

การนำการปรากฏร่วมกันของเทอมมาประยุกต์ใช้ในการขยายข้อสอบถามแบบอัตโนมัติด้วยการ
จัดกลุ่มแบบลำดับชั้นในการค้นคืนสารสนเทศ



นายพีรภัทร กุลประดิษฐ์

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาการพัฒนาซอฟต์แวร์ด้านธุรกิจ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2556


ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR) are the thesis authors' files submitted through the University Graduate School.

Applying Term Co-occurrence in Automatic Query Expansion with Hierarchical
Clustering in Information Retrieval



Mr. Pheeraphat Kulpradit

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Business Software Development

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การนำการปรากฏร่วมกันของเทอมาประยุกต์ในการการ
ขยายข้อสอบถามแบบอัตโนมัติด้วยการจัดกลุ่มแบบลำดับ
ชั้นในการค้นคืนสารสนเทศ

โดย

นายพีรภัทร กุลประดิษฐ์

สาขาวิชา

การพัฒนาซอฟต์แวร์ด้านธุรกิจ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร. อัมภาพร ทรัพย์สมบูรณ์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ถาวร อานุกาฬไตรรงค์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร. อัมภาพร ทรัพย์สมบูรณ์)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. จันทร์เจ้า มงคลนาวิน)

..... กรรมการภายนอกมหาวิทยาลัย

(รองศาสตราจารย์ ดร. สมพร พุทธาพิทักษ์ผล)



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5381853026 : MAJOR BUSINESS SOFTWARE DEVELOPMENT

KEYWORDS: THAI INFORMATION RETRIEVAL / AUTOMATIC QUERY EXPANSION /
TERM CO-OCCURRENCE

PHEERAPHAT KULPRADIT: APPLYING TERM CO-OCCURRENCE IN
AUTOMATIC QUERY EXPANSION WITH HIERARCHICAL CLUSTERING IN
INFORMATION RETRIEVAL. ADVISOR: ASSOC. PROF. ASSADAPORN
SAPSOMBOON, Ph.D., pp.

This thesis presents the application of query expansion using term clustering from hierarchical clustering, with information retrieval system for Thai language. The comparison of retrieval performance of before and after applying query expansion, with precision at 10 documents, precision at 20 documents, precision at 30 documents, and F-measure. In addition, an information retrieval system and a query expansion system is implemented using term clustering, with information retrieval toolkit, Lucene, and data mining toolkit, Weka

This research includes 100 abstracts of Senior projects in the Bachelor program in Information Technology for Business at the Faculty of Commerce and Accountancy, Department of Statistics in the academic year 2009-2011 25 4th year-Students in the academic year 2012 at the Faculty of Commerce and Accountancy, Department of Statistics provide test collection. A query and assess relevancy of each document.

With Retrieval performance : precision at 10 documents, precision at 20 documents, precision at 30 documents, and F-measure, it can be concluded that applying query expansion with term clustering leads to better retrieval performance.

Department: Statistics Student's Signature

Field of Study: Business Software Development
Advisor's Signature

Academic Year: 2013



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

กิตติกรรมประกาศ

วิทยานิพนธ์นี้จะสำเร็จลุล่วงและสมบูรณ์ไปได้ด้วยดีต้องขอกราบขอบพระคุณรองศาสตราจารย์ ดร. อัมภพร ทรัพย์สมบูรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างที่ได้ให้คำแนะนำและข้อคิดเห็นต่างๆ ตรวจสอบแก้วิทยานิพนธ์ฉบับนี้อย่างละเอียด ตลอดจนแนวทางในการวิจัยด้วยดีตลอดมา ขอขอบพระคุณอาจารย์ ผู้ช่วยศาสตราจารย์ ดร.ถาวร อานุกาฬไตรรงค์ และผู้ช่วยศาสตราจารย์ ดร. จันทร์เจ้า มงคลนาวิณ และผู้ช่วยศาสตราจารย์ ดร. สมพร พุทธาพิทักษ์ผล กรรมการวิทยานิพนธ์ที่กรุณาเสียสละเวลาให้คำแนะนำ เพื่อแก้ไขรูปแบบและเนื้อหาวิทยานิพนธ์ฉบับนี้จนเสร็จสมบูรณ์

ขอบคุณเพื่อนๆ ที่ให้คำปรึกษาและความช่วยเหลือในด้านต่างๆ ซึ่งทำให้งานวิจัยเป็นไปได้ อย่างราบรื่นตลอดจนกำลังใจที่มอบให้เสมอ

สุดท้ายนี้ ผู้วิจัยใคร่กราบขอบพระคุณบิดา มารดาและครอบครัวที่คอยช่วยเหลือให้การสนับสนุนและคอยเป็นแรงกระตุ้นให้แก่ผู้วิจัยเสมอจนสำเร็จการศึกษา



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

หน้า

บทคัดย่อภาษาไทย	จ
บทคัดย่อภาษาอังกฤษ	ช
กิตติกรรมประกาศ	ซ
สารบัญ	ฅ
สารบัญรูปภาพ	ฐ
สารบัญตาราง	ฒ
บทที่ 1 บทนำ	1
1.1. ที่มาและความสำคัญของปัญหา.....	1
1.2. วัตถุประสงค์ของงานวิจัย.....	19
1.3. ขอบเขตของการวิจัย.....	20
1.4. ตัวแปรสำคัญที่ศึกษา	20
1.5. ข้อจำกัดการวิจัย	21
1.6. ประโยชน์ที่คาดว่าจะได้รับ	22
บทที่ 2 การทบทวนวรรณกรรม	23
2.1. การค้นคืนสารสนเทศ (Information Retrieval).....	23
2.2. รูปแบบและขั้นตอนการค้นคืนสารสนเทศ	25
2.3. การขยายข้อสอบถาม (Query Expansion).....	31
2.3.1. การนำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศ	32
2.3.2. รูปแบบการขยายข้อสอบถาม.....	37
2.3.3. การวิเคราะห์ข้อมูลในการขยายข้อสอบถาม	41
2.3.4. การทำเหมืองข้อมูล (Data Mining)ในการขยายข้อสอบถาม.....	47
2.4. งานวิจัยในอดีตที่นำการจัดกลุ่มแบบลำดับชั้นประยุกต์กับการขยายข้อสอบถามในการค้นคืนสารสนเทศ.....	67
2.5. การประเมินประสิทธิผลการค้นคืน (Retrieval Effectiveness Evaluation).....	70
2.5.1. ค่าระลึกได้ (Recall) และค่าความแม่นยำ (Precision).....	70
2.5.2. การประยุกต์รำระลึกได้และค่าความแม่นยำ.....	71
2.6. การนำการประเมินประสิทธิผลการค้นคืนสารสนเทศ ในงานวิจัย.....	77

2.7.	ชุดทดสอบ (Test Collection) สำหรับการทดสอบประสิทธิผลการค้นคืนสารสนเทศ	79
	2.7.1. ลักษณะของชุดทดสอบ	79
	2.7.2. การสร้างชุดทดสอบ	80
2.7.	เครื่องมือการค้นคืนสารสนเทศลูซึน	85
2.8.	เครื่องมือการทำเหมืองข้อมูลเวก้า	86
บทที่ 3	ระเบียบวิธีวิจัย	87
3.1.	แนวทางการวิจัยและสมมติฐานงานวิจัย	87
3.2.	ประชากรการเลือกกลุ่มตัวอย่างและจำนวนตัวอย่าง	89
3.3.	ขั้นตอนการดำเนินงานวิจัย	91
	3.3.1. การพัฒนาระบบการขยายข้อสอบถามด้วยการจัดกลุ่มคำ	92
	3.3.2. การสร้างชุดทดสอบสำหรับภาษาไทย	106
	3.3.3. การประเมินประสิทธิภาพการค้นคืน	109
3.4.	เครื่องมือในการเก็บข้อมูล	111
3.5.	ความถูกต้อง (Validity) และความน่าเชื่อถือ (Reliability) ของข้อมูลที่เก็บ	114
3.6.	กรอบการวิเคราะห์ข้อมูล	115
บทที่ 4	ผลการวิเคราะห์ข้อมูล	117
4.1.	การเก็บข้อมูลสำหรับงานวิจัย	117
	4.1.1. การประเมินประสิทธิภาพการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้	118
	4.1.2. การประเมินประสิทธิภาพการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อ สอบถาม 119	
4.2.	ผลการวิเคราะห์ประสิทธิภาพการค้นคืนสารสนเทศ	122
	4.2.1. การวิเคราะห์ค่าความแม่นยำที่สิบ	124
	4.2.2. การวิเคราะห์ค่าความแม่นยำที่ยี่สิบ	128
	4.2.3. การวิเคราะห์ค่าความแม่นยำที่สามสิบ	132
	4.2.4. การวิเคราะห์ค่ามาตรวัดเอฟ	136
4.3.	สรุปผลการวิเคราะห์ข้อมูล	139
4.4.	สรุปการอภิปรายผล	140

บทที่ 5 สรุปผลการวิจัย	142
5.1. สรุปผลการวิจัย	142
5.2. การนำงานวิจัยมาประยุกต์ใช้	145
5.2.1. การนำงานวิจัยมาประยุกต์ใช้ในเชิงทฤษฎี	145
5.2.2. การนำงานวิจัยมาประยุกต์ใช้ในเชิงปฏิบัติ	145
5.3. ข้อจำกัดของงานวิจัย	146
5.4. แนวทางศึกษาต่อ	147
รายการอ้างอิง	149
ภาคผนวก ก ตัวอย่างเอกสารและรายการเอกสารที่นำมาใช้ในงานวิจัย	180
ภาคผนวก ข รายการคำหยุดภาษาไทย (Thai Stop-words List)	195
ภาคผนวก ค รายการคำที่ได้จากการวิเคราะห์เอกสาร	208
ภาคผนวก ง รายการกลุ่มคำที่ได้จากการวิเคราะห์กลุ่ม	216
ภาคผนวก จ ผลการวิเคราะห์เมทริกซ์แสดงคำที่ปรากฏในเอกสาร	225
ภาคผนวก ฉ รายการข้อสอบถามที่ได้จากการกำหนดจากผู้ประเมินความเกี่ยวข้อง	297
ประวัติผู้เขียนวิทยานิพนธ์	300



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญรูปลูกภาพ

รูปที่ 1.1 ขั้นตอนการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันระหว่างคำด้วยการจัดกลุ่มแบบลำดับขั้น	10
รูปที่ 1.2 ผลการสำรวจการเข้าชมหน้าเว็บเพจจำแนกตามจำนวนคำในข้อสอบถามที่ผู้ใช้กำหนด คัดลอกจาก A Trellian Company (2013)	15
รูปที่ 2.1 รูปแบบการสืบค้นข้อมูลสารสนเทศตัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551)	23
รูปที่ 2.2 แสดงขั้นตอนการประมวลผลการค้นคืนสารสนเทศแบบปริภูมิเวกเตอร์ ตัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551)	27
รูปที่ 2.3 รูปแบบการขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion)ตัดแปลงจาก Shapira, Taieb-Maimon และ Nemeth,2005; Kanaan และคณะ, 2007,2008; Ruthven, 2003	38
รูปที่ 2.4 รูปแบบการขยายข้อสอบถามแบบโต้ตอบ (Interactive Query Expansion)ตัดแปลงจาก Shapira และคณะ,2005; Kanaan และคณะ, 2007,2008; Ruthven, 2003	40
รูปที่ 2.5 ข้อมูลที่นำมาใช้ในการขยายข้อสอบถามตัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์(2551).....	42
รูปที่ 2.6 ตัวอย่างการจัดกลุ่มด้วยขั้นตอนวิธีอีจโกลเมราทีฟ ตัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551),Yin และคณะ (2011),Han และคณะ (2012).....	62
รูปที่ 2.7 ตัวอย่างกราฟค่าระลึกได้และค่าความแม่นยำ ตัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551) .	74
รูปที่ 2.8 ผลการสำรวจการเข้าชมหน้าเว็บเพจจำแนกตามจำนวนคำในข้อสอบถามที่ผู้ใช้กำหนด คัดลอกจาก A Trellian Company (2013)	83
รูปที่ 2.9 ตัวอย่างสารสนเทศที่นำมาใช้ในการสร้างชุดทดสอบ (MacDonald, 2006).....	84
รูปที่ 2.10 ตัวอย่างสารสนเทศที่นำมาใช้ในการสร้างชุดทดสอบ คัดลอกจาก (Ircing และคณะ, 2007)	84
รูปที่ 3.1 แสดงแนวทางการดำเนินงานวิจัย	92
รูปที่ 3.3 การวิเคราะห์รายการคำในเอกสาร	95
รูปที่ 3.4 แสดงการวิเคราะห์รายการดัชนีและเชื่อมโยงกับรายการเอกสาร	97
รูปที่ 3.5 แสดงการประเมินคุณภาพการจัดกลุ่ม จากขั้นตอนการจัดกลุ่มคำ	101
รูปที่ 3.6 ขั้นตอนวิธีการสร้างชุดทดสอบ (Test Collection).....	106

รูปที่ 3.7 ตัวอย่างแบบฟอร์มการประเมินความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร.....	108
รูปที่ 3.9 แสดงรูปแบบการทำงานของเครื่องมือระบบการค้นคืนสารสนเทศคลูซึน	112
รูปที่ 3.10 แสดงตัวอย่างข้อมูลก่อนการวิเคราะห์กลุ่มคำด้วยเครื่องมือการทำเหมืองข้อมูลเวก้า. 113	
รูปที่ 3.11 แสดงตัวอย่างผลลัพธ์การจัดกลุ่มคำที่ได้	114
รูปที่ 3.12 แสดงการกำหนดสถิติที่จะนำมาใช้ในการวิเคราะห์ข้อมูล	116
รูปที่ ก.1 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2552 (หน้าที่หนึ่ง).....	181
รูปที่ ก.2 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2552 (หน้าที่สอง).....	182
รูปที่ ก.3 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2553 (หน้าที่หนึ่ง).....	183
รูปที่ ก.4 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2553 (หน้าที่สอง).....	184
รูปที่ ก.5 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2554 (หน้าที่หนึ่ง).....	185
รูปที่ ก.6 ตัวอย่างเอกสารบทความวิจัยที่จัดทำในปี 2554 (หน้าที่สอง).....	186

สารบัญตาราง

ตารางที่ 2.1 แสดงตัวอย่างการทำเหมืองข้อมูล ดัดแปลงจาก Han และคณะ (2011)และ Tuffery (2011).....	49
ตารางที่ 2.2การนำการจำแนกประเภทมาใช้กับการขยายข้อสอบถามและตัวอย่างการนำไปใช้ ดัดแปลงจาก Han และคณะ (2011) และ Tuffery(2011).....	52
ตารางที่ 2.3 การนำการจัดกลุ่มมาใช้ในการขยายข้อสอบถามและตัวอย่างการนำไปใช้ดัดแปลงจาก Han และคณะ (2011) และ Tuffery (2011).....	55
ตารางที่ 2.4 แสดงรายการสมการการวัดค่าความเหมือนที่พบในงานวิจัยในอดีต.....	64
ตารางที่ 2.5 แสดงสมการการคำนวณของเกณฑ์การเชื่อมโยงที่พบในงานวิจัยในอดีต.....	66
ตารางที่ 2.6 ตัวอย่างระดับมาตรฐานของค่าระลีกได้.....	72
ตารางที่ 2.7 ตัวอย่างกรณีการคำนวณค่าระลีกได้และค่าความแม่นยำ.....	73
ตารางที่ 3.1 ตัวอย่างการวิเคราะห์ค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร.....	99
ตารางที่ 3.2 ตัวอย่างการวิเคราะห์เมทริกซ์แสดงการปรากฏร่วมกันของคำ.....	100
ตารางที่ 3.3 แสดงผลการวิเคราะห์คุณภาพการจัดกลุ่มที่ได้จากการกำหนดรูปแบบการจัดกลุ่มที่แตกต่างกัน 54 รูปแบบ เมื่อใช้ชุดรายการผลลัพธ์การค้นคืนชุดหนึ่ง.....	104
ตารางที่ 4.1 แสดงรายการเอกสารที่เกี่ยวข้องกับข้อสอบถาม Q1.....	117
ตารางที่ 4.2 ผลลัพธ์การค้นคืนของการใช้ข้อสอบถาม Q1 ในระบบการค้นคืนสารสนเทศและผลการประเมินผลลัพธ์การค้นคืน.....	118
ตารางที่ 4.3 แสดงผลการประเมินประสิทธิภาพการค้นคืน เมื่อใช้ข้อสอบถาม Q1.....	118
ตารางที่ 4.4 ผลลัพธ์การค้นคืนของการใช้ข้อสอบถาม Q2 ในระบบการค้นคืนสารสนเทศ รวมทั้งผลการประเมินผลลัพธ์การค้นคืน.....	120
ตารางที่ 4.5 แสดงผลการประเมินประสิทธิภาพการค้นคืน เมื่อใช้ข้อสอบถาม Q2.....	122
ตารางที่ 4.6 แสดงค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรฐานวัดเอฟ ที่คำนวณได้จากรายการผลลัพธ์การค้นคืน.....	123
ตารางที่ 4.7 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่สิบ.....	125
ตารางที่ 4.8 แสดงผลการวิเคราะห์ค่าความแม่นยำที่สิบที่ได้จากการวิเคราะห์.....	128

ตารางที่ 4.9 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วย ข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถามด้วยแบบทดสอบการจัดลำดับ เครื่องหมายวิลคอกซอน	128
ตารางที่ 4.10 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่สืบ	129
ตารางที่ 4.11 แสดงผลการวิเคราะห์ค่าความแม่นยำที่สืบที่ได้จากการวิเคราะห์	131
ตารางที่ 4.12 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศ ด้วยข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการจัดลำดับ เครื่องหมายวิลคอกซอน	132
ตารางที่ 4.13 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่สามสืบ	133
ตารางที่ 4.14 แสดงผลการวิเคราะห์ค่าความแม่นยำที่สามสืบที่ได้จากการวิเคราะห์	135
ตารางที่ 4.15 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่สามสืบที่ได้จากการค้นคืน สารสนเทศด้วยข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการ จัดลำดับเครื่องหมายวิลคอกซอน	135
ตารางที่ 4.16 แสดงค่าสถิติทดสอบการแจกแจงของค่ามาตรวัดเอฟ	136
ตารางที่ 4.17 แสดงผลการวิเคราะห์ค่ามาตรวัดเอฟที่ได้จากการวิเคราะห์	138
ตารางที่ 4.18 ผลการวิเคราะห์การเปรียบเทียบค่ามาตรวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อ สอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการจัดลำดับ เครื่องหมายวิลคอกซอน	139
ตารางที่ 4.19 สรุปผลการวิเคราะห์ค่าความแม่นยำที่สืบ ค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ สามสืบ และค่ามาตรวัดเอฟ	140
ตารางที่ ก.1 แสดงรายการเอกสารที่นำมาใช้ในงานวิจัย	187
ตารางที่ จ.1 ตารางแสดงค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร	225
ตารางที่ ฉ.1 แสดงรายการข้อสอบถามที่ได้จากการกำหนดจากผู้ประเมินความเกี่ยวข้อง	297

บทที่ 1 บทนำ

1.1. ที่มาและความสำคัญของปัญหา

การค้นคืนสารสนเทศ (Information retrieval) คือ การค้นหาสารสนเทศที่มีลักษณะที่ตรงกับข้อสอบถาม (Query) ที่ผู้ใช้ได้กำหนดไว้ เครื่องมือการค้นหาสารสนเทศหนึ่งที่ได้รับคามนิยมเป็นอย่างมากในปัจจุบัน คือ โปรแกรมค้นหา (Search Engine) ซึ่งนอกจากจะพบได้ตามเว็บไซต์ โปรแกรมค้นหาแล้ว ยังสามารถพบได้ตามเว็บไซต์ทั่วไป อาทิ เว็บไซต์ร้านหนังสือ ซึ่งนำมาใช้ในการค้นหาข้อมูลหนังสือที่ตนเองต้องการอยู่ในการค้นหาสารสนเทศนั้น เพื่อให้ผู้เข้าชมสามารถค้นหาสินค้าที่ตนเองต้องการได้ องค์ประกอบที่สำคัญของการค้นคืนสารสนเทศนั้นคือข้อสอบถาม (Query) (วาริน นาราวิทย์ และศิริภิญโญ จันทมณี, 2555) เนื่องจากข้อสอบถามในระบบการค้นคืนสารสนเทศเป็นสิ่งที่ผู้ใช้ได้ระบุเข้าไปใช้ในการค้นหาสารสนเทศเพื่อให้ได้รายการผลลัพธ์การค้นคืน (Retrieval Result List) ตามที่ผู้ใช้ต้องการอย่างไรก็ตาม การกำหนดข้อสอบถามจึงเป็นปัญหาหลักในการค้นคืนสารสนเทศ โดยมีสาเหตุที่พบเจอ ได้แก่ (1) ผู้ใช้ไม่สามารถสะกดคำได้อย่างถูกต้อง ตัวอย่างเช่นในงานวิจัยของรัตน ชุมแก้ว และสุพจน์ นิตยสุวัฒน์ (2549) (2) ปัญหาคำศัพท์ (Vocabulary Problem) มีคำพ้องความหมายหรือคำที่มีความหมายได้มากกว่าหนึ่งความหมาย (คำเหมือน, Synonym) หรือคำกำกวม (คำพ้องรูป, Polysemy) ซึ่งหมายถึงคำที่มีรูปแบบแตกต่างกันแต่มีความหมายเหมือนกัน (Carpineto และ Romano, Giovanni, 2012) และ (3) ข้อสอบถามที่ผู้ใช้กำหนดไม่สามารถนำมาค้นคืนได้ เนื่องจากไม่มีคำในข้อสอบถามอยู่ในเอกสารที่ต้องการค้นคืน หรือมีจำนวนมากหรือน้อยเกินไป จึงไม่สามารถค้นคืนเอกสารได้ตามที่ต้องการ (Sen, 2009; A Trellian Company, 2012)

เพื่อลดปัญหาดังที่กล่าวมาข้างต้น จึงมักนำการขยายข้อสอบถาม (Query Expansion) มาประกอบใช้ในโปรแกรมค้นหา การขยายข้อสอบถามเป็นการเพิ่มคำหรือทอมในข้อสอบถามของระบบการค้นคืน เพื่อให้ได้ผลลัพธ์เป็นข้อสอบถามใหม่สำหรับการนำมาใช้ในการค้นคืนเอกสารครั้งต่อไป วัตถุประสงค์ของการขยายข้อสอบถามคือเพื่อการค้นคืนเอกสารที่ได้ผลลัพธ์ที่ดีขึ้นและประสิทธิผลที่ดีมากขึ้นมีงานวิจัยจำนวนมากที่ศึกษาขั้นตอนวิธีการขยายข้อสอบถาม ซึ่งทำให้ประสิทธิผลการค้นคืนสารสนเทศเพิ่มขึ้น Ogilvie และ Callan (2001) ทดสอบการนำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศแบบกระจาย ซึ่งหมายถึงการค้นหา

สารสนเทศที่มากกว่าหนึ่งชุดข้อมูล ผลการทดลองพบว่า เมื่อนำการขยายข้อสอบถามมาประยุกต์ใช้ ประสิทธิภาพของระบบการค้นคืนเพิ่มขึ้น Klink และคณะ (2002) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศด้วยผลสะท้อนกลับความสัมพันธ์เทียม (Pseudo Relevance Feedback) กับการเรียนรู้เชิงคำ (Learning Term-based Concepts) พบว่า การขยายข้อสอบถามช่วยให้ประสิทธิภาพการค้นคืนสารสนเทศเพิ่มขึ้น โดยที่ชุดข้อมูลจำนวน 3 ชุด จากชุดข้อมูลทั้งหมด 4 ชุด มีประสิทธิภาพเพิ่มขึ้นจากการใช้การเรียนรู้เชิงคำ Daiz-Galino และคณะ (2009) ได้ทดสอบการนำการขยายข้อสอบถามประยุกต์ใช้ในระบบการค้นคืนภาพการแพทย์ (Medical Image Retrieval) โดยการนำยูเอ็มแอล (Unified Modeling Language, UML) และออนโทโลยี (Ontology) มาใช้ในการขยายข้อสอบถาม ผลการทดลองพบว่า การใช้ออนโทโลยีจะทำให้ประสิทธิภาพการค้นคืนเพิ่มขึ้น เมื่อเทียบกับการใช้ยูเอ็มแอล หรือการไม่ใช้การขยายข้อสอบถาม จากตัวอย่างงานวิจัยจะเห็นได้ว่า การขยายข้อสอบถามจึงถือได้ว่าเป็นกระบวนการที่มีบทบาทสำคัญเป็นอย่างมากในการพัฒนาระบบการค้นคืนสารสนเทศแต่อย่างไรก็ตามจากที่กล่าวมาข้างต้นนั้น การขยายข้อสอบถามนั้นจะประสบความสำเร็จหรือไม่ขึ้นอยู่กับวิธีการขยายข้อสอบถามดังนั้น การกำหนดวิธีที่นำมาใช้ในการขยายข้อสอบถามจึงมีความสำคัญเป็นอย่างยิ่ง งานวิจัยนี้จึงมุ่งประเด็นการกำหนดขั้นตอนวิธีการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนสารสนเทศเพื่อเพิ่มประสิทธิภาพการค้นคืน

รูปแบบของการค้นคืนสารสนเทศนั้นมีผลต่อการกำหนดขั้นตอนการขยายข้อสอบถาม รูปแบบการค้นคืนสารสนเทศรูปแบบหนึ่งที่น่าสนใจในการขยายข้อสอบถาม คือการค้นคืนสารสนเทศภาษาไทย เพราะเป็นรูปแบบหนึ่งของระบบการค้นคืนสารสนเทศที่เริ่มนำการขยายข้อสอบถามมาใช้ เนื่องจากการขยายขอบเขตของสารสนเทศภาษาไทยที่เกิดขึ้นในปัจจุบัน อย่างไรก็ตาม ภาษาไทยมีลักษณะที่แตกต่างจากภาษาอังกฤษซึ่งเป็นภาษาที่พบได้มากในงานวิจัยการค้นคืนสารสนเทศและการขยายข้อสอบถาม เนื่องจาก (1) ภาษาไทยไม่มีการแยกส่วนระหว่างคำสองคำ ในขณะที่ภาษาอังกฤษจะแยกส่วนระหว่างคำสองคำ (2) ภาษาไทยไม่มีการเปลี่ยนรูปแบบคำศัพท์ ในขณะที่ภาษาอังกฤษจะมีการเปลี่ยนรูปแบบคำศัพท์ในหลากหลายรูปแบบ อาทิ กาล (tense)(3) คำหนึ่งคำสามารถทำหน้าที่ได้หลายชนิด อาทิ ที่ สามารถเป็นคำบุพบทเชื่อมคำกับประโยค และสามารถทำหน้าที่เป็นคำนาม ซึ่งหมายถึงสถานที่ได้ (4) ในบางกรณี ภาษาไทยสามารถละประธานในประโยคได้ และ (5) การสร้างคำประสมในภาษาไทยจะใช้คำนามและคำกริยาและให้กำเนิดคำประสมซึ่งเป็นคำนาม (Khankasikam, 2010) ด้วยเหตุนี้ จึงมีงานวิจัยจำนวนหนึ่งที่เกี่ยวข้องกับการขยายข้อสอบถามภาษาไทย เพื่อนำมาใช้ในการพัฒนาระบบการค้นคืน

ภาษาไทยให้มีประสิทธิผลมากยิ่งขึ้น Facundes และ Theva-aksorn (2007) ได้ประยุกต์การขยายข้อสอบถามกับโปรแกรมค้นหาแบบสองภาษา Boonchom และ Soonthornphisaj (2012) จะศึกษาการสร้างออนโทโลยี (Ontology) สำหรับระบบการค้นค้นคำพิพากษาไทย โดยมีการนำขั้นตอนวิธีกลุ่มมด (Ant Colony Algorithm) มาประยุกต์ใช้ในการขยายข้อสอบถาม พบว่าถึงแม้ว่าการสร้างออนโทโลยีด้วยมือ (ซึ่งในงานวิจัยนี้สร้างโดยผู้เชี่ยวชาญด้านกฎหมาย) มีประสิทธิผลสูงกว่าการสร้างออนโทโลยีแบบอัตโนมัติซึ่งสร้างโดย ATOB (Automatic Thai Legal Ontology Building) แต่การนำขั้นตอนวิธีกลุ่มมดในการขยายข้อสอบถามช่วยเพิ่มประสิทธิภาพของการค้นค้น

นอกจากนี้แล้ว โครงสร้างและขั้นตอนการค้นค้นสารสนเทศนั้นก็ส่งผลต่อการขยายข้อสอบถามด้วยเช่นกัน ในงานวิจัยบางชิ้นผู้วิจัยได้คิดค้นทั้งระบบการค้นค้นสารสนเทศและระบบการขยายข้อสอบถามไปพร้อมกัน ดังเช่น งานวิจัยของ Boonchom และ Soonthornphisaj (2012) ที่กล่าวไปแล้วข้างต้น คือการนำออนโทโลยีมาประยุกต์ใช้ในการค้นค้นและการนำขั้นตอนวิธีกลุ่มมดซึ่งต้องนำออนโทโลยีในการค้นค้นมาประยุกต์ใช้ในการขยายข้อสอบถาม หรืองานวิจัยของ Horng และคณะ (2005) ที่ได้ใช้การจัดกลุ่มเอกสารในการค้นค้นและใช้การอนุมานเชิงคลุมเครือในการปรับเปลี่ยนข้อสอบถาม ในงานวิจัยการขยายข้อสอบถามส่วนใหญ่จะนิยมใช้รูปแบบปริภูมิเวกเตอร์ (Vector Space Model) เป็นรูปแบบหลักของการค้นค้นสารสนเทศที่นำมาใช้ในการขยายข้อสอบถาม เนื่องจากรูปแบบนี้สามารถกำหนดค่าน้ำหนักของแต่ละคำในเอกสารได้ จึงสามารถนำมาประยุกต์ใช้ในการขยายข้อสอบถามได้ เพื่อวิเคราะห์หาคำสำหรับการขยายข้อสอบถามได้ดีขึ้น

นอกจากลักษณะของการค้นค้นสารสนเทศแล้ว รูปแบบการขยายข้อสอบถามมีผลต่อการกำหนดขั้นตอนวิธีการขยายข้อสอบถามเช่นกัน การขยายข้อสอบถามสามารถแบ่งได้สองรูปแบบตามลักษณะการดำเนินงานกับการค้นค้นสารสนเทศและการมีส่วนร่วมจากผู้ใช้งาน คือ การขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion, Retrieval System-Automatic Query Expansion, AQE) และการขยายข้อสอบถามแบบโต้ตอบ (Interactive Query Expansion, User-Interactive Query Expansion, IQE) (Shapira, Taieb-Maimon and Nemeth, 2005; Kanaan และคณะ, 2007, 2008; Ruthven, 2003) การขยายข้อสอบถามแบบอัตโนมัตินั้นจะเป็นการขยายข้อสอบถามที่กำหนดข้อสอบถามใหม่เพียงหนึ่งข้อสอบถามและนำไปใช้ในการค้นค้นสารสนเทศต่อไป ผู้ใช้จะไม่มีส่วนร่วมในการกำหนดข้อสอบถามใหม่ ซึ่งจะทำ

ให้ผู้ใช้ไม่สามารถเรียนรู้การกำหนดข้อสอบถามที่ตรงกับความต้องการข้อมูลของตนเอง แต่ข้อสอบถามใหม่ที่นำไปใช้นั้นอาจจะไม่ตรงต่อความต้องการของผู้ใช้ก็เป็นได้ ส่วนการขยายข้อสอบถามแบบโต้ตอบเป็นการขยายข้อสอบถามที่ต้องกำหนดข้อสอบถามใหม่ได้มากกว่าหนึ่งข้อสอบถามเพื่อให้ผู้ใช้มีส่วนร่วมในการเลือกข้อสอบถามใหม่สำหรับการค้นคืนสารสนเทศในรูปแบบการขยายข้อสอบถามนี้ ผู้ใช้จะได้เรียนรู้การกำหนดข้อสอบถามที่นำไปค้นคืนสารสนเทศได้ตรงกับความต้องการข้อมูลของตนเอง แต่ในรายการข้อสอบถามใหม่ จะต้องมีข้อสอบถามหนึ่งข้อสอบถามเท่านั้นที่ตรงกับความต้องการของผู้ใช้ ซึ่งอาจจะต้องใช้เวลาในการตัดสินใจเลือกหรือกำหนดข้อสอบถามสำหรับการนำไปค้นคืนสารสนเทศ

สิ่งหนึ่งที่มีผลต่อการกำหนดขั้นตอนวิธีการขยายข้อสอบถาม คือ การวิเคราะห์และข้อมูลเพื่อการกำหนดคำที่ใช้ในการขยายข้อสอบถาม จากงานวิจัยในอดีต การวิเคราะห์ข้อมูลในการขยายข้อสอบถามมีสิ่งงานวิจัย ได้แก่ (1) วิธีผลป้อนกลับความเกี่ยวข้อง (Relevance Feedback) คือการนำรายการผลลัพธ์การค้นคืนมาวิเคราะห์หาความเกี่ยวข้องแล้วนำไปใช้ในการกำหนดคำใหม่ (2) การวิเคราะห์เฉพาะส่วน (Local Analysis) คือการนำเอกสารจากรายการผลลัพธ์การค้นคืนสารสนเทศมาใช้ในการวิเคราะห์คำที่ปรากฏร่วมกัน เพื่อนำมาใช้ในการกำหนดคำใหม่ (3) การวิเคราะห์ครอบคลุม (Global Analysis) คือการวิเคราะห์เอกสารจากคอร์ปัสเอกสารมาเพื่อกำหนดคำใหม่ และ (4) การนำข้อสอบถามใช้ซ้ำ (Query re-use) คือการนำข้อสอบถามจากอดีตที่บันทึกลงในล็อกข้อสอบถาม (Query Log) (ศุภชัย ตั้งวงศ์ศานต์, 2551) งานวิจัยในอดีตจะศึกษาการวิเคราะห์ข้อมูลเพื่อการกำหนดคำใหม่ในการขยายข้อสอบถาม นอกจากนี้แล้ว ยังมีการศึกษาเปรียบเทียบวิธีการวิเคราะห์ข้อมูล พบว่า การวิเคราะห์เฉพาะส่วนและการวิเคราะห์ครอบคลุมจะช่วยให้ประสิทธิภาพการค้นคืนที่ดีกว่าผลป้อนกลับความเกี่ยวข้องและการนำข้อสอบถามใช้ซ้ำ เนื่องจากการนำคำที่ปรากฏอยู่ในเอกสารมาใช้ในการขยายข้อสอบถาม จึงทำให้ค้นคืนสารสนเทศได้ตรงกับความต้องการของผู้ใช้และลักษณะของเอกสาร นอกจากนี้แล้ว ผลป้อนกลับความเกี่ยวข้องและการนำข้อสอบถามมาใช้ นั้น จะต้องใช้การตัดสินใจของผู้ใช้มาช่วยในการวิเคราะห์ โดยเฉพาะถ้ามีการบันทึกความสัมพันธ์ระหว่างข้อสอบถามจากผู้ใช้และเอกสารที่ผู้ใช้ได้เลือกดูจากรายการผลลัพธ์การค้นคืนที่ระบบได้นำเสนอต่อผู้ใช้ การตัดสินใจของผู้ใช้จะมีผลต่อการวิเคราะห์ข้อมูลเพื่อการขยายข้อสอบถาม เนื่องจากผู้ใช้ส่วนใหญ่จะศึกษาเฉพาะเอกสารในลำดับต้นของรายการผลลัพธ์การค้นคืนเท่านั้น เอกสารในลำดับต้นของรายการผลลัพธ์การค้นคืนจึงได้เพิ่มความเกี่ยวข้องกับข้อสอบถามมากขึ้น (ศุภชัย ตั้งวงศ์ศานต์, 2551; Cui และคณะ, 2002; Kunpeng, Xiaolong, และ Yuanchao, 2006) มีงานวิจัยที่ศึกษาและเปรียบเทียบความ

แตกต่างระหว่างการวิเคราะห์เฉพาะส่วนและการวิเคราะห์ครอบคลุม Xu และ Hu (2010) ได้ศึกษาการนำการจัดกลุ่มเอกสารมาประยุกต์ใช้กับผลป้อนกลับความเกี่ยวข้อง (Relevance Feedback) โดยจำแนกเอกสารที่มำจัดกลุ่มเป็นเอกสารที่เป็นผลลัพธ์และเอกสารทั้งหมดในคอร์ปัสเอกสาร พบว่าการจัดกลุ่มเอกสารที่เป็นผลลัพธ์จะให้ประสิทธิผลที่มากกว่าเมื่อจัดกลุ่มเอกสารในคอร์ปัสเอกสาร Xu และคณะ (2006) ได้ศึกษาการเปรียบเทียบการนำการวิเคราะห์เฉพาะส่วนการวิเคราะห์ครอบคลุมและออนโทโลยีมาประยุกต์กับการขยายข้อสอบถามพบว่าการนำการวิเคราะห์เฉพาะส่วนมาประยุกต์ใช้ในขั้นตอนการขยายข้อสอบถาม จะเพิ่มประสิทธิภาพการค้นคืนได้สูงกว่า การนำการวิเคราะห์ครอบคลุมมาประยุกต์ใช้ งานวิจัยนี้จึงสนใจการนำการวิเคราะห์เฉพาะส่วนมาประยุกต์ใช้ในการกำหนดขั้นตอนการขยายข้อสอบถาม

จากงานวิจัยในอดีต การทำเหมืองข้อมูล (Data Mining) นั้นเริ่มมีบทบาทในการขยายข้อสอบถาม เพื่อนำมาใช้ในการวิเคราะห์และกำหนดคำสำหรับการขยายข้อสอบถาม การทำเหมืองข้อมูลนั้นคือกระบวนการสร้างสารสนเทศหรือองค์ความรู้ใหม่จากข้อมูลที่มีอยู่เดิมแล้วด้วยขั้นตอนวิธีต่างๆ (Linoff และ Berry, 2011; Yin และคณะ, 2011; Han, Kamber และ Pei, 2012; Taffery, 2011) การทำเหมืองข้อมูลที่นิยมนำมาใช้ในการขยายข้อสอบถามนั้น มีสามรูปแบบ ได้แก่ (1) กฎความสัมพันธ์ (Association Rule) เพื่อหาความสัมพันธ์ระหว่างข้อมูลสองหน่วยที่เกิดขึ้นพร้อมกัน Fonseca และคณะ (2005) ได้นำกฎความสัมพันธ์มาสร้างแนวคิดสำหรับการกำหนดคำใหม่ในการขยายข้อสอบถาม Lee, Huang และ Chao (2007) ได้นำกฎความสัมพันธ์มาสร้างคู่ของคำเพื่อกำหนดคำใหม่ในการขยายข้อสอบถาม Song และคณะ (2006) ได้นำกฎความสัมพันธ์มาประยุกต์กับออนโทโลยีและการประมวลผลธรรมชาติ (Natural Language Processing, NLP) ในการขยายข้อสอบถาม และลดปัญหาการกำหนดคำที่มีความหมายใกล้เคียงกัน (2) การจำแนกประเภท (Classification) เป็นการนำข้อมูลไปจัดไว้ในประเภทของข้อมูล หรือกลุ่มที่ได้กำหนดคุณสมบัติของข้อมูลไว้แล้ว Campos, Fernandez และ Huete (1998) ได้นำการจำแนกแบบเบย์ (Bayes Classification) มาใช้ในการสร้างโครงสร้างเครือข่ายภายในอรรถาภิธาน (Thesaurus) เพื่อนำมาใช้ในการขยายข้อสอบถาม Meshram, Mishra และ Sharma (2007) นำซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ร่วมกับการป้อนความเกี่ยวข้องกลับ เพื่อลดขั้นตอนการขยายข้อสอบถามในระบบการค้นคืนภาพวาริน นาราวิทย์และศิริภิญโญ จันทมณี (2554) ได้นำขั้นตอนวิธีเจเนติก (Genetic Algorithm) มาใช้กับการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาไทย การจำแนกประเภทนั้น จะต้องทราบถึงประเภทของข้อมูล หรือเกณฑ์การจำแนกประเภทและ (3) การจัดกลุ่ม (Clustering) จะจัดข้อมูลเข้าสู่กลุ่มที่มีคุณสมบัติคล้ายคลึงกัน

Lin, Lin และ He (2012) ได้นำการจัดกลุ่มด้วยขั้นตอนวิธีค่าเฉลี่ยเค (K-means Algorithm) มาสร้างความสัมพันธ์ระหว่างคำ Rahimi และ Zahedi (2010) ได้นำขั้นตอนการจัดกลุ่มประยุกต์ร่วมกับการวิเคราะห์เชิงความหมายแฝง (Latent Semantic Analysis, LSA) เพื่อกำหนดคำที่นำไปใช้ในการขยายข้อสอบถาม Aono และ Doi (2005) ได้นำเสนอขั้นตอนการจัดกลุ่มแบบลำดับขั้นสำหรับการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศ จากที่กล่าวมานั้น การทำเหมืองข้อมูลนั้นนอกจากจะวิเคราะห์ข้อมูลได้ดีแล้ว ยังสามารถนำมาประยุกต์ใช้ในการขยายข้อสอบถาม อย่างไรก็ตาม เมื่อพิจารณาจากการขยายตัวของสารสนเทศที่เกิดขึ้นในปัจจุบัน จะพบว่าการนำกฎความสัมพันธ์และการจำแนกประเภทมาประยุกต์ใช้ในการขยายข้อสอบถาม อาจจะทำให้ผลการวิเคราะห์ข้อมูลนั้นคลาดเคลื่อนได้ ในการสร้างกฎความสัมพันธ์นั้น จะต้องกำหนดค่าระดับความสัมพันธ์น้อยสุด ยังมีงานวิจัยจำนวนหนึ่งที่ศึกษาการกำหนดค่าระดับความสัมพันธ์น้อยสุด ซึ่งจะขึ้นอยู่กับลักษณะของข้อมูล (Li และคณะ, 1999; Lai; Mueyba, Khan, และ Coenen, 2009; Jayalakshmi และ Rao) ส่วนการจำแนกประเภทข้อมูลนั้น จะต้องทราบและกำหนดประเภทของข้อมูล การขยายตัวของสารสนเทศนั้น จะทำให้ประเภทของข้อมูลที่กำหนดไว้ไม่เพียงพอต่อการนำไปวิเคราะห์ข้อมูล และส่งผลต่อผลลัพธ์ที่ได้จากการจำแนกประเภท ในขณะที่การจัดกลุ่มนั้น สามารถนำมาใช้ในการวิเคราะห์ข้อมูลที่ไม่สามารถจำแนกประเภทได้อย่างชัดเจน มีงานวิจัยจำนวนหนึ่งที่นำการจัดกลุ่มมาเปรียบเทียบกับการทำเหมืองข้อมูลรูปแบบอื่น Rahimi และ Zahedi (2010) ได้นำการจัดกลุ่มมาประยุกต์ร่วมกับการวิเคราะห์เชิงความหมายแฝง (Latent Semantic Analysis) ในการขยายข้อสอบถาม ซึ่งนอกจากนี้ได้นำขั้นตอนวิธี Rocchio (1971) ซึ่งจะต้องจำแนกความเกี่ยวข้องของระหว่างเอกสารและข้อสอบถาม มาใช้ในการเปรียบเทียบกับขั้นตอนที่ได้นำเสนอ พบว่าการจัดกลุ่มให้ประสิทธิภาพการค้นคืนที่ดีกว่า เมื่อเทียบกับการนำขั้นตอน Rocchio (1971) มาใช้ ดังนั้นผู้วิจัยจึงนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถาม

ขั้นตอนวิธีการจัดกลุ่มแบบลำดับขั้นอัลโกลเมิร์ฟ เป็นขั้นตอนการจัดกลุ่มที่นิยมนำมาใช้ในการวิเคราะห์ข้อมูลต่างๆ โดยเฉพาะอย่างยิ่งการขยายข้อสอบถาม ซึ่งจะต้องนำมาใช้ในการวิเคราะห์หาคำสำหรับการนำไปเพิ่มเติมข้อสอบถาม ขั้นตอนการจัดกลุ่มนี้จะเริ่มต้นด้วยการกำหนดกลุ่มตามจำนวนของข้อมูล โดยที่แต่ละกลุ่มจะประกอบด้วยสมาชิกเพียงหนึ่งตัว และกำหนดศูนย์กลางของกลุ่มเท่ากับสมาชิกที่อยู่ในกลุ่มนั้น จากนั้นก็จะเริ่มวิเคราะห์ค่าความใกล้เคียงจากศูนย์กลางของกลุ่ม หากกลุ่มคู่ใดมีค่าความเหมือน (Similarity Measure) มากที่สุดให้รวมคู่ของกลุ่มนั้น เป็นกลุ่มเดียวกันและวิเคราะห์และคำนวณหาค่าศูนย์กลางของกลุ่มใหม่ด้วย

เกณฑ์การเชื่อมโยง (Linkage Criteria) จากนั้นตรวจสอบเงื่อนไขการหยุด (Stopping Criteria) หากเงื่อนไขเป็นเท็จ ให้เริ่มต้นวิเคราะห์ค่าความเหมือนอีกครั้ง แต่หากเป็นจริง ให้ยุติการจัดกลุ่ม (Razmara และ Leila, 2007; Huang, 2008; Bailey, Craswell, และ Soboroff, 2008) จากขั้นตอนที่กล่าวมานั้น จะเห็นได้ว่าการจัดกลุ่ม จะต้องกำหนดจำนวนกลุ่ม (Number of Cluster) สมการการวัดค่าความเหมือน และเกณฑ์การเชื่อมโยง จะมีผลต่อขั้นตอนการจัดกลุ่มแบบลำดับขั้นอัลกอริทึม จากงานวิจัยในอดีต ผู้วิจัยจะกำหนดจำนวนกลุ่มจำนวนหนึ่ง จากนั้นจะนำมาประเมินและเปรียบเทียบคุณภาพการจัดกลุ่มที่ได้จากการจัดกลุ่มด้วยค่าจำนวนกลุ่มที่แตกต่างกัน (Steinbach และคณะ, 2001; Na และคณะ, 2006) ส่วนต่อมา การวัดค่าความเหมือนนั้น มีสมการการวัดค่าความเหมือนที่นิยมใช้ คือ สมการจาคคาร์ด (Jaccard) สมการไดซ์ (Dice) และสมการโคไซน์ (Cosine) (Huang, 2008; Morita และคณะ, 2011; Sesli และ Yegenoglu, 2010; Morita และคณะ, 2011)

จากงานวิจัยในอดีต จะเห็นได้ว่าการขยายข้อสอบถามนั้นมีความสำคัญอย่างยิ่งต่อระบบการค้นคืนสารสนเทศในปัจจุบัน โดยเฉพาะอย่างยิ่งการกำหนดข้อสอบถามให้ตรงกับความต้องการข้อมูลของผู้ใช้ระบบ การค้นคืนสารสนเทศหนึ่งที่มีบทบาทต่อผู้ใช้ในปัจจุบันและริเริ่มมีการนำการขยายข้อสอบถามมาประยุกต์ใช้คือการค้นคืนสารสนเทศภาษาไทย Theva-aksorn และ Facundes (2007) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศแบบภาษาไทยและภาษาอังกฤษ Theva-aksorn (2006) ได้นำเสนอการนำอรรถาภิธานมาสร้างความสัมพันธ์เชิงความหมายมาใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย รัตนา ชุมแก้ว และสุพจน์ นิตย์สุวัฒน์ (2552) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนพระคาถาธรรมบทภาษาบาลีอักษรไทย นอกจากนี้แล้วยังมีงานวิจัยจำนวนหนึ่งที่น่าการทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย Boonchom และ Soonthornprisaj (2012) นำเสนอการค้นคืนสารสนเทศด้วยออนโทโลยี และการขยายข้อสอบถามที่นำมาประยุกต์ด้วยขั้นตอนวิธีกลุ่มมด (Ant Colony Algorithm) วาริน นาราวิทย์ และศิริภิญโญ จันทมูณี (2554) ได้นำขั้นตอนวิธีเจเนติก (Genetic Algorithm) มาประยุกต์ใช้ในการขยายข้อสอบถาม

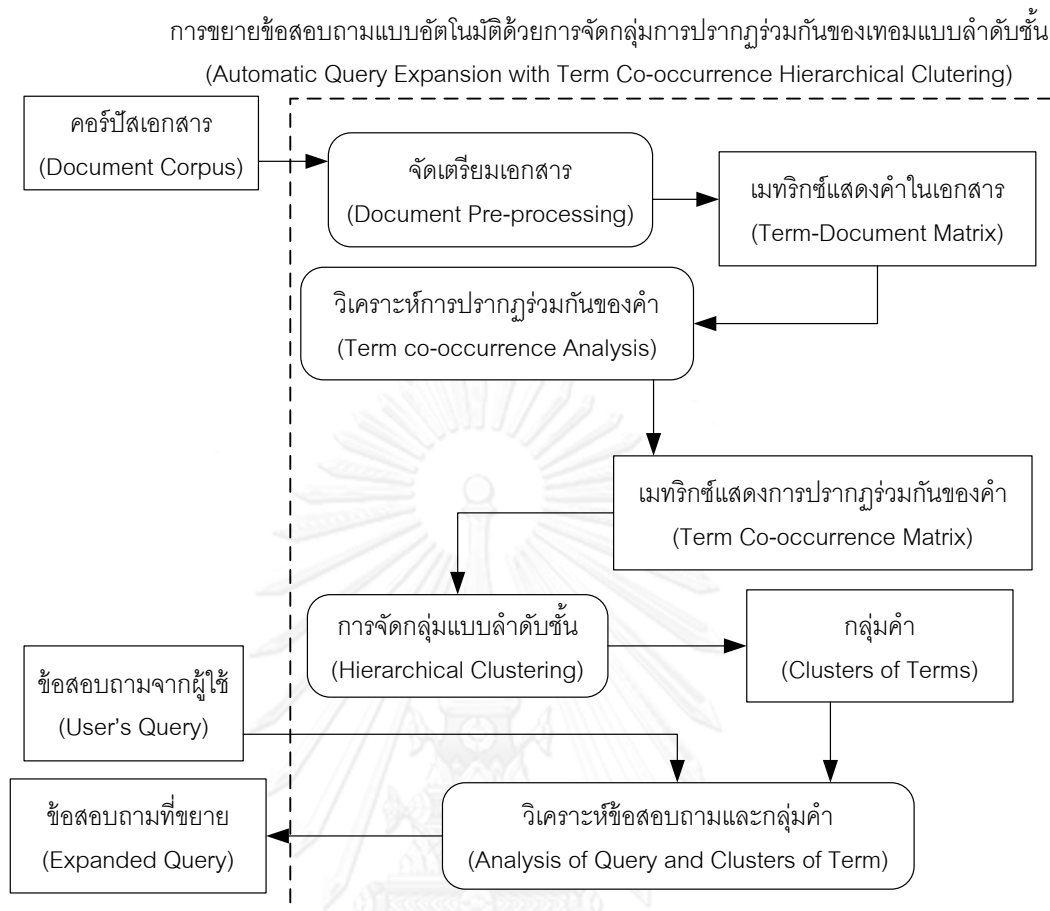
ในขณะที่มีงานวิจัยจำนวนมากที่น่าการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาอังกฤษ ทั้งนี้ยังมีการนำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถามดังกล่าว Kimura และ Araki (2006) ได้นำขั้นตอนวิธีเจเนติก (Genetic Algorithm) มาประยุกต์ใช้ในการ

ขยายข้อสอบถามในระบบการค้นคืนคำถามเชิงสาระภาษาญี่ปุ่น Zhang และคณะ (2007) ได้นำตัวพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เพื่อกำหนดความเกี่ยวข้องของระหว่างเอกสารและข้อสอบถาม สำหรับการกำหนดค่าในผลการประเมินความเกี่ยวข้องของย้อนกลับ Yang และ Zhang (2002) ได้นำการจำแนกแบบเบย์ (Bayes Classification) มาคัดเลือกคำสำหรับการขยายข้อสอบถามในระบบการค้นคืนแบบกระจาย หนึ่งในขั้นตอนวิธีการทำเหมืองข้อมูลที่นำมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศ คือ การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) Aono และ Doi (2005) ได้นำเสนอการขยายข้อสอบถามด้วยการจัดกลุ่มแบบลำดับขั้น Dvorsky, Martinovic และ Snasel (2004) ได้นำการจัดกลุ่มแบบลำดับขั้น มาใช้ในการกำหนดดัชนี เพื่อกำหนดคำสำหรับการขยายข้อสอบถาม Han และ Chen (2009) ได้นำการจัดกลุ่มมาประยุกต์ร่วมกับออนโทโลยีมาจัดกลุ่มคำ เพื่อนำมาใช้ในการวิเคราะห์คำอย่างไรก็ตาม ยังไม่มีการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบสารสนเทศภาษาไทย ทั้งที่มีงานวิจัยการจัดกลุ่มเอกสารหรือสารสนเทศภาษาไทยจำนวนหนึ่ง ฆนาศัย กริ่งไกร และชุตีรัตน์ จรัสกุลชัย (2544) ได้นำเสนอการจัดกลุ่มเอกสารภาษาไทย ด้วยขั้นตอนวิธีการจัดกลุ่มแบบค่าเฉลี่ยเค (K-means Clustering) Santipong และคณะ (2012) ได้นำเสนอการจัดกลุ่มข้อความภาษาไทยด้วยการจัดกลุ่มต้นไม้ต่อท้าย (Suffix Tree Clustering) พัทธระ นาเสงี่ยม, สิริภัทร เขียวชาญวัฒนา และคำรณ สุนิติ (พ.ศ.2553) ได้นำเสนอการจัดกลุ่มมาใช้ในการจำแนกคำจากเอกสารในรายการผลลัพธ์การค้นคืนสารสนเทศ

งานวิจัยนี้จึงนำเสนอการจัดกลุ่มการปรากฏร่วมกันของคำบนระบบการค้นคืนสารสนเทศภาษาไทย ทั้งนี้กำหนดนำการจัดกลุ่มแบบลำดับขั้นมาใช้ในการจัดกลุ่มการปรากฏร่วมกันของคำ ผู้วิจัยจึงกำหนดขอบเขตการศึกษา เพื่อลดปัจจัยที่มีผลต่อการศึกษา (สุทธิติ ชัตติยะ และวิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554) จากงานวิจัยในอดีต ขอบเขตที่ผู้วิจัยจะสามารถกำหนดได้สำหรับงานวิจัยนี้ได้แก่ (1) ขั้นตอนการค้นคืนสารสนเทศ ในงานวิจัยการขยายข้อสอบถาม ขั้นตอนการค้นคืนสารสนเทศเป็นปัจจัยหนึ่งที่มีผลต่อการกำหนดขั้นตอนวิธีการขยายข้อสอบถาม จากการศึกษาในงานวิจัยในอดีต ผู้วิจัยจึงได้กำหนดใช้ รูปแบบปริภูมิเวกเตอร์ (Vector Space Model) มาใช้เป็นรูปแบบการค้นคืนสารสนเทศสำหรับงานวิจัยนี้ เนื่องจากเป็นรูปแบบที่นิยมนำมาใช้ในการกำหนดขั้นตอนการค้นคืนสารสนเทศ (2) รูปแบบการขยายข้อสอบถาม มีงานวิจัยจำนวนมากที่ได้ศึกษาและเปรียบเทียบรูปแบบการขยายข้อสอบถาม รวมทั้งการเลือกกำหนดให้รูปแบบหนึ่ง เพื่อนำมาใช้ศึกษา ในงานวิจัยนี้ จึงได้เลือกกำหนด รูปแบบการขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion) เพื่อลดการนำการตัดสินใจของผู้ใช้ ซึ่ง

อาจจะมีผลต่อการศึกษาในงานวิจัยนี้ และ (3) ข้อมูลที่นำมาใช้ในการวิเคราะห์คำสำหรับการขยายข้อสอบถาม งานวิจัยจำนวนมากได้ศึกษาการนำข้อมูลมาเลือกใช้ในการขยายข้อสอบถาม รวมทั้งเปรียบเทียบการนำข้อมูลต่างๆ มาเปรียบเทียบ เพื่อศึกษาการนำข้อมูลมาใช้ในการขยายข้อสอบถาม จากการศึกษางานวิจัยในอดีตพบว่า การวิเคราะห์ครอบคลุม (Local Analysis) ในการขยายข้อสอบถาม จะสามารถเพิ่มประสิทธิภาพการค้นคืนได้ดีกว่าการวิเคราะห์ข้อมูลอื่น (ศุภชัย ตั้งวงศ์ศานต์, 2551; Cui และคณะ, 2002; Kunpeng, Xiaolong, และ Yuanchao, 2006; Xu และHu, 2010, Xu และคณะ, 2006) ในงานวิจัยนี้ ผู้วิจัยได้เลือกการนำรายการผลลัพธ์การค้นคืน (Retrieval Result List) มาใช้ในการวิเคราะห์ครอบคลุม ซึ่งเป็นข้อมูลที่ยอมรับนำมาใช้ในการวิเคราะห์คำสำหรับการขยายข้อสอบถาม

งานวิจัยนี้จึงต้องการศึกษาการนำการจับกลุ่มการปรากฏร่วมกันของคำมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทยด้วยการจับกลุ่มแบบลำดับชั้น ทั้งนี้ ผู้วิจัยได้นำรูปแบบการค้นคืนแบบปริภูมิเวกเตอร์ การขยายข้อสอบถามแบบอัตโนมัติ และการนำรายการผลลัพธ์การค้นคืนมาใช้ในการวิเคราะห์ครอบคลุมมาใช้ในการวิจัยนี้ ผู้วิจัยจึงได้กำหนดขั้นตอนการขยายข้อสอบถามดังรูปที่ 1.1 ขั้นตอนการขยายข้อสอบถามนี้สามารถแบ่งออกเป็นสี่ขั้นตอนหลักในการขยายข้อสอบถาม ได้แก่ (1) ขั้นตอนการจัดเตรียมเอกสาร (Document Pre-processing) ขั้นตอนนี้จะวิเคราะห์ลักษณะของเอกสารจากรายการผลลัพธ์การค้นคืน ซึ่งได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ผลลัพธ์ที่ได้จากขั้นตอนนี้คือ เมทริกซ์แสดงคำในเอกสาร (Term-Document Matrix) (2) ขั้นตอนการวิเคราะห์การปรากฏร่วมกันของคำ (Term Co-occurrence Analysis) ด้วยการนำเมทริกซ์แสดงคำในเอกสารมาใช้ในการวิเคราะห์ จะได้ผลลัพธ์เป็นเมทริกซ์แสดงการปรากฏร่วมกันของคำ (Term Co-occurrence Analysis) (3) การจับกลุ่มแบบลำดับชั้น เพื่อกำหนดกลุ่มคำ (Clusters of Terms) สำหรับการขยายข้อสอบถาม และ (4) การวิเคราะห์ข้อสอบถามและกลุ่มคำ (Analysis of query and clusters of terms) ทั้งนี้นำข้อสอบถามจากผู้ใช้มาร่วมวิเคราะห์กับกลุ่มคำที่ได้ เพื่อกำหนดคำสำหรับการขยายข้อสอบถาม จนได้ผลลัพธ์เป็นข้อสอบถามที่ขยายแล้ว



รูปที่ 1.1 ขั้นตอนการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันระหว่างคำด้วยการจัดกลุ่มแบบลำดับขั้น

ในการประเมินประสิทธิภาพของขั้นตอนการขยายข้อสอบถาม จากงานวิจัยในอดีต พบว่า ผู้วิจัยต้องกำหนดระบบการค้นคืนสารสนเทศเพื่อนำข้อสอบถามจากผู้ใช้งานและข้อสอบถามที่ได้จากขั้นตอนการขยายข้อสอบถามที่ผู้วิจัยต้องการศึกษา มาใช้ในการค้นคืนสารสนเทศและประเมินรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามทั้งสองชุดด้วยการคำนวณค่าวัดประสิทธิผลการค้นคืนเพื่อเปรียบเทียบความแตกต่างของผลลัพธ์การค้นคืนก่อนหรือหลังการขยายข้อสอบถาม จากรูปที่ 1.1 จะเห็นได้ว่า ระบบการค้นคืนสารสนเทศจะค้นคืนสารสนเทศสองครั้งด้วยข้อสอบถามจากผู้ใช้งานและข้อสอบถามจากการขยายข้อสอบถาม ดังนั้นผู้วิจัยจะเก็บบันทึกรายการผลลัพธ์การค้นคืน เพื่อคำนวณค่าวัดและประเมินผลการค้นคืนสารสนเทศ

จากงานวิจัยการค้นคืนสารสนเทศ รวมทั้งการขยายข้อสอบถาม ซึ่งนำระบบการค้นคืนสารสนเทศมาใช้ในการทดสอบการขยายข้อสอบถาม ผู้วิจัยในอดีตได้นำเสนอหรือเปรียบเทียบ

ขั้นตอนต่างๆ ที่นำมาประยุกต์ใช้ในการค้นคืนสารสนเทศ ด้วยการนำข้อสอบถามมาใช้ในระบบการค้นคืนสารสนเทศที่นำมาศึกษา จากนั้นจึงนำรายการผลลัพธ์การค้นคืนมาประเมินประสิทธิผลการค้นคืนสารสนเทศ (Retrieval Effectiveness Evaluation) ด้วยการคำนวณค่าระลึกได้ (Recall) และค่าความแม่นยำ (Precision) ค่าระลึกได้ คือ สัดส่วนของเอกสารที่อยู่ในรายการผลลัพธ์การค้นคืนที่ปรากฏอยู่ในรายการเอกสารที่เกี่ยวข้องกับข้อสอบถาม ส่วนค่าความแม่นยำ คือ สัดส่วนของเอกสารที่เกี่ยวข้องกับข้อสอบถามที่ปรากฏอยู่ในรายการผลลัพธ์การค้นคืน (ศุภชัย ตั้งวงศ์ศานต์, พ.ศ.2551, Buttcher และคณะ, 2009)

