16 เทคนิคการจัดกลุ่มเอกสาร DoC-C

1) หาช่วงของอันตรายภคชั้นของแต่ละกลุ่มโดยใช้สมการ (3.7)

$$Class\ Interval = \frac{Max - Min}{g} \quad (3.7)$$

โดยที่ Class Interval คือ ความกว้างของอันตรายภคชั้น

Max คือ คะแนนของเอกสารที่มีค่ามากที่สุด

Min คือ คะแนนของเอกสารที่มีค่าน้อยที่สุด

g คือ จำนวนกลุ่ม หรือจำนวนชั้นที่ต้องการแบ่งเอกสาร และในงานวิจัยนี้กำหนดให้ g เท่ากับ 3 กลุ่ม คือ Easy Medium และ Hard

2) หากค่า Range เป็นจำนวนจริงที่น้อยกว่า 0.5 ให้ลดค่า Range เป็นจำนวนเต็ม และหากมากกว่าหรือเท่ากับ 0.5 ให้เพิ่มค่า Range เป็นจำนวนเต็ม

3) นำค่า Range แบ่งช่วงของอันตรภาคชั้นในแต่ละกลุ่ม ซึ่งเขตข้อมูลจะเริ่มต้นที่ Min ถึง Max โดยเริ่มแบ่งช่วงให้กับ Easy และ Medium ให้เท่ากันตามลำดับ และสุดท้ายแบ่งช่วงให้กับ Hard ตามจำนวนช่วงที่เหลือ

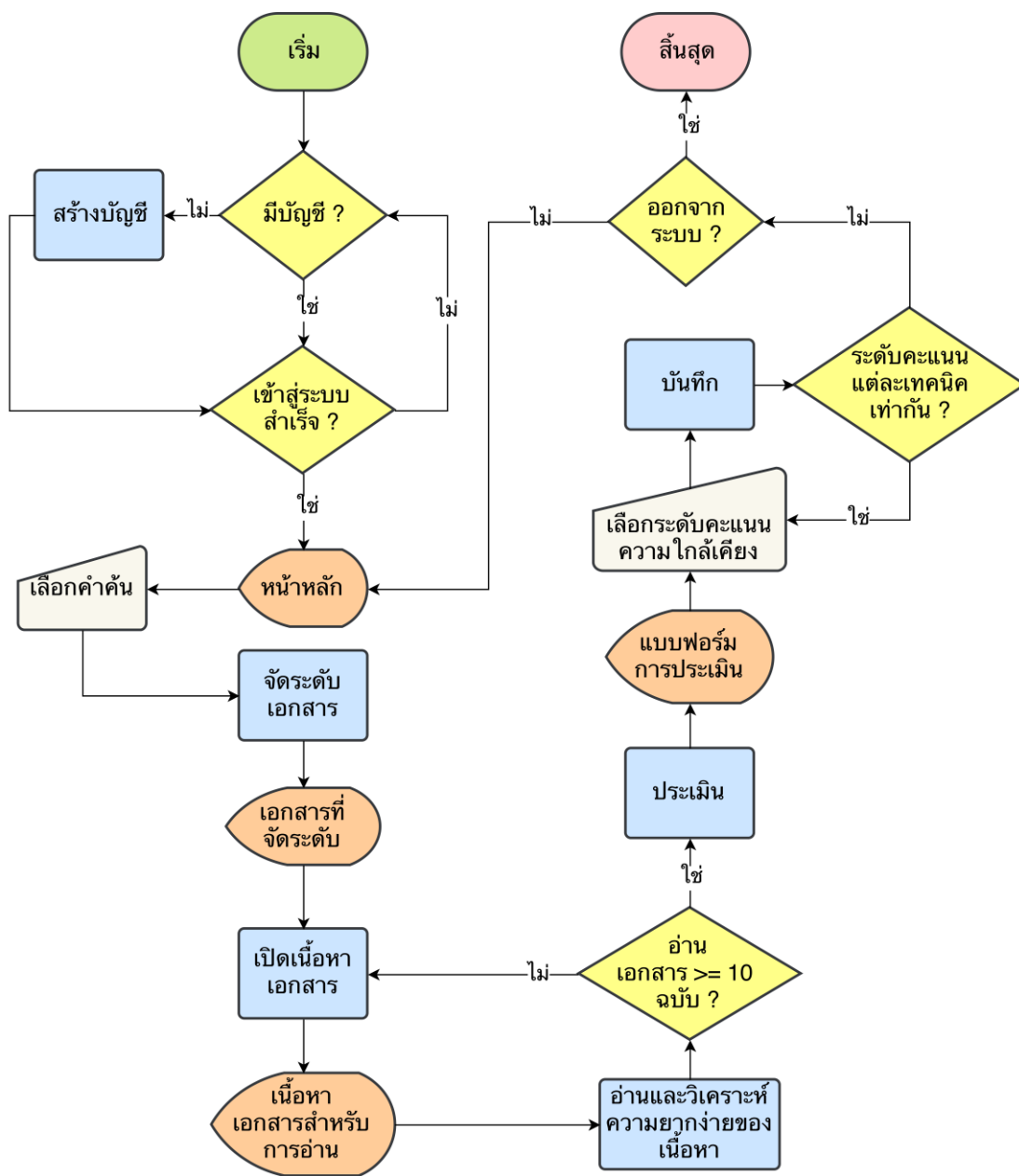
4) แบ่งเอกสารที่มีคะแนนของคำยืมตามช่วงของอันตรภาคชั้นที่สร้างขึ้น

ผลลัพธ์ที่ได้จากแต่ละเทคนิค คือ กลุ่มเอกสารตามระดับความยากง่าย โดยมีจำนวนกลุ่มตามที่ได้กำหนด ซึ่งจำนวนกลุ่มที่ได้กำหนดในการวิจัยนี้คือ 3 กลุ่ม กระบวนการต่อมา คือ การกำหนดระดับความยากง่ายให้กับเอกสารที่ได้ถูกแบ่งกลุ่มในกระบวนการก่อนหน้า ซึ่งในงานวิจัยนี้ได้กำหนดระดับความยากง่ายออกเป็น 3 ระดับ คือ Easy Medium และ Hard และนิยามของความยากง่าย คือ ระดับความยากง่ายของเอกสารขึ้นกับจำนวนของคำยืมในเอกสาร ผลลัพธ์ที่ได้จากกระบวนการนี้ คือ ชุดเอกสารแต่ละชุดที่ถูกระบุระดับความยากง่าย โดยเอกสาร 1 กลุ่มจะถูกกำหนดระดับความยากง่ายเพียง 1 ระดับ

3.2.3 การประเมินผลการจัดกลุ่มข้อมูล

ขั้นตอนนี้ คือ การประเมินประสิทธิภาพการจัดกลุ่มข้อมูลของเทคนิคการจัดระดับความยากง่ายของเอกสารแต่ละเทคนิค เนื่องจากงานวิจัยนี้ต้องการหาเทคนิคในการจัดระดับความยากง่ายของเอกสารที่เหมาะสมและใกล้เคียงกับระดับความยากง่ายในการอ่านเอกสารของมนุษย์มากที่สุด ด้วยเหตุนี้ผู้วิจัยจึงจำเป็นต้องประเมินประสิทธิภาพในการจัดกลุ่มเอกสารของเทคนิคการจัดกลุ่มเอกสารทั้ง 3 เทคนิค โดยการทดลองเพื่อหาเทคนิคการจัดกลุ่มเอกสารที่มีผลการจัดกลุ่มเอกสารใกล้เคียงกับผลการจัดกลุ่มเอกสารของกลุ่มตัวอย่างมากที่สุด

เนื่องจากพื้นความรู้ (Background knowledge) ของมนุษย์แต่ละคนไม่ได้มีผลต่อการจัดเรียงลำดับเอกสารตามความยากง่าย ดังนั้นงานวิจัยนี้จึงไม่พิจารณาพื้นความรู้ของกลุ่มตัวอย่าง



รูปที่ 3.17 ขั้นตอนการใช้งานระบบประเมินผลการจัดกลุ่มเอกสารของแต่ละเทคนิค

ผู้วิจัยได้สร้างระบบสำหรับประเมินผลของแต่ละเทคนิคสำหรับกลุ่มตัวอย่างจำนวน 20 คน (ผู้ประเมิน) สำหรับขั้นตอนการประเมินดังที่ได้แสดงในรูปที่ 3.17 ประกอบด้วย 6 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ให้ผู้ประเมินเข้าสู่ระบบประเมิน ถัดจากนั้น ให้ผู้ประเมินสร้างบัญชีผู้ใช้เพื่อลงชื่อเข้าสู่ระบบประเมิน แสดงดังรูปที่ 3.18

ขั้นตอนที่ 2 ให้ผู้ประเมินเลือกคำค้น ซึ่งหมายถึง หัวเรื่องของเอกสารที่นำมาทดสอบ โดยงานวิจัยนี้มีหัวเรื่องสำหรับทดสอบทั้งหมด 2 หัวเรื่อง คือ Data mining และ iOS operating system ในแต่ละหัวเรื่องประกอบด้วยเอกสารจำนวน 10 ฉบับ

ขั้นตอนที่ 3 ให้ผู้ประเมินกดปุ่มจัดระดับ และระบบจะแสดงผลพัทธ์การจัดกลุ่มในแต่ละเทคนิค แสดงดังรูปที่ 3.19

ขั้นตอนที่ 4 ให้ผู้ประเมินเปิดอ่านเอกสารแต่ละฉบับเพื่อพิจารณาความใกล้เคียงในการจัดกลุ่มเอกสารของแต่ละเทคนิคกับตนเองมากที่สุด แสดงดังรูปที่ 3.20 โดยกำหนดให้ผู้ประเมินอ่านเอกสารอย่างน้อยทุกฉบับของแต่ละหัวเรื่องเพื่อพิจารณาถึงความใกล้เคียงของผลการจัดกลุ่มเอกสารระหว่างตนเองและเทคนิค ถัดจากนั้นระบบจะแสดงปุ่มประเมิน

ขั้นตอนที่ 5 ให้ผู้ประเมินกดปุ่มประเมิน ระบบจะแสดงส่วนต่อประสานกับผู้ใช้ของการประเมินผล แสดงดังรูปที่ 3.21

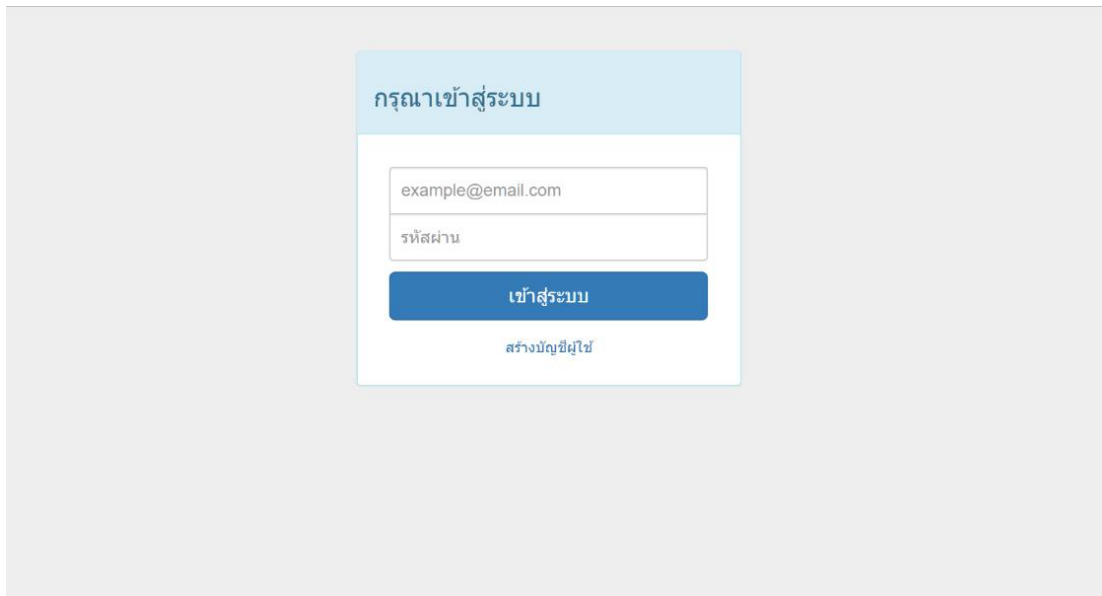
ขั้นตอนที่ 6 ให้ผู้ประเมินดำเนินการให้ระดับคะแนนความใกล้เคียงผลการจัดกลุ่มของเทคนิคทั้ง 3 เทคนิค 1 ถึง 3 โดยกำหนดให้ผู้ประเมินให้ระดับคะแนนแต่ละเทคนิคไม่เท่ากัน ซึ่งระดับคะแนนความใกล้เคียงแต่ละระดับมีความหมายดังนี้

- ระดับ 3 หมายถึง ผลการจัดกลุ่มเอกสารของเทคนิคมีความใกล้เคียงกับผลการจัดกลุ่มเอกสารของผู้ประเมินมากที่สุด

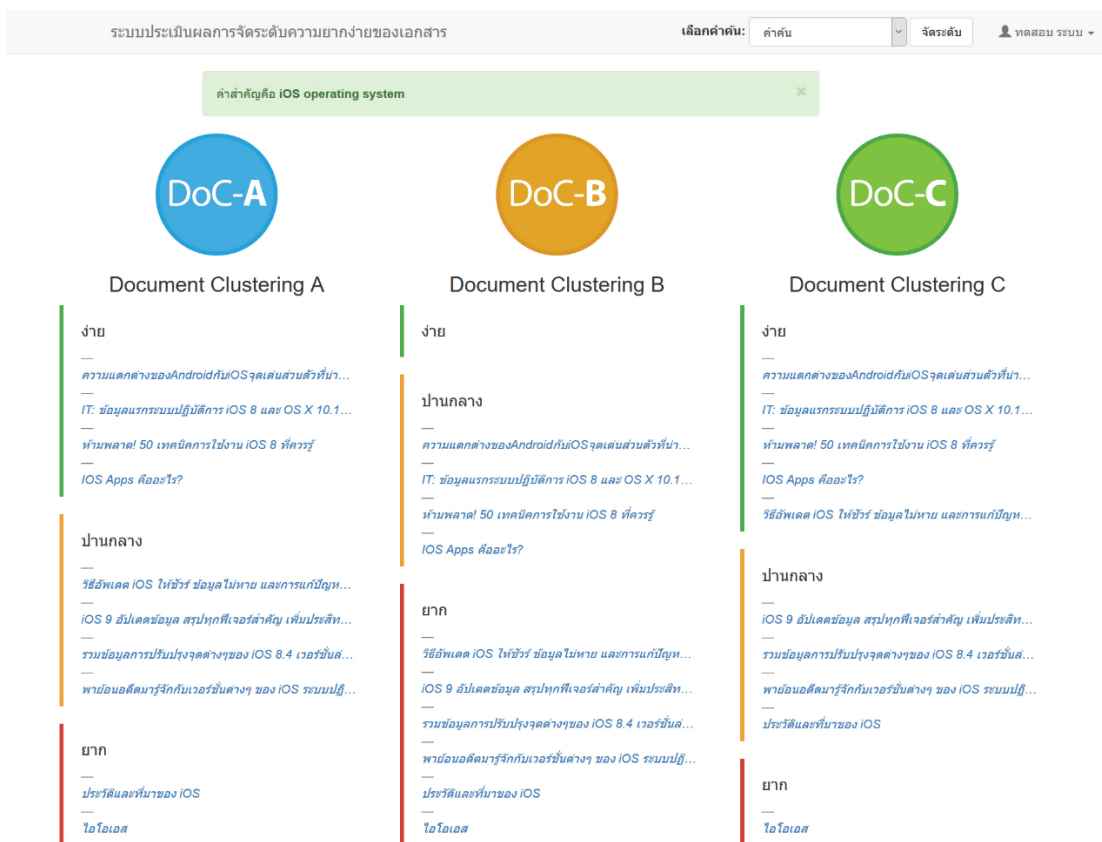
- ระดับ 2 หมายถึง ผลการจัดกลุ่มเอกสารของเทคนิค และผลการจัดกลุ่มเอกสารของผู้ประเมิน มีความใกล้เคียงในการจัดกลุ่มเอกสารน้อยกว่าระดับ 3 แต่ไม่น้อยกว่าระดับ 1 (ปานกลาง)

- ระดับ 1 หมายถึง ผลการจัดกลุ่มเอกสารของเทคนิคมีความใกล้เคียงกับผลการจัดกลุ่มเอกสารของผู้ประเมินน้อยที่สุด

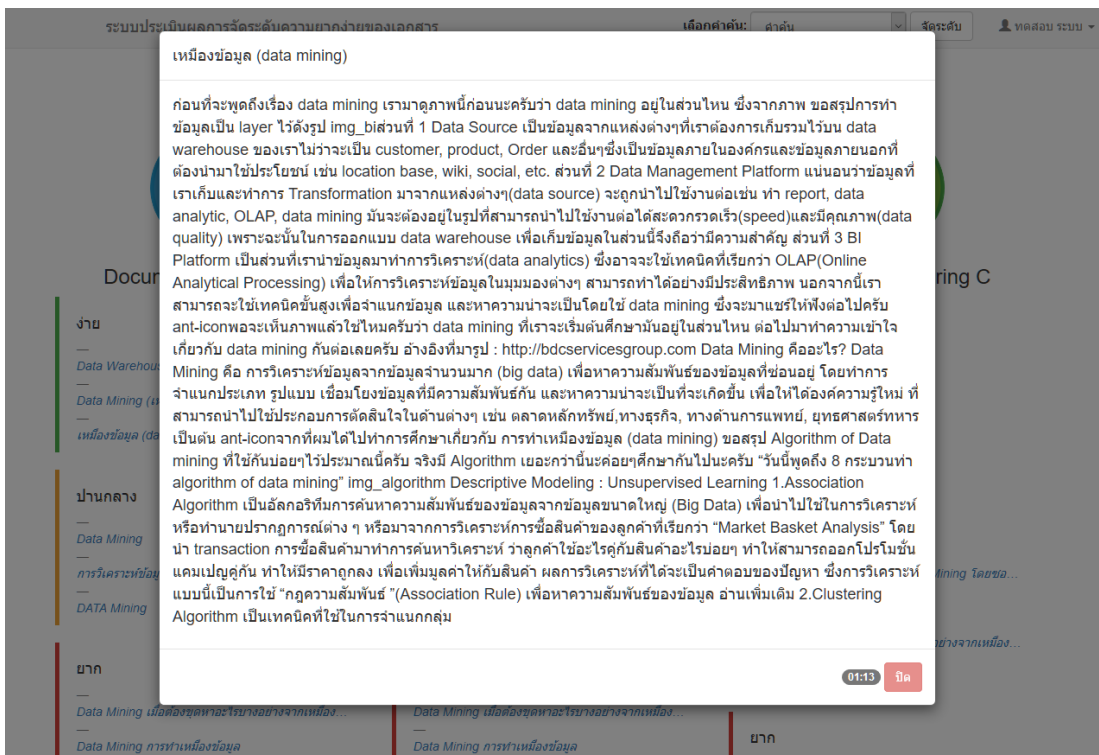
ผลลัพธ์ที่ได้จากกระบวนการนี้ คือ ระดับคะแนนความใกล้เคียงในการจัดกลุ่มระหว่างเทคนิคการจัดกลุ่มและกลุ่มตัวอย่าง กระบวนการต่อมา คือ การนำผลการประเมินในข้างต้นมาคำนวณหาผลรวมแต่ละเทคนิคและค่าเฉลี่ยทั้งหมด เพื่อนำผลมาสรุปและดูแนวโน้มของผลลัพธ์



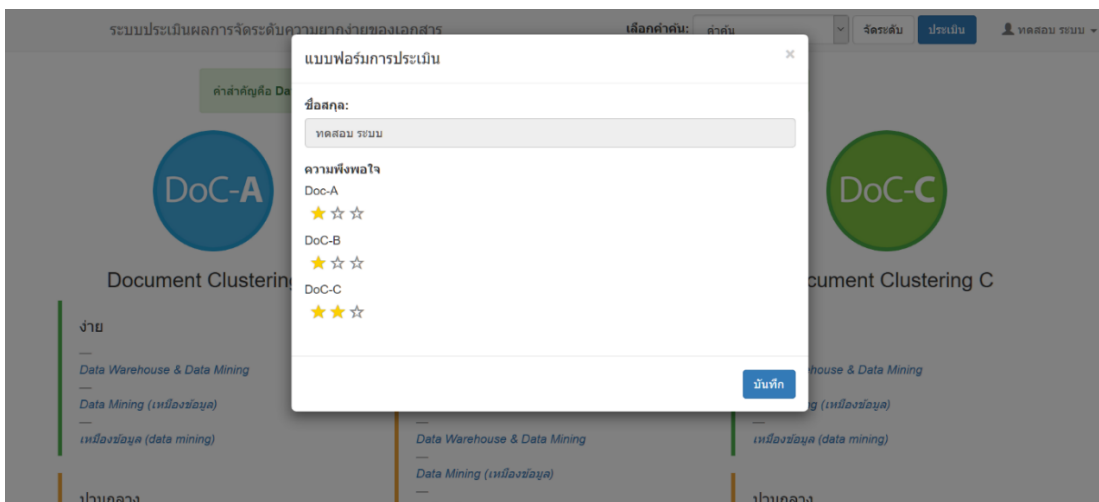
รูปที่ 3.18 ส่วนต่อประสานกับผู้ใช้ในการเข้าใช้งานระบบประเมิน



รูปที่ 3.19 ส่วนต่อประสานกับผู้ใช้ของผลการจัดกลุ่มเอกสารแต่ละเทคนิคในระบบประเมิน



รูปที่ 3.20 ส่วนต่อประสานกับผู้ใช้ของเนื้อหาในเอกสารในระบบประเมิน



รูปที่ 3.21 ส่วนต่อประสานกับผู้ใช้ของการประเมินผลในระบบประเมิน

3.3 การประเมินความยากง่ายของข้อความแบบอัตโนมัติ

ส่วนนี้สามารถดำเนินการโดยการนำขั้นตอนและกระบวนการในการสกัดคำยืม และการจัดกลุ่มเอกสารเพื่อพัฒนาการประเมินความยากง่ายของข้อความในเอกสารภาษาไทยแบบอัตโนมัติ แสดงดังรูปที่ 3.1 และได้อธิบายขั้นตอนและกระบวนการต่าง ๆ ในข้อที่ 3.1 และ 3.2

เพื่อนำงานวิจัยไปใช้ให้เกิดประโยชน์ ดังนั้นงานวิจัยนี้ได้พัฒนาระบบจากการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยแบบอัตโนมัติ โดยมีขั้นตอนการทำงานของระบบที่แสดงในรูปที่ 3.2 ประกอบด้วย 7 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 ผู้ใช้สร้างบัญชีผู้ใช้ สำหรับผู้ที่ไม่มีบัญชีผู้ใช้ ในกรณีที่ผู้ใช้มีบัญชีอยู่แล้วสามารถทำขั้นตอนถัดไป

ขั้นตอนที่ 2 ผู้ใช้ลงชื่อเข้าใช้ระบบ (Sign in) โดยกรอกที่อยู่อีเมลและรหัสผ่านที่ได้กำหนดในขั้นตอนก่อนหน้า ในกรณีที่ผู้ใช้จำอีเมลและรหัสผ่านไม่ได้ให้ดำเนินการสมัครใหม่ เมื่อผู้ใช้งานการลงชื่อเข้าใช้สำเร็จระบบแสดงหน้าหลักและแจ้งผู้ใช้สำหรับการลงชื่อเข้าใช้สำเร็จ

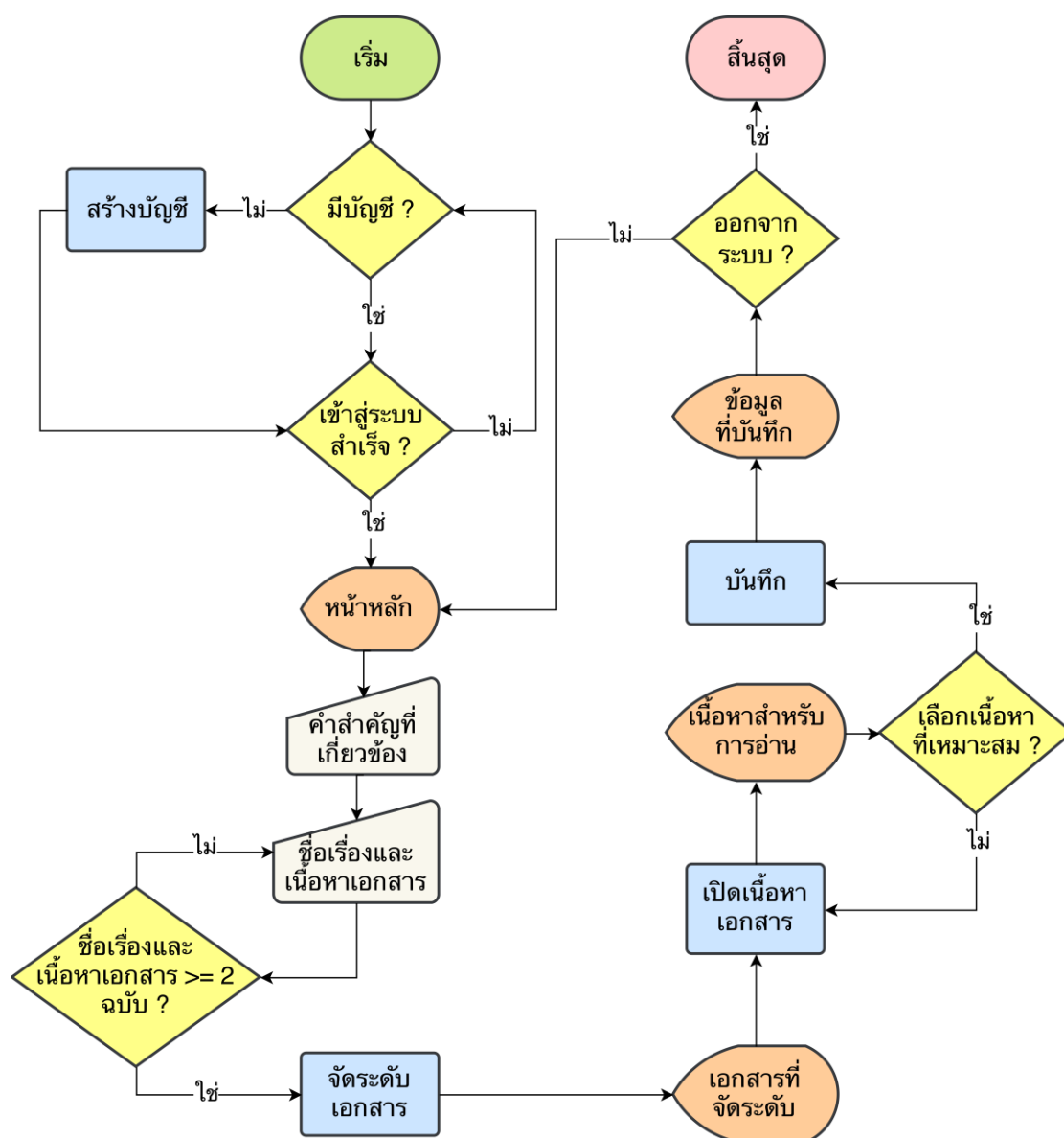
ขั้นตอนที่ 3 ผู้ใช้ใส่คำสำคัญที่เกี่ยวข้อง ชื่อเรื่อง และเนื้อหาของชุดเอกสารแต่ละฉบับสำหรับการประเมินระดับความยากง่าย โดยชื่อเรื่องและเนื้อหาของเอกสารจะต้องมากกว่า 1 ฉบับ