นอกจากการคำนวณค่าระลึกได้และค่าความแม่นยำแล้ว ผู้วิจัยในอดีตนำค่าระลึกได้และค่าความแม่นยำมาประยุกต์เป็นรูปแบบการประเมินประสิทธิผลการค้นคืนจำนวนหนึ่ง ทั้งนี้ผู้วิจัยจะเลือกรูปแบบการประเมินที่เหมาะสมกับวัตถุประสงค์ของงานวิจัย รูปแบบการประเมินที่พบได้จากงานวิจัยในอดีต ได้แก่ (1) กราฟค่าระลึกได้-ค่าความแม่นยำ รูปแบบนี้จะนำเอกสารแต่ละรายการในรายการผลลัพธ์การค้นคืนมาคำนวณค่าระลึกได้และค่าความแม่นยำตามลำดับรายการเอกสาร จากนั้นจึงนำค่าระลึกได้และค่าความแม่นยำของรายการเอกสารมาสร้างกราฟ (2) มาตรฐานของค่าระลึกได้ รูปแบบนี้จะนำเอกสารแต่ละรายการในรายการผลลัพธ์การค้นคืนมาคำนวณค่าระลึกได้และค่าความแม่นยำตามลำดับ จากนั้นจะจัดลำดับเอกสารตามช่วงค่าระลึกได้ไว้ 10 ช่วง และกำหนดลำดับเอกสารลงในแต่ละช่วงและประเมินค่าความแม่นยำในแต่ละช่วงค่าระลึกได้ (3) ค่าความแม่นยำที่อาร์ คือค่าความแม่นยำที่คำนวณจากรายการผลลัพธ์การค้นคืนจำนวนหนึ่งที่มีผู้วิเคราะห์ได้กำหนดไว้ (4) ค่าความแม่นยำอาร์ คือค่าความแม่นยำที่คำนวณจากจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามและปรากฏอยู่ในรายการผลลัพธ์การค้นคืน (5) ค่าความแม่นยำเฉลี่ยกลาง คือค่าเฉลี่ยของค่าความแม่นยำที่อาร์ของเอกสารทั้งหมดในรายการผลลัพธ์การค้นคืน และ (6) ค่าเฉลี่ยฮาร์โมนิกระหว่างค่าระลึกได้และค่าความแม่นยำ หรือค่ามาตรวัดเอฟ

จากการทบทวนงานวิจัยในอดีต ผู้วิจัยนิยมนำค่าความแม่นยำที่อาร์ (Precision at R documents, P@R) และค่ามาตรวัดเอฟ (F-measure) มาใช้ในการวิเคราะห์ประสิทธิผลการค้นคืนด้วยรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามและชุดเอกสารที่นำมาใช้ในการทดลอง โดยเฉพาะอย่างยิ่งการนำระบบการขยายข้อสอบถามมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศ มีงานวิจัยจำนวนมากที่เลือกนำค่าความแม่นยำที่อาร์และค่ามาตรวัดเอฟมาใช้ในการวิเคราะห์ระบบการค้นคืนสารสนเทศ เมื่อมีการนำการขยายข้อสอบถามมาประยุกต์ใช้ร่วม (Collins-Thompson, 2009; Massoudi และคณะ, 2011; Song และคณะ, 2007; Lu และคณะ, 2009)

โดยที่ค่ามาตรฐานวัดเอฟ จะนำมาใช้ในการวิเคราะห์หรือรายการผลลัพธ์การค้นคืนทั้งหมด ส่วนค่าความแม่นยำที่อาร์นั้น จะเลือกเพียงลำดับต้นของรายการผลลัพธ์การค้นคืนจำนวนหนึ่งหรือค่าอาร์ จากการใช้โปรแกรมค้นหา (Search Engine) ซึ่งเป็นระบบการค้นคืนสารสนเทศบนเครือข่ายอินเทอร์เน็ต พบว่าผู้ใช้งานระบบร้อยละ 99 จะพิจารณาเฉพาะรายการผลลัพธ์การค้นคืนหน้าแรกเท่านั้น และมีจำนวนผู้ใช้งานที่เลือกชมรายการผลลัพธ์การค้นคืนหน้าถัดไปจำนวนที่ลดลงตามจำนวนหน้าที่โปรแกรมค้นหานำเสนอต่อผู้ใช้ อย่างไรก็ตาม เมื่อผู้ใช้งานมีความตั้งใจในการค้นคืนสารสนเทศสูง จำนวนหน้าของรายการผลลัพธ์การค้นคืนที่ผู้ใช้เรียกดูนั้นจึงได้เพิ่มขึ้นเป็น 3 หน้า (Liu และคณะ, 2010; Kim, 2008) ผู้วิจัยจึงสนใจเพียงรายการผลลัพธ์การค้นคืนจำนวน 3 หน้า โดยที่แต่ละหน้านั้นจะมีจำนวนรายการเอกสารอย่างน้อย 10 รายการ (Miller, 2010; StatOwl, 2012; Buscher และคณะ, 2012) ดังนั้นจึงกำหนดค่าอาร์สำหรับการวิเคราะห์หรือรายการผลลัพธ์การค้นคืนด้วยค่าความแม่นยำที่อาร์ มีค่าเท่ากับ 10, 20 และ 30 ตามลำดับ

ในงานวิจัยการค้นคืนสารสนเทศ รวมทั้งการขยายข้อสอบถาม ผู้วิจัยจะนำชุดทดสอบ (Test Collection) มาใช้เพื่อประเมินประสิทธิภาพการค้นคืนสารสนเทศ ในแต่ละชุดทดสอบนั้น จะประกอบไปด้วยข้อสอบถาม ชุดเอกสารหรือสารสนเทศ และรายการผลการตัดสินความเกี่ยวข้อง (Relevance Judgement) ระหว่างข้อสอบถามและเอกสาร ผู้วิจัยจะนำข้อสอบถามและเอกสารมาใช้ในการค้นคืนสารสนเทศที่ผู้วิจัยได้นำเสนอ ศึกษาหรือเปรียบเทียบ จากนั้นนำรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนมาใช้ในการวิเคราะห์ด้วยรายการผลการตัดสินความเกี่ยวข้องและการคำนวณค่าระลีกได้และค่าความแม่นยำ เพื่อนำค่าที่คำนวณได้มาใช้ในการวิเคราะห์การค้นคืนสารสนเทศที่ผู้วิจัยต้องการศึกษามีหน่วยงานต่างๆ ที่ให้บริการชุดทดสอบสำหรับการนำมาใช้ในงานวิจัยการค้นคืนสารสนเทศ อาทิ ชุดข้อมูลการประชุมร่วมการค้นคืนข้อความ (Text Retrieval Conference, TREC) การประชุมนี้ได้รับการอุปถัมภ์โดยสถาบันมาตรฐานและเทคโนโลยีแห่งชาติ (National Institute of Standards and Technology, NIST) และสำนักงานวิจัยและพัฒนาขั้นสูง (Advanced Research and Development Activity, ARDA) ประเทศสหรัฐอเมริกา นอกจากนี้แล้ว ผลการตัดสินความเกี่ยวข้องก็มีผลต่อการประเมินประสิทธิภาพการค้นคืน ซึ่งกำหนดโดยผู้เชี่ยวชาญเฉพาะด้านและแตกต่างกันไปตามชุดทดสอบ จึงมีงานวิจัยจำนวนมากที่ได้ศึกษาการกำหนดผลการตัดสินความเกี่ยวข้องของ Xu and Chen (2005,2006) ศึกษาองค์ประกอบที่ผู้ใช้ได้ตัดสินใจถึงความเกี่ยวข้องระหว่างข้อสอบถามและ

เอกสาร Zhang (2002) ได้ศึกษาการตัดสินความเกี่ยวข้องด้วยขั้นตอนวิธีความคิดเห็นของคนส่วนใหญ่ในกลุ่ม (Group Consensus Method) Wu และ McClean (2006) ได้กำหนดระดับความเกี่ยวข้องจำนวนมากกว่าสองระดับ (Relevance Level)

ชุดทดสอบที่นิยมนำมาใช้ไม่สามารถนำมาใช้ได้กับการประเมินกับระบบการค้นคืนสารสนเทศทุกระบบ เนื่องจากลักษณะของการค้นคืนสารสนเทศและขอบเขตของสารสนเทศที่แตกต่างกับสารสนเทศที่อยู่ในชุดทดสอบ Oard และคณะ (2004) ได้สร้างชุดทดสอบสำหรับคำพูดเชิงสนทนาตามปกติ (Spontaneous Conversational Speech) โดยมีสารสนเทศเป็น โดยรวบรวมแฟ้มข้อมูลเสียงการสัมภาษณ์ สร้างชุดหัวข้อ และกำหนดผลการตัดสินความเกี่ยวข้องระหว่างแฟ้มข้อมูลเสียงและหัวข้อตั้งแต่ 0 ถึง 4 ทั้งนี้ 0 หมายถึงไม่เกี่ยวข้องมากที่สุด และ 4 หมายถึงเกี่ยวข้องกันมากที่สุด Macdonald และ Qunis (2006) ได้สร้างชุดทดสอบเว็บส่วนบุคคล (Blog) ซึ่งในแต่ละหน้าเว็บจะมีหัวข้อ (Topic) และประกาศ (Post) Ritchie, Teufel และ Robertson (2006) ได้สร้างชุดทดสอบสำหรับการทดลองการค้นคืนสารสนเทศด้านหมายเรียกบุคคลโดยนำหลักการสร้างชุดทดสอบแครนฟิลด์สอง (Cranfield 2 Test Collection) ซึ่งจะสอบถามความเกี่ยวข้องระหว่างเอกสารและข้อสอบถามด้วยการสอบถามผู้จัดทำเอกสาร

งานวิจัยนี้มีวัตถุประสงค์เพื่อการนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศภาษาไทย จึงต้องใช้ชุดทดสอบภาษาไทย เพื่อนำมาใช้ในการประเมินและเปรียบเทียบประสิทธิภาพการค้นคืนสารสนเทศก่อนและหลังการขยายข้อสอบถาม อย่างไรก็ตามไม่มีหน่วยงานใดที่ริเริ่มพัฒนาชุดทดสอบภาษาไทย ดังเช่นชุดทดสอบภาษาอังกฤษที่มีหน่วยงานจำนวนหนึ่งได้พัฒนาและนำเสนอให้ผู้วิจัยระบบการค้นคืนสารสนเทศมาใช้ (Li, Zu และ Geva, 2008; Jaruskulchai, 2001) ทั้งที่มีงานวิจัยระบบการค้นคืนสารสนเทศภาษาไทยจำนวนมาก (Chuleerat, 2001; Khankasikam, 2010; Tantisripreecha และ Soonthornphisaj, 2010; Thava-askorn, 2003; วาริน นาราวิทย์ และศิริกัญญา จันทมณี, 2554) ดังนั้นผู้วิจัยจึงต้องสร้างชุดทดสอบภาษาไทยสำหรับการประเมินประสิทธิภาพการค้นคืนในงานวิจัยนี้ ในการสร้างชุดทดสอบภาษาไทย ผู้วิจัยต้อง (1) รวบรวมเอกสาร (2) คัดเลือกผู้ประเมินความเกี่ยวข้อง (3) กำหนดให้ผู้ประเมินความเกี่ยวข้องกำหนดข้อสอบถามหนึ่งข้อสอบถามต่อหนึ่งคน (4) กำหนดให้ผู้ประเมินความเกี่ยวข้องศึกษาเอกสาร และประเมินความเกี่ยวข้องระหว่างเอกสารและข้อสอบถาม

ถามของตนเอง และ (5) ผู้วิจัยรวบรวมรายการข้อสอบถามและรายการผลการประเมินความเกี่ยวข้อง จากผู้ประเมินความเกี่ยวข้อง เพื่อกำหนดดัชนีจากรายการเอกสารและสรุปผลการตัดสินความเกี่ยวข้อง (Oard และคณะ, 2004; Ritchie, Teufel และ Robertson, 2006; Macdonald และ Qunis, 2006) ในแต่ละขั้นตอนนั้น ผู้วิจัยจะต้องศึกษาและประเมินขอบเขตของเอกสารหรือสารสนเทศ ผู้ประเมินความเกี่ยวข้อง และขั้นตอนวิธีการเก็บข้อมูล เพื่อลดปัจจัยที่อาจจะมีผลต่องานวิจัยนี้

ระบบการค้นคืนสารสนเทศที่มีจำนวนผู้ใช้งานมากที่สุด คือโปรแกรมค้นหา (Search Engine) (สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2553) ซึ่งเป็นระบบการค้นคืนสารสนเทศบนเครือข่ายอินเทอร์เน็ต ซึ่งผู้ใช้งานทั่วไปสามารถเรียกใช้งานได้ ผู้วิจัยจึงคัดเลือกผู้ประเมินความเกี่ยวข้อง ลักษณะของเอกสาร และขอบเขตการจัดเก็บข้อมูลจากพฤติกรรมการใช้โปรแกรมการค้นหา

มีหน่วยงานต่างๆ ของภาครัฐและเอกชนที่ได้สำรวจ ศึกษาและประเมินพฤติกรรมการใช้อินเทอร์เน็ต รวมทั้งการใช้โปรแกรมค้นหา รวมทั้งข้อมูลทางประชากรศาสตร์ (Demographics) ของผู้ใช้อินเทอร์เน็ต สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (2553) และสำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (2556) ได้สำรวจพฤติกรรมการใช้อินเทอร์เน็ต พบว่าผู้ร่วมตอบแบบสอบถามนั้นมีช่วงอายุ 20-34 ปี โดยร้อยละ 40 มีการใช้โปรแกรมค้นหา เพื่อตอบสนองความต้องการข้อมูล มีจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์มากที่สุด อีกทั้งนักเรียนนักศึกษา มีจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์ 34.2 ชั่วโมง ซึ่งใกล้เคียงกับค่าเฉลี่ยของจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์ทั้งหมด 32.3 ชั่วโมง ขณะที่ ผู้วิจัยจึงกำหนดลักษณะของผู้ประเมินความเกี่ยวข้อง เป็นนิสิตนักศึกษาที่มีช่วงอายุ 20-34 ปี ผู้วิจัยจึงกำหนดให้นิสิตที่กำลังศึกษาชั้นปีที่ 4 คณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2555 เป็นกลุ่มตัวอย่างผู้ประเมินความเกี่ยวข้องสำหรับงานวิจัยนี้ มีนิสิตกลุ่มดังกล่าวสนใจเข้าร่วมศึกษาและประเมินเอกสารจำนวน 25 คน

เมื่อกำหนดผู้เข้าร่วมประเมินเอกสารเรียบร้อยแล้ว ผู้ประเมินความเกี่ยวข้องต้องกำหนดข้อสอบถามหนึ่งข้อสอบถามต่อหนึ่งคน เพื่อนำมาใช้ในการศึกษาและประเมินเอกสาร เพื่อให้ผู้ประเมินได้กำหนดข้อสอบถามได้อย่างชัดเจน ผู้วิจัยจะต้องกำหนดจำนวนคำในข้อสอบถาม (Query Size) มีหน่วยงานที่ได้สำรวจลักษณะของข้อสอบถามที่ผู้ใช้โปรแกรมนำมาใช้ในการค้นคืนสารสนเทศ รูปที่ 1.2 แสดงถึงผลการสำรวจของ A Trellian Company (2013) เกี่ยวกับการมีส่วนร่วมของการเข้าชมหน้าเว็บเพจที่ได้จากรายการผลลัพธ์การค้นหาที่ได้จากข้อสอบถามที่มีจำนวนคำใน

ข้อสอบถามที่แตกต่างกัน โดยพิจารณาจากสัดส่วนการคลิกเลือกชมหน้าเว็บเพจ (Organic), สัดส่วนการคลิกโฆษณา (PPC) และสัดส่วนการคลิกเลือกชมทั้งหมด (Total) ในหน้าผลลัพธ์โปรแกรมค้นหา (Search Engine Result Page, SERP) จากรูปที่ 1.2 จะเห็นได้ว่าเมื่อกำหนดข้อสอบถามที่มีจำนวนคำในข้อสอบถามเท่ากับเจ็ดหรือแปดคำ จะได้สัดส่วนการเข้าชมหน้าเว็บเพจจากรายการผลลัพธ์การค้นหาที่มีค่าสูงสุด ผู้วิจัยจึงกำหนดให้ผู้ประเมินความเกี่ยวข้องของทุกคนต้องกำหนดข้อสอบถามที่มีจำนวนคำในข้อสอบถามเท่ากับเจ็ดหรือแปดคำ

Words:	Total(%)	Organic(%)	PPC(%)
Total	34.76%	33.91%	0.85%
1	21.74%	21.41%	0.33%
2	71.36%	69.81%	1.55%
3	65.84%	63.82%	2.02%
4	63.43%	60.57%	2.86%
5	56.83%	54.08%	2.75%
6	55.48%	52.79%	2.69%
7	112.98%	108.26%	4.72%
8	111.27%	106.93%	4.34%
9	96.75%	93.06%	3.69%
10+	22.47%	20.99%	1.48%

รูปที่ 1.2 ผลการสำรวจการเข้าชมหน้าเว็บเพจจำแนกตามจำนวนคำในข้อสอบถามที่ผู้ใช้กำหนด คัดลอกจาก A Trellian Company (2013)

หลังจากกำหนดจำนวนคำในข้อสอบถามและกำหนดข้อสอบถาม ผู้วิจัยจึงกำหนดให้ผู้ประเมินความเกี่ยวข้องได้ศึกษาและประเมินเอกสาร เอกสารที่นำมาใช้ในการค้นคืนสารสนเทศนั้นจะต้องเป็นเอกสารที่ไม่มีโครงสร้าง (ศุภชัย ตั้งวงศ์วานต์, 2551) ผู้วิจัยในอดีตได้กำหนดลักษณะและจำนวนของรายการเอกสารที่นำมาใช้ในการศึกษา ประเมิน และสร้างชุดทดสอบสำหรับการประเมินประสิทธิภาพการค้นคืนสารสนเทศ ซึ่งต้องใช้ระยะเวลาและความตั้งใจของผู้ร่วมประเมินความเกี่ยวข้อง อย่างไรก็ตามจำนวนหรือความยาวของเอกสารที่นำมาใช้ในการประเมินนั้นอาจจะส่งผลต่อผลการประเมินความเกี่ยวข้องที่ผู้ร่วมประเมินได้กำหนด ในงานวิจัยนี้ ผู้วิจัยจึง

ต้องกำหนดลักษณะและจำนวนของรายการเอกสารที่เหมาะสมต่อการประเมิน งานวิจัยในอดีตได้นำเสนอรูปแบบตัวอย่างของเอกสารหรือสารสนเทศที่นำมาใช้ในการสร้างชุดทดสอบ งานวิจัยบางส่วนได้นำเสนอลักษณะของเอกสารในรูปแบบเอกสารเอ็กซ์ตรีมแมล (Extreme Markup Language, XML) ซึ่งมีการกำหนดหัวข้อย่อยในเอกสาร อาทิ MacDonald (2006) และ Ircing และคณะ (2007) ดังนั้นผู้วิจัยจึงกำหนดลักษณะของเอกสารที่นำมาใช้ จะต้องเป็นเอกสารที่ไม่มีโครงสร้าง มีความยาวไม่เกินหนึ่งหน้ากระดาษ และมีการกำหนดหัวข้อย่อยในเอกสาร ผู้วิจัยจึงเลือกใช้บทความของโครงการชั้นปีที่ 4 ที่จัดทำโดยนิสิตที่กำลังศึกษาชั้นปีที่ 4 คณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ทั้งนี้ผู้วิจัยได้รับความอนุเคราะห์จากภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ให้ใช้บทความโครงการชั้นปีที่ 4 ที่จัดทำในช่วงปีการศึกษา 2552-2554 จำนวนทั้งสิ้น 100 ฉบับ

หลังจากได้เอกสารที่จะนำมาใช้ในการประเมินแล้ว ผู้วิจัยจะต้องกำหนดรูปแบบการศึกษาและประเมินเอกสารของผู้ประเมินความเกี่ยวข้อง โดยลดปัจจัยอื่นๆ ที่อาจจะส่งผลกระทบต่อผลการตัดสินความเกี่ยวข้อง มีงานวิจัยจำนวนหนึ่งได้กำหนดรูปแบบการศึกษาและประเมินเอกสารและข้อสอบถาม Oard และคณะ (2004) ได้ใช้นักเรียนจำนวนสี่คนในการประเมินสารสนเทศจำนวน 404 รายการด้วยระยะเวลา 700 ชั่วโมง ในช่วงระยะเวลามากกว่า 3 เดือน Ritchie และคณะ (2006) ได้รวบรวมรายการเอกสารและรายการข้อสอบถาม จากนั้นจึงนำรายการข้อสอบถามส่งไปให้ผู้เขียนเอกสารที่ปรากฏในเอกสารนั้น เพื่อให้ประเมินความเกี่ยวข้องระหว่างเอกสารและข้อสอบถาม แต่ระยะเวลาที่ใช้ในการประเมินนั้นจะแตกต่างกันไปตามความสะดวกของผู้เขียนเอกสาร จากข้อความดังกล่าว จะเห็นได้ว่าผู้วิจัยในอดีตจะกำหนดจำนวนครั้ง การนัดหมายการศึกษาและประเมินเอกสาร รวมทั้งจำนวนเอกสารที่นำมาใช้ในการประเมินแต่ละครั้ง ในงานวิจัยนี้ได้ผู้ประเมินจำนวน 25 คน และรายการเอกสาร 100 ฉบับ เพื่อลดปัจจัยที่มีผลต่อการประเมินความเกี่ยวข้องและระยะเวลาที่ใช้ในการประเมินเอกสารแต่ละฉบับ ผู้วิจัยจึงกำหนดให้นัดหมายการศึกษาและประเมินเอกสารทั้งสิ้นสี่ครั้ง ในแต่ละครั้ง กลุ่มผู้ประเมินจะศึกษาเอกสารจำนวน 25 ฉบับพร้อมกัน โดยเริ่มต้น ผู้ประเมินจะได้รับเอกสารคนละหนึ่งฉบับและแบบฟอร์มจากผู้วิจัย หลังจากนั้น เมื่อได้ศึกษาเอกสารเรียบร้อยแล้ว ผู้ประเมินจะบันทึกรหัสเอกสารและผลการประเมินความเกี่ยวข้องลงในแบบฟอร์ม จากนั้นผู้ประเมินจะส่งต่อเอกสารให้ผู้ประเมินคนอื่น และรับเอกสารฉบับใหม่จากผู้ประเมินอีกคน เพื่อศึกษาและประเมินเอกสารต่อไป จนกระทั่งได้ศึกษาและประเมินครบ 25 ฉบับ ในการประเมินครั้งนั้น

หลังจากผู้วิจัยได้รวบรวมผลการตัดสินความเกี่ยวข้องจากผู้ร่วมประเมิน ผู้วิจัยต้องกำหนดดัชนีในชุดทดสอบ (ศุภชัย ตั้งวงศ์วานต์, 2551) จากงานวิจัยการสร้างชุดทดสอบในอดีต ผู้วิจัยได้กำหนดวิธีการสร้างดัชนีสำหรับชุดทดสอบที่แตกต่างกันออกไป Oard และคณะ (2004) และ Caterette, Allan และ Sitataman (2006) ได้กำหนดดัชนีด้วยการใช้หัวข้อ (Topic) หรือคำสำคัญ (Keyword) จากเอกสารทั้งหมดหรือความสนใจของผู้ศึกษาเอกสาร ในการวิเคราะห์ดัชนีที่นำมาใช้ในการสร้างชุดทดสอบสำหรับงานวิจัยนี้ ผู้วิจัยได้นำรายการคำหยุด (Stop-words List) จากวรรณีย์ เกษร (2553) และพจนานุกรมอิเล็กทรอนิกส์ (Electronic Dictionary) จากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronic and Computer Technology Center) (ปรับปรุงข้อมูลล่าสุดปี 2552) ร่วมวิเคราะห์ ทั้งนี้ผู้วิจัยได้นำเครื่องมือไทยอานาไลเซอร์ (ThaiAnalyzer) ซึ่งเป็นส่วนขยายของเครื่องมือระบบการค้นคืนสารสนเทศอะพาเช่ลูซีน 3.5 (Apache Lucene 3.5) มาใช้ในการแยกคำในเอกสารภาษาไทย ขั้นตอนในการวิเคราะห์และกำหนดดัชนีนั้น ผู้วิจัยจะนำเครื่องมือไทยอานาไลเซอร์มาใช้ในการแยกคำในเอกสาร และนำคำที่ปรากฏอยู่ในรายการคำหยุดภาษาไทยที่นำมาใช้ออกจากรายการคำที่ได้จากการแยกคำภาษาไทย หลังจากนั้น ผู้วิจัยจึงรายการคำมาวิเคราะห์ความสัมพันธ์ระหว่างคำ เพื่อกำหนดดัชนีด้วยการนำคำที่ได้จากการค้นคืนและคำพ้องความหมาย (Synonym) ซึ่งมีมากกว่าหนึ่งคำในแต่ละรายการของพจนานุกรมอิเล็กทรอนิกส์มาประกอบการวิเคราะห์

เพื่อกำหนดดัชนีสำหรับชุดทดสอบผู้วิจัยจึงได้นำพจนานุกรมอิเล็กทรอนิกส์มาประกอบการวิเคราะห์ ในพจนานุกรมอิเล็กทรอนิกส์ อยู่ในรูปแบบเอกซ์เอ็มแอล (XML, EXtensible Markup Language) ซึ่งคำหนึ่งคำสามารถมีความหมายที่เหมือนกันหรือใกล้เคียงกับคำอื่นๆ มากกว่าหนึ่งคำ ด้วยรายการคำพ้องความหมาย ผู้วิจัยจึงวิเคราะห์ความสัมพันธ์ของคำที่มีความหมายเหมือนกัน เพื่อกำหนดดัชนีสำหรับชุดทดสอบนี้จากพจนานุกรมอิเล็กทรอนิกส์นี้ หลังจากนั้น ผู้วิจัยจึงสามารถสรุปผลการตัดสินความเกี่ยวข้องระหว่างข้อสอบถามและข้อสอบถามด้วยดัชนีที่ได้

ในการพัฒนาระบบการค้นคืนสารสนเทศสำหรับงานวิจัยนี้ ผู้วิจัยจึงได้เลือกใช้ระบบการค้นคืนสารสนเทศโอเพนซอร์ส (Open-sourced Information Retrieval System) ซึ่งเป็นระบบการค้นคืนสารสนเทศที่เจ้าของลิขสิทธิ์ได้เปิดโอกาสให้ผู้พัฒนาหรือผู้วิจัยได้พัฒนาระบบการค้นคืนสารสนเทศ และสามารถควบคุมปัจจัยอื่นๆ ที่มีผลต่อการค้นคืนสารสนเทศและประสิทธิภาพการค้นคืนสารสนเทศ (สุทธิ ชัดติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554) เมื่อผู้วิจัยได้ประเมินและคัดเลือกระบบการค้นคืนสารสนเทศโอเพนซอร์ส ผู้วิจัยพบว่า เครื่องมือ

การค้นคืนสารสนเทศอะพาเช่ลูซีน 3.5 (Apache Lucene 3.5) ซึ่งจะใช้รูปแบบการค้นคืนปริภูมิเวกเตอร์ (Vector Space Model) แล้วยังมีส่วนขยายที่สามารถนำมาใช้ร่วมกับเครื่องมือนี้คือ ส่วนขยายการวิเคราะห์การแยกคำภาษาไทย ไทยอนาไลเซอร์ (Thai Analyzer) ซึ่งจะช่วยวิเคราะห์คำภาษาไทยที่มีอยู่ในเอกสารได้

ในขั้นตอนการขยายข้อสอบถามที่ผู้วิจัยได้นำเสนอนั้น จะมีการจัดกลุ่มคำ (Term Clustering) ด้วยขั้นตอนวิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) แต่เนื่องจากผู้วิจัยจะสนใจเฉพาะการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามนั้น ผู้วิจัยจึงเลือกใช้เครื่องมือการทำเหมืองข้อมูลมาประยุกต์ใช้ในขั้นตอนการขยายข้อสอบถาม (สุทธิติ ชัตติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554) จากที่กล่าวมาข้างต้น ผู้วิจัยจะได้รายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ที่แตกต่างกันจำนวน 25 ชุด ซึ่งจะนำมาใช้ในการวิเคราะห์เมทริกซ์แสดงการปรากฏร่วมกันของคำและเซตของกลุ่มคำตามลำดับ ด้วยเหตุนี้ ผู้วิจัยไม่สามารถกำหนดรูปแบบการจัดกลุ่มคำที่ชัดเจนได้สำหรับงานวิจัยนี้

มีงานวิจัยจำนวนมากที่ศึกษาการกำหนดรูปแบบการจัดกลุ่ม รูปแบบการจัดกลุ่มที่ผู้วิจัยในอดีตสนใจศึกษา ได้แก่ (1) จำนวนกลุ่ม (Number of Clusters) (Tibshirani, Walther, และ Hastle, 2001; Dudiot และ Fridlyand, 2002; Yan, 2005) (2) สมการการวัดค่าความเหมือน (Similarity Measure) (Huang, 2008; Balcan, Blum และ Vempala, 2008) (3) เกณฑ์การเชื่อมโยง (Linkage Criteria) (Rashid และคณะ, 2011) ผู้วิจัยจึงศึกษารูปแบบการจัดกลุ่มเพื่อนำมาใช้ในขั้นตอนการจัดกลุ่มคำในการขยายข้อสอบถาม ดังที่ผู้วิจัยได้นำเสนอไว้ข้างต้น จำนวนกลุ่มที่นิยมนำมาใช้ในการจัดกลุ่มนั้นมีจำนวนเท่ากับ 5, 10, 15, 20, 25 และ 30 สมการค่าความเหมือนที่นิยมนำมาใช้ได้แก่ สมการโคไซน์ (Cosine) สมการยูคลิเดียน (Euclidean) และสมการจาคคาร์ด (Jaccard) สมการเกณฑ์การเชื่อมโยงที่นิยมนำมาใช้ได้แก่ สมการเกณฑ์การเชื่อมโยงเชิงเดี่ยว (Single Linkage Criteria) สมการเกณฑ์การเชื่อมโยงเชิงสมบูรณ์ (Complete Linkage Criteria) และสมการเกณฑ์การเชื่อมโยงเชิงค่าเฉลี่ย (Average Linkage Criteria) ผู้วิจัยจึงกำหนดรูปแบบการจัดกลุ่มที่กำหนดจำนวนกลุ่ม สมการค่าความเหมือน และสมการเกณฑ์การเชื่อมโยงที่แตกต่างกัน 54 รูปแบบ มาวิเคราะห์การจัดกลุ่มคำจากเมทริกซ์การปรากฏร่วมกันของคำที่แตกต่างกัน 25 เมทริกซ์ เพื่อวิเคราะห์และกำหนดรูปแบบการจัดกลุ่มสำหรับขั้นตอนการขยายข้อสอบถามดังที่ได้นำเสนอไว้ข้างต้น ผู้วิจัยจึงกำหนดให้นำเมทริกซ์การปรากฏร่วมกันของคำแต่ละเมทริกซ์มาวิเคราะห์การจัดกลุ่มด้วยรูปแบบการจัดกลุ่มที่แตกต่างกัน 54 รูปแบบ จากนั้น ผู้วิจัยจึง

วิเคราะห์เลือกรูปแบบการจัดกลุ่มที่ให้คุณภาพการจัดกลุ่มที่ดีที่สุด ซึ่งจะได้รูปแบบการจัดกลุ่มหนึ่งรูปแบบสำหรับเมทริกซ์การปรากฏร่วมกันของคำหนึ่งเมทริกซ์ จากนั้นผู้วิจัยจึงวิเคราะห์รูปแบบการจัดกลุ่ม 25 รูปแบบที่ได้จากการจัดกลุ่มคำในเมทริกซ์ 25 เมทริกซ์ เพื่อวิเคราะห์หารูปแบบการจัดกลุ่มที่เหมาะสมสำหรับงานวิจัยนี้

ในการกำหนดรูปแบบการจัดกลุ่มสำหรับขั้นตอนการขยายข้อสอบถามนั้น ผู้วิจัยจะต้องวิเคราะห์คุณภาพการจัดกลุ่ม ที่ได้จากการกำหนดรูปแบบการจัดกลุ่มที่แตกต่างกัน การจัดกลุ่มเป็นขั้นตอนวิธีการรวบรวมวัตถุหรืออ็อบเจกต์ที่คล้ายคลึงเป็นกลุ่มเดียวกัน โดยวิเคราะห์ความเหมือนระหว่างวัตถุ (Evertt และคณะ, 2011; Jacob, 2007; Han, Kamber และ Pei, 2012; Weiss และคณะ, 2005; Tuffey, 2011) ผู้วิจัยในอดีตจึงวิเคราะห์คุณภาพการจัดกลุ่มด้วย ความเหมือนระหว่างกลุ่ม (Inter-Cluster Similarity) และความเหมือนภายในกลุ่ม (Intra-Cluster Similarity) ความเหมือนระหว่างกลุ่ม หมายถึงความสัมพันธ์ระหว่างสมาชิกที่ไม่อยู่ในกลุ่มเดียวกัน ซึ่งควรมีระดับความสัมพันธ์ที่ต่ำ ส่วนความเหมือนภายในกลุ่ม หมายถึงความสัมพันธ์ระหว่างสมาชิกที่อยู่ในกลุ่มเดียวกัน ซึ่งควรมีระดับความสัมพันธ์ที่สูง Raskutti และ Leckie (1999) ได้นำเสนอการประเมินคุณภาพการจัดกลุ่ม โดยวิเคราะห์ความกะทัดรัด (Compactness) และ ความเป็นอิสระ (Isolation) เมื่อวิเคราะห์เซตของกลุ่มที่ได้จากการกำหนดจำนวนกลุ่มที่แตกต่างกัน ผู้วิจัยจึงเลือกใช้หลักการการจัดกลุ่มมาใช้ในการประเมินคุณภาพการจัดกลุ่ม เช่นเดียวกับงานวิจัยในอดีตผู้วิจัยจึงได้เลือกใช้ค่าเบี่ยงเบนมาตรฐานของรากของผลรวมกำลังสอง (Root Mean Square Standard Deviation) ระยะทางสูงสุดจากกลุ่ม (Maximum Distance from Cluster) และระยะทางสู่กลุ่มที่ใกล้ที่สุด (Distance to Nearest Cluster) มาใช้ในการวิเคราะห์และประเมินคุณภาพการจัดกลุ่ม เพื่อนำไปใช้ในการกำหนดการจัดกลุ่มคำ ด้วยเครื่องมือการทำเหมืองข้อมูลเวก้า โดยที่ค่าเบี่ยงเบนมาตรฐานของรากของผลรวมกำลังสองและระยะทางสูงสุดจากกลุ่มนำมาใช้ในการประเมินความเหมือนภายในกลุ่ม ส่วนระยะทางสู่กลุ่มที่ใกล้ที่สุดมาใช้ในการวิเคราะห์ความเหมือนระหว่างกลุ่ม

1.2. วัตถุประสงค์ของงานวิจัย

เพื่อเปรียบเทียบประสิทธิภาพการค้นคืนระหว่างการใช้ข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถามด้วยการจัดกลุ่มแบบลำดับชั้นในระบบการค้นคืนสารสนเทศด้วยคำวัดสี่ตัว คือ (1) ค่าความแม่นยำของผลลัพธ์จำนวนสิบรายการแรก (2) ค่าความแม่นยำจำนวนยี่สิบรายการแรก (3) ค่าความแม่นยำจำนวนสามสิบรายการแรก และ (4) ค่ามาตรวัดเอฟ

1.3. ขอบเขตของการวิจัย

1.3.1. ผู้วิจัยได้กำหนดกลุ่มตัวอย่างผู้ประเมินความเกี่ยวข้องเป็นนิสิตที่กำลังศึกษาชั้นปีที่ 4 คณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2555 จำนวน 25 คน

1.3.2. ผู้วิจัยได้กำหนดกลุ่มตัวอย่างเอกสาร เป็นบทความของโครงการชั้นปีที่ 4 ที่จัดทำโดยนิสิตคณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ในช่วงปีการศึกษา 2552-2554 จำนวน 100 ฉบับ

1.3.3. ในการวิเคราะห์กำหนดดัชนีสำหรับชุดทดสอบนั้น ผู้วิจัยจะไม่คำนึงถึงไวยากรณ์ของภาษา ทั้งนี้ ผู้วิจัยได้นำรายการคำหยุดและรายการคำในพจนานุกรมอิเล็กทรอนิกส์มาใช้ในการคัดเลือกคำที่นำมาใช้ในการขยายข้อสอบถาม

1.3.4. ผู้วิจัยจะนำเครื่องมือการทำเหมืองข้อมูล และเครื่องมือระบบการค้นคืนสารสนเทศมาใช้ในการพัฒนาระบบการขยายข้อสอบถาม

1.3.5. ระบบการค้นคืนสารสนเทศภาษาไทยที่ผู้วิจัยสนใจนำมาใช้นั้นมีรูปแบบปริภูมิเวกเตอร์ (Vector Space Model) เท่านั้น

1.3.6. ผู้วิจัยจะกำหนดรูปแบบการจัดกลุ่มสำหรับขั้นตอนการจัดกลุ่มแบบลำดับชั้นในการขยายข้อสอบถาม โดยพิจารณาถึงรูปแบบการจัดกลุ่มที่ให้คุณภาพการจัดกลุ่มที่ดีที่สุด เมื่อนำรายการผลลัพธ์การค้นคืนหนึ่งชุดมาวิเคราะห์

1.4. ตัวแปรสำคัญที่ศึกษา

1.4.1. ตัวแปรต้นหรือตัวแปรอิสระ (Independent Variable) มี 1 ตัวแปร คือ

- 1) การใช้ระบบการขยายข้อสอบถามด้วยการจัดกลุ่มแบบลำดับชั้นสามารถจำแนกค่าของตัวแปรได้เป็น 2 ค่า คือ (1) มีการนำระบบการขยายข้อสอบถามไปใช้ และ (2) ไม่มีการนำระบบการขยายข้อสอบถามไปใช้

1.4.2. ตัวแปรตาม (Dependent Variable) มี 4 ตัวแปร คือ

- 1) ค่าความแม่นยำจำนวนสิบรายการแรก (Precision at 10 documents) โดยวัดจากความสามารถในการค้นคืนเอกสาร คือ ร้อยละของจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามในรายการผลลัพธ์การค้นคืนตั้งแต่ลำดับที่หนึ่งถึงสิบ เพื่อนำมาใช้ในการประเมินประสิทธิภาพการทำงานของระบบการค้นคืนสารสนเทศ
- 2) ค่าความแม่นยำจำนวนยี่สิบรายการแรก (Precision at 20 documents) โดยวัดจากความสามารถในการค้นคืนเอกสาร คือ ร้อยละของจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามในรายการผลลัพธ์การค้นคืนตั้งแต่ลำดับที่หนึ่งถึงยี่สิบ เพื่อนำมาใช้ในการประเมินประสิทธิภาพการทำงานของระบบการค้นคืนสารสนเทศ
- 3) ค่าความแม่นยำจำนวนสามสิบรายการแรก (Precision at 30 documents) โดยวัดจากความสามารถในการค้นคืนเอกสาร คือ ร้อยละของจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามในรายการผลลัพธ์การค้นคืนตั้งแต่ลำดับที่หนึ่งถึงสามสิบ เพื่อนำมาใช้ในการประเมินประสิทธิภาพการทำงานของระบบการค้นคืนสารสนเทศ
- 4) ค่ามาตรวัดเอฟ (F-measure) คือค่าเฉลี่ยฮาร์โมนิระหว่างค่าระลอกี้ได้ และค่าความแม่นยำจะใช้ค่าระลอกี้ได้และค่าความแม่นยำของรายการผลลัพธ์ทั้งหมด เพื่อนำมาใช้ในการประเมินประสิทธิภาพการทำงานของระบบการค้นคืนสารสนเทศ

1.5. ข้อจำกัดการวิจัย

1.5.1. ผู้ประเมินความเกี่ยวข้องและรายการเอกสารที่นำมาใช้ในการสร้างชุดทดสอบนั้นมีจำนวนน้อย เมื่อเทียบกับงานวิจัยการค้นคืนสารสนเทศในอดีต รวมทั้งการขยายข้อสอบถามและการสร้างชุดทดสอบ ซึ่งอาจจะส่งผลกระทบต่อผลลัพธ์ที่ได้จากการประเมินระบบการค้นคืนสารสนเทศ เมื่อนำระบบการขยายข้อสอบถามที่พัฒนามาประยุกต์ใช้ (องอาจ นัยพัฒน์, 2554)

1.5.2. ในงานวิจัยนี้ เนื่องจากต้องควบคุมปัจจัยอื่นที่ส่งผลต่อการกำหนดข้อสอบถาม อาทิ การประเมินความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร การกำหนดรูปแบบการจัดกลุ่ม การกำหนดค่าสำหรับการนำมาใช้ในการขยายข้อสอบถาม จึงจำกัดทดสอบในห้องปฏิบัติการ (Laboratory) (สุทธิติ ชัตติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554)

1.5.3. เนื่องด้วยงานวิจัยนี้ เป็นงานวิจัยเชิงทดลอง จึงเลือกใช้กลุ่มตัวอย่างด้วยการสุ่มมาจากประชากร จึงอาจเกิดความลำเอียงในการเลือกตัวอย่าง ดังนั้น ผู้วิจัยต้องกำหนดและควบคุมปัจจัยต่างๆ ที่ส่งผลต่อการเข้าร่วมการทดลอง อาทิ การกำหนดคุณสมบัติของหน่วยตัวอย่าง การเลือกช่วงเวลาการเก็บข้อมูล เพื่อให้กลุ่มตัวอย่างมีความเท่าเทียมกันในการเข้าร่วมทดลอง (สุทธิ ชาติติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554)

1.6. ประโยชน์ที่คาดว่าจะได้รับ

1.6.1. เพื่อทราบถึงประสิทธิผลของการนำการจัดกลุ่มมาประยุกต์กับการขยายคำสอบถามในการค้นคืนสารสนเทศภาษาไทย

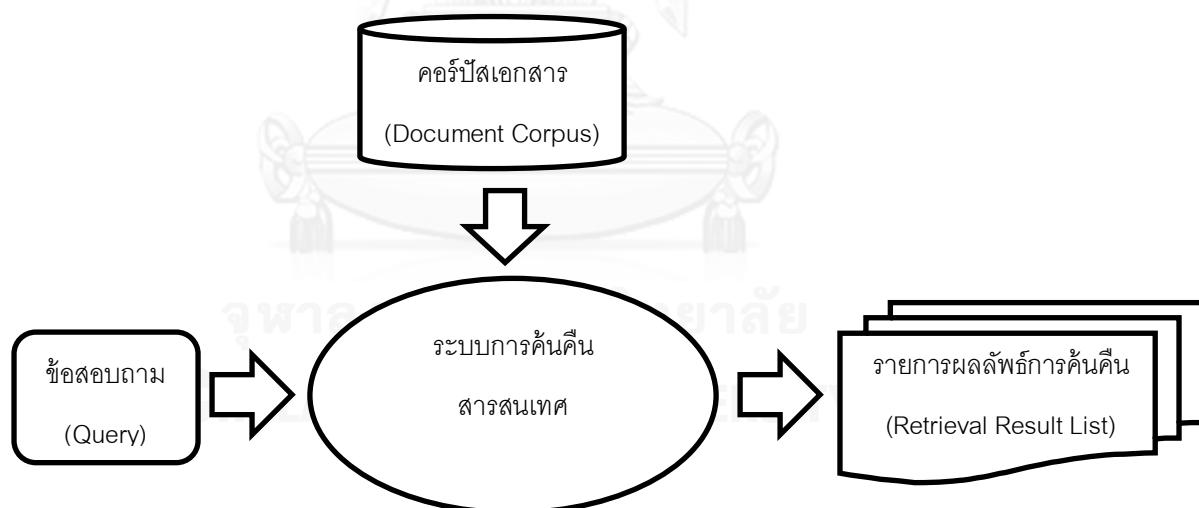
1.6.2. เพื่อนำไปใช้ในแนวทางการพัฒนาขั้นตอนวิธีการขยายข้อสอบถามภาษาไทยในการค้นคืนสารสนเทศภาษาไทย

1.6.3. เพื่อนำไปใช้ในแนวทางการพัฒนาขั้นตอนการค้นคืนสารสนเทศภาษาไทย

บทที่ 2 การทบทวนวรรณกรรม

2.1. การค้นคืนสารสนเทศ (Information Retrieval)

การค้นคืนสารสนเทศ (Information Retrieval, IR) เป็นศาสตร์ของการค้นหาข้อความหรือสารสนเทศในคอร์ปัสเอกสารต่างๆ เพื่อให้ได้เอกสารตามความต้องการของผู้ใช้เองด้วยเหตุนี้จึงมีงานวิจัยจำนวนมากที่เกี่ยวข้องกับการค้นคืนสารสนเทศ ที่ต้องการให้ได้รายการผลลัพธ์การค้นคืน (Retrieval Result List) ให้เป็นที่พึงพอใจต่อผู้ใช้งานระบบการค้นคืนสารสนเทศ มีขั้นตอนการทำงานในรูปแบบพื้นฐานทั่วไป ดังที่แสดงในรูปที่ 2.1 โดยมีรายละเอียดขั้นตอนการทำงานดังต่อไปนี้ ผู้ใช้จะกำหนดข้อสอบถาม (Query) เข้าสู่ระบบการค้นคืน โดยคำที่กำหนดลงไปนั้น จะมีลักษณะแตกต่างกันไปตามหลักภาษาของเอกสารที่ต้องการค้นหา หรือจากคอร์ปัสเอกสาร (Document Corpus) และค้นคืนตามข้อสอบถามที่ได้รับ เมื่อค้นคืนสารสนเทศเป็นที่เรียบร้อยแล้ว ระบบจะแสดงรายการผลลัพธ์การค้นคืน (Retrieved Result List) หรือเอกสารที่ระบบคัดเลือกว่ามีความเกี่ยวข้องกับข้อสอบถาม (ศุภชัย ตั้งวงศ์ศานต์, 2551)



รูปที่ 2.1 รูปแบบการสืบคืนข้อมูลสารสนเทศดัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551)

เนื่องจากการค้นคืนสารสนเทศนั้น มีบทบาทต่อชีวิตประจำวันของผู้คนเป็นจำนวนมากอันเนื่องมาจากความต้องการข้อมูลของผู้คน จะเห็นได้ว่า หนึ่งในกระบวนการค้นคืนสารสนเทศที่ได้รับความนิยม

มากที่สุด คือโปรแกรมค้นหา (Search Engine) เนื่องจากการค้นคืนสารสนเทศผ่านระบบเครือข่ายอินเทอร์เน็ต อีกทั้งสามารถใช้งานได้อย่างง่าย รวดเร็ว และสามารถทำความเข้าใจขั้นตอนการใช้งานได้อย่างรวดเร็ว ในปี พ.ศ.2553 กิจกรรมที่เกิดขึ้นมากที่สุด คือการใช้ไปรษณีย์อิเล็กทรอนิกส์ (E-mail หรือ Electronic Mail) และการค้นหาข้อมูล (Information Search) ซึ่งมีอัตราการใช้นี้ถึงร้อยละ 26.1 (สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2553) ในปี 2555 กิจกรรมการใช้อินเทอร์เน็ตที่เกิดขึ้นมากที่สุด คือการค้นหาข้อมูลต่างๆ อาทิ ข้อมูลจากภาครัฐ ร้อยละ 33.7 ข้อมูลสินค้าและการบริการ ร้อยละ 43.9 (สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร, 2556)

แต่อย่างไรก็ตามยังพบปัญหาการค้นคืนสารสนเทศ คือการไม่พบเอกสารหรือข้อมูลที่ไม่ตรงกับความต้องการของผู้ใช้งาน สาเหตุหลักของปัญหาที่เกิดขึ้นคือ(1) ผู้ใช้ไม่สามารถสะกดคำได้อย่างถูกต้อง Willson และ Given (2008, 2010)ศึกษาเกี่ยวกับการสะกดคำผิดด้วยการเปรียบเทียบความสามารถในการใช้งานของผู้ใช้และลักษณะของเนื้อหาที่ต้องการค้นคืน Rush (2002) ศึกษาความผิดพลาดในการสะกดคำเนื่องจากความกำกวมของคำศัพท์ แต่อย่างไรก็ตามก็มีงานวิจัยที่มีการพัฒนาระบบเพื่อป้องกันการเกิดปัญหาข้างต้น ตัวอย่างเช่นในงานวิจัยของ รัตนา ชุมแก้ว และสุพจน์ นิตยสุวัฒน์ (2549) ศึกษาภาษาบาลีอักษรไทยที่เป็นภาษาที่ใช้ในเฉพาะด้านพระคาถาธรรมบทและเป็นภาษาที่ผู้ใช้ทั่วไปไม่ได้นำมาใช้ในชีวิตประจำวัน จึงอาจเกิดความผิดพลาดในการสะกดคำศัพท์ ผู้วิจัยจึงได้พัฒนาระบบค้นคืนพระคาถาธรรมบทซึ่งมีระบบแนะนำการสะกดคำบาลีอักษรไทย (2)การเลือกใช้คำศัพท์ในการค้นคืนที่เหมาะสมต่อความต้องการของตัวผู้ใช้ เพราะเนื่องจากที่มาและการใช้ภาษาของผู้กำหนดข้อมูลสารสนเทศแตกต่างกัน โดยเฉพาะอย่างยิ่ง คำศัพท์ที่มีอยู่ในปัจจุบันจะมีคำพ้องความหมาย(Synonym)หรือคำที่มีรูปแบบที่แตกต่างกันแต่มีความหมายที่เหมือนกัน และในทางกลับกัน ก็มีคำกำกวม (Polysemy) หรือคำศัพท์คำหนึ่งที่สามารถมีได้หลายความหมาย(Carpinetoและคณะ,2012)และ (3)ในข้อสอบถามที่ผู้ใช้กำหนดไม่สามารถนำมาค้นคืนได้ เนื่องจากไม่มีคำที่อยู่ในข้อสอบถามปรากฏอยู่ในเอกสารที่ต้องการค้นคืน หรือมีจำนวนมากหรือน้อยเกินไป จึงไม่สามารถค้นคืนเอกสารได้ตามที่

ต้องการ ถึงแม้ว่า Sen (2009) จะสำรวจแนวโน้มการกำหนดข้อสอบถามและพบว่าจำนวนคำในข้อสอบถามเริ่มเพิ่มขึ้น A Trellian Company (2012) ได้แสดงสถิติขนาดข้อสอบถามในโปรแกรมค้นหา พบว่าผู้ใช้โดยทั่วไปนิยมกำหนดขนาดข้อสอบถาม 1-3 คำอย่างไรก็ตาม มีงานวิจัยที่วิเคราะห์ขนาดของข้อสอบถามที่มีผลต่อประสิทธิภาพการค้นคืนสารสนเทศ Lu และ Keefer (1995) ได้เปรียบเทียบประสิทธิภาพการค้นคืนสารสนเทศระหว่างการขยายข้อสอบถามและการลดข้อสอบถาม พบว่าการขยายข้อสอบถามจะช่วยเพิ่มประสิทธิภาพการค้นคืน แต่ในทางตรงกันข้าม การลดข้อสอบถามจะลดประสิทธิภาพการค้นคืนแทนนอกจากนี้แล้ว ผู้ใช้เองก็ไม่อาจถ่ายทอดความต้องการข้อมูลออกมาเป็นคำศัพท์ที่ใช้ในการสอบถามข้อมูลได้ดี จึงเป็นอุปสรรคในการค้นหาข้อมูลของผู้ใช้อย่างมาก

ผลลัพธ์ที่ระบบได้ค้นคืนออกมานั้นไม่อาจจะบอกได้ว่าเป็นคำตอบในการค้นหาหรือสร้างความพึงพอใจให้กับผู้ใช้ได้ เนื่องจากผู้ใช้เป็นผู้กำหนดข้อสอบถามเอง อีกทั้งเป็นผลสืบเนื่องจากสาเหตุข้างต้น ผู้ใช้อาจจะไม่สามารถทำความเข้าใจกับเอกสารที่เป็นผลลัพธ์และสรุปได้ว่าตรงกับความต้องการของตนเองหรือไม่ เมื่อผู้ใช้ตัดสินใจว่าเอกสารชิ้นนี้ไม่ตรงกับความต้องการของตนเอง ผู้ใช้จะเรียกดูผลลัพธ์ที่เป็นเอกสารอื่นๆ แล้วท้ายที่สุด หากผู้ใช้ไม่สามารถพบเอกสารที่ตรงกับความต้องการของตนเองได้ตามผลลัพธ์ที่ปรากฏ ผู้ใช้จะทำการปรับเปลี่ยนแบบสอบถามใหม่ (Query Modification) เพื่อให้ระบบทำการค้นหาเอกสารใหม่

2.2. รูปแบบและขั้นตอนการค้นคืนสารสนเทศ

รูปแบบสำหรับการค้นคืนสารสนเทศ เป็นรูปแบบที่แสดงในทรรคนะเชิงตรรกะ (Logical View) เพื่อจำลองรูปแบบที่ใช้ในการค้นคืนข้อมูลสารสนเทศ ลักษณะโดยทั่วไปจะเป็นแผนภาพแสดงสัญลักษณ์ต่างๆ รวมทั้งอาจจะเขียนในรูปแบบสัญลักษณ์หรือสมการคณิตศาสตร์รูปแบบสำหรับการค้นคืนสารสนเทศ ได้กำเนิดจากหลักคณิตศาสตร์โดยทำให้เกิดรูปแบบการค้นคืนเบื้องต้น 3 รูปแบบ ดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551)

1) รูปแบบตรรกะเชิงแบบแผน (Classical Boolean Model, CBM) รูปแบบนี้เกิดจากการนำทฤษฎีเซต (Set Theory) มาใช้ในการสร้างเซตของคำในเอกสารและข้อสอบถาม จากนั้นจึงเป็นการเปรียบเทียบคำในข้อสอบถามในเซตของคำในเอกสาร โดยผลลัพธ์ของเอกสารจะเป็นจริงก็ต่อเมื่อทุกคำในข้อสอบถามปรากฏในเซตของคำในเอกสาร ผลลัพธ์จึงมีเพียงใช่หรือไม่เท่านั้น นอกจากนี้ สามารถกำหนดตัวดำเนินการเชิงตรรกะ (Boolean Operator) ได้แก่ OR, AND และ NOT มาร่วมใช้ในการประมวลผลด้วย

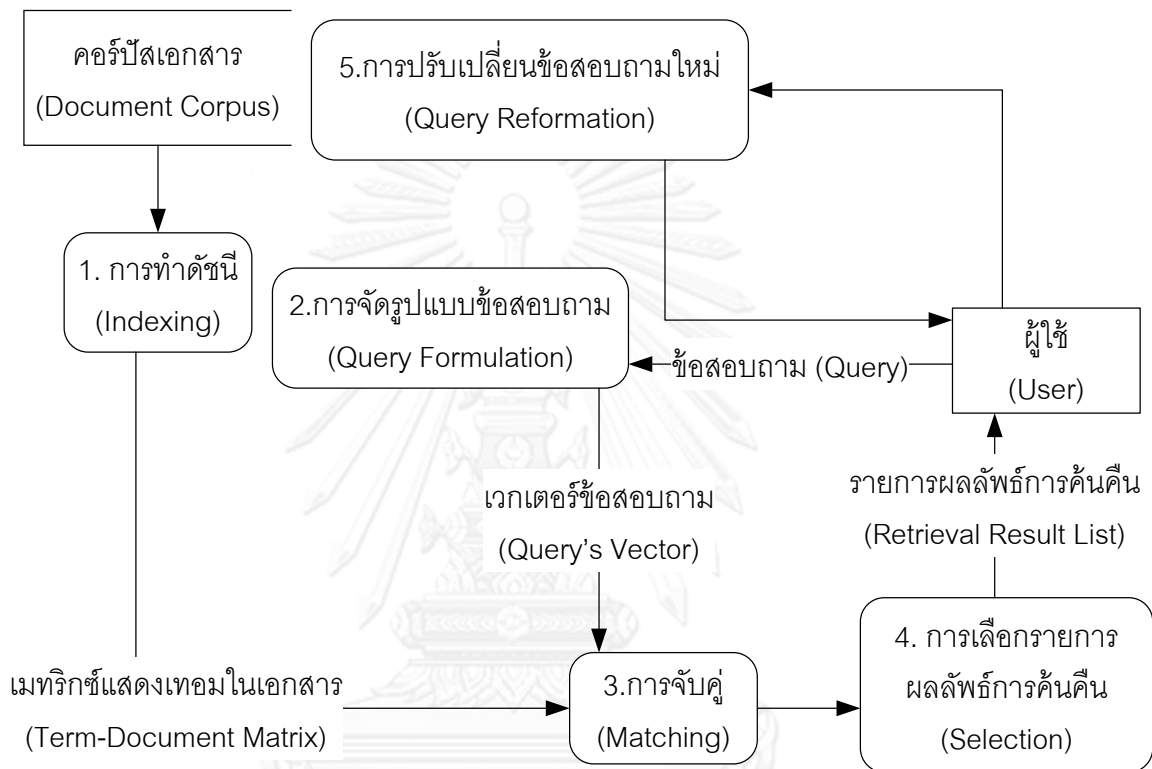
2) รูปแบบปริภูมิเวกเตอร์ (Vector Space Model, VSM) รูปแบบนี้จะมีความแตกต่างจากรูปแบบตรรกะเชิงแบบแผนเพราะสามารถกำหนดผลลัพธ์ที่เป็นค่าการเปรียบเทียบระหว่างเอกสารและข้อสอบถามได้ และนำไปใช้ในการจัดลำดับความสำคัญ โดยนำหลักการของพีชคณิตเชิงเมทริกซ์ (Matrix Algebra) มาใช้ในการสร้างเมทริกซ์คำ-เอกสาร (Term-Document Matrix)

3) รูปแบบความเกี่ยวพันเชิงความน่าจะเป็น (Probabilistic Relevance Model, PRM) จะนำทฤษฎีความน่าจะเป็น (Probability Theory) มาใช้ โดยการคำนวณอัตราส่วนระหว่างความน่าจะเป็นที่เอกสารมีความเกี่ยวข้องกับข้อสอบถามเทียบกับความน่าจะเป็นที่เอกสารไม่มีความเกี่ยวข้องกับข้อสอบถาม

นอกจากรูปแบบของการค้นคืนสารสนเทศแล้ว ยังมีขั้นตอนการประมวลผลของการค้นคืนสารสนเทศ มีขั้นตอนหลัก 5 ขั้นตอน ได้แก่ (1) การทำดัชนี (Indexing) เพื่อสร้างตัวแทนของเอกสาร (2) การจัดรูปแบบข้อสอบถาม (Query Formulation) เพื่อสร้างตัวแทนของข้อสอบถาม (3) การจับคู่ (Matching) ระหว่างตัวแทนเอกสารและตัวแทนข้อสอบถาม (4) การเลือก (Selection) ผลลัพธ์การค้นคืนสารสนเทศ และ (5) การปรับเปลี่ยนข้อสอบถามใหม่ (Query Reformulation) สำหรับการค้นคืนสารสนเทศในครั้งต่อไป (ศุภชัย ตั้งวงศ์ศานต์, 2551)

ในงานวิจัยการขยายข้อสอบถามส่วนใหญ่จะนิยมใช้รูปแบบปริภูมิเวกเตอร์ (Vector Space Model) เป็นรูปแบบหลักของการค้นคืนสารสนเทศที่นำมาใช้ในการขยายข้อสอบถาม เนื่องจากรูปแบบนี้สามารถกำหนดค่าน้ำหนักของแต่ละคำในเอกสารได้ จึงสามารถนำมาประยุกต์ใช้ในการขยายข้อสอบถามได้ เพื่อประเมินความแตกต่างระหว่างเอกสารหรือคำที่มีอยู่ในคอร์ปัสเอกสารได้ นอกจากนี้แล้วรูปแบบนี้จะไม่มีการนำการตัดสินใจจากผู้ใช้งานมาร่วมใช้ในขั้นตอนการค้นคืนสารสนเทศด้วย ผู้วิจัยจึงเลือกใช้รูปแบบปริภูมิเวกเตอร์เป็นรูปแบบหลัก รูปแบบปริภูมิ

เวกเตอร์จะมีขั้นตอนการประมวลผลดังรูปที่ 2.2 ซึ่งมีรายละเอียดดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551)



รูปที่ 2.2 แสดงขั้นตอนการประมวลผลการค้นคืนสารสนเทศแบบปริภูมิเวกเตอร์ ดัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551)

1) การทำดัชนีเพื่อสร้างตัวแทนเอกสาร แต่ละหน่วยของเมทริกซ์จะเป็นค่าน้ำหนักที่ขึ้นอยู่กับความถี่ของคำที่ปรากฏอยู่ในเอกสาร (Term Frequency, TF) และอัตราส่วนของจำนวนเอกสารทั้งหมดต่อจำนวนเอกสารที่มีค่านั้นปรากฏ หรือที่เรียกว่าค่าผกผันความถี่ที่คำปรากฏ (Inverse Document Frequency, IDF) นอกจากนี้ ข้อสอบถามจะกำหนดเป็นเวกเตอร์ข้อสอบถาม โดยให้สอดคล้องกับแนวตั้งของ Term-Document Matrix และค่าของเวกเตอร์ข้อสอบถามจะกำหนดตามความสำคัญในการค้นคืน โดยที่จะกำหนดให้ 0 แทนการไม่ปรากฏของค่านั้น

2) กำหนดให้ d_j แทนเวกเตอร์ของเอกสารลำดับที่ j ใน D และ $w_{i,j}$ แทนค่าน้ำหนักของคำที่ i ในเอกสาร j จึงสามารถเขียน d_j ในรูปแบบของเวกเตอร์เอกสารได้ดังนี้

$$d_j = [w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{i,j}] \text{ โดยที่ } w_{i,j} = tf_{i,j} \times idf_i$$

โดยที่ $tf_{i,j}$ แทนความถี่ของคำ i ที่ปรากฏในเอกสาร j ในขณะที่ idf_i แทนค่าผกผันของความถี่ของคำที่ i ปรากฏในการกำหนดค่า $tf_{i,j}$ เนื่องจากแต่ละคำที่ปรากฏในเอกสารนั้นจะมีความถี่ที่แตกต่างกัน จึงอาจจะเกิดความเลื่อมล้ำได้ เมื่อนำไปใช้ในการคำนวณ อาทิ คำหนึ่งมีความถี่ 10 ในขณะที่อีกคำหนึ่งมีความถี่ 1,000 เพื่อลดปัญหาดังกล่าว จึงควรทำรูปแบบการสร้างบรรทัดฐาน (Normalized Form) ดังนี้

$$tf_{i,j} = \frac{f_{i,j}}{F_j}$$

โดยที่ $f_{i,j}$ แทนความถี่ของคำ i ในเอกสาร j ในขณะที่ F_j แทนความถี่ของคำหนึ่งในเอกสาร j ที่มีค่าสูงที่สุดในเอกสาร (ในบางกรณี F_j อาจจะใช้ค่าเฉลี่ยของคำในเอกสารแทน)

ส่วนการกำหนดค่า idf_i นี้เนื่องจากอัตราส่วนของจำนวนเอกสารทั้งหมด (N) ต่อจำนวนเอกสารที่ปรากฏคำ i อยู่ (n_i) อาจมีค่าที่สูงหรือจนเกินไป ซึ่งอาจจะส่งผลกระทบต่อค่าคำนวณ $w_{i,j}$ ได้ ซึ่งนำฟังก์ชันลอการิทึม (Logarithm) มาใช้เพื่อกำหนดไม่ให้ค่าที่ได้นั้นสูงจนเกินไป ดังนั้น จึงคำนวณหา idf_i ได้จาก

$$idf_i = \log_a \frac{N}{n_i}$$

ในงานวิจัยการค้นคืนข้อมูลสารสนเทศนั้น สมการการคำนวณค่าน้ำหนักจะมีความแตกต่างกันไป สมการที่ได้รับความนิยม เป็นสมการของระบบการค้นคืนสารสนเทศสมาร์ท (System for the Mechanical Analysis and Retrieval of Text, SMART) แต่ในงานวิจัยบางส่วนก็ได้ศึกษา วิเคราะห์หรือสร้างสมการการคำนวณค่าน้ำหนักในรูปแบบต่างๆ กัน อาทิ Horng และ

Chen (2003) ศึกษาขั้นตอนการกำหนดค่าน้ำหนักในแต่ละคำของการค้นคืนสารสนเทศแบบคลุมเคลือ

หลังจากการคำนวณหาค่าน้ำหนักของคำในเอกสารทั้งหมด จะได้เมทริกซ์แสดงความสัมพันธ์ระหว่างคำและเอกสารดังต่อไปนี้

$$M = \begin{bmatrix} w_{1,1} & \cdots & w_{t,1} \\ \vdots & \ddots & \vdots \\ w_{1,d} & \cdots & w_{t,d} \end{bmatrix}_{t \times d}$$

โดยที่ t แทนจำนวนคำทั้งหมดที่มีอยู่ในเอกสาร

d แทนจำนวนเอกสารทั้งหมดที่มีอยู่ในคอร์ปัสเอกสาร

M คือเมทริกซ์แสดงความสัมพันธ์ระหว่างคำและเอกสาร ซึ่งสร้างจากรายการเอกสารที่เป็นผลลัพธ์การค้นคืนสารสนเทศ ซึ่งใช้วิธีการสร้างเมทริกซ์จากขั้นตอนการสร้าง

$w_{i,j}$ คือค่าน้ำหนักของคำลำดับที่ i ที่ปรากฏอยู่ในเอกสารลำดับที่ j

3) การจัดรูปแบบข้อสอบถามเพื่อสร้างตัวแทนข้อสอบถาม ในรูปแบบปริภูมิเวกเตอร์ เวกเตอร์ของข้อสอบถามจะถูกกำหนดให้มีจำนวนค่าที่เหมือนกับเวกเตอร์ของเอกสาร และกำหนดให้แต่ละคำมีค่าน้ำหนักขึ้นอยู่กับนักวิจัยกำหนด โดยทั่วไป จะกำหนดให้ 1 แทนคำที่ปรากฏในข้อสอบถาม และ 0 แทนคำที่ไม่ปรากฏในข้อสอบถาม

$$q = [w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{t,q}]$$

งานวิจัยบางส่วนจะมีวิธีในการกำหนดค่าน้ำหนักแตกต่างกันออกไป Horng และคณะ (2005) ได้กำหนดให้นำเอกสารที่เป็นผลลัพธ์ของการค้นคืนด้วยข้อสอบถามมาคำนวณหาค่าน้ำหนักของข้อสอบถาม เพื่อนำไปใช้ในการปรับเปลี่ยนข้อสอบถาม Chen, Lin และ Chang

(2006) ได้ศึกษาการกำหนดค่าน้ำหนักให้เวกเตอร์ของข้อสอบถามด้วยเครือข่ายประสาท (Neural Network)

4) การจับคู่ระหว่างข้อสอบถามและเอกสาร ในรูปแบบปริภูมิเวกเตอร์ การเปรียบเทียบค่าของเวกเตอร์ข้อสอบถามและเอกสารแต่ละชุดจะคำนวณค่าโคไซน์ของมุม θ (cosine) โดยที่ θ จะแทนมุมระหว่างเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสารแต่ละชุด หาก θ มีค่าเป็น 0 องศา ค่าโคไซน์จะมีค่าเป็น 1 และในทางกลับกัน หาก θ มีค่าเป็น 90 องศา ค่าโคไซน์จะมีค่าเป็น 0 โดยที่การคำนวณจะนำหลักการคูณภายใน (Inner Product) ของเวกเตอร์มาใช้ในการคำนวณหาค่าโคไซน์

เมื่อได้เวกเตอร์ของเอกสารและข้อสอบถามแล้ว จะมีสูตรในการคำนวณหาค่าความเหมือน (Similarity measure) ดังต่อไปนี้

$$Sim(q, d_j) = \frac{q \times d_j}{\|q\| \|d_j\|} = \frac{\sum_{i=1}^M (w_{i,q} \times w_{i,j})}{\sqrt{\sum_{i=1}^M w_{i,q}^2} \sqrt{\sum_{i=1}^M w_{i,j}^2}}$$

โดยที่ M แทนจำนวนค่าทั้งหมดในเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถาม ค่าความเหมือนที่ได้ หากมีค่าใกล้หรือเท่ากับ 1 แสดงว่า ข้อสอบถามคล้ายคลึงกับเอกสารมาก แต่หากค่าความเหมือนที่ได้มีค่าใกล้หรือเท่ากับ 0 แสดงว่า ข้อสอบถามไม่คล้ายคลึงกับเอกสาร

5) การเลือกผลลัพธ์การค้นคืนสารสนเทศ ในขั้นตอนนี้ ระบบจะแสดงรายการสารสนเทศซึ่งเป็นผลลัพธ์การค้นคืนสารสนเทศ เพื่อให้ผู้ใช้ได้เลือกสารสนเทศที่ตรงกับความต้องการการค้นคืน ในบางระบบผู้พัฒนาได้ทำการจัดลำดับ (Ranking) การจำแนก (Classification) และการจัดกลุ่ม (Clustering) สารสนเทศที่เป็นผลลัพธ์ของการค้นคืนด้วยค่าความเหมือน เพื่ออำนวยความสะดวกในการเลือกสารสนเทศ Lu, Kim และ Wilbur (2009) ได้พัฒนากลยุทธ์การจัดลำดับความเกี่ยวข้องในระบบการวิเคราะห์และค้นคืนข้อมูลทางการแพทย์ (Medical Literature Analysis and Retrieval System, MEDLINE) โดยใช้ลำดับเวลาเชิงโครโนโลจิกเคล็ดค้นกลับ ค่าความถี่คำและค่าถ่วงน้ำหนักและการเกิดเหตุการณ์ระดับประโยชน์ร่วมกัน Carpineto และคณะ (2009) ได้ศึกษาสำรวจการนำการจัดกลุ่มมาประยุกต์ใช้ในการแสดงรายการผลลัพธ์การปรับข้อสอบถามใหม่ ดังรายละเอียดที่กล่าวในหัวข้อที่ 2.1 เมื่อผู้ใช้ต้องการค้นคืน

สารสนเทศและยังไม่พบสารสนเทศที่ต้องการ ผู้ใช้จะปรับข้อสอบถามใหม่ เพื่อค้นคืนสารสนเทศที่ต้องการ ในบางระบบการค้นคืนสารสนเทศ ผู้พัฒนาได้แทรกระบบการขยายข้อสอบถาม เพื่อให้ผู้ใช้ได้ปรับข้อสอบถามที่ตรงกับขอบเขตของสารสนเทศ

2.3. การขยายข้อสอบถาม (Query Expansion)

การขยายข้อสอบถาม (Query Expansion) หมายถึงการกำหนดคำเพิ่มเติมลงในข้อสอบถามเดิมที่ผู้ค้นคืนได้กำหนดครั้งแรก เพื่อให้ผู้ใช้ได้เลือกข้อสอบถามใหม่ที่นำไปสู่การค้นหาค้นหาใหม่ และผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้ได้ ทำให้ผู้ใช้ได้รับความสะดวกมากยิ่งขึ้นในการค้นหา (ศุภชัย ตั้งวงศ์ศานต์, 2551; Bernhard, 2010; Campos Fernandez, และ Huete, 1998; Alipanah, Khan and Thuraisingham, 2011; Shiri and Revie, 2006) วัดประสิทธิภาพการขยายข้อสอบถามคือ ประสิทธิภาพการค้นคืน (Retrieval Effectiveness) ที่เพิ่มขึ้น ด้วยข้อสอบถามที่ได้รับการขยายแล้วนำไปใช้ในการค้นคืน

มีงานวิจัยจำนวนมากที่ศึกษาขั้นตอนวิธีการขยายข้อสอบถาม ซึ่งทำให้ประสิทธิภาพการค้นคืนสารสนเทศเพิ่มขึ้นมีงานวิจัยจำนวนมากที่ศึกษาขั้นตอนวิธีการขยายข้อสอบถาม ซึ่งทำให้ประสิทธิภาพการค้นคืนสารสนเทศเพิ่มขึ้น Ogilvie และ Callan(2001)ทดสอบการนำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศแบบกระจาย ซึ่งหมายถึงการค้นคืนสารสนเทศที่มากกว่าหนึ่งชุดข้อมูล ผลการทดลองพบว่า เมื่อนำการขยายข้อสอบถามมาประยุกต์ใช้ ประสิทธิภาพของระบบการค้นคืนเพิ่มขึ้น Klink และคณะ(2002) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศด้วยผลสะท้อนกลับความสัมพันธ์เทียม (Pseudo Relevance Feedback) กับการเรียนรู้เชิงคำ (Learning Term-based Concepts) พบว่า การขยายข้อสอบถามช่วยให้ประสิทธิภาพการค้นคืนสารสนเทศเพิ่มขึ้น โดยที่ชุดข้อมูลจำนวน 3 ชุด จากชุดข้อมูลทั้งหมด 4 ชุด มีประสิทธิภาพสูงสุดจากการใช้การเรียนรู้เชิงคำ Daiz-Galino และคณะ (2009) ได้ทดสอบการนำการขยายข้อสอบถามประยุกต์ใช้ในระบบการค้นคืนภาพการแพทย์ โดยการนำยูเอ็มแอลและออนโทโลยีมาใช้ในการขยายข้อสอบถาม ผลการทดลอง

พบว่า การใช้ออนโทโลยีจะทำให้ประสิทธิผลการค้นคืนเพิ่มขึ้น เมื่อเทียบกับการใช้ยูเอ็มแอล (Undefined Modeling Language, UML) หรือการไม่ใช้การขยายข้อสอบถาม Liu และ Chu (2007) ได้ศึกษาการนำการขยายข้อสอบถามบนพื้นฐานความรู้ประยุกต์ใช้กับการค้นคืนสารสนเทศทางการแพทย์ ได้พบว่าประสิทธิผลการค้นคืนสารสนเทศนั้นเพิ่มขึ้น เนื่องจากนำข้อมูลในสารสนเทศมาใช้ในการกำหนดค่าในข้อสอบถามเพิ่มเติม Riezler, Liu และ Vasserman (2008) ได้ศึกษาการพัฒนาการขยายข้อสอบถามด้วยการศึกษาความสัมพันธ์ระหว่างข้อสอบถามและบทความสั้นหนึ่ง (Snippet) ของเอกสาร ด้วยรูปแบบภาษาเอ็นแกรม (N-gram language model)

จากการยกตัวอย่างงานวิจัยจะเห็นได้ว่า การขยายข้อสอบถามจึงถือได้ว่าเป็นระบบหนึ่งที่มีบทบาทสำคัญเป็นอย่างมากในการพัฒนาการค้นคืนสารสนเทศอย่างไรก็ตามการขยายข้อสอบถามนั้นจะประสบความสำเร็จหรือไม่ ขึ้นอยู่กับข้อสอบถามที่ระบบได้ขยายแล้ว ดังนั้นการกำหนดขั้นตอนวิธีที่นำมาใช้ในการขยายข้อสอบถามจึงมีความสำคัญเป็นอย่างยิ่งเพื่อหาขั้นตอนวิธีการขยายข้อสอบถามใหม่ๆ ที่มีประสิทธิผลการค้นคืนมากยิ่งขึ้นงานวิจัยนี้จึงมุ่งประเด็นการนำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนสารสนเทศเพื่อเพิ่มประสิทธิผลการค้นคืน

2.3.1. การนำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศ

การขยายข้อสอบถามนั้นจะแตกต่างกันไปตามลักษณะของระบบการค้นคืนสารสนเทศที่ผู้พัฒนาได้นำมาประยุกต์ใช้ เพราะความแตกต่างทางด้านลักษณะขององค์ความรู้และภาษาของผู้กำหนดและรวบรวมข้อมูลจึงมีงานวิจัยจำนวนมากที่ได้ศึกษาการนำการขยายข้อสอบถามมาใช้ในระบบการค้นคืนสารสนเทศที่มีรูปแบบต่างๆ กันไป โดยประเด็นที่นักวิจัยสนใจได้แก่ ขอบเขตของสารสนเทศในคอร์ปัส ภาษาที่นำมาใช้ในการสร้างสารสนเทศ และโครงสร้างและขั้นตอนของการค้นคืนสารสนเทศ

(1) **ขอบเขตของสารสนเทศในคอร์ปัส** เนื่องจากการใช้ค่าในการกำหนดสารสนเทศแตกต่างกัน ซึ่งจะมีผลต่อการกำหนดค่านำหนักของคำที่จะนำไปใช้ในการวิเคราะห์การ

ค้นคืนสารสนเทศรวมทั้งการขยายข้อสอบถาม นอกจากนี้เพราะผู้ใช้อาจจะไม่กำหนดข้อสอบถามได้ตรงกับความต้องการการค้นคืน การขยายข้อสอบถามจึงนำมาใช้ในการเพิ่มทางเลือกให้ผู้ใช้สามารถระบุข้อสอบถามลงไปได้ Daiz-Galino และคณะ (2009) ได้ทดสอบการนำการขยายข้อสอบถามประยุกต์ใช้ในระบบการค้นคืนภาพการแพทย์ โดยการนำยูเอ็มแอลและออนโทโลยีมาใช้ในการขยายข้อสอบถาม ผลการทดลองพบว่า การใช้ออนโทโลยีจะทำให้ประสิทธิผลการค้นคืนเพิ่มขึ้น เมื่อเทียบกับการใช้ยูเอ็มแอล(Undefined Modeling Language, UML)หรือการไม่ใช้การขยายข้อสอบถาม Bandyopadhyayและคณะ (2012) ได้พัฒนาระบบการค้นคืนไมโครบล็อก ซึ่งเป็นข้อความแสดงความคิดเห็นที่มีความยาวไม่เกิน 140 ตัวพบว่าประสิทธิผลการค้นคืนเพิ่มขึ้นเมื่อนำการขยายข้อสอบถามมาใช้ ไม่ว่าจะป็นขั้นตอนวิธีความเกี่ยวข้องย้อนกลับ (ซึ่งจะกล่าวในหัวข้อ 2.2.3) หรือการใช้ Google Search API มาใช้Peetz และ Lopatka (2010) ได้นำการขยายข้อสอบถามในระบบการค้นคืนในสารานุกรมออนไลน์ พบว่า ประสิทธิผลการค้นคืนเพิ่มขึ้นเมื่อใช้รูปแบบภาษา (language Modeling) มาวิเคราะห์เอกสารและค้นหาคำที่ปรากฏร่วมกันมาใช้ในการขยายข้อสอบถาม

(2) **ภาษาที่ใช้ในสารสนเทศ** เนื่องจากความแตกต่างในการเลือกใช้ภาษาในการกำหนดข้อมูล ในงานวิจัยมีการแบ่งภาษาออกเป็น 2 ประเภท คือ การเขียนอย่างต่อเนื่องและการเขียนแบบไม่ต่อเนื่อง นอกจากนี้แล้วปัจจัยที่แตกต่างกันระหว่างภาษา ได้แก่ ไวยากรณ์ ระดับการใช้ภาษา จึงทำให้การขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศภาษาต่างๆ จึงแตกต่างกันออกไป

นอกจากนี้แล้วในงานวิจัยบางส่วนนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศเชิงข้ามภาษา (Cross-Language Information Retrieval, CLIR)โดยนำอรรถาภิธาน (Thesaurus)แบบพจนานุกรมสองภาษา (Bilingual Dictionary) มาใช้ในการแปลคำศัพท์จากภาษาหนึ่งเป็นอีกภาษาหนึ่ง แล้วนำคำศัพท์มาใช้ในการขยายข้อสอบถามอีกครั้ง Li, Zu และ Geva (2008) ได้นำการทำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถามภาษาจีน Jaruskulchai (2001) ได้สำรวจการนำการขยายข้อสอบถามมาใช้ในระบบการค้นคืนสารสนเทศภาษาไทยข้ามภาษาด้วยพจนานุกรมChuleerat (2001) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศข้ามภาษา

ในงานวิจัยการขยายข้อสอบถามนี้ ผู้วิจัยสนใจการนำประยุกต์การขยายข้อสอบถามในเอกสารภาษาไทย เนื่องจากเป็นภาษาหนึ่งที่มีลักษณะเป็นการเขียนอย่างต่อเนื่อง ซึ่งแตกต่างจากภาษาอังกฤษที่มีลักษณะเป็นการเขียนอย่างไม่ต่อเนื่อง อีกทั้งภาษาไทยเองก็แตกต่างจาก

ภาษาอังกฤษในด้านของไวยากรณ์และระดับการใช้ภาษา เนื่องจาก (1) ภาษาไทยไม่มีการแยกส่วนระหว่างคำสองคำ ในขณะที่ภาษาอังกฤษจะแยกส่วนระหว่างคำสองคำ (2) ภาษาไทย มีตัวพยางค์ 44 ตัว สระ 32 ตัว และวรรณยุกต์ 4 ตัว ในขณะที่ภาษาอังกฤษมีตัวพยางค์ 21 ตัว และสระ 5 ตัว รวมทั้งมีตัวอักษรใหญ่และเล็ก (3) ภาษาไทยไม่มีการเปลี่ยนรูปแบบคำศัพท์ ในขณะที่ภาษาอังกฤษจะมีการเปลี่ยนรูปแบบคำศัพท์ในหลากหลายรูปแบบ อาทิ เทนส์ (tense) (4) ภาษาไทยมีคำพ้องเสียง ซึ่งเป็นคำที่ออกเสียงเหมือนกัน แต่มีรูปแบบการเขียนที่แตกต่างกัน และมีความหมายที่ต่างกัน (5) คำหนึ่งคำสามารถทำหน้าที่ได้หลายชนิด อาทิ ที่ สามารถเป็นคำบุพบทเชื่อมคำกับประโยค และสามารถทำหน้าที่เป็นคำนาม ซึ่งหมายถึงสถานที่ได้ (6) ในบางกรณี ภาษาไทยสามารถละประธานในประโยคได้ และ (7) การสร้างคำประสมในภาษาไทยจะใช้คำนามและคำกริยาและให้กำเนิดคำประสมซึ่งเป็นคำนาม (Khankasikam, 2010) นอกจากนี้แล้ว ภาษาอังกฤษเป็นภาษาที่นิยมนำมาใช้ในการวิจัยการค้นคว้าสารสนเทศ เนื่องจากขั้นตอนก่อนการประมวลผล (Pre-processing) ในสารสนเทศภาษาอังกฤษจะมีปัจจัยที่น้อยกว่าภาษาไทย

อย่างไรก็ตาม การค้นคว้าสารสนเทศเริ่มมีบทบาทต่อผู้ใช้อินเทอร์เน็ตเป็นอย่างมาก ดังที่กล่าวในหัวข้อที่ 2.1 อีกทั้งขอบเขตของสารสนเทศภาษาไทยที่ต้องการค้นคว้านั้นเริ่มกว้างขวางมากยิ่งขึ้น จึงมีการนำการขยายข้อสอบถามภาษาไทยมาประยุกต์ใช้กันมากยิ่งขึ้น Tantisripreecha และ Soonthornphisaj (2010) ได้นำการขยายข้อสอบถามด้วยอนโทโลยีมาประยุกต์ใช้ในระบบการค้นคว้าพิพาทศาสตร์ฎีกาวิชา รศ. ชุมแก้ว และ สุพจน์ นิตย์สุวัฒน์ (2552) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในระบบการค้นคว้าพระคาถาธรรมบท Facundes และ Theva-aksorn (2007) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคว้าสารสนเทศแบบสองภาษา นอกจากงานวิจัยที่กล่าวมาข้างต้น ยังมีงานวิจัยอีกจำนวนหนึ่งที่ได้้นำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคว้าสารสนเทศ

นอกจากนี้ ผู้วิจัยในอดีตสนใจนำพจนานุกรมมาประยุกต์ใช้เพื่อลดปัญหาการสะกดคำผิด Facundes และ Theva-aksorn (2007) นำพจนานุกรมมาประยุกต์ใช้ในการขยายข้อสอบถาม รศ. ชุมแก้ว และ สุพจน์ นิตย์สุวัฒน์ (2552) ได้นำการขยายข้อสอบถามมาช่วยในการแก้ไขปัญหาการสะกดคำในข้อสอบถามผิด เนื่องจากภาษาบาลีอักษรไทยเป็นภาษาที่ใช้พระคาถาธรรมบท ซึ่งเป็น

ภาษาที่ผู้ใช้ทั่วไปไม่ได้ใช้ในชีวิตประจำวัน ประเด็นสุดท้ายคือการนำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถามภาษาไทย Tantisripreecha และ Soonthornphisaj (2010) ได้นำออนโทโลยีมาประยุกต์ใช้ในระบบการค้นคืนคำพิพากษาภาษาไทย และนำขั้นตอนวิธีกลุ่มมด (Ant Colony Algorithm) มาประยุกต์กับออนโทโลยีที่สร้างขึ้นใช้ในการขยายข้อสอบถาม เพื่อให้ได้ข้อสอบถามใหม่ที่ตรงประเด็น Thava-askorn (2003) ได้นำความสัมพันธ์เชิงความหมายคำศัพท์ (Lexical Semantic Relation) ประยุกต์กับพจนานุกรมอิงคอร์ปัส (Corpus-based Dictionary) ในการขยายข้อสอบถาม วาริน นาราวิทย์ และศิริภิญโญ จันทมณี (2554) ได้นำเจเนติกอัลกอริทึม (Genetic Algorithm) มาประยุกต์ใช้ในการขยายข้อสอบถาม

(3) **โครงสร้างของระบบการค้นคืนสารสนเทศ** จากงานวิจัยในอดีตบางส่วน ผู้วิจัยได้คิดค้นทั้งระบบการค้นคืนสารสนเทศและระบบการขยายข้อสอบถามไปพร้อมกัน Boonchom และ Soonthornphisaj (2012) นำออนโทโลยีมาประยุกต์ใช้ในการค้นคืนสารสนเทศและนำขั้นตอนวิธีกลุ่มมด (Ant Colony Algorithm) เพื่อนำมาใช้ในการขยายข้อสอบถาม Horng และคณะ (2005) ใช้การจัดกลุ่มเอกสารมาประยุกต์ใช้ในการค้นคืนสารสนเทศ และนำการอนุมานแบบคลุมเครือ (Fuzzy inference) เพื่อนำมาใช้ในการปรับเปลี่ยนข้อสอบถาม แต่งานวิจัยในอดีตส่วนใหญ่ มักจะใช้รูปแบบสำหรับการค้นคืนสารสนเทศหลัก 3 รูปแบบ

รูปแบบสำหรับการค้นคืนสารสนเทศ เป็นรูปแบบที่แสดงในทฤษฎีเซตเชิงตรรกะ (Logical View) เพื่อจำลองรูปแบบที่ใช้ในการค้นคืนข้อมูลสารสนเทศ ลักษณะโดยทั่วไปจะเป็นแผนภาพแสดงสัญลักษณ์ต่างๆ รวมทั้งอาจจะเขียนในรูปแบบสัญลักษณ์หรือสมการคณิตศาสตร์รูปแบบสำหรับการค้นคืนสารสนเทศ ได้กำเนิดจากหลักคณิตศาสตร์โดยทำให้เกิดรูปแบบการค้นคืนเบื้องต้น 3 รูปแบบ ดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551)

1) รูปแบบบูลีนคลาสสิก (Classical Boolean Model, CBM) รูปแบบนี้เกิดจากการนำทฤษฎีเซต (Set Theory) มาใช้ในการสร้างเซตของคำในเอกสารและข้อสอบถาม จากนั้นจึงเป็นการเปรียบเทียบคำในข้อสอบถามในเซตของคำในเอกสาร โดยผลลัพธ์ของเอกสารจะเป็นจริงก็ต่อเมื่อทุกคำในข้อสอบถามปรากฏในเซตของคำในเอกสาร ผลลัพธ์จึงมีเพียงใช่หรือไม่เท่านั้น นอกจากนี้ สามารถกำหนดตัวดำเนินการบูลีน (Boolean Operator) ได้แก่ OR, AND และ NOT มาร่วมใช้ในการประมวลผลด้วยมีงานวิจัยจำนวนหนึ่งที่ได้้นำการขยายข้อสอบถามนั้นมา

ประยุกต์ใช้ในการค้นคืนสารสนเทศรูปแบบนี้ Kwon และคณะ (1994) ได้นำการสร้างอรรถาภิธานด้วยการกำหนดน้ำหนักสำหรับการขยายข้อสอบถามในรูปแบบการค้นคืนบูลีนขยาย (Extended Boolean Model) Nguyen และคณะ (2008) ได้นำเสนอการขยายข้อสอบถามในการค้นคืนสารสนเทศรูปแบบบูลีนขยายอย่างไรก็ตาม รูปแบบการค้นคืนสารสนเทศนี้ไม่ได้รับความนิยมในปัจจุบัน เนื่องจากประสิทธิภาพการค้นคืนสารสนเทศจึงลดลงเมื่อเทียบกับรูปแบบการค้นคืนสารสนเทศแบบอื่น (Khankasikam, 2010) การขยายข้อสอบถามสำหรับการค้นคืนรูปแบบนี้จึงไม่ได้นำมาศึกษาต่อเพิ่มเติม

2) รูปแบบปริภูมิเวกเตอร์ (Vector Space Model, VSM) รูปแบบนี้จะมีความแตกต่างจากรูปแบบตรรกะเชิงแบบแผนเพราะสามารถกำหนดผลลัพธ์ที่เป็นค่าการเปรียบเทียบระหว่างเอกสารและข้อสอบถามได้ และนำไปใช้ในการจัดลำดับความสำคัญ โดยนำหลักการของพีชคณิตเชิงเมทริกซ์ (Matrix Algebra) มาใช้ในการสร้างเมทริกซ์คำ-เอกสาร (Term-Document Matrix) เพื่อนำมาใช้ในการวิเคราะห์เอกสารเทียบกับข้อสอบถามสามารถดูรายละเอียดรูปแบบการค้นคืนสารสนเทศแบบปริภูมิเวกเตอร์ได้จากภาคผนวก ก รูปแบบนี้เป็นรูปแบบการค้นคืนสารสนเทศหนึ่งที่ได้้นำการขยายข้อสอบถามมาประยุกต์ใช้ Bratsas และคณะ (2007) ได้นำรูปแบบปริภูมิเวกเตอร์เป็นหลักในการพัฒนาการค้นคืนสารสนเทศการแก้ปัญหาทางการแพทย์ร่วมกับบอคโนโทโลยี อีกทั้งยังนำการขยายข้อสอบถามแบบคลุมเครือมาประยุกต์ร่วม SanJuan และคณะ (2007) ได้นำรูปแบบปริภูมิเวกเตอร์มาประยุกต์ร่วมกับการขยายคำในการขยายข้อสอบถามเชิงความหมาย (Semantic Query Expansion) Tao, Gao และ Zhang (2011) ได้ศึกษาการนำการขยายข้อสอบถามเชิงความหมายมาประยุกต์ใช้ในการค้นคืนสารสนเทศด้วยรูปแบบปริภูมิเวกเตอร์ Rahman และ Bhattacharya (2009) ได้นำเสนอการค้นคืนภาพด้วยรูปแบบดังกล่าว โดยนำการขยายข้อสอบถามมาประยุกต์ใช้ร่วม นอกจากการนำการขยายข้อสอบถามมาประยุกต์ใช้แล้ว Xhao, Wang และ Liu (2004) ยังนำวิธีการค้นคืนรูปแบบนี้มาประยุกต์ใช้ในการสร้างอรรถาภิธานสำหรับการขยายข้อสอบถามอีกด้วย Efron (2008) ได้นำศึกษาการนำการขยายข้อสอบถามและการลดมิติ (Dimensionality Reduction) ในรูปแบบการค้นคืนนี้ด้วยการป้อนความเกี่ยวข้องย้อนกลับร็อคชิโอ (Rocchio) และการกำหนดดัชนีเชิงความหมายแฝง (Latent Semantic Indexing) SanJuan และคณะ (2007) ได้กำหนดขั้นตอนการขยายข้อสอบถามที่นำรูปแบบปริภูมิเวกเตอร์ และการแยกคำที่มีหลายคำมาประยุกต์ใช้ Zhao, Wang และ Liu ได้นำรูปแบบปริภูมิเวกเตอร์แบบการเกิดขึ้นมาประยุกต์ในการขยายข้อสอบถาม Campos, Fernandez และ Huete (1998) ได้นำการใช้อรรถาภิธานรูปแบบเครือข่ายของเบย์มาใช้ในการขยายข้อสอบถามสำหรับ

การค้นคืนสารสนเทศ จะเห็นได้ว่ามีงานวิจัยจำนวนมากที่นำการขยายข้อสอบถามด้วยวิธีการต่างๆ มาประยุกต์ใช้กับการค้นคืนสารสนเทศรูปแบบปริภูมิเวกเตอร์ เนื่องจากสามารถกำหนดน้ำหนักของคำในเอกสารได้ รูปแบบนี้จึงได้รับความนิยมเป็นอย่างมากในการนำมาใช้ในการขยายข้อสอบถาม

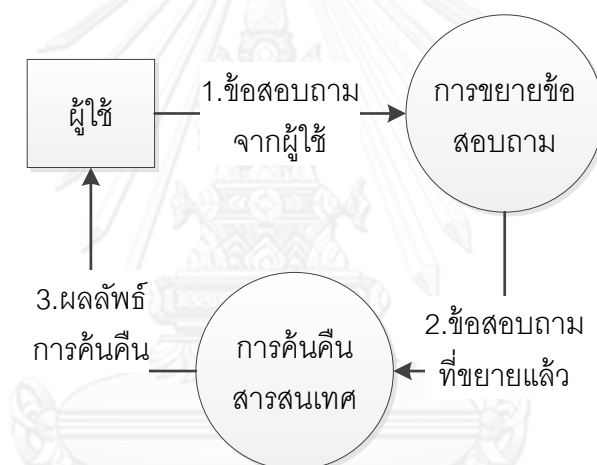
3) รูปแบบความเกี่ยวพันเชิงความน่าจะเป็น (Probabilistic Relevance Model, PRM) จะนำทฤษฎีความน่าจะเป็น (Probability Theory) มาใช้ โดยการคำนวณอัตราส่วนระหว่างความน่าจะเป็นที่เอกสารมีความเกี่ยวพันกับข้อสอบถามเทียบกับความน่าจะเป็นที่เอกสารไม่มีความเกี่ยวพันกับข้อสอบถาม การค้นคืนแบบความเกี่ยวพันเชิงความน่าจะเป็นที่นิยมนำมาศึกษา คือ การค้นคืนแบบเครือข่ายเบย์เอสเสียน (Bayesian Network retrieval) รูปแบบการค้นคืนนี้ได้นำมาใช้ในการศึกษาการประยุกต์กับการขยายข้อสอบถาม Zhao, Wu และ Liu (2012) ได้นำเสนอการค้นคืนแบบเครือข่ายเบย์เอสเสียนพร้อมทั้งการขยายข้อสอบถามด้วยออนโทโลยีโดเมน (Domain Ontology) Campos, Fernandez-Luna และ Huete (2001) ได้นำการปรับความเกี่ยวข้องย้อนกลับมาประยุกต์ร่วมกับการค้นคืนแบบเครือข่ายเบย์เอสเสียน Eguchi (2005) ได้ทดสอบการนำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนนี้ด้วยการนำรูปแบบความเกี่ยวข้องของคำ (Term Dependence Model)

จากที่กล่าวมาข้างต้นนั้น เห็นได้ว่ารูปแบบการค้นคืนสารสนเทศเป็นส่วนหนึ่งในการกำหนดวิธีการขยายข้อสอบถาม เนื่องจากความแตกต่างของวิธีการค้นคืนสารสนเทศ รูปแบบการค้นคืนสารสนเทศที่ได้นิยมเป็นอย่างมากในปัจจุบันจึงมีรูปแบบปริภูมิเวกเตอร์ และรูปแบบความเกี่ยวพันเชิงความน่าจะเป็น จากงานวิจัยการขยายข้อสอบถามในอดีต รูปแบบการค้นคืนสารสนเทศที่นิยมนำการขยายข้อสอบถามมาประยุกต์ใช้ คือรูปแบบปริภูมิเวกเตอร์ นอกจากนี้แล้วผู้วิจัยต้องการลดการนำการตัดสินใจของผู้ใช้ในการค้นคืนสารสนเทศ รูปแบบความเกี่ยวพันเชิงความน่าจะเป็นที่ใช้การตัดสินใจของผู้ใช้จึงไม่สามารถนำมาใช้ในงานวิจัยนี้ได้ ดังนั้นรูปแบบปริภูมิเวกเตอร์ จึงนำมาใช้ในงานวิจัย

2.3.2. รูปแบบการขยายข้อสอบถาม

การขยายข้อสอบถามจะสามารถแบ่งได้สองรูปแบบตามลักษณะการดำเนินงานในระบบการค้นคืนสารสนเทศและการมีส่วนร่วมจากผู้ใช้ระบบคือ (Shapira, Taleb-Maimon และ Nemeth, 2005; Kanaan และคณะ, 2007,2008; Ruthven, 2003)

1) การขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion, Retrieval System-Automatic Query Expansion, AQE) เป็นการขยายข้อสอบถามที่กำหนดข้อสอบถามใหม่เพียงหนึ่งข้อสอบถามและนำไปใช้ในการค้นคืนสารสนเทศต่อไป โดยที่ผู้ใช้ไม่ต้องมีส่วนร่วมในการกำหนดข้อสอบถามใหม่ จะมีขั้นตอนการทำงานดังรูปที่ 2.3 ขั้นตอนการขยายข้อสอบถามรูปแบบนี้จะมีดังนี้ (1) ผู้ใช้จะกำหนดข้อสอบถามเข้าสู่ระบบ (2) ระบบจะขยายข้อสอบถามจากผู้ใช้ ได้ผลลัพธ์เป็นข้อสอบถามที่ขยายแล้วและ (3) นำข้อสอบถามที่ได้ไปใช้ในการค้นคืนสารสนเทศ จากนั้นจะได้ผลลัพธ์การค้นคืนส่งให้ผู้ใช้ การขยายข้อสอบถามรูปแบบนี้เป็นรูปแบบที่นิยมอย่างมากในการขยายข้อสอบถาม (CarpinetoและRomano,2012)

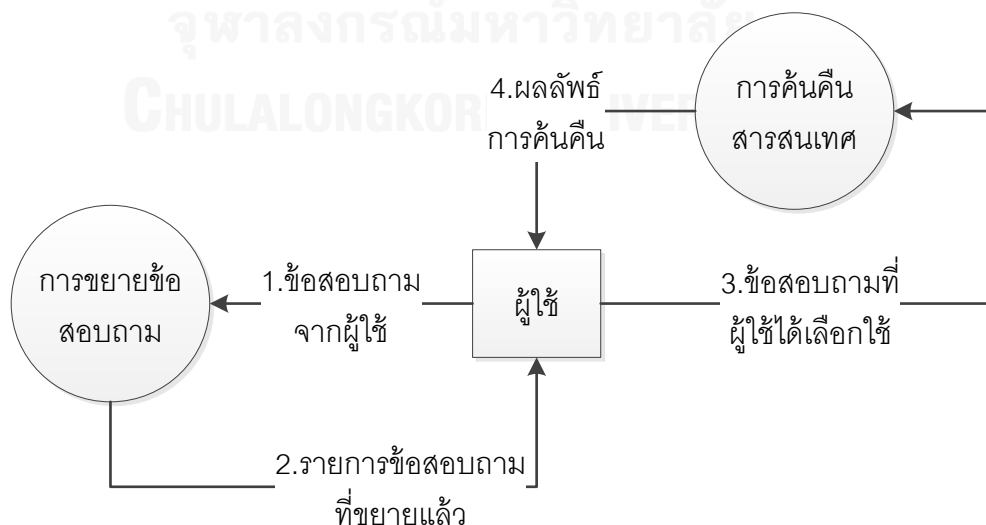


รูปที่ 2.3 รูปแบบการขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion)ดัดแปลงจาก Shapira, Taieb-Maimon และ Nemeth,2005; Kanaan และคณะ, 2007,2008; Ruthven, 2003

จากงานวิจัยในอดีตการขยายข้อสอบถามแบบอัตโนมัติเป็นรูปแบบการขยายข้อสอบถามที่ผู้วิจัยนิยมนำมาใช้ในการศึกษาและพัฒนา เนื่องจากมีจุดประสงค์ในการกำหนดคำหรือคำสำหรับการขยายข้อสอบถาม Chumและคณะ (2007) ได้ศึกษาและพัฒนาการขยายข้อสอบถามแบบอัตโนมัติในระบบการค้นคืนสารสนเทศเชิงวัตถุ ด้วยการวิเคราะห์รูปแบบลักษณะโดยรวม (Generative Feature Model)กับวัตถุเชิงรูปภาพ Greenberg (2001) ได้นำการขยายข้อสอบถามแบบอัตโนมัติมาศึกษาเกี่ยวกับการใช้คำเชิงความหมาย (Lexical Semantic Term)Perez-Aquera

และ Araujo (2008) ได้นำขั้นตอนวิธีทฤษฎีข้อมูล (Information-Theoretic Approach) มาประยุกต์ใช้กับขั้นตอนวิธีของ Rocchio ในการขยายข้อสอบถามแบบอัตโนมัติ Vilares, Vilares และ Alonso (2001) ได้นำการสร้างหน่วยคำจากคำศัพท์หนึ่งคำในการขยายข้อสอบถามแบบอัตโนมัติสำหรับการค้นคืนสารสนเทศภาษาสเปน Yuan, Zhang และ Xing (2007) ได้กำหนดการสร้างขอบเขตศัพท์ทางวิชาการเฉพาะทางและความสัมพันธ์ระหว่างขอบเขต เพื่อนำมากำหนดศัพท์ในการขยายข้อสอบถาม

2) การขยายข้อสอบถามแบบโต้ตอบ (Interactive Query Expansion, User-Interactive Query Expansion, IQE) เป็นการขยายข้อสอบถามที่กำหนดข้อสอบถามใหม่ได้มากกว่าหนึ่งข้อสอบถาม และผู้ใช้มีส่วนร่วมในการเลือกข้อสอบถามใหม่ก่อนที่ระบบจะนำไปใช้ในการค้นคืนสารสนเทศต่อไป ซึ่งการแสดงผลรายการข้อสอบถามที่ขยายแล้วจะสามารถแสดงได้ทั้งก่อนหรือหลังการขยายข้อสอบถาม รูปที่ 2.4 แสดงถึงรูปแบบการขยายข้อสอบถามแบบโต้ตอบ ขั้นตอนการขยายข้อสอบถามรูปแบบนี้มีดังนี้ (1) ผู้ใช้กำหนดข้อสอบถามเข้าสู่ระบบ (2) ระบบจะขยายข้อสอบถามจากผู้ใช้ ออกมาเป็นรายการผลลัพธ์ข้อสอบถามใหม่ โดยมีจำนวนตั้งแต่หนึ่งข้อสอบถามขึ้นไป แล้วนำไปเสนอต่อผู้ใช้ (3) ผู้ใช้จะเลือกหรือกำหนดข้อสอบถามที่จะนำไปใช้สำหรับการค้นคืนสารสนเทศ และ (4) ข้อสอบถามที่ผู้ใช้กำหนดจะส่งไปใช้ในการค้นคืนสารสนเทศ จึงจะได้รายการผลลัพธ์การค้นคืน (Fonseca และคณะ, 2005; Sahib, Tombros, และ Ruthven, 2010; Lee, Lin, และ Huang, 2001)



รูปที่ 2.4 รูปแบบการขยายข้อสอบถามแบบโต้ตอบ (Interactive Query Expansion) ดัดแปลงจาก Shapira และคณะ, 2005; Kanaan และคณะ, 2007, 2008; Ruthven, 2003

มีงานวิจัยจำนวนหนึ่งที่ศึกษาการนำการขยายข้อสอบถามแบบโต้ตอบมาใช้ในการชี้แนะผู้ใช้ในการกำหนดข้อสอบถามสำหรับการค้นคืนสารสนเทศ Kaczmarek (2011) ได้นำขั้นตอนวิธีการจัดกลุ่มด้วยทิศทาง (Clustering-by-directions) มาใช้ในการขยายข้อสอบถามแบบโต้ตอบเพื่อให้ผู้ใช้สามารถกำหนดข้อสอบถามสำหรับการค้นคืนสารสนเทศตามความต้องการของตนเอง Joho, Sanderson และ Beaulieu (2004) ได้นำ CiQuest (Concept-based Interactive Query Expansion Support Tool) ซึ่งเป็นเครื่องมือสำหรับการขยายข้อสอบถามแบบโต้ตอบ ด้วยการสร้างลำดับชั้นเชิงความคิด (Concept Hierarchy) มาใช้ในการประเมินประสิทธิภาพการค้นคืนสารสนเทศและการประเมินการรับรู้ของผู้ใช้ เพื่อศึกษาการโต้ตอบจากผู้ใช้ Bast, Majudar และ Weber (2007) นำการจัดกลุ่มคำที่เกี่ยวข้องมาใช้ในการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศแบบการค้นหาสมบูรณ์ (Complete Search) โดยระบบจะแสดงรายการคำที่เกี่ยวข้องกับคำในข้อสอบถามเมื่อผู้ใช้งานกำลังกำหนดข้อสอบถามเข้าสู่ระบบ เพื่อให้ผู้ใช้ได้เลือกคำที่ตรงกับความต้องการข้อมูลของตนเอง Meiyappan และ Iyengar (2011) ได้นำการขยายข้อสอบถามแบบโต้ตอบมาประยุกต์ใช้วารสารนุกรมออนไลน์ เพื่อเพิ่มความแม่นยำในการค้นคืนสารสนเทศ

จะเห็นได้ว่า การขยายข้อสอบถามแบบอัตโนมัติจะเห็นการกำหนดคำลงในข้อสอบถามเพื่อสร้างข้อสอบถามใหม่แล้วนำไปใช้ในการค้นคืนสารสนเทศทันที ซึ่งผู้ใช้จะไม่มีบทบาทในการขยายข้อสอบถามและไม่ทราบถึงข้อสอบถามใหม่ที่ขยายก่อนนำไปใช้ในการค้นคืนสารสนเทศ แต่ช่วยลดการตัดสินใจของผู้ใช้ ซึ่งจะลดเวลาในการขยายข้อสอบถาม ส่วนการขยายข้อสอบถามแบบโต้ตอบนั้น จะต้องใช้การตัดสินใจจากผู้ใช้ ซึ่งระบบจะต้องนำเสนอรายการผลลัพธ์ข้อสอบถามใหม่ และรับข้อสอบถามใหม่ที่ผู้ใช้ได้เลือกหรือกำหนด เพื่อนำไปใช้ในการค้นคืนสารสนเทศในรูปแบบการขยายข้อสอบถามนี้ ผู้ใช้จะได้เรียนรู้การกำหนดข้อสอบถามที่สามารถนำไปค้นคืนสารสนเทศที่ตรงกับความต้องการของตนเองได้ ในงานวิจัยนี้ เพื่อการกำหนดคำสำหรับข้อสอบถาม จึงลดบทบาทการมีส่วนร่วมหรือการตัดสินใจจากผู้ใช้ในการขยายข้อสอบถาม เพื่อศึกษาถึงประสิทธิภาพการค้นคืนสารสนเทศด้วยข้อสอบถามที่ได้จากการขยายแล้ว อีกทั้งเพื่อลดจำนวนข้อสอบถามจะต้องนำไปใช้ในการค้นคืนสารสนเทศและเวลาที่ใช้ในการขยายข้อสอบถาม

ผู้วิจัยจึงเลือกใช้รูปแบบการขยายข้อสอบถามแบบอัตโนมัติในการกำหนดขั้นตอนการขยายข้อสอบถาม (สุทธิ ชัตติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554)

2.3.3. การวิเคราะห์ข้อมูลในการขยายข้อสอบถาม

จากงานวิจัยในอดีต ขั้นตอนวิธีการขยายข้อสอบถาม จะมีวิธีการวิเคราะห์ข้อมูลสำหรับการกำหนดคำใหม่ในการขยายข้อสอบถามที่แตกต่างกันไป รูปที่ 2.5 แสดงข้อมูลที่นำมาใช้ในการขยายข้อสอบถาม จะเห็นได้ว่า นอกจากข้อสอบถามที่ผู้ใช้ได้กำหนดเข้าสู่ระบบการค้นคืนสารสนเทศแล้ว ยังมีข้อมูลอื่นๆ ที่นำมาใช้ในการกำหนดคำใหม่ที่นำมาใช้ในการขยายข้อสอบถาม ได้แก่ (1) รายการผลลัพธ์การค้นคืนสารสนเทศ (Retrieval Result list) เป็นรายการเอกสารที่ได้จากการค้นคืนสารสนเทศ (2) คอร์ปัสเอกสาร (Document Corpus) เป็นรายการเอกสารทั้งหมดที่มีอยู่ในระบบ (3) ล็อกข้อสอบถาม (Query Logs) เป็นรายการข้อสอบถามที่ผู้ใช้ในอดีตได้กำหนดไว้แล้วระบบบันทึกเก็บไว้ และ (4) อรรถภูมิตาน (Thasaurus) เป็นรายการคำศัพท์ที่ได้จากภายนอกหรือวิเคราะห์จากเอกสารในคอร์ปัสเอกสาร แล้วบันทึกไว้ในระบบ (ศุภชัย ตั้งวงศ์ศานต์, 2551) วิธีการวิเคราะห์ข้อมูลสำหรับการกำหนดคำใหม่ในการขยายข้อสอบถามที่พบในงานวิจัยมีวิธีดังต่อไปนี้

$\sum \bar{D}_{relevant}$ คือผลรวมเวกเตอร์ของเอกสารทั้งหมดที่เกี่ยวข้องกับ
ข้อสอบถาม

$\sum \bar{D}_{non-relevant}$ คือผลรวมเวกเตอร์ของเอกสารทั้งหมดที่ไม่เกี่ยวข้อง
ข้อสอบถาม

a,b,c คือค่าความสำคัญ

งานวิจัยผลป้อนกลับความเกี่ยวข้อง (อาทิ Lv และ Zhai,2009; Ruthven และ Lalmas,2003;Huiskesและ Lew,2008; Belkin และคณะ, 1995; Yinและคณะ, 2007) จะศึกษาถึงขั้นตอนวิธีการที่เกี่ยวข้องกับการกำหนดผลรวมเวกเตอร์ของเอกสารซึ่งนำมาใช้ในการกำหนดค่าน้ำหนักของเวกเตอร์ข้อสอบถามใหม่หรือการนำขั้นตอนวิธีดังกล่าวมาประยุกต์ใช้ ผลการวิจัยพบว่า ประสิทธิภาพการค้นคืนเพิ่มขึ้น

2) **วิธีการวิเคราะห์เฉพาะส่วน (Local Analysis)** จะใช้เอกสารที่เป็นผลลัพธ์มาใช้ในการวิเคราะห์เพื่อนำมาใช้ในการหาคำใหม่มาเพิ่มในข้อสอบถามเดิมเป็นข้อสอบถามใหม่การวิเคราะห์เฉพาะส่วนนั้น จะวิเคราะห์เพื่อหาคำที่ปรากฏร่วมกันในเอกสาร เพื่อนำมาใช้ในการขยายข้อสอบถาม ในรูปแบบการค้นคืนแบบปริภูมิเวกเตอร์ ผู้วิจัยสามารถคำนวณหาค่าน้ำหนักของคู่คำที่ปรากฏร่วมกัน ดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551)

$$M = \begin{bmatrix} w_{1,1} & \cdots & w_{t,1} \\ \vdots & \ddots & \vdots \\ w_{1,d} & \cdots & w_{t,d} \end{bmatrix}_{t \times d}$$

$$M^T = \begin{bmatrix} w_{1,1} & \cdots & w_{1,t} \\ \vdots & \ddots & \vdots \\ w_{d,1} & \cdots & w_{d,t} \end{bmatrix}_{d \times t}$$

$$S = M^T \times M$$

$$S = \begin{bmatrix} l_{1,1} & l_{2,1} & \dots & l_{t,1} \\ l_{1,2} & l_{2,2} & \dots & l_{t,2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1,t} & l_{2,t} & \dots & l_{t,t} \end{bmatrix}_{t \times t}$$

โดยที่ t แทนจำนวนคำทั้งหมดที่มีอยู่ในเอกสาร

d แทนจำนวนเอกสารทั้งหมดที่ปรากฏอยู่ในคอร์ปัสเอกสารหรือรายการเอกสารที่ต้องการวิเคราะห์

M คือเมทริกซ์แสดงค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร ซึ่งสร้างจากรายการเอกสารที่เป็นผลลัพธ์การค้นคืนสารสนเทศ ซึ่งใช้วิธีการสร้างเมทริกซ์จากขั้นตอนการสร้างตัวแทนเอกสารในหัวข้อที่ 2.2.1 ที่ผ่านมา

$w_{i,j}$ คือค่าน้ำหนักของคำลำดับที่ i ที่ปรากฏอยู่ในเอกสารลำดับที่ j

M^T คือเมทริกซ์สลับเปลี่ยน (A transpose of a matrix) ของเมทริกซ์ M กล่าวคือเป็นเมทริกซ์แสดงค่าน้ำหนักของเอกสารที่มีคำปรากฏอยู่

S คือเมทริกซ์ของการปรากฏร่วมกันของคำ (Term Co-occurrence Matrix) ซึ่งจะแสดงถึงค่าน้ำหนักระหว่างคำสองคำในรายการคำทั้งหมดที่ปรากฏอยู่ในคอร์ปัสเอกสาร

$l_{i,j}$ คือค่าน้ำหนักของการปรากฏร่วมกันระหว่างคำลำดับที่ i และคำลำดับที่ j

เมื่อได้เมทริกซ์ของการปรากฏร่วมกันของคำแล้ว ระบบจะวิเคราะห์เพื่อกำหนดคำสำหรับการขยายข้อสอบถาม

มีงานวิจัยบางส่วนที่ได้นำการวิเคราะห์ข้อมูลเฉพาะส่วน Rahman และ Bhattacharya (2009) ได้นำการวิเคราะห์เฉพาะส่วนมาใช้ในการขยายข้อสอบถามบนระบบการค้นคืน

สารสนเทศ ทั้งนี้ผู้วิจัยได้เปรียบเทียบกับขั้นตอนวิธี Rocchio พบว่าประสิทธิผลการค้นคืนสารสนเทศเพิ่มขึ้น และมากกว่าเมื่อเทียบกับการใช้ขั้นตอนวิธี Rocchio งานวิจัยของ Luke, Schaer และ Mayr (2012) ได้ศึกษาการนำการวิเคราะห์เฉพาะส่วนมาประยุกต์ใช้กับระบบการแนะนำคำค้นหาและการกำหนดคำเฉพาะทางในการขยายข้อสอบถามแบบอัตโนมัติ เพื่อให้ได้คำสำหรับการค้นคืนสารสนเทศที่ตรงกับความต้องการของผู้ใช้ Rahman, Antani, และ Thoma (2011) ได้ศึกษาการนำการวิเคราะห์เฉพาะส่วนและการวิเคราะห์ครอบคลุม (ซึ่งจะกล่าวในหัวข้อถัดไป) มาประยุกต์ใช้ในระบบการค้นคืนรูปภาพ Arcoverde และคณะ (2006) ได้นำเอกสารจากรายการผลลัพธ์การค้นคืนมาวิเคราะห์เฉพาะส่วนเพื่อหาวลีหรือกลุ่มคำสำหรับการขยายข้อสอบถาม Yan และ Hauprmann (2007) ได้นำการวิเคราะห์เฉพาะส่วนมาประยุกต์กับการสะท้อนความเกี่ยวข้องกลับเชิงความน่าจะเป็นในการค้นคืนสื่อประสม (Multimedia Retrieval)

3) **วิธีการวิเคราะห์ครอบคลุม (Global Analysis)** จะใช้เอกสารจากคอร์ปัสเอกสารทั้งหมดมาใช้ในการวิเคราะห์เพื่อนำมาใช้ในการหาคำใหม่มาเพิ่มในข้อสอบถามเดิมเป็นข้อสอบถามใหม่ (วิธีการนี้จะแตกต่างจากวิธีการวิเคราะห์เฉพาะส่วนในส่วนที่มาของเอกสารเท่านั้น) ทั้งนี้อาจจะมีการนำอรรถาภิธาน (Thesaurus) มาประกอบการวิเคราะห์ ซึ่งอรรถาภิธานจะสร้างมาจากคอร์ปัสเอกสาร หรือนำมาจากอรรถาภิธานที่มีอยู่จากภายนอก (ศุภชัย ตั้งวงศ์ ศานต์, 2551; Rahman, Antani, และ Thoma, 2011) ในงานวิจัยการค้นคืนสารสนเทศและการขยายข้อสอบถามภาษาอังกฤษ จะนิยมใช้เวิร์คเน็ต (WordNet) ส่วนภาษาไทยนั้น นิยมใช้ไทยเล็กซิตรอน (ThaiLexitron)

มีงานวิจัยที่เสนอการนำการวิเคราะห์ครอบคลุมมาประยุกต์ใช้กับการขยายข้อสอบถามร่วมกับเครื่องมือต่างๆ Zu และ Croft (1996) คิดค้นการวิเคราะห์การวิเคราะห์เฉพาะส่วนและการวิเคราะห์ครอบคลุม ซึ่งทำให้ประสิทธิภาพการค้นคืนสารสนเทศเพิ่มขึ้นจากเดิม นอกจากนี้แล้วยังมีงานวิจัยจำนวนมากที่ศึกษาและเปรียบเทียบขั้นตอนวิธีทั้งสองวิธี Xu และคณะ (2006) ได้ศึกษาเปรียบเทียบการขยายข้อสอบถามด้วยการวิเคราะห์เฉพาะส่วน การวิเคราะห์ครอบคลุมและออนโทโลยี Lei และคณะ (2011) ได้สำรวจการนำการวิเคราะห์เฉพาะส่วนมาประยุกต์ใช้ในการขยายข้อสอบถาม Wang และ Niu ได้นำการวิเคราะห์ครอบคลุมมาประยุกต์กับออนโทโลยีในการขยายข้อสอบถามแบบผสม (Hybrid Query Expansion) แล้วพบว่าประสิทธิผลการค้นคืนเพิ่มขึ้นและ

มากกว่า เมื่อเทียบกับการนำการวิเคราะห์เฉพาะส่วนและการขยายข้อสอบถามลูกผสม Yurekil และคณะ (2009) ได้นำการวิเคราะห์ครอบคลุมมาประยุกต์ใช้กับการเสนอแนวทางการเลือกใช้ข้อสอบถามสำหรับระบบการค้นคืนสารสนเทศในเว็บไซต์การพาณิชย์อิเล็กทรอนิกส์ Dinh และ Tamine (2011) ได้นำการวิเคราะห์เอกสารแบบครอบคลุมในการขยายเอกสาร เพื่อกำหนดคำสำหรับการขยายข้อสอบถาม

4) **วิธีการนำข้อสอบถามมาใช้ (Query re-use)** จะนำข้อสอบถามจากล็อกข้อสอบถามมาวิเคราะห์หาคำใหม่ในข้อสอบถามนอกจากนี้แล้ว อาจมีการนำข้อสอบถามซึ่งระบบได้บันทึกเก็บไว้ในล็อกข้อสอบถามมาใช้ในการวิเคราะห์หาความสัมพันธ์ระหว่างข้อสอบถามและเอกสารจากผู้ใช้ระบบค้นคืนสารสนเทศ Zhu และคณะ (2007) ได้ศึกษาการขยายข้อสอบถามด้วยการทำเหมืองข้อสอบถามและ Cui และคณะ (2002) ได้วิจัยล็อกข้อสอบถามมาประยุกต์ใช้กับการขยายข้อสอบถามเชิงความน่าจะเป็น งานวิจัยทั้งสองพบว่า ได้ประสิทธิผลที่มากกว่าการวิเคราะห์เนื้อหาเฉพาะส่วน (Local Content Analysis, LCA) พบว่า มีประสิทธิผลที่ดีกว่าการวิเคราะห์เนื้อหาเฉพาะส่วน

ในงานวิจัยนี้ ผู้วิจัยสนใจนำขั้นตอนการวิเคราะห์เฉพาะส่วน (Local Analysis) มาประยุกต์ใช้ เนื่องจากการนำผลป้อนกลับความเกี่ยวข้อง (Relevance Feedback) และการนำข้อสอบถามมาใช้ (Query Reuse) มาประยุกต์ใช้ในการขยายข้อสอบถามนั้น อาจเกิดความลำเอียงขึ้น (bias) ได้ด้วยสาเหตุ 2 สาเหตุ คือ (1) ศุภชัย ตั้งวงศ์ศานต์ (2551) ได้สังเกตเห็นว่า เนื่องจากการใช้ขั้นตอนวิธีผลป้อนกลับความเกี่ยวข้องจะทำให้เอกสารจำนวนหนึ่งที่อยู่อันดับต้นในรายการผลลัพธ์การค้นคืนจะได้รับความนิยมเป็นอย่างมาก เปรียบได้กับการให้คะแนนมาก แต่ในทางตรงกันข้าม เอกสารอีกส่วนหนึ่งจะไม่นำมาแสดงรายละเอียด เปรียบได้กับการให้คะแนนน้อย และ (2) การนำข้อสอบถามมาใช้นั้นอาจจะสร้างความสัมพันธ์ระหว่างข้อสอบถามที่ผู้ใช้กำหนดและเอกสารซึ่งเป็นผลลัพธ์การค้นคืนและถูกเรียกเพื่อแสดงรายละเอียดโดยผู้ใช้ ซึ่งล้วนแล้วแต่เป็นการตัดสินใจโดยผู้ใช้ Cui และคณะ (2002) และ Kunpeng, Xiaolong, และ Yuanchao (2006) ได้มีการบันทึกการเข้าชมเว็บไซต์พร้อมทั้งข้อสอบถามที่นำมาใช้ในการค้นคืนและเข้าชมเว็บไซต์นั้น แล้วนำบันทึกดังกล่าวมาใช้ในการวิเคราะห์และสร้างข้อสอบถามใหม่ ดังนั้นเพื่อลดการตัดสินใจจากผู้ใช้ ผู้วิจัยจึงเลือกใช้การวิเคราะห์จากเอกสาร (สุจิตติ ชัตติยะ และ

วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยวัฒน์, 2554)งานวิจัยจำนวนหนึ่งพบว่า การขยายข้อสอบถามด้วยการวิเคราะห์เฉพาะส่วนทำให้ประสิทธิผลการค้นคืนดีกว่าการขยายข้อสอบถามด้วยการวิเคราะห์ครอบคลุม Xu และ Hu (2010) ได้ศึกษาการนำการจัดกลุ่มเอกสารมาประยุกต์ใช้กับผลป้อนกลับความเกี่ยวข้อง โดยจำแนกเอกสารที่มาจัดกลุ่มเป็นเอกสารที่เป็นผลลัพธ์และเอกสารทั้งหมดในคอร์ปัสเอกสาร พบว่า การจัดกลุ่มเอกสารที่เป็นผลลัพธ์จะให้ประสิทธิผลที่มากกว่าเมื่อจัดกลุ่มเอกสารในคอร์ปัสเอกสาร Xu และคณะ (2006) ได้ศึกษาการเปรียบเทียบการนำการวิเคราะห์เฉพาะส่วน การวิเคราะห์ครอบคลุม และออนโทโลยีมาประยุกต์กับการขยายข้อสอบถาม พบว่า การวิเคราะห์เฉพาะส่วนทำให้ประสิทธิผลการค้นคืนดีกว่าการวิเคราะห์ครอบคลุม ดังนั้น ในงานวิจัยนี้จะนำการวิเคราะห์เฉพาะส่วนเป็นรูปแบบหลักในการขยายข้อสอบถาม ซึ่งจะเอกสารที่เป็นผลลัพธ์การค้นคืนมาสร้างเมทริกซ์ของการปรากฏร่วมกันของคำ เพื่อนำมาใช้ในการวิเคราะห์หาคำที่มีลักษณะใกล้เคียงกับคำในข้อสอบถาม

2.3.4. การทำเหมืองข้อมูล (Data Mining) ในการขยายข้อสอบถาม

มีงานวิจัยการขยายข้อสอบถามจำนวนมากที่นำการทำเหมืองข้อมูลมาประยุกต์ใช้ในการวิเคราะห์หาคำสำหรับการขยายข้อสอบถาม ดังนั้น การศึกษาและกำหนดขั้นตอนวิธีสำหรับการนำมาใช้ในการขยายข้อสอบถามนั้น จึงมีความสำคัญอย่างยิ่งต่อการกำหนดขั้นตอนการขยายข้อสอบถาม ในหัวข้อนี้จะอธิบายถึงการทำเหมืองข้อมูล การจัดกลุ่ม และการจัดกลุ่มแบบลำดับชั้น รวมทั้งการกำหนดขั้นตอนวิธีการจัดกลุ่มแบบลำดับชั้น เพื่อสร้างความเข้าใจถึงการนำการจัดกลุ่มแบบลำดับชั้นมาประยุกต์ใช้ในการขยายข้อสอบถาม