ขั้นตอนที่ 4 ผู้ใช้ดำเนินการกดปุ่มจัดระดับ ต่อมาระบบแสดงผลการจัดกลุ่มเอกสารตามระดับความยากง่ายของข้อความในเอกสาร

ขั้นตอนที่ 5 ผู้ใช้เปิดเนื้อหาเอกสารเพื่อค้นหาเอกสารที่เหมาะสมกับตนเอง ต่อมาระบบแสดงเนื้อหาของเอกสารที่ผู้ใช้เปิด ในขั้นตอนนี้ผู้ใช้ต้องอ่านเอกสารจนกระทั่งค้นพบเอกสารที่เหมาะสมกับตนเอง คำแนะนำ ผู้ใช้ไม่จำเป็นต้องอ่านเอกสารหมดทุกฉบับ แต่ผู้ใช้สามารถอ่านเอกสารฉบับใดฉบับหนึ่งในแต่ละกลุ่มเพื่อความรวดเร็วในการเลือก เมื่อค้นพบเอกสารที่เหมาะสมกับตนเอง ผู้ใช้ต้องดำเนินการคลิกเลือกเอกสาร และดำเนินการในขั้นตอนต่อไป

ขั้นตอนที่ 6 ผู้ใช้บันทึกข้อมูลทั้งหมด เช่น คำสำคัญที่เกี่ยวข้อง ชื่อเรื่องของเอกสาร เนื้อหาของเอกสาร เอกสารที่ผู้ใช้เลือกอ่าน ต่อมาระบบดำเนินการบันทึกข้อมูลทั้งหมด และแจ้งผู้ใช้ถึงการดำเนินการบันทึกข้อมูลเสร็จสิ้น และรอดำเนินการในขั้นตอนต่อไป

ขั้นตอนที่ 7 ผู้ใช้สามารถดำเนินการในขั้นตอนที่ 3 ถึง 6 ซ้ำได้ไม่จำกัดจำนวนครั้ง หากต้องการหยุดการใช้งานระบบ ผู้ใช้ต้องดำเนินการกดปุ่มออกจากระบบ



รูปที่ 3.22 ขั้นตอนการใช้งานระบบการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยแบบอัตโนมัติ

บทที่ 4

ผลการทดลองและการอภิปรายผล

ในบทนี้กล่าวถึงผลการทดลองและการอภิปรายเกี่ยวกับผลการทดลองของงานวิจัยนี้ โดยผลการทดลองของงานวิจัยนี้ คือ ผลลัพธ์การประเมินการสกัดคำยืม และการประเมินความใกล้เคียงในการจัดกลุ่มเอกสารของเทคนิคการจัดกลุ่มเอกสาร โดยแบ่งการอธิบายผลการทดลองและการอภิปรายของทั้งสองในหัวข้อ 4.1 และ 4.2 ตามลำดับ

4.1 ผลการทดลองการสกัดคำยืมและการอภิปรายผล

ในหัวข้อนี้อธิบายเกี่ยวกับผลการทดลองของการสกัดคำยืม ซึ่งเป็นการทดลองเพื่อหาค่าความถูกต้องและความแม่นยำของแบบจำลองการสกัดคำยืม โดยผลการทดลองของการหาความแม่นยำในการสกัดคำยืม แสดงดังตารางที่ 4.1 ถึงตารางที่ 4.3 และผลการทดลองของการหาความถูกต้องในการสกัดคำยืม แสดงดังตารางที่ 4.4

ตารางที่ 4.1 ค่าเฉลี่ย Precision ของการสกัดคำยืม

คำยืม	ร้อยละของค่าเฉลี่ย Precision ของชุดเอกสารในแต่ละหัวเรื่อง		ร้อยละของค่าเฉลี่ย Precision ของคำยืม
	Data mining	iOS operating system	
บาลี	26.67	60.00	43.33
สันสกฤต	90.00	95.00	92.50
ตัวการ์นต์	100.00	100.00	100.00
บาลีและสันสกฤต	50.00	100.00	75.00
บาลีและตัวการ์นต์	100.00	60.00	80.00
สันสกฤตและตัวการ์นต์	100.00	60.00	80.00
บาลี สันสกฤต และตัวการ์นต์	100.00	100.00	100.00
ค่าเฉลี่ย Precision ของหัวเรื่อง	80.95	82.14	

จากตารางที่ 4.1 ได้แสดงค่า Precision ซึ่งแบ่งออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ร้อยละของค่าเฉลี่ย Precision ของชุดเอกสารในแต่ละหัวเรื่อง คือ ร้อยละของค่าเฉลี่ย Precision จากชุดเอกสารในแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ในหัวเรื่อง Data mining ได้แก่ ประเภทตัวการ์นต์ และประเภทบาลี สันสกฤต และตัวการ์นต์ และในหัวเรื่อง iOS operating system ได้แก่ ประเภทตัวการ์นต์ ประเภทบาลีและสันสกฤต และประเภทบาลี สันสกฤต และตัวการ์นต์ ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุดในหัวเรื่อง Data mining ได้แก่ ประเภทบาลี ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 26.67 และในหัวเรื่อง iOS operating system ได้แก่ ประเภทบาลี ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 60.00

ส่วนที่ 2 ร้อยละของค่าเฉลี่ย Precision ของคำยืม คือ ค่าเฉลี่ย Precision จากหัวเรื่องของเอกสารแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง ซึ่งแสดงผลเป็นร้อยละ โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ได้แก่

ประเภทตัวการ์นต์ และประเภทบาลี สันสกฤต และการ์นต์ ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุด ได้แก่ ประเภทบาลี ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 43.33

ส่วนที่ 3 ค่าเฉลี่ย Precision ของหัวเรื่อง คือ ค่าเฉลี่ย Precision จากการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลองในแต่ละหัวเรื่องของเอกสาร โดยหัวเรื่องที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 82.14 ได้แก่ iOS operating system

ตารางที่ 4.2 ค่าเฉลี่ย Recall ของการสกัดคำยืม

คำยืม	ร้อยละของค่าเฉลี่ย Recall ของชุดเอกสารในแต่ละหัวเรื่อง		ร้อยละของค่าเฉลี่ย Recall ของคำยืม
	Data mining	iOS operating system	
บาลี	100.00	100.00	100.00
สันสกฤต	100.00	100.00	100.00
ตัวการ์นต์	40.00	81.00	60.50
บาลีและสันสกฤต	100.00	100.00	100.00
บาลีและตัวการ์นต์	100.00	100.00	100.00
สันสกฤตและตัวการ์นต์	100.00	100.00	100.00
บาลี สันสกฤต และตัวการ์นต์	100.00	100.00	100.00
ค่าเฉลี่ย Recall ของหัวเรื่อง	91.43	97.29	

จากตารางที่ 4.2 ได้แสดงค่า Recall ซึ่งแบ่งออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ร้อยละของค่าเฉลี่ย Recall ของชุดเอกสารในแต่ละหัวเรื่อง คือ ร้อยละของค่าเฉลี่ย Recall จากชุดเอกสารในแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ในหัวเรื่อง Data mining และ iOS operating system ได้แก่ ประเภทบาลี ประเภทสันสกฤต ประเภทบาลีและสันสกฤต ประเภทบาลีและตัวการ์นต์ ประเภทสันสกฤตและตัวการ์นต์ และประเภทบาลี สันสกฤต

และตัวการ์นต์ ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุดในหัวเรื่อง Data mining และ iOS operating system ได้แก่ ประเภทตัวการ์นต์ ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 40.00 และ 81.00 ตามลำดับ

ส่วนที่ 2 ร้อยละของค่าเฉลี่ย Recall ของคำยืม คือ ค่าเฉลี่ย Recall จากหัวเรื่องของเอกสารแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง ซึ่งแสดงผลเป็นร้อยละ โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ได้แก่ ประเภทบาลี ประเภทสันสกฤต ประเภทบาลีและสันสกฤต ประเภทบาลีและตัวการ์นต์ ประเภทสันสกฤตและตัวการ์นต์ และประเภทบาลี สันสกฤต และตัวการ์นต์ ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุดคือ ตัวการ์นต์ ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 60.50

ส่วนที่ 3 ค่าเฉลี่ย Recall ของหัวเรื่อง คือ ค่าเฉลี่ย Recall จากการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลองในแต่ละหัวเรื่องของเอกสาร โดยหัวเรื่องที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 97.29 ได้แก่ iOS operating system

ตารางที่ 4.3 ค่าเฉลี่ย F-measure ของการสกัดคำยืม

คำยืม	ร้อยละของค่าเฉลี่ย F-measure ของชุดเอกสารในแต่ละหัวเรื่อง		ร้อยละของค่าเฉลี่ย F-measure ของคำยืม
	Data mining	iOS operating system	
บาลี	80.00	100.00	90.00
สันสกฤต	94.18	97.14	95.66
ตัวการ์นต์	100.00	88.25	94.13
บาลีและสันสกฤต	88.89	100.00	94.44
บาลีและตัวการ์นต์	100.00	100.00	100.00
สันสกฤตและตัวการ์นต์	100.00	100.00	100.00
บาลี สันสกฤต และตัวการ์นต์	100.00	100.00	100.00
ค่าเฉลี่ย F-measure ของหัวเรื่อง	94.72	97.91	

จากตารางที่ 4.3 ได้แสดงค่า F-measure ซึ่งแบ่งออกเป็น 3 ส่วน ดังนี้

ส่วนที่ 1 ร้อยละของค่าเฉลี่ย F-measure ของชุดเอกสารในแต่ละหัวเรื่อง คือ ร้อยละของค่าเฉลี่ย F-measure จากชุดเอกสารในแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ในหัวเรื่อง Data mining ได้แก่ ประเภทตัวการ์ตูน ประเภทบาติและตัวการ์ตูน ประเภทสันสกฤตและตัวการ์ตูน และประเภทบาติ สันสกฤต และตัวการ์ตูน และในหัวเรื่อง iOS operating system ได้แก่ ประเภทบาติ ประเภทบาติและสันสกฤต ประเภทบาติและตัวการ์ตูน ประเภทสันสกฤตและตัวการ์ตูน และประเภทบาติ สันสกฤต และตัวการ์ตูน ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุดในหัวเรื่อง Data mining ได้แก่ ประเภทบาติ ที่มีค่าเฉลี่ยเพียงร้อยละ 80.00 และในหัวเรื่อง iOS operating system ได้แก่ ประเภทตัวการ์ตูน ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 88.25

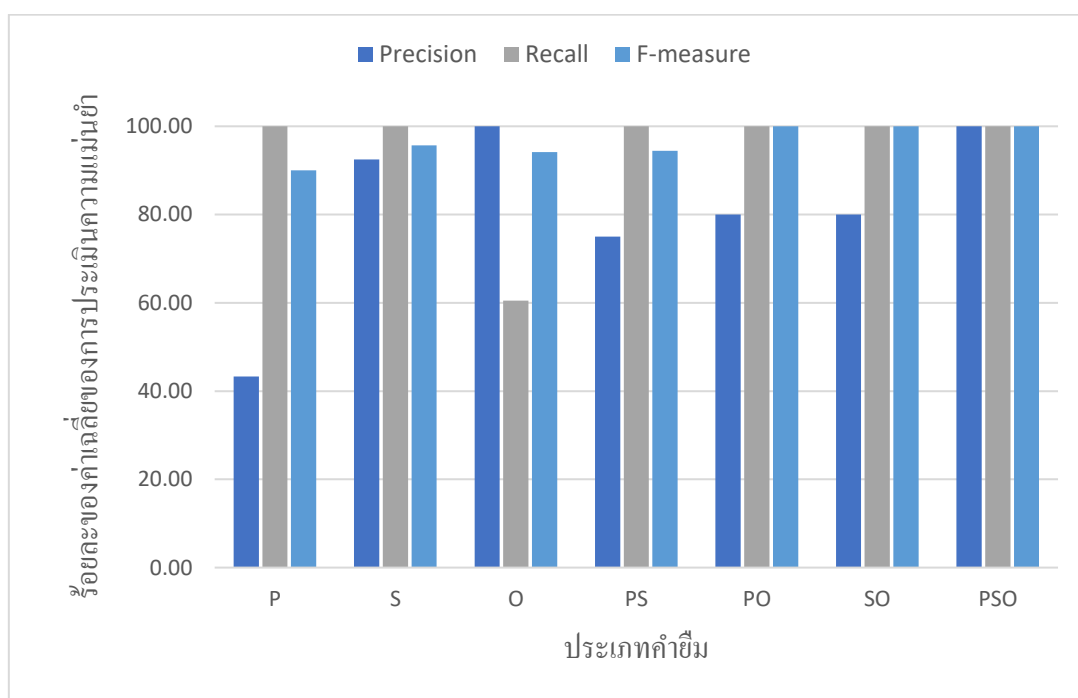
ส่วนที่ 2 ร้อยละของค่าเฉลี่ย F-measure ของคำยืม คือ ค่าเฉลี่ย F-measure จากหัวเรื่องของเอกสารแต่ละหัวเรื่องของการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลอง ซึ่งแสดงผลเป็นร้อยละ โดยการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 100.00 ได้แก่ ประเภทบาติและตัวการ์ตูน ประเภทสันสกฤตและตัวการ์ตูน และประเภทบาติ สันสกฤต และตัวการ์ตูน ส่วนการจำแนกประเภทคำยืมที่มีค่าเฉลี่ยน้อยที่สุด ได้แก่ ประเภทบาติ ซึ่งมีค่าเฉลี่ยเพียงร้อยละ 90.00

ส่วนที่ 3 ค่าเฉลี่ย F-measure ของหัวเรื่อง คือ ค่าเฉลี่ย F-measure จากการจำแนกประเภทคำยืมแต่ละประเภทของแบบจำลองในแต่ละหัวเรื่องของเอกสาร โดยหัวเรื่องที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 97.91 คือ iOS operating system

ตารางที่ 4.4 ค่าเฉลี่ยความถูกต้องของการสกัดคำยืม

หัวเรื่อง	ร้อยละของค่าเฉลี่ยความถูกต้อง
Data mining	98.52
iOS operating system	99.10
ค่าเฉลี่ยความถูกต้องของหัวเรื่องทั้งหมด	98.81

จากตารางที่ 4.4 แสดงค่าเฉลี่ยของค่าความถูกต้องเป็นร้อยละ โดยแบ่งตามหัวเรื่องของเอกสาร ซึ่งค่าเฉลี่ยความถูกต้องนี้คำนวณจากค่าความถูกต้องในการสกัดคำยืมของเอกสารในแต่ละหัวเรื่อง หัวเรื่องที่มีค่าเฉลี่ยความถูกต้องที่สูงที่สุด คือ หัวเรื่อง iOS operating system ซึ่งมีค่าเฉลี่ยถึงร้อยละ 99.10 ส่วนค่าเฉลี่ยความถูกต้องของหัวเรื่องทั้งหมดเท่ากับร้อยละ 98.81



รูปที่ 4.1 การสรุปผลของการประเมินความแม่นยำในการสกัดคำยืมแต่ละประเภท

จากรูปที่ 4.1 คือ การสรุปผลความแม่นยำในการสกัดคำยืมแต่ละประเภทในภาษาไทยด้วยแบบจำลองที่งานวิจัยนี้ได้พัฒนาขึ้น (สามารถดูความหมายของอักษรย่อในตารางที่ 3.4) โดยสังเกตได้จากค่า F-measure เป็นสิ่งสำคัญ ซึ่งการสกัดคำยืมในภาษาไทยที่มีค่า F-measure ที่สูงที่สุดถึงร้อยละ 100.00 ได้แก่ คำยืมจากภาษาบาลีที่มีตัวการ์นต์ ภาษาสันสกฤตที่มีตัวการ์นต์ และทั้งภาษาบาลีและสันสกฤตที่มีตัวการ์นต์ ซึ่งเฉลี่ยจากทั้งสองหัวเรื่อง และในทางกลับกัน การสกัดคำยืมในภาษาไทยที่มีค่า F-measure ที่น้อยที่สุดเพียงร้อยละ 90.00 ได้แก่ คำยืมจากภาษาบาลี ซึ่งเฉลี่ยจากทั้งสองหัวเรื่อง

จากผลการทดลองทั้งหมดของการสกัดคำยืมสำหรับการประเมินความยากง่ายของข้อความในเอกสารภาษาไทย สามารถสรุปประเด็นสำคัญได้ 3 ประเด็น ดังนี้

ประเด็นที่ 1 ผลของค่า Precision สำหรับการจำแนกประเภทคำบาลี (P) น้อยกว่าร้อยละ 50.00 เนื่องจากเกิดการจำแนกผิดประเภทของแบบจำลอง โดยที่รูปแบบสำหรับตรวจจับคำบาลีได้ตรวจจับคำในประเภทอื่น ๆ เช่น **ต้องการ** **ดึงความรู้** **มุมมอง** **ขั้นตอน** ซึ่งคำในข้างต้นถูกตรวจจับด้วยรูปแบบตัวสะกดตัวตามของบาลี ซึ่งสังเกตได้จากพยัญชนะที่ขีดเส้นใต้ ส่งผลให้ค่า Precision ที่ได้รับของการจำแนกคำบาลีน้อยลง

ประเด็นที่ 2 ผลค่า Recall สำหรับการจำแนกประเภทคำที่มีตัวการันต์น้อยกว่าค่า Precision โดยปกติแล้วการตรวจจับเครื่องหมายทัศนธาตุในข้อความจะต้องมีความถูกต้องร้อยละ 100.00 เนื่องจากในงานวิจัยนี้มีการแบ่งประเภทคำที่มีตัวการันต์ออกเป็น 4 ประเภท คือ ตัวการันต์ (O) บาลีและการันต์ (PS) สันสกฤตและการันต์ (SO) และบาลี สันสกฤต และการันต์ (PSO) ส่งผลให้แบบจำลองเกิดการจำแนกคำที่มีตัวการันต์ผิดประเภท ตัวอย่างคำที่ตรวจจับผิดพลาด เช่น **คอมพิวเตอร์** แบบจำลองได้จำแนกประเภทเป็นคำยืมจากภาษาบาลีที่มีตัวการันต์ ซึ่งความเป็นจริงคำนี้จะต้องจำแนกเป็นประเภทตัวการันต์ โดยบริเวณพยัญชนะที่ขีดเส้นใต้จะเป็นส่วนที่แบบจำลองตรวจจับได้ ซึ่งบริเวณ มพ คือ รูปแบบตัวสะกดตัวตามของบาลี ส่งผลให้ค่า Recall ที่ได้รับของการจำแนกประเภทตัวการันต์น้อยกว่าค่า Precision

ประเด็นที่ 3 ประเด็นผลค่า Precision Recall และ F-measure สำหรับการจำแนกประเภทคำยืมจากภาษาบาลี สันสกฤต และการันต์ (PSO) สูงเท่ากันถึงร้อยละ 100.00 เนื่องจากการสกัดคำยืมประเภทนี้เกิดขึ้นจากการตรวจจับตำแหน่งคำยืมด้วยชุดรูปแบบเงื่อนไขของการสกัดตำแหน่งคำบาลี ชุดรูปแบบเงื่อนไขของการสกัดตำแหน่งคำสันสกฤต และรูปแบบเงื่อนไขของการสกัดตำแหน่งตัวการันต์ ส่งผลให้การจำแนกคำประเภทนี้มีประสิทธิภาพสูงสุดจากการจำแนกประเภทคำอื่น ๆ สังเกตได้จากค่า Precision ของการจำแนกคำบาลีก่อนข้างน้อย แต่เมื่อนำรูปแบบเงื่อนไขในการสกัดตำแหน่งคำบาลีไปใช้ร่วมกับรูปแบบเงื่อนไขในการสกัดตำแหน่งคำสันสกฤต และรูปแบบเงื่อนไขในการสกัดตำแหน่งตัวการันต์ สำหรับตรวจจับคำประเภทบาลีและสันสกฤตประเภทบาลีและตัวการันต์ และประเภทบาลี สันสกฤต และตัวการันต์ จะมีค่า Precision เพิ่มขึ้น ซึ่งเกินร้อยละ 50.00

4.2 ผลการทดลองการจัดกลุ่มเอกสารและการอภิปรายผล

ในหัวข้อนี้อธิบายเกี่ยวกับผลการทดลองของการจัดกลุ่มเอกสาร ซึ่งเป็นการทดลองเพื่อหาความใกล้เคียงในการจัดกลุ่มเอกสารระหว่างการจัดกลุ่มเอกสารด้วยวิธีที่งานวิจัยนี้พัฒนาขึ้น (DoC-A DoC-B และ DoC-C) และการจัดกลุ่มด้วยกลุ่มตัวอย่าง (กลุ่มผู้อ่าน) โดยผลการทดลองของการประเมินความใกล้เคียงในการจัดกลุ่ม แสดงดังตารางที่ 4.5

ตารางที่ 4.5 ผลการประเมินความใกล้เคียงในการจัดกลุ่มของกลุ่มตัวอย่าง

กลุ่ม ตัวอย่าง	Data mining			iOS operating system		
	DoC-A	DoC-B	DoC-C	DoC-A	DoC-B	DoC-C
1	2	1	3	2	1	3
2	3	2	1	3	1	2
3	3	1	2	2	3	1
4	2	1	3	2	1	3
5	2	1	3	2	1	3
6	2	1	3	2	1	3
7	2	3	1	1	2	3
8	3	1	2	2	1	3
9	2	1	3	2	1	3
10	3	1	2	2	1	3
11	3	1	2	2	1	3
12	2	3	1	3	1	2
13	2	3	1	3	1	2
14	1	3	2	2	3	1

ตารางที่ 4.5 ผลการประเมินความใกล้เคียงในการจัดกลุ่มของกลุ่มตัวอย่าง (ต่อ)

กลุ่ม ตัวอย่าง	Data mining			iOS operating system		
	DoC-A	DoC-B	DoC-C	DoC-A	DoC-B	DoC-C
15	1	3	2	2	3	1
16	1	2	2	2	3	1
17	3	1	2	3	1	2
18	2	3	1	3	2	1
19	2	1	3	2	1	3
20	1	3	2	3	2	1

จากตารางที่ 4.5 แสดงผลการประเมินความใกล้เคียงในการจัดกลุ่มของกลุ่มตัวอย่าง โดยกลุ่มตัวอย่างจำนวน 20 คน ประเมินระดับความใกล้เคียงต่อผลการจัดกลุ่มเอกสารของเทคนิคที่พัฒนาขึ้น (DoC-A Doc-B และ Doc-C) บนเนื้อหาเอกสารที่แบ่งออกเป็นสองหัวข้อเรื่อง คือ Data mining และ iOS operating system

ตารางที่ 4.6 สรุปการประเมินผลและการคำนวณผลความใกล้เคียงในการจัดกลุ่มของหัวข้อเรื่อง Data mining

เทคนิคการจัดกลุ่ม	จำนวนกลุ่มตัวอย่างของระดับความ ใกล้เคียงของการจัดกลุ่ม			การคำนวณผล	
	3	2	1	ผลรวม	ฐานนิยม
DoC-A	6	10	4	42	2
DoC-B	7	2	11	36	1
DoC-C	6	9	5	41	2