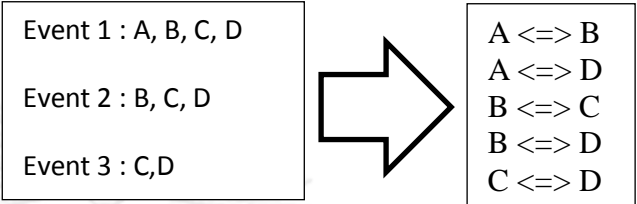
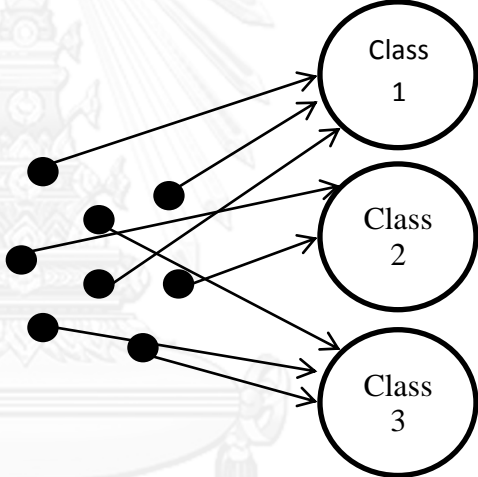
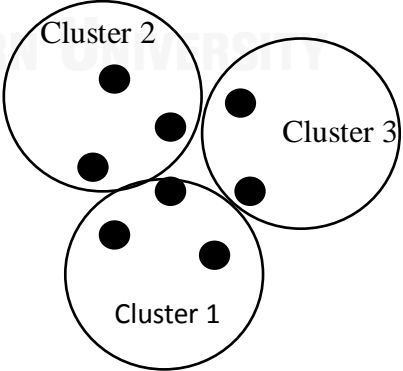
(1) แนวคิดการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการสร้างสารสนเทศใหม่จากข้อมูลที่มีอยู่เดิมด้วยขั้นตอนวิธีต่างๆ (Linoff และ Berry, 2011; Yin, Kaku, Tang, และ Zhu, 2011; Han, Kamber และ Pei, 2012) จากงานวิจัยในอดีต การขยายข้อสอบถามจะนำการทำเหมืองข้อมูลมาประยุกต์ใช้เพื่อวิเคราะห์หาคำสำหรับข้อสอบถามใหม่ การทำเหมืองข้อมูลที่นำมาใช้ในการขยาย

ข้อสอบถามนั้นจะสามารถจำแนกออกได้เป็นสามรูปแบบตารางที่ 2.1 แสดงรูปแบบการทำเหมือง
ข้อมูล



ตารางที่ 2.1 แสดงตัวอย่างการทำเหมืองข้อมูล ดัดแปลงจาก Han และคณะ (2011)และ Tuffery (2011)

รูปแบบการทำเหมืองข้อมูล	ตัวอย่างการทำเหมืองข้อมูล
<p>กฎความสัมพันธ์ (Association Rule)</p>	
<p>การจำแนกประเภท (Classification)</p>	
<p>การจัดกลุ่ม (Clustering)</p>	

(2) **กฎความสัมพันธ์ (Association Rule)** ขั้นตอนวิธีนี้จะหาความสัมพันธ์ระหว่างวัตถุสองชิ้น โดยกำหนดจำนวนความสัมพันธ์อย่างน้อยที่สุด เพื่อสรุปเป็นความสัมพันธ์แล้วนำมาใช้ในการประเมินความเป็นไปได้ของสิ่งที่น่าจะเกิดขึ้น (Han และคณะ, 2011) ในกฎความสัมพันธ์นั้น จะต้องกำหนดตัววัดระดับความสัมพันธ์ คือ (1) ค่าสนับสนุน (support) เป็นค่าความน่าจะเป็นที่จะเกิดเหตุการณ์สองเหตุการณ์พร้อมกัน และ (2) ค่าความเชื่อมั่น (Confidence) เป็นค่าความน่าจะเป็นที่จะเกิดเหตุการณ์หนึ่งภายใต้เงื่อนไขการเกิดเหตุการณ์หนึ่ง ค่าสนับสนุนและค่าความเชื่อมั่นนั้นมีค่าตั้งแต่ 0 ถึง 1 หากตัววัดระดับความสัมพันธ์ทั้งสองค่านี้มีค่ามากกว่าค่าคงที่ (Threshold) ผู้วิจัยสามารถกำหนดความสัมพันธ์ระหว่างวัตถุสองตัวนี้ได้

งานวิจัยจำนวนหนึ่งนำกฎความสัมพันธ์มาประยุกต์ใช้ในการขยายข้อสอบถาม Fonseca และคณะ(2005) ได้นำกฎความสัมพันธ์มาใช้ในการป้องกันความเสี่ยงของย้อนกลับด้วยการสร้างแนวคิดRungsawangและคณะ ได้ทดสอบนำกฎความสัมพันธ์มาประยุกต์ใช้ในการป้องกันความเสี่ยงของย้อนกลับ Lee, Huang และ Chao(2007) ได้นำกฎความสัมพันธ์มาใช้ในการสร้างคู่ของคำที่สัมพันธ์กันจากเอกสาร เพื่อนำมาใช้ในการวิเคราะห์คำสำหรับการป้องกันความเสี่ยงของย้อนกลับ Song และคณะ(2006) ได้นำกฎความสัมพันธ์มาประยุกต์ใช้กับออนไลน์ีและการประมวลผลภาษาธรรมชาติ (Natural Language Processing, NLP) เพื่อลดปัญหาการกำหนดคำที่มีความหมายใกล้เคียงกัน เพื่อนำมาใช้ในการป้องกันความเสี่ยงของย้อนกลับ Wang และ Niu (2012) ได้นำเสนอการวิเคราะห์ครอบคลุมด้วยออนไลน์ี เพื่อสร้างกฎความสัมพันธ์ระหว่างคำเพื่อนำมาใช้ในการขยายข้อสอบถาม Latiri และคณะ (2003) และ Wei (2000) ได้นำเอกสารจากคอร์ปัสเอกสารมาวิเคราะห์หาความสัมพันธ์สำหรับการขยายข้อสอบถามด้วยกฎความสัมพันธ์เชิงคลุมเครือ Xu และคณะ (2006) ได้ศึกษาการนำกฎความสัมพันธ์และการสร้างดัชนีเชิงความหมายแฝง (Latent Semantic Indexing) มาประยุกต์ใช้ในการวิเคราะห์เฉพาะส่วน การวิเคราะห์ครอบคลุม และการขยายข้อสอบถามเชิงออนไลน์ี ในระบบการค้นหาวรรณกรรมทางชีวแพทย์Han และ Chen (2008) ได้นำกฎความสัมพันธ์มาสร้างออนไลน์ีของกลุ่มคำจากเอกสารทั้งหมดในดคอร์ปัสเอกสาร มาใช้ในการกำหนดคำขยาย Lee, Huang และ Chao (2007) ได้นำล็อกข้อสอบถาม ซึ่งประกอบด้วยข้อสอบถามและเอกสารที่ผู้ใช้ได้กำหนดและเลือกเปิด มาสร้างอรรถาภิธานด้วยกฎความสัมพันธ์Gulati และ Sharma (2010) ได้นำกฎความสัมพันธ์มา

ประยุกต์ใช้ในการสร้างความสัมพันธ์ของคำจากข้อสอบถามในล๊อคข้อสอบถาม เพื่อนำมาใช้ในการขยายข้อสอบถามในระบบการค้นคืนรูปภาพ จากงานวิจัยในอดีตที่ได้กล่าวมานั้น จะเห็นได้ว่ากฎความสัมพันธ์นั้นจะนำมาใช้ในการสร้างความสัมพันธ์ระหว่างคำสองคำจากเอกสารในคอร์ปัสเอกสารหรือข้อสอบถามจากล๊อคข้อสอบถาม จากนั้นจึงนำคำที่สัมพันธ์กับคำในข้อสอบถามมาขยายข้อสอบถามเดิม เพื่อสร้างข้อสอบถามใหม่

(3) **การจำแนกประเภท(Classification)**จะนำข้อมูลไปไว้ในกลุ่มที่ตรงกับคุณสมบัติของข้อมูลตามที่ผู้วิจัยได้กำหนดไว้แล้ว ทั้งนี้ผู้วิจัยสามารถกำหนดลักษณะของกลุ่มข้อมูลแต่ละกลุ่ม (Han และคณะ, 2011; Tuffery, 2011)จากงานวิจัยในอดีต การขยายข้อสอบถามนั้นจะนำการจำแนกประเภทหลากหลายขั้นตอนวิธี ตารางที่ 2.2แสดงตัวอย่างการจำแนกประเภทที่พบนำมาใช้ในการขยายข้อสอบถาม

ตารางที่ 2.2 การนำการจำแนกประเภทมาใช้ในการขยายข้อสอบถามและตัวอย่างการนำไปใช้
ดัดแปลงจาก Han และคณะ (2011) และ Tuffery(2011)

ขั้นตอนการจำแนกประเภท	ตัวอย่างงานวิจัยในอดีต
การจำแนกแบบเบย์ (Bayes Classification)	<ul style="list-style-type: none"> ● Yang และ Zhang (2002) ได้ใช้การจำแนกแบบเบย์ (Bayes) ในการคัดเลือกคำสำหรับการขยายข้อสอบถามในระบบการค้นหาแบบกระจาย ● Campos, Fernandez และ Huete (1998) ได้นำการจำแนกแบบเบย์มาใช้ในการสร้างโครงสร้างเครือข่ายภายในอรรถาภิธาน (Thesaurus) เพื่อนำมาใช้ในการขยายข้อสอบถาม
เครือข่ายประสาท (Neural Network)	<ul style="list-style-type: none"> ● Han และ Chen (2009) นำเครือข่ายประสาท (Neural Network) และ ออนโทโลยี (Ontology) มาประยุกต์ใช้ในการขยายข้อสอบถาม ● Takagi และ Tajima (2001) ได้นำเครือข่ายประสาทมาสร้างเซตของคำเพื่อนำมาใช้ในการขยายข้อสอบถาม
ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM)	<ul style="list-style-type: none"> ● Meshram, Mishra และ Sharma (2007) ใช้เครื่องมือสนับสนุนเวกเตอร์ร่วมกับการป้อนความเกี่ยวข้องกลับ เพื่อลดขั้นตอนการขยายข้อสอบถามในระบบการค้นหาค้นคืนภาพ ● Zhang และคณะ (2007) ได้นำซัพพอร์ตเวกเตอร์แมชชีนสำหรับการกำหนดความเกี่ยวข้องระหว่างเอกสารและข้อสอบถาม เพื่อนำมาใช้ในผลป้อนกลับความเกี่ยวข้อง ● Rahman และ Bhattacharya (2009) ได้นำซัพพอร์ตเวกเตอร์แมชชีนร่วมกับรูปแบบทางสถิติมาใช้ในการสร้างแนวคิดหรือกลุ่มคำที่แทนความหมายของสารสนเทศ

ขั้นตอนการ จำแนกประเภท	ตัวอย่างงานวิจัยในอดีต
<p>ขั้นตอนวิธีเจเนติก หรือขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm, GA)</p>	<ul style="list-style-type: none"> ● วาริน นาราวิทย์และศิริภิญโญ จันทมูณี (2554) ได้นำขั้นตอนวิธีเจเนติกมาใช้ในการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาไทย ● Hmeidi, Najadat และ Al-Sha'or ได้ศึกษาการนำขั้นตอนวิธีเจเนติกมาประยุกต์ใช้ในการขยายข้อสอบถามแบบอัตโนมัติ ● Kimura และ Araki (2006) ได้นำขั้นตอนวิธีเจเนติกมาใช้ในการขยายข้อสอบถามในระบบการค้นคืนคำถามเชิงสาระภาษาญี่ปุ่น (Contextual Question Retrieval)
<p>ตรรกะคลุมเครือหรือตรรกะวิชันัย (Fuzzy Logic)</p>	<ul style="list-style-type: none"> ● Lee, Lin และ Huang (2001) ได้นำตรรกะคลุมเครือมาใช้ในการสร้างอรรถาภิธานแบบความสัมพันธ์เชิงคลุมเครือ สำหรับการกำหนดคำมาใช้ในการขยายข้อสอบถาม ● Kim, Kim และ Kim (2001) และ Chang, Chen และ Liao (2007) ได้นำตรรกะคลุมเครือมาประยุกต์กับการป้อนความสัมพันธ์ย้อนกลับ ด้วยการคำนวณค่าน้ำหนักของคำในเอกสาร ● Cock และ Cornelis นำตรรกะคลุมเครือมาใช้ในการกำหนดคำสำหรับการขยายข้อสอบถาม

(4) **การจัดกลุ่ม (Clustering) หรือ การวิเคราะห์กลุ่ม (Cluster Analysis)** จะจัดข้อมูลเข้าสู่กลุ่มที่มีคุณสมบัติที่เหมือนหรือคล้ายคลึงกัน โดยพิจารณาจากลักษณะของข้อมูล (Han และคณะ, 2011) จากงานวิจัยในอดีต การจัดกลุ่มสามารถนำมาใช้ในการขยายข้อสอบถาม ตารางที่ 2.3 แสดงตัวอย่างงานวิจัยในอดีตที่นำการจัดกลุ่มมาใช้ในการขยายข้อสอบถาม นอกจากนี้ยังมีการนำการจัดกลุ่มมาประยุกต์ใช้กับการเหมืองข้อมูล



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 2.3 การนำการจัดกลุ่มมาใช้ในการขยายข้อสอบถามและตัวอย่างการนำไปใช้ดัดแปลง
จาก Han และคณะ (2011) และ Tuffery (2011)

ขั้นตอนการจัดกลุ่ม	ตัวอย่างงานวิจัยในอดีต
<p>การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering)</p>	<ul style="list-style-type: none"> ● Aono และ Doi (2005) ได้นำเสนอขั้นตอนการจัดกลุ่มแบบลำดับขั้นสำหรับการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศ ● Dvorsky, Martinovic และ Snasel (2004) ได้นำการจัดกลุ่มแบบลำดับขั้น มาใช้ในการกำหนดดัชนี ● Han และ Chen (2009) ได้นำการจัดกลุ่มและอนโทโลยีมาจัดกลุ่มคำ เพื่อนำมาใช้ในการวิเคราะห์หาคำ ● Xu และ Hu (2010) ได้นำการจัดกลุ่มแบบลำดับขั้นมาใช้ในการจัดกลุ่มเอกสารร่วมกับรูปแบบภาษา (Language Modeling) เพื่อค้นหาคำ ● Wang และ Tanaka (2006) ได้นำการจัดกลุ่มแบบเชื่อมโยงด้วยค่าเฉลี่ยมาใช้ในการจัดกลุ่มคำร่วมกับทฤษฎีคอขวดสารสนเทศ (Information Bottleneck Theory) เพื่อสร้างความสัมพันธ์ระหว่างคำ
<p>การจัดกลุ่มแบบการแบ่งส่วน (Partitional Clustering)</p>	<ul style="list-style-type: none"> ● Lin, Lin และ He (2012) ได้นำการจัดกลุ่มด้วยขั้นตอนวิธีค่าเฉลี่ยเค (K-means Algorithm) มาสร้างความสัมพันธ์ระหว่างคำ
<p>การจัดกลุ่มเชิงความหนาแน่น (Density Based Clustering)</p>	<ul style="list-style-type: none"> ● Lee, Croft และ Allan (2008) ได้นำการจัดกลุ่มแบบเพื่อนบ้านใกล้ที่สุด (Nearest Neighbor Clustering) มาประยุกต์ใช้กับการป้อนการสุ่มตัวอย่างย้อนกลับ (Resampling Feedback)
<p>การจัดกลุ่มเชิง</p>	<ul style="list-style-type: none"> ● Akrivas, Wallace, Stamou และ Kollias (2002) ได้นำการจัด

ขั้นตอนการจัดกลุ่ม	ตัวอย่างงานวิจัยในอดีต
<p>รูปแบบความน่าจะเป็น (Probability Model Based Clustering)</p>	<p>กลุ่มคลุมเครือมาใช้ในการสร้างกลุ่มคำที่มีความสัมพันธ์ต่อกัน เพื่อนำไปเปรียบเทียบกับข้อสอบถามจากผู้ใช้</p> <ul style="list-style-type: none"> ● Amini และ Usunier (2007) ได้นำการจัดกลุ่มด้วยขั้นตอนวิธีความคาดหวังสูงสุด (Expectation-Maximum Algorithm) มาใช้ในการจัดกลุ่มคำ ● Lita และ Carbonnell (2008) นำขั้นตอนวิธีความคาดหวังสูงสุดมาจัดกลุ่มคำถาม มาใช้ในผลป้อนกลับความเกี่ยวข้อง
<p>การจัดกลุ่มเชิงกราฟและเครือข่ายข้อมูล (Clustering Graph and Network Data)</p>	<ul style="list-style-type: none"> ● Bernhard (2010) ได้นำการจัดกลุ่มแบบการนิยาม (Definition Clustering) มาประยุกต์ใช้กับผลป้อนกลับความเกี่ยวข้อง ● Rahimi และ Zahedi (2010) ได้นำขั้นตอนการจัดกลุ่มมาทำงานร่วมกับการวิเคราะห์เชิงความหมายแฝง (Latent Semantic Analysis, LSA) เพื่อกำหนดคำที่นำไปใช้ในการขยายข้อสอบถาม ทั้งนี้ใช้การจัดกลุ่มแบบลำดับชั้น และการจัดกลุ่มแบบการแบ่งส่วนร่วมด้วย
<p>การจัดกลุ่มเชิงคลุมเครือ (Fuzzy Clustering)</p>	<ul style="list-style-type: none"> ● Akrivas และคณะ (2002) ได้นำการจัดกลุ่มเชิงคลุมเครือมาใช้ในการจัดกลุ่มดัชนีของเอกสารจากคอร์ปัสเอกสารทั้งหมด เพื่อนำมาใช้ในการขยายข้อสอบถาม

จากงานวิจัยในอดีตนั้น การกำหนดขั้นตอนวิธีการทำเหมืองข้อมูลมาใช้ในการขยายข้อสอบถามนั้นจึงมีความสำคัญอย่างยิ่ง อย่างไรก็ตาม เมื่อพิจารณาจากการขยายตัวของขอบเขต

สารสนเทศที่เกิดขึ้นในปัจจุบันแล้วและความต้องการข้อมูลของผู้ใช้ที่เปลี่ยนแปลงไป การนำกฎความสัมพันธ์และการจำแนกประเภทข้อมูลมาใช้ในการขยายข้อสอบถามนั้นอาจจะสร้างผลลัพธ์ในการค้นคืนได้ไม่ตรงกับความต้องการของผู้ใช้ การสร้างกฎความสัมพันธ์จะต้องใช้ค่าคงที่น้อยสุดของค่าสนับสนุน (Minimum Support Threshold) และค่าคงที่น้อยสุดของค่าความเชื่อมั่น (Minimum Confidence Threshold) ในการพิจารณาความสัมพันธ์ระหว่างวัตถุสองชิ้น (Li และ คณะ, 1999; Lai; Mueyba, Khan, และ Coenen, 2009; Jayalakshmi และ Rao) หากลักษณะของสารสนเทศเปลี่ยนแปลงจนส่งผลกระทบต่อค่าสนับสนุนและค่าความเชื่อมั่น ค่าคงที่ทั้งสองค่าที่ได้กำหนดไว้นั้นอาจจะส่งผลกระทบต่อผลลัพธ์ที่ได้จากการค้นคืนด้วยเช่นกัน นอกจากนี้แล้วยังมีงานวิจัยจำนวนหนึ่งที่นำความคิดเห็นจากผู้ใช้ระบบการค้นคืนสารสนเทศไปใช้ในการประเมินและสร้างกฎความสัมพันธ์ขึ้นมา Lee, Huang และ Chao (2007) ได้นำเสนอการนำรายการข้อสอบถามจากลิ้งค์ข้อสอบถามและรายการผลลัพธ์การค้นคืนที่ได้จากแต่ละข้อสอบถามและผู้ใช้ได้เลือกศึกษามาใช้ในการสร้างกฎความสัมพันธ์ แต่ผลลัพธ์การค้นคืนที่ผู้ใช้ไม่ได้เลือกนั้น อาจจะตรงกับความต้องการของผู้ใช้ได้เช่นกัน

ส่วนการจำแนกประเภทนั้น ผู้วิจัยจะต้องทราบประเภทหรือเกณฑ์การจำแนกประเภทข้อมูล แต่เมื่อลักษณะของสารสนเทศได้ขยายตัวมากขึ้น ประเภทหรือเกณฑ์การจำแนกประเภทข้อมูลเดิมที่มีอยู่อาจจะไม่เพียงพอต่อกรณีวิเคราะห์ข้อมูล และอาจจะส่งผลกระทบต่อความต้องการข้อมูลของผู้ใช้และผลลัพธ์การค้นคืนสารสนเทศ (Han และคณะ, 2011; Tuffery, 2011) จากที่กล่าวมาข้างต้นนั้น การนำการสร้างกฎความสัมพันธ์และการจำแนกประเภทข้อมูลนั้น อาจจะทำให้ผลลัพธ์การค้นคืนที่ได้นั้นไม่ตรงกับความต้องการของผู้ค้นคืน

ในขณะที่การจัดกลุ่มนั้นจะไม่ทราบประเภทหรือเกณฑ์ในการจำแนกประเภทข้อมูล จึงเหมาะสำหรับการนำมาใช้ในการวิเคราะห์ข้อมูลที่ไม่สามารถกำหนดประเภทหรือคุณสมบัติของข้อมูลได้มีงานวิจัยจำนวนหนึ่งที่เปรียบเทียบการใช้การจัดกลุ่มกับการทำเหมืองข้อมูลด้วยขั้นตอนวิธีต่างๆ Rahimi และ Zahedi (2010) ได้นำเสนอการขยายข้อสอบถามซึ่งได้นำการวิเคราะห์เชิงความหมายแฝง (Latent Semantic Analysis) และการจัดกลุ่มมาใช้ในการกำหนดคำ และกำหนดให้เปรียบเทียบกับขั้นตอนวิธี Rocchio ซึ่งต้องใช้หลักการจำแนกเอกสารจากรายการผลลัพธ์ที่ได้ด้วยความเกี่ยวข้องกับข้อสอบถาม (จากหัวข้อ 2.2.3) ผลการทดลองพบว่าผลลัพธ์ที่ได้จากการขยายข้อสอบถามที่ผู้วิจัยเสนอนั้นให้ประสิทธิผลการค้นคืนที่ดีกว่าผลลัพธ์ที่ได้จากขั้นตอนวิธี Rocchio นอกจากนี้ Aono และ Doi (2005) ได้นำเสนอขั้นตอนวิธีลำดับชั้นของกลุ่มมาใช้ในการขยายข้อสอบถาม โดยเปรียบเทียบกับซัพพอร์ตเวกเตอร์แมกซิม (Support Vector

Machine) พบว่าวิธีการจัดกลุ่มที่นำเสนอขึ้นได้ให้ผลลัพธ์การค้นคืนที่ดีกว่าซอฟต์แวร์แมกชีนเพื่อศึกษาขั้นตอนวิธีการขยายข้อสอบถาม ผู้วิจัยต้องการลดการนำการตัดสินใจของผู้ใช้มาช่วยในการขยายข้อสอบถามด้วยจากที่กล่าวมาข้างต้น ผู้วิจัยสนใจการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถาม (สุทธิ ชัตติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554)

จากงานวิจัยในอดีต มีการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามร่วมด้วย เพื่อนำมาใช้ในการกำหนดค่า ตารางที่ 2.3 แสดงตัวอย่างงานวิจัยการนำการจัดกลุ่มมาใช้ในการขยายข้อสอบถามการจัดกลุ่มนั้นสามารถจำแนกได้สองรูปแบบ

รูปแบบแรกจะจำแนกการจัดกลุ่มตามความสัมพันธ์ระหว่างกลุ่มและขั้นตอนการจัดกลุ่ม รูปแบบนี้มีขั้นตอนสองขั้นตอนวิธีหลัก ได้แก่ การจัดกลุ่มแบบการแยกส่วน (Partitional Clustering, Flat Clustering) และการจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering) (ศุภชัย ตั้งวงศ์ศานต์, 2551; Yin, Kaku และ Zhu, 2011; Han และคณะ, 2011) การจัดกลุ่มแบบการแยกส่วนจะกำหนดจำนวนกลุ่มและจุดศูนย์กลางที่จะนำมาใช้เป็นเกณฑ์ในการจัดกลุ่ม ทั้งนี้อาจคำนวณค่าจุดศูนย์กลางใหม่ เพื่อนำมาใช้เป็นเกณฑ์ในการจัดกลุ่มใหม่ การจัดกลุ่มแบบนี้จะไม่มีความสัมพันธ์ระหว่างกลุ่ม ขั้นตอนวิธีการจัดกลุ่มแบบการแยกส่วนที่นิยมนำมาใช้นั้น คือ ค่าเฉลี่ยเค (k-means) และค่าเฉลี่ยซีเชิงคลุมเครือ (Fuzzy C-Means) ส่วนการจัดกลุ่มแบบลำดับชั้นนั้น จะเน้นการเชื่อมโยงข้อมูลที่มีลักษณะที่ใกล้เคียงกัน จนกระทั่งเกิดโครงสร้างต้นไม้ (Tree Structure) รูปแบบการจัดกลุ่มนี้จะสร้างความสัมพันธ์ระหว่างกลุ่ม ขั้นตอนวิธีการจัดกลุ่มแบบลำดับชั้นที่นิยมใช้นั้นคือ การจัดกลุ่มแบบลำดับชั้นอ็กโกลเมราทีฟ (Agglomerative Hierarchical Clustering)

รูปแบบที่สองนั้นจะจำแนกตามความสัมพันธ์ระหว่างข้อมูลและกลุ่ม รูปแบบนี้มีสองขั้นตอนวิธีหลัก ได้แก่ การจัดกลุ่มแบบแข็ง (Hard Clustering) และ การจัดกลุ่มเชิงคลุมเครือ (Fuzzy Clustering) หรือการจัดกลุ่มแบบอ่อน (Soft Clustering) การจัดกลุ่มแบบแข็งนั้น สมาชิกต้องมีความสัมพันธ์กับกลุ่มเพียงกลุ่มหนึ่งกลุ่มเท่านั้น ขั้นตอนวิธีการจัดกลุ่มแบบแข็งที่พบเจอได้ชัด คือ ค่าเฉลี่ยเคและการจัดกลุ่มอ็กโกลเมราทีฟ ในขณะที่การจัดกลุ่มเชิงคลุมเครือนั้น สมาชิกมีความสัมพันธ์กับกลุ่มได้ตั้งแต่หนึ่งกลุ่มขึ้นไป ขั้นตอนวิธีการจัดกลุ่มเชิงคลุมเครือที่พบเจอได้ คือ ค่าเฉลี่ยซีเชิงคลุมเครือ

นอกจากการขยายข้อสอบถามแล้ว การจัดกลุ่มยังนำไปใช้ในการทำเหมืองข้อมูลรูปแบบต่างๆ รวมทั้งการค้นคืนสารสนเทศ Steinbach, Karypis และ Kumar (2000) ได้ศึกษาการจัดกลุ่มแบบการแยกส่วนและการจัดกลุ่มแบบลำดับชั้นในการจัดกลุ่มเอกสาร พบว่าการจัดกลุ่มแบบลำดับชั้นไม่ว่าจะเป็นขั้นตอนวิธีอัลกอริทึมที่ฟหรือค่าเฉลี่ยเคแบบแบ่งครึ่ง (Bisecting k-means) จะให้คุณภาพการจัดกลุ่ม (Clustering Quality) ที่ดีกว่าขั้นตอนวิธีค่าเฉลี่ยเค Horng และคณะ (2005) ได้ออกแบบการจัดกลุ่มเชิงคลุมเครือแบบลำดับชั้นอัลกอริทึมที่ฟ (Fuzzy Agglomerative Hierarchical Clustering) สำหรับการกำหนดข้อสอบถามใหม่ (Query Reformulation) พบว่าขั้นตอนวิธีการจัดกลุ่มนี้ช่วยให้การค้นคืนสารสนเทศนั้นมีประสิทธิผลการค้นคืนที่ดีกว่าการใช้ขั้นตอนวิธีค่าเฉลี่ยเค ขั้นตอนการจัดกลุ่มแบบอัลกอริทึมที่ฟ และขั้นตอนวิธีค่าเฉลี่ยซีเชิงคลุมเครือ Zamir และ Etzioni (1998) ได้นำเสนอการจัดกลุ่มต้นไม้ต่อท้าย (Suffix Tree Clustering) ซึ่งมีรูปแบบการจัดกลุ่มแบบลำดับชั้น ซึ่งเมื่อเทียบกับการจัดกลุ่มแบบการแยกส่วน พบว่าขั้นตอนวิธีการจัดกลุ่มที่นำเสนอให้ประสิทธิผลการค้นคืนสารสนเทศที่ดีกว่าการจัดกลุ่มแบบการแบ่งส่วน Kashef และ Kamel ได้นำเสนอขั้นตอนวิธีการจัดกลุ่มแบบแข็งและอ่อนร่วมกันแบบกระจาย (Distributed Cooperative Hard-Fuzzy Document Clustering) โดยประยุกต์จากขั้นตอนวิธีค่าเฉลี่ยเคและค่าเฉลี่ยซีเชิงคลุมเครือ และเปรียบเทียบขั้นตอนวิธีการจัดกลุ่มทั้งสามวิธี พบว่าคุณภาพการจัดกลุ่มของขั้นตอนวิธีที่นำเสนอที่ดีกว่าของขั้นตอนวิธีอีกสองวิธีที่เหลือ Chen, Wang และ Jiang (2009) ได้นำเสนอขั้นตอนวิธีการจัดกลุ่ม R-FPC (Fuzzy Projected Clustering) ซึ่งมีการนำขั้นตอนวิธี R-Greedy มาใช้ พบว่า คุณภาพการจัดกลุ่มที่นำขั้นตอนการจัดกลุ่มที่นำเสนอมาใช้ นั้น ดีกว่าการนำค่าเฉลี่ยซีเชิงคลุมเครือมาใช้

จากงานวิจัยในอดีต มีการศึกษาเปรียบเทียบระหว่างการจัดกลุ่มแบบการแยกส่วนและการจัดกลุ่มแบบลำดับชั้น Steinbach และคณะ (2000) ได้ศึกษาเปรียบเทียบรูปแบบการจัดกลุ่มทั้งสองแบบ พบว่า การจัดกลุ่มแบบลำดับชั้นจะนำไปใช้ในการจัดกลุ่มได้คุณภาพการจัดกลุ่มที่ดีกว่าการจัดกลุ่มแบบการแยกส่วน Horng และคณะ (2005), Abbas (2008) และ ชุสรีรัตน์ จรัสกุลชัย และคณะ (2544) ได้เปรียบเทียบกับการจัดกลุ่มแบบการแย่งส่วนและการจัดกลุ่มแบบลำดับชั้นในการประเมินการจัดกลุ่มกับขั้นตอนวิธีที่ได้นำเสนอ คือการจัดกลุ่มแบบลำดับชั้นอัลกอริทึมที่ฟเชิงคลุมเครือ (Fuzzy Agglomerative Hierarchical Clustering) พบว่า ถึงแม้ว่าขั้นตอนวิธีที่ได้นำเสนอนั้นจะให้คุณภาพการจัดกลุ่มที่ดีกว่าขั้นตอนวิธีการจัดกลุ่มสองขั้นตอนวิธีที่นำมาเปรียบเทียบ การจัดกลุ่มแบบลำดับชั้นนั้นก็ให้ประสิทธิผลการค้นคืนที่ดีกว่าการจัดกลุ่มแบบการแยกส่วน นอกจากการเปรียบเทียบคุณภาพการจัดกลุ่มนี้แล้วยังนำการจัดกลุ่มที่นำเสนอนั้นมา

เปรียบเทียบกับค่าเฉลี่ยซีเชิงคลุมเครือในการกำหนดข้อสอบถามใหม่ พบว่าขั้นตอนการจัดกลุ่มที่นำเสนอนั้น ให้ประสิทธิผลที่ดีกว่าค่าเฉลี่ยซีเชิงคลุมเครือ จากที่กล่าวมานั้น พบว่าการจัดกลุ่มแบบลำดับขั้นนั้นให้คุณภาพการจัดกลุ่มที่ดีกว่าการจัดกลุ่มแบบการแยกส่วน

นอกจากรูปแบบการจัดกลุ่มตามความสัมพันธ์ระหว่างกลุ่มแล้วยังมีการศึกษาการเปรียบเทียบระหว่างการจัดกลุ่มแบบแข็งและการจัดกลุ่มเชิงคลุมเครือ จากข้อความข้างต้นที่ได้กล่าวมานั้น Horng และคณะ (2005) ได้นำเสนอการจัดกลุ่มแบบลำดับขั้นเชิงคลุมเครือ (Fuzzy Agglomerative Hierarchical Clustering) เปรียบเทียบกับการจัดกลุ่มแบบแข็งสองขั้นตอนวิธี พบว่าขั้นตอนที่ได้นำเสนอนั้นให้คุณภาพการจัดกลุ่มที่ดีกว่าการจัดกลุ่มแบบแข็ง Jipkate และ Goholar (2008) ได้นำเสนอการเปรียบเทียบขั้นตอนวิธีค่าเฉลี่ยเคและค่าเฉลี่ยซีเชิงคลุมเครือ Shroavastava และคณะ (2010) ได้นำเสนอและเปรียบเทียบการนำการจัดกลุ่มมาประยุกต์ใช้กับการค้นคืนภาพเชิงเนื้อหา (Content Based Image Retrieval, CBIR) ด้วยขั้นตอนวิธีค่าเฉลี่ยเคและขั้นตอนวิธีค่าเฉลี่ยซีเชิงคลุมเครือ Schenker และคณะ (2004) ได้นำการจัดกลุ่มด้วยขั้นตอนวิธีต่างๆ มาประยุกต์ใช้ในการจัดกลุ่มเอกสารเว็บ พบว่าขั้นตอนวิธีค่าเฉลี่ยซีเชิงคลุมเครือ นั้น จะให้คุณภาพการจัดกลุ่มที่ดีกว่าการจัดกลุ่มแบบแข็ง ถึงแม้ว่าจะใช้ขั้นตอนการจัดกลุ่มแบบลำดับขั้นเชิงเดี่ยว (Simple Hierarchical Clustering) และการจัดกลุ่มแบบลำดับขั้นเชิงสมบูรณ์ (Complete Hierarchical Clustering) ในบางงานวิจัยยังมีการนำเสนองานวิจัยการการจัดกลุ่มเชิงคลุมเครือ Jamshidi และ Pilevar (2013) ได้นำเสนอการวิเคราะห์และจัดกลุ่มรูปภาพด้วยขั้นตอนวิธีค่าเฉลี่ยซีเชิงคลุมเครือ และขั้นตอนวิธีเจเนติก (Genetic Algorithm) Li และคณะ (2008) ได้นำเสนอขั้นตอนการจัดกลุ่มแบบลำดับขั้นเชิงคลุมเครือ ด้วยการนำขั้นตอนวิธีค่าเฉลี่ยเคมาประยุกต์ใช้ นอกจากนี้แล้วยังมีการเปรียบเทียบกับขั้นตอนวิธีการจัดกลุ่มอื่นๆ Zhang และคณะ (2010) ได้นำเสนอการจัดกลุ่มที่ใช้กับการกำหนดจำนวนกลุ่มอัตโนมัติ Zhang และคณะ (2011) ได้นำการจัดกลุ่มมาใช้ในการสร้างกฎการจำแนกข้อมูล ถึงแม้ว่าจะมีงานวิจัยเกี่ยวกับการจัดกลุ่มเชิงคลุมเครือ ซึ่งให้คุณภาพการจัดกลุ่มที่ดีกว่าการจัดกลุ่มแบบคลุมเครือ แต่งานวิจัยการจัดกลุ่มในการขยายข้อสอบถามส่วนใหญ่นั้นนิยมนำการจัดกลุ่มแบบแข็งมาประยุกต์ใช้ในการขยายข้อสอบถามมากกว่า อีกทั้งการจัดกลุ่มเชิงคลุมเครือที่นำมาใช้ในการขยายข้อสอบถามนั้น มักจะเป็นการจัดกลุ่มที่ผู้วิจัยในอดีตนำเสนอขึ้น ซึ่งไม่มีการนำมาใช้ในการศึกษาต่อเพิ่มเติม ผู้วิจัยจึงเลือกใช้การจัดกลุ่มแบบอัลโกลเมราทีฟซึ่งเป็นการจัดกลุ่มแบบลำดับขั้นและการจัดกลุ่มแบบแข็ง นำมาใช้ในการศึกษาการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถาม

(2) การจัดกลุ่มแบบลำดับชั้น:การจัดกลุ่มแบบลำดับชั้นอัลไกลเมราทีฟ (Agglomerative Hierarchical Clustering) และการกำหนดการจัดกลุ่ม

ขั้นตอนวิธีที่ได้นิยมนำมาใช้ในการจัดกลุ่มแบบลำดับชั้นคือขั้นตอนวิธีการจัดกลุ่มอัลไกลเมราทีฟ (Agglomerative Clustering) การจัดกลุ่มอัลไกลเมราทีฟมีขั้นตอนวิธีดังต่อไปนี้ (Razmara และ Leila, 2007; Huang, 2008; Bailey, Craswell, และ Soboroff, 2008; Horng และคณะ, 2005; Vega-Pons และ Ruiz-Shulcloper, 2011; Auffarth, 2010; Schenker และคณะ, 2004; Sesli และ Yegenoglu, 2010; Liu, Pennell, Liu และ Liu, 2009; Buscher และคณะ, 2012; Morita และคณะ, 2011)

ขั้นตอนที่ 1 กำหนดกลุ่มที่มีสมาชิกข้อมูลเป็นข้อมูลหนึ่งหน่วย จำนวนกลุ่มเท่ากับจำนวนข้อมูล และกำหนดให้ค่าตัวแทนของกลุ่ม คือค่าของข้อมูลในกลุ่ม และกำหนดจำนวนกลุ่ม (Number of Clusters) ในการจัดกลุ่ม มีค่าตั้งแต่สองกลุ่มขึ้นไป

ขั้นตอนที่ 2 คำนวณการวัดค่าความเหมือน (Similarity measure) ระหว่างคู่กลุ่มทั้งหมด ด้วยค่าตัวแทนของกลุ่มแต่ละกลุ่ม

ขั้นตอนที่ 3 พิจารณาความเหมือนของทุกคู่กลุ่ม หากค่าความเหมือนของคู่กลุ่มใดมีค่ามากที่สุด ให้รวมคู่กลุ่มเป็นกลุ่มเดียวกัน

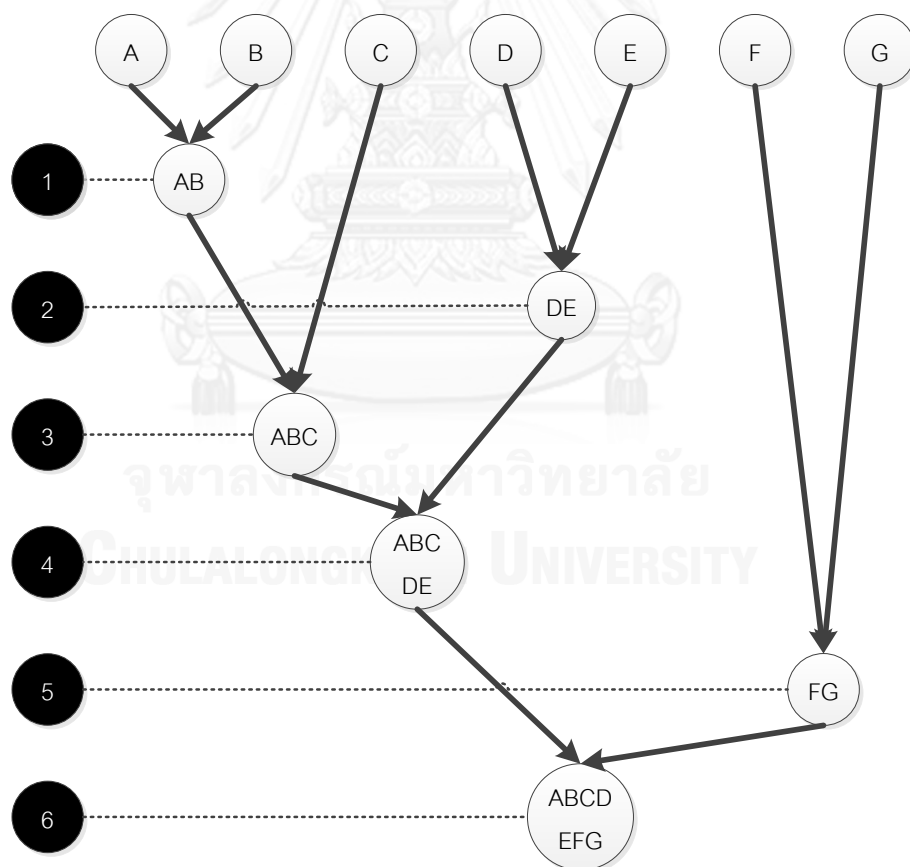
ขั้นตอนที่ 4 คำนวณค่าศูนย์กลางของกลุ่มใหม่ด้วยสมการของเกณฑ์การเชื่อมโยง (Linkage Criteria)

ขั้นตอนที่ 5 ตรวจสอบเกณฑ์การหยุด (Stopping Criteria) หากเป็นเท็จให้กลับไปขั้นตอนที่ 2 หากเป็นจริง ให้ยุติการจัดกลุ่ม

ผลลัพธ์ที่ได้จะมีรูปแบบตามตัวอย่างในรูปที่ 2.6 โดยในวงกลมสีขาวแสดงกลุ่มของข้อมูลที่นำมาใช้ในการจัดกลุ่ม จะเห็นได้ว่าเริ่มต้นนั้น จะกำหนดกลุ่มข้อมูล โดยมีสมาชิกเพียงหนึ่งตัว จากนั้น จะมีการรวมกลุ่มต่อไปตามลำดับหมายเลขในวงกลมสีดำและเส้นประ จนกระทั่งสามารถรวมกลุ่มข้อมูลได้ทั้งหมด หรือตามเงื่อนไขที่กำหนดไว้ในเกณฑ์การหยุดจากขั้นตอนที่กล่าวมาข้างต้น การกำหนดจำนวนกลุ่ม (Number of Cluster) สมการการวัดค่าความเหมือน (Similarity

measure) เกณฑ์การเชื่อมโยง(Linkage Criteria) และเกณฑ์การหยุด (Stopping Criteria) มีผลต่อประสิทธิภาพการจัดกลุ่ม โดยมีรายละเอียดดังต่อไปนี้

(1) **จำนวนกลุ่ม (Number of Clusters)** เนื่องจากการจัดกลุ่มนี้ ผู้วิจัยจะไม่ทราบถึงจำนวนกลุ่มที่ต้องใช้ เพื่อกำหนดผลลัพธ์การจัดกลุ่มที่ต้องการ ดังนั้นจึงมีงานวิจัยจำนวนหนึ่งที่ศึกษาการกำหนดจำนวนกลุ่มสำหรับการจัดกลุ่ม (Rousseeuw, 1987; Tibshirani, Walther และ Hastie; Yan, 2005) อย่างไรก็ตาม ในงานวิจัยส่วนใหญ่ นั้น จะนิยมกำหนดค่าจำนวนกลุ่มจำนวนหนึ่ง แล้วนำมาใช้ในการประเมินและเปรียบเทียบคุณภาพการจัดกลุ่มที่ได้จากการจัดกลุ่มด้วยค่าจำนวนกลุ่มที่แตกต่างกัน เพื่อหาค่าจำนวนกลุ่มที่ให้คุณภาพการจัดกลุ่มที่ดีกว่า (อาทิ Steinbach และคณะ, 2001; Na และคณะ, 2006)



รูปที่ 2.6 ตัวอย่างการจัดกลุ่มด้วยขั้นตอนวิธีอัลกอริทึมที่ฟ ดัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551), Yin และคณะ (2011), Han และคณะ (2012)

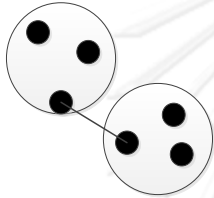
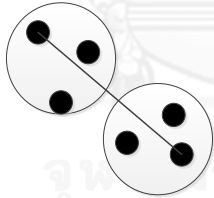
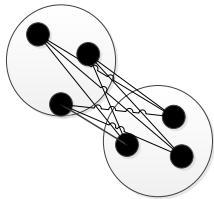
(2) **ค่าความเหมือน (Similarity measure)** คือการเปรียบเทียบระหว่างกลุ่มสองกลุ่ม เพื่อนำมาใช้ในการพิจารณาว่า กลุ่มสองกลุ่มมีลักษณะที่ใกล้เคียงกันหรือไม่ มีสมการการวัดค่าความเหมือนจำนวนหนึ่งที่น่ามาใช้ในการทำเหมืองข้อมูล ดังตารางที่ 2.4 แสดงสมการการวัดค่าความเหมือนและตัวอย่างงานวิจัยที่ปรากฏ โดยที่ A และ B แทนกลุ่มสองกลุ่มที่ต้องการนำมาประเมินความเหมือน a_i และ b_i แทนลำดับสมาชิกในกลุ่มลำดับที่ i ใน A และ B ตามลำดับและ t แทนจำนวนค่าที่ปรากฏในเอกสารทั้งหมด จากงานวิจัยในอดีต ผู้วิจัยบางส่วนจะนำเลือกสมการหนึ่งตัวสำหรับการนำมาใช้ในการวัดค่าความเหมือน ในขณะที่ผู้วิจัยบางส่วนนำสมการมาเปรียบเทียบ เพื่อทดสอบว่าสมการใดที่ให้ประสิทธิผลการจัดการข้อมูล หรือการค้นคืนสารสนเทศที่ดีกว่า จากงานวิจัยในอดีต สมการที่นิยมนำมาใช้ในการวัดค่าความเหมือน คือ สมการจาคคาร์ด (Jaccard) สมการไดซ์ (Dice) สมการโคไซน์ (Cosine) และสมการความสัมพันธ์ร่วมกันเพียร์สัน (Pearson Correlation) (Huang, 2008; Morita และคณะ, 2011; Sesli และ Yegenoglu, 2010; Morita และคณะ, 2011)

ตารางที่ 2.4 แสดงรายการสมการการวัดค่าความเหมือนที่พบในงานวิจัยในอดีต

ชื่อและสมการการวัดค่าความเหมือน	ตัวอย่างงานวิจัยที่พบ
<p style="text-align: center;">Cosine</p> $\frac{A \cdot B}{\ A\ \ B\ } = \frac{\sum_{i=1}^t (a_i \times b_i)}{(\sum_{i=1}^t a_i^2) \times (\sum_{i=1}^t b_i^2)}$	<p>Huang, 2008; Liu และคณะ, 2009; Buscher และคณะ, 2012; Morita และคณะ, 2011</p>
<p style="text-align: center;">Jaccard</p> $\frac{ A \cap B }{ A \cup B } = \frac{A \cdot B}{\ A\ ^2 + \ B\ ^2 - A \cdot B}$ $\frac{\sum_{i=1}^t (a_i \times b_i)}{(\sum_{i=1}^t a_i^2) + (\sum_{i=1}^t b_i^2) - (\sum_{i=1}^t (a_i \times b_i))}$	<p>Huang, 2008; Bailey และคณะ, 2008; Horng และคณะ, 2005; Vega-Pons และ Ruiz-Shulcloper, 2011; Auffarth, 2010; Schenker และคณะ, 2004; Sesli และ Yegenoglu, 2010; Dechang และ Xiaolin, 2008</p>
<p style="text-align: center;">Dice</p> $\frac{2 A \cap B }{ A + B } = \frac{2 \sum_{i=1}^t (a_i \cdot b_i)}{\sum_{i=1}^t a_i^2 + \sum_{i=1}^t b_i^2}$	<p>Morita และคณะ, 2011; Dalirsefat, Meyer และ Mirhoseini, 2009; Sesli และ Yegenoglu, 2010</p>
<p><u>Pearson Correlation</u></p> $\frac{\sum_{i=1}^t (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^t (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^t (b_i - \bar{b})^2}}$	<p>Huang, 2008; Morita และคณะ, 2011; Bhattacharya และ De, 2009, 2010; Langfelder, Peter และ Horvath, 2012; Ben-Hur และ Guyon, 2010</p>

(3) **เกณฑ์การเชื่อมโยง (Linkage Criteria)**คือการนำกลุ่มสองกลุ่มมารวมกันและกำหนดค่าตัวแทนของกลุ่มใหม่ ตารางที่ 2.5 จะแสดงสมการการคำนวณค่าตัวแทนของกลุ่ม (ศุภชัย ตั้งวงศ์วานศ์, 2551) โดยที่ a และ b เป็นสมาชิกในกลุ่ม A และ B ส่วน $d(a,b)$ คือระยะห่างระหว่าง a และ b จากงานวิจัยในอดีต นิยมใช้เกณฑ์การเชื่อมโยงทั้งสามเกณฑ์ โดยเฉพาะอย่างยิ่งเกณฑ์การเชื่อมโยงแบบง่าย (Simple Linkage Criteria) และเกณฑ์การเชื่อมโยงแบบสมบูรณ์ (Complete Linkage) อย่างไรก็ตามก็พบว่าเกณฑ์การเชื่อมโยงแบบค่าเฉลี่ย (Average Linkage Criteria) จะให้ประสิทธิภาพการจัดกลุ่มที่ดีกว่าเกณฑ์อื่นๆ (Sesli และ Yegenoglu, 2010; Fritsch และ Ickstadt, 2009) อย่างไรก็ตาม เกณฑ์การเชื่อมโยงแบบค่าเฉลี่ยนั้น อาจจะไม่ให้คุณภาพการจัดกลุ่มที่ดีกว่าเกณฑ์การเชื่อมโยงอื่นได้เสมอ ทั้งนี้ขึ้นอยู่กับลักษณะของข้อมูลที่นำมาใช้ในการวิเคราะห์

ตารางที่ 2.5 แสดงสมการการคำนวณของเกณฑ์การเชื่อมโยงที่พบในงานวิจัยในอดีต

ชื่อและสมการการคำนวณ	งานวิจัยที่พบในอดีต
<p>Minimum or Simple Linkage</p> $\min\{Sim(a, b): a \in A, b \in B\}$ 	<p>Hornig และคณะ, 2005; Mingoti และ Lima, 2006; Slonim และ Tishby, 2000; Bank และ Lima, 2010; Sesli และ Yegenoglu, 2010; Ferreira และ Hitchcock</p>
<p>Maximum or Complete Linkage</p> $\max\{Sim(a, b): a \in A, b \in B\}$ 	<p>Hornig และคณะ, 2005; Mingoti และ Lima, 2006; Slonim และ Tishby, 2000; Bank และ Lima, 2010; Sesli และ Yegenoglu, 2010; Ferreira และ Hitchcock</p>
<p>Mean or Average Linkage</p> $\frac{1}{ A B } \sum_{a \in A} \sum_{b \in B} Sim(a, b)$ 	<p>Steinbach และคณะ, 2000; Mingoti และ Lima, 2006; Razmara และ Kosseim, 2007; Bank และ Lima, 2010; Sesli และ Yegenoglu, 2010; Ferreira และ Hitchcock</p>

(4) **เกณฑ์การหยุด (Stopping Criteria)** เนื่องจากการจัดกลุ่มนั้น จะต้องพิจารณา ข้อมูลจำนวนมาก เพื่อลดขั้นตอนระยะเวลาในการจัดกลุ่ม จึงได้มีการกำหนดเกณฑ์การหยุด เป็น เงื่อนไขการหยุดการจัดกลุ่ม ที่ทำให้ลักษณะของกลุ่มที่จัดนั้นเป็นที่พอใจและเหมาะสม ผู้ใช้ ขั้นตอนวิธีการจัดกลุ่มรวมทั้งผู้วิจัยจำนวนมากจะกำหนดเกณฑ์การหยุดที่แตกต่างกันไป โดย ส่วนมากจะพิจารณาลักษณะของการจัดกลุ่ม อาทิ จำนวนกลุ่ม ค่าความเหมือนอย่างน้อย เกณฑ์ การหยุดนี้ นอกจากการจัดกลุ่มแบบลำดับขั้นแล้ว ยังนำมาใช้กับการจัดกลุ่มแบบการแบ่งส่วน ด้วยเช่นกัน Steinbach และคณะ (2000) และ Razmara และ Leila (2007) กำหนดให้ได้จำนวน กลุ่มตามที่กำหนดไว้, Horng และคณะ(2005) ได้กำหนดให้ค่าความเหมือนของทุกคู่กลุ่มที่ได้ น้อยกว่าค่าคงที่ ซึ่งผู้วิจัยได้กำหนดไว้ Oh, Honda และ Ichihashi (2009) และ Mahmoodi และ Mansoori (2012) กำหนดให้ค่าทดสอบมีค่าเท่ากับค่าคงที่ที่กำหนด ในบางงานวิจัย ผู้วิจัยกำหนด เงื่อนไขมากกว่าหนึ่งเงื่อนไข Liu, Wang และ Liu (2004) ตรวจสอบจำนวนกลุ่มและลักษณะของ กลุ่ม

2.4. งานวิจัยในอดีตที่นำการจัดกลุ่มแบบลำดับขั้นประยุกต์กับการขยาย ข้อสอบในการค้นคืนสารสนเทศ

จากงานวิจัยในอดีตที่กล่าวมานั้น การขยายข้อสอบถามนั้นมีความสำคัญอย่างยิ่งต่อ ระบบการค้นคืนสารสนเทศในปัจจุบัน เนื่องจากมีบทบาทในการช่วยกำหนดข้อสอบถามตรงกับ ความต้องการของผู้ใช้ระบบ แต่เนื่องจากการขยายตัวของขอบเขตสารสนเทศที่มีอยู่ในปัจจุบัน รวมทั้งรูปแบบการค้นคืนสารสนเทศ การขยายข้อสอบถามจึงนำมาใช้และศึกษาเพื่อประยุกต์ใช้ กับระบบการค้นคืนสารสนเทศที่มีหลายรูปแบบรูปแบบการค้นคืนสารสนเทศแบบหนึ่งที่น่าสนใจ คือการค้นคืนสารสนเทศภาษาไทย

การค้นคืนสารสนเทศภาษาไทยนั้น มีบทบาทอย่างยิ่งต่อผู้ค้นคืนสารสนเทศชาวไทยที่มี ความต้องการข้อมูลเพื่อนำมาใช้ในการดำรงชีวิตประจำวัน Theeramunkong และคณะ (2000) ได้ นำเสนอการค้นคืนสารสนเทศภาษาไทย โดยมุ่งประเด็นการศึกษารูปแบบการใช้ภาษาไทยกับ ขั้นตอนวิธีการทางคณิตศาสตร์และการทำเหมืองข้อมูล นพดล หมื่นไพ (พศ. 2548) ได้นำเสนอ การค้นคืนสารสนเทศด้วยออนโทโลยีและการแปรผันกลุ่มคำนาม (Noun Phrase Variant)

Khankasikam (2010) ได้ศึกษาประสิทธิภาพการค้นคืนสารสนเทศเมื่อนำใช้รูปแบบการค้นคืนสารสนเทศจำนวนสามรูปแบบ Chumwatana, Wong, และ Xie (2009) ได้นำเสนอการสร้างดัชนีที่นำมาใช้ในการค้นคืนสารสนเทศภาษาไทยด้วยความถี่ของส่วนย่อยของข้อความ Jaruskulchai (2001), Chuleerat (2001) และไกรศักดิ์ เกษร (พ.ศ. 2556) ได้นำเสนอแนวคิดการพัฒนาระบบการค้นคืนสารสนเทศแบบข้ามภาษาระหว่างภาษาไทยและภาษาอังกฤษจากงานวิจัยที่กล่าวมานั้น จะเห็นได้ว่าบทบาทของการค้นคืนสารสนเทศภาษาไทยเริ่มขยายตัวมากขึ้น

อย่างไรก็ตาม การกำหนดข้อสอบถามสำหรับการค้นคืนนั้น ยังคงเป็นปัญหาหลัก เช่นเดียวกับการค้นคืนสารสนเทศรูปแบบอื่นๆ ดังนั้นในปัจจุบันจึงเริ่มมีงานวิจัยการนำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนสารสนเทศภาษาไทยมากขึ้น Theva-aksorn และ Facundes (2007) ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนสารสนเทศแบบสองภาษา (ไทย-อังกฤษ) Theva-aksorn (2006) ได้นำเสนอการนำอรรถาภิธานมาสร้างความสัมพันธ์เชิงความหมายมาใช้ เพื่อนำมาใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทยรัตนากุมแก้วและสุพจน์นิศย์สุวัฒน์ (2552) ได้นำเสนอการค้นคืนพระคาถาธรรมบทภาษาบาลีอักษรไทย ทั้งนี้ได้นำเสนอการนำการขยายข้อสอบถาม เพื่อนำมาช่วยในการค้นคืน

นอกจากการศึกษากำหนดการขยายข้อสอบถามมาประยุกต์ใช้ในการค้นคืนสารสนเทศภาษาไทยแล้ว ยังมีงานวิจัยจำนวนหนึ่งที่นำขั้นตอนวิธีต่างๆ ทางคณิตศาสตร์ หรือการนำการทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้ในการขยายข้อสอบถามดังกล่าว Boonchom และ Soonthornprisaj (2012) ได้นำเสนอการนำข้อสอบถามมาประยุกต์ใช้ระบบการค้นคืนคีติความ โดยนำเสนอรูปแบบการค้นคืนด้วยขั้นตอนออนโทโลยี และนำขั้นตอนวิธีกลุ่มมด (Ant Colony Algorithm) มาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนวาริน นาราวิทย์ และศิริภิญโญ จันทมณี (2554) ได้นำเสนอการนำขั้นตอนวิธีเจเนติก (Genetic Algorithm) มาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย ด้วยรูปแบบปริภูมิเวกเตอร์จากงานวิจัยที่กล่าวมานั้น การทำเหมืองข้อมูลนั้นเริ่มมีบทบาทต่อการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาไทย

จากบทความข้างต้นนั้น ได้กล่าวถึงงานวิจัยการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาไทย ในขณะที่มีงานวิจัยการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาต่างๆ โดยเฉพาะอย่างยิ่ง การขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาอังกฤษ

จากงานวิจัยที่กล่าวมานั้น มีการนำการทำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถาม ด้วยขั้นตอนวิธีต่างๆ อาทิ ขั้นตอนวิธีเจเนติก ซัพพอร์ตเวกเตอร์แมกซิม การจำแนกแบบเบย์ หรือ การจัดกลุ่มแบบลำดับชั้น (สามารถศึกษาตัวอย่างการนำการทำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถามได้จากหัวข้อที่ 2.3.4 การทำเหมืองข้อมูลในการขยายข้อสอบถาม)

จะเห็นได้ว่า ยังมีงานวิจัยจำนวนน้อยที่ศึกษาการนำการทำเหมืองข้อมูลมาประยุกต์ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย เมื่อเทียบกับงานวิจัยแนวทางเดียวกันบนระบบการค้นคืนสารสนเทศภาษาอังกฤษ ในขณะที่ปัญหาการกำหนดข้อสอบถามในการค้นคืนสารสนเทศยังคงเกิดขึ้นอยู่เสมอ ไม่ว่าจะสารสนเทศที่ต้องการค้นคืนนั้นจะเป็นภาษาไทยหรือภาษาอังกฤษ อาทิ การสะกดคำ คำพ้องเสียง คำพ้องความหมาย (Khankasikam, 2010) ดังนั้นการศึกษานำการทำเหมืองข้อมูลมาใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทยนั้น จึงสำคัญอย่างมาก

จากงานวิจัยการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศภาษาอังกฤษ วิธีการทำเหมืองข้อมูลหนึ่งที่ได้นิยมนำมาใช้ในการขยายข้อสอบถามคือการจัดกลุ่ม (Clustering) ขั้นตอนวิธีนี้จะนำคำที่ได้จากแหล่งข้อมูลต่างๆ มาใช้ในการจัดกลุ่มคำ (Cluster of Terms) เพื่อนำกลุ่มคำที่ได้มาวิเคราะห์และกำหนดกลุ่มคำมาใช้ในการขยายข้อสอบถาม อย่างไรก็ตาม ผู้วิจัยไม่พบงานวิจัยการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทยทั้งที่มีงานวิจัยจำนวนหนึ่งที่น่าเสนอการนำการจัดกลุ่มมาประยุกต์ใช้ในการจัดการเอกสารหรือสารสนเทศภาษาไทย สมนาศัย กริ่งไกร และชุลีรัตน์ จรัสกุลชัย (2544) ได้นำเสนอการจัดกลุ่มเอกสารด้วยขั้นตอนวิธีการจัดกลุ่มค่าเฉลี่ยเค (K-means Clustering Algorithm) Santipong และคณะ (2012) ได้นำเสนอการจัดกลุ่มข้อความด้วยการจัดกลุ่มต้นไม้ต่อท้าย (Suffix Tree Clustering) พัชระ นาเสงี่ยม, สิริภัทร เชี่ยวชาญวัฒนา และคำรณ สุนิติ (พ.ศ.2553) ได้นำเสนอการจัดกลุ่มมาใช้ในการจำแนกคำจากเอกสารในรายการผลลัพธ์การค้นคืนสารสนเทศ Theeramunkong และคณะ (2000) ได้นำเสนอการค้นคืนเอกสารภาษาไทยด้วยกลุ่มตัวอักษร ดังนั้นงานวิจัยนี้จึงนำเสนอการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย เพื่อศึกษาถึงการนำการจัดกลุ่มมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย

ขั้นตอนการจัดกลุ่มที่นิยมนำมาใช้ในการขยายข้อสอบถาม คือการจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering) เนื่องจากผู้วิจัยจะใช้ความใกล้เคียงระหว่างคู่เวกเตอร์ในการวิเคราะห์

และกำหนดกลุ่มคำที่มีลักษณะที่ใกล้เคียงกัน มีงานวิจัยการนำการจัดกลุ่มมาใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศที่นำเสนอหรือศึกษาการนำการจัดกลุ่มแบบลำดับขั้นมาประยุกต์ใช้ในการขยายข้อสอบถาม (สามารถศึกษาตัวอย่างการนำการทำเหมืองข้อมูลมาประยุกต์ใช้ในการขยายข้อสอบถามได้จากตารางที่ 2.3 การนำการจัดกลุ่มมาใช้ในการขยายข้อสอบถามและตัวอย่างการนำไปใช้ที่ดัดแปลง) ทั้งนี้ผู้วิจัยในอดีตจะนำคำจากเอกสารที่ได้จากรายการผลลัพธ์การค้นคืนมาใช้ในการวิเคราะห์กลุ่ม เพื่อกำหนดกลุ่มคำสำหรับการขยายข้อสอบถาม ดังนั้นงานวิจัยนี้จึงนำการจัดกลุ่มการปรากฏร่วมกันของคำมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย ด้วยการจัดกลุ่มแบบลำดับขั้น

2.5. การประเมินประสิทธิผลการค้นคืน (Retrieval Effectiveness Evaluation)

2.5.1. ค่าระลึกได้ (Recall) และค่าความแม่นยำ (Precision)

งานวิจัยการค้นคืนสารสนเทศ รวมทั้งการขยายข้อสอบถามที่นำมาประยุกต์ใช้ในระบบการค้นคืนสารสนเทศ ได้นำเสนอหรือเปรียบเทียบขั้นตอนวิธีต่างๆ ที่นำมาใช้ในการค้นคืนสารสนเทศ จากนั้นจึงนำรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศมาใช้ในการประเมินประสิทธิผลการค้นคืน (Retrieval Effectiveness Evaluation) เพื่อประเมินถึงประสิทธิภาพของเครื่องมือการค้นคืนสารสนเทศหรือแนวคิดที่นำมาประยุกต์ใช้กับเครื่องมือดังกล่าว ตามวัตถุประสงค์ของผู้วิจัยหรือแนวทางการวิจัย ในการประเมินประสิทธิผลการค้นคืนนั้น ผู้วิจัยจะคำนวณค่าระลึกได้ (Recall) และค่าความแม่นยำ (Precision) เพื่อนำมาใช้ในการประเมินและศึกษาประสิทธิภาพการค้นคืน (ศุภชัย ตั้งวงษ์ศานต์, 2551; Buttcher และคณะ, 2009)

ค่าระลึกได้ (Recall) คือ อัตราส่วนระหว่างจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามและเป็นผลลัพธ์การค้นคืนต่อจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถาม ค่าระลึกได้นี้เป็นค่าที่แสดงถึงประสิทธิผลการดึงเอกสารที่เกี่ยวข้องกับข้อสอบถาม ค่าที่ได้สูงสุดคือ 1 ซึ่งแสดงว่าเอกสารที่ดึงออกมาทั้งหมดนั้นมีความเกี่ยวข้องกับข้อสอบถาม และในทางกลับกันค่าที่ต่ำสุดคือ 0 ซึ่งแสดงว่าเอกสารที่ดึงออกมานั้นไม่เกี่ยวข้องกับข้อสอบถาม (ศุภชัย ตั้งวงษ์ศานต์, 2551; Buttcher และคณะ, 2009)

ค่าความแม่นยำ (Precision) คือ อัตราส่วนระหว่างจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามและเป็นผลลัพธ์การค้นคืนต่อจำนวนเอกสารที่เป็นผลลัพธ์การค้นคืน ค่าความแม่นยำนี้ เป็นค่าที่แสดงถึงประสิทธิภาพในการดึงเอกสารที่สอดคล้องและตรงประเด็นกับข้อสอบถาม ค่าสูงสุดคือ 1 ซึ่งแสดงว่าเอกสารที่ดึงออกมานั้นมีความเกี่ยวข้องกับข้อสอบถาม และในทางกลับกัน ค่าต่ำสุดเป็น 0 ซึ่งจะแสดงว่า การไม่พบเอกสารใดที่เกี่ยวข้องกับข้อสอบถาม (ศุภชัย ตั้งวงศ์ศานต์, 2551; Buttcher และคณะ, 2009)

ค่าระลึกได้และค่าความแม่นยำจะสามารถคำนวณได้ดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551) กำหนดให้ R แทนเซตของเอกสารที่เกี่ยวข้องกับข้อสอบถาม A แทนเซตของเอกสารที่เป็นผลลัพธ์การค้นคืนสารสนเทศและ RA แทนเซตของเอกสารที่เป็นผลลัพธ์การค้นคืนสารสนเทศและเกี่ยวข้องกับข้อสอบถาม ผู้วิจัยจะสามารถคำนวณได้ด้วยสมการดังต่อไปนี้

$$\text{Recall} = \frac{|RA|}{|R|}$$

$$\text{Precision} = \frac{|RA|}{|A|}$$

2.5.2. การประยุกต์รำระลึกได้และค่าความแม่นยำ

จากหัวข้อที่กล่าวมา ในงานวิจัยการค้นคืนสารสนเทศและการขยายข้อสอบถาม ผู้วิจัยนำค่าระลึกได้และค่าความแม่นยำมาใช้ในการประเมินประสิทธิภาพการค้นคืนสารสนเทศ ด้วยรูปแบบการประเมินประสิทธิภาพการค้นคืนที่มีหลายรูปแบบ ผู้วิจัยจะเลือกกำหนดขั้นตอนวิธีที่เหมาะสมกับวัตถุประสงค์ของงานวิจัย รูปแบบการประเมินผลที่นำมาใช้ในงานวิจัยส่วนใหญ่มีรายละเอียดดังต่อไปนี้ (ศุภชัย ตั้งวงศ์ศานต์, 2551; Buttcher และคณะ, 2009; Perez-Aquera และ Araujo, 2008; Wang และ Niu, 2012)

1) **กราฟค่าระลอกใต้-ค่าความแม่นยำ** รูปแบบนี้จะนำเอกสารในรายการผลลัพธ์การค้นคืนสารสนเทศมาประเมินเอกสารตามลำดับรายการจากลำดับที่หนึ่งถึงลำดับสุดท้าย กำหนดให้ m แทนลำดับที่ของเอกสารในรายการผลลัพธ์การค้นคืนคำนวณค่าระลอกใต้และค่าความแม่นยำตั้งแต่เอกสารในรายการผลลัพธ์การค้นคืนลำดับที่ 1 ถึงลำดับที่ m แล้วนำค่าที่ได้มาจัดแสดงเป็นกราฟในกราฟนั้นแนวนอนเป็นค่าที่ระลอกใต้ ส่วนแนวตั้งเป็นค่าความแม่นยำในรูปแบบนี้นิยมนำมาประเมินข้อสอบถามที่นำมาใช้ในการค้นคืนสารสนเทศ ดังตัวอย่างต่อไปนี้

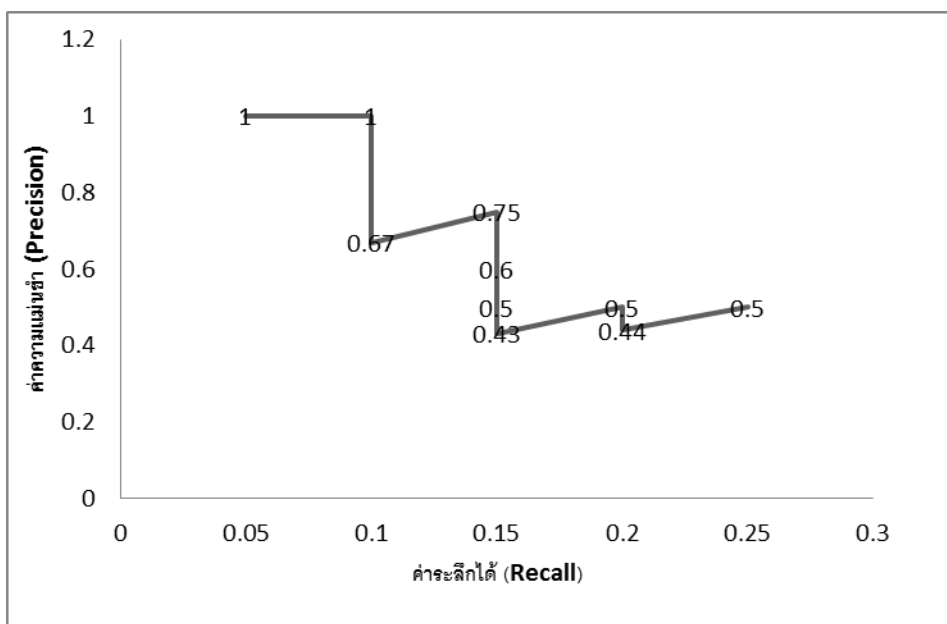
ตัวอย่างที่ 2.1 กำหนดให้ตารางที่ 2.6 แสดงผลลัพธ์การค้นคืนสารสนเทศ และมีจำนวนเอกสารในคอร์ปัสเอกสารที่เกี่ยวข้องกับข้อสอบถามทั้งหมดเท่ากับ 20 และจำนวนผลลัพธ์การค้นคืนเท่ากับ 10 จากตารางที่ 2.7 แสดงค่าระลอกใต้และค่าความแม่นยำในแต่ละลำดับเอกสาร จากตารางนี้ นำค่าระลอกใต้และค่าความแม่นยำในตารางไปสร้างกราฟ จะได้ดังรูปที่ 2.7

ตารางที่ 2.6 ตัวอย่างระดับมาตรฐานของค่าระลอกใต้

ช่วงค่าระลอกใต้	ค่าความแม่นยำ
0.00-0.10	1.00
0.10-0.20	0.75
0.20-0.30	0.50

ตารางที่ 2.7 ตัวอย่างกรณีการคำนวณค่าระลึกได้และค่าความแม่นยำ

ลำดับเอกสารที่	เกี่ยวข้องกับข้อสอบถามหรือไม่	ค่าระลึกได้ (Recall)	ค่าความแม่นยำ (Precision)
1	เกี่ยวข้อง	$\frac{1}{20} = 0.05$	$\frac{1}{1} = 1$
2	เกี่ยวข้อง	$\frac{2}{20} = 0.1$	$\frac{2}{2} = 1$
3	ไม่เกี่ยวข้อง	$\frac{2}{20} = 0.1$	$\frac{2}{3} = 0.67$
4	เกี่ยวข้อง	$\frac{3}{20} = 0.15$	$\frac{3}{4} = 0.75$
5	ไม่เกี่ยวข้อง	$\frac{3}{20} = 0.15$	$\frac{3}{5} = 0.6$
6	ไม่เกี่ยวข้อง	$\frac{3}{20} = 0.15$	$\frac{3}{6} = 0.5$
7	ไม่เกี่ยวข้อง	$\frac{3}{20} = 0.15$	$\frac{3}{7} = 0.43$
8	เกี่ยวข้อง	$\frac{4}{20} = 0.2$	$\frac{4}{8} = 0.5$
9	ไม่เกี่ยวข้อง	$\frac{4}{20} = 0.2$	$\frac{4}{9} = 0.44$
10	เกี่ยวข้อง	$\frac{5}{20} = 0.25$	$\frac{5}{10} = 0.5$



รูปที่ 2.7 ตัวอย่างกราฟค่าระลึกได้และค่าความแม่นยำ ดัดแปลงจาก ศุภชัย ตั้งวงศ์ศานต์ (2551)

2) **ระดับมาตรฐานของค่าระลึกได้**เป็นการเปรียบเทียบประสิทธิภาพของระบบการค้นคืนสารสนเทศที่ต่างกัันตั้งแต่ระบบ 2 ระบบขึ้นไป โดยจะเปรียบเทียบค่าความแม่นยำในช่วงค่าระลึกได้ที่ได้กำหนดไว้ โดยทั่วไปจะมีการกำหนดช่วงค่าระลึกได้ไว้ 10 ช่วง ได้แก่ ค่าระลึกได้ 0.00-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40, 0.40-0.50, 0.50-0.60, 0.60-0.70, 0.70-0.80, 0.80-0.90, และ 0.90-1.00 ตามลำดับ โดยจะให้ค่าความแม่นยำสูงสุดในแต่ละช่วงเป็นค่าความแม่นยำของแต่ละช่วงไปดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 2.2 จากตารางที่ 2.7 สามารถคำนวณหาค่าความแม่นยำได้ดังตารางที่ 2.7

3) **ค่าความแม่นยำของผลลัพธ์อาร์รายการแรก (Precision at R Document, P@R)** คือค่าความแม่นยำที่คำนวณจากผลลัพธ์การค้นคืนสารสนเทศจำนวนหนึ่งซึ่งผู้วิจัยได้กำหนดไว้ กำหนดให้ R แทนลำดับเอกสารในรายการผลลัพธ์การค้นคืนค่าความแม่นยำที่อาร์มีสมการการคำนวณดังต่อไปนี้

$$P@R = \frac{r}{R}$$

โดยที่ r แทนจำนวนเอกสารที่เกี่ยวข้องในรายการผลลัพธ์การค้นคืนตั้งแต่ลำดับที่ 1 ถึงลำดับที่ R ค่านี้จะนำมาใช้ในการเปรียบเทียบประสิทธิภาพการค้นคืนสารสนเทศที่ต่างระบบกัน มีงานวิจัยจำนวนมากที่นำมาใช้ในการเปรียบเทียบและประเมินผลลัพธ์การค้นคืน จะเห็นตัวอย่างการคำนวณได้จากตารางที่ 2.7 ค่าความแม่นยำที่ 5 ($P@5$) มีค่าเท่ากับ 0.6 ในขณะที่ค่าความแม่นยำที่ 9 ($P@9$) มีค่าเท่ากับ 0.44

4) **ค่าความแม่นยำอาร์ (R-Precision, P-R)** คือค่าความแม่นยำที่คำนวณจากจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามและปรากฏอยู่ในรายการผลลัพธ์การค้นคืน กำหนดโดยผู้ประเมิน กำหนดให้ R เป็นจำนวนเอกสารในรายการผลลัพธ์การค้นคืนลำดับต้นที่เกี่ยวข้องกับข้อสอบถามและ n เป็นจำนวนเอกสารในรายการผลลัพธ์การค้นคืนที่สามารถได้จำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามจำนวน R รายการ ค่าความแม่นยำอาร์จึงสามารถคำนวณได้ด้วยสมการต่อไปนี้

$$P - R = \frac{R}{n}$$

จะเห็นตัวอย่างการนำค่าความแม่นยำอาร์มาใช้ในการคำนวณจากรายการผลลัพธ์การค้นคืนจากตัวอย่างที่ 2.3 ดังต่อไปนี้

ตัวอย่างที่ 2.3 จากตารางที่ 2.7 ผู้วิจัยจะสามารถคำนวณค่าความแม่นยำสอง (2-Precision) และค่าความแม่นยำสี่ (4-Precision) ในการคำนวณค่าความแม่นยำสอง เมื่อกำหนดให้ค่าอาร์ (R) มีค่าเท่ากับ 2 จากตารางที่ 2.7 ลำดับเอกสารในรายการผลลัพธ์การค้นคืนที่สามารถพบจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามจำนวน 2 รายการคือลำดับที่ 2 ดังนั้น ค่าความแม่นยำสอง จึงมีค่าเท่ากับ $\frac{2}{2} = 1$ ส่วนในการคำนวณค่าความแม่นยำสาม จากตารางที่ 2.7 เมื่อผู้วิจัยกำหนดค่าอาร์มีค่าเท่ากับ 4 จากตารางที่ 2.7 ลำดับเอกสารในรายการผลลัพธ์การค้นคืนที่สามารถพบ

จำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามจำนวน 4 รายการ คือลำดับที่ 8 ค่าความแม่นยำจึงมีค่าเท่ากับ $\frac{4}{8} = 0.5$

5) **ค่าความแม่นยำเฉลี่ยกลาง (Mean Average Precision, MAP)** คือค่าเฉลี่ยของค่าความแม่นยำที่เอกสารลำดับที่ 1 ถึงลำดับสุดท้ายของรายการผลลัพธ์การค้นคืน โดยมีสมการการคำนวณดังต่อไปนี้

$$MAP = \frac{\sum_{i=1}^n P@i}{n}$$

โดยที่ n แทนจำนวนเอกสารทั้งหมดในรายการผลลัพธ์การค้นคืน และ $P@i$ แทนค่าความแม่นยำที่ i ค่าความแม่นยำเฉลี่ยกลางนี้ จะใช้ในการประเมินประสิทธิผลการค้นคืนโดยรวมของระบบการค้นคืนสารสนเทศ

6) **ค่าเฉลี่ยฮาร์โมนิระหว่างค่าระลึกได้และค่าความแม่นยำ (Harmonic mean of recall and precision, F-measure) หรือค่ามาตรวัดเอฟ (F-measure)** เป็นค่าที่ได้รับความนิยมอย่างมากในงานวิจัยการค้นคืนประสิทธิผลการค้นคืนสารสนเทศค่าเฉลี่ยฮาร์โมนิกนี้จะนำค่าระลึกได้และค่าความแม่นยำมาคำนวณด้วยสมการดังนี้

$$F - measure = \frac{(1 + B^2) \times p \times r}{B^2 \times p + r}$$

โดยที่ p คือค่าความแม่นยำของระบบ r คือค่าระลึกได้ของระบบและ B คือค่าความสำคัญระหว่างค่าระลึกได้และค่าความแม่นยำ นักวิจัยส่วนใหญ่มักจะกำหนดให้ค่า $B=1$ ค่าเฉลี่ยฮาร์โมนิกที่ได้จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยที่ค่าเท่ากับ 0 หมายถึงระบบไม่สามารถค้นคืนเอกสารที่เกี่ยวข้องกับข้อสอบถาม และค่าเฉลี่ยฮาร์โมนิกมีค่าเท่ากับ 1 แสดงว่าระบบสามารถค้นคืนเอกสารที่เกี่ยวข้องกับข้อสอบถาม ค่าเฉลี่ยนี้จึงบอกได้ถึงประสิทธิผลของระบบการค้นคืนสารสนเทศที่ทดสอบ

2.6. การนำการประเมินประสิทธิผลการค้นคืนสารสนเทศ ในงานวิจัย

หลังจากที่ผู้วิจัยได้นำเสนอการจัดกลุ่มการปรากฏร่วมกันของคำมาประยุกต์ใช้ในการขยายข้อสอบถามบนระบบการค้นคืนสารสนเทศภาษาไทย ผู้วิจัยจะต้องพัฒนาระบบการค้นคืนสารสนเทศ รวมทั้งระบบการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันของคำ เพื่อศึกษาถึงผลลัพธ์ที่ได้เมื่อนำแนวคิดที่นำเสนอมาประยุกต์ใช้ จากแนวคิดที่ได้นำเสนอนั้น ผู้วิจัยจะได้รายการผลลัพธ์การค้นคืนสารสนเทศจำนวนสองชุดต่อข้อสอบถามจากผู้ใช้หนึ่งข้อสอบถาม ได้แก่ รายการผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากผู้ใช้ และรายการผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามที่ขยายแล้ว ดังนั้นผู้วิจัยจึงสนใจนำผลการประเมินรายการผลลัพธ์การค้นคืนทั้งสองชุดมาเปรียบเทียบกัน เพื่อศึกษาถึงความแตกต่าง เมื่อนำการขยายข้อสอบถามด้วยวิธีการที่นำเสนอมาประยุกต์ใช้ในการค้นคืนสารสนเทศ จากงานวิจัยในอดีต ผู้วิจัยจึงต้องกำหนดค่าที่นำมาใช้ในการประเมินประสิทธิผลการค้นคืน

ผู้วิจัยจึงได้กำหนดค่าความแม่นยำที่อาร์ (Precision at R documents, P@R) มาใช้ในการประเมินประสิทธิผลการค้นคืน เพื่อประเมินถึงรายการผลลัพธ์การค้นคืนที่นำไปเสนอต่อผู้ค้นคืนสารสนเทศ (Collins-Thompson, 2009; Carpineto และคณะ, 2001; Massoudi และคณะ, 2011; Asfari และคณะ, 2010; Liu และคณะ, 2012; Arguello, 2008) ระบบการค้นคืนสารสนเทศที่นิยมมาใช้มากที่สุด คือโปรแกรมการค้นหา(Search Engine)(สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร, 2556) มีการสำรวจการใช้โปรแกรมค้นหา (Search engine)บนเครือข่ายอินเทอร์เน็ต พบว่าผู้ที่จะเข้าชมหน้าผลลัพธ์ของโปรแกรมค้นหา (Search Engine Result Page, SERP)เฉพาะหน้าแรกถึงร้อยละ 99 แต่ในหน้าผลลัพธ์ของโปรแกรมค้นหาตั้งแต่หน้าที่สอง มีผู้เข้าชมร้อยละ 3 และในหน้าผลลัพธ์ของโปรแกรมค้นหาลำดับถัดไป จำนวนผู้เข้าชมจะลดลง โดยในแต่ละหน้าผลลัพธ์ของโปรแกรมค้นหานั้น จะมีจำนวนผลลัพธ์ต่อหน้าเว็บเพจ (Results per page)จำนวน 10 ผลลัพธ์เป็นอย่างน้อย (Miller, 2010; StatOwl, 2012; Buscher และคณะ, 2012) ด้วยเหตุนี้ผู้ใช้อินเทอร์เน็ตส่วนใหญ่จะพิจารณาผลลัพธ์การค้นคืน

อย่างน้อย 10 ผลลัพธ์แรก อย่างไรก็ตาม เมื่อภารกิจการค้นหา (Search Task) เปลี่ยนแปลงความตั้งใจการค้นหาสารสนเทศจะเพิ่มขึ้น โดยเฉพาะอย่างยิ่ง จำนวนรายการผลลัพธ์ที่ผู้ใช้เรียกดูเพิ่มขึ้นเป็น 3 หน้า (Liu และคณะ, 2010; Kim, 2008) ดังนั้นผู้วิจัยจึงสนใจวิเคราะห์รายการผลลัพธ์การค้นหาจากจำนวนหน้าเว็บเพจ 3 หน้า จากจำนวนหน้าเว็บเพจที่แสดงรายการผลลัพธ์การค้นหาทั้งหมด ซึ่งหมายถึงรายการผลลัพธ์การค้นหาจำนวน 30 รายการจากรายการผลลัพธ์การค้นหาทั้งหมด ผู้วิจัยจึงกำหนดให้ค่าความแม่นยำที่อาร์ โดยกำหนดค่าอาร์เท่ากับ 10, 20 และ 30 ตามลำดับรายการผลลัพธ์ที่แสดงต่อผู้ใช้ เพื่อศึกษาถึงประสิทธิผลการค้นหาที่มีต่อผู้ใช้ระบบ

จากบทความที่กล่าวมานั้น แสดงถึงการประเมินประสิทธิผลการค้นหาด้วยรายการผลลัพธ์การค้นหาลำดับต้นที่นำเสนอต่อผู้ใช้ระบบ แต่เพื่อศึกษาถึงประสิทธิผลการค้นหาทั้งหมด ผู้วิจัยจึงสนใจนำเอกสารในรายการผลลัพธ์การค้นหาทั้งหมดมาประเมินประสิทธิผลการค้นหาสารสนเทศ ค่าที่นิยมนำมาใช้ในการประเมินการค้นหาสารสนเทศ รวมทั้งการขยายข้อสอบถามคือค่ามาตรวัดเอฟ (F-measure) ซึ่งเป็นค่าเฉลี่ยฮาร์โมนิก (Harmonic Means) ระหว่างค่าความแม่นยำและค่าที่ระลึกได้ที่คำนวณได้จากรายการผลลัพธ์การค้นหาทั้งหมด มีงานวิจัยจำนวนมากที่นิยมนำค่ามาตรวัดเอฟมาใช้ในการวิเคราะห์รายการผลลัพธ์การค้นหาที่ได้จากการนำระบบการขยายข้อสอบถามร่วมกับระบบการค้นหาสารสนเทศ Liu และคณะ (2011) ได้นำค่ามาตรวัดเอฟร่วมกับค่าวัดอื่นๆ ในการประเมินการนำการจัดกลุ่มรวมกับการขยายข้อสอบถาม Song และคณะ (2007) ได้นำค่ามาตรวัดเอฟมาวิเคราะห์การนำกฎความสัมพันธ์และอนโทโลยีมาประยุกต์ร่วมกับการขยายข้อสอบถามกับค่าความแม่นยำที่ยี่สิบ Lu และคณะ (2009) นำค่ามาตรวัดเอฟมาวิเคราะห์การนำการขยายข้อสอบถามบนระบบการค้นหาสารสนเทศทางการแพทย์ ดังนั้นงานวิจัยนี้จึงเลือกใช้ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรวัดเอฟ มาเลือกใช้ในการวิเคราะห์และประเมินประสิทธิผลการค้นหา เมื่อนำการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันของคำและการจัดกลุ่มแบบลำดับชั้นมาร่วมประยุกต์บนระบบการค้นหาสารสนเทศภาษาไทย

2.7. ชุดทดสอบ (Test Collection) สำหรับการทดสอบประสิทธิภาพการค้นคืนสารสนเทศ

จากการศึกษางานวิจัยการค้นคืนสารสนเทศ รวมทั้งการขยายข้อสอบถามร่วมกับการค้นคืนสารสนเทศรูปแบบต่างๆ ผู้วิจัยจะต้องประเมินประสิทธิภาพการค้นคืนสารสนเทศ เพื่อเปรียบเทียบหรือประเมินผลลัพธ์ที่ได้จากการทดลองการค้นคืนสารสนเทศด้วยการกำหนดค่าระลึกได้หรือค่าความแม่นยำ ในขั้นตอนนี้ ผู้วิจัยจะต้องใช้ชุดทดสอบ (Test Collection) เพื่อนำมาใช้ในการคำนวณค่าการประเมินประสิทธิภาพการค้นคืน

2.7.1. ลักษณะของชุดทดสอบ

ในการค้นคืนสารสนเทศนั้น ดังที่กล่าวในหัวข้อที่ 2.1 ผู้ใช้จะต้องกำหนดข้อสอบถามนำเข้าสู่ระบบการค้นคืน จากนั้นระบบจะแสดงผลลัพธ์เป็นรายการผลลัพธ์การค้นคืนออกมา เพื่อให้ผู้ใช้ได้ประเมินและเลือกดูเอกสารแต่ละรายการ ดังนั้นชุดทดสอบที่นำมาใช้ในการประเมินประสิทธิภาพการค้นคืนนั้น จึงประกอบด้วย (1) เอกสาร (Document) หรือสารสนเทศที่อยู่ในระบบการค้นคืน แตกต่างกันตามการขยายขอบเขตของสารสนเทศ (2) ข้อสอบถาม (Query) ที่ผู้ใช้ได้กำหนดสำหรับการค้นคืน ข้อสอบถามที่นำมาใช้ในระบบการค้นคืนนั้นมีจำนวนมากและหลากหลายรูปแบบ ตามความต้องการข้อมูลของผู้ใช้ และ (3) ผลการตัดสินความเกี่ยวข้อง (Relevance Judgment) ระหว่างข้อสอบถามกับเอกสารจากผู้ใช้ ซึ่งผู้ใช้สามารถประเมินได้จากรายการผลลัพธ์การค้นคืนที่ระบบได้นำเสนอต่อผู้ใช้

มีหน่วยงานต่างๆ ที่ให้บริการชุดทดสอบเพื่อนำมาใช้ในการงานวิจัยการค้นคืนสารสนเทศมากมาย ชุดทดสอบหนึ่งที่ผู้วิจัยนิยมนำมาใช้ในการทดสอบระบบการค้นคืนสารสนเทศ คือ ชุดข้อมูลการประชุมร่วมการค้นคืนข้อความ (Text Retrieval Conference, TREC) การประชุมนี้ได้รับการสนับสนุนจากสถาบันมาตรฐานและเทคโนโลยีแห่งชาติ (National Institute of Standards and Technology, NIST) และสำนักงานวิจัยและพัฒนาขั้นสูง (Advanced Research and Development Activity, ARDA) ประเทศสหรัฐอเมริกา นอกจากชุดทดสอบนี้แล้ว ยังมีชุดข้อมูลการประชุมร่วมการประเมินแบบข้ามภาษา (Cross-Language Evaluation Forum, CLEF) และ

ชุดทดสอบนิยมนำมาใช้ในการค้นคืนสารสนเทศ (NII Test Collection for IR Systems, NTCIR) ที่นิยมนำมาใช้ในการทดสอบประสิทธิภาพของขั้นตอนการค้นคืนสารสนเทศ ชุดทดสอบดังกล่าวมีจำนวนมากและแตกต่างกันในแต่ละองค์ความรู้และช่วงระยะเวลาในการเก็บข้อมูล

ในชุดทดสอบ นอกจากเอกสารจำนวนมากและข้อสอบถามจำนวนมาก ผลการตัดสินความเกี่ยวข้อง ก็เป็นส่วนหนึ่งที่สำคัญต่อการประเมินประสิทธิภาพการค้นคืนสารสนเทศอย่างมาก ในชุดทดสอบนั้น ผลการตัดสินความเกี่ยวข้อง จะกำหนดโดยผู้เชี่ยวชาญเฉพาะด้าน และมีความแตกต่างกัน จึงมีงานวิจัยจำนวนหนึ่งที่ได้ศึกษาการกำหนดผลการตัดสินความเกี่ยวข้อง Xu and Chen (2005,2006) ศึกษาองค์ประกอบที่ผู้ใช้ได้ตัดสินใจถึงความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร ได้แก่ ความน่าสนใจของหัวข้อ (Topicality) ความแปลกใหม่ (Novelty) ความเชื่อถือได้ (reliability) ความสามารถในการเข้าใจ (understandability) และขอบเขต (scope) ส่วน Zhang (2002) ได้ศึกษาการตัดสินความเกี่ยวข้องด้วยขั้นตอนวิธีความคิดเห็นของคนส่วนใหญ่ในกลุ่ม (Group Consensus Method) นอกจากนี้แล้ว ยังมีการกำหนดระดับความเกี่ยวข้อง (Relevance Level) ซึ่งเป็นการกำหนดจำนวนตัวเลือกแบบสเกลลิเกิร์ต (Likert Scale) โดยกำหนดให้ค่าที่น้อยที่สุดหมายถึงไม่มีความเกี่ยวข้องเลย และค่าที่มากที่สุดหมายถึงมีความเกี่ยวข้องมากที่สุด เพื่อนำมาใช้ในการประเมินความเกี่ยวข้อง งานวิจัยในอดีตนิยมกำหนดระดับความเกี่ยวข้องสองระดับ ซึ่งหมายถึงเกี่ยวข้องและไม่เกี่ยวข้อง ตามลำดับ Wu และ McClean (2006) และ Oard และคณะ (2004) ได้นำเสนอการกำหนดระดับความเกี่ยวข้องจำนวนมากกว่าสองระดับ อาทิ ระดับความเกี่ยวข้อง สาม ห้า หรือเจ็ดระดับ

2.7.2. การสร้างชุดทดสอบ

ถึงแม้จะมีชุดทดสอบที่นำมาใช้ในการประเมินประสิทธิภาพการค้นคืนจำนวนมาก ชุดทดสอบดังกล่าวนั้นไม่สามารถนำมาใช้ในการประเมินประสิทธิภาพการค้นคืนทั้งหมด เนื่องจากลักษณะของการค้นคืนสารสนเทศแตกต่างกันไป โดยเฉพาะอย่างยิ่งเมื่อขอบเขตของสารสนเทศและการค้นคืนสารสนเทศเริ่มขยายกว้างขึ้นไป มีงานวิจัยจำนวนไม่น้อยที่นำเสนอการสร้างชุดทดสอบ เพื่อตอบสนองความต้องการการประเมินประสิทธิภาพการค้นคืนรูปแบบต่างๆ Oard และคณะ (2004) ได้สร้างชุดทดสอบสำหรับคำพูดเชิงสนทนาตามปกติ

(Spontaneous Conversational Speech) โดยมีสารสนเทศเป็นแฟ้มข้อมูลเสียงการสัมภาษณ์ สร้างชุดหัวข้อ (Topic) และกำหนดผลการตัดสินความเกี่ยวข้องระหว่างแฟ้มข้อมูลเสียงและหัวข้อ โดยใช้ 5 จุดมาตราส่วน (Scale point) ตั้งแต่ 0 ถึง 4 ทั้งนี้ 0 หมายถึงไม่เกี่ยวข้องเลย และ 4 หมายถึงเกี่ยวข้องกันมากที่สุด โดยการสร้างระบบการค้นคืนสารสนเทศ เพื่อให้ผู้ประเมินได้ ประเมินความเกี่ยวข้อง Macdonald และ Qunis (2006) ได้สร้างชุดทดสอบเว็บส่วนบุคคล (Blog) ซึ่งในแต่ละหน้าเว็บจะมีหัวข้อ (Topic) และประกาศ (Post) เริ่มแรก จะคัดเลือกหน้าเว็บส่วนบุคคลและดึงข้อมูลจากหน้าเว็บออกมา เพื่อสร้างเป็นชุดสารสนเทศสำหรับการค้นคืน จากนั้นจึง วิเคราะห์ชุดสารสนเทศที่นำมากับหัวข้อภายในเว็บส่วนบุคคล Ritchie, Teufel และ Robertson (2006) ได้สร้างชุดทดสอบสำหรับการทดลองการค้นคืนสารสนเทศด้านหมายเรียกบุคคล โดยนำ หลักการสร้างชุดทดสอบ Cranfield 2 Test Collection) ซึ่งจะสอบถามความ เกี่ยวข้องระหว่างเอกสารและข้อสอบถามด้วยการสอบถามผู้จัดทำเอกสาร อีกทั้งกำหนดให้มี ระดับความเกี่ยวข้องทั้งหมดสี่ระดับ โดยเริ่มต้นจาก 1 ถึง 4 รูปแบบการสร้างชุดทดสอบที่นิยมจะ กำหนดหัวข้อ (Topic) เพื่อสร้างความสัมพันธ์ระหว่างข้อสอบถามและเอกสารหรือสารสนเทศที่ สร้างขึ้น Efron (2009) ได้นำลักษณะของข้อสอบถาม (Query Aspect) ได้สร้างความเกี่ยวข้อง ระหว่างข้อสอบถามและเอกสาร

จากงานวิจัยในอดีตดังกล่าวข้างต้น ผู้วิจัยใช้เอกสาร และผู้ประเมินความเกี่ยวข้องใน การรวบรวมข้อมูลสำหรับการสร้างชุดทดสอบภาษาไทย ซึ่งมีขั้นตอน คือ (1) รวบรวมเอกสาร (2) คัดเลือกผู้ประเมินความเกี่ยวข้อง (3) กำหนดให้ผู้ประเมินความเกี่ยวข้อง กำหนดข้อสอบถามหนึ่ง ข้อสอบถาม (4) ศึกษาเอกสาร และประเมินความเกี่ยวข้องระหว่างเอกสารและข้อสอบถาม (5) กำหนดดัชนี (Index)จากรายการเอกสาร และสรุปผลการตัดสินความเกี่ยวข้องระหว่างข้อ สอบถาม (Oard และคณะ, 2004;Ritchie, Teufel และ Robertson, 2006; Macdonald และ Qunis, 2006) ทั้งนี้ ผู้วิจัยต้องกำหนดข้อมูล รวมทั้งลักษณะของข้อมูล ที่นำมาใช้ในการสร้างชุด ทดสอบ ดังต่อไปนี้

(1) ลักษณะของผู้กำหนดข้อสอบถามหรือผู้ประเมินความเกี่ยวข้อง

มีหน่วยงานต่างๆ ของภาครัฐและเอกชนที่ได้สำรวจพฤติกรรมการใช้อินเทอร์เน็ตของคน ไทย โดยในการสำรวจนั้น จะมีการสอบถามข้อมูลทางประชากรศาสตร์ (Demographics) รวมทั้ง อายุของผู้ร่วมสำรวจด้วย สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (2553)ได้สำรวจ

ลักษณะการใช้อินเทอร์เน็ต พบว่า ผู้ร่วมตอบแบบสอบถามช่วงอายุ 20-29 ปี จะนิยมใช้อินเทอร์เน็ตมากที่สุด ถึงร้อยละ 40 โดยประมาณ ในการค้นคืนสารสนเทศสำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์(2556) ได้สำรวจลักษณะการใช้อินเทอร์เน็ต พบว่า ผู้ร่วมตอบแบบสอบถามช่วงอายุ 20-34 ปี ซึ่งมีจำนวนชั่วโมงการใช้งานอินเทอร์เน็ตมากที่สุด นิยมใช้อินเทอร์เน็ตในการค้นคืนสารสนเทศจะเห็นได้ว่าประชากรชาวไทยช่วงอายุ 20-34 ปี มีความรู้ความสามารถในการใช้ระบบการค้นคืนสารสนเทศบนเครือข่ายอินเทอร์เน็ตผู้วิจัยจึงเลือกกำหนดช่วงอายุ 20-34 ปี เป็นผู้ประเมินความเกี่ยวข้อง

นอกจากช่วงอายุของผู้ร่วมตอบแบบสอบถามที่ได้กล่าวมาข้างต้นแล้ว อาชีพของผู้ร่วมตอบแบบสอบถาม ถือเป็นส่วนหนึ่งของการสำรวจพฤติกรรมการใช้อินเทอร์เน็ตได้กล่าวถึง เพื่อนำมาใช้ในการศึกษาพฤติกรรมของผู้ใช้เครือข่ายอินเทอร์เน็ต ศูนย์วิจัยนวัตกรรมอินเทอร์เน็ตไทย (2555) ได้สำรวจพฤติกรรมการใช้อินเทอร์เน็ตของผู้ใช้ชาวไทยในช่วงปี 2554-2555 มีผู้ร่วมตอบแบบสอบถาม ร้อยละ 34.2 เป็นนักเรียนนักศึกษา ขณะที่ผลการสำรวจของสำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (2556) พบว่า นักเรียนนักศึกษามีจำนวนชั่วโมงการใช้งานอินเทอร์เน็ตต่อสัปดาห์ 34.2 ชั่วโมง ซึ่งใกล้เคียงกับค่าเฉลี่ยของจำนวนชั่วโมงการใช้งานอินเทอร์เน็ตต่อสัปดาห์ของผู้ร่วมตอบแบบสำรวจทั้งหมด 32.3 ชั่วโมง

(2) ลักษณะของข้อสอบถาม

มีหน่วยงานหลายแห่งที่ได้สำรวจการกำหนดข้อสอบถาม A Trellian Company (2013) ได้สำรวจจำนวนคำในข้อสอบถามที่ใช้ค้นคืนสารสนเทศ โดยจำแนกด้วยภาษา รูปที่ 2.9 แสดงถึงผลการสำรวจของ A Trellian Company (2013) เกี่ยวกับสัดส่วนของการเข้าชมหน้าเว็บเพจที่ได้จากรายการผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามที่มีจำนวนคำในข้อสอบถามที่แตกต่างกัน โดยพิจารณาจากสัดส่วนการคลิกเลือกชมหน้าเว็บเพจ (Organic), สัดส่วนการคลิกโฆษณา (PPC) และสัดส่วนการคลิกเลือกชมทั้งหมด (Total) ในหน้าผลลัพธ์โปรแกรมค้นหา (Search Engine Result Page, SERP) จากรูปที่ 2.8 จะเห็นได้ว่าเมื่อกำหนดข้อสอบถามที่มีจำนวนคำในข้อสอบถามเท่ากับเจ็ดหรือแปดคำ จะได้สัดส่วนการเข้าชมหน้าเว็บเพจจากรายการผลลัพธ์การค้นคืนที่มีค่าสูงสุด ผู้วิจัยจึงกำหนดให้ผู้ประเมินความเกี่ยวข้องทุกคนต้องกำหนดข้อสอบถามที่มีจำนวนคำในข้อสอบถามเท่ากับเจ็ดหรือแปดคำ

Words:	Total(%)	Organic(%)	PPC(%)
Total	34.76%	33.91%	0.85%
1	21.74%	21.41%	0.33%
2	71.36%	69.81%	1.55%
3	65.84%	63.82%	2.02%
4	63.43%	60.57%	2.86%
5	56.83%	54.08%	2.75%
6	55.48%	52.79%	2.69%
7	112.98%	108.26%	4.72%
8	111.27%	106.93%	4.34%
9	96.75%	93.06%	3.69%
10+	22.47%	20.99%	1.48%

รูปที่ 2.8 ผลการสำรวจการเข้าชมหน้าเว็บเพจจำแนกตามจำนวนคำในข้อสอบถามที่ผู้ใช้กำหนด คัดลอก
จาก A Trellian Company (2013)

(3) ลักษณะของเอกสาร

ในการสร้างชุดทดสอบหรือการค้นคืนสารสนเทศนั้น เอกสารหรือสารสนเทศที่นำมาใช้นั้น จะมีลักษณะของรูปแบบ ภาษาและขอบเขตที่แตกต่างกัน ตามวัตถุประสงค์ของงานวิจัย สารสนเทศที่นิยมนำมาใช้ในงานวิจัยในอดีต คือ สารสนเทศที่เป็นข้อความ อย่างไรก็ตามในการสร้างชุดทดสอบนั้น ผู้วิจัยจะต้องกำหนดระยะเวลาในการประเมินเอกสารหนึ่งฉบับ เพื่อควบคุมปัจจัยที่มีผลต่อการประเมินความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร รวมทั้งระยะเวลาในการประเมินเอกสาร การกำหนดเอกสารที่ง่ายต่อการทำความเข้าใจจึงมีผลต่อการประเมินความเกี่ยวข้องของ งานวิจัยในอดีตนั้นนิยมนำเอกสารที่กำหนดหัวข้อย่อย ดังตัวอย่างในรูปที่ 2.9 และรูปที่ 2.10

```

<DOC>
<FEEDNO>BLOG06-feed-001002</FEEDNO>
<FEEDURL>http://www.henrikbennetsen.com/wp-rss2.php#</FEEDURL>
<BLOGHPNO>BLOG06-bloghp-001002</BLOGHPNO>
<BLOGHPURL>http://www.henrikbennetsen.com/#</BLOGHPURL>
<PERMALINKS>
http://www.henrikbennetsen.com/?p=85#
BLOG06-20051206-012-0001942855
http://www.henrikbennetsen.com/?p=83#
BLOG06-20051206-012-0001954556
</PERMALINKS>
<DOCHDR>
http://www.henrikbennetsen.com/wp-rss2.php# 200512663735 25595
Date: Tue, 06 Dec 2005 20:37:26 GMT
Server: Apache
Content-Type: text/xml; charset=UTF-8
Last-Modified: Tue, 06 Dec 2005 18:55:22 GMT
X-Pingback: http://www.henrikbennetsen.com/xmlrpc.php

</DOCHDR>
<?xml version="1.0" encoding="UTF-8"?>
<!-- generator="wordpress/1.5" -->
<rss version="2.0"
...
</DOC>

```

รูปที่ 2.9 ตัวอย่างสารสนเทศที่นำมาใช้ในการสร้างชุดทดสอบ (MacDonald, 2006)

```

<DOC>
<DOCNO>VHF10325-1080.34</DOCNO>
<INTERVIEWDATA>Alexej H...</INTERVIEWDATA>
<ENGLISHMANUKEYWORD>social relations in prisons</ENGLISHMANUKEYWORD>
<CZECHMANUKEYWORD>společenská vztahy v vězení</CZECHMANUKEYWORD>
<ASRSYSTEM>2006</ASRSYSTEM>
<CHANNEL>left</CHANNEL>
<ASRTEXT> PĚKNĚ TAKŽE NĚKDY I TY I TY HLÍDAČI NE </s> <s> TO MYSLÍ
POSLOUCHALI POSLOUCHALI TO CO ZPÍVÁM HO NECHAL CELKEM ASI TO ZPÍVAL ONI NÁS JAKO NE
</s> <s> KDE MÁTE NA MYSLI </s> <s> NO TAM JSME
BYLI ASI </s> <s> DO JARA ROKU ČTYŘICET </s> <s> NO TO SEM NÁS
VOZILI NA NA NA SILNICI A NÁM SE Podařilo NĚKOLIKA ...</ASRTEXT>
<ENGLISHAUTOKEYWORD>fate of loved ones | living conditions in the
camps | Poland 1941 (June 21) - 1944 (July 21) | Germany 1945
(January 1 - May 7) ...</ENGLISHAUTOKEYWORD>
<CZECHAUTOKEYWORD>osudy blízkých | životní podmínky v táborech |
Polsko 1941 (21. červen) - 1944 (21. červen) | Německo 1945
(1. leden - 7. květen) ...</CZECHAUTOKEYWORD>
</DOC>

```

รูปที่ 2.10 ตัวอย่างสารสนเทศที่นำมาใช้ในการสร้างชุดทดสอบ คัดลอกจาก (Ircing และคณะ, 2007)

(4) รูปแบบการประเมินความเกี่ยวข้อง

เมื่อผู้วิจัยสามารถรวบรวมเอกสารและผู้ประเมินแล้ว ผู้วิจัยจะต้องกำหนดรูปแบบการศึกษาและประเมินเอกสาร โดยที่ลดปัจจัยอื่นๆ ที่มีผลต่อผลการตัดสินความเกี่ยวข้อง มีงานวิจัยจำนวนหนึ่งที่ได้กำหนดจำนวนเอกสารที่ส่งผลให้ระยะเวลาในการประเมินมากขึ้น Oard

และคณะ (2004) ได้ให้นักเรียนสี่คนในการประเมินสารสนเทศจำนวน 404 รายการด้วยระยะเวลา 700 ชั่วโมงในช่วงระยะเวลามากกว่า 3 เดือน โดยที่สารสนเทศที่ใช้เป็นข้อมูลเสียงบันทึกบทสัมภาษณ์ พร้อมบทสรุปการสัมภาษณ์สั้น Ritchie และ คณะ (2006) ได้กำหนดให้รวบรวมเอกสารงานวิจัยที่จะนำมาใช้ในการสร้างชุดทดสอบและรายการข้อสอบถาม จากนั้นจึงนำรายการข้อสอบถามส่งไปให้ผู้เขียนงานวิจัย เพื่อให้ผู้เขียนงานวิจัยได้ประเมินรายการข้อสอบถามว่าเกี่ยวข้องกับงานวิจัยของตนเองหรือไม่ แล้วส่งผลการประเมินกลับมาให้ผู้วิจัย ข้อดีของงานวิจัยนี้คือ ผู้ประเมินข้อสอบถามนั้นมีความรู้ความเข้าใจในเอกสาร ซึ่งหมายถึงงานวิจัยของผู้ประเมินเอง แต่ข้อเสียคืออาจจะต้องใช้ระยะเวลาในการประเมินที่ต่างจากกัน ตามความสะดวกของผู้วิจัย

2.7. เครื่องมือการค้นคืนสารสนเทศลูชัน

เครื่องมือการค้นคืนสารสนเทศลูชัน เป็นเครื่องมือการค้นคืนสารสนเทศ โดยใช้รูปแบบปริภูมิเวกเตอร์ (Cui, Chen และ Li, 2012) ผู้วิจัยในอดีตนิยมนำเครื่องมือนี้มาใช้ในการนำเสนอแนวคิดสำหรับการพัฒนาระบบการค้นคืนสารสนเทศเนื่องจาก (1) สามารถรองรับภาษาอื่นๆ ที่นอกเหนือจากภาษาอังกฤษ Zhang, Zhang และ Chen (2005) ได้นำเสนอระบบการค้นคืนสารสนเทศแบบข้ามภาษาระหว่างภาษาจีนและภาษาอังกฤษ โดยนำเสนอการเปลี่ยนข้อสอบถามภาษาจีนเป็นข้อสอบถามภาษาอังกฤษ จากนั้นจึงนำข้อสอบถามที่ได้เข้าสู่ระบบการค้นคืนสารสนเทศ ซึ่งได้ใช้เครื่องมือการค้นคืนสารสนเทศลูชัน ทั้งนี้ได้นำเสนอการนำการป้อนความเกี่ยวข้องย้อนกลับ (Relevance Feedback) มาประยุกต์ร่วม เพื่อประสิทธิภาพการค้นคืน (2) สามารถวิเคราะห์และกำหนดดัชนี ในการค้นคืนสารสนเทศได้ เมื่อนำเอกสารเข้าสู่ระบบการค้นคืนสารสนเทศ Chen และคณะ (2012) ได้นำเสนอการนำดัชนีแบบแจกจ่าย (Distributed Index) มาประยุกต์ใช้กับเครื่องมือการค้นคืนสารสนเทศลูชัน ซึ่งใช้ดัชนีผกผัน (Inverted Index) เพื่อเพิ่มประสิทธิภาพการค้นคืน (3) สามารถรองรับสารสนเทศได้ นอกเหนือจากสารสนเทศที่เป็นลายลักษณ์อักษร Lux และ Savvas (2008) ได้นำเครื่องมือการค้นคืนสารสนเทศลูชันมาใช้ในการพัฒนาระบบการค้นคืนรูปภาพ (4) สามารถรองรับสารสนเทศที่มีขนาดใหญ่ได้ Tianyuan, Meina และ Xiaoqi (2010) ได้นำเครื่องมือการค้นคืนสารสนเทศลูชันมาใช้ในการค้นคืนสารสนเทศบนระบบบูรณาการข้อมูลที่แตกต่างกัน (5) รูปแบบการค้นคืนสารสนเทศที่ใช้เป็นรูปแบบปริภูมิเวกเตอร์ (Vector Space Model) Turney และ Pantel (2010) ได้นำเสนอการนำรูปแบบปริภูมิเวกเตอร์ในระบบการค้นคืนสารสนเทศในปัจจุบัน ทั้งนี้ได้กล่าวถึงนำเครื่องมือการค้นคืนสารสนเทศลูชันมาใช้ในการพัฒนาระบบการค้นคืนสารสนเทศ และ (6) สามารถนำการขยายข้อสอบถามมา

ประยุกต์ใช้ได้ Zhu คณະ (2006) และ Xu และคณະ (2006) ได้นำเครื่องมือการค้นคืนสารสนเทศสูงขึ้นมาประกอบการศึกษาการนำการวิเคราะห์เฉพาะส่วน (Local Analysis), การวิเคราะห์ครอบคลุม (Global Analysis) และ การกำหนดค่าน้ำหนักของคำเชิงออนโทโลยี

2.8. เครื่องมือการทำเหมืองข้อมูลเวก้า

เครื่องมือการทำเหมืองข้อมูลเวก้า (WEKA, The Waikato Environment for Knowledge Analysis) เป็นชุดคำสั่งภาษาจาวาที่ประกอบด้วยขั้นตอนการทำเหมืองข้อมูล (Holmes, Donkin และ Written, 1994; Written และคณະ, 1999) ที่พัฒนาโดยมหาวิทยาลัยแห่งไวคาโต (The University of Waikato) เครื่องมือนี้สามารถนำมาใช้ได้ในการทำเหมืองข้อมูล เพื่อศึกษาลักษณะหรือทิศทางของข้อมูลที่ได้รับ และสามารถนำมาใช้ในการพัฒนาระบบต่างๆ เพื่อให้ผู้วิจัยหรือผู้พัฒนาอื่นสามารถนำขั้นตอนการทำเหมืองข้อมูลไปประยุกต์ใช้กับระบบการประมวลผลต่างๆ หรือพัฒนาขั้นตอนการทำเหมืองข้อมูลเพิ่มเติม เพื่อรองรับกับลักษณะของสารสนเทศต่างๆ มีงานวิจัยจำนวนมากที่นำเครื่องมือการทำเหมืองข้อมูลเวก้ามาใช้ในการศึกษาและนำเสนอแนวคิดการทำเหมืองข้อมูลใหม่ Jun-qiang และคณະ (2012) ได้นำเครื่องมือการทำเหมืองข้อมูลเวก้ามาใช้ในการพัฒนาขั้นตอนการจัดกลุ่มแบบคลุมเครือ Shekarpour และคณະ (2013) และ Araujo และ Perez-Iglesias (2010) ได้นำเครื่องมือการทำเหมืองข้อมูลเวก้ามาใช้ในการวิเคราะห์คำสำหรับการขยายข้อสอบถามด้วยขั้นตอนซัพพอร์ตเวกเตอร์แมกซิม Aboulhaga และ Clarke (2012) ได้นำเสนอการขยายข้อสอบถามด้วยขั้นตอนการทำเหมืองเซตถี่ (Frequent Itemset Mining) ซึ่งได้นำเครื่องมือการทำเหมืองข้อมูลเวก้าอย่างไรก็ตาม การใช้เครื่องมือการทำเหมืองข้อมูลนั้น นอกเหนือจากการเลือกขั้นตอนวิธีการทำเหมืองข้อมูลแล้ว ผู้วิจัยจะต้องกำหนดรูปแบบการทำเหมืองข้อมูลด้วย Sharma, Bajpai และ Litoriya (2012) และ Chaudhari และ Parikh (2012) ได้นำเสนอขั้นตอนวิธีการจัดกลุ่มต่างๆ ที่มีอยู่ในเครื่องมือการทำเหมืองข้อมูลเวก้า จะเห็นได้ว่าในการเลือกขั้นตอนวิธีการจัดกลุ่มนั้น ผู้วิจัยจะต้องกำหนดรูปแบบการจัดกลุ่มที่จะต้องนำมาใช้ อาทิ จำนวนกลุ่ม สมการการวัดค่าความเหมือน หรือสมการเกณฑ์การเชื่อมโยง

บทที่ 3 ระเบียบวิธีวิจัย

3.1. แนวทางการวิจัยและสมมติฐานงานวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงประยุกต์ (Applied Research) ที่ต้องการนำการจัดกลุ่มเอกสารมาประยุกต์ใช้ในการขยายข้อสอบถาม

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบ (1) ค่าความแม่นยำที่สิบ (Precision at 10 documents) (2) ค่าความแม่นยำที่ยี่สิบ (Precision at 20 documents) (3) ค่าความแม่นยำที่สามสิบ (Precision at 30 documents) และ (4) ค่ามาตรวัดเอฟ (F-measure) ที่ได้จาก (1) รายการผลลัพธ์การค้นหาสารสนเทศ ที่ได้จากการค้นหาสารสนเทศด้วยข้อสอบถามจากผู้ใช้ หรือก่อนการนำการขยายข้อสอบถามมาประยุกต์ใช้ กับ (2) รายการผลลัพธ์การค้นหาสารสนเทศที่ได้จากการค้นหาสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม หรือหลังการนำการขยายข้อสอบถามมาประยุกต์ใช้ ผู้วิจัยจึงตั้งสมมติฐานไว้ดังต่อไปนี้

3.1.1. การเปรียบเทียบค่าความแม่นยำจำนวนสิบรายการแรกระหว่างผลลัพธ์การค้นหาสารสนเทศที่ใช้ข้อสอบถามจากผู้ใช้และใช้ข้อสอบถามที่ได้จากการขยายข้อสอบถาม

ในงานวิจัยนี้ ผู้วิจัยจะต้องเปรียบเทียบค่าความแม่นยำที่สิบที่คำนวณจากผลลัพธ์การค้นหาที่ได้จากข้อสอบถามจากผู้ใช้ และผลลัพธ์การค้นหาที่ได้จากข้อสอบถามจากการขยายข้อสอบถาม จากงานวิจัยในอดีตพบว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นหาสารสนเทศนั้น ประสิทธิภาพการค้นหาจะเพิ่มขึ้น ดังนั้นการตั้งสมมติฐานแย้ง (H_1) จึงไม่สามารถกำหนดทิศทางได้ สมมติฐานที่ว่า คือ

H_0 : ผลลัพธ์การค้นหาเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนสิบรายการแรก ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นหาเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

H_1 : ผลลัพธ์การค้นหาเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนสิบรายการแรก **สูงกว่า** ผลลัพธ์การค้นหาเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

3.1.2. การเปรียบเทียบค่าความแม่นยำจำนวนยี่สิบรายการแรกระหว่างผลลัพธ์การค้นคืนสารสนเทศที่ใช้ข้อสอบถามจากผู้ใช้และใช้ข้อสอบถามที่ได้จากการขยายข้อสอบถาม

ในงานวิจัยนี้ ผู้วิจัยจะต้องเปรียบเทียบค่าความแม่นยำที่ยี่สิบที่คำนวณจากผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากผู้ใช้ และผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากการขยายข้อสอบถาม จากงานวิจัยในอดีตพบว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศนั้น ประสิทธิภาพการค้นคืนจะเพิ่มขึ้น ดังนั้นการตั้งสมมติฐานแย้ง (H_1) จึงไม่สามารถกำหนดทิศทางได้ สมมติฐานที่ว่า คือ

H_0 : ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนยี่สิบรายการแรก ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

H_1 : ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนยี่สิบรายการแรก **สูงกว่า** ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

3.1.3. การเปรียบเทียบค่าความแม่นยำจำนวนสามสิบรายการแรกระหว่างผลลัพธ์การค้นคืนสารสนเทศที่ใช้ข้อสอบถามจากผู้ใช้และใช้ข้อสอบถามที่ได้จากการขยายข้อสอบถาม

ในงานวิจัยนี้ ผู้วิจัยจะต้องเปรียบเทียบค่าความแม่นยำที่สามสิบที่คำนวณจากผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากผู้ใช้ และผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากการขยายข้อสอบถาม จากงานวิจัยในอดีตพบว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศนั้น ประสิทธิภาพการค้นคืนจะเพิ่มขึ้น ดังนั้นการตั้งสมมติฐานแย้ง (H_1) จึงไม่สามารถกำหนดทิศทางได้ สมมติฐานที่ว่า คือ

H_0 : ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนสามสิบรายการแรก ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

H_1 : ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำจำนวนสามสิบรายการแรก **สูงกว่า** ผลลัพธ์การค้นคืนเอกสารเมื่อใช้ข้อสอบถามจากผู้ใช้

3.1.4. การเปรียบเทียบค่ามาตรฐานวัดเอพระหว่างผลลัพธ์การค้นคืนสารสนเทศที่ใช้ข้อสอบถามจากผู้ใช้และใช้ข้อสอบถามที่ได้จากการขยายข้อสอบถาม

ในงานวิจัยนี้ ผู้วิจัยจะต้องเปรียบเทียบค่ามาตรฐานวัดเอพที่คำนวณจากผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากผู้ใช้ และผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากการขยายข้อสอบถาม จากงานวิจัยในอดีตพบว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศนั้น ประสิทธิภาพการค้นคืนจะเพิ่มขึ้น ดังนั้นการตั้งสมมติฐานแย้ง (H_1) จึงไม่สามารถกำหนดทิศทางได้ สมมติฐานที่ว่า คือ

H_0 : ระบบการค้นคืนสารสนเทศที่มีการขยายข้อสอบถาม จะมีค่าเฉลี่ยฮาร์โมนิระหว่างค่าระลึกได้และค่าความแม่นยำระหว่างระบบ น้อยกว่าหรือเท่ากับระบบการค้นคืนสารสนเทศที่ไม่มีการขยายข้อสอบถาม

H_1 : ระบบการค้นคืนสารสนเทศที่มีการขยายข้อสอบถาม จะมีค่าเฉลี่ยฮาร์โมนิระหว่างค่าระลึกได้และค่าความแม่นยำระหว่างระบบ **สูงกว่า** ระบบการค้นคืนสารสนเทศที่ไม่มีการขยายข้อสอบถาม

3.2. ประชากรการเลือกกลุ่มตัวอย่างและจำนวนตัวอย่าง

งานวิจัยนี้ต้องการทดสอบการนำการจัดกลุ่มการปรากฏร่วมกันของค่ามาประยุกต์ใช้ในการขยายข้อสอบถาม โดยการวัดค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยี่สืบ ค่าความแม่นยำที่สามสืบ และค่ามาตรฐานวัด ซึ่งการคำนวณค่านี้ จำเป็นต้องมีการนำชุดทดสอบ (Test Collection) ด้วย ซึ่งประกอบด้วยข้อสอบถาม ชุดเอกสาร และความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร ดังนั้นเพื่อตอบวัตถุประสงค์ของงานวิจัย จึงจำเป็นต้องสร้างชุดทดสอบสำหรับการทดสอบ ดังนั้นงานวิจัยนี้ต้องใช้ (1) ชุดเอกสารที่นำมาใช้ในการสร้างชุดทดสอบและการค้นคืนสารสนเทศ และ (2) ผู้ประเมินความเกี่ยวข้องระหว่าง(Relevance Assessor) แต่เนื่องจากผู้ประเมินความเกี่ยวข้องและชุดเอกสารเป็นประชากรขนาดใหญ่ จึงไม่สามารถเก็บประชากรได้ทุกหน่วย (กัลยา วานิชย์บัญชา, 2550, 2554) จึงต้องเก็บข้อมูลจากตัวอย่าง หรือหน่วยทดลองที่เป็นตัวแทนของ

ประชากรทั้งนี้ อาจมีปัจจัยต่างๆ ที่กระทบต่อประสิทธิภาพการค้นคืนและการศึกษาในงานวิจัยนี้ ผู้วิจัยจึงควบคุมปัจจัยดังกล่าวด้วยการเก็บข้อมูลและการทดลองในห้องปฏิบัติการ (สุทธิ ชัดติยะ และ วิไลลักษณ์ สุวจิตตานนท์, 2554; องอาจ นัยพัฒน์, 2554) โดยมีรายละเอียดดังต่อไปนี้

3.2.1. เอกสาร

เอกสารที่สามารถนำมาใช้ในการค้นคืนสารสนเทศเป็นเอกสารที่ไม่มีโครงสร้างหรือกึ่งมีโครงสร้าง (ศุภชัย ตังวงศ์ศานต์, 2553) ลักษณะข้อความของเอกสารแต่ละชุดนั้นจึงต้องมีความยาวเหมาะสมสำหรับการนำมาใช้ในการค้นคืนสารสนเทศ อีกทั้งงานวิจัยนี้มุ่งประเด็นที่เอกสารภาษาไทย ดังนั้นประชากรเอกสารที่นำมาใช้ในงานวิจัย คือเอกสารภาษาไทยที่มีลักษณะกึ่งโครงสร้างหรือไม่มีโครงสร้าง