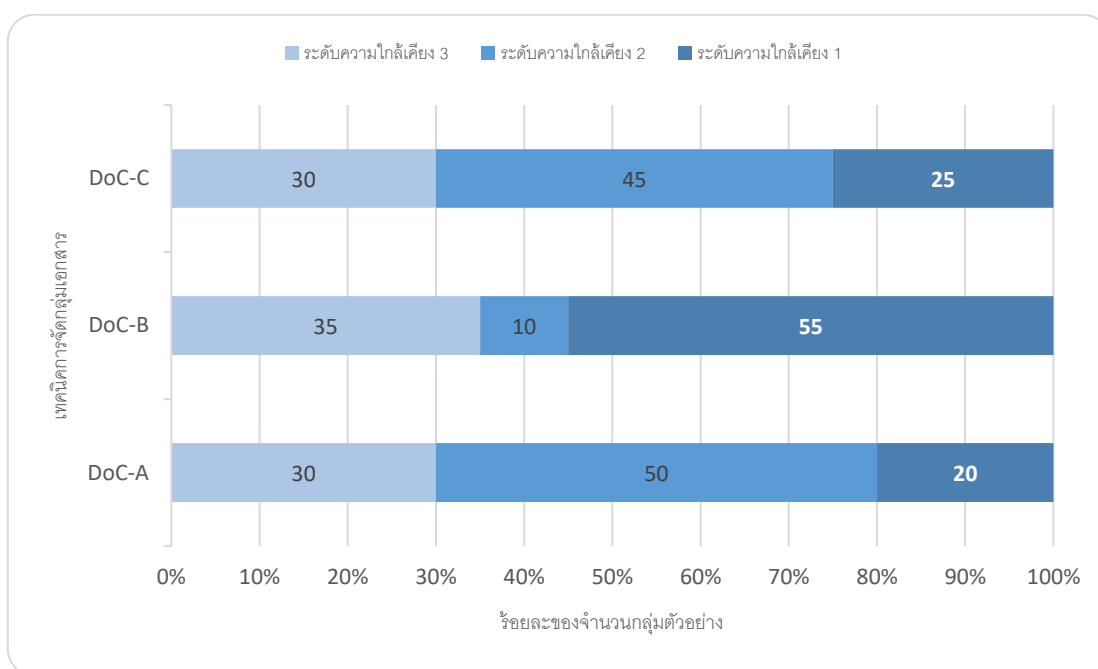
ตารางที่ 4.7 สรุปการประเมินผลและการคำนวณผลความใกล้เคียงในการจัดกลุ่มของหัวข้อเรื่อง iOS operating system

เทคนิคการจัดกลุ่ม	จำนวนกลุ่มตัวอย่างของระดับความใกล้เคียงของการจัดกลุ่ม			การคำนวณผล	
	3	2	1	ผลรวม	ฐานนิยม
DoC-A	6	13	1	45	2
DoC-B	4	3	13	31	1
DoC-C	10	4	6	44	3

จากตารางที่ 4.6 และตารางที่ 4.7 แสดงการสรุปผลการประเมินระดับความใกล้เคียงในการจัดกลุ่มเอกสารและผลการคำนวณความใกล้เคียงการจัดกลุ่มเอกสาร โดยการสรุปผลการประเมินระดับใกล้เคียงในการจัดกลุ่มนั้นได้แสดงในรูปแบบจำนวนของกลุ่มตัวอย่างที่ดำเนินการประเมินโดยให้ระดับความใกล้เคียง 1 ถึง 3 ในแต่ละเทคนิค ซึ่งแสดงให้เห็นถึงการประเมินของจำนวนกลุ่มตัวอย่างในการให้ระดับความใกล้เคียงแก่เทคนิคการจัดกลุ่มเอกสาร และผลการคำนวณความใกล้เคียงการจัดกลุ่มเอกสารแสดงให้เห็นถึงผลรวมของระดับความใกล้เคียงในแต่ละเทคนิคเพื่อต้องการแสดงความมากน้อยของผลรวมในแต่ละเทคนิค และฐานนิยมของระดับความใกล้เคียงในแต่ละเทคนิคเพื่อต้องการแสดงการเลือกระดับความใกล้เคียงของกลุ่มตัวอย่างส่วนใหญ่ในแต่ละเทคนิค จากตารางที่ 4.6 เป็นผลประเมินของเอกสารหัวข้อเรื่อง Data mining กลุ่มตัวอย่างส่วนใหญ่ให้ระดับความใกล้เคียง 1 สำหรับเทคนิค DoC-B จำนวน 11 คน และกลุ่มตัวอย่างส่วนน้อยประเมินระดับความใกล้เคียง 2 สำหรับเทคนิค DoC-B เช่นกัน จำนวน 2 คน โดยผลสรุปการคำนวณหาผลรวมของแต่ละเทคนิค เทคนิคที่ได้รับผลรวมมากที่สุดถึง 42 คือ DoC-A และเทคนิคที่ได้รับผลรวมน้อยที่สุดเพียง 36 คือ DoC-B ในส่วนการสรุปผลของค่าฐานนิยมในเทคนิค DoC-A เท่ากับ 2 DoC-B เท่ากับ 1 และ DoC-C เท่ากับ 2

จากตารางที่ 4.7 เป็นผลประเมินของเอกสารหัวข้อเรื่อง iOS operating system กลุ่มตัวอย่างส่วนใหญ่ให้ระดับความใกล้เคียง 2 และ 1 สำหรับเทคนิค DoC-A และ DoC-B ตามลำดับ

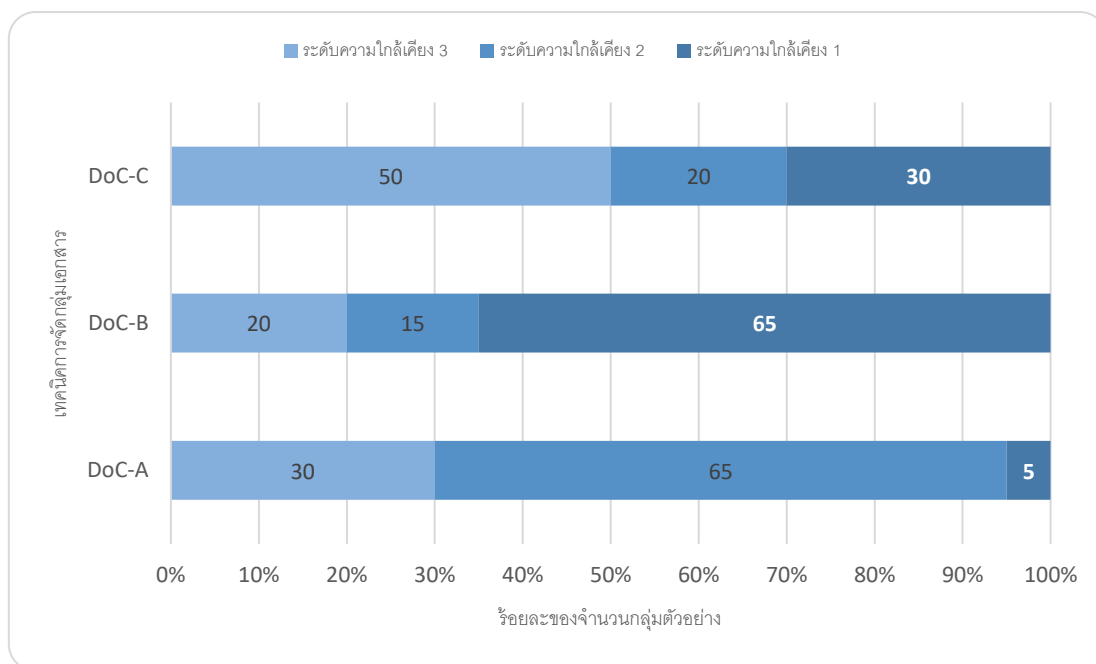
จำนวน 13 คน และกลุ่มตัวอย่างส่วนน้อยประเมินระดับความใกล้เคียง 1 สำหรับเทคนิค DoC-A จำนวน 1 คน โดยผลสรุปการคำนวณหาผลรวมของแต่ละเทคนิค เทคนิคที่ได้รับผลรวมมากที่สุดถึง 45 คือ DoC-A และเทคนิคที่ได้รับผลรวมน้อยที่สุดเพียง 31 คือ DoC-B ในส่วนการสรุปผลของค่าฐานนิยมในเทคนิค DoC-A เท่ากับ 2 DoC-B เท่ากับ 1 และ DoC-C เท่ากับ 3



รูปที่ 4.2 ร้อยละจำนวนกลุ่มตัวอย่างในการประเมินความใกล้เคียงสำหรับการจัดกลุ่มเอกสารของเทคนิคในหัวเรื่อง Data mining

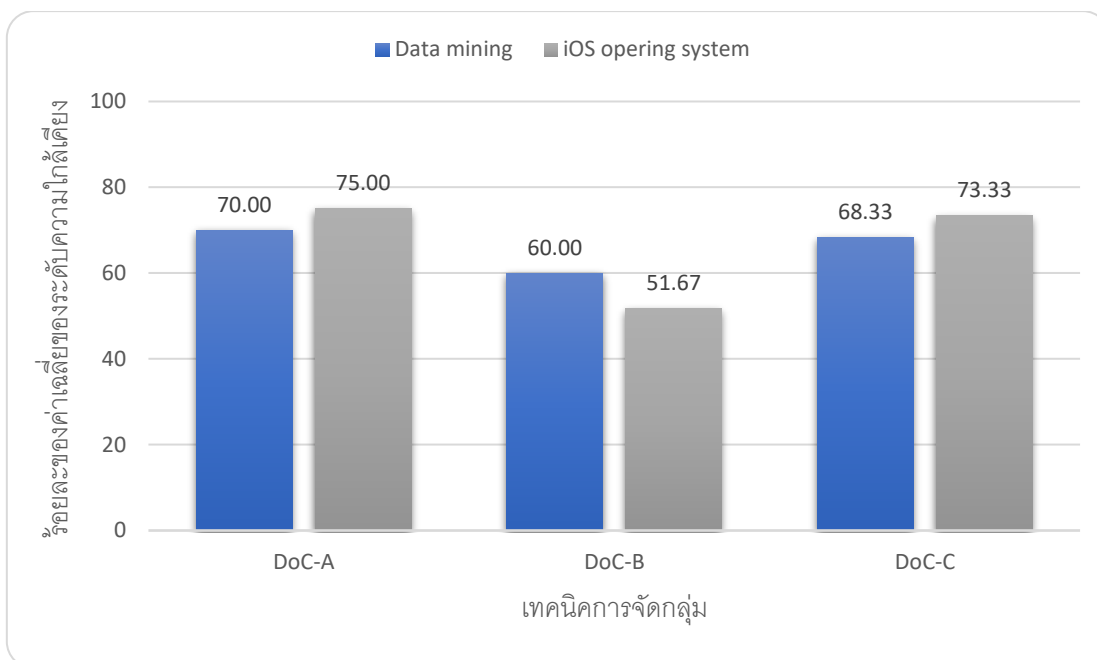
จากรูปที่ 4.2 แสดงถึงสัดส่วนของจำนวนกลุ่มตัวอย่างในการประเมินระดับความใกล้เคียงในการจัดกลุ่มของเทคนิคการจัดกลุ่มเอกสารหัวเรื่อง Data mining โดยสัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-A ที่สูงที่สุดถึงร้อยละ 50.00 คือ ระดับ 2 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-A ที่น้อยที่สุดเพียงร้อยละ 20.00 คือ ระดับ 1 สัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-B ที่สูงที่สุดถึงร้อยละ 55.00 คือ ระดับ 1 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-B ที่น้อยที่สุดเพียงร้อยละ 10.00 คือ ระดับ 2 และสัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-C ที่สูงที่สุด

ถึงร้อยละ 45.00 คือ ระดับ 2 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-C ที่น้อยที่สุดเพียงร้อยละ 25.00 คือ ระดับ 1



รูปที่ 4.3 ร้อยละจำนวนกลุ่มตัวอย่างในการประเมินความใกล้เคียงสำหรับการจัดกลุ่มเอกสารของเทคนิคในหัวเรื่อง iOS operating system

จากรูปที่ 4.3 แสดงถึงสัดส่วนของจำนวนกลุ่มตัวอย่างในการประเมินระดับความใกล้เคียงในการจัดกลุ่มของเทคนิคการจัดกลุ่มเอกสารหัวเรื่อง iOS operating system โดยสัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-A ที่สูงที่สุดถึงร้อยละ 65.00 คือ ระดับ 2 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-A ที่น้อยที่สุดเพียงร้อยละ 5.00 คือ ระดับ 1 สัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-B ที่สูงที่สุดถึงร้อยละ 65.00 คือ ระดับ 1 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-B ที่น้อยที่สุดเพียงร้อยละ 15.00 คือ ระดับ 2 และสัดส่วนของระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-C ที่สูงที่สุดถึงร้อยละ 50.00 คือ ระดับ 3 และส่วนระดับความใกล้เคียงในการจัดกลุ่มของเทคนิค DoC-C ที่น้อยที่สุดเพียงร้อยละ 20.00 คือ ระดับ 2



รูปที่ 4.4 ค่าเฉลี่ยระดับความใกล้เคียงในการจัดกลุ่มเอกสารของเทคนิคในแต่ละหัวเรื่อง

จากรูปที่ 4.4 แสดงถึงค่าเฉลี่ยระดับความใกล้เคียงในการจัดกลุ่มเอกสารของเทคนิคในแต่ละหัวเรื่อง ซึ่งนำผลรวมของแต่ละเทคนิคในแต่ละหัวเรื่องมาแสดงผลเป็นร้อยละ โดยในส่วนของหัวเรื่อง Data mining เทคนิคที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 70.00 คือ DoC-A และเทคนิคที่มีค่าเฉลี่ยน้อยที่สุดเพียงร้อยละ 60.00 คือ DoC-B ส่วนหัวเรื่อง iOS operating system เทคนิคที่มีค่าเฉลี่ยสูงที่สุดถึงร้อยละ 75.00 คือ DoC-A และเทคนิคที่มีค่าเฉลี่ยน้อยที่สุดเพียงร้อยละ 51.67 คือ DoC-B

จากผลการทดลองการประเมินระดับความยากง่ายด้วยเทคนิค DoC-A มีความใกล้เคียงกับผลการประเมินระดับความยากง่ายของกลุ่มตัวอย่างมากที่สุด เนื่องจากการพิจารณาความยากง่ายของมนุษย์จะประกอบด้วยหลายปัจจัย เช่น ความซับซ้อนของโครงสร้างประโยค คำศัพท์เทคนิค คำที่ไม่คุ้นเคย การจัดรูปแบบข้อความ ซึ่งแตกต่างกับ DoC-B และ DoC-C ที่พิจารณาเฉพาะคำยืม ส่งผลให้ผลการทดลองได้น้อยกว่า DoC-A อย่างไรก็ตามแม้ว่าคำยืม (บาลี สันสกฤต ตัวการ์ตูน) จะเป็นคุณสมบัติสำคัญที่ช่วยในการประเมินความยากง่ายของข้อความ ผู้วิจัยเห็นว่าจำเป็นต้องพิจารณาปัจจัยอื่น ๆ เพิ่มเติม เช่น คำยืมจากภาษาจีน คำไทยแท้ คำที่ไม่คุ้นเคย ความ

ซับซ้อนของโครงสร้างประโยค เพื่อให้ได้รับผลการประเมินระดับความยากง่ายที่ใกล้เคียงกับกลุ่มตัวอย่างเพิ่มมากขึ้น

ตารางที่ 4.8 การเปรียบเทียบงานวิจัยด้านการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย

ผู้วิจัย	วัตถุประสงค์	คุณสมบัติการประเมิน ความยากง่าย	เทคนิค	จำนวนระดับความยากง่าย	ร้อยละของความแม่นยำสูงสุด
Daowadung and Chen (2011)	จำแนกเอกสารตามระดับชั้นประถมศึกษาในประเทศไทย	คำศัพท์ในแต่ละระดับชั้นการศึกษา	การจำแนกประเภท	2	83.00
Tongtep, et al. (2014)	จำแนกเอกสารตามระดับชั้นมัธยมศึกษาตอนต้น มัธยมศึกษาตอนปลาย และปริญญาตรี	ความซับซ้อนของข้อความ	การจำแนกประเภท	3	46.53
Daowadung and Chen (2015)	จำแนกเอกสารตามระดับชั้นประถมศึกษาในประเทศไทย	คำศัพท์ในแต่ละระดับชั้นการศึกษาและความซับซ้อนของข้อความ	การจำแนกประเภท	6	50.97

ตารางที่ 4.8 การเปรียบเทียบงานวิจัยด้านการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย (ต่อ)

ผู้วิจัย	วัตถุประสงค์	คุณสมบัติการประเมินความยากง่าย	เทคนิค	จำนวนระดับความยากง่าย	ร้อยละของความแม่นยำสูงสุด
งานวิจัยนี้	จัดกลุ่มเอกสารตามระดับความยากง่ายของข้อความสำหรับบุคคลทั่วไป	คำยืม (บาลี สันสกฤต และตัวการ์นต์)	การจัดกลุ่ม	3	75.00

จากตารางที่ 4.8 สรุปข้อมูลงานวิจัยที่เกี่ยวกับการประเมินระดับความยากง่ายของข้อความสำหรับเอกสารภาษาไทยตั้งแต่อดีตถึงปัจจุบัน ซึ่งพบว่างานวิจัยที่มีการกำหนดจำนวนประเภทสำหรับจำแนกจำนวน 2 ประเภท ส่งผลให้ได้รับค่าความแม่นยำสูงถึงร้อยละ 83.00 แต่ตรงกันข้ามหากกำหนดจำนวนประเภทสำหรับจำแนกมากกว่า 2 ประเภทขึ้นไป ส่งผลให้ค่าความแม่นยำในการจำแนกลดน้อยลง จากข้อเท็จจริงข้างต้น งานวิจัยนี้สามารถกำจัดข้อเท็จจริงนี้ได้ด้วยการใช้เทคนิคในการจัดกลุ่ม ซึ่งจะเห็นได้ว่างานวิจัยนี้มีการกำหนดจำนวนกลุ่มที่มากกว่า 2 กลุ่ม และค่าความแม่นยำสูงที่สุดเมื่อเปรียบเทียบกับงานวิจัยที่มีการกำหนดจำนวนประเภทที่มากกว่า 2 ประเภท นอกจากนี้ระดับความยากง่าย (กลุ่ม) ไม่เจาะจงเฉพาะกลุ่มบุคคล ซึ่งสามารถปรับใช้กับบุคคลทั่วไปได้ อย่างไรก็ตามเทคนิคในการจัดกลุ่มที่ได้นำเสนอต้องอาศัยจำนวนเอกสารที่มากกว่า 1 ฉบับขึ้นไปในการจัดกลุ่มเอกสาร และการจัดกลุ่มเอกสารจะมีประสิทธิภาพมากขึ้นหากมีการใช้คำยืมในชุดเอกสารนั้น ๆ

บทที่ 5

สรุปและข้อเสนอแนะ

ในบทนี้กล่าวถึงการสรุปเกี่ยวกับงานวิจัยนี้ทั้งหมด ปัญหาและอุปสรรคที่เกิดขึ้นระหว่างการดำเนินการวิจัย ข้อเสนอแนะต่าง ๆ ที่เกี่ยวกับงานวิจัยนี้ และลักษณะการนำงานวิจัยนี้ไปใช้ประโยชน์

5.1 สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอวิธีการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยโดยใช้คำยืมในภาษาไทยเป็นคุณสมบัติในการประเมินระดับความยากง่ายของข้อความ และใช้การจัดกลุ่มเอกสารเป็นเทคนิคในการจัดระดับความยากง่ายของเอกสาร ในส่วนของคุณสมบัติการประเมินระดับความยากง่ายของข้อความได้นำเสนอวิธีการสกัดคำยืมในภาษาไทย และผลลัพธ์ความถูกต้องในการทดลองการจำแนกประเภทของคำยืมทั้ง 7 ประเภท โดยดำเนินการทดลองโดยใช้เอกสารทั้งหมด 2 หัวเรื่อง จำนวน 30 ฉบับ ประกอบด้วยหัวเรื่อง Data mining จำนวน 11 ฉบับ และหัวเรื่อง iOS operating system จำนวน 19 ฉบับ ซึ่งปรากฏว่าการจำแนกประเภทคำยืมที่มาจากภาษาบาลี สันสกฤต และตัวการ์นต์ได้รับค่า F-measure สูงที่สุดถึงร้อยละ 100.00 และจากเอกสารทั้งหมดจำนวน 2 หัวเรื่อง ปรากฏว่าหัวเรื่อง iOS operating system ได้รับค่า Accuracy สูงที่สุดถึงร้อยละ 99.10 และส่วนของเทคนิคในการจัดระดับความยากง่ายของเอกสารได้นำเสนอเทคนิคการจัดกลุ่มเอกสาร ประกอบด้วย 3 เทคนิค คือ 1) การจัดกลุ่มเอกสารที่คำนวณจาก

สัดส่วนของจำนวนเอกสารทั้งหมด (DoC-A) 2) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุด (DoC-B) 3) การจัดกลุ่มเอกสารด้วยอันตรภาคชั้นที่คำนวณจากค่าที่มากที่สุดและน้อยที่สุดที่เกิดขึ้นจริง (DoC-C) โดยทั้งหมดมีพื้นฐานการจัดกลุ่มออกเป็น 3 กลุ่ม คือ กง่าย ปานกลาง และยาก และได้นำเสนอผลลัพธ์การทดลองโดยสร้างระบบการประเมินความใกล้เคียงของผลการจัดกลุ่มเอกสารระหว่างเทคนิคการจัดกลุ่มเอกสารที่ได้นำเสนอและกลุ่มตัวอย่าง (ผู้ประเมิน) จำนวน 20 คน โดยดำเนินการทดลองโดยการใช้เอกสารทั้งหมด 2 หัวเรื่อง หัวเรื่องละ 10 ฉบับ ประกอบด้วยหัวเรื่อง Data mining และหัวเรื่อง iOS operating system ซึ่งปรากฏว่าเทคนิคการจัดกลุ่มเอกสาร DoC-A มีผลลัพธ์ความใกล้เคียงในการจัดกลุ่มเอกสารกับผู้ประเมินมากที่สุดโดยเฉลี่ยถึงร้อยละ 70.00 ของหัวเรื่อง Data mining และร้อยละ 75.00 ของหัวเรื่อง iOS operating system

5.2 ปัญหาและอุปสรรค

- 1) เนื่องจากหลักสังเกตคำบาลีและสันสกฤตที่ศึกษา เมื่อนำไปสร้างกฎเพื่อสกัดคำ ยืมพบว่ายังมีข้อบกพร่อง ส่งผลให้การตรวจจับเกิดความผิดพลาด
- 2) เนื่องจากทรัพยากรพจนานุกรมคำศัพท์บาลีและสันสกฤตมีจำนวนน้อยส่งผลให้จำนวนพจนานุกรมที่พร้อมใช้งานสำหรับงานวิจัยนี้น้อยลงเช่นกัน

5.3 ข้อเสนอแนะ

- 1) ควรมีความหลากหลายของหัวเรื่องเอกสารสำหรับการทดลอง
- 2) ควรหาคุณสมบัติอื่น ๆ นอกจากคำยืม ในการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทย เช่น คำยืมจากภาษาจีน คำไทยแท้ซึ่งสะกดตรงตามมาตรา

3) ควรนำวิธีการประเมินระดับความยากง่ายของข้อความในเอกสารภาษาไทยที่ได้ นำเสนอทดลองกับภาษาอื่น ๆ ในกลุ่มระบบการเขียนแบบ Syllabic alphabets ซึ่งภาษาไทยเป็นหนึ่งในระบบการเขียนนี้ เช่น ลาว เขมร พม่า

4) ควรทดลองการประเมินระดับความยากง่ายโดยใช้เทคนิคการจัดกลุ่มเอกสารของงานวิจัยก่อนหน้า เพื่อเปรียบเทียบผลการจัดกลุ่มเอกสารระหว่างเทคนิคการจัดกลุ่มเอกสารที่ได้นำเสนอ

5.4 ลักษณะการนำไปใช้ประโยชน์

สำหรับตัวอย่างการนำไปใช้ประโยชน์ในเชิงวิชาการ เช่น การนำไปประยุกต์ใช้ สำหรับการประเมินระดับความยากง่ายของเนื้อหาเอกสารดิจิทัลภาษาไทยที่เกี่ยวข้องกับการเรียนของนักเรียนและนักศึกษาทุกระดับชั้น เพื่อช่วยให้ผู้สอนจัดเอกสารการเรียนรู้ให้เหมาะสมกับระดับชั้นการศึกษาแต่ละระดับชั้นตามที่ต้องการได้อย่างรวดเร็วขึ้น ซึ่งอาจจะต้องประยุกต์โดยปรับจำนวนกลุ่มให้เหมาะสมกับจำนวนระดับชั้นการศึกษา และสามารถนำวิธีการดำเนินงานในงานวิจัยนี้สร้างเป็นฟังก์ชันเสริมให้กับระบบค้นหาเอกสารในห้องสมุดหรือร้านหนังสือ โดยฟังก์ชันดังกล่าวจะช่วยจัดเรียงผลการค้นหาเอกสารตามความยากง่ายของเนื้อหาข้อความ ทำให้ผู้อ่านเลือกอ่านเอกสารที่เหมาะสมกับตนเองได้รวดเร็วขึ้น สำหรับการประยุกต์ใช้กับเอกสารดิจิทัลภาษาอื่น ๆ ที่อยู่ในกลุ่ม Syllabic Alphabets จำเป็นต้องพิจารณาเกณฑ์การสกัดคำยืมและคุณสมบัติการประเมินระดับความยากง่ายเพื่อให้เหมาะสมกับโครงสร้างของภาษานั้น ๆ

เอกสารอ้างอิง

- กำชัย ทองหล่อ. (2550). *หลักภาษาไทย*, รวมสาส์น, กรุงเทพฯ.
- รำเนา พลอยขาว. (2549). “ภาษาไทย.” (ออนไลน์) เข้าถึงได้ที่ <http://www.prakan.ac.th/Link-Data/e-learning-49/e-learning/index.html> (วันที่ 4 พฤษภาคม 2559).
- ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (2559). “โปรแกรมตัดคำภาษาไทย LexToPlus.” (ออนไลน์) เข้าถึงได้ที่ <https://www.nectec.or.th/innovation/innovation-software/lextoplus.html> (วันที่ 3 ธันวาคม 2559).
- สำนักงานสถิติแห่งชาติ. (2555). “สำรวจการอ่านหนังสือของประชากร ปี 2558.” (ออนไลน์) เข้าถึงได้ที่ <http://service.nso.go.th/nso/web/survey/surpop2-3-1.html> (วันที่ 20 สิงหาคม 2559).
- สุธีวงศ์ พงศ์ไพบูลย์. (2532). *หลักภาษาไทย*, ไทยวัฒนาพานิช, กรุงเทพฯ.
- สุพัฒนา ปัญญา และ ทศนัย ชุ่มวัฒนะ. (2559). “เทคนิคการเทียบโอนผลการเรียนจากคำอธิบายรายวิชาภาษาไทยโดยใช้เทคนิคการจัดกลุ่ม ข้อมูล กรณีศึกษา มหาวิทยาลัยรังสิต.”, *เอกสารการประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 12 ประจำปี 2559*, โรงแรมเซ็นทาราแอนด์คอนเวนชันเซ็นเตอร์ ขอนแก่น: 7-8 กรกฎาคม 2559.
- สุภาพร มากแจ้ง. (2535). *ภาษาบาลี-สันสกฤตในภาษาไทย*, โอเดียนสโตร์, กรุงเทพฯ.
- Abedi, J., Leon, S., Kao, J., Bayley, R., Ewers, N., Herman, J., and Mundhenk, K. (2011). “Accessible Reading Assessments for Students with Disabilities: The Role of Cognitive, Grammatical, Lexical, and Textual/ Visual Features.” (Online) Available on <http://files.eric.ed.gov/fulltext/ED520492.pdf> (5 December 2016).
- Ager, S. (2017). “Types of writing system.” (Online) Available on <http://omniglot.com/writing/types.htm> (15 May 2016).
- Al-Khalifa, H. S., and Al-Ajlan, A. A. (2010). “Automatic readability measurements of the Arabic text: An exploratory study.” *Arabian Journal for Science and Engineering*, 35(2 C), 103-124.