เนื่องจาก ไม่พบหน่วยงานใดที่นำเสนอหรือพัฒนาชุดทดสอบสำหรับภาษาไทย ผู้วิจัยจึงพัฒนาชุดทดสอบเพื่อนำมาใช้ในการประเมินประสิทธิภาพการค้นคืนสารสนเทศ งานวิจัยการสร้างชุดทดสอบและระบบการค้นคืนสารสนเทศในอดีต ได้นำเสนอรูปแบบของเอกสารหรือสารสนเทศที่นำมาใช้ในการประเมินประสิทธิภาพการค้นคืน มีรูปแบบเอ็กซ์เอ็มแอล (EXtreme Markup Language: XML) ซึ่งมีการกำหนดหัวข้อในเอกสาร ซึ่งง่ายต่อการศึกษาและประเมินเอกสารนอกจากนี้ เพื่อควบคุมระยะเวลาที่ใช้ในการศึกษาและประเมินความเกี่ยวข้อง ผู้วิจัยจึงกำหนดให้ความยาวเอกสารไม่เกินหนึ่งหน้ากระดาษ และเป็นเอกสารภาษาไทย ผู้วิจัยจึงกำหนดกลุ่มตัวอย่างเอกสารเป็นบทความย่อโครงการปี 4 ของนิสิตสาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชีซึ่งจัดทำตั้งแต่ปีการศึกษา 2552 ถึง 2554 เนื่องจากผู้ประเมินความเกี่ยวข้องที่ได้จากกำหนดไว้ข้างต้นนั้น มีความรู้ความเข้าใจเกี่ยวข้องกับเทคโนโลยีสารสนเทศสำหรับธุรกิจ และมีความต้องการในการศึกษาบทความดังกล่าว เพื่อประกอบการจัดทำโครงการชั้นปี 4 ของตนเอง ทั้งนี้ผู้วิจัยได้รับความอนุเคราะห์จาก ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย นำบทความย่อจำนวน 100 ฉบับมาใช้ในงานวิจัยนี้ ตัวอย่างและรายการบทความย่อที่นำมาใช้เป็นตัวอย่างเอกสารในงานวิจัยนี้ สามารถดูได้ที่ภาคผนวก ก

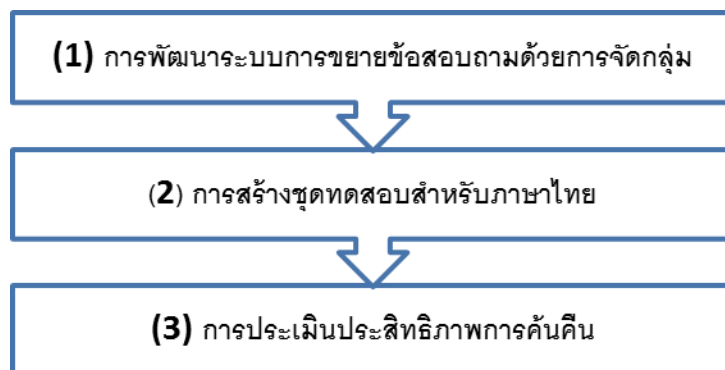
3.2.2. ผู้ประเมินความเกี่ยวข้อง

ในงานวิจัยนี้ จำเป็นต้องใช้ผู้ประเมินเพื่อนำมาใช้ในการกำหนดผลความเกี่ยวข้องระหว่างเอกสารและข้อสอบถาม เพื่อให้ได้ชุดข้อมูลที่สามารถนำไปใช้ในการประเมินประสิทธิผลการค้นคืน ผู้ประเมินจะต้องมีความรู้ความเข้าใจในเนื้อหาของเอกสาร อีกทั้งยังสามารถวิเคราะห์และแสดงความคิดเห็นได้หลากหลาย (Xu and Chen,2005,2006)

ชุดทดสอบที่นำมาใช้นั้น จะนำมาใช้ในการประเมินประสิทธิภาพการค้นคืนสารสนเทศ ในปัจจุบัน โปรแกรมค้นหา (Search Engine) เป็นระบบการค้นคืนสารสนเทศหนึ่งที่ใช้อินเทอร์เน็ตนิยมใช้มากที่สุด ผู้วิจัยจึงนำพฤติกรรมการใช้อินเทอร์เน็ตมาใช้ในการกำหนดลักษณะของผู้ประเมินความเกี่ยวข้อง จากการสำรวจของสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (2553) พบว่าผู้ใช้อินเทอร์เน็ตร้อยละ 40 มีอายุช่วง 20-29 ปี สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (2556) พบว่าผู้ใช้อินเทอร์เน็ต จะอยู่ในช่วงอายุ 20-34 ปี โดยเฉพาะอย่างยิ่ง นิสิตนักศึกษา มีจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์ถึง 34.2 ชั่วโมง ซึ่งใกล้เคียงกับค่าเฉลี่ยของจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์ 32.3 ชั่วโมงของผู้ใช้อินเทอร์เน็ตทั่วไป ผู้วิจัยจึงกำหนดให้ผู้ประเมินความเกี่ยวข้อง เป็นนิสิตนักศึกษาที่มีช่วงอายุ 20-34 ปี ทั้งนี้ผู้วิจัยจึงเลือกนิสิตคณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ที่กำลังศึกษาชั้นปีที่ 4 ในปีการศึกษา 2555 มาเป็นกลุ่มตัวอย่างผู้ประเมินความเกี่ยวข้อง นิสิตกลุ่มนี้ยังมีความต้องการใช้การค้นคืนข้อมูล เพื่อประกอบการจัดทำโครงการปี 4 ของตน ในงานวิจัยนี้มีนิสิตจำนวน 25 คนสนใจเข้าร่วมการวิจัยในการประเมินเอกสาร

3.3. ขั้นตอนการดำเนินงานวิจัย

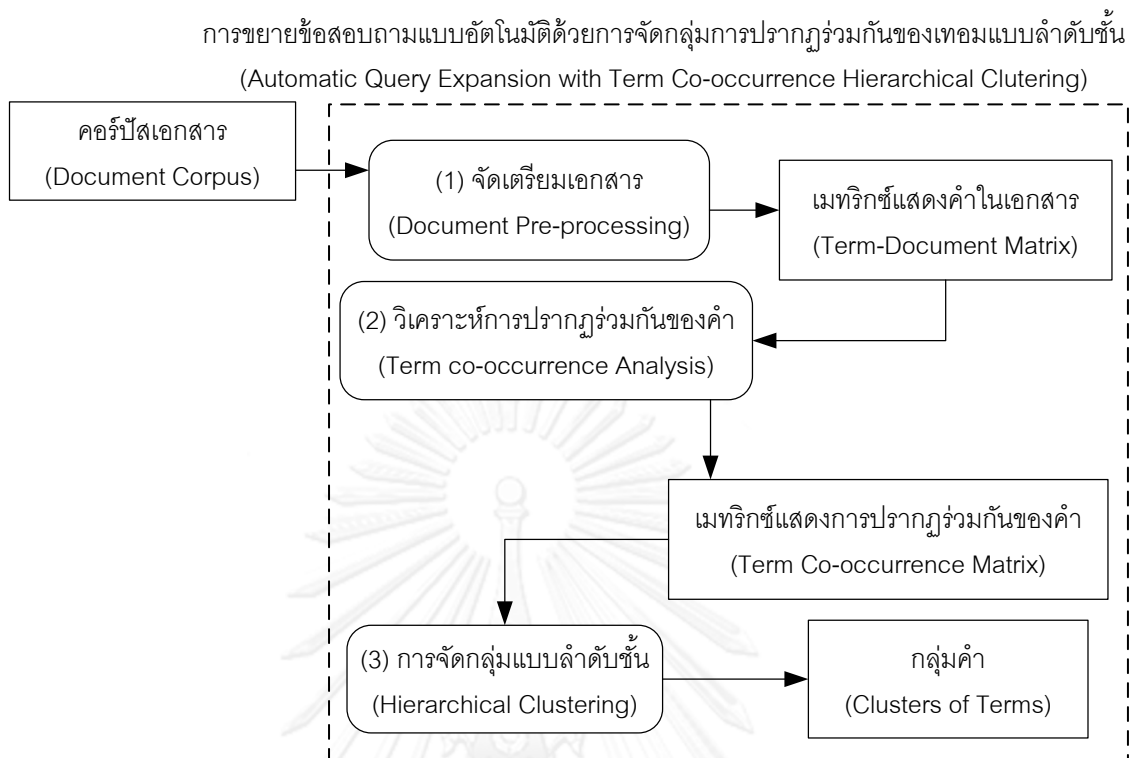
งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการนำการขยายข้อสอบถามด้วยการจัดกลุ่มคำถามประยุกต์ใช้กับระบบการค้นคืนสารสนเทศภาษาไทย ด้วยการเปรียบเทียบประสิทธิภาพการค้นคืนก่อนและหลังการนำการขยายข้อสอบถามมาประยุกต์ใช้ ดังนั้นงานวิจัยนี้จึงกำหนดขั้นตอนการดำเนินงานวิจัยดังรูปที่ 3.1



รูปที่ 3.1 แสดงแนวทางการดำเนินงานวิจัย

3.3.1. การพัฒนาระบบการขยายข้อสอบถามด้วยการจัดกลุ่มคำ

ผู้วิจัยได้นำเสนอการขยายข้อสอบถามด้วยการจัดกลุ่มคำ ด้วยการวิเคราะห์หากลุ่มคำจากรายการเอกสารจากคอร์ปัสเอกสารและข้อสอบถามจากผู้ใช้ จนกระทั่งได้ข้อสอบถามใหม่ ซึ่งเกิดจากการนำกลุ่มคำมาเพิ่มเติมลงในข้อสอบถามเดิม โดยมีขั้นตอนวิธีดังรูปที่ 3.2 ผู้วิจัยจึงพัฒนาระบบการขยายข้อสอบถามด้วยการจัดกลุ่มคำซึ่งมีรายละเอียดตามขั้นตอนในการขยายข้อสอบถามดังต่อไปนี้



รูปที่ 3.2 ขั้นตอนการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันระหว่างคำด้วยการจัดกลุ่มแบบลำดับชั้น

(1) การจัดเตรียมเอกสาร (Document Pre-processing)

ในขั้นตอนวิธีนี้ จะวิเคราะห์รายการเอกสารในคอร์ปัสเอกสาร เพื่อกำหนดรายการคำที่จะนำมาใช้ในการขยายข้อสอบถาม รวมทั้งการวิเคราะห์ในขั้นตอนถัดไป จากรูปที่ 3.2 ผลลัพธ์ที่ได้คือ เมทริกซ์แสดงคำในเอกสาร (Term-document Matrix) หรือ D ซึ่งมีลักษณะดังสมการต่อไปนี้

$$D = [w_{i,j}]_{n \times m}$$

โดยที่ $w_{i,j}$ แทนค่าน้ำหนักของค่าของคำลำดับที่ i ในเอกสารฉบับที่ j

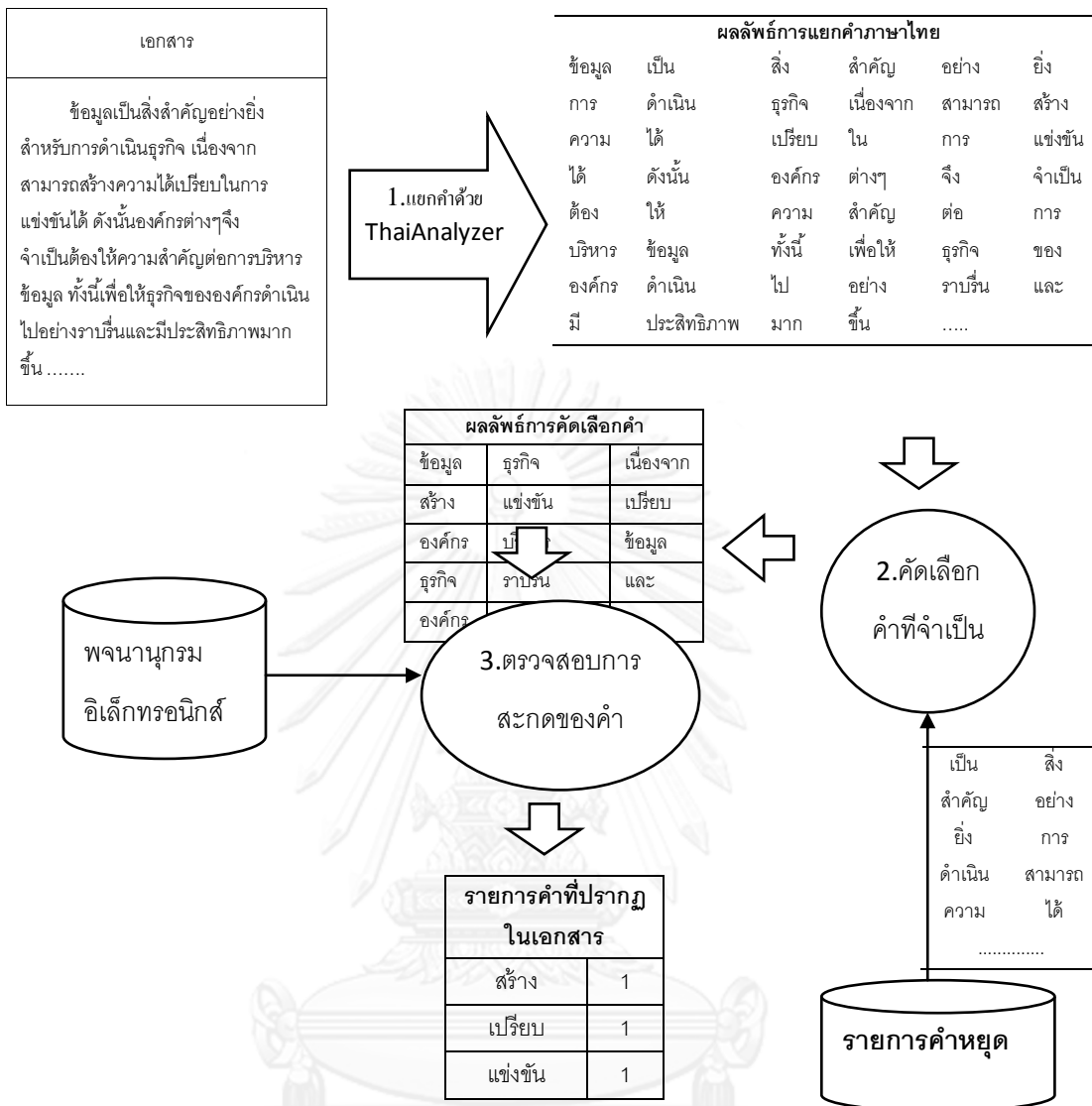
n แทนจำนวนเอกสารที่นำมาใช้ในงานวิจัยนี้ โดยที่มีค่าเท่ากับ 100

m แทนจำนวนคำที่ปรากฏอยู่ในเอกสารทั้งหมด

จะเห็นได้ว่า เมทริกซ์แสดงคำในเอกสารนั้น จะประกอบด้วยรายการเอกสาร รายการคำที่ปรากฏอยู่ในเอกสารทั้งหมด และค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสารแต่ละรายการ ซึ่งรายการคำที่ใช้นั้นจะนำไปใช้ในการวิเคราะห์คำ เพื่อการขยายข้อสอบถาม อย่างไรก็ตาม ปัญหาคำศัพท์ (Vocabulary Problem) ยังคงเป็นปัญหาที่เกิดขึ้นในการวิจัยหรือวิเคราะห์คำจากเอกสารรวมทั้งเอกสารภาษาไทย ดังนั้น ผู้วิจัยจึง (1) วิเคราะห์รายการคำที่ปรากฏในเอกสาร (2) วิเคราะห์รายการดัชนีและเชื่อมโยงกับรายการเอกสาร และ (3) คำนวณค่าน้ำหนักของคำในเอกสาร เพื่อให้ได้เมทริกซ์แสดงคำในเอกสาร สำหรับการวิเคราะห์ขั้นต่อไป โดยมีรายละเอียดดังต่อไปนี้

(1.1) การวิเคราะห์รายการคำในเอกสาร

ในขั้นตอนนี้ ผู้วิจัยจะต้องนำข้อมูลในเอกสารแต่ละฉบับมาวิเคราะห์ เพื่อกำหนดรายการคำจากเอกสาร โดยมีขั้นตอนดังรูปที่ 3.3 โดยที่ระบบ (1) แยกคำด้วยไทยอนาไลเซอร์ (ThaiAnalyzer) ซึ่งเป็นส่วนขยายของเครื่องมือการค้นคืนสารสนเทศ มาแยกคำภาษาไทยจากเอกสาร (2) คัดเลือกคำที่จำเป็น ด้วยรายการคำหยุดภาษาไทย (Thai Stop words List) ของวรวิโรภษร (2553) ดังภาคผนวก ข ซึ่งปรากฏอยู่ในภาคผนวก หากเป็นคำที่ปรากฏอยู่ในรายการคำหยุด ระบบจะตัดคำนั้นออกจากรายการคำจากเอกสาร และ (3) ตรวจสอบการสะกดของคำที่นำมาใช้ในการวิเคราะห์ ด้วยพจนานุกรมอิเล็กทรอนิกส์ ที่จัดทำโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronic and Computer Technology Center, NECTEC) ซึ่งปรับปรุงล่าสุดปี พ.ศ. 2552 หากพบว่าเป็นคำที่ไม่ปรากฏอยู่ในพจนานุกรม ระบบจะตัดคำนั้นออกจากรายการคำจากเอกสาร เมื่อวิเคราะห์เสร็จสิ้นแล้ว ผลลัพธ์ที่ได้จากการวิเคราะห์เอกสารคือรายการคำที่ปรากฏในเอกสาร รวมทั้งจำนวนคำที่ปรากฏ นอกจากนี้ ระบบจะบันทึกรายการคำที่ได้จากการวิเคราะห์เอกสารเข้าสู่ฐานข้อมูล ซึ่งจะปรากฏอยู่ในภาคผนวก ค

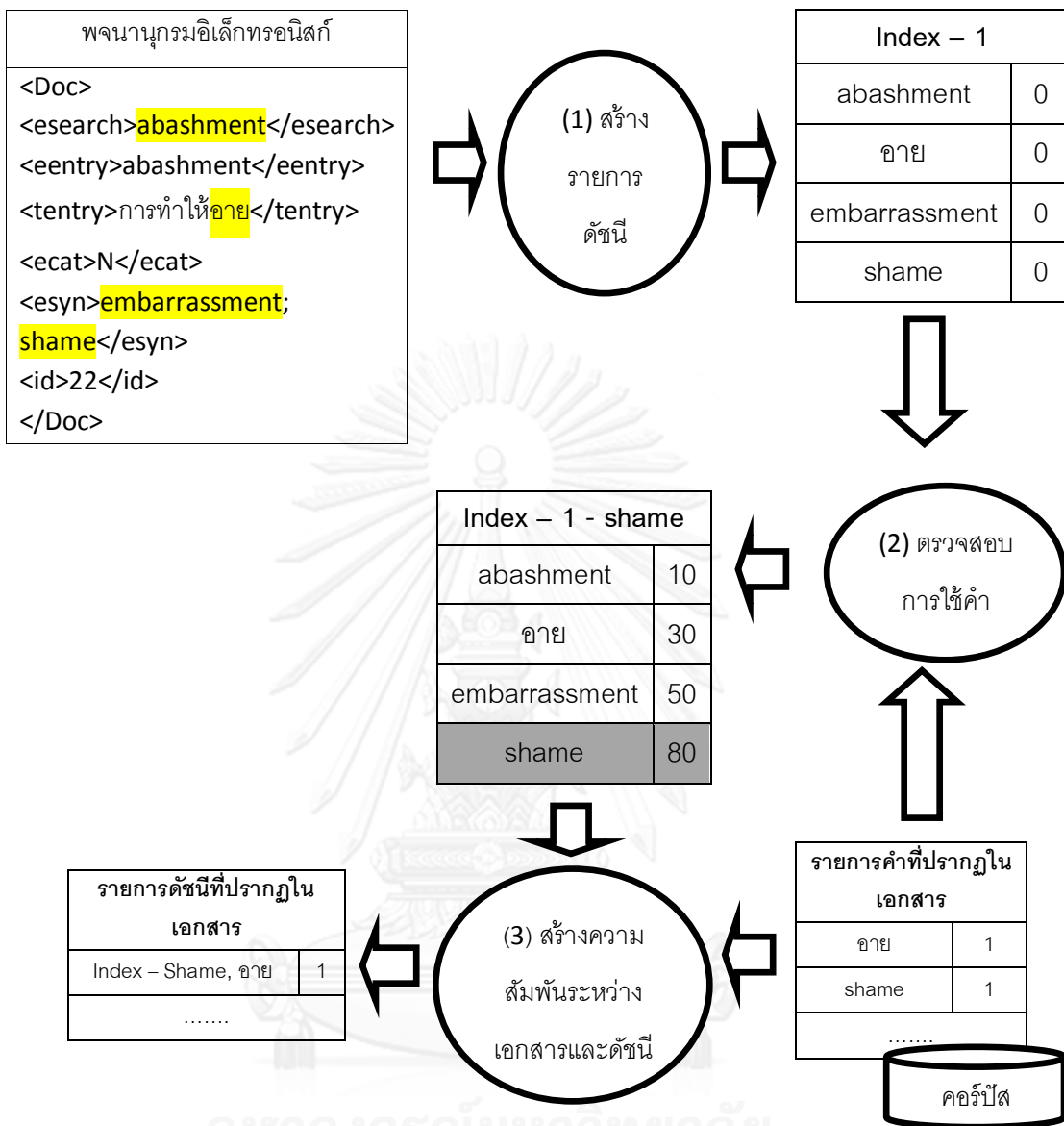


รูปที่ 3.3 การวิเคราะห์รายการคำในเอกสาร

(1.2) การวิเคราะห์รายการดัชนีและเชื่อมโยงกับรายการเอกสาร

ในขั้นตอนนี้ ผู้วิจัยจะต้องสร้างรายการดัชนีเพื่อลดปัญหาคำศัพท์ อาทิ คำพ้อง ความหมาย หรือคำกำกวม โดยที่ระบบนั้น (1) สร้างรายการดัชนี จากพจนานุกรมอิเล็กทรอนิกส์ มาประกอบการวิเคราะห์ ดังรูปที่ 3.4 โดยที่รายการคำศัพท์ในพจนานุกรมอิเล็กทรอนิกส์ มีลักษณะเป็นโครงสร้างเอ็กซ์เอ็มแอล (Extreme Markup Language, XML) แต่ละรายการจะ

ประกอบแตกต่างกันที่สนใจดังนี้ (1) esearch หรือ tsearch หมายถึงคำที่นำมาใช้ในการค้นคืน (2) eentry หรือ tentry หมายถึงผลลัพธ์ที่ได้จากการค้นคืน และ (3) esys หรือ tsys หมายถึงคำพ้องหรือคำที่เกี่ยวข้อง มาใช้ในการวิเคราะห์รายการคำ จากรูปที่ 3.4 ผลลัพธ์ที่ได้จากการสร้างดัชนีเป็นรายการคำในดัชนี รวมทั้งจำนวนที่ปรากฏอยู่ในเอกสาร (2) ตรวจสอบการใช้คำในดัชนี โดยผู้วิจัยจะนำรายการดัชนีแต่ละรายการมาตรวจสอบจำนวนครั้งที่ปรากฏในรายการเอกสารทุกฉบับ หลังจากนั้น ระบบจะเลือกคำที่ปรากฏมากที่สุดมาเป็นตัวแทนของคำในดัชนีแต่ละรายการ เพื่อกำหนดเป็นคำสำคัญที่จะนำมาใช้ในการวิเคราะห์คำและ (3) สร้างความสัมพันธ์ระหว่างเอกสารและดัชนี หากพบว่าคำที่ปรากฏในดัชนี ตรงกับรายการคำในเอกสาร ถือได้ว่าดัชนีนี้ปรากฏอยู่ในเอกสาร จากรูปที่ 3.4 จะเห็นได้ว่า คำสองคำที่ปรากฏอยู่ในรายการคำของเอกสาร จึงถือได้ว่าดัชนี Shame ปรากฏอยู่ในเอกสารฉบับนี้ ทั้งนี้ในการวิเคราะห์คำ ระบบจะตั้งเฉพาะคำว่า Shame ที่เป็นตัวแทนของกลุ่มคำในดัชนีใกล้เคียงกัน



รูปที่ 3.4 แสดงการวิเคราะห์รายการดัชนีและเชื่อมโยงกับรายการเอกสาร

(1.3) การคำนวณค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร

เมื่อผู้วิจัยวิเคราะห์การแยกคำออกจากเอกสารแล้ว ผู้วิจัยจึงนำรายการคำที่ได้จากการวิเคราะห์เอกสารทุกฉบับมาวิเคราะห์เมทริกซ์แสดงคำในเอกสาร ดังต่อไปนี้

กำหนดให้ d_j แทนเวกเตอร์ของเอกสารลำดับที่ j ใน D และ $w_{i,j}$ แทนค่าน้ำหนักของคำที่ i ในเอกสาร j จึงสามารถเขียน d_j ในรูปแบบของเวกเตอร์เอกสารได้ดังนี้

$$d_j = [w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{i,j}] \text{ โดยที่ } w_{i,j} = tf_{i,j} \times idf_i$$

โดยที่ $tf_{i,j}$ แทนความถี่ของคำ i ที่ปรากฏในเอกสาร j ในขณะที่ idf_i แทนค่าผกผันของความถี่ของคำที่ i ปรากฏในการกำหนดค่า $tf_{i,j}$ เนื่องจากแต่ละคำที่ปรากฏในเอกสารนั้นจะมีความถี่ที่แตกต่างกัน จึงอาจจะเกิดความเลื่อมล้ำได้ เมื่อนำไปใช้ในการคำนวณ อาทิ คำหนึ่งมีความถี่ 10 ในขณะที่อีกคำหนึ่งมีความถี่ 1,000 เพื่อลดปัญหาดังกล่าว จึงควรทำรูปแบบการสร้างบรรทัดฐาน (Normalized Form) ดังนี้

$$tf_{i,j} = \frac{f_{i,j}}{F_j}$$

โดยที่ $f_{i,j}$ แทนความถี่ของคำ i ในเอกสาร j ในขณะที่ F_j แทนความถี่ของคำหนึ่งในเอกสาร j ที่มีค่าสูงที่สุด (ในบางกรณี F_j อาจจะใช้ค่าเฉลี่ยของคำในเอกสารแทน)

ส่วนการกำหนดค่า idf_i นี้เนื่องจากอัตราส่วนของจำนวนเอกสารทั้งหมด (N) ต่อจำนวนเอกสารที่ปรากฏคำ i อยู่ (n_i) อาจมีค่าที่สูงหรือจนเกินไป ซึ่งอาจจะส่งผลกระทบต่อค่าการคำนวณ $w_{i,j}$ ได้ซึ่งนำฟังก์ชันลอการิทึม (Logarithm) มาใช้เพื่อกำหนดไม่ให้ค่าที่ได้นั้นสูงจนเกินไป ดังนั้นจึงคำนวณหาค่า idf_i ได้จาก

$$idf_i = \log \frac{N}{n_i}$$

โดยที่ N แทนจำนวนเอกสารที่มีอยู่ในคอร์ปัสเอกสารหรือจำนวนรายการเอกสารทั้งหมด ส่วน n_i แทนจำนวนเอกสารที่มีคำที่ i ปรากฏอยู่ทั้งหมดในคอร์ปัสเอกสาร

ผู้วิจัยขอยกตัวอย่างการคำนวณค่าน้ำหนักของคำ "A" ที่ปรากฏอยู่ในเอกสาร ดังตารางที่ 3.1 โดยกำหนดให้มีรายการเอกสารในคอร์ปัสเอกสารจำนวนทั้งสิ้น 3 ฉบับ

ตารางที่ 3.1 ตัวอย่างการวิเคราะห์ค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร

ตัวอย่างเอกสาร	รายการดัชนีที่ปรากฏในเอกสาร B		รายการดัชนีที่ปรากฏในเอกสาร C		รายการดัชนีที่ปรากฏในเอกสาร D	
	A	1	A	2	B	1
	B	1	B	1		
	C	3				
F_j	3		2		1	
$f_{A,j}$	1		2		0	
$tf_{i,j} = \frac{f_{i,j}}{F_j}$	$\frac{1}{3} = 0.33$		$\frac{2}{2} = 1$		0	
N	3					
n_A	2					
$idf_i = \log \frac{N}{n_i}$	$\log \frac{3}{2} = 0.1761$					
$w_{A,j} = tf_{i,j} \times idf_i$	0.33 x 0.1761 = 0.0581		1 x 0.1761 = 0.1761		0	

ในการจัดเตรียมเอกสาร ผู้วิจัยได้นำรายการเอกสารมาใช้จำนวน 100 ฉบับ สามารถวิเคราะห์รายการคำได้ จำนวนทั้งสิ้น 794 คำ และได้ค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร ดังที่ปรากฏอยู่ในภาคผนวก จ

(2) การคำนวณเมทริกซ์แสดงการปรากฏร่วมกันของคำ (Term Co-occurrence Matrix)

หลังจากการวิเคราะห์ค่าน้ำหนักของคำในเอกสาร ผู้วิจัยได้พัฒนาชุดคำสั่งการวิเคราะห์เมทริกซ์แสดงการปรากฏร่วมกันของคำ โดยได้ใช้สูตรดังต่อไปนี้

$$S = D^T \times D$$

โดยที่ $S = [s_{k,l}]_{m \times m}$ แทนเมทริกซ์การปรากฏร่วมกันของคำ มีขนาด $m \times m$

$s_{k,l}$ แทนค่าสัมประสิทธิ์ในเมทริกซ์ S ซึ่งแสดงถึงค่าน้ำหนักระหว่างคำ k และคำ l โดยที่ $s_{k,l} = s_{l,k}$

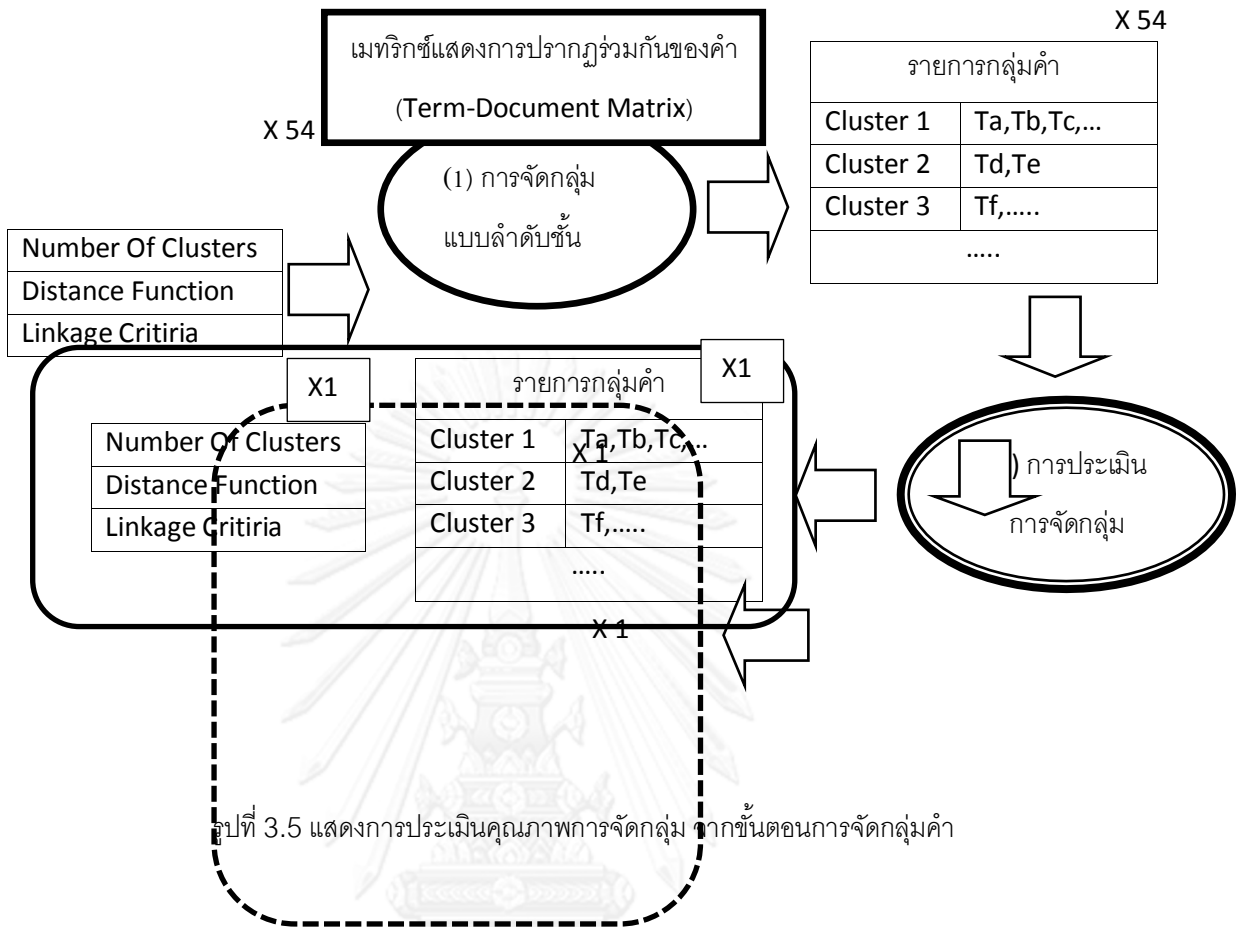
ผู้วิจัยขอยกตัวอย่างการนำเมทริกซ์แสดงค่าที่ปรากฏในเอกสารมาวิเคราะห์เมทริกซ์แสดงการปรากฏร่วมกันของคำ ดังตารางที่ 3.2 ตามขั้นตอนการวิเคราะห์ข้างต้น

ตารางที่ 3.2 ตัวอย่างการวิเคราะห์เมทริกซ์แสดงการปรากฏร่วมกันของคำ

D	$\begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.67 & 0.1 & 0.4 \end{bmatrix}$
D^T	$\begin{bmatrix} 0.3 & 0.67 \\ 0.5 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}$
$S = D^T \times D$	$\begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.67 & 0.1 & 0.4 \end{bmatrix} \times \begin{bmatrix} 0.3 & 0.67 \\ 0.5 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}$ $= \begin{bmatrix} 0.38 & 0.331 \\ 0.331 & 0.6189 \end{bmatrix}$

(3) การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering)

หลังจากที่ผู้วิจัยสามารถวิเคราะห์ได้เมทริกซ์แสดงการปรากฏร่วมกันของคำแล้ว ผู้วิจัยจึงนำเมทริกซ์ดังกล่าวมาวิเคราะห์กลุ่มคำ (Term Cluster Analysis) หรือการจัดกลุ่มคำ (Term Clustering) เพื่อสร้างกลุ่มคำที่จะนำมาใช้ในการขยายข้อสอบถาม ทั้งนี้ผู้วิจัยได้กำหนดรูปแบบการจัดกลุ่ม ดังที่กล่าวมาข้างต้น ผลลัพธ์ที่ได้จากการจัดกลุ่มนี้ เป็นเซตของกลุ่มคำ (Sets of Clusters) ผู้วิจัยจึงกำหนดขั้นตอนดังต่อไปนี้



รูปที่ 3.5 แสดงการประเมินคุณภาพการจับกลุ่ม จากขั้นตอนการจับกลุ่มคำ

(3.1) การจับกลุ่มแบบลำดับขั้น (Hierchical Clustering)

จากขั้นตอนการวิธีการขยายข้อสอบถามดังที่ผู้วิจัยได้นำเสนอ จะต้องจับกลุ่มคำ ด้วยขั้นตอนวิธีการจับกลุ่มแบบลำดับขั้น ซึ่งจะต้องกำหนดรูปแบบการจับกลุ่มที่จะนำมาใช้ จากการนำเสนองานวิจัยข้างต้น จำนวนกลุ่มที่นำมาใช้ได้แก่ 5, 10, 15, 20, 25, และ 30 ส่วนสมการค่าความเหมือนที่นิยมนำมาใช้ได้แก่ โคไซน์ (Cosine) ยูคลิดเดียน (Euclidean) และ จาคคาร์ด (Jaccard) และสมการเกณฑ์การเชื่อมโยงที่นำมาใช้ได้แก่ เกณฑ์การเชื่อมโยงเดี่ยว (Simple Linkage) เกณฑ์การเชื่อมโยงทับซ้อน (Complete Linkage) และ เกณฑ์การเชื่อมโยงเฉลี่ย (Average Linkage) ดังนั้นรูปแบบการจับกลุ่มที่นำมาใช้ในงานวิจัยนี้ มีจำนวนทั้งสิ้น 54 รูปแบบ

ผู้วิจัยจึงนำเมทริกซ์แสดงการปรากฏร่วมกันของคำมาวิเคราะห์หากลุ่มคำ ด้วยขั้นตอนวิธีการจับกลุ่มแบบลำดับขั้น โดยที่นำรูปแบบการจับกลุ่มทั้ง 54 รูปแบบมาใช้ จนกระทั่งได้เซตของกลุ่มคำจำนวน 54 เซต

(3.2) การประเมินการจัดกลุ่ม

การจัดกลุ่มคำด้วยรูปแบบการจัดกลุ่มที่แตกต่างกัน 54 รูปแบบ ส่งผลให้ผลลัพธ์ที่ได้จากการจัดกลุ่มคำนั้นแตกต่างกันออกไป อย่างไรก็ตาม ในการขยายข้อสอบถามด้วยการจัดกลุ่มคำนั้น เซตของกลุ่มคำส่งผลให้ข้อสอบถามและประสิทธิภาพการค้นคืนนั้นแตกต่างกันออกไป ดังนั้น ผู้วิจัยจึงต้องคัดเลือกเซตของกลุ่มคำที่นำมาใช้ โดยเทียบกับรูปแบบการจัดกลุ่ม เพื่อให้ข้อสอบถามและประสิทธิภาพการค้นคืนนั้นเหมาะสมกับการนำเซตของกลุ่มคำไปใช้ ทั้งนี้ในการประเมินคุณภาพการจัดกลุ่มนั้น ผู้วิจัยจะกำหนดค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสอง ค่าระยะทางสูงสุดจากกลุ่ม และค่าระยะทางสู่กลุ่มที่ใกล้ที่สุด มาใช้ (สุนันทา เปี่ยมพริ้ง, 2550) โดยที่

- ค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสอง (Root Means Square Standard Deviation, RMSSD) เป็นค่าแสดงการกระจายของข้อมูลภายในกลุ่ม การจัดกลุ่มที่ดีนั้น ควรมีการกระจายตัวของข้อมูลที่น้อย
- ค่าระยะทางสูงสุดจากกลุ่ม (Maximum Distance from Cluster, MDC) เป็นค่าสถิติแสดงค่าระยะทางที่มากที่สุดระหว่างจุดศูนย์กลางของกลุ่มไปยังข้อมูลภายในกลุ่มเดียวกันในการจัดกลุ่มที่ดีนั้น ค่าระยะทางสูงสุดจากกลุ่มนั้น ควรมีค่าน้อย
- ค่าระยะทางสู่กลุ่มที่ใกล้ที่สุด (Distance to Nearest Cluster, DNC) เป็นค่าระยะทางระหว่างกลุ่มข้อมูลที่อยู่ใกล้ที่สุด ในการจัดกลุ่ม ค่านี้นควรมีค่ามาก

ดังนั้น ผู้วิจัยต้องคำนวณค่าวัดคุณภาพการจัดกลุ่มสามค่าดังที่กล่าวมา โดยที่ผู้วิจัยจะต้องกำหนดจุดศูนย์กลางกลุ่ม (Cluster Seed) กำหนดให้ C_p แทนกลุ่มของคำกลุ่มที่ p ที่ได้จากการจัดกลุ่ม มี $T_{i,p}$ แทนเทอมลำดับที่ i ของรายการคำทั้งหมด และอยู่ในกลุ่มคำกลุ่มที่ p โดยที่ $1 \leq i \leq m$ ดังนั้นผู้วิจัยสามารถหาจุดศูนย์กลางของกลุ่มได้จากสมการต่อไปนี้

$$CS_p = [cs_k]_{1 \times m}$$

$$cs_k = \frac{\sum_{T_{i,p} \in C_p} S_{i,k}}{|C_p|}$$

โดยที่ CS_p แทนเวกเตอร์ของจุดศูนย์กลางของกลุ่มที่ p

CS_k แทนค่าสัมประสิทธิ์ของค่าลำดับที่ k

เมื่อผู้วิจัยได้เวกเตอร์ของจุดศูนย์กลางของกลุ่มแล้ว ผู้วิจัยจึงสามารถคำนวณค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสอง ค่าระยะทางสูงสุดจากกลุ่ม และระยะทางสู่กลุ่มที่ใกล้ที่สุด กำหนดให้ nc แทนจำนวนกลุ่มที่ได้ทั้งหมด ระบบสามารถคำนวณค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสองจาก

$$RMSSD = \sqrt{\frac{\sum_{p=1}^{nc} \sum_{T_{i,p} \in C_p} (CS_k - s_{i,k})^2}{m}}$$

ส่วนค่าระยะทางสูงสุดจากกลุ่มและระยะทางสู่กลุ่มที่ใกล้ที่สุด สามารถคำนวณได้จาก

$$MDC = \text{Max}_{p=1}^{nc} (\text{Max}_{T_{i,p} \in C_p} \text{distance}(CS_p, T_{i,p}))$$

$$DNC = \text{Min}_{p=1}^{nc} (\text{Min}_{q=p+1}^{nc} \text{distance}(CS_p, CS_q))$$

โดยที่ $\text{distance}(A, B)$ เป็นสมการการคำนวณระยะทางระหว่างเวกเตอร์ A และ B

เมื่อผู้วิจัยได้ค่าการประเมินการจัดกลุ่มทั้งสามค่า จากการประเมินการจัดกลุ่มตามรูปแบบการจัดกลุ่มที่แตกต่างกัน 54 รูปแบบ ผู้วิจัยจึงนำค่าที่ได้ทั้งหมดมาเปรียบเทียบดังตารางที่ 3.3 เพื่อเลือกรูปแบบการจัดกลุ่ม

ตารางที่ 3.3 แสดงผลการวิเคราะห์คุณภาพการจัดกลุ่มที่ได้จากการกำหนดรูปแบบการจัดกลุ่มที่แตกต่างกัน 54 รูปแบบ เมื่อใช้ชุดรายการผลลัพธ์การค้นคืนชุดหนึ่ง

การกำหนดการจัดกลุ่ม			ผลลัพธ์การวิเคราะห์ที่ได้		
K	DISTANCE	LINKAGE	RMSSD	MDCS	DNC
5	COSINE	SIMPLE	1.0951	7.1154	0.0049
5	COSINE	COMPLETE	1.0983	7.1114	0.0163
5	COSINE	AVERAGE	1.0953	7.1154	0.0060
5	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
5	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000
5	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000
5	JACCARD	SIMPLE	1.0941	7.0375	0.0020
5	JACCARD	COMPLETE	1.0941	7.0375	0.0020
5	JACCARD	AVERAGE	1.0941	7.0375	0.0020
10	COSINE	SIMPLE	1.0975	7.1154	0.0034
10	COSINE	COMPLETE	1.0998	7.1114	0.0077
10	COSINE	AVERAGE	1.0984	7.1154	0.0043
10	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
10	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000
10	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000
10	JACCARD	SIMPLE	1.0952	7.0384	0.0005
10	JACCARD	COMPLETE	1.0957	7.0441	0.0015
10	JACCARD	AVERAGE	1.0950	7.0394	0.0011
15	COSINE	SIMPLE	1.0992	7.1154	0.0027
15	COSINE	COMPLETE	1.1021	7.1154	0.0043
15	COSINE	AVERAGE	1.1000	7.1154	0.0042
15	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
15	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000
15	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000

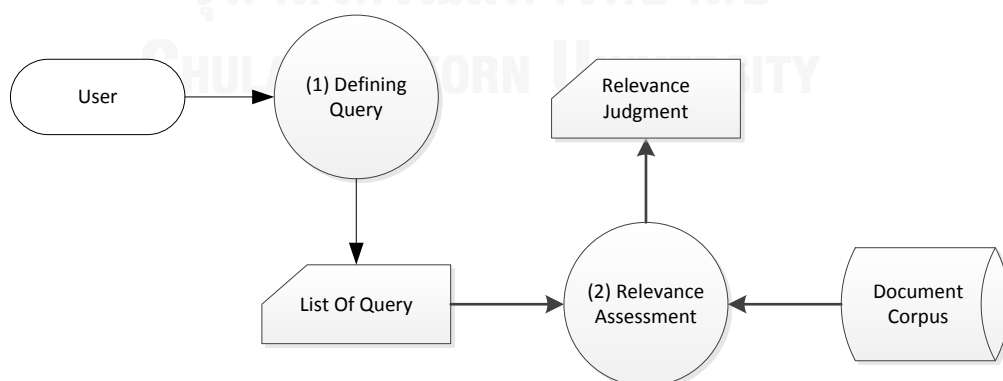
การกำหนดการจัดกลุ่ม			ผลลัพธ์การวิเคราะห์ที่ได้		
K	DISTANCE	LINKAGE	RMSSD	MDCS	DNC
15	JACCARD	SIMPLE	1.0967	7.0389	0.0005
15	JACCARD	COMPLETE	1.0973	7.0466	0.0009
15	JACCARD	AVERAGE	1.0961	7.0398	0.0005
20	COSINE	SIMPLE	1.1009	7.1154	0.0027
20	COSINE	COMPLETE	1.1037	7.1154	0.0042
20	COSINE	AVERAGE	1.1009	7.1154	0.0027
20	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
20	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000
20	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000
20	JACCARD	SIMPLE	1.0987	7.0411	0.0005
20	JACCARD	COMPLETE	1.0979	7.0469	0.0009
20	JACCARD	AVERAGE	1.0973	7.0404	0.0005
25	COSINE	SIMPLE	1.1014	7.1154	0.0023
25	COSINE	COMPLETE	1.1093	7.1154	0.0038
25	COSINE	AVERAGE	1.1032	7.1154	0.0024
25	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
25	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000
25	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000
25	JACCARD	SIMPLE	1.0999	7.0444	0.0005
25	JACCARD	COMPLETE	1.1053	7.1036	0.0005
25	JACCARD	AVERAGE	1.0991	7.0441	0.0005
30	COSINE	SIMPLE	1.1023	7.1154	0.0023
30	COSINE	COMPLETE	1.1097	7.1154	0.0038
30	COSINE	AVERAGE	1.1041	7.1154	0.0024
30	EUCLIDEAN	SIMPLE	1.0929	7.0367	0.0000
30	EUCLIDEAN	COMPLETE	1.0929	7.0367	0.0000

การกำหนดการจัดกลุ่ม			ผลลัพธ์การวิเคราะห์ที่ได้		
K	DISTANCE	LINKAGE	RMSSD	MDCS	DNC
30	EUCLIDEAN	AVERAGE	1.0929	7.0367	0.0000
30	JACCARD	SIMPLE	1.1003	7.0446	0.0005
30	JACCARD	COMPLETE	1.1058	7.1039	0.0005

เมื่อผู้วิจัยได้พิจารณาค่าที่ได้ทั้งหมด ค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสอง (RMSSD) และค่าระยะทางสูงสุดจากกลุ่ม (MDCS) มีค่าที่ใกล้เคียงกัน ในขณะที่ค่าระยะทางสู่กลุ่มที่ใกล้ที่สุด (DNC) มีค่าที่แตกต่างอย่างเห็นได้ชัด ผู้วิจัยจึงได้เลือกจำนวนกลุ่ม 30 กลุ่ม สมการค่าความเหมือนโคไซน์ และสมการเกณฑ์การเชื่อมโยงเชิงสมบูรณ์ เป็นรูปแบบการจัดกลุ่มที่จะนำมาใช้ในการวิเคราะห์กลุ่มคำ ซึ่งในภาคผนวก

3.3.2. การสร้างชุดทดสอบสำหรับภาษาไทย

ผู้วิจัยต้องรวบรวมผลการประเมินความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร เพื่อประเมินประสิทธิภาพการค้นคืนสารสนเทศ จึงต้องมีการสร้างชุดทดสอบภาษาไทย ขั้นตอนการสร้างชุดทดสอบสำหรับภาษาไทยดังรูปที่ 3.6 มีรายละเอียดต่อไปนี้



รูปที่ 3.6 ขั้นตอนวิธีการสร้างชุดทดสอบ (Test Collection)

(1) การกำหนดรายการข้อสอบถาม (Defining Query)

ผู้วิจัยรับสมัครผู้ประเมินความเกี่ยวข้องของข้อสอบถามและเอกสารจำนวน 25 คน จากนั้น ผู้วิจัยได้กำหนดให้ผู้ประเมินกำหนดข้อสอบถามคนละหนึ่งข้อสอบถาม โดยที่ข้อสอบถามมีขนาดตั้งแต่ห้าถึงแปดคำ และตรงกับหัวข้อโครงการงานชั้นปี 4 ของนิสิตคณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีเพื่อธุรกิจ รายการข้อสอบถามที่ผู้ประเมินได้กำหนดและนำมาใช้ในงานวิจัยนี้ อยู่ในภาคผนวก ฉ

(2) การประเมินความเกี่ยวข้อง (Relevance Assessment)

จากงานสำรวจในอดีต (Miller, 2010; StatOwl, 2012; Buscher และคณะ, 2012; Liu และคณะ, 2010; Kim, 2008) พบว่า ผู้ใช้ระบบการค้นคืนสารสนเทศนั้นจะศึกษารายการเอกสารจากรายการผลลัพธ์การค้นคืนเพียง 30 รายการแรก ดังนั้นผู้วิจัยจึงต้องกำหนดปริมาณเอกสารในแต่ละครั้งไม่เกิน 30 ฉบับต่อครั้ง จากจำนวนรายการเอกสาร 100 ฉบับและผู้ประเมินความเกี่ยวข้องจำนวน 25 คนที่ผู้วิจัยได้กำหนดนั้น ผู้วิจัยจึงกำหนดให้ผู้ประเมินความเกี่ยวข้องประเมินเอกสารจำนวน 25 ฉบับต่อครั้ง จำนวนสี่ครั้ง

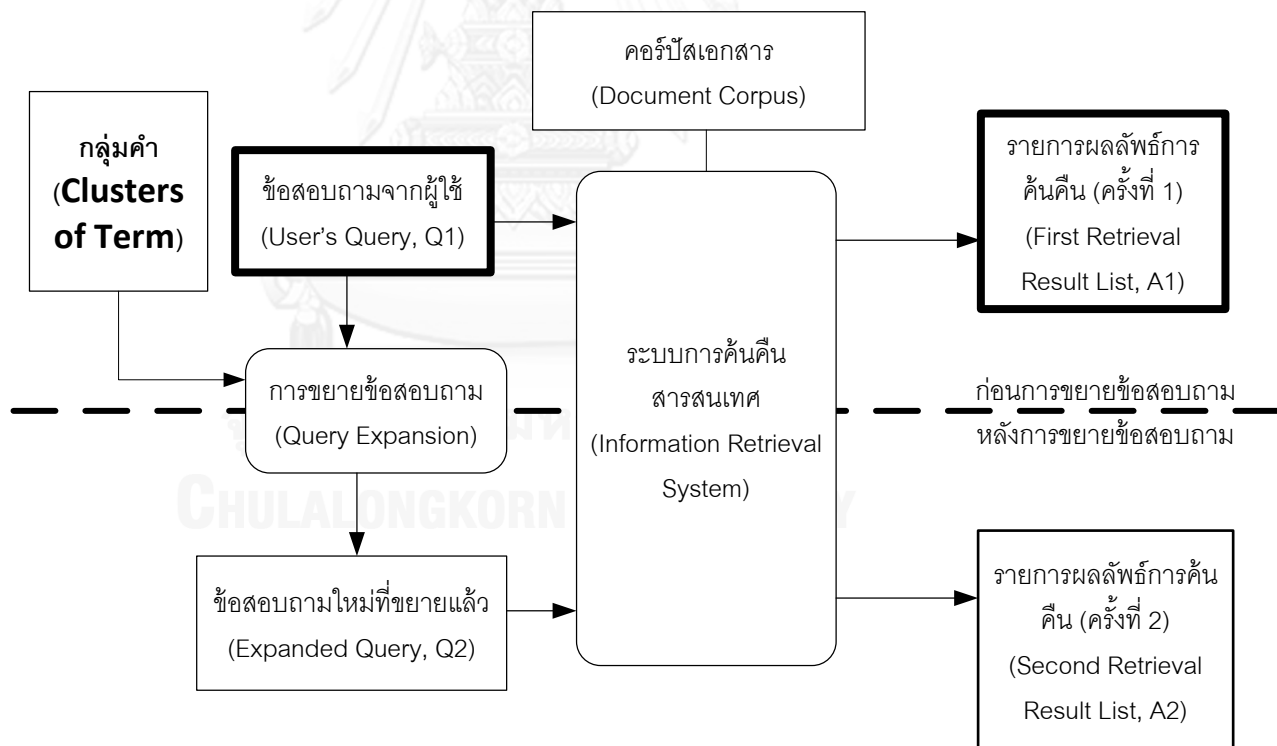
ในการประเมินความเกี่ยวข้อง ผู้วิจัยได้กำหนดแบบฟอร์มการบันทึกความเกี่ยวข้อง ดังรูปที่ 3.7 โดยเริ่มต้นผู้วิจัยจะบันทึกข้อสอบถาม หลังจากนั้นผู้ประเมินจะรับเอกสารมาเพื่อศึกษาและประเมินเอกสาร ก่อนที่จะบันทึกรหัสเอกสารของเอกสารที่กำลังประเมินความเกี่ยวข้องในครั้งนั้น และผลการประเมินความเกี่ยวข้อง ก่อนที่จะประเมินเอกสารฉบับถัดไป ในแบบฟอร์มนี้ การทำเครื่องหมายลงในช่องว่างใน “ความเกี่ยวข้อง” หมายถึงเอกสารฉบับนี้ จะเกี่ยวข้องข้อสอบถามที่ผู้ประเมินได้กำหนด ในขณะที่การไม่ทำเครื่องหมายลงในช่องดังกล่าว จะแสดงถึงว่า เอกสารไม่เกี่ยวข้องกับข้อสอบถาม

<p>ชื่อ..... วัลภา ภาสันเพ็ญ ..</p> <p>รหัสประจำตัวนิสิต52xxxxxx26.....</p> <p>ครั้งที่3.....</p> <p><u>จงใส่คำศัพท์ที่เกี่ยวข้องกับโครงการปี 4 ที่นิสิตกำลัง</u> <u>ดำเนินงาน</u></p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>การเปรียบเทียบ</td></tr> <tr><td>คุณภาพ</td></tr> <tr><td>เว็บไซต์</td></tr> <tr><td>องค์การท่องเที่ยว</td></tr> <tr><td>AEC</td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> </table>	การเปรียบเทียบ	คุณภาพ	เว็บไซต์	องค์การท่องเที่ยว	AEC				<p><u>จงประเมินเอกสารว่ามีความเกี่ยวข้องกับโครงการปี 4</u> <u>ของนิสิตหรือไม่</u></p> <table border="1" style="width: 100%;"> <thead> <tr> <th>รหัสเอกสาร</th> <th>ความเกี่ยวข้อง</th> </tr> </thead> <tbody> <tr><td>53B15</td><td></td></tr> <tr><td>53B14</td><td>/</td></tr> <tr><td>53B10</td><td>/</td></tr> <tr><td>53B07</td><td></td></tr> <tr><td>53B06</td><td></td></tr> <tr><td>53B04</td><td></td></tr> <tr><td>53B03</td><td></td></tr> <tr><td>53B02</td><td></td></tr> <tr><td>53B01</td><td></td></tr> <tr><td>53B26</td><td></td></tr> <tr><td>53B25</td><td></td></tr> <tr><td>53B24</td><td>/</td></tr> <tr><td>53B22</td><td>/</td></tr> <tr><td>52B23</td><td></td></tr> <tr><td>53B21</td><td>/</td></tr> <tr><td>53B20</td><td>/</td></tr> <tr><td>53B19</td><td></td></tr> <tr><td>53B17</td><td>/</td></tr> <tr><td>53B12</td><td>/</td></tr> <tr><td>53B13</td><td></td></tr> <tr><td>53B16</td><td>/</td></tr> <tr><td>53B11</td><td></td></tr> <tr><td>53B18</td><td></td></tr> <tr><td>53B09</td><td></td></tr> <tr><td>53B08</td><td></td></tr> </tbody> </table>	รหัสเอกสาร	ความเกี่ยวข้อง	53B15		53B14	/	53B10	/	53B07		53B06		53B04		53B03		53B02		53B01		53B26		53B25		53B24	/	53B22	/	52B23		53B21	/	53B20	/	53B19		53B17	/	53B12	/	53B13		53B16	/	53B11		53B18		53B09		53B08	
การเปรียบเทียบ																																																													
คุณภาพ																																																													
เว็บไซต์																																																													
องค์การท่องเที่ยว																																																													
AEC																																																													
รหัสเอกสาร	ความเกี่ยวข้อง																																																												
53B15																																																													
53B14	/																																																												
53B10	/																																																												
53B07																																																													
53B06																																																													
53B04																																																													
53B03																																																													
53B02																																																													
53B01																																																													
53B26																																																													
53B25																																																													
53B24	/																																																												
53B22	/																																																												
52B23																																																													
53B21	/																																																												
53B20	/																																																												
53B19																																																													
53B17	/																																																												
53B12	/																																																												
53B13																																																													
53B16	/																																																												
53B11																																																													
53B18																																																													
53B09																																																													
53B08																																																													

รูปที่ 3.7 ตัวอย่างแบบฟอร์มการประเมินความเกี่ยวข้องระหว่างข้อสอบถามและเอกสาร

3.3.3. การประเมินประสิทธิภาพการค้นคืน

จากงานวิจัยที่ได้นำเสนอไว้ ผู้วิจัยจะต้องเปรียบเทียบประสิทธิภาพการค้นคืนก่อนและหลังการนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศ ด้วยรายการผลลัพธ์การค้นคืนของการใช้ข้อสอบถามจากผู้ใช้ในระบบการค้นคืนสารสนเทศ และรายการผลลัพธ์การค้นคืนของการใช้ข้อสอบถามจากการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศ ผู้วิจัยจึงกำหนดขั้นตอนการสร้างรายการผลลัพธ์การค้นคืนสองรายการดังกล่าว ดังรูปที่ 3.8 ซึ่งมีขั้นตอนดังต่อไปนี้ (1) ผู้วิจัยนำรายการข้อสอบถามจากผู้ใช้เข้าสู่ระบบการค้นคืนสารสนเทศ และบันทึกรายการผลลัพธ์การค้นคืนที่ได้ครั้งที่ 1 ทั้งหมด (2) ผู้วิจัยนำรายการข้อสอบถามจากผู้ใช้ไปขยายข้อสอบถาม และบันทึกรายการข้อสอบถามจากการขยายข้อสอบถามทั้งหมด และ (3) ผู้วิจัยนำรายการข้อสอบถามจากการขยายข้อสอบถามเข้าสู่ระบบการค้นคืนสารสนเทศ และบันทึกรายการผลลัพธ์การค้นคืนที่ได้ครั้งที่ 2 ทั้งหมด



รูปที่ 3.8 รูปแบบการขยายข้อสอบถามร่วมกับการค้นคืนเอกสาร

เมื่อผู้วิจัยได้รายการผลลัพธ์การค้นคืนจากข้อสอบถามจากผู้ใช้ และรายการผลลัพธ์การค้นคืนจากการขยายข้อสอบถามแล้ว ผู้วิจัยจึงนำรายการผลลัพธ์การค้นคืนทั้งหมดมาประเมินประสิทธิภาพการค้นคืน เพื่อศึกษาการนำการขยายข้อสอบถามมาใช้ ตามวัตถุประสงค์ของงานวิจัย ในการประเมินประสิทธิภาพการค้นคืนนั้น ผู้วิจัยกำหนดนำค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรวัดเอฟ มาใช้ในการประเมินประสิทธิภาพการค้นคืน ในการดำเนินงานวิจัยนี้ ผู้วิจัยจึงได้พัฒนาชุดคำสั่งการประเมินประสิทธิภาพการค้นคืน ด้วยการคำนวณค่าความแม่นยำที่อาร์ และค่ามาตรวัดเอฟ ซึ่งมีสมการการคำนวณดังต่อไปนี้

(4) การวิเคราะห์ข้อสอบถามและกลุ่มคำ

จากรูปที่ 3.2 ผู้วิจัยนำรายการข้อสอบถามเข้าสู่การวิเคราะห์ร่วมกับกลุ่มคำ เพื่อเลือกกลุ่มคำสำหรับการขยายข้อสอบถาม ทั้งนี้ผู้วิจัยได้กำหนดไว้ หากคำในข้อสอบถามทั้งหมด ตรงกับรายการคำในกลุ่มคำใดมากที่สุด ผู้วิจัยจะเลือกกลุ่มคำนั้นมาใช้ในการขยายข้อสอบถาม ทั้งนี้ผู้วิจัยได้นำคำทั้งหมดในกลุ่มคำที่เลือกมาเพิ่มเติมลงในข้อสอบถาม ยกเว้นคำที่มีอยู่แล้วในข้อสอบถาม

1.1. การประเมินประสิทธิภาพการค้นคืนสารสนเทศ (Retrieval Evaluation)

เมื่อวิเคราะห์หารายการเอกสารที่เกี่ยวข้องกับข้อสอบถามได้แล้ว ผู้วิจัยจึงนำรายการผลลัพธ์การค้นคืนที่ได้มานั้น มาร่วมประเมินประสิทธิภาพการค้นคืน ด้วยค่าความแม่นยำที่สิบ (Precision at 10 document, P@10) ค่าความแม่นยำที่ยี่สิบ (Precision at 20 document, P@20) ค่าความแม่นยำที่สามสิบ (Precision at 30 document, P@30) และค่ามาตรวัดเอฟ (F-measure, F) ดังที่งานวิจัยได้กำหนดไว้

$$P@R = \frac{r}{R}$$

$$F = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{precision} = \frac{|A \cap RT|}{|A|}$$

$$recall = \frac{|A \cap RT|}{|RT|}$$

โดยที่	R	เป็นจำนวนเอกสารลำดับต้นในรายการผลลัพธ์การค้นคืนที่ต้องการศึกษา
	r	เป็นจำนวนเอกสารที่เกี่ยวข้องกับข้อสอบถามและปรากฏในจำนวนเอกสารลำดับต้นในรายการผลลัพธ์การค้นคืนจำนวน R ฉบับ
	<i>precision</i>	เป็นค่าความแม่นยำที่คำนวณจากรายการผลลัพธ์การค้นคืนทั้งหมด
	<i>recall</i>	เป็นค่าที่ระลึกได้ที่คำนวณได้จากรายการผลลัพธ์การค้นคืนทั้งหมด
	A	เป็นเซตของรายการเอกสารที่เป็นรายการผลลัพธ์การค้นคืน
	RT	เป็นเซตของรายการเอกสารที่เป็นรายการเอกสารที่เกี่ยวข้องกับข้อสอบถามที่นำมาใช้ในการค้นคืน

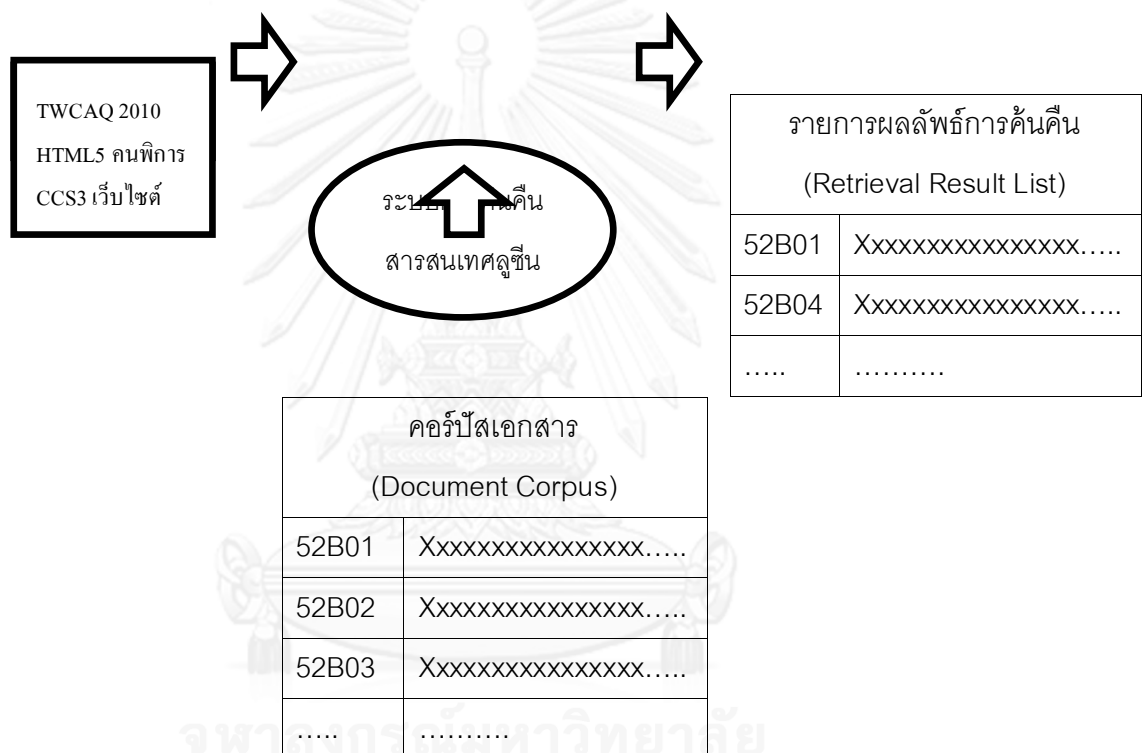
อย่างไรก็ตาม ในการคำนวณค่าความแม่นยำที่สืบ ค่าความแม่นยำที่สืบและค่าความแม่นยำที่สามสืบ หากจำนวนเอกสารในรายการผลลัพธ์การค้นคืนนั้นมีจำนวนน้อยกว่าจำนวนที่กำหนด ผู้วิจัยจะไม่สามารถคำนวณค่าความแม่นยำดังกล่าว กล่าวคือ ผู้วิจัยจะไม่สามารถประเมินประสิทธิภาพการค้นคืนได้ด้วยการคำนวณค่าความแม่นยำที่สืบ ค่าความแม่นยำที่สืบและค่าความแม่นยำที่สามสืบ

3.4. เครื่องมือในการเก็บข้อมูล

ในงานวิจัยนี้ ผู้วิจัยจึงต้องจัดเตรียมเครื่องมือที่นำมาใช้ในการศึกษาตามวัตถุประสงค์ของงานวิจัยนี้ เครื่องมือที่นำมาใช้ในงานวิจัยนี้ได้แก่

3.4.1. **เครื่องมือระบบการค้นคืนสารสนเทศลูซีน (Lucene)** เป็นระบบการค้นคืนสารสนเทศที่ผู้วิจัยได้นำมาใช้ในการค้นคืนสารสนเทศ ซึ่งสามารถรองรับการค้นคืนสารสนเทศภาษาไทยได้ อีกทั้งมีรูปแบบการค้นคืนแบบปริภูมิเวกเตอร์ การทำงานของระบบการค้นคืน

สารสนเทศนี้มีดังรูปที่ 3.9 ซึ่งข้อมูลที่น่าเข้าระบบการค้นคืนสารสนเทศนั้นจะเป็น (1) ข้อสอบถามที่นำมาใช้ในการค้นคืน ในงานวิจัยนี้ได้หมายถึงข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ลักษณะของข้อสอบถามที่นำมาใช้ เป็นข้อความ ดังรูปที่ 3.9 ข้อสอบถามที่นำมาใช้เป็น “TWCAQ 2010 HTML5 คนพิการ CCS3 เว็บไซต์” และ (2) รายการเอกสารทั้งหมดในคอร์ปัสเอกสาร ในงานวิจัยนี้ ผู้วิจัยได้นำข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม รวมทั้งรายการเอกสารในคอร์ปัสมาใช้ในการค้นคืนสารสนเทศ จนกระทั่งได้รายการผลลัพธ์การค้นคืน



รูปที่ 3.9 แสดงรูปแบบการทำงานของเครื่องมือระบบการค้นคืนสารสนเทศคู่ขึ้น

3.4.2. ส่วนขยายการวิเคราะห์ภาษาไทย ไทยอนาไลเซอร์ (ThaiAnalyzer) เป็นส่วนขยายหนึ่งของระบบการค้นคืนสารสนเทศคู่ขึ้น ที่นำมาใช้ในการวิเคราะห์แยกคำภาษาไทย ในงานวิจัยนี้ได้นำมาใช้กับ (1) ระบบการค้นคืนสารสนเทศคู่ขึ้น เพื่อวิเคราะห์การแยกคำในสารสนเทศภาษาไทย และ (2) ยังนำมาใช้ในการวิเคราะห์คำในเอกสารภาษาไทย เพื่อวิเคราะห์รายการดัชนีและค่านำหนักของคำในแต่ละเอกสาร สำหรับการขยายข้อสอบถามด้วย

3.4.3. รายการคำหยุดภาษาไทย (Thai Stop-word list) ของวรวีร์ เกษร (พ.ศ. 2553) นำมาใช้ในการคัดกรองคำในเอกสาร เพื่อเลือกมาวิเคราะห์และขยายข้อสอบถาม หากคำที่ตรวจสอบปรากฏในรายการคำหยุดนี้ จะไม่นำคำดังกล่าวมารวมวิเคราะห์และขยายข้อสอบถาม

3.4.4. พจนานุกรมอิเล็กทรอนิกส์ (Electronic Dictionary) จากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronic and Computer Technology Center) ที่จัดทำในปี พ.ศ.2552 ผู้วิจัยได้นำมาใช้ในการตรวจสอบการสะกดของคำ ในขั้นตอนการจัดเตรียมเอกสาร เพื่อการขยายข้อสอบถาม

3.4.5. เครื่องมือการทำเหมืองข้อมูลเวก้า (Weka) ในงานวิจัยนี้ ผู้วิจัยได้นำเครื่องมือนี้มาใช้ในการวิเคราะห์กลุ่มคำ เพื่อนำมาใช้ในการขยายข้อสอบถาม จากเมทริกซ์แสดงการปรากฏร่วมกันของคำ รูปที่ 3.10 แสดงตัวอย่างข้อมูลก่อนการวิเคราะห์กลุ่มคำ ซึ่งจำลองจากเมทริกซ์แสดงการปรากฏร่วมกันของคำ เป็นไฟล์ซีเอสวี (Comma-Separated Value, CSV) โดยที่ข้อมูลในแถวแรกและคอลัมน์แรก เป็นรายการคำที่นำมาใช้ในการวิเคราะห์ ส่วนค่าตัวเลขในตารางแต่ละตัว แสดงระดับความสัมพันธ์ระหว่างคำ ส่วนผลลัพธ์ที่ได้จากการวิเคราะห์กลุ่มคำนั้น จะเป็นดังรูปที่ 3.11 ซึ่งในแต่ละกลุ่มคำนั้น จะแสดงรายการคำ ตามความสัมพันธ์ของคำ

term	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
business	0.094869	0.142137	0.334642	0.107501	0.116325	0.044288	0.310924	0.239934	0.133597	0.108845	0.167321	0.131386	0.158283	0.127405	0.142						
home	0.090008	0.166071	0.339092	0.119433	0.053868	0.013571	0.332142	0.238866	0.097748	0.134238	0.195496	0.148858	0.086595	0.16178	0.166						
original	0.246042	0.355342	0.836606	0.272588	0.123211	0.035408	0.710684	0.511103	0.209151	0.272112	0.418303	0.318512	0.137533	0.329572	0.355						
design	0.024582	0.468509	1.002471	1.041714	0.383443	0.176541	0.325729	0.766886	0.234255	0.123139	0.029226	0.651458	0.468509	0.191721	0.249435	0.383443	0.291969	0.16165	0.291969	0.325	
business	0.167321	0.195496	0.418303	0.383443	0.16	0.073666	0.135918	0.32	0.097748	0.040198	0.007128	0.271835	0.195496	0.08	0.104082	0.16	0.12183	0.052606	0.12183	0.135	
home	0.094869	0.090008	0.246042	0.176541	0.073666	0.411479	0.062578	0.147331	0.045004	0.053572	0.021863	0.125155	0.123761	0.036833	0.247589	0.073666	0.056092	0.093874	0.056092	0.078	
original	0.142137	0.166071	0.355342	0.325729	0.135918	0.062578	0.11546	0.271835	0.083035	0.034147	0.006056	0.23092	0.166071	0.067959	0.088416	0.135918	0.103493	0.044688	0.103493	0.11	
design	0.334642	0.390992	0.836606	0.766886	0.32	0.147331	0.271835	0.64	0.195496	0.080396	0.014257	0.54367	0.390992	0.16	0.208165	0.32	0.243661	0.105212	0.243661	0.271	
business	0.107501	0.119433	0.272588	0.234255	0.097748	0.045004	0.083035	0.195496	0.100254	0.027287	0.006481	0.166071	0.119433	0.048874	0.063586	0.097748	0.074429	0.032138	0.074429	0.083	
home	0.116325	0.053868	0.123211	0.123139	0.040198	0.053572	0.034147	0.080396	0.027287	0.052115	0.017913	0.068295	0.074305	0.035791	0.064193	0.040198	0.031689	0.034728	0.03238	0.04	
original	0.044288	0.013571	0.035408	0.029226	0.007128	0.021863	0.006056	0.014257	0.006481	0.017913	0.014304	0.012489	0.016549	0.021612	0.0102	0.007128	0.006799	0.028679	0.005899	0.008	
design	0.310924	0.332142	0.710684	0.651458	0.271835	0.125155	0.23092	0.54367	0.166071	0.068295	0.012489	0.490705	0.332142	0.135918	0.176833	0.271835	0.206986	0.089376	0.206986	0.23	
business	0.239934	0.238866	0.511103	0.468509	0.195496	0.123761	0.166071	0.390992	0.119433	0.074305	0.016549	0.332142	0.388204	0.097748	0.127173	0.195496	0.148858	0.10078	0.148858	0.228	
home	0.133597	0.097748	0.209151	0.191721	0.08	0.036833	0.067959	0.16	0.048874	0.035791	0.021612	0.135918	0.097748	0.465264	0.074629	0.08	0.074135	0.026303	0.060915	0.067	
original	0.088416	0.208165	0.104082	0.247589	0.088416	0.208165	0.063586	0.064193	0.0102	0.176833	0.127173	0.074629	0.510155	0.104082	0.086132	0.034221	0.079252	0.034221	0.079252	0.088	
design	0.167321	0.195496	0.418303	0.383443	0.16	0.073666	0.135918	0.32	0.097748	0.040198	0.007128	0.271835	0.195496	0.08	0.104082	0.16	0.12183	0.052606	0.12183	0.135	
business	0.131386	0.148858	0.318512	0.291969	0.12183	0.056092	0.103493	0.243661	0.074429	0.031689	0.006799	0.206986	0.148858	0.074135	0.086132	0.12183	0.102591	0.040056	0.092766	0.103	
home	0.158283	0.086595	0.137533	0.16165	0.052606	0.093874	0.044688	0.105212	0.032138	0.034275	0.028679	0.089376	0.10078	0.026303	0.034221	0.052606	0.040056	0.229992	0.040056	0.105	
original	0.127405	0.16178	0.329572	0.291969	0.12183	0.056092	0.103493	0.243661	0.074429	0.03238	0.005899	0.206986	0.148858	0.060915	0.079252	0.12183	0.092766	0.040056	0.108872	0.103	
design	0.142137	0.166071	0.355342	0.325729	0.135918	0.062578	0.11546	0.271835	0.083035	0.04553	0.008326	0.23092	0.228347	0.067959	0.088416	0.135918	0.103493	0.05586	0.103493	0.144	
business	0.233797	0.060575	0.165047	0.134128	0.044116	0.090264	0.037476	0.088232	0.04155	0.057084	0.029963	0.089005	0.097449	0.034153	0.120871	0.044116	0.03505	0.058356	0.036507	0.056	
home	0.155854	0.065993	0.140913	0.180049	0.046789	0.038519	0.039746	0.093578	0.057169	0.070738	0.046517	0.114271	0.072851	0.188171	0.030437	0.046789	0.037853	0.07154	0.038719	0.044	
original	0.116952	0.136646	0.292381	0.268015	0.111835	0.05149	0.095002	0.22367	0.102484	0.039398	0.011726	0.190005	0.136646	0.055918	0.07275	0.111835	0.085156	0.051964	0.085156	0.095	
design	0.228668	0.132758	0.244718	0.23108	0.093604	0.080583	0.079515	0.187209	0.064333	0.061398	0.039847	0.16897	0.163878	0.11993	0.060891	0.093604	0.073501	0.09053	0.071274	0.099	

รูปที่ 3.10 แสดงตัวอย่างข้อมูลก่อนการวิเคราะห์กลุ่มคำด้วยเครื่องมือการทำเหมืองข้อมูลเวก้า

ลำดับกลุ่ม	รายการค่า
1	$W_{1,1}, W_{1,2}, W_{1,3}, W_{1,4}$
2	$W_{2,1}, W_{2,2}, W_{2,3}, W_{2,4}, W_{2,5}, W_{2,6}, W_{2,7}, W_{2,8}, W_{2,9}, W_{2,10}, \dots$
3	$W_{3,1}, W_{3,2}, W_{3,3}$
4	$W_{4,1}$
5	$W_{5,1}, W_{5,2}, W_{5,3}, W_{5,4}, W_{5,5}, W_{5,6}, W_{5,7}, W_{5,8}, W_{5,9}, W_{5,10}, \dots$

รูปที่ 3.11 แสดงตัวอย่างผลลัพธ์การจัดกลุ่มค่าที่ได้

3.5. ความถูกต้อง (Validity) และความน่าเชื่อถือ (Reliability) ของข้อมูลที่เก็บ

งานวิจัยนี้เป็นการทดลองในห้องปฏิบัติการ เพื่อตอบวัตถุประสงค์ทั้งหมดของงานวิจัย ดังนั้น การควบคุมการทดลองจึงมีความสำคัญมาก เพื่อให้ได้ข้อมูลที่ดีที่สุดสำหรับการนำไปสรุปผลที่ถูกต้องและน่าเชื่อถือ จึงต้องควบคุมปัจจัยต่างๆ

3.5.1. **การกำหนดกลุ่มตัวอย่าง** ผู้วิจัยเลือกนิสิตและเอกสารโครงงานพิเศษของภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย เป็นหน่วยทดลอง เนื่องจากกลุ่มนิสิตจะมีความสามารถและความรู้พื้นฐานที่ทัดเทียมกัน และสามารถทำความเข้าใจในเอกสารโครงงานชั้นปี 4 ที่นำมาใช้ในการทดลองได้ เนื่องจากผู้ประเมินนั้นจะต้องมีความรู้ความเข้าใจในเอกสารที่จะต้องนำมาประเมิน (Xu and Chen, 2005, 2006, Azzopardi)

3.5.2. **เครื่องมือที่นำมาใช้ในการพัฒนาระบบการค้นคืน** เพื่อนำผลลัพธ์การค้นคืนมาใช้ในการประเมินประสิทธิภาพการค้นคืน ผู้วิจัยจึงต้องพัฒนาระบบการค้นคืนเพื่อนำข้อสอบถามที่ได้มาใช้ในการดำเนินงานวิจัย เครื่องมือที่นำมาใช้ในการพัฒนาระบบการค้นคืนนั้น เป็นเครื่องมือที่นิยมนำมาใช้ในงานวิจัยในอดีต (ต้นติมมา เวพูกานนท์ และ สมชาย ปราการเจริญ, 2554; วารินนารวิทย์และศิริภิญโญจันทร์, 2554; Olzye; Costa Pereira, Célia da, Dragoni, Mauro and Pasi, Gabriella, 2010; Burrows and Malheiro, 2008)

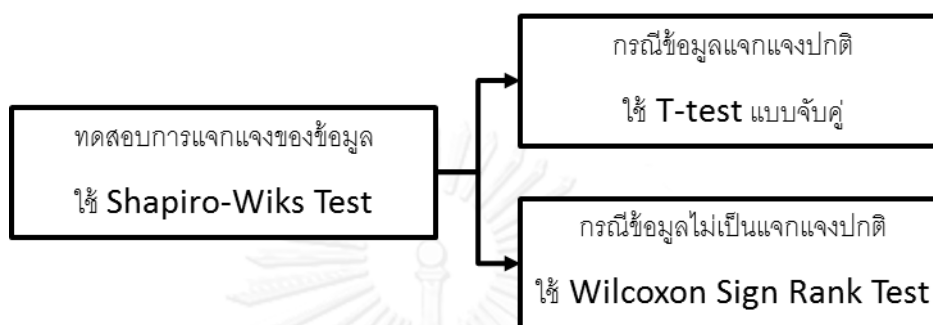
3.6. กรอบการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเพื่อทดสอบสมมติฐานและตอบวัตถุประสงค์ของงานวิจัย และการเก็บรวบรวมข้อมูลของงานวิจัยเป็นการเก็บข้อมูลจากหน่วยทดลอง โดยสร้างชุดทดสอบและระบบการค้นคืนเอกสาร เพื่อเก็บข้อมูลสำหรับการวิเคราะห์หาค่าความแม่นยำอาร์และค่ามาตรวัดเอฟ ซึ่งจะกระทำดังต่อไปนี้

3.6.1. ในงานวิจัยนี้ ได้แบ่งตัวแปรออกเป็นสองกลุ่ม คือกลุ่มที่หนึ่งเป็นตัวแปรต้นหรือตัวแปรอิสระ (Independent Variable) ได้แก่ การใช้ระบบการขยายข้อสอบถามด้วยการจัดกลุ่มแบบลำดับขั้น กลุ่มที่สองเป็นตัวแปรตาม (Dependent Variable) ได้แก่ (1) ค่าความแม่นยำจำนวนสิบรายการแรก (Precision at 10)(2) ค่าความแม่นยำจำนวนยี่สิบรายการแรก (Precision at 20)(3) ค่าความแม่นยำจำนวนสามสิบรายการแรก (Precision at 30)(1) และ (4) ค่ามาตรวัดเอฟโดยผู้วิจัยสนใจว่าตัวแปรต้นจะมีอิทธิพลต่อตัวแปรตามหรือไม่

3.6.2. ในการตอบวัตถุประสงค์ของงานวิจัย จะมีกรอบการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ โดยจะตรวจสอบว่า สามารถนำมาใช้วิธีการทางสถิติแบบอิงพารามิเตอร์หรือไม่อิงพารามิเตอร์ เพื่อนำมาใช้ในการวิเคราะห์ข้อมูล จากรูปที่ 3. จะเห็นได้ว่า ข้อสอบถามจากผู้ใช้นั้นจะนำไปใช้ในการค้นคืนสารสนเทศและการขยายข้อสอบถาม แล้วจึงได้ข้อสอบถามที่ขยายแล้วไปใช้ในการค้นคืนสารสนเทศอีกครั้ง จนได้รายการผลลัพธ์ในการค้นคืนจำนวนสองชุด ได้แก่ รายการผลลัพธ์การค้นคืนด้วยข้อสอบถามจากผู้ใช้ และรายการผลลัพธ์การค้นคืนด้วยข้อสอบถามที่ขยายแล้ว ในการเปรียบเทียบผลการประเมินการค้นคืนนั้นต่อหนึ่งข้อสอบถามจากผู้ใช้ จะสามารถนำรายการผลลัพธ์การค้นคืนทั้งสองชุดมาเปรียบเทียบได้ ซึ่งเกิดจากข้อสอบถามจากผู้ใช้เพียงหนึ่งข้อสอบถาม กล่าวคือรายการผลลัพธ์การค้นคืนทั้งสองชุดนั้นไม่ได้เป็นอิสระกัน นอกจากนี้แล้ว หน่วยตัวอย่างที่นำมาใช้ในงานวิจัยนี้มีขนาด 25 ข้อสอบถาม ดังนั้นผู้วิจัยเลือกใช้การทดสอบวิลส์ชาปิโร (Shapiro-Wilk Test) ในการทดสอบการแจกแจงของข้อมูล หากพบว่าการแจกแจงของตัวแปรตามเป็นการแจกแจงปกติ ผู้วิจัยจะเลือกใช้วิธีการทางสถิติอิงพารามิเตอร์ (Parametric Statistic Technique) คือการทดสอบทีแบบจับคู่ (Paired T-Test) แต่หากไม่เป็นการแจกแจงแบบปกติ ผู้วิจัยจะเลือกใช้สถิติแบบไม่อิงพารามิเตอร์ (Non-Parametric Statistic Technique) คือการใช้การทดสอบลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test) (กัลยา วานิชย์

บัญชา, 2550, 2554; ปรภายรัตน์ สุวรรณ และอมรวิทย์ วิเศษสงวน, 2555) การนำการทดสอบทางสถิติในงานวิจัยนี้มีรูปแบบการตัดสินใจดังรูปที่ 3.12



รูปที่ 3.12 แสดงการกำหนดสถิติที่จะนำมาใช้ในการวิเคราะห์ข้อมูล

บทที่ 4 ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีจุดประสงค์เพื่อศึกษาการนำการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันของคำด้วยขั้นตอนวิธีการจัดกลุ่มแบบลำดับมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศภาษาไทย

ทั้งนี้ผู้วิจัยได้กำหนดการใช้ค่าความแม่นยำที่สิบ (Precision at 10 documents) ค่าความแม่นยำที่ยี่สิบ (Precision at 20 documents) ค่าความแม่นยำที่สามสิบ (Precision at 30 documents) และค่ามาตรวัดเอฟ (F-measure) มาใช้ในการวิเคราะห์และประเมินประสิทธิภาพการค้นคืนสารสนเทศก่อนและหลังการนำการขยายข้อสอบถามมาประยุกต์ใช้กับการค้นคืนสารสนเทศ

4.1. การเก็บข้อมูลสำหรับงานวิจัย

ในหัวข้อนี้ ผู้วิจัยขอยกตัวอย่างข้อสอบถามหนึ่งข้อสอบถามที่ได้จากผู้ประเมินความเกี่ยวข้องเอกสารและข้อสอบถาม 1 รายคือ “Phone app html5 java” ซึ่งผู้วิจัยกำหนดเป็นข้อสอบถาม Q1 ซึ่งประกอบด้วยสี่คำ (term) ผู้ประเมินความเกี่ยวข้องได้ประเมินไว้ว่ามีเอกสารทั้งหมด 48 ฉบับจากเอกสารทั้งหมด 100 ฉบับ มีความเกี่ยวข้องกับข้อสอบถาม Q1 นี้ ดังตาราง 4.1

ตารางที่ 4.1 แสดงรายการเอกสารที่เกี่ยวข้องกับข้อสอบถาม Q1

54B06	54B23	54B27	54B28	54B31	52B15	53B26	54B01	54B10	54B11
52B06	52B14	52B18	52B23	52B25	52B34	53B08	54B05	52B20	52B27
52B30	53B09	53B12	53B15	53B19	53B27	54B18	52B03	52B04	52B31
54B37	54B12	52B02	52B10	52B33	53B01	53B02	53B03	53B17	53B22
53B25	53B29	54B09	54B14	54B20	54B24	54B25	54B29		

4.1.1. การประเมินประสิทธิภาพการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

ผู้วิจัยได้นำข้อสอบถาม Q1 เข้าสู่ระบบการค้นคืนสารสนเทศ ผลลัพธ์การค้นคืนแสดงดังตารางที่ 4.2 มีจำนวน 10 ฉบับ ดังนั้นผลการค้นคืนจากระบบการค้นคืนสารสนเทศตาม Q1 ถูกนำไปเปรียบเทียบกับรายการเอกสารที่เกี่ยวข้องกับข้อสอบถาม Q1 ดังตารางที่ 4.2 โดยที่ Y หมายถึง เอกสารฉบับนี้มีความเกี่ยวข้องกับข้อสอบถาม Q1 และ N หมายถึง เอกสารฉบับนี้ไม่มีความเกี่ยวข้องกับข้อสอบถาม Q1

ตารางที่ 4.2 ผลลัพธ์การค้นคืนของการใช้ข้อสอบถาม Q1 ในระบบการค้นคืนสารสนเทศและผลการประเมินผลลัพธ์การค้นคืน

ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)	ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)
1	54B31	Y	6	54B23	Y
2	54B04	N	7	54B06	Y
3	54B12	Y	8	53B08	Y
4	52B04	Y	9	54B27	Y
5	52B15	Y	10	53B26	Y

ผู้วิจัยจึงสามารถคำนวณค่าความแม่นยำที่สิบ (P@10) ค่าความแม่นยำที่ยี่สิบ (P@20) ค่าความแม่นยำที่สามสิบ (P@30) และค่ามาตรวัดเอฟ (F) ได้ดังตารางที่ 4.3 โดยที่ค่าความแม่นยำที่ยี่สิบและค่าความแม่นยำที่สามสิบ ไม่สามารถหาได้ เนื่องจากจำนวนผลลัพธ์การค้นคืนมีเพียงสิบรายการเท่านั้น

ตารางที่ 4.3 แสดงผลการประเมินประสิทธิภาพการค้นคืน เมื่อใช้ข้อสอบถาม Q1

	P@10	P@20	P@30	F
ผลการประเมิน	0.9	-	-	0.1552

4.1.2. การประเมินประสิทธิภาพการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

ผู้วิจัยได้นำข้อสอบถาม Q1 เข้าสู่กระบวนการขยายข้อสอบถามด้วยกลุ่มคำ เพื่อให้ระบบดังกล่าวได้คัดเลือกกลุ่มคำที่จะนำมาใช้ในการขยายข้อสอบถาม ด้วยการเปรียบเทียบรายการคำในข้อสอบถามกับรายการคำในแต่ละกลุ่มคำ ดังที่ปรากฏอยู่ในภาคผนวก ง จากนั้นผู้วิจัยจึงนำกลุ่มที่ได้มาเพิ่มเติมลงในข้อสอบถาม Q1 จนกระทั่งได้ข้อสอบถามใหม่ เป็น

"Phone app html5 java โครงงาน ภายใน สร้าง ควบคุม ทำ ไร่ ศึกษา ประยุกต์ study สำหรับ หรือ ดึง รูป รายงาน เรื่อง ตรวจ สอบ ยาน ผล หลัก การนำ สูง ประจำ สอง สะดวก คล่องตัว ช่อง ใหม่ visual studio c ชื่อ ค่า สรุป สารสนเทศ เข้าถึง หา สังคม บัณฑิต ตอบ โอกาส นิยม ชั่วโม่ง ชัดเจน ลด control traffic หนาแน่น อ้างอิง ฟัง สู่ สำเร็จ reader เด่น technology report สปีดาร์ห์ แดง เขียว hardware tutorial"

โดยข้อสอบถามใหม่ที่ได้มานั้น ผู้วิจัยกำหนดเป็นข้อสอบถาม Q2 ซึ่งประกอบด้วย 67 คำ (terms)

ผู้วิจัยได้นำข้อสอบถาม Q2 เข้าสู่ระบบการค้นคืนสารสนเทศ ได้ผลลัพธ์การค้นคืนดังตารางที่ 4.4 โดยที่รายการผลลัพธ์การค้นคืนที่ได้มีจำนวน 100 ฉบับ หลังจากนั้น และเมื่อนำไปเปรียบเทียบกับรายการเอกสารที่เกี่ยวข้องกับข้อสอบถาม Q1 จนกระทั่งได้รายการเอกสารที่เกี่ยวข้องกับข้อสอบถามดังตารางที่ 4.1 ผู้วิจัยจึงสามารถประเมินความเกี่ยวข้องของเอกสารแต่ละฉบับในรายการผลลัพธ์การค้นคืนทั้งหมด ดังที่กำหนดในตารางที่ 4.4 โดยที่ Y หมายถึง เอกสารฉบับนี้มีความเกี่ยวข้องกับข้อสอบถาม Q1 และ N หมายถึง เอกสารฉบับนี้ไม่มีความเกี่ยวข้องกับข้อสอบถาม Q1

ตารางที่ 4.4 ผลลัพธ์การค้นคืนของการใช้ข้อสอบถาม Q2 ในระบบการค้นคืนสารสนเทศ รวมทั้ง
ผลการประเมินผลลัพธ์การค้นคืน

ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)	ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)
1	54B23	Y	25	53B15	Y
2	54B31	Y	26	54B01	Y
3	54B18	Y	27	52B04	Y
4	54B15	N	28	53B21	N
5	52B03	Y	29	53B12	Y
6	53B25	Y	30	54B10	Y
7	54B06	Y	31	52B25	Y
8	54B29	Y	32	54B19	N
9	53B02	Y	33	53B17	Y
10	53B04	N	34	53B18	N
11	54B04	N	35	54B14	Y
12	54B36	N	36	52B15	Y
13	52B09	N	37	53B29	Y
14	52B23	Y	38	52B31	Y
15	53B28	N	39	54B38	N
16	53B13	N	40	53B10	N
17	52B16	N	41	53B31	N
18	53B16	N	42	52B33	Y
19	52B27	Y	43	53B09	Y
20	54B16	N	44	54B08	N
21	53B30	N	45	53B27	Y
22	52B32	N	46	52B34	Y
23	52B02	Y	47	54B32	N
24	52B22	N	48	52B07	N

ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)
49	54B25	Y
50	52B28	N
51	54B33	N
52	53B26	Y
53	53B24	N
54	53B11	N
55	54B28	Y
56	52B11	N
57	54B30	N
58	53B19	Y
59	52B18	Y
60	52B20	Y
61	53B22	Y
62	54B22	N
63	53B07	N
64	54B13	N
65	52B08	N
66	53B03	Y
67	54B17	N
68	53B01	Y
69	54B34	N
70	52B30	Y
71	52B10	Y
72	54B37	Y
73	52B06	Y
74	54B11	Y

ลำดับ	เอกสาร	ความเกี่ยวข้อง (Y/N)
75	54B24	Y
76	54B12	Y
77	54B21	N
78	54B03	N
79	54B27	Y
80	54B05	Y
81	53B14	N
82	54B02	N
83	52B05	N
84	52B14	Y
85	53B37	N
86	53B06	N
87	54B09	Y
88	53B23	N
89	54B35	N
90	52B12	N
91	54B20	Y
92	54B07	N
93	53B20	N
94	53B08	Y
95	52B01	N
96	52B26	N
97	52B17	N
98	52B21	N
99	52B13	N
100	52B35	N

ดังนั้น ผู้วิจัยจึงสามารถคำนวณค่าความแม่นยำที่สิบ ($P@10$) ค่าความแม่นยำที่ยี่สิบ ($P@20$) ค่าความแม่นยำที่สามสิบ ($P@30$) และค่ามาตรวัดเอฟ (F) ได้ดังตารางที่ 4.5

ตารางที่ 4.5 แสดงผลการประเมินประสิทธิภาพการค้นคืน เมื่อใช้ข้อสอบถาม Q2

	P@10	P@20	P@30	F
ผลการประเมิน	0.8	0.5	0.533	0.3242

4.2. ผลการวิเคราะห์ประสิทธิภาพการค้นคืนสารสนเทศ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการนำการขยายข้อสอบถามด้วยการจัดกลุ่มการปรากฏร่วมกันของคำด้วยขั้นตอนวิธีการจัดกลุ่มแบบลำดับขั้นมาประยุกต์ใช้ร่วมกับระบบการค้นคืนสารสนเทศภาษาไทย ด้วยการนำรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามที่ได้จากผู้ใช้ก่อนการขยายข้อสอบถาม และรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามที่ได้จากการขยายข้อสอบถาม มาคำนวณและประเมินประสิทธิผลการค้นคืนด้วยค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรวัดเอฟ ตารางที่ 4.6 แสดงค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรวัดเอฟ ที่คำนวณได้จากรายการผลลัพธ์จำนวน 50 ชุด ซึ่งเป็นรายการผลลัพธ์การค้นคืนที่ได้จากการใช้ข้อสอบถามจากผู้ใช้จำนวน 25 ชุดและรายการผลลัพธ์การค้นคืนที่ได้จากข้อสอบถามจากการขยายข้อสอบถามจำนวน 25 ชุด โดยที่ $P@R_1$ และ F_1 แทนค่าความแม่นยำที่ R และค่ามาตรวัดเอฟที่ได้จากการคำนวณรายการผลลัพธ์การค้นคืนที่ได้จากผู้ใช้งาน ส่วน $P@R_2$ และ F_2 แทนค่าความแม่นยำที่ R และค่ามาตรวัดเอฟที่ได้จากการคำนวณรายการผลลัพธ์การค้นคืนที่ได้จากการขยายข้อสอบถามด้วยข้อสอบถามจากผู้ใช้และรายการผลลัพธ์การค้นคืนที่ได้ก่อนการขยายข้อสอบถาม

ตารางที่ 4.6 แสดงค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรฐานวัดเอฟ ที่คำนวณได้จากรายการผลลัพธ์การค้นคืน

ลำดับ	Q_1^1	Q_2^2	A_1^3	A_2^4	$P@10_1^5$	$P@10_2^6$	$P@20_1^7$	$P@20_2^8$	$P@30_1^9$	$P@30_2^{10}$	F_1^{11}	F_2^{12}
1	6	467	70	100	0.8500	0.9111	0.6625	0.8151	0.6833	0.8429	0.3933	0.4481
2	7	657	100	100	0.9197	0.9716	0.9399	0.9476	0.9085	0.9381	0.4609	0.4767
3	7	514	75	100	0.9040	0.9329	0.8995	0.9279	0.8702	0.9258	0.4060	0.4766
4	6	142	14	100	0.9459	0.8418	-	0.8187	-	0.8590	0.2003	0.4536
5	7	446	48	100	0.8608	0.9364	0.9100	0.9233	0.8572	0.9283	0.3193	0.4776
6	7	653	100	100	0.8792	0.9629	0.9158	0.9409	0.9114	0.9401	0.4737	0.4811
7	5	42	8	100	-	0.7258	-	0.7567	-	0.7494	0.1301	0.4382
8	3	183	22	100	0.9709	0.9547	0.9167	0.9460	-	0.9398	0.2540	0.4779
9	6	67	8	100	-	0.7889	-	0.7764	-	0.7998	0.1724	0.4472
10	7	305	39	100	0.9408	0.9548	0.9217	0.9458	0.8978	0.9452	0.3164	0.4780

¹ Q_1 หมายถึงขนาดของข้อสอบถามจากผู้ใช้นำมาใช้ในการค้นคืนสารสนเทศ

² Q_2 หมายถึงขนาดของข้อสอบถามจากการขยายข้อสอบถาม นำมาใช้ในการค้นคืนสารสนเทศ

³ A_1 หมายถึงจำนวนเอกสารในรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

⁴ A_2 หมายถึงจำนวนเอกสารในรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

⁵ $P@10_1$ หมายถึงค่าความแม่นยำที่สิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

⁶ $P@10_2$ หมายถึงค่าความแม่นยำที่สิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

⁷ $P@20_1$ หมายถึงค่าความแม่นยำที่ยี่สิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

⁸ $P@20_2$ หมายถึงค่าความแม่นยำที่ยี่สิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

⁹ $P@30_1$ หมายถึงค่าความแม่นยำที่สามสิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

¹⁰ $P@30_2$ หมายถึงค่าความแม่นยำที่สามสิบ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

¹¹ F_1 หมายถึงค่ามาตรวัดเอฟ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

¹² F_2 หมายถึงค่ามาตรวัดเอฟ ที่ประเมินจากรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

ลำดับ	Q_1^1	Q_2^2	A_1^3	A_2^4	$P@10_1^5$	$P@10_2^6$	$P@20_1^7$	$P@20_2^8$	$P@30_1^9$	$P@30_2^{10}$	F_1^{11}	F_2^{12}
11	7	519	70	100	0.8192	0.9499	0.8429	0.9305	0.8369	0.9199	0.4143	0.4777
12	4	477	63	100	0.8967	0.9304	0.8779	0.9188	0.8440	0.9244	0.3862	0.4811
13	5	97	11	100	0.6456	0.8842	-	0.8237	-	0.8579	0.1230	0.4490
14	3	8	1	1	-	-	-	-	-	-	0.0288	0.0288
15	4	13	2	76	-	0.7000	-	0.6500	-	0.7000	0.0667	0.3819
16	4	505	72	100	0.9000	0.9514	0.8500	0.9540	0.8333	0.9490	0.4104	0.4781
17	4	66	10	100	0.8514	0.9010	-	0.9280	-	0.9196	0.1537	0.4699
18	6	119	9	100	-	0.9248	-	0.9281	-	0.9121	0.0975	0.4712
19	5	654	100	100	0.9000	0.9522	0.7667	0.9539	0.7431	0.9390	0.4368	0.4767
20	9	622	93	100	0.9292	0.9501	0.8480	0.9359	0.8556	0.9306	0.4297	0.4775
21	4	209	22	100	0.8722	0.9081	0.5389	0.9066	-	0.9052	0.2202	0.4722
22	5	566	80	100	0.8817	0.9392	0.8121	0.7672	0.7803	0.7872	0.3851	0.4551
23	3	381	46	100	0.9333	0.9297	0.7917	0.9272	0.6222	0.9293	0.2884	0.4775
24	5	639	95	100	0.8850	0.9612	0.8498	0.9394	0.8363	0.9365	0.4346	0.4813
25	3	61	7	100	-	0.9269	-	0.9410	-	0.9265	0.2846	0.4786

4.2.1. การวิเคราะห์ค่าความแม่นยำที่ลิบ

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะต้องตรวจสอบว่า ค่าความแม่นยำที่ลิบที่ได้จากรายการผลลัพธ์การค้นคืนนั้น แจกแจงแบบปกติหรือไม่ จึงตั้งสมมติฐานดังนี้

- 1) ทดสอบการแจกแจงค่าความแม่นยำที่ลิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

H_0 : ข้อมูลค่าความแม่นยำที่ลิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่ลิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ไม่แจกแจงแบบปกติ

- 2) ทดสอบการแจกแจงค่าความแม่นยำที่สิบของการคั่นคั้นสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

H_0 : ข้อมูลค่าความแม่นยำที่สิบของการคั่นคั้นสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่สิบของการคั่นคั้นสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่แจกแจงแบบปกติ

ผู้วิจัยได้เลือกใช้แบบทดสอบซิปโร-วิลค์ ในการตรวจสอบการแจกแจงค่าความแม่นยำที่สิบ ผู้วิจัยจะยอมรับสมมติฐาน H_0 ถ้าค่านัยสำคัญ (Significance, Sig) มีค่ามากกว่า 0.05 ตารางที่ 4.7 แสดงการทดสอบการนำค่าความแม่นยำที่สิบของการคั่นคั้นสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และจากการขยายข้อสอบถามมาทดสอบการแจกแจงแบบปกติ จะแสดงให้เห็นว่า

ตารางที่ 4.7 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่สิบ

ค่าตัวแปร	การนำข้อสอบถามมาใช้ในการคั่นคั้นสารสนเทศ	Shapiro-Wilk		
		Statistic	Df	Sig.
ค่าความแม่นยำที่สิบ	ข้อสอบถามจากผู้ใช้	0.783	19	0.001
	ข้อสอบถามจากการขยายข้อสอบถาม	0.872	19	0.016

- 1) การคั่นคั้นสารสนเทศด้วยข้อสอบถามจากผู้ใช้ มีค่านัยสำคัญเท่ากับ 0.001 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0
- 2) การคั่นคั้นสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มีค่านัยสำคัญเท่ากับ 0.016 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้น จึงสรุปได้ว่า ค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่ได้แจกแจงแบบปกติ ผู้วิจัยจึงเลือกใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non-Parameter Test) ด้วยการเลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test)

การเปรียบเทียบค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และข้อสอบถามจากการขยายข้อสอบถาม

จากการวิเคราะห์การแจกแจงของค่าความแม่นยำที่สืบข้างต้น ผู้วิจัยจึงทดสอบสมมติฐาน โดยที่เลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอนด้วยโปรแกรม IBM SPSS Statistic 20

กำหนดให้ $P@10_1$ แทนค่าความแม่นยำที่สืบที่คำนวณได้จากรายการผลลัพธ์การค้นคืนซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ หรือรายการผลลัพธ์การค้นคืนก่อนการขยายข้อสอบถาม และ $P@10_2$ แทนค่าความแม่นยำที่สืบที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม หรือรายการผลลัพธ์การค้นคืนหลังการขยายข้อสอบถามจากบทที่ 3 ผู้วิจัยได้กำหนดสมมติฐานการเปรียบเทียบค่าความแม่นยำที่สืบได้ดังนี้

H_0 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่สืบ ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_0 : P@10_2 \leq P@10_1$$

H_1 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่สืบ สูงกว่าผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_1 : P@10_2 > P@10_1$$

ผู้วิจัยได้นำค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มาทดสอบด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test) จะได้ผลลัพธ์การวิเคราะห์ทางสถิติซึ่งจะได้ผลลัพธ์ดังตารางที่ 4.8 และตารางที่ 4.9 จากผลการวิเคราะห์ค่าความแม่นยำที่สืบทางสถิติ คำนัยสำคัญที่ได้มีค่าเท่ากับ $\frac{0.003}{2} = 0.015$ ซึ่งมีค่าน้อยกว่าระดับค่านัยสำคัญ 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0 จึงสามารถสรุปได้ว่า ค่าความแม่นยำที่สืบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามนั้นมีค่าสูงกว่าค่าความแม่นยำที่สืบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

ตารางที่ 4.8 แสดงผลการวิเคราะห์ค่าความแม่นยำที่สืบที่ได้จากการวิเคราะห์

การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ		ข้อสอบถามจากผู้ใช้	ข้อสอบถามจากการขยายข้อสอบถาม
N		19	19
Means		0.8834	0.9328
Std. Deviation		0.0689	0.0319
Minimum		0.6456	0.8418
Maximum		0.9708	0.9716
Percentage	25 th	0.8607	0.9111
	50 th	0.8967	0.9392
	75 th	0.9217	0.9567

ตารางที่ 4.9 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่สืบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถามด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน

	ค่าความแม่นยำที่สืบ
Wilcoxon Sign Rank Test	148.00
Z	-2.978
Asymp. Sig (2-tailed)	0.003

4.2.2. การวิเคราะห์ค่าความแม่นยำที่ยี่สิบ

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะต้องตรวจสอบว่า ค่าความแม่นยำที่ยี่สิบที่ได้จากรายการผลลัพธ์การค้นคืนนั้น แจกแจงแบบปกติหรือไม่ จึงตั้งสมมติฐานดังนี้

1) ทดสอบการแจกแจงค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

H_0 : ข้อมูลค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ไม่แจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

H_0 : ข้อมูลค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่แจกแจงแบบปกติ

ผู้วิจัยได้เลือกใช้แบบทดสอบซapiro-วิลค ในการตรวจสอบการแจกแจงของค่าความแม่นยำที่ยี่สิบผู้วิจัยจะยอมรับสมมติฐาน H_0 ถ้าค่านัยสำคัญ (Significance, Sig) มีค่ามากกว่า 0.05 ตารางที่ 4.10 แสดงการทดสอบการนำค่าความแม่นยำที่ยี่สิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และจากการขยายข้อสอบถามมาทดสอบการแจกแจงแบบปกติ จะแสดงให้เห็นว่า

ตารางที่ 4.10 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่ยี่สิบ

ค่าตัวแปร	การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ	Shapiro-Wilk		
		Statistic	Df	Sig.
ค่าความแม่นยำที่ยี่สิบ	ข้อสอบถามจากผู้ใช้	0.819	16	0.005
	ข้อสอบถามจากการขยายข้อสอบถาม	0.647	16	0.000

1) การค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ มีค่านัยสำคัญเท่ากับ 0.005 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มีค่านัยสำคัญเท่ากับ 0.000 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้น จึงสรุปได้ว่า ค่าความแม่นยำที่ยีลึบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่ยีลึบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่ได้ແจกແจงแบบปกติ ผู้วิจัยจึงเลือกใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non-Parameter Test) ด้วยการเลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test)

การวิเคราะห์เปรียบเทียบค่าความแม่นยำที่ยีลึบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และข้อสอบถามจากการขยายข้อสอบถาม

จากการวิเคราะห์การແจกແจงของค่าความแม่นยำที่ยีลึบข้างต้น ผู้วิจัยจึงทดสอบสมมติฐาน โดยที่เลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอนด้วยโปรแกรม IBM SPSS Statistic 20

กำหนดให้ $P@20_1$ แทนค่าความแม่นยำที่ยีลึบที่คำนวณได้จากรายการผลลัพธ์การค้นคืนซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ หรือรายการผลลัพธ์การค้นคืนก่อนการขยายข้อสอบถาม และ $P@20_2$ แทนค่าความแม่นยำที่ยีลึบที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม หรือรายการผลลัพธ์การค้นคืนหลังการขยายข้อสอบถาม จากบทที่ 3 ผู้วิจัยได้กำหนดสมมติฐานการเปรียบเทียบค่าความแม่นยำที่ยีลึบได้ดังนี้

H_0 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่ยีลึบ ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_0: P@20_2 \leq P@20_1$$

H_1 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่ยีลึบ มากกว่าผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_1: P@20_2 > P@20_1$$

ผู้วิจัยได้นำค่าความแม่นยำที่ยีลึบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่ยีลึบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามมาทดสอบด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test) จะได้ผลลัพธ์การวิเคราะห์ทางสถิติซึ่งจะได้ผลลัพธ์ดังตารางที่ 4.11 และตารางที่ 4.12 จากผลการวิเคราะห์ค่าความแม่นยำที่ยีลึบทางสถิติ คำนัยสำคัญที่ได้มีค่าเท่ากับ

$\frac{0.002}{2} = 0.001$ ซึ่งมีค่าน้อยกว่าระดับค่านัยสำคัญ 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0 จึงสามารถสรุปได้ว่า ค่าความแม่นยำที่ยีลึบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามนั้นมีค่ามากกว่าค่าความแม่นยำที่ยีลึบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

ตารางที่ 4.11 แสดงผลการวิเคราะห์ค่าความแม่นยำที่ยีลึบที่ได้จากการวิเคราะห์

การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ		ข้อสอบถามจากผู้ใช้	ข้อสอบถามจากการขยายข้อสอบถาม
N		16	16
Means		0.8340	0.9175
Std. Deviation		0.1060	0.0517
Minimum		0.5389	0.7672
Maximim		0.9399	0.9540
Percentage	25 th	0.7968	0.9199
	50 th	0.8499	0.9332
	75 th	0.9144	0.9459

ตารางที่ 4.12 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่ยีสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน

	ค่าความแม่นยำที่ยีสิบ
Wilcoxon Sign Rank Test	120.00
Z	-3.103
Asymp. Sig (2-tailed)	0.002

4.2.3. การวิเคราะห์ค่าความแม่นยำที่สามสิบ

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะต้องตรวจสอบว่า ค่าความแม่นยำที่สามสิบที่ได้จากรายการผลลัพธ์การค้นคืนนั้น แจกแจงแบบปกติหรือไม่ จึงตั้งสมมติฐานดังนี้

1) ทดสอบการแจกแจงค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

H_0 : ข้อมูลค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ไม่แจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

H_0 : ข้อมูลค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม แจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่แจกแจงแบบปกติ

ผู้วิจัยได้นำแบบทดสอบซapiro-วิลค์ ในการตรวจสอบการแจกแจงของค่าความแม่นยำที่สามสิบ ผู้วิจัยจะยอมรับสมมติฐาน H_0 ถ้าค่านัยสำคัญ (Significance, Sig) มีค่ามากกว่า 0.05 ตารางที่ 4.13 แสดงการทดสอบการนำค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และจากการขยายข้อสอบถามมาทดสอบการแจกแจงแบบปกติ จะแสดงให้เห็นว่า

ตารางที่ 4.13 แสดงค่าสถิติทดสอบการแจกแจงของค่าความแม่นยำที่สามสิบ

ค่าตัวแปร	การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ	Shapiro-Wilk		
		Statistic	Df	Sig.
ค่าความแม่นยำที่สามสิบ	ข้อสอบถามจากผู้ใช้	0.866	14	0.037
	ข้อสอบถามจากการขยายข้อสอบถาม	0.622	14	0.000

1) การค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ มีค่านัยสำคัญเท่ากับ 0.037 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มีค่านัยสำคัญเท่ากับ 0.000 ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้น จึงสรุปได้ว่า ค่าความแม่นยำที่สามสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่สามสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่ได้แจกแจงแบบปกติ ผู้วิจัยจึงเลือกใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non-Parameter Test) ด้วยการเลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test)

การวิเคราะห์เปรียบเทียบค่าความแม่นยำที่สามสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และข้อสอบถามจากการขยายข้อสอบถาม

จากการวิเคราะห์การแจกแจงของค่าความแม่นยำที่สามสิบข้างต้น ผู้วิจัยจึงทดสอบสมมติฐาน โดยที่เลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกชันด้วยโปรแกรม IBM SPSS Statistic 20

กำหนดให้ $P@30_1$ แทนค่าความแม่นยำที่สามสิบที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ หรือรายการผลลัพธ์การค้นคืนก่อนการขยายข้อสอบถาม และ $P@30_2$ แทนค่าความแม่นยำที่สามสิบที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม หรือรายการผลลัพธ์การค้นคืนหลังการขยายข้อสอบถาม จากบทที่ 3 ผู้วิจัยได้กำหนดสมมติฐานการเปรียบเทียบค่าความแม่นยำที่สามสิบได้ดังนี้

H_0 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่สามสิบ ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_0: P@30_2 \leq P@30_1$$

H_1 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่าความแม่นยำที่สามสิบ มากกว่าผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_1: P@30_2 > P@30_1$$

ผู้วิจัยได้นำค่าความแม่นยำที่สิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่าความแม่นยำที่สามสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามมาทดสอบด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test) จะได้ผลลัพธ์การวิเคราะห์ทางสถิติซึ่งจะได้ผลลัพธ์ดังตารางที่ 4.14 และตารางที่ 4.15 จากผลการวิเคราะห์ค่าความแม่นยำที่สามสิบทางสถิติ คำนัยสำคัญที่ได้มีค่าเท่ากับ

$\frac{0.001}{2} = 0.0005$ ซึ่งมีค่าน้อยกว่าระดับค่านัยสำคัญ 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0 จึงสามารถสรุปได้ว่า ค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการ

ขยายข้อสอบถามนั้นมีค่ามากกว่าค่าความแม่นยำที่สามสิบของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

ตารางที่ 4.14 แสดงผลการวิเคราะห์ค่าความแม่นยำที่สามสิบที่ได้จากการวิเคราะห์

การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ		ข้อสอบถามจากผู้ใช้	ข้อสอบถามจากการขยายข้อสอบถาม
N		14	14
Means		0.8200	0.9169
Std. Deviation		0.0850	0.0452
Minimum		0.6222	0.7872
Maximim		0.9114	0.9490
Percentage	25 th	0.7710	0.9233
	50 th	0.8405	0.9299
	75 th	0.8771	0.9393

ตารางที่ 4.15 ผลการวิเคราะห์การเปรียบเทียบค่าความแม่นยำที่สามสิบที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน

	ค่าความแม่นยำที่สามสิบ
Wilcoxon Sign Rank Test	105
Z	-3.296
Asymp. Sig (2-tailed)	0.001

4.2.4. การวิเคราะห์ค่ามาตรฐานวัดเอฟ

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะต้องตรวจสอบว่า ค่ามาตรฐานวัดเอฟที่ได้จากรายการผลลัพธ์การค้นคือนั้น แจกแจงแบบปกติหรือไม่ จึงตั้งสมมติฐานดังนี้

1) ทดสอบการแจกแจงค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากผู้ใช้

H_0 : ข้อมูลค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากผู้ใช้ แจกแจงแบบปกติ

H_1 : ข้อมูลค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ไม่แจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม

H_0 : ข้อมูลค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม แจกแจงแบบปกติ

H_1 : ข้อมูลค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่แจกแจงแบบปกติ

ผู้วิจัยใช้แบบทดสอบซapiro-วิลค์ ในการทดสอบการแจกแจงของค่ามาตรฐานวัดเอฟ ผู้วิจัยจะยอมรับสมมติฐาน H_0 ถ้าค่านัยสำคัญ (Significance, Sig) มีค่ามากกว่า 0.05 ตารางที่ 4.16 แสดงการทดสอบการนำค่ามาตรฐานวัดเอฟของการค้นคือนิตสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และจากการขยายข้อสอบถามมาทดสอบการแจกแจงแบบปกติ จะแสดงให้เห็นว่า

ตารางที่ 4.16 แสดงค่าสถิติทดสอบการแจกแจงของค่ามาตรฐานวัดเอฟ

ค่าตัวแปร	การนำข้อสอบถามมาใช้ในการค้นคือนิตสารสนเทศ	Shapiro-Wilk		
		Statistic	Df	Sig.