เอกสารอ้างอิง (ต่อ)

- Beinborn, L., Zesch, T., and Gurevych, I. (2012). “Towards fine-grained readability measures for self-directed language learning.”, *Proceedings of the SLTC 2012 workshop on NLP for CALL*, Lund, Sweden: 25 October; 2012.
- Calvillo, E. A., Padilla, A., Muñoz, J., Ponce, J., and Fernandez, J. T. (2013). “Searching research papers using clustering and text mining.”, *Proceedings of 2013 International Conference on Electronics, Communications and Computing (CONIELECOMP)*, Universidad de las Americas Puebla, Maxico: 11-13 March, 2013.
- Chen, Y.-H., and Daowadung, P. (2015). “Assessing readability of Thai text using support vector machines.” *Maejo International Journal of Science and Technology*, 9(3), 355.
- Chen, Y.-T., Chen, Y.-H., and Cheng, Y.-C. (2013). “Assessing chinese readability using term frequency and lexical chain.” *International Journal of the Association for Computational Linguistics and Chinese Language Processing*, 18(2), 1-18.
- Coleman, M., and Liau, T. L. (1975). “A computer readability formula designed for machine scoring.” *Journal of Applied Psychology*, 60(2), 283.
- Collins-Thompson, K. (2014). “Computational assessment of text readability: A survey of current and future research.” *ITL-International Journal of Applied Linguistics*, 165(2), 97-135.
- Collins-Thompson, K., Bennett, P. N., White, R. W., De La Chica, S., and Sontag, D. (2011). “Personalizing web search results by reading level.”, *Proceedings of The 20th ACM international conference on Information and knowledge management*, Glasgow, Scotland: 24-28 October, 2011.
- Collins-Thompson, K., and Callan, J. (2004). “A Language Modeling Approach to Predicting Reading Difficulty.”, *Proceedings of Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, Boston, Massachusetts, United States of America: 2-7 May, 2004.

เอกสารอ้างอิง (ต่อ)

- Collins-Thompson, K., and Callan, J. (2005). "Predicting reading difficulty with statistical language models." *Journal of the Association for Information Science and Technology*, 56(13), 1448-1462.
- Daowadung, P., and Chen, Y.-H. (2011). "Using word segmentation and SVM to assess readability of Thai text for primary school students.", *Proceedings of The 2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Nakhon Pathom, Thailand: 11-13 May, 2011.
- Daowadung, P., and Chen, Y.-H. (2012). "Stop Word in Readability Assessment of Thai Text.", *Proceedings of The 12th International Conference on Advanced Learning Technologies (ICALT)*, Rome, Italy: 4-6 July, 2012.
- Dascalu, M. (2014). *Analyzing Discourse and Text Complexity for Learning and Collaborating*, Springer International Publishing, Switzerland.
- Dave, C., and Ruth, C. (2017). "The Flesch Reading Ease and Flesch-Kincaid Grade Level." (Online) Available on <https://readable.io/content/the-flesch-reading-ease-and-flesch-kincaid-grade-level/> (5 December 2016).
- Feng, L., Elhadad, N., and Huenerfauth, M. (2009) "Cognitively motivated features for readability assessment.", *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, Athens, Greece: 2 April, 2009.
- Flesch, R. (1948). "A new readability yardstick." *Journal of applied psychology*, 32(3), 221.
- François, T., and Fairon, C. (2012). "An AI readability formula for French as a foreign language.", *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju Island, Korea: 12-14 July, 2012.

เอกสารอ้างอิง (ต่อ)

- Graesser, A. C., McNamara, D. S., Louwrese, M. M., and Cai, Z. (2004). "Coh-Metrix: Analysis of text on cohesion and language." *Behavior Research Methods*, 36(2), 193-202.
- Han, J., Pei, J., and Kamber, M. (2011). *Data Mining: Concepts and Techniques*, Elsevier Inc., United States of America.
- Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M., Juffs, A., and Wilson, L. (2010). "Personalization of reading passages improves vocabulary acquisition." *International Journal of Artificial Intelligence in Education*, 20(1), 73-98.
- Kidwell, P., Lebanon, G., and Collins-Thompson, K. (2011). "Statistical estimation of word acquisition with application to readability prediction." *Journal of the American Statistical Association*, 106(493), 21-30.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated readability index, fog count, and flesch reading ease formula) for navy enlisted personnel, Naval Technical Training Command, United States of America.
- Kitson, H. D. (1921). *The mind of the buyer: A psychology of selling*, Macmillan, London, England.
- Ma, J., Xu, W., Sun, Y.-h., Turban, E., Wang, S., and Liu, O. (2012). "An ontology-based text-mining method to cluster proposals for research project selection." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(3), 784-790.
- Mitchell, J. (1987). "The Ninth Mental Measurements Yearbook." *Journal of Educational Measurement*, 24(1), 90-96.
- Pang, L. T. (2006). "Chinese readability analysis and its applications on the internet.", Master of Philosophy in Computer Science and Engineering, Chinese University of Hong Kong.
- Patteera, T. (2009). "A new vocabulary-based readability index for Thai university students.", Doctor of Philosophy in English Language Studies, Suranaree University of Technology.

เอกสารอ้างอิง (ต่อ)

- Pilán, I., Volodina, E., and Johansson, R. (2014). “Rule-based and machine learning approaches for second language sentence-level readability.”, *Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Applications*, Baltimore, Maryland, United States of America: 26 June, 2014.
- Pitler, E., and Nenkova, A. (2008). “Revisiting readability: A unified framework for predicting text quality.”, *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, Honolulu, Hawaii: 25-27 October, 2008.
- Saddiki, H., Bouzoubaa, K., and Cavalli-Sforza, V. (2015). “Text readability for Arabic as a foreign language.”, *Proceedings of the 2015 IEEE/ ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, Marrakech, Morocco: 17-20 November, 2015.
- Sato, S., Matsuyoshi, S., and Kondoh, Y. (2008). “Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus.”, *Proceedings of the 6th edition of the Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco: 28-30 May, 2008.
- Senter, R., and Smith, E. A. (1967). *Automated readability index*, University of Cincinnati, Ohio, United States of America.
- Sitbon, L., and Bellot, P. (2008). “A readability measure for an information retrieval process adapted to dyslexics.”, *Proceedings of the Second International Workshop on Adaptive Information Retrieval (AIR 2008)*, London, England: 18 October, 2008.
- Sjöholm, J. (2012). “Probability as readability: A new machine learning approach to readability assessment for written Swedish.”, Master Thesis in Human Centered Systems Division, Department of Computer and Information Science, Linköping University.
- Sung, Y.-T., Chen, J.-L., Cha, J.-H., Tseng, H.-C., Chang, T.-H., and Chang, K.-E. (2015). “Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning.” *Behavior research methods*, 47(2), 340-354.

เอกสารอ้างอิง (ต่อ)

- Svadasa, T., and Jhab, J. (2014). "A Literature Survey on Text Document Clustering and Ontology based techniques. " *International Journal of Innovative and Emerging Research in Engineering*, 1(2), 8-11.
- Tan, C., Gabrilovich, E., and Pang, B. (2012). "To each his own: personalized content selection based on text comprehensibility.", *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM)*, Washington, United States of America: 08-12 February, 2012.
- Todirascu, A., François, T., Gala, N., Fairon, C., Ligozat, A.-L., and Bernhard, D. (2013). "Coherence and cohesion for the assessment of text readability.", *Proceedings of the 10th International Workshop on Natural Language Processing and Cognitive Science (NLPCS)*, Marseille, France: 15-16 October, 2013.
- Tongtep, N., Coenen, F., and Theeramunkong, T. (2014). "Content-Based Readability Assessment: A Study Using A Syllabic Alphabetic Language (Thai).", *Proceedings of 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI)*, Queensland, Australia: 1-5 December, 2014.
- Tongtep, N., and Theeramunkong, T. (2011). "Simultaneous character-cluster-based word segmentation and named entity recognition in Thai language.", *Proceedings of the 6th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, Chiang Mai, Thailand: 25-27 November, 2010.
- Vor Der Brück, T., and Hartrumpf, S. (2007). "A semantically oriented readability checker for German.", *Proceedings of the 3rd Language & Technology Conference (LTC)*, Poznań, Poland: 5-7 October, 2007.
- Wang, Y. (2006). "Automatic recognition of text difficulty from consumers health information.", *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, Utah, United States of America: 22-23 June, 2006.

เอกสารอ้างอิง (ต่อ)

Zhang, L., Liu, Z., and Ni, J. (2013). "Feature-based assessment of text readability.", *Proceedings of the 2013 Seventh International Conference on Internet Computing for Engineering and Science (ICICSE)*, Shanghai, China: 20-22 September, 2013.

ภาคผนวก

ภาคผนวก ก

เอกสารหัวเรื่อง Data mining

ในภาคผนวก ก เป็นเอกสารข้อความดิจิทัลหัวเรื่อง Data mining ที่ค้นคืนจากโปรแกรมค้นหาของกูเกิล ซึ่งเป็นข้อความที่ได้จัดเก็บจากเอกสารต้นฉบับที่ผ่านกระบวนการจำกัดจำนวนอักขระ (สามารถดูรายละเอียดที่ข้อ 3.1.2) ทั้งนี้คำที่สะกดผิด หรือการใช้เครื่องหมายวรรคตอนที่ไม่ถูกต้อง เกิดจากการเรียงเรียงของเจ้าของข้อความนั้น ๆ ซึ่งถือเป็นปรากฏการณ์ทางภาษาดังนั้นผู้วิจัยจึงไม่ได้แก้ไขคำหรือข้อความใด ๆ

ฉบับที่	1
ชื่อเรื่อง	เหมืองข้อมูล (data mining)
เนื้อหา	

ก่อนที่จะพูดถึงเรื่อง data mining เรามาดูภาพนี้ก่อนนะครับว่า data mining อยู่ใน ส่วนไหน ซึ่งจากภาพ ขอสรุปการทำข้อมูลเป็น layer ไว้ดังรูป img_bi ส่วนที่ 1 Data Source เป็น ข้อมูลจากแหล่งต่าง ๆ ที่เราต้องการเก็บรวมไว้บน data warehouse ของเราไม่ว่าจะเป็น customer, product, Order และอื่น ๆ ซึ่งเป็นข้อมูลภายในองค์กรและข้อมูลภายนอกที่ต้องนำมาใช้ประโยชน์ เช่น location base, wiki, social, etc. ส่วนที่ 2 Data Management Platform แน่นอนว่าข้อมูลที่เราก็ก๊เก็บ และทำการ Transformation มาจากแหล่งต่าง ๆ (data source) จะถูกนำไปใช้งานต่อเช่น ทำ report, data analytic, OLAP, data mining มันจะต้องอยู่ในรูปที่สามารถนำไปใช้งานต่อได้สะดวกรวดเร็ว (speed) และมีคุณภาพ (data quality) เพราะฉะนั้นในการออกแบบ data warehouse เพื่อเก็บข้อมูล ในส่วนนี้จึงถือว่ามีความสำคัญ ส่วนที่ 3 BI Platform เป็นส่วนที่เรานำข้อมูลมาทำการวิเคราะห์ (data analytics) ซึ่งอาจจะใช้เทคนิคที่เรียกว่า OLAP (Online Analytical Processing) เพื่อให้การ วิเคราะห์ข้อมูลในมุมมองต่าง ๆ สามารถทำได้อย่างมีประสิทธิภาพ นอกจากนี้เราสามารถจะใช้ เทคนิคขั้นสูงเพื่อจำแนกข้อมูล และหาความน่าจะเป็น โดยใช้ data mining ซึ่งจะมาแชร์ให้ฟังต่อไป ครับ ant-icon พอจะเห็นภาพแล้วใช่ไหมครับว่า data mining ที่เราจะเริ่มต้นศึกษามันอยู่ในส่วน ไหน ต่อไปมาทำความเข้าใจเกี่ยวกับ data mining กันต่อเลยครับ อ้างอิงที่มา รูป : <http://bdcservicesgroup.com> Data Mining คืออะไร? Data Mining คือ การวิเคราะห์ข้อมูลจากข้อมูล จำนวนมาก (big data) เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้น เพื่อให้ได้อรรถความรู้ใหม่ ที่ สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่าง ๆ เช่น ตลาดหลักทรัพย์,ทางธุรกิจ, ทางด้าน การแพทย์, ยุทธศาสตร์ทหาร เป็นต้น ant-iconจากที่ผมได้ไปทำการศึกษาเกี่ยวกับ การทำเหมือง ข้อมูล (data mining) ขอสรุป Algorithm of Data mining ที่ใช้กันบ่อย ๆ ไว้ประมาณนี้ครับ จริงมี Algorithm เยอะกว่านี้ละค่อยๆศึกษากัน ไปนะครับ “วันนี้พูดถึง 8 กระบวนท่า algorithm of data mining” img_algorithm Descriptive Modeling : Unsupervised Learning 1. Association Algorithm เป็นอัลกอริทึมการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ (Big Data) เพื่อนำไปใช้ใน การวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ หรือมาจากการวิเคราะห์การซื้อสินค้าของลูกค้าที่ เรียกว่า “Market Basket Analysis” โดยนำ transaction การซื้อสินค้ามาทำการค้นหาวิเคราะห์ ว่า ลูกค้าใช้อะไรคู่กับสินค้าอะไรบ่อย ๆ ทำให้สามารถออกโปรโมชั่นแคมเปญคู่กัน ทำให้มีราคาถูกลง เพื่อเพิ่มมูลค่าให้กับสินค้า ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้

เป็นการใช้ “กฎความสัมพันธ์” (Association Rule) เพื่อหาความสัมพันธ์ของข้อมูล อ่านเพิ่มเติม
 2. Clustering Algorithm เป็นเทคนิคที่ใช้ในการจำแนกกลุ่ม

ฉบับที่	2
ชื่อเรื่อง	การทำเหมืองข้อมูล
เนื้อหา	

การทำเหมืองข้อมูล (อังกฤษ: data mining) หรืออาจจะเรียกว่า การค้นหาความรู้
 ในฐานข้อมูล (อังกฤษ: Knowledge Discovery in Databases - KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ
 (pattern) ของจากข้อมูลจำนวนมากมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของ
 เครื่อง และ การรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับ
 ข้อมูล(โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ใน
 ชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ ความรู้ที่ได้
 จากการทำเหมืองข้อมูลมีหลายรูปแบบ ได้แก่ กฎความสัมพันธ์ (Association rule) แสดง
 ความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง
 เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้า
 ออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวีดีโอ มักจะซื้อเทป
 กาวด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่า
 หลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะ
 ซื้อหนังสือ ก ได้ การจำแนกประเภทข้อมูล (Data classification) หากฎเพื่อระบุประเภทของวัตถุ
 จากคุณสมบัติของวัตถุ เช่น หากความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดย
 ใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัย
 ทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้หรือหนี้เสีย เพื่อประกอบการ
 พิจารณาการอนุมัติเงินกู้ การแบ่งกลุ่มข้อมูล (Data clustering) แบ่งข้อมูลที่มีลักษณะคล้ายกัน
 ออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการ
 วิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน การสร้างมโนภาพ
 (Visualization) สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทน
 การใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ่อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์
 ขั้นตอนการทำเหมืองข้อมูล ทำความเข้าใจปัญหา ทำความเข้าใจข้อมูล เตรียมข้อมูล สร้าง
 แบบจำลอง ประเมิน นำไปใช้งาน ประโยชน์จากการทำเหมืองข้อมูล การทำเหมืองข้อมูล

จำเป็นต้องอาศัยบุคลากรจากหลายฝ่าย และต้องอาศัยความรู้จำนวนมาก ถึงจะได้รับประโยชน์อย่างแท้จริง เพราะสิ่งที่ได้จากขั้นตอนวิธีเป็นเพียงตัวเลข และข้อมูล ที่อาจจะนำไปใช้ประโยชน์ได้หรือใช้ประโยชน์อะไรไม่ได้เลยก็เป็นได้ ผู้ที่ศึกษาการทำเหมืองข้อมูลจึงควรมีความรู้รอบด้านและต้องติดต่อกับทุก ๆ ฝ่าย เพื่อให้เข้าใจถึงขอบเขตของปัญหาโดยแท้จริงก่อน เพื่อให้การทำเหมืองข้อมูลเกิดประโยชน์อย่างแท้จริง การประยุกต์ใช้ Data Mining การขายปลีกและขายส่ง การธนาคาร การประดิษฐ์และการผลิต การประกันภัย การทำงานของตำรวจ การดูแลสุขภาพ การตลาด

ฉบับที่ 3

ชื่อเรื่อง Data Mining เมื่อต้องหาคะไรบางอย่างจากเหมืองข้อมูล?

เนื้อหา

เดต้าไมนิ่ง (Data Mining) ถ้าแปลตรง ๆ คือเหมืองข้อมูล คล้ายกับเหมืองแร่ ที่ขุดดินมาเยอะแต่ได้แร่ชนิดเดียว เดต้าไมนิ่งเป็นศาสตร์แขนงหนึ่งทาง AI (Artificial Intelligence) ดังนั้นข้อมูลที่จะพูดถึงต่อไป ท่านที่เคยศึกษา ด้าน AI (Artificial Intelligence) มาบ้างอาจคุ้นเคยเพราะหลายเทคนิคใน AI (Artificial Intelligence) ได้นำมาใช้กับเดต้าไมนิ่ง เดต้าไมนิ่งเป็นการค้นหาความสัมพันธ์และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล ซึ่งสัมพันธ์และรูปแบบเหล่านั้นได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมากที่มีอยู่ เดต้าไมนิ่งจะทำการสำรวจและวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎ โดยความสัมพันธ์หน่วยนี้แสดงให้เห็นถึงความรู้ต่าง ๆ (Knowledge) ที่มีประโยชน์ในฐานข้อมูลในปัจจุบันองค์กรธุรกิจส่วนใหญ่เผชิญกับปัญหาของ ข้อมูลดิบจำนวนมากแต่ข้อมูลที่ประยุกต์ใช้ได้มีน้อย เดต้าไมนิ่งจึงเป็นสาขาที่คาดว่าจะเป็นที่รู้จักและนำมาใช้ประยุกต์ใช้อย่างแพร่หลายเนื่องจากเดต้าไมนิ่งสามารถดึงความรู้ออกมาจากข้อมูลจำนวนมากที่ถูกเก็บสะสมและซ่อนไว้ ในโลกปัจจุบันธุรกิจต่าง ๆ จะพยายามหาเทคนิคที่สามารถนำความสำเร็จมาสู่บริษัท เช่น ในโลกธุรกิจขนาดย่อมจะสร้างความสัมพันธ์กับลูกค้า โดยสังเกตจากความต้องการ ความชอบและความสนใจของลูกค้า และอาจมีการเรียนรู้ได้จากผลสะท้อนในอดีตว่าจะทำอย่างไรให้การบริการลูกค้ามีประสิทธิภาพดีขึ้นในอนาคต เช่น บริษัทที่เป็นผู้ออกบัตรเครดิตและธนาคารต่าง ๆ จะมีขบวนการที่ใช้เดต้าไมนิ่งให้เป็นประโยชน์ ในการตัดสินใจว่าลูกค้ากลุ่มใดเป็นกลุ่มที่ดี ทำความเข้าใจลูกค้า ช่วยในการแยกประเภทของลูกค้าและจะทำนายกลุ่มของประชากรที่คาดว่าจะมาเป็นลูกค้าในอนาคต เป็นต้น อย่างไรก็ตามการเรียนรู้นั้นต้องมากกว่าการเก็บสะสมข้อมูลอย่างตรงไปตรงมา ซึ่งจะทำให้การทำงานเป็นไป อย่างมีประสิทธิภาพ วัฏจักรขั้นตอนการทำงานของเดต้าไมนิ่ง วัฏจักรขั้นตอนการทำงานของเดต้าไมนิ่ง

ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้ 1. การระบุโอกาสทางธุรกิจหรือการระบุปัญหาที่เกิดขึ้นกับธุรกิจ เป็นการระบุขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดหรือเพื่อนำมาทำการแก้ไขปัญหา 2. ส่วนของเดต้าไมนิ่ง เป็นการนำเทคนิคของเดต้าไมนิ่ง ไปใช้ ถ้ายทอดหรือทำการเปลี่ยนแปลงข้อมูลคิบบให้อยู่ในรูปของข้อมูลที่จะนำไปใช้ได้จริงในทางธุรกิจ 3. การปฏิบัติตามข้อมูล คือการนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วนเดต้าไมนิ่งมาลองปฏิบัติจริงกับธุรกิจ 4. การวัดประสิทธิภาพจากผลลัพธ์ การวัดประสิทธิภาพของเทคนิคของเดต้าไมนิ่งที่จะนำมาใช้จากผลลัพธ์ ซึ่งสามารถตรวจสอบได้หลายทาง เช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ เป็นต้น จากทั้ง 4 ขั้นตอนที่กล่าวมาข้างต้นคือการนำเอาเดต้าไมนิ่งไปใช้กับระบบทางธุรกิจ โดยแต่ละขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายมาเป็นอินพุตจากอีกขั้นตอนต่อไป ซึ่งเดต้าไมนิ่งจะเปลี่ยนข้อมูลคิบบให้เป็นข้อ

ฉบับที่	4
ชื่อเรื่อง	Data Warehouse & Data Mining
เนื้อหา	

ครั้งนี้เป็นการสรุปครั้งที่ 3 แล้ว คิดใจจริง ๆ แหะละ ถ้าสรุปกดยังไงก็ comment ได้เนาะครับ ไม่ว่างกัน ฮ่า ๆ Business Intelligence เพิ่มความสามารถในการทำธุรกิจ นำข้อมูลมาวิเคราะห์เพื่อช่วยในการตัดสินใจกลยุทธ์ Data Warehousing : คลังข้อมูล จะแบ่งโครงสร้างตามเนื้อหา เก็บข้อมูลไปเพื่อวิเคราะห์ช่วยในการตัดสินใจ เป็นข้อมูลที่จำเป็น (ตัดสิ่งที่ไม่จำเป็นออก) คุณสมบัติ จัดเก็บตามเนื้อหาลูกค้า การรวมเป็นหนึ่งในข้อมูล มีการปรับข้อมูลให้สอดคล้องการ เช่น ค่าตัวแปร ความสัมพันธ์กับเวลา (ข้อมูลยอดขายแต่ละปี) จะไม่มีการเปลี่ยนแปลง (Operation จะเปลี่ยน) สามารถโหลดข้อมูลมาดูเท่านั้น Data Mart : คลังข้อมูลขนาดเล็กที่ออกแบบสำหรับธุรกิจเชิงกลยุทธ์ ความแตกต่างระหว่าง Warehouse กับ Mart ขอบเขต : M จะเก็บข้อมูลย่อย ๆ แหล่งข้อมูล : M จะมีเอาแต่แหล่งข้อมูลของเรื่องนั้น ๆ ขนาด : M ขนาดเล็ก ระยะเวลาการพัฒนา : M สร้างง่ายกว่า OLAP : การประมวลผลออนไลน์เชิงวิเคราะห์ (Online Analytical Processing) : วิเคราะห์ข้อมูลในหลายมิติ เช่น วิเคราะห์ข้อมูลแต่ละแผนก แต่ละเรื่อง วิเคราะห์มุมมองทางด้าน time series และ trend analysis ซึ่งเป็นการวิเคราะห์แบบตาราง และกราฟ CRM Analysis Data Mining : กระบวนการค้นหาข้อมูลและวิเคราะห์ และตัดสินใจอย่างรวดเร็ว Market Management : จัดกระเข้าปีใหม่ ส่งจดหมายตรงกลุ่มลูกค้า รับรู้เพิ่มขึ้นว่าลูกค้าต้องการอะไร Risk Analysis : Forecasting (Uncertainty) Fraud Detection : ตรวจจับกม โกง Web mining : เอกสาร ขนาดข้อมูล

เติบโตเกินความสามารถของมนุษย์ในการวิเคราะห์ด้วยตนเอง แก้ปัญหาที่ซับซ้อนที่เกิดขึ้นใหม่ แหล่งข้อมูลในการทำ Data Mining 1. Database, 2. Data Warehouses, 3. Flat File Data mining คือ ? Data mining คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น หรือวิเคราะห์ข้อมูลเชิงลึก, แยกข้อมูลที่ไม่ทราบมาก่อน มาใช้ตัดสินใจในธุรกิจ Knowledge Discovery in Database (KDD) กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น การบวนการทำงาน Data Mining 1. Business Obj. Objective Determination กำหนดวัตถุประสงค์ เช่น เพื่อแบ่งกลุ่มลูกค้า เพื่อหาลักษณะลูกค้า กำหนดปัญหา และ โอกาส 2. Selection กำหนดข้อมูลทั้งหมด แล้วเลือกกลุ่มของข้อมูลที่ต้องการ 3. Data Preparation เตรียมข้อมูล กำหนดเป้าหมายในการเตรียมข้อมูลที่มีคุณภาพ เพื่อมั่นใจว่าเลือกข้อมูลถูกต้อง และ Transformation สร้างข้อมูลชุดใหม่ Data Selection เลือกข้อมูล เพศ อายุ รายได้ Data Preprocessing ทำข้อมูลให้มีคุณภาพดี Data Transformation สร้างข้อมูลชุดใหม่ เพื่อให้สอดคล้องกับ model ที่จะมา map 4. Data mining Analysis of Results เลือก model technique Predictive : สร้างจากข้อมูลเก่า เพื่อทำนายข้อมูลอนาคต = Neural network Segmentation : กลุ่ม

ฉบับที่	5
ชื่อเรื่อง	Data Mining การทำเหมืองข้อมูล
เนื้อหา	

ดาต้าไมนิ่ง(Data Mining) ถ้าแปลตรง ๆ คือเหมืองข้อมูล คล้ายกับเหมืองแร่ ที่ขุดดินมาเยอะแต่ได้แร่ชนิดเดียว Data Mining เป็นศาสตร์แขนงหนึ่งทาง AI (Artificial Intelligence) ดังนั้นข้อมูลที่จะพูดถึงต่อไป ท่านที่เคยศึกษา ด้าน AI (Artificial Intelligence) มาบ้างอาจคุ้นเคย เพราะหลายเทคนิคใน AI (Artificial Intelligence) ได้นำมาใช้กับ Data Mining ซึ่งเป็นการค้นหาความสัมพันธ์และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล ซึ่งสัมพันธ์และรูปแบบเหล่านั้นได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมากที่มีอยู่ Data Mining จะทำการสำรวจและวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎ โดยความสัมพันธ์หน่วยนี้แสดงให้เห็นถึงความรู้ต่าง ๆ (Knowledge) ที่มีประโยชน์ในฐานข้อมูลในปัจจุบันองค์กรธุรกิจส่วนใหญ่เผชิญกับปัญหาของ ข้อมูลดิบจำนวนมากแต่ข้อมูลที่ประยุกต์ใช้ได้นั้นมีน้อย Data Mining จึงเป็นสาขาที่คาดว่าจะเป็นที่รู้จักและนำมาใช้ประยุกต์ใช้อย่างแพร่หลายเนื่องจาก Data Mining สามารถดึงความรู้ ออกมาจากข้อมูลจำนวนมากที่ถูกเก็บสะสมและซ่อนไว้ ในโลกปัจจุบันธุรกิจต่างๆจะพยายามหา

เทคนิคที่สามารถนำความสำเร็จมาสู่บริษัท เช่น ในโลกธุรกิจขนาดย่อมจะสร้างความสัมพันธ์กับลูกค้า โดยสังเกตจากความต้องการ ความชอบและความสนใจของลูกค้า และอาจมีการเรียนรู้ได้จากผลสะท้อนในอดีตว่าจะทำอย่างไรให้การบริการลูกค้ามีประสิทธิภาพดีขึ้นในอนาคต เช่น บริษัทที่เป็นผู้ออกบัตรเครดิตและธนาคารต่าง ๆ จะมีขบวนการที่ใช้ Data Mining ให้เป็นประโยชน์ ในการตัดสินใจว่าลูกค้ากลุ่มใดเป็นกลุ่มที่ดี ทำความเข้าใจลูกค้า ช่วยในการแยกประเภทของลูกค้าและจะทำนายกลุ่มของประชากรที่คาดว่าจะมาเป็นลูกค้าในอนาคต เป็นต้น อย่างไรก็ตามการเรียนรู้นี้ต้องมากกว่าการเก็บสะสมข้อมูลอย่างตรงไปตรงมา ซึ่งจะทำให้การทำงานเป็นไป อย่างมีประสิทธิภาพ วัฏจักรขั้นตอนการทำงานของ Data Mining วัฏจักรขั้นตอนการทำงานของ Data Mining ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้ 1. การระบุโอกาสทางธุรกิจหรือการระบุปัญหาที่เกิดขึ้นกับธุรกิจ เป็นการระบุขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดหรือเพื่อนำมาทำการแก้ไขปัญหา 2. ส่วนของ Data Mining เป็นการนำเทคนิคของ Data Mining ไปใช้ถ่ายทอดหรือทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปของข้อมูลที่จะนำไปใช้ได้จริงในทางธุรกิจ 3. การปฏิบัติตามข้อมูล คือการนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วน Data Mining มาลองปฏิบัติตามจริงกับธุรกิจ 4. การวัดประสิทธิภาพจากผลลัพธ์ การวัดประสิทธิภาพของเทคนิคของ Data Mining ที่จะนำมาใช้จากผลลัพธ์ ซึ่งสามารถตรวจสอบได้หลายทาง เช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ เป็นต้น จากทั้ง 4 ขั้นตอนดังกล่าวมาข้างต้นคือการนำเอา Data Mining ไปใช้กับระบบทางธุรกิจ โดยแต่ละขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายมาเป็นอินพุตจากอีกขั้นตอนต่อไป ซึ่ง Data Mining จะเปลี่ยนข้อมูลดิบให้

ฉบับที่	6
ชื่อเรื่อง	Data Mining
เนื้อหา	

Data Mining คืออะไร? Data Mining คือ การสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Databases : KDD) เป็นเทคนิคที่ใช้จัดการกับข้อมูลขนาดใหญ่ โดยจะนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้ หรือสิ่งสำคัญออกมา เพื่อใช้ในการวิเคราะห์ หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้น ซึ่งการค้นหาคำความรู้ ความจริงที่แฝงอยู่ในข้อมูล (Knowledge Discovery) เป็นกระบวนการขุดค้นสิ่งที่น่าสนใจในกองข้อมูลที่เรามีอยู่ Data Mining มีความหมายแตกต่างกันใน 2 แง่มุม คือ ในมุมมองทางวิชาการและในมุมมองเชิงธุรกิจ ในมุมมองเชิงวิชาการนั้น นักวิจัยจะอ้างถึงกระบวนการทั้งหมดในการทำ Data

Mining ว่า “Knowledge discovery in database (KDD)” และใช้คำว่า “Data Mining” แทนขั้นตอน
 ขั้นหนึ่งของกระบวนการ ที่เกี่ยวข้องกับการค้นหารูปแบบ ความสัมพันธ์ของข้อมูลเท่านั้น ในแง่มุม
 เชิงธุรกิจแล้ว จะใช้คำว่า “Data Mining” แทนความหมายของ ขั้นตอนทั้งหมด วิวัฒนาการของ
 Data Mining ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่
 นำเชื่อถือ และป้องกันการสูญหายได้เป็นอย่างดี ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมา
 สร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อประโยชน์ในการนำไปวิเคราะห์และการตัดสินใจอย่างมี
 คุณภาพ ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปใน
 ฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกแง่มุมขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ ปี
 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการสร้าง
 แบบจำลอง และความสัมพันธ์ ทางสถิติ ข้อแตกต่างระหว่างระบบฐานข้อมูล กับ Data Mining
 ระบบ Data Mining เราไม่ต้องเป็นคนกำหนดคำสั่ง (เช่น SQL) เพื่อค้นหาข้อมูลที่ต้องการ แต่ระบบ
 Data Mining จะมีขบวนการ/วิธีการ (ซึ่งปกติจะเป็น machine learning tools) เพื่อทำหน้าที่นี้ นั่นคือ
 เราแค่บอกว่าเราต้องการอะไร (what to be mined) แต่ไม่จำเป็นต้อง ระบุว่า ทำอย่างไร (How to
 mine) ระบบฐานข้อมูลทั่วไป จะบังคับให้เราต้องทำทั้งสองหน้าที่นี้ คือคิดก่อนว่าจะ ค้นหาอะไร
 แล้วก็ไปประคิษฐ์คำสั่ง SQL เพื่อค้นหาข้อมูลนั้น ดังนั้นถ้าเราคิดไม่รอบคอบ หรือคิดผิดแล้วแต่แปล
 เป็นคำสั่งผิด ก็จะได้ข้อมูลผิด ๆ หรือไม่ตรงกับความต้องการ หลักการของ Knowledge Discovery
 from very large Databases (KDD) และ Data Mining จากที่ Data Mining เป็นขั้นตอนหนึ่งที่สำคัญ
 ในกระบวนการค้นหาลักษณะแฝงของข้อมูล ที่มีประโยชน์ในฐานข้อมูล (Knowledge Discovery in
 Database : KDD) กระบวนการของ KDD นั้นประกอบด้วยขั้นตอนต่าง ๆ ดังนี้ การคัดเลือกข้อมูล
 (Data Selection) ที่จะนำมาใช้ในการทำ mining รวมถึง การนำข้อมูลที่ต้องการออกมาจาก
 ฐานข้อมูลเพื่อทำการพิจารณาในเบื้องต้นว่าควรจะต้องเลือกข้อมูลใดมาใช้งาน การกรองข้อมูล (Data
 Cleaning) เป็นการกรั่นกรองข้อมูลที่ไม่ถูกต้องออกไปจากกลุ่มข้อมูลที่เลือกใช้ ทำให้มั่นใจใน
 คุณภาพของข้อมูลว่าถูก

ฉบับที่	7
ชื่อเรื่อง	DATA Mining
เนื้อหา	

Data Mining ถ้าแปลตรง ๆ คือเหมืองข้อมูล คล้ายกับเหมืองแร่ ที่ขุดดินมาเยอะแต่ได้แร่นิดเดียว Data Mining เป็นศาสตร์แขนงหนึ่งทาง AI (Artificial Intelligence) AI (Artificial Intelligence) ได้นำมาใช้กับ Data Mining ซึ่งเป็นการค้นหาความสัมพันธ์ และ รูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล ซึ่งความสัมพันธ์และรูปแบบเหล่านั้น ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมากที่มีอยู่ Data Mining จะทำการสำรวจและวิเคราะห์ข้อมูลให้อยู่ในรูปแบบที่เต็มไปด้วย ความหมาย และอยู่ในรูปของกฎ โดยความสัมพันธ์หน่วยนี้แสดงให้เห็นถึงความรู้ต่าง ๆ (Knowledge) ที่ มีประโยชน์ในฐานข้อมูลในปัจจุบันองค์กรธุรกิจส่วนใหญ่เผชิญกับปัญหาของ ข้อมูลดิบจำนวนมาก แต่ข้อมูลที่ประยุกต์ใช้ได้มีน้อย Data Mining จึงเป็นสาขาที่คาดว่าจะเป็นที่รู้จักและนำมาใช้ประยุกต์ใช้ อย่างแพร่หลายเนื่องจาก Data Mining สามารถดึงความรู้ออกมาจากข้อมูลจำนวนมากที่ ถูกเก็บสะสม และซ่อนไว้ วัฏจักรขั้นตอนการทำงานของ Data Mining วัฏจักรขั้นตอนการทำงานของ Data Mining ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้ 1. การระบุโอกาสทางธุรกิจหรือการระบุ ปัญหาที่เกิดขึ้นกับธุรกิจ เป็นการระบุขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์เพื่อหาความ ได้เปรียบทางการตลาด หรือเพื่อนำมาทำการแก้ไขปัญหา 2. ส่วนของ Data Mining เป็นการนำ เทคนิคของ Data Mining ไปใช้ถ่ายทอดหรือทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปแบบของข้อมูลที่จะนำไปใช้ได้จริงในทางธุรกิจ 3. การปฏิบัติตามข้อมูล คือการนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วน Data Mining มาลองปฏิบัติจริงกับธุรกิจ 4. การวัดประสิทธิภาพจากผลลัพธ์ การวัดประสิทธิภาพ ของเทคนิคของ Data Mining ที่จะนำมาใช้จากผลลัพธ์ ซึ่งสามารถ ตรวจสอบได้หลายทาง เช่น วัด จากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ เป็นต้น จากทั้ง 4 ขั้นตอนนี้ กล่าวมาข้างต้นคือการนำเอา Data Mining ไปใช้กับระบบทางธุรกิจ โดยแต่ละ ขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายมาเป็นอินพุตจากอีกขั้นตอนต่อไป ซึ่ง Data Mining จะเปลี่ยนข้อมูลดิบให้เป็นข้อมูลประยุกต์ ดังนั้นการระบุแหล่งข้อมูลที่ถูกต้องจึงเป็นสิ่งที่สำคัญ อย่างยิ่งต่อผลลัพธ์ที่ได้จากการวิเคราะห์ งานของ Data Mining (Task of data mining) ในทางปฏิบัติจริง Data Mining จะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ภายใต้ภาวะที่ จำกัดปัญหา เหมาะสมกับการใช้เทคนิค Data Mining จะเป็นปัญหาที่ต้องใช้เหตุผลในการแก้ เป็นปัญหาที่ เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ในรูปแบบของงาน ทั้ง 6 งานได้ ดังนี้ 1. การจัดหมวดหมู่ (Classification) 2. การประเมินค่า (Estimation) 3. การทำนายล่วงหน้า (Prediction) 4. การจัดกลุ่มโดยอาศัยความใกล้ชิด (Affinity Group) 5. การรวมตัว

(Clustering) 6. การบรรยาย (Description) ไม่มีเทคนิคหรือเครื่องมือเพียงชนิดเดียวของ Data Mining ที่เหมาะสมกับงานทุกชนิด งานในแต่ละชนิดก็จะมีเทคนิคของ

ฉบับที่	8
ชื่อเรื่อง	Data Mining
เนื้อหา	

ในปัจจุบันสภาวะการแข่งขันเพื่อให้ได้ชัยชนะทางธุรกิจจำเป็นต้องมีกลยุทธ์หรือยุทธวิธี (Business Strategies) ที่เชื่อมั่นได้ว่าจะลดความเสี่ยงขององค์กรลงได้ กลยุทธ์วิธีการต่าง ๆ จำเป็นต้องมีฐานความรู้ (Knowledge Base) เพื่อใช้ในการสร้างกรอบการทำงาน (Framework) ที่สนองตอบกับกลยุทธ์ทางธุรกิจ การที่จะได้มาซึ่งฐานความรู้และกรอบการทำงานที่มีประโยชน์ จำเป็นต้องมีเทคโนโลยีสารสนเทศที่สามารถวิเคราะห์ข้อมูลทางธุรกิจ เทคโนโลยีสารสนเทศที่สามารถกลั่นกรองข้อมูลทางธุรกิจที่มีปริมาณมหาศาล (Very large data in business information systems) เพื่อให้ได้ข้อมูลที่มีประโยชน์ในการคิดกลยุทธ์ (Usefull information from very large data) ดังนั้นในขณะนี้ Data Mining จึงเป็นเทคโนโลยีสารสนเทศที่ได้รับการกล่าวถึงมากที่สุด เหตุผลที่สำคัญสำหรับทำไมถึงมี Data Mining และทำไมถึงต้องทำ Data Mining นั้นก็เพราะว่า Data Mining เป็นเทคโนโลยีสารสนเทศที่สามารถกลั่นกรอง วิเคราะห์ ข้อมูลที่มีปริมาณมหาศาล เพื่อให้ได้ข้อมูลที่มีประโยชน์หรือได้ข้อมูลที่ซ่อนเร้นอยู่ในข้อมูลที่มีปริมาณมหาศาล และนำข้อมูลที่มีประโยชน์มาใช้เป็นฐานความรู้เพื่อช่วยในการบริหารงาน เช่น การบริหาร CRM (Customer Relationship Management) Data Mining คือขบวนการทำงานที่เรียกว่า process ที่สกัดข้อมูล (Extract data) จากฐานข้อมูลขนาดใหญ่ (Large Information) เพื่อให้ได้สารสนเทศ (Usefull Information) ที่เรายังไม่รู้ (Unknown data) โดยเป็นสารสนเทศที่มีเหตุผล (Valid) และสามารถนำไปใช้ได้ (Actionable) ซึ่งเป็นสิ่งสำคัญในการที่จะช่วยการตัดสินใจในการทำธุรกิจ Data Mining เป็นโปรเซสที่สำคัญในการทำ Knowledge Discovery in Database ที่เราเรียกสั้น ๆ ว่า KDD ส่วน Data Mining สามารถเรียกสั้น ๆ ว่า DM ขั้นตอนการทำ Data Mining มี 4 ขั้นตอนหลักดังนี้ 1. Business Object Determination เป็นตัวจักรที่สำคัญในการทำ KDD เนื่องจากเป็นกำหนด ขอบเขต เป้าหมาย ของการทำ KDD ซึ่งจะมีผลต่อทุก ๆ ขั้นตอนของการทำ KDD โดยนักวิเคราะห์ธุรกิจ (Business Analyst) จะต้อง Identify ปัญหาที่เกิดขึ้นในการทำธุรกิจให้ครอบคลุมและชัดเจนรวมทั้งวัตถุประสงค์ด้วย 2.Data Preparation หน้าที่ของขั้นตอนนี้คือจัดการข้อมูลให้สามารถนำเข้าสู่ อัลกอริทึมของ Data Mining ได้ เช่น การทำ Data Cleaning, Data Integration, Data Reduction เป็น

ต้น ซึ่ง Data Preparation สามารถแบ่งออกเป็น 3 ส่วนได้แก่ Data Selection, Data Preprocessing และ Data Transformation 3. Data Mining เป็นขั้นตอนการทำ Mining โดยมี operation ในการทำ Data Mining หลายแบบ เช่น Database Segmentation, Predictive Modeling, Link Analysis เป็นต้น แต่ละ Data Mining Operation จะมีอัลกอริทึมให้เลือกใช้ เช่น การทำ Database Segmentation อาจใช้ K-Mean Algorithms หรืออาจใช้ Unsupervised Learning Neural Networks เช่น โมเดล Kohonen Neural Net ถ้าเป็นการทำ Predictive Modeling อาจใช้ CART (Classification

ฉบับที่	9
ชื่อเรื่อง	Data Mining (การทำเหมืองข้อมูล) และ โปรแกรม Weka-Orange_Canvas-Matlab

เนื้อหา

การทำเหมืองข้อมูล (data mining) หรืออาจจะเรียกว่า การค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases – KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อมูลจำนวนมากโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ ความรู้ที่ได้จากการทำเหมืองข้อมูลมีหลายรูปแบบ ได้แก่ กฎความสัมพันธ์ (Association rule) แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุ ที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวีดีโอมักจะซื้อเทปกาวยด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้ การจำแนกประเภทข้อมูล (Data classification) หากกฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้ การแบ่งกลุ่มข้อมูล (Data clustering) แบ่งข้อมูลที่มีลักษณะคล้ายกัน ออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการ

วิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน จินตทัศน์(Visualization) สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทนการใช้ข้อความ นำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ่อนเร้นเมื่อคุณข้อมูลชุดนั้นด้วยจินตทัศน์ ขั้นตอนการทำเหมืองข้อมูล ทำความเข้าใจปัญหา ทำความเข้าใจข้อมูล เตรียมข้อมูล สร้างแบบจำลอง ประเมินนำไปใช้งาน ประโยชน์จากการทำเหมืองข้อมูล การทำเหมืองข้อมูล จำเป็นต้องอาศัยบุคลากรจากหลายฝ่าย และต้องอาศัยความรู้จำนวนมาก ถึงจะได้รับประโยชน์อย่างแท้จริง เพราะสิ่งที่ได้จากขั้นตอนวิธีเป็นเพียงตัวเลข และข้อมูล ที่อาจจะนำไปใช้ประโยชน์ได้หรือใช้ประโยชน์อะไรไม่ได้เลยก็เป็นได้ ผู้ที่ศึกษาการทำเหมืองข้อมูลจึงควรมีความรู้รอบด้านและต้องติดต่อกับทุก ๆ ฝ่าย เพื่อให้เข้าใจถึงขอบเขตของปัญหาโดยแท้จริงก่อน เพื่อให้การทำเหมืองข้อมูลเกิดประโยชน์อย่างแท้จริง WEKA โปรแกรม Weka (Waikato Environment for Knowledge Analysis) เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จรูป อยู่

ฉบับที่ 10
ชื่อเรื่อง การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining โดยซอฟต์แวร์ RapidMiner Studio 6

เนื้อหา

การพัฒนาด้าน Business Intelligenceเป็นการวิเคราะห์ข้อมูลในคลังข้อมูลเพื่อช่วยสนับสนุนการตัดสินใจในการบริหารจัดการองค์กร ประกอบด้วยระบบข้อมูลและโปรแกรม แอปพลิเคชันด้านการวิเคราะห์หลายระบบ เช่น ดาต้าแวร์เฮ้าส์ (Data Warehouse) ดาต้ามาร์ท (Data Mart) การทำเหมืองข้อมูล (Data Mining) การแก้ปัญหาทางคณิตศาสตร์ (Operations Research & Numerical Methods) เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลในหลายมิติ (OLAP) ระบบสืบค้นและออกรายงานต่าง ๆ การวิเคราะห์ข้อมูลด้วย Data Mining กำลังเป็นที่นิยมไปทั่วโลกด้วยแรงขับเคลื่อนอย่างหนึ่งคือ การมีซอฟต์แวร์ที่ช่วยให้ทำการวิเคราะห์ได้ง่ายขึ้น เหมาะกับการวิเคราะห์ข้อมูลที่มีขนาดใหญ่และจำนวนมาก เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล สำหรับการนำมาใช้เพื่อการบริหารและการตัดสินใจขององค์กร ซึ่ง Data Mining Tools ที่ได้รับความนิยมในประสิทธิภาพการใช้งานมากที่สุดในปี 2013 ได้แก่ซอฟต์แวร์ RapidMiner Studio 6 ซึ่งเป็น open source software ทำให้สามารถประหยัดค่าใช้จ่าย กรณีต้องการใช้งานเพื่อทดสอบการวิเคราะห์ข้อมูล การทำความสะอาดข้อมูล (Data cleansing) และการรายงานข้อมูลในรูปแบบต่าง ๆ เช่น ตาราง กราฟ เป็นต้น ความแตกต่างของ Database & Warehouse & Mining

Database คือ ฐานข้อมูลใช้ในการจัดเก็บข้อมูล ลดความซ้ำซ้อนของข้อมูล เน้นการจัดเก็บ เพิ่มแก้ไข และลบข้อมูล Data Warehouse คือ คลังข้อมูลรวบรวมข้อมูลจากหลาย ๆ ฐานข้อมูล แปลงข้อมูลให้มีความเหมือนกัน เหมาะสำหรับการเรียกดู (view) เพื่อสร้างรายงานสรุป Data Mining คือ การวิเคราะห์ข้อมูลเพื่อค้นหาความสัมพันธ์หรือรูปแบบที่มีประโยชน์ในฐานข้อมูล Data Mining คือ อะไร หนังสือ Data Mining Techniques ให้ความหมายไว้ว่า “เป็นการวิเคราะห์ข้อมูล เพื่อหารูปแบบ (patterns) หรือความสัมพันธ์ (relation) ระหว่างข้อมูลในฐานข้อมูลขนาดใหญ่” หนังสือ Data Mining Concepts & Techniques ให้ความหมายไว้ว่า “เป็นกระบวนการดึงข่าวสารที่น่าสนใจ และมีประโยชน์แต่ไม่เคยรู้มาก่อนจากฐานข้อมูลขนาดใหญ่” ความสามารถและการทำงานของซอฟต์แวร์ RapidMiner Studio 6 การนำเข้าข้อมูลได้หลายลักษณะ เช่น การเชื่อมโยงจากฐานข้อมูลโดยตรง, ไฟล์ Excel, ไฟล์ CSV เป็นต้น การเขียนไฟล์ให้อยู่ในรูปแบบของ Excel และ CSV การแสดงข้อมูลในกราฟแบบต่าง ๆ เช่น scatter plot, time series การเตรียมข้อมูล (Data Preparation) เพื่อให้การนำเข้าข้อมูลมีความถูกต้อง ก่อนการนำเข้าจึงต้องมีการจัดการข้อมูล (preprocessing) ดังนี้ 4.1 จัดการข้อมูลที่มีความผิดพลาดในชุดข้อมูล เช่น • ข้อมูลมีค่าไม่ตรง • ข้อมูลขาดหายไป (missing value) • ข้อมูลแปลกแยก (outlier) 4.2 แปลงข้อมูล เช่น • Discretization แปลงข้อมูล numeric ให้เป็น nominal โดยการแบ่งข้อมูลออกเป็นช่วง ๆ ได้แก่ แบ่งตามเงื่อนไขที่กำหนด แบ่งตามช่วงของข้อมูลที่เท่ากัน (equal width) และ แบ่งตามข้อมูลที่มีความถี่เท่ากัน (equal frequency) • Norma

ฉบับที่	11
ชื่อเรื่อง	Data Mining
เนื้อหา	

Data Mining การทำเหมืองข้อมูล คุณลักษณะและเป้าหมายหลักของ Data Mining คือ ค้นหา Pattern ของข้อมูลที่ฝังลึกและซ่อนเร้นอยู่ภายในฐานข้อมูลขนาดใหญ่ Knowledge Discovery Data (KDD) something infeasible for an human being to perform manually ขั้นตอนการทำ Data Mining 1. เรียนรู้และศึกษาเกี่ยวกับโปรแกรมที่จะใช้ เช่น Weka, Orange 2. Data selection (เลือกมาจาก DB หรือ Information Retrieve) 3. Data cleaning 4. Data reduction and transformation 5. เลือก Functions ของ data mining เช่น summarization, classification, regression, association และ clustering เป็นต้น 6. เลือก Algorithm ของ data mining 7. ทำการค้นหา Patterns ที่เราสนใจ 8. วิเคราะห์ผลลัพธ์ที่ได้ 9. ใช้อ็องค์ความรู้ที่ค้นพบ Functions ของ Data Mining มี ดังนี้ 1.

การวิเคราะห์คุณสมบัติและการแยกแยะข้อมูล (Characterization and discrimination) 2. การหาความสัมพันธ์ของข้อมูล (Association) - unsupervised 3. การจัดหมวดหมู่และการวิเคราะห์การถดถอย (Classification/ Regression) - supervised 4. การวิเคราะห์การรวมกลุ่ม หรือ การแบ่งแยกข้อมูล (Cluster analysis/ Segmentation) - unsupervised 5. การประเมินและการพยากรณ์ (Estimation/Prediction) 6. การบรรยายและการแสดงภาพของข้อมูล (Description / Visualization)

จะอธิบายเฉพาะ Functions ที่นิยมใช้บ่อยๆ → การจำแนกประเภทข้อมูล (Data Classification) เป็นการเรียนรู้แบบ “Supervised Learning” เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด (output) ไปได้ โดยจะนำข้อมูลส่วนหนึ่งมาสอน (Training Data) เพื่อให้เครื่องเรียนรู้ และจำแนกผลลัพธ์ตามที่กำหนดไว้ และนำข้อมูลส่วนที่เหลือ (Test Data) มาทดสอบตามโมเดลที่กำหนดไว้ ทำใหม่เรื่อยๆ จนกว่าจะได้ผลลัพธ์ ความถูกต้องในระดับที่พอใจ Algorithms ที่ใช้ในการแก้ปัญหาแบบ Classification ได้แก่ 1. Decision Tree 2. เครือข่ายประสาท (Neural Network) เป็นวิธีการให้เครื่องเรียนรู้จากตัวแบบ แล้วฝึก (Train) ให้ระบบได้รู้จักที่จะคิดแก้ปัญหาที่กว้างขึ้นได้ เลียนแบบวิธีคิดแบบมนุษย์ 3. Naive-Bayes จะใช้วิเคราะห์หาความน่าจะเป็นของสิ่งที่ยังไม่เคยเกิดขึ้น โดยการคาดเดาจากสิ่งที่เคยเกิดขึ้นมาก่อน นิยมใช้ เนื่องจากเป็นรูปแบบการหาความสัมพันธ์ที่ไม่ซับซ้อน ได้ผลลัพธ์ดี แต่ใช้กับข้อมูลแบบ continuous ไม่ได้ ต้องแปลงข้อมูลให้เป็น category ก่อน คือจัดกลุ่มให้ข้อมูลก่อน [เช่น ถ้าเป็นตัวเลข 1-3 ให้อยู่ในกลุ่มที่ 1 , ตัวเลข 4-6 ให้อยู่ในกลุ่มที่ 2] ตัวอย่าง Train on Time (วันฝนตก ลมแรง โอกาสที่ train on time จะเป็นเท่าไร)

4. K-nearest neighbor (K-NN) ไม่ต้องมี Training data ใช้ในกรณีที่เป็นค่าต่อเนื่อง (ตัวเลข) กำหนดกลุ่ม K กลุ่มไว้ก่อน แล้วกำหนดศูนย์กลางของแต่ละกลุ่ม แล้วดูว่า unseen instance อยู่ใกล้กลุ่มไหนมากที่สุด ด้วยการ plot กราฟ วัดค่าแบบ distance measures ทำไปเรื่อย ๆ

ภาคผนวก ข

เอกสารหัวเรื่อง iOS operating system

ในภาคผนวก ข เป็นเอกสารข้อความดิจิทัลหัวเรื่อง iOS operating system ที่ค้นคืนจากโปรแกรมค้นหาของกูเกิล ซึ่งเป็นข้อความที่ได้จัดเก็บจากเอกสารต้นฉบับที่ผ่านกระบวนการจำกัดจำนวนอักขระ (สามารถดูรายละเอียดที่ข้อ 3.1.2) ทั้งนี้คำที่สะกดผิด หรือการใช้เครื่องหมายวรรคตอนที่ไม่ถูกต้อง เกิดจากการเรียงเรียงของเจ้าของข้อความนั้น ๆ ซึ่งถือเป็นปรากฏการณ์ทางภาษา ดังนั้นผู้วิจัยจึง ไม่ได้แก้ไขคำหรือข้อความใด ๆ

ฉบับที่ 1

ชื่อเรื่อง iOS Apps คืออะไร?

เนื้อหา

คำว่า ไอโอเอส (iOS) โดยมีชื่อเดิมที่เรียกกันคือ ไอโฟนโอเอส (iPhone OS) คือระบบปฏิบัติการสำหรับสมาร์ตโฟนของบริษัท แอปเปิล (Apple Inc.) เป็นระบบที่ได้พัฒนาขึ้นสำหรับใช้ในโทรศัพท์ไอโฟน (iPhone) และต่อมาได้มีการพัฒนาต่อเพื่อใช้สำหรับไอพอดทัช (iPod touch) และไอแพด (iPad) ซึ่งระบบ iOS สามารถเชื่อมต่อไปยัง Apps Store สำหรับการเข้าถึงแอปพลิเคชันที่สามารถใช้งานได้บนระบบปฏิบัติการ iOS หรือที่เรียกกันว่า iOS Application หรือ iOS Apps ซึ่งมีการแบ่งเป็นหมวดการใช้งานสำหรับผู้ใช้อุปกรณ์ไอโฟนสามารถทำการโหลด iOS App มาใช้งานได้ตามต้องการ และนอกจากนี้ในป

ฉบับที่ 2

ชื่อเรื่อง iOS 9 อัปเดตข้อมูล สรุปทุกฟีเจอร์สำคัญ เพิ่มประสิทธิภาพ Siri, Apple Pay, Notes, Apple Maps ส่วน iPad รองรับการใช้งานสองแอปฯพร้อมกัน

เนื้อหา

หลังจากการเปิดงาน WWDC 2015 ที่มีการเปิดตัว iOS 9 อย่างเป็นทางการเมื่อหลายเดือนก่อน และในตอนนี้ก็ถึงคิวของพระเอกอย่าง iPhone 6s และ iPhone 6s Plus ที่ทำการเปิดตัวอย่างเป็นทางการไปเรียบร้อยแล้วในคืนวันที่ 9 กันยายนที่ผ่านมา และเป็นที่น่าอนว่า 2 รุ่นใหม่นี้ จะมาพร้อมกับระบบปฏิบัติการล่าสุดอย่าง iOS 9 รวมไปถึง iPad Pro กับ iPad mini 4 ที่ได้มีการเปิดตัวในวันดังกล่าวด้วย โดย iOS 9 นั้นมีการพัฒนาฟีเจอร์ที่เพิ่มขึ้นมาจำนวนไม่น้อย รวมถึงการปรับปรุงในแอปพลิเคชันเดิมที่มีอยู่แล้วให้เพิ่มประสิทธิภาพดียิ่งขึ้น ซึ่งทางทีมงาน Thaimobilecenter ได้รวบรวม

ฉบับที่	3
ชื่อเรื่อง	IT: ข้อมูลแรกระบบปฏิบัติการ iOS 8 และ OS X 10.10 เวอร์ชันต่อไปจาก Apple!

เนื้อหา

IT อัพเดทข่าวล่าสุดกับ ป้าเอก TechXcite ยังไม่ทันจะถึงงาน WWDC 2014 ที่ Apple จะจัดขึ้นในเดือนมิถุนายนโน่นเลยแต่ตอนนี้ปรากฏว่าเริ่มมีข้อมูลเกี่ยวกับ iOS 8 และ OS X 10.10 มาให้ได้ติดตามกันบ้างแล้วว่าจะมีอะไรใหม่ ๆ เพิ่มเข้ามาหรือปรับปรุงให้ดีขึ้นในระบบปฏิบัติการเวอร์ชันต่อไปของ Apple ทั้งบน iPhone, iPad, iPod Touch และผลิตภัณฑ์ตระกูล Mac ทั้งหมดครับ ส่วนจะจริงหรือไม่ไปรอลุ้นกันเดือนมิถุนายนกันได้นะจ๊ะ :D ข้อมูลแรก ระบบปฏิบัติการ iOS 8 iOS 8 สำหรับ iPhone, iPad และ iPod Touch จะใช้ชื่อโค้ดเนมว่า Okemo iOS 8 Healthbook iOS 8 จะรองรับแอปที่ซ

ฉบับที่	4
ชื่อเรื่อง	ความแตกต่างของ Android กับ iOS จุดเด่นส่วนตัวที่น่าสนใจของแต่ละระบบ

เนื้อหา

หลายคนคงเคยได้ยินคำถามที่ว่าจะไปเล่นมือถือค่ายไหนดี ณ ตอนนี้มี 2 ระบบหลัก ๆ คือระบบ Android และระบบ iOS ที่เป็นคู่แข่งกันอยู่ ซึ่งถ้าจะถามว่าระบบไหนดีกว่ากันนั้นมันก็จะเป็คำถามที่ตอบยากเลยทีเดียว แต่ถ้าหากจะพูดถึงจุดเด่นของแต่ละระบบแล้วอาจจะดูน่าฟังกว่า อีกทั้งยังสามารถวิเคราะห์ความต้องการทางด้านการใช้งานที่แท้จริงของผู้ใช้ได้ การจัดการเกี่ยวกับตัวเครื่องและระบบ Android - จะมีความเป็นอิสระค่อนข้างมาก เปรียบเสมือนคอมพิวเตอร์เครื่องหนึ่งได้เลย เชื่อมต่อภายนอกได้ง่ายเสียบคอมที่ไหนก็ได้ ถ่ายโอนข้อมูลง่ายเหมือนเสียบแฟลชไดร์ฟ ลงแอฟจากภายนอกได้

ฉบับที่	5
ชื่อเรื่อง	ห้ามพลาด! 50 เทคนิคการใช้งาน iOS 8 ที่ควรรู้
เนื้อหา	

มีเทคนิคการใช้งานและฟีเจอร์ที่เพิ่มเข้ามาใน iOS หลังจากที่ได้อัปเดตกันไป 8 ที่ผ่านมา สำหรับบางฟีเจอร์นั้นเราอาจจะยังไม่ทราบว่าคืออะไรบ้าง ไปดูกันครับว่า 17 แล้วเมื่อวันที่ iOS 8.0 นั้นจะมีอะไรซ่อนอยู่ในแบบที่เราไม่รู้กันบ้างครับ 8QuickType Suggestions บน iOS จะมีฟีเจอร์เดาคำศัพท์ล่วงหน้า สำหรับคน 8ที่เปิดใช้งานไว้สามารถซ่อนไม่ให้ชั่วคราวโดยการปิดแถบเดาคำศัพท์ด้านล่างเพื่อเป็นการซ่อนการใช้งานชั่วคราว หรือจะปิดการใช้งานไปเลย ดูวิธีการปิดที่นี้ Hide QuickType Suggestions การส่งรูปภาพและวิดีโอใน .2Message สามารถส่งได้ง่ายขึ้น

ฉบับที่	6
ชื่อเรื่อง	ประวัติและที่มาของ iOS
เนื้อหา	

iOS (ก่อนหน้า iPhone OS) เป็นระบบปฏิบัติการมือถือที่พัฒนาและจัดจำหน่ายโดยบริษัท Apple Incออกจำหน่ายในปี 2007 สำหรับ iPhone และ iPod Touch ของมันได้รับการขยายเพื่อรองรับอุปกรณ์ที่แอปเปิ้ลอื่น ๆ เช่น iPad และ โทรศัพท์แอปเปิ้ล ซึ่งแตกต่างจากไมโครซอฟท์ 's Windows CE (Windows โทรศัพท์) และ Google 's Android, แอปเปิ้ลไม่ได้ใบอนุญาตสำหรับการติดตั้ง iOS บนฮาร์ดแวร์ที่ไม่ใช่แอปเปิ้ล ณ วันที่ 12 มิถุนายน 2012, แอปเปิ้ลที่ App Storeมีมากกว่า 650,000 โปรแกรม iOS ซึ่งได้รับการเรียกรวมดาวน์โหลดมากกว่า 30 ล้านครั้ง มันมีส่วนแบ่ง 16% จากมาร์ทโฟนระบบปฏิบัติการ

ฉบับที่ 7

ชื่อเรื่อง ไอโอเอส

เนื้อหา

ไอโอเอส (ก่อนหน้าใช้ชื่อ ไอโฟนโอเอส) คือระบบปฏิบัติการบนอุปกรณ์พกพา (สมาร์ทโฟน, แท็บเล็ตคอมพิวเตอร์) พัฒนาและจำหน่ายโดยแอปเปิล (บริษัท) เปิดตัวครั้งแรกในปี เพื่อใช้บนไอโฟน และได้มีการพัฒนาเพิ่มเติมเพื่อใช้บนอุปกรณ์พกพาอื่น ๆ ของแอปเปิล 2007 (2007 เช่น ไอพอดทัช (ในเดือนกันยายน, ไอแพด (ในเดือนมกราคม (2010, ไอแพด มินิ (พฤศจิกายน 2012) แอปเปิลทีวี รุ่นที่ 2) ไอโอเอสแตกต่างจาก วินโดวส์โฟนของไมโครซอฟท์และแอนดรอยด์ (ระบบปฏิบัติการ) ของกูเกิล ตรงที่แอปเปิลไม่อนุญาตให้นำไอโอเอสไปติดตั้งบนอุปกรณ์ที่ไม่ใช่อุปกรณ์ของแอปเปิล ใน

ฉบับที่ 8

ชื่อเรื่อง รวมข้อมูลการปรับปรุงจุดต่าง ๆ ของ iOS 8.4 เวอร์ชันล่าสุด

เนื้อหา

แนะนำบริการฟังเพลง Apple Music - ประสบการณ์ฟังเพลงที่ปฏิวัติใหม่ สถานีวิทยุทั่วโลกตลอดเวลา - และที่สำหรับแฟนเพลงที่จะได้ใกล้ชิดกับศิลปินที่พวกเขาชื่นชอบ 7/24 ทั้งหมดนี้มีอยู่ในแอปเพลงที่ออกแบบใหม่หมด นอกจากนี้ iOS ยังมีการปรับปรุงและการแก้ไข 4.8 ข้อบกพร่องสำหรับ iBooks ด้วย รวมข้อมูลการปรับปรุงจุดต่างๆของ iOS เวอร์ชันล่าสุด 4.8 Apple Music เข้าร่วมเป็นสมาชิกของ Apple Music เพื่อฟังเพลงหลายล้านเพลงจากแคตตาล็อกใน Apple Music หรือจะเก็บไว้เล่นออฟไลน์ภายหลังก็ได้ สำหรับคุณ: สมาชิกสามารถดูเพลย์ลิสต์และคำแนะนำเกี่ยวกับอัลบั้มที่เลือกสรรโดย

ฉบับที่	9
ชื่อเรื่อง	วิธีอัปเดต iOS ให้ซัวร์ ข้อมูลไม่หาย และการแก้ปัญหาหน้าจอขึ้นให้เชื่อมต่อกับ iTunes
เนื้อหา	

ในช่วงที่ผ่านมา Apple ได้ปล่อย iOS 8 ให้ผู้ใช้ iPhone, iPad, iPod touch ได้ทำการอัปเดตกันแล้ว ซึ่งล่าสุดก็มีการปล่อยเวอร์ชัน 8.0.2 มาให้อัปเดตแก้บั๊กต่าง ๆ โดยผู้อุปกรณ์ที่รองรับการอัปเดตครั้งนี้ก็คง ได้ทำการอัปเดตกันไปบ้าง แต่ก็ยังมีผู้ใช้หลายคนที่ยังไม่ได้อัปเดตเนื่องจากกังวล กลัวปัญหา “ข้อมูลหาย” ซึ่งเป็นปัญหาที่มีผู้ใช้หลายคนเจอหลังจากการอัปเดตเสร็จสิ้น และที่หนักไปกว่านั้น ผู้ใช้บางรายอาจจะเจอปัญหา ระหว่างการอัปเดตที่ไม่สมบูรณ์ทำให้ปรากฏหน้าจอขึ้นให้เสียบกับ iTunes แต่สิ่งเหล่านี้ไม่ใช่เรื่องน่ากลัว หากมีการอัปเดตที่ถูกต้อง แล

ฉบับที่	10
ชื่อเรื่อง	พาย้อนอดีตมารู้จักกับเวอร์ชันต่าง ๆ ของ iOS ระบบปฏิบัติการบนอุปกรณ์พกพาจาก Apple กัน
เนื้อหา	

iOS ถือได้ว่าเป็นระบบปฏิบัติการบนสมาร์ตโฟนและแท็บเล็ตที่ได้รับความนิยมอย่างมากจากผู้บริโภคทั่วโลกในปัจจุบันและยังเป็นคู่แข่งคนสำคัญในตลาดสมาร์ตโฟนและแท็บเล็ตโลกของ Android OS อีกด้วย ซึ่งหลายคนอาจกำลังใช้งานผ่าน iPhone, iPad, iPod อยู่ แต่ก็อาจมีอีกหลายคนที่ยังไม่รู้จักรู้จักหรือรู้จักแบบคร่าว ๆ ดังนั้นเรามารู้จัก iOS ระบบปฏิบัติการจาก Apple กันให้มากขึ้นอีกนิดผ่านบทความนี้กัน iOS (iPhoneOS) คืออะไร? iOS หรือเดิมชื่อ iPhone OS เป็นระบบปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาขึ้นมาโดยบริษัท Apple Inc. (Apple Computer Inc.) โดยมีพื้นฐานในการพัฒนา

ฉบับที่	11
ชื่อเรื่อง	ไม่มี
เนื้อหา	

ระบบปฏิบัติการ Android ถ้าพูดถึงระบบปฏิบัติการ (Operating System) ทั้ง Windows, Linux และ Mac OS X ซึ่งเป็นระบบปฏิบัติการที่ใช้งานอยู่บนคอมพิวเตอร์ แต่รู้หรือไม่ โทรศัพท์มือถือแบบสมาร์ทโฟนที่เราใช้งานกันอยู่ในปัจจุบันก็มีระบบปฏิบัติการเป็นตัวขับเคลื่อนในการทำงานเช่นกัน ยกตัวอย่างเช่น ระบบปฏิบัติการ iOS ที่ใช้งานบน iPhone, iPad และ iPod Touch ที่พัฒนาโดยแอปเปิล, ระบบปฏิบัติการ Windows Phone ที่พัฒนาโดยไมโครซอฟท์ และระบบปฏิบัติการยออดิตอีกหนึ่งตัวที่จะแนะนำให้รู้จักนั่นก็คือ ระบบปฏิบัติการแอนดรอยด์ (Android) ที่พัฒนาโดยกูเกิล เรามาทำความรู้

ฉบับที่	12
ชื่อเรื่อง	รวม 15 ความสามารถใหม่บน iOS 8 ที่คุณควรรู้
เนื้อหา	

แอปเปิลเปิดตัว iOS 8 ในครั้งนี้ แม้จะไม่ได้เปลี่ยนดีไซน์ครั้งใหญ่เหมือนครั้ง iOS 7 แต่เป็นการปรับปรุงและพัฒนาการใช้งาน UX (User eXperience) ให้ดียิ่งขึ้น อ่านเพิ่ม : รวมทุกเรื่องต้องรู้และวิธีอัปเดต iOS 8 สำหรับ iPhone, iPad, iPod Touch ios-8 การรีวิวนี้จะเป็นการเล่าประสบการณ์การใช้งานผ่านมุมมองที่ใช้ iPhone แทบจะติดมือและทำหลายอย่างมากในชีวิตประจำวัน เพื่อให้เห็นการใช้งานจริงมากที่สุด โดยเวอร์ชันที่รีวิวจะเป็น iOS 8 Gold Master คือถ้าไม่มีปัญหาอะไร เวอร์ชันนี้จะเป็น iOS 8 ตัวเต็มนั่นเอง First Look-02 1. ดีไซน์ อย่างแรกที่ตั้งเกต

ฉบับที่	13
ชื่อเรื่อง	Android กับ iOS จุดเด่นและความแตกต่าง
เนื้อหา	

ถูกถามอยู่บ่อย ๆ ว่าจะเอาอะไรดีระหว่างมือถือซัมซุงกับไอโฟน มันต่างกันตรงไหน...มันต่างกันโครตเยอะเลยนะ มือถือระดับไฮเอนด์สเปกมันใกล้เคียงกันหมด แต่ที่แตกต่างคือ “ระบบ” ดังนั้นการเลือกมือถือไม่ใช่การเลือกสเปกอย่างเดียว แต่เป็นการเลือกระบบที่ใช้ ซึ่งถ้าถามว่าอันไหนดีกว่า บอกไว้ก่อนเลยว่า ซัมซุงที่เป็น Android กับไอโฟนที่เป็น iOS มันไม่ได้ดีไปกว่ากัน Android กับ iOS เหมือนอาหารไทยกับอาหารญี่ปุ่น ทั้งสองไม่ได้ร่อยไปกว่ากันหรอก แต่อยู่ที่ว่าใครจะชอบกินแบบไหน Android - iOS อีสาระ Android เหมือนคอมพิวเตอร์เลย ติดตั้งแอปจากไหนก็ได้ เชื่อมต่อกับอุปกรณ์

ฉบับที่	14
ชื่อเรื่อง	"ระบบปฏิบัติการบนมือถือชั้นนำของโลก"
เนื้อหา	

ความยอดเยี่ยมของ iPhone, iPad และ iPod touch มีพื้นฐานมาจาก iOS 9 ที่ประกอบด้วยอินเทอร์เฟซที่ใช้งานง่าย คุณสมบัติที่น่าทึ่งมากมาย รวมทั้ง ระบบความปลอดภัยซึ่งเป็นสิ่งที่สำคัญที่สุด นอกจากนี้ยังได้รับการออกแบบ ให้ดูสวยงามและทำงานได้อย่างราบรื่น โดยคุณ จะรู้สึกได้แม้กระทั่งจากการทำงาน ที่เรียบง่ายมาก ๆ และเนื่องจาก iOS 9 ได้รับการสร้างขึ้นเพื่อนำเทคโนโลยีสุดล้ำ ที่อยู่ในฮาร์ดแวร์ของ Apple มาใช้ประโยชน์ได้อย่างเต็มที่ อุปกรณ์ของคุณจึงล้ำหน้าไปไกลกว่าคนอื่นเสมอ อินเทอร์เฟซที่สวยงาม ใช้งานง่ายไม่ยุ่งยาก ตั้งแต่วินาทีที่หยิบขึ้นมา คุณก็แทบจะสามารถใช้

ฉบับที่ 15
ชื่อเรื่อง ระบบปฏิบัติการ MAC

เนื้อหา

เป็นระบบปฏิบัติการของเครื่องแมคอินทอช เป็นผลิตภัณฑ์แรกที่ประสบความสำเร็จเกี่ยวกับการทำงานแบบ GUI ในปี ค.ศ. 1978 ของบริษัท Apple ต่อมาได้มีการเปลี่ยนชื่อเป็นระบบปฏิบัติการ Mac OS โดยเวอร์ชันล่าสุดมีชื่อเรียกว่า Mac OS X เหมาะสมกับคอมพิวเตอร์ที่ผลิตโดยบริษัท Apple และมีความสามารถในการทำงานหลายโปรแกรมพร้อมกัน (Multitasking) เหมาะกับงานในด้านเดสก์ทอปพับลิชชิ่ง (Desktop Publishing) ตัวอย่างหน้าต่างระบบปฏิบัติการ MAC Mac รุ่นของ MAC OS X Mac OS X 10.0 Cheetah (รุ่นแรก) Mac-OS-X-10-0-Cheetah ในปี 2001 ก็ได้ปรากฏ Mac OS X เวอร์ชันเต็มตัว

ฉบับที่ 16
ชื่อเรื่อง iOS คืออะไร และ iOS Device คืออะไร

เนื้อหา

iOS คือ ระบบปฏิบัติการ ที่เป็นของบริษัท Apple ในตอนแรกถูกพัฒนามาเพื่อ iPhone เปิดตัวครั้งแรก 29 มิถุนายน 2550 และมีการพัฒนาต่อไปที่ iPod Touch และ iPad ระบบปฏิบัติการนี้ สามารถเชื่อมต่อกับ itunes store เพื่อสามารถดาวน์โหลด Application ได้ ในเวอร์ชัน ios 2.0 เป็นครั้งแรก ในตอนนี้เป็นที่ฮือฮามากสำหรับระบบปฏิบัติการนี้ จนตอนนี้มีทั้งหมด ถึง ios 7 แล้ว Z9dCA7D [sws_divider_basic] iOS Device คือ อุปกรณ์ที่สามารถใช้ระบบปฏิบัติการ iOS ได้ นั่นก็คือ iPhone, iPod, iPad โดยเราสามารถเรียก อุปกรณ์ทั้งหมดโดยรวมที่เรียกว่า iOS Device เช่น ลำโพงตัวนี้

ฉบับที่	17
ชื่อเรื่อง	"IOS คืออะไร" สั้น ๆ ง่าย ๆ คนไม่เคยใช้ แนะนำให้อ่าน
เนื้อหา	

IOS หรือ ที่เรียกกันว่า ผลไม้...แอปเปิ้ล ... IOS คืออะไร? : ไอโอเอส (ก่อนหน้านี้ใช้ชื่อ ไอโฟน ไอเอส) คือระบบปฏิบัติการบนอุปกรณ์พกพา (สมาร์ทโฟน, แท็บเล็ตคอมพิวเตอร์) พัฒนาและจำหน่ายโดยแอปเปิล (บริษัท) เปิดตัวครั้งแรกในปี 2007 เพื่อใช้บนไอโฟน และได้มีการพัฒนาเพิ่มเติมเพื่อใช้บนอุปกรณ์พกพาอื่น ๆ ของแอปเปิล เช่น ไอพอดทัช (ในเดือนกันยายน 2007), ไอแพด (ในเดือนมกราคม 2010), ไอแพด มินิ (พฤศจิกายน 2012) และ แอปเปิลทีวี รุ่นที่ 2 (ในเดือนกันยายน 2010) ไอโอเอสแตกต่างจากวินโดวส์โฟนของไมโครซอฟท์และแอนดรอยด์ (ระบบปฏิบัติการ) ของกูเกิล ตรงที่แอปเปิลไม่อนุ

ฉบับที่	18
ชื่อเรื่อง	iOS คืออะไร
เนื้อหา	

ไอโอเอส (iOS) หรือในชื่อเดิมคือ ไอโฟน ไอเอส (iPhone OS) เป็นระบบปฏิบัติการสำหรับสมาร์ทโฟนของบริษัทแอปเปิล (Apple Inc.) หรือในชื่อเดิม แอปเปิลคอมพิวเตอร์ (Apple Computer Inc.) โดยเริ่มต้นพัฒนาสำหรับใช้ในโทรศัพท์ไอโฟน (iPhone) เปิดตัวครั้งแรกเมื่อ 29 มิถุนายน พ.ศ. 2550 และได้พัฒนาต่อในเวอร์ชัน iOS 2.x ใช้สำหรับ ไอพอดทัช (iPod Touch) และ ไอแพด (iPad) เปิดตัวเมื่อ 11 กรกฎาคม พ.ศ. 2551 โดยระบบปฏิบัติการนี้สามารถเชื่อมต่อไปยังแอปสโตร์ (App Store) สำหรับเข้าถึงแอปพลิเคชัน (Application) มากมาย โดยใช้งานได้ ในเวอร์ชัน iOS 2.x เป็นครั้งแรกอีกด้วย

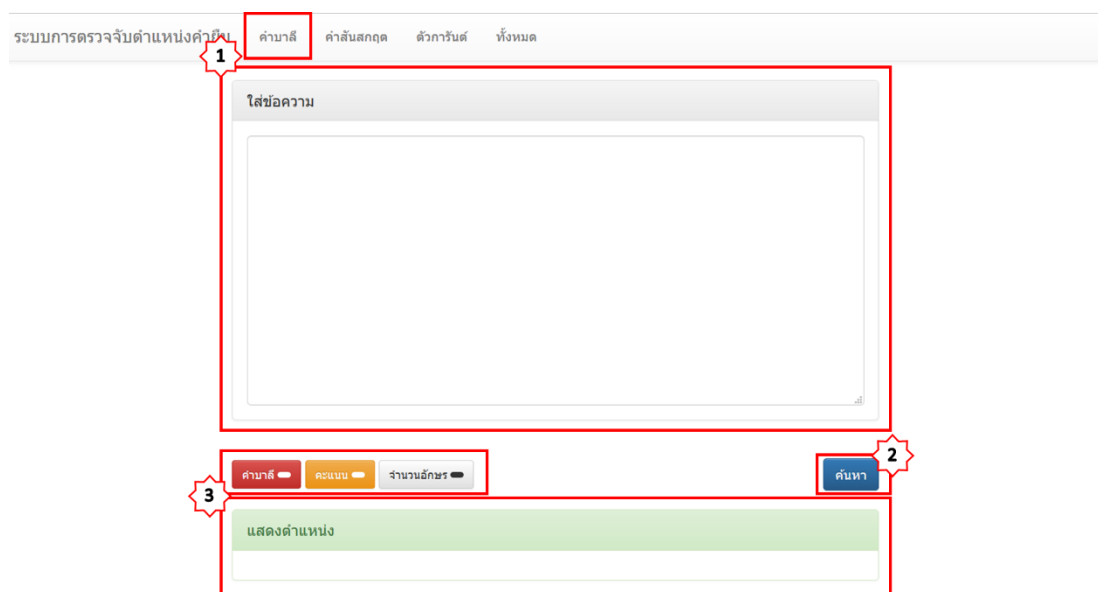
ฉบับที่	19
ชื่อเรื่อง	iOS ย่อมาจากอะไร หมายถึง...
เนื้อหา	

iOS ย่อมาจาก iPhone Operating System iOS หมายถึง ระบบปฏิบัติการบนอุปกรณ์พกพาที่ถูกพัฒนาเพื่อใช้สำหรับโทรศัพท์สมาร์ทโฟนของบริษัท แอปเปิล ซึ่งเริ่มจากการพัฒนาเพื่อใช้สำหรับโทรศัพท์ iPhone จากนั้นได้นำไปใช้กับอุปกรณ์พกพาอื่นๆของ แอปเปิล เช่น iPad, iPod โดย iOS มีความพิเศษคือจะมีระบบปฏิบัติการนี้บนอุปกรณ์ของ แอปเปิลเท่านั้น iOS นั้นจะสามารถเข้าถึง App Store ซึ่งผู้ใช้งานสามารถทำการ Download โปรแกรมหรือ Application ที่ต้องการมาใช้งานได้ผ่านทางระบบ Internet รวมถึงระบบการถ่ายโอนข้อมูลระหว่างอุปกรณ์กับ Server ของบริษัท แอปเปิล โดย iOS นั้นมีต้น

ภาคผนวก ค

แนะนำการใช้งานระบบการตรวจจับตำแหน่งคำยืม

ในส่วนนี้จะแนะนำการใช้งานอย่างละเอียดของระบบการตรวจจับตำแหน่งคำยืม ซึ่งเป็นระบบสำหรับตรวจจับตำแหน่งของคำ บาลี สันสกฤต และตัวการ์นต์ โดยสามารถอธิบายการใช้งานส่วนต่อประสานกับผู้ใช้ของระบบในแต่ละหน้าได้ดังนี้



ภาพที่ ค.1 ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำบาลี

จากภาพที่ ค.1 คือ ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำบาลี โดยจะแบ่งเป็น 3 ส่วนหลัก คือ 1) ส่วนนำเข้า คือ ผู้ใช้ต้องใส่ข้อความภาษาไทยที่ผ่านการตัดคำ 2) ส่วนการกระทำ คือ เมื่อผู้ใช้งานดำเนินการเสร็จสิ้นในส่วนที่ 1 ผู้ใช้จะต้องกดปุ่ม “ค้นหา” เพื่อให้ระบบค้นหาตำแหน่งคำบาลีจากข้อความที่ผู้ใช้ใส่ในตอนต้น 3) ส่วนการแสดงผล เมื่อระบบดำเนินการประมวลผลเสร็จสิ้นในส่วนที่ 2 ระบบจะแสดงผลลัพธ์การค้นหาตำแหน่ง คือ จำนวนตำแหน่งคำ

บาลีที่ตรวจพบ จำนวนอักขระของข้อความทั้งหมด และสุดท้ายแสดงตำแหน่งคำบาลีที่ตรวจพบในข้อความในข้างต้น ดังที่ได้แสดงตัวอย่างการดำเนินการค้นหาตำแหน่งคำบาลีในภาพที่ ค.2

ใส่ข้อความ

ไอโอเอส |(ก่อนหน้า|นี้|ใช่|ชื่อ| |ไอโฟน|ไอเอส)| |คือ|ระบบ|ปฏิบัติ|การ|บน|อุปกรณ์|พกพา| |(สมาร์ตโฟน|,|แท็บเล็ต|คอมพิวเตอร์)| |พัฒนา|และ|จำหน่าย|โดย|แอปเปิล| |(บริษัท)| |เปิดตัว|ครั้งแรก|ในปี| |2007| |เพื่อ|ใช้|บน|ไอโฟน| |และ|ได้|มี|การ|พัฒนา|เพิ่มเติม|เพื่อ|ใช้|บน|อุปกรณ์|พกพา|อื่นๆ| |ของ|แอปเปิล| |เช่น| |ไอพอด|ทัช| |(ใน|เดือน|กันยายน| |2007)|,| |ไอแพด| |(ใน|เดือน|มกราคม| |2010)|,| |ไอแพด| |มินิ| |(พฤศจิกายน| |2012)| |และ| |แอปเปิล|ทีวี| รุ่น|ที่| |2| |(ใน|เดือน|กันยายน| |2010)| |ไอโอเอส|แตกต่าง|จาก|วินโดวส์|โฟน|ของ|ไมโครซอฟท์|และ|แอนดรอยด์| |(ระบบ|ปฏิบัติการ)|ของ|กูเกิล| |ตรง|ที่|แอปเปิล|ไม่|อนุญาต|ให้|นำ|ไอโอเอส|ไป|ติดตั้ง|บน|อุปกรณ์|ที่|ไม่|ใช่|อุปกรณ์|ของ|แอปเปิล| |ใน

คำบาลี 2

คะแนน 2

จำนวนอักขระ 704

ค้นหา

แสดงตำแหน่ง

ไอโอเอส |(ก่อนหน้า|นี้|ใช่|ชื่อ| |ไอโฟน|ไอเอส)| |คือ|ระบบ|ปฏิบัติ|การ|บน|อุปกรณ์|พกพา| |(สมาร์ตโฟน|,|แท็บเล็ต|คอมพิวเตอร์)| |พัฒนา|และ|จำหน่าย|โดย|แอปเปิล| |(บริษัท)| |เปิดตัว|ครั้งแรก|ในปี| |2007| |เพื่อ|ใช้|บน|ไอโฟน| |และ|ได้|มี|การ|พัฒนา|เพิ่มเติม|เพื่อ|ใช้|บน|อุปกรณ์|พกพา|อื่นๆ| |ของ|แอปเปิล| |เช่น| |ไอพอด|ทัช| |(ใน|เดือน|กันยายน| |2007)|,| |ไอแพด| |(ใน|เดือน|มกราคม| |2010)|,| |ไอแพด| |มินิ| |(พฤศจิกายน| |2012)| |และ| |แอปเปิล|ทีวี| รุ่น|ที่| |2| |(ใน|เดือน|กันยายน| |2010)| |ไอโอเอส|แตกต่าง|จาก|วินโดวส์|โฟน|ของ|ไมโครซอฟท์|และ|แอนดรอยด์| |(ระบบ|ปฏิบัติการ)|ของ|กูเกิล| |ตรง|ที่|แอปเปิล|ไม่|อนุญาต|ให้|นำ|ไอโอเอส|ไป|ติดตั้ง|บน|อุปกรณ์|ที่|ไม่|ใช่|อุปกรณ์|ของ|แอปเปิล| |ใน

ภาพที่ ค.2 ตัวอย่างการค้นหาตำแหน่งคำบาลี

ระบบการตรวจจับตำแหน่งคำยืม คำบาลี คำสันสกฤต ตัวการ์นต์ ทั้งหมด

ใส่ข้อความ

คำสันสกฤต
คะแนน
จำนวนอักษร
ค้นหา

แสดงตำแหน่ง

ภาพที่ ค.3 ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำสันสกฤต

จากภาพที่ ค.3 คือ ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำสันสกฤต ซึ่งแบ่งออกเป็น 3 ส่วนเช่นเดียวกับการค้นหาตำแหน่งคำบาลี สามารถดูได้ที่ภาพ ค.1 และสามารถดูตัวอย่างการดำเนินการค้นหาตำแหน่งคำสันสกฤตได้ที่ภาพ ค.4

ใส่ข้อความ

ไอโอเอส ((ก่อนหน้านี้ใช้ชื่อ ไอโฟนไอเอส)) คือระบบปฏิบัติการบนอุปกรณ์พกพา ((สมาร์ตโฟน,แท็บเล็ตคอมพิวเตอร์)) พัฒนาและจำหน่ายโดยแอปเปิล ((บริษัท)) เปิดตัวครั้งแรกในปี 2007 เพื่อใช้บนไอโฟน และได้มีการพัฒนาเพิ่มเติมเพื่อใช้บนอุปกรณ์พกพาอื่นๆ ของแอปเปิล เช่น ไอพอดทัช ((ในเดือนกันยายน 2007)), ไอแพด ((ในเดือนมกราคม 2010)), ไอแพด มินิ ((พฤศจิกายน 2012)) และ แอปเปิลทีวี รุ่นที่ 2 ((ในเดือนกันยายน 2010)) ไอโอเอสแตกต่างจากรินโดสโฟนของไมโครซอฟท์และแอนดรอยด์ ((ระบบปฏิบัติการ)) ของกูเกิล ตรงที่แอปเปิลไม่อนุญาตให้นำไอโอเอสไปติดตั้งบนอุปกรณ์ที่ไม่ใช่อุปกรณ์ของแอปเปิล ใน

คำสั้นสกฤต 13
คะแนน 13
จำนวนอักษร 704
ค้นหา

แสดงตำแหน่ง

ไอโอเอส ((ก่อนหน้านี้ใช้ชื่อ ไอโฟนไอเอส)) คือระบบปฏิบัติการบนอุปกรณ์พกพา ((สมาร์ตโฟน,แท็บเล็ตคอมพิวเตอร์)) พัฒนาและจำหน่ายโดยแอปเปิล ((บริษัท)) เปิดตัวครั้งแรกในปี 2007 เพื่อใช้บนไอโฟน และได้มีการพัฒนาเพิ่มเติมเพื่อใช้บนอุปกรณ์พกพาอื่นๆ ของแอปเปิล เช่น ไอพอดทัช ((ในเดือนกันยายน 2007)), ไอแพด ((ในเดือนมกราคม 2010)), ไอแพด มินิ ((พฤศจิกายน 2012)) และ แอปเปิลทีวี รุ่นที่ 2 ((ในเดือนกันยายน 2010)) ไอโอเอสแตกต่างจากรินโดสโฟนของไมโครซอฟท์และแอนดรอยด์ ((ระบบปฏิบัติการ)) ของกูเกิล ตรงที่แอปเปิลไม่อนุญาตให้นำไอโอเอสไปติดตั้งบนอุปกรณ์ที่ไม่ใช่อุปกรณ์ของแอปเปิล ใน

ภาพที่ ค.4 ตัวอย่างการค้นหาตำแหน่งคำสั้นสกฤต

ระบบการตรวจจับตำแหน่งคำยืม คำบาลี คำสั้นสกฤต ตัวการ์นต์ ทั้งหมด

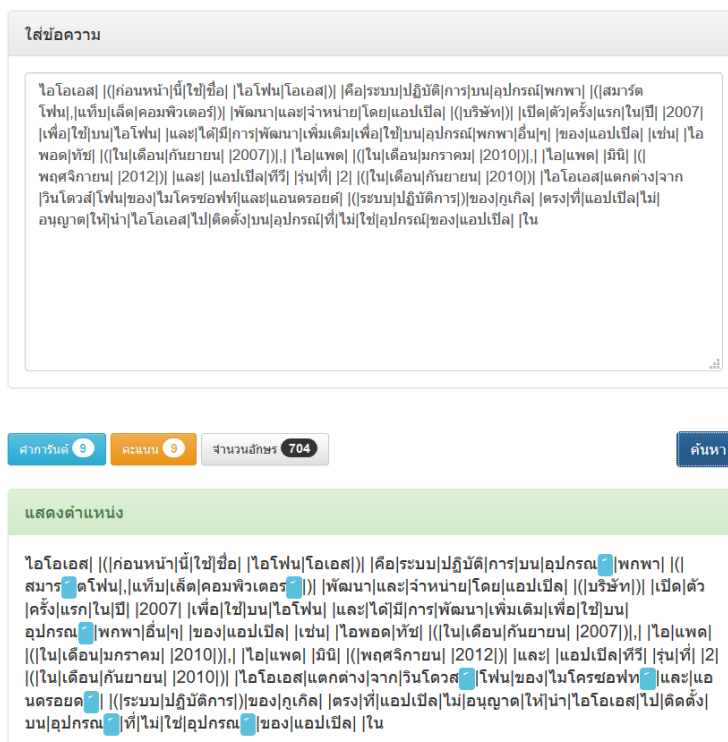
ใส่ข้อความ

ตัวการ์นต์
คะแนน
จำนวนอักษร
ค้นหา

แสดงตำแหน่ง

ภาพที่ ค.5 ส่วนต่อประสานกับผู้ใช้สำหรับค้นตำแหน่งตัวการ์นต์

จากภาพที่ ค.5 คือ ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งตัวการ์นต์ ซึ่งแบ่งออกเป็น 3 ส่วนเช่นเดียวกับการค้นหาตำแหน่งคำบาลี สามารถดูได้ที่ภาพ ค.1 และสามารถดูตัวอย่างการดำเนินการค้นหาตำแหน่งคำสันสกฤตได้ที่ภาพ ค.6



ภาพที่ ค.6 ตัวอย่างการค้นหาตำแหน่งตัวการ์นต์

ระบบการตรวจจับตำแหน่งคำยืม คำบาลี คำสันสกฤต ตัวการ์นต์ ทั้งหมด

ใส่ข้อความ

คำบาลี
คำสันสกฤต
คำราชาศัพท์
ตัวการ์นต์
คะแนน
จำนวนอักษร

ค้นหา

แสดงตำแหน่ง

ภาพที่ ค.7 ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำทั้งหมด

จากภาพที่ ค.7 คือ ส่วนต่อประสานกับผู้ใช้สำหรับค้นหาตำแหน่งคำทั้งหมด คือ คำบาลี สันสกฤต และตัวการ์นต์ ซึ่งแบ่งออกเป็น 3 ส่วนเช่นเดียวกับการค้นหาตำแหน่งคำบาลี สามารถดูได้ที่ภาพ ค.1 และสามารถดูตัวอย่างการดำเนินการค้นหาตำแหน่งคำสันสกฤตได้ที่ภาพ ค.8

ใส่ข้อความ

ไอโอเอส ((ก่อนหน้านี้|ใช่|ชื่อ |ไอโฟน|ไอเอส)) |คือระบบ|ปฏิบัติ|การ|บน|อุปกรณ์|พกพา |((|สมาร์ตโฟน|,|แท็บเล็ต|คอมพิวเตอร์|)) |พัฒนา|และ|จำหน่าย|โดย|แอปเปิล |((|บริษัท|)) |เปิดตัว|ครั้งแรก|ในปี |2007| |เพื่อ|ใช้|บน|ไอโฟน |และ|ได้|มี|การพัฒนา|เพิ่มเติม|เพื่อ|ใช้|บน|อุปกรณ์|พกพา|อื่นๆ |ของ|แอปเปิล |เช่น| |ไอพอด|ทัช |((|ในเดือน|กันยายน |2007|))|, |ไอแพด |((|ในเดือน|มกราคม |2010|))|, |ไอแพด |มินิ |((|พฤศจิกายน |2012|)) |และ| |แอปเปิล|ทีวี |รุ่น|ที่ |2| |((|ในเดือน|กันยายน |2010|)) |ไอโอเอส|แตกต่าง|จาก|ริโนโดวส์|โฟน|ของ|ไมโครซอฟท์|และ|แอนดรอยด์ |((|ระบบ|ปฏิบัติการ|))|ของ|กูเกิล |ตรง|ที่|แอปเปิล|ไม่|อนุญาต|ให้|นำ|ไอโอเอส|ไป|ติดตั้ง|บน|อุปกรณ์|ที่|ไม่|ใช่|อุปกรณ์|ของ|แอปเปิล |ใน

คำวลี 2
คำสั้นที่สุด 13
คำราชาศัพท์ 0
คำกรันต์ 9
คะแนน 24
จำนวนอักษร 704

จำนวนคำ 128
ค้นหา

แสดงตำแหน่ง

ไอโอเอส |((ก่อนหน้านี้|ใช่|ชื่อ |ไอโฟน|ไอเอส)) |คือระบบ|ปฏิบัติ|การ|บน|อุปกรณ์|พกพา |((|สมาร์ตโฟน|,|แท็บเล็ต|คอมพิวเตอร์|)) |พัฒนา|และ|จำหน่าย|โดย|แอปเปิล |((|บริษัท|)) |เปิดตัว|ครั้งแรก|ในปี |2007| |เพื่อ|ใช้|บน|ไอโฟน |และ|ได้|มี|การพัฒนา|เพิ่มเติม|เพื่อ|ใช้|บน|อุปกรณ์|พกพา|อื่นๆ |ของ|แอปเปิล |เช่น| |ไอพอด|ทัช |((|ในเดือน|กันยายน |2007|))|, |ไอแพด |((|ในเดือน|มกราคม |2010|))|, |ไอแพด |มินิ |((|พฤศจิกายน |2012|)) |และ| |แอปเปิล|ทีวี |รุ่น|ที่ |2| |((|ในเดือน|กันยายน |2010|)) |ไอโอเอส|แตกต่าง|จาก|ริโนโดวส์|โฟน|ของ|ไมโครซอฟท์|และ|แอนดรอยด์ |((|ระบบ|ปฏิบัติการ|))|ของ|กูเกิล |ตรง|ที่|แอปเปิล|ไม่|อนุญาต|ให้|นำ|ไอโอเอส|ไป|ติดตั้ง|บน|อุปกรณ์|ที่|ไม่|ใช่|อุปกรณ์|ของ|แอปเปิล |ใน

ภาพที่ ค.8 ตัวอย่างการค้นหาตำแหน่งคำทั้งหมด

ภาคผนวก ง

แนะนำการใช้งานระบบการประเมินผลการจัดระดับความยากง่ายของเอกสาร

ในส่วนนี้จะแนะนำการใช้งานอย่างละเอียดของระบบการประเมินผลการจัดระดับความยากง่ายของเอกสาร ซึ่งเป็นระบบสำหรับการประเมินผลความใกล้เคียงในการจัดกลุ่มระหว่างเทคนิคการจัดกลุ่มเอกสารที่ได้สร้างขึ้นทั้งหมด 3 เทคนิค และกลุ่มตัวอย่าง¹ โดยสามารถอธิบายการใช้งานส่วนต่อประสานกับผู้ใช้ของระบบในแต่ละหน้าได้ดังนี้

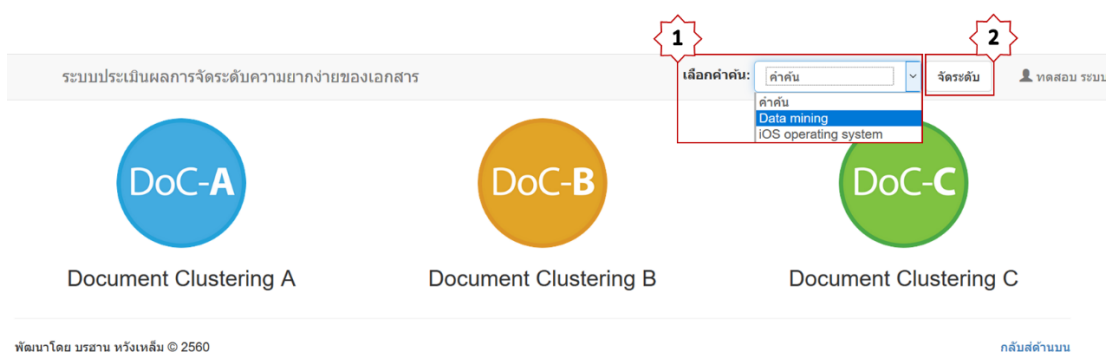
ภาพที่ ง.1 ส่วนต่อประสานกับผู้ใช้สำหรับเข้าสู่ระบบ

จากภาพที่ ง.1 คือ หน้าเริ่มต้นของระบบ โดยผู้ประเมินจะต้องกรอกอีเมลและรหัสผ่าน หากเป็นผู้ใช้ครั้งแรก ผู้ประเมินจะต้องสร้างบัญชีผู้ใช้ก่อนเพื่อสร้างรหัสผ่านเข้าสู่ระบบ โดยกดที่ลิงค์ข้อความ “สร้างบัญชีผู้ใช้”

¹ ผู้ประเมิน

ภาพที่ ง.2 ฟอรั่มสำหรับสร้างบัญชีผู้ใช้

จากภาพที่ ง.2 คือ แบบฟอร์มสำหรับกรอกข้อมูลในการสร้างบัญชีผู้ใช้งานระบบ โดยจะต้องพิมพ์ข้อมูลให้ครบถ้วนทั้งหมด จากนั้นกดปุ่ม “สร้าง” หากกดปุ่มดังกล่าวโดยที่กรอกข้อมูลไม่สมบูรณ์ จะมีข้อความแจ้งเตือนว่า “Please fill out this field.” เพื่อต้องการให้ผู้ประเมินพิมพ์ข้อมูลให้ครบถ้วนทั้งหมด



ภาพที่ ง.3 ส่วนต่อประสานหน้าแรกของระบบ

จากภาพที่ ง.3 คือ หน้าแรกของระบบ จะแสดงเมื่อเข้าสู่ระบบได้สำเร็จ ผู้ประเมินจะพบกับหน้าแรกของระบบ 1) ให้ผู้ประเมินเลือกคำค้น โดยจะมีสองคำค้น คือ Data mining และ iOS operating system หลังจากนั้น 2) ให้กดปุ่ม “จัดระดับ” หลังจากนั้นระบบจะแสดงผลการจัดระดับเอกสาร (จัดกลุ่มเอกสาร) ทั้ง 3 เทคนิค แสดงดังภาพที่ ง.4

ระบบประเมินผลการจัดระดับความยากของเอกสาร

เลือกคำค้น: คำค้น จัดระดับ ทดสอบ ระบบ

คำสำคัญคือ Data mining

DoC-A

Document Clustering A

ง่าย

- Data Warehouse & Data Mining
- Data Mining (เหมืองข้อมูล)
- เหมืองข้อมูล (data mining)

ปานกลาง

- Data Mining
- การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining โดยชอ...
- DATA Mining

ยาก

- Data Mining เมื่อต้องขุดหาอะไรบางอย่างจากเหมือง...
- Data Mining การทำเหมืองข้อมูล
- การทำเหมืองข้อมูล
- Data Mining(การทำเหมืองข้อมูล)และ โปรแกรม W...

DoC-B

Document Clustering B

ง่าย

ปานกลาง

- Data Warehouse & Data Mining
- Data Mining (เหมืองข้อมูล)
- เหมืองข้อมูล (data mining)
- Data Mining

ยาก

- การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining โดยชอ...
- DATA Mining
- Data Mining เมื่อต้องขุดหาอะไรบางอย่างจากเหมือง...
- Data Mining การทำเหมืองข้อมูล
- การทำเหมืองข้อมูล
- Data Mining(การทำเหมืองข้อมูล)และ โปรแกรม W...

DoC-C

Document Clustering C

ง่าย

- Data Warehouse & Data Mining
- Data Mining (เหมืองข้อมูล)
- เหมืองข้อมูล (data mining)

ปานกลาง

- Data Mining
- การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining โดยชอ...
- DATA Mining
- Data Mining เมื่อต้องขุดหาอะไรบางอย่างจากเหมือง...
- Data Mining การทำเหมืองข้อมูล

ยาก

- การทำเหมืองข้อมูล
- Data Mining(การทำเหมืองข้อมูล)และ โปรแกรม W...

พัฒนาโดย บุรฮาน หวังเหล็ม © 2560

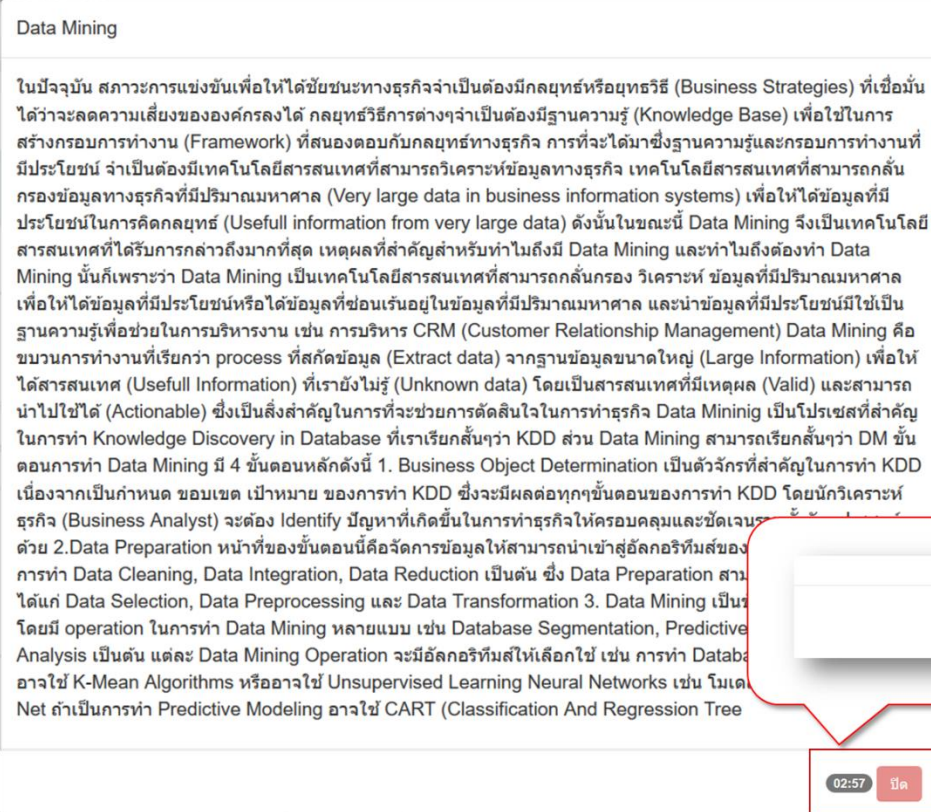
กลับสู่ด้านบน

ภาพที่ ง.4 ส่วนต่อประสานในหน้าแรกแสดงผลการจัดระดับ

ผู้ประเมินจะต้องเปิดเอกสารเพื่ออ่านเอกสารให้ครบทั้งหมด 10 ฉบับ เพื่อเข้าสู่ขั้นตอนการให้คะแนนความใกล้เคียง จากภาพที่ ง.4 แสดงให้เห็นถึง เอกสารในแต่ละเทคนิคที่ได้จัดระดับ ซึ่งจะแสดงเป็นชื่อเรื่องของเอกสารแต่ละฉบับที่เป็นลิงก์นำไปสู่ภาพที่ ง.5 นั้น เป็นเอกสารชุดเดียวกัน ดังนั้นผู้ประเมินสามารถเลือกอ่านเอกสารจากเทคนิคใดก็ได้ อย่างไรก็ตามเพื่อไม่ให้เกิดความสับสนระหว่างการอ่าน แนะนำให้ผู้ประเมินยึดการเปิดอ่านเอกสารจากเทคนิคใดเทคนิคหนึ่งและจดบันทึกชื่อเรื่องที่อ่านแล้วทั้งหมดไว้

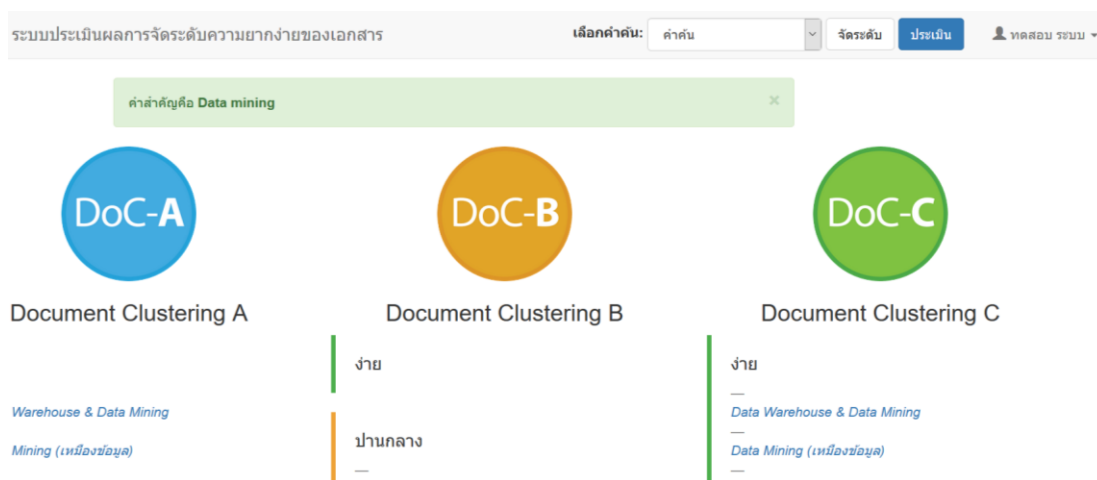
Data Mining

ในปัจจุบัน สถานะการแข่งขันเพื่อให้ได้ชัยชนะทางธุรกิจจำเป็นต้องมีกลยุทธ์หรือยุทธวิธี (Business Strategies) ที่เชื่อมโยงได้ว่าจะลดความเสี่ยงขององค์กรลงได้ กลยุทธ์วิธีการต่างๆจำเป็นต้องมีฐานความรู้ (Knowledge Base) เพื่อใช้ในการสร้างกรอบการทำงาน (Framework) ที่สนองตอบกับกลยุทธ์ทางธุรกิจ การที่จะได้มาซึ่งฐานความรู้และกรอบการทำงานที่มีประโยชน์ จำเป็นต้องมีเทคโนโลยีสารสนเทศที่สามารถวิเคราะห์ข้อมูลทางธุรกิจ เทคโนโลยีสารสนเทศที่สามารถกลั่นกรองข้อมูลทางธุรกิจที่มีปริมาณมหาศาล (Very large data in business information systems) เพื่อให้ได้ข้อมูลที่มีประโยชน์ในการคิดกลยุทธ์ (Usefull information from very large data) ดังนั้นในขณะนี้ Data Mining จึงเป็นเทคโนโลยีสารสนเทศที่ได้รับการกล่าวถึงมากที่สุด เหตุผลที่สำคัญสำหรับทำไมถึงมี Data Mining และทำไมถึงต้องทำ Data Mining นั้นก็เพราะว่า Data Mining เป็นเทคโนโลยีสารสนเทศที่สามารถกลั่นกรอง วิเคราะห์ ข้อมูลที่มีปริมาณมหาศาล เพื่อให้ได้ข้อมูลที่มีประโยชน์หรือได้ข้อมูลที่ซ่อนเร้นอยู่ในข้อมูลที่มีปริมาณมหาศาล และนำข้อมูลที่มีประโยชน์มีใช้เป็นฐานความรู้เพื่อช่วยในการบริหารงาน เช่น การบริหาร CRM (Customer Relationship Management) Data Mining คือ ขบวนการทำงานที่เรียกว่า process ที่สกัดข้อมูล (Extract data) จากฐานข้อมูลขนาดใหญ่ (Large Information) เพื่อให้ได้สารสนเทศ (Usefull Information) ที่เรายังไม่รู้ (Unknown data) โดยเป็นสารสนเทศที่มีเหตุผล (Valid) และสามารถนำไปใช้ได้ (Actionable) ซึ่งเป็นสิ่งสำคัญในการที่จะช่วยการตัดสินใจในการทำธุรกิจ Data Mining เป็นโปรเซสที่สำคัญในการทำ Knowledge Discovery in Database ที่เราเรียกสั้นๆว่า KDD ส่วน Data Mining สามารถเรียกสั้นๆว่า DM ขั้นตอนการทำ Data Mining มี 4 ขั้นตอนหลักดังนี้ 1. Business Object Determination เป็นตัวจักรที่สำคัญในการทำ KDD เนื่องจากเป็นกำหนด ขอบเขต เป้าหมาย ของการทำ KDD ซึ่งจะส่งผลต่อทุกขั้นตอนของการทำ KDD โดยนักวิเคราะห์ธุรกิจ (Business Analyst) จะต้อง Identify ปัญหาที่เกิดขึ้นในการทำธุรกิจให้ครอบคลุมและชัดเจนขึ้นด้วย 2.Data Preparation หน้าที่ของขั้นตอนนี้คือจัดการข้อมูลให้สามารถนำไปเข้าสู่อัลกอริทึมของการทำ Data Cleaning, Data Integration, Data Reduction เป็นต้น ซึ่ง Data Preparation สามารถทำได้แก่ Data Selection, Data Preprocessing และ Data Transformation 3. Data Mining เป็นขั้นตอนที่มี operation ในการทำ Data Mining หลายแบบ เช่น Database Segmentation, Predictive Analysis เป็นต้น แต่ละ Data Mining Operation จะมีอัลกอริทึมให้เลือกใช้ เช่น การทำ Database อาจใช้ K-Mean Algorithms หรืออาจใช้ Unsupervised Learning Neural Networks เช่น โมเดล Net ถ้าเป็นการทำ Predictive Modeling อาจใช้ CART (Classification And Regression Tree



ภาพที่ ง.5 ส่วนต่อประสานกับผู้ใช้สำหรับแสดงเนื้อหาของเอกสารสำหรับผู้ประเมิน

จากภาพที่ ง.5 คือ ส่วนต่อประสานกับผู้ใช้สำหรับแสดงเนื้อหาของเอกสารสำหรับผู้ประเมิน ซึ่งผู้ประเมินจะต้องอ่านเนื้อหาผ่านในเอกสารแต่ละฉบับเป็นเวลาอย่างน้อย 3 นาที ซึ่งผู้ประเมินจะไม่สามารถกดปุ่ม “ปิด” ได้ในช่วงเวลาดังกล่าว เมื่อเวลาผ่านไป 3 นาที จะสังเกตเห็นว่าเวลาที่นับถอยหลังจะหายไปและระดับสีของปุ่ม “ปิด” จะเพิ่มขึ้น ซึ่งผู้ประเมินสามารถกดปุ่ม “ปิด” ได้ตามปกติ



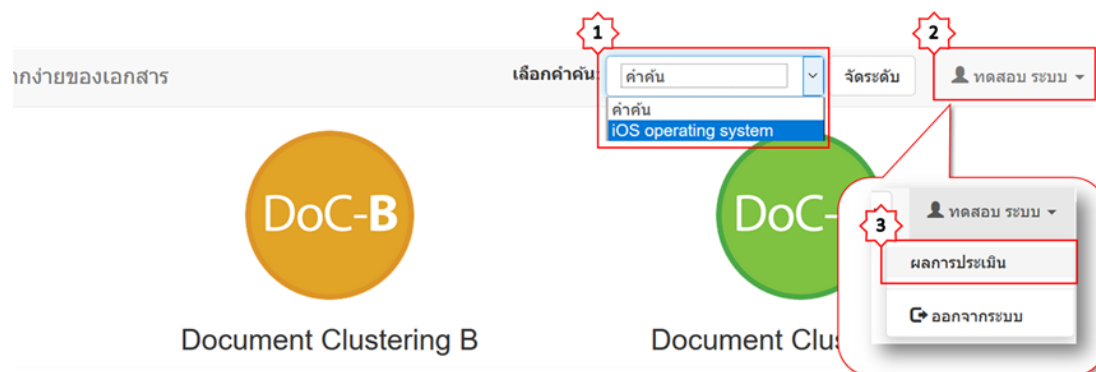
ภาพที่ ๖.6 ส่วนต่อประสานกับผู้ใช้หลังอ่านเอกสาร จำนวน 10 ฉบับ

จากภาพที่ ๖.6 แสดงให้เห็นส่วนต่อประสานกับผู้ใช้เมื่อผู้ประเมินอ่านเอกสารครบทุกฉบับ ปุ่ม “ประเมิน” จะปรากฏขึ้นอัตโนมัติ หากผู้ประเมินต้องการประเมินผลการจัดระดับ ต้องกดปุ่มดังกล่าว เพื่อเข้าสู่ส่วนต่อประสานกับผู้ใช้ แสดงดังภาพที่ ๖.7

ภาพที่ ๖.7 ฟอรั่มสำหรับประเมินระดับคะแนนความใกล้เคียง

จากภาพที่ ๖.7 คือ ฟอรั่มสำหรับประเมินระดับคะแนนความใกล้เคียง โดยผู้ประเมินจะต้องประเมินให้ระดับความใกล้เคียงในการจัดกลุ่ม แบ่งออกเป็น 3 ระดับ สำหรับแต่ละ

เทคนิค ซึ่งผู้ประเมินไม่สามารถเลือกระดับความใกล้เคียงซ้ำกัน ได้จากทั้ง 3 เทคนิค หากผู้ประเมินเลือกระดับดังกล่าวซ้ำกันใน 3 เทคนิคนี้ เมื่อผู้ประเมินกดปุ่ม “บันทึก” ระบบจะแสดงข้อความเพื่อแจ้งเตือนว่า “ไม่สามารถเลือกค่าที่ซ้ำกันได้” ดังนั้นผู้ประเมินจะต้องเลือกระดับใหม่อีกครั้ง โดยที่ไม่ให้ซ้ำกันจึงจะสามารถกดปุ่ม “บันทึก” แล้วระบบจะบันทึกข้อมูลการประเมินได้สำเร็จ



ภาพที่ ๖.8 ส่วนต่อประสานกับผู้ใช้หลังประเมินสำเร็จ

เมื่อผู้ประเมินผ่านขั้นตอนการประเมิน ระบบจะแสดงหน้าแรกของระบบ แสดงดังภาพที่ ๖.8 ซึ่งผู้ประเมินจะต้องเริ่มการประเมินในคำค้นถัดไป สังเกตว่าคำค้นที่ผู้ประเมินเลือกประเมินไปในตอนต้นจะหายไป เนื่องจากผู้ประเมินจะไม่สามารถประเมินในคำค้นที่ได้ประเมินก่อนหน้านี้ซ้ำได้ เว้นแต่ผู้ประเมินจะเลือกเมนูในกล่องที่ 2 และ 3 “ผลการประเมิน” ซึ่งสามารถดูผลประเมินที่ผ่านของผู้ประเมิน และสามารถเลือกแก้ไขผลการประเมิน

ระบบประเมินผลการจัดซื้อความยากง่ายของเอกสาร

เลือกคำค้น: คำค้น จัดระดับ ทดสอบ ระบบ

1

Data mining

#	เทคนิค	ผลการประเมิน
1	DoC-A	☆☆☆
2	DoC-B	☆☆☆
3	DoC-C	☆☆☆

Used Time : 0 day 2 hours 28 mins 3 secs

iOS operating system

#	เทคนิค	ผลการประเมิน
1	DoC-A	☆☆☆
2	DoC-B	☆☆☆
3	DoC-C	☆☆☆

Used Time : 0 day 0 hours 57 mins 6 secs

2 แก้ไขผลการประเมิน : Data mining iOS operating system

ภาพที่ ง.9 ส่วนต่อประสานกับผู้ใช้สำหรับแสดงผลการประเมิน

จากภาพที่ ง.9 คือ ส่วนต่อประสานกับผู้ใช้สำหรับแสดงผลการประเมิน ประกอบด้วย 2 ส่วน คือ 1) ส่วนแสดงผลการประเมินที่ผ่านมามากที่สุด รวมถึงระยะเวลาที่ใช้ในการประเมิน 2) ส่วนการแก้ไขผลการประเมิน โดยในส่วนนี้ผู้ประเมินจะต้องเลือกคำค้นที่ต้องการแก้ไขผลประเมิน จากนั้นผู้ประเมินต้องกดปุ่ม “ประเมิน” ระบบจะนำผู้ประเมินไปยังหน้าแรก ส่วนคำค้นที่เลือกเพื่อแก้ไขผลจะแสดงขึ้นอีกครั้ง เพื่อให้ผู้ประเมินสามารถเริ่มประเมินในคำค้นที่ต้องการแก้ไขผลการประเมินอีกครั้ง

ภาคผนวก จ

รายการคำยืมที่จำแนกโดยผู้เชี่ยวชาญ

รายการคำยืมที่จำแนกโดยผู้เชี่ยวชาญจากเอกสารหัวเรื่อง Data mining

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
1	องค์กร	1	O
	ประโยชน์	1	SO
	สามารถ	3	S
	วิเคราะห์	2	SO
	ประสิทธิภาพ	1	SP
3	ศาสตร์	1	SO
	ศึกษา	1	S
	สัมพันธ์	3	PO
	วิเคราะห์	1	SO
	ประโยชน์	1	SO
	ปัจจุบัน	2	P
	องค์กร	1	O
	ประยุกต์	2	SO
	สามารถ	2	S
4	สามารถ	2	S
	วิเคราะห์	6	SO
	กลยุทธ์	2	PO
	ออนไลน์	1	O
	สัมพันธ์	1	PO

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
6	ประโยชน์	1	SO
	วิเคราะห์	2	SO
	สัมพันธ์	1	PO
	อุปกรณ์	1	SO
8	ปัจจุบัน	1	P
	กลยุทธ์	4	PO
	ยุทธวิธี	1	P
	องค์กร	1	O
	ประโยชน์	4	SO
	สารสนเทศ	3	S
	เทคโนโลยีสารสนเทศ	1	S
	สามารถ	3	S
	วิเคราะห์	2	SO
	มหาสถา	3	S

รายการคำยืมที่จำแนกโดยผู้เชี่ยวชาญจากเอกสารหัวเรื่อง iOS operating system

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
2	กันยายน	1	S
	พัฒนาพีเจอร์	1	O
	ประสิทธิภาพ	1	PS
4	สามารถ	1	S
	วิเคราะห์	1	SO
	คอมพิวเตอร์	1	O
	แฟลชไดรฟ์	1	O

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
7	อุปกรณ์	4	SO
	สมาร์ทโฟน	1	O
	คอมพิวเตอร์	1	O
	บริษัท	1	S
	กันยายน	2	S
	มกราคม	1	S
	พฤศจิกายน	1	S
	วินโดวส์	1	O
	ไมโครซอฟท์	1	O
	แอนดรอยด์	1	O
8	ประสบการณ์	1	O
	วิทยุ	1	S
	ศิลปิน	1	S
	เวอร์ชัน	1	O
	ออฟไลน์	1	O
	สามารถ	1	S
	เพลย์ลิสต์	1	O
10	สมาร์ทโฟน	1	O
	ปัจจุบัน	1	P
	ตลาดสมาร์ทโฟน	1	O
	อุปกรณ์	1	SO
	บริษัท	1	S
	สามารถ	3	S
	วิเคราะห์	2	SO
	มหาศาล	3	S

ภาคผนวก จ

รายการคำยืมที่จำแนกโดยแบบจำลองตรวจจับรูปแบบคำยืม

รายการคำยืมที่จำแนกโดยแบบจำลองตรวจจับรูปแบบคำยืมจากเอกสารหัวเรื่อง

Data mining

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
1	ต้องการ	1	P
	องค์กร	1	SO
	ประโยชน์	1	SO
	สามารถ	3	S
	วิเคราะห์	2	SO
	มุมมอง	1	P
	ประสิทธิภาพ	1	SP
3	ศาสตร์	1	SO
	ศึกษา	1	S
	สัมพันธ์	3	PO
	วิเคราะห์	1	SO
	ประโยชน์	1	SO
	ปัจจุบัน	2	P
	องค์กร	1	PO
	ประยุกต์	2	SO
	สามารถ	2	S
	ถึงความ	1	P

*หมายเหตุ รายการคำยืมนี้เป็นผลจากการตรวจจับจากแบบจำลอง ส่งผลให้มีบางคำที่ไม่ใช่คำยืม

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
4	สามารถ	2	S
	ประมวล	1	S
	วิเคราะห์	6	SO
	กลยุทธ์	2	PO
	ออนไลน์	1	O
	สัมพันธ	1	PO
	มุมมอง	1	P
6	ประโยชน์	1	SO
	วิเคราะห์	2	SO
	มุมมอง	3	P
	ขั้นตอน	2	P
	สัมพันธ	1	PO
	อุปกรณ์	1	SO
	ป้องกัน	1	P
8	ปัจจุบัน	1	P
	แข่งขัน	1	P
	กลยุทธ์	4	PO
	ยุทธวิธี	1	P
	องค์กร	1	PO
	ประโยชน์	4	SO
	สารสนเทศ	3	S
	สามารถ	3	S
	วิเคราะห์	2	SO
	เทคโนโลยีสารสนเทศ	1	S
	กลั่นกรอง	2	S
	มหาศาล	3	S

*หมายเหตุ รายการคำยืมนี้เป็นผลจากการตรวจจับจากแบบจำลอง ส่งผลให้มีบางคำที่ไม่ใช่คำยืม

รายการคำขี้มที่จำแนกโดยแบบจำลองตรวจจับรูปแบบคำขี้มจากเอกสารหัวเรื่อง
iOS operating system

ฉบับที่	ตำแหน่งรูปแบบคำขี้ม	จำนวนคำขี้มในเอกสาร	ประเภทคำขี้ม
2	ทางการ	1	P
	กันยายน	1	S
	พัฒนาพีเจอร์	1	O
	ประสิทธิภาพ	1	SP
4	สามารถ	1	S
	วิเคราะห์	1	SO
	ต้องการ	1	P
	คอมพิวเตอร์	1	PO
	แฟลชไดร์ฟ	1	O
7	อุปกรณ์	4	SO
	สมาร์ตโฟน	1	O
	คอมพิวเตอร์	1	PO
	บริษัท	1	S
	กันยายน	2	S
	มกราคม	1	S
	พฤศจิกายน	1	S
	วินโดวส์	1	O
	ไมโครซอฟท์	1	O
	แอนดรอยด์	1	O
8	ประสบการณ์	1	SO
	วิทยุ	1	S
	ศิลปิน	1	S
	เวอร์ชัน	1	O

* หมายเหตุ รายการคำขี้มนี้เป็นผลจากการตรวจจับจากแบบจำลอง ส่งผลให้มีบางคำที่ไม่ใช่คำขี้ม

ฉบับที่	ตำแหน่งรูปแบบคำยืม	จำนวนคำยืมในเอกสาร	ประเภทคำยืม
8	ออฟไลน์	1	O
	สามารถ	1	S
	เพลย์ลิสต์	1	O
	เลือกสรร	1	S
10	สมาร์ทโฟน	1	O
	ปัจจุบัน	1	P
	ตลาดสมาร์ทโฟน	1	O
	อุปกรณ์	1	SO
	บริษัท	1	S

*หมายเหตุ รายการคำยืมนี้เป็นผลจากการตรวจจับจากแบบจำลอง ส่งผลให้มีบางคำที่ไม่ใช่คำยืม

ประวัติผู้เขียน

ชื่อ สกุล นายบูรฮาน หวังเหล็ม

รหัสประจำตัวนักศึกษา 5830223004

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิทยาศาสตรบัณฑิต (เทคโนโลยีสารสนเทศและ การสื่อสารเพื่อการจัดการ)	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี	2557

การตีพิมพ์เผยแพร่ผลงาน

Wanglem, B., and Tongtep, N. (2017). "Pattern-Sensitive Loanword Estimation for Thai Text Clustering". *Walailak Journal of Science and Technology*, 14(10), 813-823.

บูรฮาน หวังเหล็ม และ ฉัฐพงศ์ ทองเทพ. (2560). "การประเมินระดับความยากง่ายของข้อความด้วยวิธีการสกัดคำยืมจากข้อความภาษาไทย.", *เอกสารการประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 13 ประจำปี 2560*, โรงแรมอโนมาแกรนด์ กรุงเทพฯ: 6-7 กรกฎาคม 2560.