ค่ามาตรฐานวัด เอฟ	ข้อสอบถามจากผู้ใช้	0.920	25	0.053
	ข้อสอบถามจากการขยายข้อ สอบถาม	0.361	25	0.000

1) การค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ มีค่านัยสำคัญเท่ากับ 0.053ซึ่งมีค่ามากกว่า 0.05 ดังนั้นจึงยอมรับสมมติฐาน H_0

2) การค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มีค่านัยสำคัญเท่ากับ 0.000ซึ่งมีค่าน้อยกว่า 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้น จึงสรุปได้ว่า ค่ามาตรฐานวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ แยกแยะแบบปกติ ส่วนค่ามาตรฐานวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม ไม่ได้แยกแยะแบบปกติ ผู้วิจัยจึงเลือกใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non-Parameter Test) ด้วยการเลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test)

การวิเคราะห์เปรียบเทียบค่ามาตรฐานวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และข้อสอบถามจากการขยายข้อสอบถาม

จากการวิเคราะห์การแจกแจงของค่ามาตรฐานวัดข้างต้น ผู้วิจัยจึงทดสอบสมมติฐาน โดยที่เลือกใช้แบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอนด้วยโปรแกรม IBM SPSS Statistic 20

กำหนดให้ F_1 แทนค่ามาตรฐานวัดเอฟที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ หรือรายการผลลัพธ์การค้นคืนก่อนการขยายข้อสอบถาม และ F_2 แทนค่ามาตรฐานวัดเอฟที่คำนวณได้จากรายการผลลัพธ์การค้นคืน ซึ่งเป็นผลลัพธ์ของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม หรือรายการผลลัพธ์การค้นคืนหลังการขยายข้อสอบถาม จากบทที่ 3 ผู้วิจัยได้กำหนดสมมติฐานการเปรียบเทียบค่ามาตรฐานวัดเอฟได้ดังนี้

H_0 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่ามาตรฐานวัดเอฟ ต่ำกว่าหรือเท่ากับผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_0: F_2 \leq F_1$$

H_1 : ผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามให้ค่ามาตรฐานวัดเอฟ มากกว่าผลลัพธ์การค้นคืนสารสนเทศเมื่อใช้ข้อสอบถามจากผู้ใช้

$$H_1: F_2 > F_1$$

ผู้วิจัยได้นำค่ามาตรฐานวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ และค่ามาตรฐานวัดเอฟที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามมาทดสอบด้วยแบบทดสอบการจัดลำดับเครื่องหมายวิลคอกซอน (Wilcoxon Sign Rank Test) จะได้ผลลัพธ์การวิเคราะห์ทางสถิติซึ่งจะได้ผลลัพธ์ดังตารางที่ 4.17 และตารางที่ 4.18 จากผลการวิเคราะห์ค่าความแม่นยำที่สามสิบทางสถิติ คำนัยสำคัญที่ได้มีค่าเท่ากับ $\frac{0.000}{2} = 0.000$ ซึ่งมีค่าน้อยกว่าระดับค่านัยสำคัญ 0.05 ดังนั้นจึงปฏิเสธสมมติฐาน H_0 จึงสามารถสรุปได้ว่า ค่ามาตรฐานวัดเอฟของการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถามนั้นมีค่ามากกว่าค่ามาตรฐานวัดเอฟของการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้

ตารางที่ 4.17 แสดงผลการวิเคราะห์ค่ามาตรฐานวัดเอฟที่ได้จากการวิเคราะห์

การนำข้อสอบถามมาใช้ในการค้นคืนสารสนเทศ	ข้อสอบถามจากผู้ใช้	ข้อสอบถามจากการขยายข้อสอบถาม
N	25	25
Means	0.2915	0.4485
Std. Deviation	0.1376	0.0901
Minimum	0.0288	0.0288

Maximim		0.4738	0.4813
Percentage	25 th	0.1630	0.4513
	50 th	0.3164	0.4767
	75 th	0.4124	0.4780

ตารางที่ 4.18 ผลการวิเคราะห์การเปรียบเทียบค่ามาตรฐานที่ได้จากการค้นคืนสารสนเทศด้วย
 ข้อสอบถามจากผู้ใช้และข้อสอบถามจากการขยายข้อสอบถาม ด้วยแบบทดสอบการจัดลำดับ
 เครื่องหมายวิลคอกซอน

	ค่ามาตรฐาน
Wilcoxon Sign Rank Test	300.00
Z	-4.286
Asymp. Sig (2-tailed)	0.000

4.3. สรุปผลการวิเคราะห์ข้อมูล

จากผลการวิเคราะห์ข้อมูลค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรฐาน จะเห็นว่ารายการผลลัพธ์การค้นคืนสารสนเทศที่ได้จากการค้นคืน
 ด้วยข้อสอบถามจากการขยายข้อสอบถามนั้น จะให้ค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยี่สิบ
 ค่าความแม่นยำที่สามสิบ และค่ามาตรฐาน ที่สูงกว่ารายการผลลัพธ์การค้นคืนสารสนเทศที่ได้
 จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ตารางที่ 4.19 แสดงผลการวิเคราะห์ค่าความ
 แม่นยำที่สืบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรฐาน ผู้วิจัยจึงสรุป
 ได้ว่า ประสิทธิภาพการค้นคืนสารสนเทศจะเพิ่มขึ้น เมื่อนำการขยายข้อสอบถามมาประยุกต์ใช้ใน
 การค้นคืนสารสนเทศภาษาไทย

ตารางที่ 4.19 สรุปผลการวิเคราะห์ค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยั่งยืน ค่าความแม่นยำที่สามสืบ และค่ามาตรวัดเอฟ

	P@10		P@20		P@30		F	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Means	0.8834	0.9328	0.8340	0.9175	0.8200	0.9169	0.2915	0.4485
Std. Deviation	0.0689	0.0319	0.1060	0.0517	0.0850	0.0452	0.1376	0.0901
Minimum	0.6456	0.8418	0.5389	0.7672	0.6222	0.7872	0.0288	0.0288
Maximim	0.9708	0.9716	0.9399	0.9540	0.9114	0.9490	0.4738	0.4813
Z	-2.978		-3.103		-3.296		-4.286	
Asymp. Sig (2-tailed)	0.003		0.002		0.001		0.000	

ผลการวิเคราะห์และสรุปการศึกษาของงานวิจัยนี้ได้สอดคล้องกับงานวิจัยในอดีตที่ศึกษาการนำการขยายข้อสอบถามด้วยขั้นตอนวิธีการจัดกลุ่มมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศ Kalmanovich และ Kurland (2009) ได้นำเสนอการการจัดกลุ่มมาประยุกต์ใช้กับขั้นตอนการขยายข้อสอบถาม พบว่า ประสิทธิภาพการค้นคืนเพิ่มขึ้น เมื่อนำขั้นตอนการจัดกลุ่มมาประยุกต์ร่วมกับระบบการขยายข้อสอบถาม Bernhard (2010) ได้นำเสนอการจัดกลุ่มคำเพื่อนำมาใช้ในการขยายข้อสอบถาม พบว่าขั้นตอนการการจัดกลุ่มคำสำหรับการขยายข้อสอบถามดังกล่าวนั้นสามารถเพิ่มประสิทธิภาพการค้นคืนสารสนเทศได้เช่นกัน เมื่อเทียบกับประสิทธิภาพการค้นคืน เมื่อไม่ได้นำการขยายข้อสอบถามมาประยุกต์ใช้ หรือเมื่อนำการขยายข้อสอบถามด้วยขั้นตอนวิธีการอื่นๆ มาประยุกต์ใช้

4.4. สรุปการอภิปรายผล

จากตารางที่ 4.6 จะเห็นได้ว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศนั้น ช่วยเพิ่มประสิทธิภาพการค้นคืนเพิ่มขึ้น จากตารางที่ 4.6 จะเห็นว่า ถึงแม้ระบบจะค้นคืนสารสนเทศทั้งหมดในคอร์ปัสเอกสารทั้ง 100 รายการ ไม่ว่าจะใช้ข้อสอบถามจากผู้ใช้หรือ

ข้อสอบถามจากการขยายข้อสอบถาม แต่จากค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยี่สิบ และค่าความแม่นยำที่สามสิบนั้น ที่ประเมินจากรายการผลลัพธ์การค้นคืนของการใช้ข้อสอบถามจากการขยายข้อสอบถามในระบบการค้นคืนสารสนเทศ เพิ่มขึ้น



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 5 สรุปผลการวิจัย

ภายหลังจากผู้วิจัยได้เก็บรวบรวมข้อมูลการทดลองและสรุปผลการทดลองแล้ว เราจึงสามารถสรุปผลการวิจัย รวมทั้งเสนอการนำงานวิจัยมาประยุกต์ใช้ในเชิงทฤษฎีและปฏิบัติ ข้อจำกัดของงานวิจัยที่เกิดขึ้น และแนวทางการศึกษาต่อได้

5.1. สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอการนำการจัดกลุ่มการปรากฏร่วมกันของคำด้วยการจัดกลุ่มแบบลำดับขั้น สำหรับการขยายข้อสอบถามในการค้นคืนสารสนเทศภาษาไทย ผู้วิจัยได้กำหนด (1) รูปแบบการค้นคืนสารสนเทศที่ใช้ในการวิจัย คือ รูปแบบปริภูมิเวกเตอร์ (Vector Space Model) (Khankasikam, 2010; Suwanapong, Theeramunkong และ Nantajeewarawat, 2010; Klabbankoh และ Pinthong, 2000) (2) รูปแบบการขยายข้อสอบถามเป็นการขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion) (Kanaan และคณะ, 2007 และ 2008) (3) การนำรายการเอกสารในคอร์ปัสเอกสาร (Document Corpus) และข้อสอบถามมาใช้ในการวิเคราะห์การกำหนดกลุ่มคำในการขยายข้อสอบถาม (Klink และคณะ, 2002; Rahman, Antani และ Thoma, 2011) และ (4) ขั้นตอนวิธีการทำเหมืองข้อมูลที่นำมาใช้ในการวิเคราะห์และกำหนดคำสำหรับการขยายข้อสอบถามคือ การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) (Aono และ Doi, 2005; Dvorsky, Martinovic และ Snasel, 2004; Xu และ Hu, 2010)

ทั้งนี้ผู้วิจัยจะนำรายการผลลัพธ์การค้นคืน (Retrieval Result List) ที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากผู้ใช้ ซึ่งได้ก่อนการขยายข้อสอบถาม และรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนสารสนเทศด้วยข้อสอบถามจากการขยายข้อสอบถาม มาวิเคราะห์เปรียบเทียบประสิทธิภาพการค้นคืนสารสนเทศ

ในการประเมินประสิทธิภาพการค้นคืนนั้น ผู้วิจัยใช้ค่าความแม่นยำที่สิบ (Precision at 10 documents) ค่าความแม่นยำที่สี่สิบ (Precision at 20 documents) ค่าความแม่นยำที่สามสิบ (Precision at 30 documents) และค่ามาตรวัดเอฟ (F-measure) โดยค่าความแม่นยำที่สิบ ค่า

ความแม่นยำที่ยี่สิบและค่าความแม่นยำที่สามสิบ ใช้วิเคราะห์ประสิทธิภาพการค้นคืนส่วนต้นของรายการผลลัพธ์การค้นคืนจำนวนสิบ ยี่สิบ และสามสิบรายการ ส่วนค่ามาตรฐานวัดเอฟนั้น จะนำมาใช้ในการวิเคราะห์ประสิทธิภาพการค้นคืนจากรายการผลลัพธ์การค้นคืนทั้งหมด

ในการสร้างชุดทดสอบ ผู้วิจัยใช้หนังสือที่กำลังศึกษาชั้นปีที่ 4 คณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จำนวน 25 คนกำหนดข้อสอบถามและศึกษาบทคัดย่อโครงงานปี 4 ที่จัดทำโดยนิสิตคณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ ในอดีตจำนวน 100 ฉบับ เป็นเอกสาร และกำหนดความเกี่ยวข้องระหว่างข้อสอบถามและบทคัดย่อ โดยที่ข้อสอบถามที่นิสิตได้กำหนดมีความยาว 5-8 คำ และบทคัดย่อที่ใช้มีมีความยาวไม่เกินหนึ่งหน้ากระดาษขนาดเอสี่ หลังจากที่ได้ผลการกำหนดความเกี่ยวข้องระหว่างคำในข้อสอบถามและบทคัดย่อ ผู้วิจัยจึงพัฒนาระบบการขยายข้อสอบถามด้วยกลุ่มคำด้วยพจนานุกรมอิเล็กทรอนิกส์ ที่จัดทำโดยศูนย์อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronic and Computer Technology Center : NECTEC) โปรแกรมตัดคำภาษาไทยไทยอานาไลเซอร์ (ThaiAnalyzer) และรายการคำหยุด (Stop-word list) จากวรรณวีร์ เกษร (2553)

ในส่วนการพัฒนากระบวนการขยายข้อสอบถามนั้น ผู้วิจัยได้กำหนดกระบวนการไว้ดังนี้ (1) วิเคราะห์และกำหนดเมทริกซ์แสดงคำในเอกสาร (Term-Document Matrix) (ศุภชัย ตั้งวงศ์ศานต์, 2551; Weiss และคณะ, 2005) (2) วิเคราะห์และกำหนดเมทริกซ์แสดงการปรากฏร่วมกันของคำ (Term Co-occurrence Matrix) (Razmara และ Kosseim, 2007) (3) วิเคราะห์กลุ่มคำ (Term Clustering) ด้วยการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) จึงจะได้เซตของกลุ่มคำ (Set of Clusters) และ (4) วิเคราะห์และกำหนดกลุ่มคำจากเซตกลุ่มคำ เพื่อนำมาใช้เพิ่มคำในข้อสอบถามจากผู้ใช้ ซึ่งได้จะข้อสอบถามใหม่ที่ได้จากการขยายข้อสอบถาม

อย่างไรก็ตาม ในส่วนการจัดกลุ่มแบบลำดับขั้นนั้น ผู้วิจัยได้เลือกใช้เครื่องมือการทำเหมืองข้อมูล เวอร์ก้า 3.7 (Weka 3.7) มาใช้เป็นเครื่องมือการจัดกลุ่มคำสำหรับการขยายข้อสอบถาม ผู้วิจัยจะต้องกำหนดจำนวนกลุ่ม (Number of Clusters) สมการค่าความเหมือน (Similarity Measure) และเกณฑ์การเชื่อมโยง (Linkage Criteria) การกำหนดรูปแบบการจัดกลุ่มมีผลต่อคุณภาพการจัดกลุ่ม (Clustering Quality) ของเซตของกลุ่มที่ได้จากการจัดกลุ่ม ผู้วิจัยจึงศึกษารูปแบบการจัดกลุ่มแบบลำดับขั้นเพื่อกำหนดจำนวนกลุ่ม สมการค่าความเหมือน และสมการเกณฑ์การเชื่อมโยงสำหรับการนำมาใช้ในการขยายข้อสอบถามนี้

จากการนำรายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามจากผู้ใช้จำนวน 25 ชุด มาวิเคราะห์การจัดกลุ่มด้วยรูปแบบการจัดกลุ่มทั้งหมด เพื่อวิเคราะห์ เปรียบเทียบ และคัดเลือกคุณภาพการจัดกลุ่ม สำหรับการนำมาใช้ในงานวิจัยนี้ ค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกำลังสอง (Root Mean Square Standard Deviation) ระยะทางที่ไกลที่สุดจากกลุ่ม (Maximum Distance from Cluster) และระยะทางไปยังกลุ่มที่ใกล้ที่สุด (Distance to Nearest Cluster) ถูกเลือกมาใช้ในการวิเคราะห์คุณภาพการจัดกลุ่มที่มีรูปแบบการจัดกลุ่มที่แตกต่างกัน (สุนันทา เปี่ยมพริ้ง, 2549; Jacob, 2007)

ในงานวิจัยนี้ กำหนดให้จำนวนกลุ่มมีค่าเท่ากับ 30 กลุ่ม สมการค่าความเหมือนยูคลิดเดียน และสมการเกณฑ์การเชื่อมโยงเชิงซับซ้อนนั้น ให้คุณภาพการจัดกลุ่มที่ดีที่สุด ผู้วิจัยจึงใช้รูปแบบการจัดกลุ่มที่นำมาใช้ในงานวิจัยนี้

ผู้วิจัยได้พัฒนาระบบการค้นคืนสารสนเทศ และระบบการขยายข้อสอบถามด้วยการจัดกลุ่มคำ ผู้วิจัยจึงได้เลือกใช้อะพาคเฮลูซีน 3.5 (Apache Lucene 3.5) เป็นเครื่องมือหลักของระบบการค้นคืนสารสนเทศ

หลังจากนั้น ผู้วิจัยจึงนำข้อสอบถามที่ได้จากการขยายข้อสอบถามทั้งหมดมาค้นคืนสารสนเทศ และนำรายการผลลัพธ์การค้นคืนที่ได้มารวมประเมินประสิทธิผลการค้นคืนเปรียบเทียบกับรายการผลลัพธ์การค้นคืนที่ได้มาจากการค้นคืนด้วยข้อสอบถามจากผู้ใช้

จากการเปรียบเทียบรายการผลลัพธ์การค้นคืนที่ได้ ผู้วิจัยพบว่ารายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามจากการขยายข้อสอบถามนั้นให้ค่าความแม่นยำที่สิบ ค่าความแม่นยำที่ยี่สิบ ค่าความแม่นยำที่สามสิบ และค่ามาตรวัดเอฟทีสูงกว่ารายการผลลัพธ์การค้นคืนที่ได้จากการค้นคืนด้วยข้อสอบถามจากผู้ใช้อย่างมีนัยสำคัญ กล่าวคือ ประสิทธิภาพการค้นคืนภาษาไทยจะเพิ่มขึ้น เมื่อนำการขยายข้อสอบถามด้วยการจัดกลุ่มคำมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศ

5.2. การนำงานวิจัยมาประยุกต์ใช้

5.2.1. การนำงานวิจัยมาประยุกต์ใช้ในเชิงทฤษฎี

งานวิจัยนี้ได้นำเสนอการนำการขยายข้อสอบถามด้วยการจัดกลุ่มมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศภาษาไทย ทั้งนี้ผู้วิจัยได้ประเมินประสิทธิภาพการค้นคืนด้วยค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ยีสืบ ค่าความแม่นยำที่สามสืบ และค่ามาตรวัดเอฟ หลังจากการศึกษา ผู้วิจัยพบว่า ประสิทธิภาพการค้นคืนสารสนเทศ เมื่อใช้ข้อสอบถามจากการขยายข้อสอบถามนั้น จะมากกว่าประสิทธิภาพการค้นคืนสารสนเทศ เมื่อใช้ข้อสอบถามจากผู้ใช้ กล่าวคือ ประสิทธิภาพการค้นคืนสารสนเทศได้เพิ่มขึ้น เมื่อนำการขยายข้อสอบถามดังกล่าวมาประยุกต์ใช้ ซึ่งสอดคล้องกับ Kalmanovich และ Kurland (2009) และ Bernhard (2010) ที่ว่า การนำการขยายข้อสอบถามมาประยุกต์ใช้ ทำให้ประสิทธิภาพการค้นคืนสารสนเทศจะเพิ่มขึ้น นอกจากนี้ยังไม่พบงานวิจัยในอดีต ที่นำเสนอการนำการขยายข้อสอบถามด้วยขั้นตอนการจัดกลุ่มคำมาประยุกต์ใช้กับระบบการค้นคืนสารสนเทศภาษาไทย ดังนั้นข้อค้นพบในงานวิจัยนี้ จึงเป็นพื้นฐานขององค์ความรู้ในการออกแบบขั้นตอนการขยายข้อสอบถาม เพื่อนำประยุกต์กับระบบการค้นคืนสารสนเทศ เพื่อเพิ่มประสิทธิภาพการค้นคืน

5.2.2. การนำงานวิจัยมาประยุกต์ใช้ในเชิงปฏิบัติ

1) **การจัดกลุ่มคำภาษาไทยร่วมกับการขยายข้อสอบถาม** มีงานวิจัยที่นำเสนอการจัดกลุ่มคำภาษาไทย หรือการจำแนกประเภทคำไทยด้วยขั้นตอนวิธีการทำเหมืองข้อมูล แต่ไม่พบการนำการจัดกลุ่มคำมาประยุกต์ใช้ร่วมกับการขยายข้อสอบถาม จากงานวิจัยนี้ จะเห็นได้ว่าการจัดกลุ่มคำภาษาไทยร่วมกับการขยายข้อสอบถาม จะสามารถช่วยเพิ่มประสิทธิภาพการค้นคืนภาษาไทยได้

2) **การขยายข้อสอบถามแบบอัตโนมัติ (Automatic Query Expansion)**
งานวิจัยนี้นำเสนอการนำการขยายข้อสอบถามแบบอัตโนมัติมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศภาษาไทย เพื่อลดการตัดสินใจของผู้ใช้ ซึ่งจะมีผลกระทบต่อประสิทธิภาพการค้นคืนสารสนเทศ การขยายข้อสอบถามแบบอัตโนมัติจะขยายข้อสอบถามจากผู้ใช้ และนำข้อสอบถามใหม่ซึ่งขยายแล้วมาใช้ในระบบการค้นคืนสารสนเทศ และนำเสนอรายการผลลัพธ์การค้นคืนต่อ

ผู้ใช้ โดยไม่มีการตัดสินใจของผู้ใช้ในการขยายข้อสอบถามนอกจากนี้แล้ว ยังช่วยลดปัญหาในการกำหนดข้อสอบถามสำหรับการค้นคืนที่จะต้องใช้คำที่ผู้ใช้ไม่ทราบ ไม่สามารถสะกดได้ หรือไม่มีองค์ความรู้ในเอกสารหรือสารสนเทศที่ตนเองต้องการ ผู้พัฒนาระบบการค้นคืนสารสนเทศที่มีองค์ความรู้เฉพาะด้าน อาทิ สารสนเทศทางการแพทย์ สามารถนำประยุกต์ใช้กับระบบสารสนเทศนั้นได้ดี

3) การนำการขยายข้อสอบถามร่วมกับการค้นคืนสารสนเทศภาษาไทย

เนื่องจากการขยายขอบเขตของสารสนเทศภาษาไทย การค้นคืนสารสนเทศภาษาไทยจึงเริ่มมีบทบาทมากยิ่งขึ้น อย่างไรก็ตามการกำหนดข้อสอบถามภาษาไทยสำหรับการค้นคืนนั้น จึงเป็นปัญหาในการใช้ระบบการค้นคืนสารสนเทศภาษาไทย การขยายข้อสอบถามจะช่วยกำหนดคำเพิ่มเติมในข้อสอบถาม เพื่อนำไปใช้ในการค้นคืนสารสนเทศ กล่าวคือ การขยายข้อสอบถามจะช่วยให้การค้นคืนสารสนเทศนั้น มีประสิทธิภาพเพิ่มขึ้น ในปัจจุบัน องค์กรหรือหน่วยงานต่างๆ ในประเทศไทยเริ่มนำการค้นคืนสารสนเทศมาใช้ในระบบเครือข่ายอินเทอร์เน็ตหรืออินเทอร์เน็ตขององค์กร การนำการขยายข้อสอบถามมาร่วมกับระบบการค้นคืนนี้ จะสามารถช่วยเพิ่มประสิทธิภาพการค้นคืนสารสนเทศได้

5.3. ข้อจำกัดของงานวิจัย

1) **จำนวนรายการเอกสารและข้อสอบถาม** เนื่องจากในชุดทดสอบ จะมีรายการเอกสารและข้อสอบถามจำนวนมาก เพื่อนำมาใช้ในการประเมินระบบการค้นคืนสารสนเทศที่ผู้วิจัยในอดีตได้พัฒนาหรือวิเคราะห์ แต่ในงานวิจัยนี้มีรายการเอกสาร 100 ฉบับ และรายการข้อสอบถามจำนวน 25 ข้อสอบถาม รายการเอกสารและข้อสอบถามที่นำมาใช้ในงานวิจัยนี้ จึงถือได้ว่ามีจำนวนน้อย เมื่อเปรียบเทียบกับจำนวนรายการข้อสอบถามและเอกสารที่ผู้วิจัยในอดีตนำมาใช้ในการทดสอบแต่ในปัจจุบันยังไม่มีหน่วยงานใดที่พัฒนาชุดทดสอบภาษาไทยขึ้นมา

2) **เครื่องมือการพัฒนาระบบการค้นคืนสารสนเทศ การขยายข้อสอบถาม และการประเมินประสิทธิภาพการค้นคืนสารสนเทศ** เครื่องมือที่นำมาใช้ในงานวิจัยนี้เป็นโอเพนซอร์ซ (Open-Sourced) ซึ่งเป็นเครื่องมือที่ได้รับอนุญาตให้ผู้พัฒนาทั่วไปสามารถนำมาพัฒนาต่อเองได้ ตามวัตถุประสงค์ในการใช้งานและไม่เสียค่าใช้จ่าย อย่างไรก็ตาม เครื่องมือเหล่านี้อาจจะขาดความน่าเชื่อถือในการทดสอบระบบการค้นคืนสารสนเทศ นอกจากนี้ยังมี

เครื่องมืออื่นๆ ที่สามารถนำมาใช้ในการพัฒนาระบบดังกล่าว อาทิ เครื่องมือการพัฒนาระบบการค้นคืนสารสนเทศลีเมอ (Lemur) (Song และคณะ, 2005; Yilmaz, Arslan และ Yilmazel, 2009; Chen, Li และ Li, 2005) พจนานุกรมอิเล็กทรอนิกส์ไทยทูอิงลิช (Thai2English) หรือระบบการวิเคราะห์ทางสถิติ (Statistical Analysis System, SAS)

3) **ลักษณะของงานวิจัย** งานวิจัยนี้เป็นงานวิจัยเชิงทดลอง ผู้วิจัยจะต้องควบคุมปัจจัยต่างๆ ที่มีผลต่อการพัฒนาระบบการค้นคืนสารสนเทศ ระบบการขยายข้อสอบถาม และการประเมินประสิทธิภาพการค้นคืนสารสนเทศโดยเฉพาะอย่างยิ่งการกำหนดรูปแบบการจัดกลุ่มแบบลำดับขั้นสำหรับการจัดกลุ่มคำ ดังนั้นผลลัพธ์การทดลองที่ได้จึงไม่สามารถนำไปสรุปเป็นแนวทางหลักในการจัดกลุ่มคำสำหรับการขยายข้อสอบถาม

4) **การกำหนดค่าที่นำมาใช้ในการประเมิน** งานวิจัยนี้ได้เลือกใช้ค่าความแม่นยำที่สืบ ค่าความแม่นยำที่ย่สืบและค่าความแม่นยำที่สามสืบ มาใช้ในการวิเคราะห์ประสิทธิภาพการค้นคืนสารสนเทศ อย่างไรก็ตาม หากจำนวนเอกสารที่อยู่ในรายการผลลัพธ์การค้นคืนนั้นมีจำนวนน้อยกว่าจำนวนที่กำหนดไว้ ผู้วิจัยจะไม่สามารถนำรายการผลลัพธ์การค้นคืนนั้นมาใช้ในการวิเคราะห์ข้อมูล

5) **ขนาดของกลุ่มคำที่นำมาใช้ในการขยายข้อสอบถาม** เนื่องจากในขั้นตอนการกำหนดกลุ่มคำสำหรับการขยายข้อสอบถาม ผู้วิจัยไม่ได้คำนึงถึงขนาดของกลุ่มคำที่นำมาใช้ จึงทำให้กลุ่มคำขนาดใหญ่ถูกนำมาใช้ในการขยายข้อสอบถาม จึงทำให้ขนาดของข้อสอบถามใหม่ซึ่งเป็นผลลัพธ์ของการขยายข้อสอบถามนั้น มีขนาดใหญ่มาก เมื่อเทียบกับขนาดของข้อสอบถามจากผู้ใช้ จึงทำให้รายการผลลัพธ์การค้นคืนของการใช้ข้อสอบถามจากการขยายข้อสอบถามมีจำนวนมากกว่ารายการผลลัพธ์การค้นคืนของการใช้ข้อสอบถามจากผู้ใช้

5.4. แนวทางศึกษาต่อ

1) **การค้นคืนสารสนเทศภาษาไทย** เริ่มมีงานวิจัยการนำการขยายข้อสอบถามมาประยุกต์ร่วมกับการค้นคืนสารสนเทศภาษาไทยมากยิ่งขึ้น การพัฒนาการขยายข้อสอบถามเพื่อนำมาประยุกต์ร่วมกับระบบการค้นคืนสารสนเทศภาษาไทย บนชุดทดสอบที่มีจำนวนและเนื้อหาหลากหลายต่อไป

2) **การนำการขยายข้อสอบถามมาประยุกต์ร่วมกับการค้นคืนสารสนเทศภาษาไทย** งานวิจัยนี้มุ่งประเด็นการนำการขยายข้อสอบถามด้วยการจัดกลุ่มคำถามมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศภาษาไทย เนื่องจากงานวิจัยในอดีต ถึงแม้ว่าจะมีการนำการขยายข้อสอบถามด้วยวิธีการทำเหมืองข้อมูลอื่นๆ มาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศภาษาไทย แต่ยังไม่มียานวิจัยที่นำการขยายข้อสอบถามด้วยการจัดกลุ่มคำถามมาประยุกต์ใช้ร่วมกับการค้นคืนสารสนเทศภาษาไทย เช่นเดียวกับงานวิจัยในอดีตที่นำมาประยุกต์ร่วมกับการค้นคืนสารสนเทศภาษาอื่นๆ อาทิ ภาษาอังกฤษ ภาษาจีน หรือภาษาอาหรับ ดังนั้นการศึกษานำการขยายข้อสอบถามด้วยการจัดกลุ่มคำถามมาประยุกต์ร่วมกับการค้นคืนสารสนเทศภาษาไทยนั้น จึงเป็นแนวทางในการศึกษาต่อไปได้

3) **การสร้างชุดทดสอบภาษาไทย** เนื่องจากการขยายขอบเขตของสารสนเทศภาษาไทย จึงมีการพัฒนาการค้นคืนสารสนเทศภาษาไทยเพิ่มขึ้น ชุดทดสอบภาษาไทยจึงนำมาใช้ในการประเมินประสิทธิภาพการค้นคืนสารสนเทศมากขึ้น แต่ไม่พบหน่วยงานหรือองค์กรใดที่พัฒนาชุดทดสอบภาษาไทยนี้ สำหรับการวิจัยการค้นคืนสารสนเทศภาษาไทย ซึ่งรวมทั้งการนำการขยายข้อสอบถามมาประยุกต์ร่วมกับการค้นคืนสารสนเทศภาษาไทย ซึ่งจะขาดความน่าเชื่อถือในชุดทดสอบและการวิจัย ดังนั้นควรมีแนวทางในการพัฒนาการสร้างชุดทดสอบภาษาไทย

รายการอ้างอิง

ภาษาไทย

กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน) สำนักงานนโยบายและส่งเสริมธุรกรรมทางอิเล็กทรอนิกส์ ส่วนงานดัชนีและสำรวจ. รายงานผลการสำรวจพฤติกรรมผู้ใช้อินเทอร์เน็ตในประเทศไทย ปี 2556.
กรุงเทพฯ : สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน). 2556

ศูนย์วิจัยนวัตกรรมอินเทอร์เน็ตไทย (www.truehits.com)

กระทรวงเทคโนโลยีสารสนเทศและการสื่อสารสำนักงานสถิติแห่งชาติ, สำรวจการมีและการใช้เทคโนโลยีสารสนเทศและการสื่อสารในครัวเรือน พ.ศ.2555. กรุงเทพฯ, 2556.

กระทรวงวิทยาศาสตร์และเทคโนโลยี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. รายงานผลการสำรวจกลุ่มผู้ใช้อินเทอร์เน็ตในประเทศไทย ปี 2553.2553.

กัลยา วานิชย์บัญชา. สถิติสำหรับงานวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ, 2550.

กัลยา วานิชย์บัญชา. การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล. พิมพ์ครั้งที่ 18, กรุงเทพฯ : ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย, 2554

ไกรศักดิ์ เกษร. แนวคิดและความท้าทายในการพัฒนาการค้นคืนข้อมูลข้ามภาษาไทย-อังกฤษ.วารสารวิทยาศาสตร์ มข. 41(1), 2556

มนาชัย กริ่งไกร และชุลีรัตน์ จรัสกุลชัย. การจัดกลุ่มเอกสารข้อความภาษาไทยด้วยขั้นตอนวิธี Spherical K-Means แบบขนานบนพิกัดลึกลับคลัสเตอร์. The Fifth National Computer Science and Engineering Conference, 2001, 7 - 9 November

ชวลีรัตน์ จรัสกุลชัย, เจษฎา กันทะเสนา, และ สถาพร คิวสุวรรณสุข. การจัดกลุ่มเอกสารสำหรับ
ข้อความภาษาไทย. [ออนไลน์]แหล่งที่มา:std.kku.ac.th/4830203385/ncsec2001-
 cluster-s.doc

ต้นติมา เวฬุกันนท์ และ สมชาย ปราการเจริญ. ระบบค้นคืนเอกสารแบบเต็มรูปแบบ :
 กรณีศึกษาสำนักงานศาลปกครอง. ใน การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิต
แห่งชาติ ครั้งที่ 23. 23-24 ธันวาคม 2554 : 190-196

นพดล หมั่นไผ่. การเพิ่มประสิทธิภาพระบบค้นคืนสารสนเทศภาษาไทยด้วยนามวลีผันแปรและ
 ออนโทโลยี. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์).
 มหาวิทยาลัยเกษตรศาสตร์. 2548

สุนันทา เปี่ยมพริ้ง. การศึกษาเปรียบเทียบประสิทธิภาพของระบบค้นคืนสารสนเทศที่ใช้เทคนิคการ
 วัดความคล้ายคลึงเชิงมุมและเทคนิคการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดีเยนที่
 กำหนดกรอบค่าความคล้ายคลึงด้วยผลลัพธ์จากการจัดกลุ่มข้อมูล. วิทยานิพนธ์
 ปริญญาวิทยาศาสตรมหาบัณฑิต (การพัฒนาซอฟต์แวร์ด้านธุรกิจ). จุฬาลงกรณ์
 มหาวิทยาลัย. 2550

ประกายรัตน์ สุวรรณ และอมรวิทย์ วิเศษสงวน. การวิจัยและวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS
 เวอร์ชัน 20. กรุงเทพฯ : ซีเอ็ดยูเคชั่น, 2555.

พัชระ นาเสี่ยม, สิริภัทร เขียวชาญวัฒนา และคำรณ สุนิติ. การประสานคำตอบสำหรับการจัด
 กลุ่มผลการสืบค้นเว็บภาษาไทยที่มีความน่าเชื่อถือ. The nationl Conference on
 Computer Information Technology. 2553, 13-15 มกราคม : 51-59

รัตนา ชุมแก้ว และ สุพจน์ นิตยส์สุวรรณ. ระบบค้นคืนพระคาถาธรรมบทภาษาบาลีอักษรไทย. ใน
Proceeding of the 5th National Conference on Computer and Information
Technology (NCCIT2009). , 2552, 1166-1171.

วรวิทย์ เกษร. การวัดความคล้ายคลึงของเอกสารภาษาไทยโดยใช้การประมวลผลภาษาธรรมชาติ.

วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์), มหาวิทยาลัยเชียงใหม่ :
2553

วาริน นาราวิทย์ และศิริภิญโญ จันทมณี. การขยายคำค้นเพื่อเพิ่มประสิทธิภาพระบบการค้นคืน
สารสนเทศภาษาไทยโดยใช้เจเนติกอัลกอริทึม. ใน การประชุมวิชาการวไลยลักษณ์ ครั้งที่
4. นครศรีธรรมราช : มหาวิทยาลัยวลัยลักษณ์, 2554

ศุภชัย ตั้งวงศ์ศานต์. ระบบการจัดเก็บและการค้นคืนสารสนเทศด้วยคอมพิวเตอร์. กรุงเทพฯ : โรง
พิมพ์พิทักษ์การพิมพ์, 2551

สุทธิ ชัดตียะ และ วิไลลักษณ์ สุวจิตตานนท์. แบบแผนการวิจัยและสถิติ. พิมพ์ครั้งที่ 2, กรุงเทพฯ :
ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย, 2554

องอาจ นัยพัฒน์. การออกแบบการวิจัย : วิธีการเชิงปริมาณ เชิงคุณภาพ และผสมผสานวิธีการ.
พิมพ์ครั้งที่ 2, กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2554

ภาษาอังกฤษ

Search Engines : SERP Traffic Analysis. (2012). from
http://www.statowl.com/search_engine_results_page.php

Abbas, O. A. (2008). Comparisons Between Data Clustering Algorithms. The
International Arab Journal of Information Technology, 5(3).

Aboulnaga, Y., and Clarke, Charles L.A. (2012). Frequent Itemset Mining for Query
Expansion in Microblog Ad-hoc Search. Paper presented at the The Twenty-First
REtrieval Conference (TREC 2012) Proceedings.

Akrivas, G., Wallace, Manolis, Stamou, Giorge, and Kollias, Stefanos. (2002). Context-
Sensitive Query Expansion Based on Fuzzy Clustering of Index Terms. Flexible
Query Answering Systems, 2522.

- Alipanah, N., Khan, Latifur, and Thuraisingham, Bhavani. (2011). Optimized Ontology-Driven Query Expansion Using Map-Reduce Framework to Facilitate Federated Queries The University of Texas, Dallas.
- Amini, M. a. U., Nicolas. (2007, 26-27 Apr 2007). A Contextual Query Expansion Approach by Term Clustering for Robust Text Summarization. Paper presented at the 7th Document Understanding Conference (DUC'07), New York, USA.
- Amini, M. R., Usunier, Nicolas and Gallinari, Patrick. (2005). Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms. *Advances in Information Retrieval*, 3408.
- Aono, M. a. D., Hironori. (2005). A Method for Query Expansion Using a Hierarchy of Clustering. *Lecture Notes in Computer Science*, 3689.
- Araujo, L., and Perez-Iglesias, joaquin. (2010). Training a Classifier for the selecteion of Goods Query Expansion Terms with a Genetic Algorithm. 2010 IEEE Congress on Evolutionary Computation (CEC).
- Arcoverde, J. M. A., Nunes, Maria D.G.V., and Scardua, Wendel. (2006). Using Noun Phrases for Local Analysis in Automatic Query Expansion. CLEF working notes for ad-hoc, monolingue, Portuguese track.
- Arguello, J., Elsas, Jonathan L., Callan, Jamie, Carbonell, Jamie G. . (2008). Document Representation and Query Expansion Models for Blog Recommendation. Paper presented at the Proceedins of the second International Conference on Weblogs and Social Media (ICWSN).
- Asfari, O., Doan, Bich-Lien, Bourda, Yolaine, and Sansonnet, Jean-Paul. (2010). Context-based Hybrid Method for User Query Expansion. Paper presented at the SEMAPRO 2010 : The Fourth International Conference on Advances in Semantic Proceeding, Florence.
- Auffarth, B. (2010). Clustering by a genetic algorithm with biased mutation operation. Paper presented at the Evolutionary Computation (CEC), 2010 IEEE Congress on.

- Azzopardi, C. (2006). Contextual information and assessor characteristics in complex question answering. Paper presented at the The Fifteenth Text REtrieval Conference (TREC 2006) Proceedings.
- Bailey, P., Craswell, Nick and Soboroff, Ian. (2008). Relevance Assessment: Are Judges Exchangeable and Does it Matter? Paper presented at the Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- Balcan, M.-F., Blum, Avrim, and Vempala, Santosh. (2008). A Discriminative Framework for Clustering via Similarity Functions. Paper presented at the STOC'08 Proceedings of the fortieth annual ACM symposium on Theory of computing.
- Bandyopadhyay, A., Ghosh, Kripabandhu, Majumder, Prasenjit, and Mitra, Mandar. (2012). Query Expansion for microblog retrieval. International Journal of Web Science (IJWS), 1(4).
- Bank, M. a. S., Friedhelm. (2010, 2010, July 21-23). Fuzzification of Agglomerative Hierarchical Crisp Clustering Algorithms. Paper presented at the Challenges at the Interface of Data Analysis, Computer Science, and Optimization.
- Bast, H., Majudar, Debapriyo, and Weber, Ingmar. (2007). Efficient interactive query expansion with complete search. Paper presented at the CIKM'07 Proceedings of the sixteenth ACM conference on Conference on information and knowledge management.
- Belkin, N., Cool, Colleen, Koenemann, Jurgen, BorNg, Kwong, and Park, Soyeon. (1995). Using Relevance Feedback and Ranking in Interactive Searching. Paper presented at the Proceedings of the Fourth Text Retrieval Conference (TREC-4).
- Ben-Hur, A. a. G., Isabelle. (2003). Detecting stable clusters using principal component analysis. Functional Genomics: Methods and Protocols.
- Bernhard, D. (2010). Query expansion based on pseudo relevance feedback from definition clusters. Paper presented at the COLING'10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters.

- Bhattacharya, A. a. D., Rajat K. (2009). Bi-Correlation Clustering Algorithm (BCCA) for determining a set of co-regulated genes. Paper presented at the Bioinformatics Advance Access.
- Bhattacharya, A. a. D., Rajat K. (2010). Average correlation clustering algorithm (ACCA) for grouping of co-regulated genes with similar pattern of variation in their expression values. *Journal of Biomedical Informatics*, 43.
- Boonchom, V. a. S., Nuanwai. (2011). ATOB algorithm: an automatic ontology construction for Thai legal sentences retrieval. *Journal of Information Science*, 38(1).
- Boubekour, F., Boughanem, Mohand, and Tamine-Lechani, Lynda. (2008). Exploiting association rules ontology for semantic document indexing. Paper presented at the International Conference on Information Proceeding and Management of Uncertainty in Knowledge-based Systems (IPMU 2008).
- Bratsas, C., Koutkias, Vassilis, Kaimakamis, Evangelos, Bamidis, Panagiotis, and Maglavera, Nicos. (2007). Ontology-based Vector Space Model and Fuzzy Query Expansion to Retrieve Knowledge on Medical Computational Problem Solutions. Paper presented at the Proceedings of the 29th Annual International Conference of the IEEE EMBS.
- Burrows, C., and Malheiro, Ricardo. (2008). Using information Retrieval Techniques for keyword evaluation and extraction. *IADIS International Conference Information Systems 2008*.
- Buscher, G., White, Ryan W., Dumais, Susan T., and Huang, Jeff. (2012). Large-Scale Analysis of Individual and Task Differences in Search Result Page Examination Strategies. Paper presented at the WSDM '12 Proceedings of the fifth ACM international conference on Web search and data mining.
- Butcher, S., Clarke, Charles L.A., and Cormark, Gordon V. (2010). Implementing and Evaluating Search Engines. *Information Retrieval*.

- Campos, L. M. d., Fernandez, Juan M. and Huete, Juan F. (1998). Query Expansion in information retrieval systems using a Bayesian network-based thesaurus. Paper presented at the UAI'98 Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence.
- Campos, L. M. d., Fernandez-Luna, Juan M., and Huete, Juan F. (2001). Relevance Feedback in the Bayesian Network Retrieval Model : An Approach Based in Term Instantiation. *Advances in intelligent Data Analysis*, 2189.
- Carpineto, C., de Mori, Renato, Romano, Giovanni, and Bigi, Brigitte. (2001). An information-theoretic approach to automatic query expansion. *ACM Transaction in Information Systems (TOIS)*, 19(1).
- Carpineto, C., Osinski, Stanislaw, Romano, Giovanni, and Weiss, Dawid. (2009). A survey of Web Clustering Results. *ACM Computing Surveys (CSUR)*, 41(3).
- Carpineto, C. a. R., Giovanni. (2012). A Survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys (CSUR)*, 44(1).
- Carterette, B., Allan, James, and Sitaraman, Ramesh. (2006). Minimal Test Collection for retrieval evaluation. Paper presented at the SIGIR'06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.
- Chang, Y.-C. a. C., Shyi-Ming. (2006). A New Query Reweighting Method for Document Retrieval Based on Genetic Algorithms. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 10(5).
- Chaudhari, B., and Parikh, Manan. (2012). A Comparative Study of clustering algorithms Using Weka tools. *International Journal of Application or Innovation in Engineering & Management (ITAEM)*, 1(2).
- Chen, C., Zhu, Chonglai, Cheng, Wei, Song, Qiulin, and Cai, Shibang. (2012). Research of Distributed Index Based on Lucene. *Advances in Electronic Engineering, Communication and Management. Lecture Notes in Electronic Engineering*, 1(139).

- Chen, G., et al. (2002). Evaluation and comparison of clustering algorithms in analyzing cell gene expression data. *Statistica Sinica* 12.
- Chen, J., Li, Rowena, and Li, Fei. (2005, 2005, December 6-9). Chinese Information Retrieval Using Lemur: NTCIR-5 CIR Experiments at UNT. Paper presented at the Proceedings of NTCIR-5 Workshop Meeting.
- Chen, L., Wang, Shengrui, and Jiang, Qingshan. (2009). A Robust Algorithm for Fuzzy Document Clustering. Paper presented at the International Conference on Advanced Information Network and Application Workshop.
- Chen, S.-M., Lin Hsi-Ching and Chang, Yu-Chuan. (2006). A New Method for Query Reweighting for Document Retrieval Based on Neural Networks. *Information and Management Sciences*, 17(4).
- Chum. Ondrej, P., James, Sivic, Josef, Isard, Michael, and Zisserman, Andrew. (2007). Total Recall : Automatic Query Expansion with a Generative Feature Model for Object Retrieval. Paper presented at the IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil.
- Chumwatana, T., Kok, Wai Wong, and Hong, Xie. (2009). An automatic indexing technique for Thai texts using frequent max substring. Paper presented at the 2009 Eighth International Symposium on Natural Language Processing.
- Chumwatana, T., Wong, Kok Wai, and Xie, Hong. (2009). An automatic indexing technique for Thai texts using frequent max substring. Paper presented at the 2009 Eighth International Symposium on Natural Language Processing.
- Cock, D. M., and Cornelis, Chris. (2005). Fuzzy Rough Set Based Web Query Expansion. Paper presented at the Proceedings of Rough Sets and Soft Computing in Intelligent Agent and Web Technology, International Workshop at WIAT2005 (2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology).
- Cock, D. M., Cornelis, Chris. (2005). Fuzzy Rough Set Based Web Query Expansion. Paper presented at the Proceedings of Rough Sets and Soft Computing in

Intelligent Agent and Web Technology, International Workshop at WIIAT2005 (2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology).

Collins-Thompson, K. (2009). Reducing the Risk of Query Expansion via Robust Constrained Optimization. Paper presented at the CIKM'09 Proceedings of the 18th ACM conference on Information and knowledge management.

Company, A. T. (2012). Keyword and Search Engines Statistics. from <http://www.keyworddiscovery.com/keyword-stats.html>

Costa Pereira, C. d., Dragoni, Mauro and Pasi, Gabriella. (2010, 2010, January 27–28). User Evaluation of Multidimensional Relevance Assessment. Paper presented at the Proceedings of the 1st Italian Information Retrieval Workshop (IIR'10).

Cui, H., Wen, Ji-Rong, Nie, Jian-Yun, and Ma, Wei-Ying. (2002). Probabilistic Query Expansion Using Query Logs. Paper presented at the Proceedings of the 11th international conference on World Wide Web.

Cui, W., Xu, Mengjia, Sun, Huaya, and Shao, Hong. (2011). Research on Application of Lucene in Medical Image Retrieval System. Paper presented at the International Conference on Computer Science and Network Technology.

Cui, Y., Chen, Yu, and Li, Jingyu. (2012). Research of Information Search Engine in Forestry Based on the Lucene. Lecture Notes in Electrical Engineering Volume 123 : Advances in Automation and Robotics Vol. 2, 123(2), 603-609.

Dalirsefat, S. B., Meyer, Andreia da Silva and Mirhoseini, Seyed Ziyaeddin. (2009). Comparison of similarity coefficients used for cluster analysis with amplified fragment length polymorphism markers in the silkworm, *Bombyx mori*. Journal of Insect Science, 9(71), 1-8.

Dan, L., Linhua, Liu, and Zhaoxin, Zhang. (2013). Research of Text Categorization on WEKA. Paper presented at the 2013 Third International Conference on Intelligent System Design and Engineering Applications.

- Dang, V., Bendersky and Croft, W. Bruce. (2010, 2010, July 19-23). Learning to Rank Query Reformulations. Paper presented at the SIGIR'10 Proceedings of the 33rd international ACM SIGIR conference on Research and Development in information retrieval.
- Dechang, P. a. X., Qin. (2008). A New Fuzzy Clustering Algorithm on Association Rules for Knowledge Management. *Information Technology Journal*, 7(1).
- Díaz-Galiano, M. C., García-Cumbreras, Miguel Angel, Martín-Valdivia, María Teresa, Ureña-López ,L. Alfonso and Montejo-Ráez,Arturo. (2009). Query Expansion on Medical Image Retrieval:MeSH vs. UMLS. *Lecture Notes in Computer Science*, 5706, 732-735.
- Ding, C. a. H., Xiaofeng. (2002). Clsuter merging and splitting in hierarchical clustering algorithms. Paper presented at the Proceedings of the 2002 IEEE International Conference on Data Mining (ICDDM'02).
- Dinh, D., and Tamine, Lynda. (2011). Combining global and local semantic contexts for improving biomedical informaiton retrieval. Paper presented at the European Conference on Information Retrieval (ECIR).
- Doi, H., Seki, Yohei, and Aono, Masaki. (2005). A Patent Retrieval Model Using a Hierarchy of Clusters at TUT. Paper presented at the Proceedings of NTCIR-5 Workshop Meeting, Tokyo.
- Dudoit, S., and Fridlyand, Jane. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. (V). *Genome Biology*.
- Dvorsky, J., Martinovic, Jan, and Snasel, Vaclav. (2004). Query Expansion and Evolution of Topic in Information Retrieval Systems. *Dateso*, 117-127.
- Efron, M. (2008). Query expansion and dimensionality reduction: Notions of optimality in Rocchio relevance feedback and latent semantic indexing. *Information Processing & Management*, 44(1), 163-180.

- Efron, M. (2009). Using Multiple Query Aspects to Build Test Collections without Human Relevance Judgments. Paper presented at the ECIR'09 Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval.
- Eguchi, K. (2005, 2005, 6-9 December). NTCIR-5 Query Expansion Experiments using Term Dependence Models. Paper presented at the Proceedings of NTCIR-5 Workshop Meeting.
- Facundes, N. a. T.-a., Ratchaneekorn. (2007). Bilingual Search Engine Augmented with Query Expansion. Paper presented at the Seventh International Conference on Intelligent Systems Design and Applications.
- Ferreira, L. a. H., David B. (2009). A comparison of hierarchical methods for clustering functional data. *Communications in Statistics - Simulation and Computation*, 38(9).
- Fonseca, B. M., Golgher, Paulo, Possas, Bruno Ribeiro-Neto, Berthier, and Ziviani, Nivio. (2005). Concept-Based Interactive Query Expansion. Paper presented at the CIKM '05 Proceedings of the 14th ACM international conference on Information and knowledge management,.
- Fritsch, A. a. I., Katja. (2009). Improved Criteria for Clustering Based on the Posterior Similarity Matrix. *Bayesian Analysis*, 4(2).
- Greenberg, J. (2001). Automatic Query Expansion via lexical-semantic relationships. *Journal of the American Society for Information Science and Technology*, 52(5).
- Gulati, P., and Sharma, A.K. (2010, August :). Ontology Driven Query Expansion for Better Image Retrieval. *International Journal of Computer Applications* (0975-8887), 5(10), 33-37.
- Han, J., Kamber, Micheline, and Pei, Jian. (2012). *Data Mining : Concepts and Techniques* (3 ed.). United States of America: Elsevier.
- Han, L., and Chen, Guihai. (2009). HQE: A hybrid method for query expansion. *Expert Systems with Application*, 36.

- Hancock-Beaulieu, M., and Walker, Stephen. (1992). An Evaluation of Automatic Query Expansion in An Online Library Catalogue. *Journal of Documentation*, 48(4).
- Haruechaiyasak, C., Srichaivattana, Prapass, Kongyoung, Sarawoot, and Damrongrat, Chaianun. (2008). Automatic Thai Keyword Extraction from Categorized Text Corpus.
- Herbert, C. W. (2005). An Application of Machine Learning Techniques to Interactive, Constraint-Based Search. (Master's Thesis of Science), University of Missouri, Columbia.
- Hmeidi, I. I., Najadat, Hassan M., and Al-Sha'or, Ahmed J. (2008). Application of genetic algorithm in automatic query expansion. Paper presented at the The 2008 International Arab Conference on Information Technology (ACIT'2008).
- Holmes, G., Donkin, Andrew, and Witten, Ian H. (1994, 1994, Nov 29 - Dec 2). WEKA : A Machine Learning Workbench. Paper presented at the Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems.
- Horng, Y.-J., Chen, Shyi-Ming Chang, Yu-Chuan, and Lee, Chia-Hoang. (2005). A New Method for Fuzzy Information Retrieval Based on Fuzzy Hierarchical Clustering and Fuzzy Inference Techniques. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 13(2).
- Horng, Y.-J. a. C., Shyi-Ming. (2003). A New Fuzzy Information Retrieval Method Based on Document Terms Reweighting Techniques. *Information and Management Sciences*, 14(4).
- Hu, Q., Huang, Jimmy Xiangji, and Hu, Xiaohua. (2012, 2012, June 13). Modelig and mining term association for improving biomedical information retrieval performance. Paper presented at the BMC Bioinformatics.
- Huang, A. (2008). Similarity Measures for Text Document Clustering. Paper presented at the the proceedings of the New Zealand Computer Science Research Student Conference 2008.

- Huiskes, M. J. a. L., Michael S. (2008). Performance Evaluation of Relevance Feedback Methods. Paper presented at the CIVR '08 Proceedings of the 2008 international conference on Content-based image and video retrieval.
- Imran, H. a. S., Aditi. (2009). Thesaurus and query expansion. International Journal of Computer science & Information Technology(IJCSIT), 1(2).
- Ircing, P., et al. (2007). Information Retrieval Test Collection for Searching Spontaneous Czech Speech. Paper presented at the Proceedings of the 10th International Conference on Text, Speech, and Dialogue.
- Jamshidi, O. a. P., Abdol. (2013). Automatic Segmentation of Medical Image Using Fuzzy c-Means and the Genetic Algorithms. Journal of Computational Medicine.
- Jaruskulchai, C. (2002). Dictionary-Based Thai CLIR: An Experimental Survey of Thai CLIR. EVALUATION OF CROSS-LANGUAGE INFORMATION RETRIEVAL SYSTEMS, 2406.
- Jayalakshmi, G. a. R., Nageswara. (2009). Mining association rules for large transaction using new support and confidence measure. Journal of Theoretical and Applied Information Technology, 7(1).
- Jingjing Liu, et al. (2010). Search behaviors in different task types. Paper presented at the JCDL '10 Proceedings of the 10th annual joint conference on Digital libraries.
- Jipkate, B. R. a. G., V.V. (2012). A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. International Journal Of Computational Engineering Research, 2(3).
- Joho, H., Sanderson, Mark, and Beaulieu Micheline. (2004). A Study of User Interaction with a Concept-Based Interaction Query Expansion Support Tool. Advance in Information Retrieval, 2997.
- Jun-qiang, Y., Jia, Du, Shi-ming, Zheng, Lei, Dong, and Bing, Qiao. (2012). Research on Fuzzy Clustering Algorithm WEKA-Based. Paper presented at the

Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011).

Kaczmarek, A. L. (2011). Interactive Query Expansion With the Use of Clustering-by-Directions Algorithm. *IEEE Transactions on Industrial Electronics*, 58(8).

Kalmanovich, I. G., and Kurland, Oren. (2009). Cluster-based query expansion. Paper presented at the SIGIR'09 Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval.

Kanaan, G., Al-Shalabi, Riyadh, Ghwanmeh, Sameh and Bani-Ismael, Basel. (2007, 2007, 18-20 Nov). A Comparison between Interactive and Automatic Query Expansion Applied on Arabic Language. Paper presented at the Fourth Innovations in Information Technology, 2007 (IIT '07).

Kanaan, G., Al-Shalabi, Riyadh, Ghwanmeh, Sameh and Bani-Ismael, Basel. (2008). Interactive and Automatic Query Expansion: A Comparative Study with an Application on Arabic. *American Journal of Applied Sciences*, 5(11).

Kaptein, R., Kamus, Jaap, and Hiemstra, Djoerd. (2008, 2008, November 18-21). The Impact of Positive, Negative and Topical Relevance Feedback. Paper presented at the Text Retrieval Conference (TREC).

Kashef, R. K., and Kamel, M.S. (2006, 8-10 November 2006). Distributed Cooperative Hard-Fuzzy Document Clustering. Paper presented at the Proceedings of the 3rd Annual Scientific Conference of the LORNET Research Network (I2LOR-06), Montreal, Canada.

Khankasikam, K. (2010). A Comparison of Information Retrieval Model to Thai Digital Library. 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE).

Kim, J. (2008). Task as a Context of Information Seeking: An Investigation of Daily Life Tasks on the Web. *Libri. International Journal of Libraries and Information Services*, 58.

- Kim, M. B., Kim, Ju Youn, and Kim, Jongwan. (2001). Query tern Expansion and reweighting using term co-occurrence similarith and a fuzzy interference. Paper presented at the IFSA World Congress and 20th NAFIPS International Conference.
- Kim, M. B., Kim, Ju Youn, and Kim, Jongwan. (2001). Query tern Expansion and reweighting using term co-occurrence similarith and afuzzy interference. Paper presented at the IFSA World Congress and 20th NAFIPS International Conference.
- Kim, Y., Seo Jangwon, and Croft, W.Bruce. (2011). Automatic Boolean Query Suggestion for Professional Search. Paper presented at the SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval.
- Kimura, Y., and Araki, Kenji. (2006). Query Expansion for Contextual Question Using Genetic Algorithms. *Information Retrieval Technology*, 4182.
- Kirkland, O., and De La Iglesia, Beatriz. (2013). Experimental Evaluation of Cluster Quality Measures. 2013 13th UK Workshop on Computational Intelligence (UKCI).
- Klink, S., Hust, Armin, Junker, Markus, and Dengel. (2002). Andreas.Improving Document Retrieval by Autometic Query Expansion Using Collaborative Learning of Term-Based Concepts. Paper presented at the DAS '02 Proceedings of the 5th International Workshop on Document Analysis Systems V.
- Kunpeng, Z., Xiaolong, Wang, and Yuanchao, Liu. (2006). A new query expansion method based on query logs mining. *International Journal on Asian Language Processing*, 19(1), 1-12.
- Kwon, O.-W., Kim, Myoung-Cheol, and Choi, Key-Sun. (1994). Query expansion using domain-adapted, weighted thesaurus in an extended Boolean model. *CIKM '94 Proceedings of the thrid international conference on Information and knowledge management*.

- Lai, K. a. C., N. (2001). Support vs Confidence in Association Rule Algorithms. Paper presented at the Proceedings of the OPTIMA Conference, Curicó.
- Langfelder, P. a. H., Steve. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software*, 46(11).
- Latiri, C. C., Yahia, S.B., Chevallet, J.P., and Jaoua, A. (2004). Query Expansion using fuzzy association rules between terms. Paper presented at the Proceedings of the 4 th International Conference Journees de l'Informatique Messine (JIM'03).
- Lee, H.-M., Lin, Sheng-Kang, and Huang, Chiung-Wei. (2001). Interactive query expansion based on fuzzy association thesaurus for Web information retrieval. Paper presented at the Fuzzy Systems. The 10th IEEE International Conference on.
- Lee, H.-M., Huang, Chi-chun, and Chao, Chun-Yen. (2007). Association Thesaurus Construction for Interactive Query Expansion Based on Association Rule Mining. *Journal of Information Science and Engineering* 23.
- Lee, K. S., Croft, W. Bruce, and Allan, James. (2008). A cluster-based resampling method for pseudo-relevance feedback. Paper presented at the SIGIR'08 Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- Lei, J., Li, Weijiang, Wang, Feng, and Deng, Hui. (2011). A Survey on Query Expansion Based on Local Analysis. 2011 Fourth International Conference on Intelligent Networks and Intelligent Systems, 1-4.
- Li, J., Zhang, Xiuzhen, Dong, Guozhu, Ramamohanarao, Kotagiri, and Sun, Qun. (1999). Efficient Mining of High Confidence Association Rules without Support Thresholds. *Principles of Data Mining and Knowledge Dsiccovery*.
- Li, M. J., Ng, Michael K., Cheung, Yiu-ming, and Huang, Joshua Zhexue. (2008). Agglomerative Fuzzy K-Means Clustering Algorith, with Selection of Number of Clusters. *IEEE Transactions on Knowledge and Data Engineering*, 20(11).

- Li, Z., Xu, Yue, and Geva, Shlomo. (2008). Text Mining Based Query Expansion for Chinese IR. Paper presented at the Proceedings of the Australasian Language Technology Association Workshop 2008.
- Lin, Y., Lin, Hongfei, and He, Li. (2012). A Cluster-based Resource Correlative Query Expansion in Distributed Information Retrieval. *Journal of Computational Information Systems*, 8(1), 31-38.
- Linoff, G. S. a. B., Michael J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). Indiana: Wiley Publishing, Inc.
- Lita, L. V., and Carbonnell, Jaime G. (2008). Cluster-Based Query Expansion for Statistic Question Answering. Paper presented at the International Joint Conference on Natural Language Processing.
- Liu, F., Pennell, Deana, Liu, Fei and Liu, Yang. (2009). Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts. *Human Language Technologies*. Paper presented at the The 2009 Annual Conference of the North American Chapter of the ACL.
- Liu, X., Fang, Hui, Chen, Fei, and Wang, Min. (2012). Entity Centric Query Expansion for Enterprise Search. Paper presented at the CIKM'12 Proceedings of the 21st ACM international conference on Information and knowledge management.
- Liu, y.-c., Wang, Xiao-long and Liu, Bing-quan. (2004, 2004, August 26-29). A feature selection algorithm for document clustering based on word co-occurrence frequency. Paper presented at the Proceedings of the Thrid International Conference on Machine Learning and Cybernetics.
- Liu, Z., and Chu, Wesley W. (2007). Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Information Retrieval*, 10(2).
- Liu, Z., Natarajan, Sivaramkrishnan, and Chen, Yi. (2011, 2011, March). Query Expansion based on clustered results. Paper presented at the Proceedings of the VLDB Endowment.

- Lu, X. A., and Keefer, Robert B. (1995). Query expansion/reduction and its impact on retrieval effectiveness. Paper presented at the Proceedings of the Thrid Text Retrieval Conference TREC-3.
- Lu, Z., Kim, Won, and Wilbur, W. John. (2009). Evaluating Relevance Ranking Strategies of MEDLINE Retrieval. *Journal of the American Medical Informatics Association*, 16(1).
- Lu, Z., Kim, Won, and Wilbur, W. John. (2009). Evaluaton of Query Expansion Using MeSH in PubMed. *NIHPA Author Manuscripts*, 12(1).
- Luke, T., Schaer, Philipp, and Mayr, Philipp. (2012). Improving retrieval result with discipline-specific query expansion. Paper presented at the TPD L'12 Proceedings of the Second international conference on Theory and Practice of Digital Libraries.
- Lux, M., and Chatzichristofis, Savvas A. (2008). Lire : lucene image retrieval : an extensible java CBIR library. Paper presented at the MM'08 Proceedings of the 16th ACM international confernece on Multimedia.
- Lv, Y. a. Z., ChengXiang. (2009). Adaptive Relevance Feedback in Information Retrieval. Paper presented at the CIKM '09 Proceedings of the 18th ACM conference on Information and knowledge management.
- Macdonald, C. a. O., Iadh. (2006). The TREC Blogs06 Collection : Creating and Analysing a Blog Test Collection. Technical Report.
- Mahmoodi, H. a. M., Eghbal. (2012). Document Clustering Based On Semi-Supervised Term Clustering. *Internation Journal of Artificial Intelligence & Applications (IJAIA)*, 3(3).
- Marrison, P. J. (2007). Tagging and Searching : Search Retrieval Effectiveness of Folksonomies on the Web. (Master of thesis's Science), College of Communication and Information of Kent State University.

- Massoudi, K., Tsagkias, Manos, de Rijke, Maarten, and Weerkamp, Wouter. (2011). Incorporating Query Expansion and Quality Indicators in Searching Microblog Posts. Paper presented at the ECIR'11 Proceedings of the 33rd European conference on Advances in information retrieval.
- Meiyappan, Y., and Iyengar, SrimanNarayana. (2011). Interactive Query Expansion Using Concept-Based Directions Finder Based on Wikipedia. The International Arab Journal of e-Technology (IAJeT).
- Meji, E., and Rijke, Maarten de. (2006, 2006, August 6-11). Deploying Lucene on the Grid. Paper presented at the Proceedings SIGIR 2006 workshop on Open Source Information Retrieval (OSIR2006).
- Meshram, V., Mishra, Nishchol, and Sharma, Sanjeev. (2012). SVM Based Navigation Patterns for Efficient Relevance Feedback in Content Based Image Retrieval. Int.J.Computer Technology & Application, 3(4).
- Metzler, D., Trevor, Strohman, Zhou, Yun, and Croft, W. Bruce. (2006). Indri at TREC 2005 : Terabyte Track. Paper presented at the Online Proceedings of 2005 Text REtrieval Conference (2006).
- Miller, A. (2010, February, 17, 2010). Finally some useful SERP click statistic. from <http://www.yoursearchadvisor.com/blog/finally-some-useful-serp-click-statistics>
- Mingoti, S. A. a. L., Joab O. (2006). Comparing SOM neural network with Fuzzt c-means, K-means and traditional hierarchical clustering algorithms. European Journal of Operational Research, 174, 1742-1759.
- Morita, K., Arai, Shuto, Kitagawa, Hiroya Fuketa, Masao and Aoe, Jun-ichi. (2011). Dynamic Construction of Hierarchical Thesaurus using Co-occurrence Information. 2011 2nd International Conference on Networking and Information Technology IPCSIT, 17.
- Muyeba, M., Khan, M. Sulaiman, and Coenen, Frans. (2009). Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework. New Frontiers in Applied Data Mining, 5433.

- Na, S.-H., Kang In-Su, Roh, Ji-Eun and Lee, Jong-Hyeok. (2007). An Empirical study of query expansion and cluster-based retrieval in language model approach. Paper presented at the Information Processing and Management.
- Nallapati, R., Allan, James and Mahadevan, Sridhar. (2004). Extraction of Key Words from News Stories. University of Massachusetts. Retrieved from <http://maroo.cs.umass.edu/getpdf.php?id=488>
- Nguyen, T.-Q., Heo, Jun-Seok, Lee, Jung-Hoon, Kim, Yi-Reun, and Whang Kyu-Young. . JCSE. (2008). Query Expansion Using Augmented Terms in an Extended Boolean Model. Joint Communications Support Element (JCSE), 2(1).
- Nixon, J. (2000). How does the UK NHS compare with European standards? A review of EU health care systems using hierarchical cluster analysis. Department of Economics, University of York. Retrieved from <http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20182.pdf>
- Oard, D. W. e. a. (2004). Building an Information Retrieval Test Collection for Spontaneous Conversational Speech. SIGIR'04 Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval.
- Oelze, I. (2009). Automatic Keyword Extraction for Database Search. (Master).
- Ogilvie, P., and Callen, Jamie. (2001). The effectiveness of query expansion for distributes information retrieval. Paper presented at the CIKM'01 Proceedings of the tenth international conference on Information and knowledge manangement.
- Ogilvie, P., and Callen, Jamie. (2005). Experiments with Language Models for Known-Item Finding of E-mail Messages. Paper presented at the Proceedings of the Fourteenth TextRetrieval Conference (TREC-14).
- Oh, C.-H., Honda, Katsuhiko and Ichihashi, Hidetomo. (2009, 2009, September). Graded Possibilistic Approach to Fuzzy Clustering for Cooccurrence Matrix with

- Regularization by K-L Information. Paper presented at the ISIS 2007 Proceedings of the 8th symposium on advanced intelligent systems.
- Park, J. H. a. C., W. Bruce. (2010, 2010, July 19-23). Query term ranking based on dependency parsing of verbose queries. Paper presented at the SIGIR'10 Proceedings of the 33rd international ACM SIGIR conference on Research and Development in information retrieval.
- Peetz, M.-H., and Lopatka, Martin. (2007). Query Expansion With Wikipedia. Paper presented at the Proceedings of IIR, University of Amsterdam.
- Perez-Aguera, J. R. a. A., Lourdes. (2008). Comparing and Combing Methods for Automatic Query Expansion. Advances in Natural Language Proceeding and Applications. Research in Computing Science 33, 177-188.
- Rahimi, M., and Zahedi. (2012). Query expansion based on contextual meaning of the query terms. International Journal of the Physical Science, 7(24).
- Rahimi, M., and Zahedi. (2012). Query expansion based on contextual meaning of the query terms. International Journal of the Physical Science, 7(24).
- Rahman, M. M., and Bhattacharya, Prabir. (2009). Image retrieval with automatic query expansion based on local analysis in a semantical concept feature space. Paper presented at the CIVR'09 Proceedings of the ACM International Conference on Image and Video Retrieval.
- Rahman, M. M., Antani, S.K., and Thoma, G.R. (2011). A Query Expansion Framework in Image Retrieval Domain Based on Local and Global Analysis. Inf Process Manag, 47(5).
- Rashid, N. R. e. a. (2011). A Comparision between single linkage and complete linkage in agglomerative hierarchical clustering analysis for identifying tourists segments. IIUM Engineering Journal, 12(6).

- Raskutti, B., and Leckie, Christopher. (1999). An Evaluation of Criteria for Measuring the Quality of Clusters. Paper presented at the IJCAI'99 Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence.
- Razmara, M. a. K., Leila. (2008). Answering List Questions using Co-occurrence and Clustering. Paper presented at the International Conference on Language Resources and Evaluation.
- Reichart, R., and Rappoport, Ari. (2009). The NLI Clustering Evaluation Measure. Paper presented at the Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL).
- Riezler, S., Liu, Yi, and Vasserman Alexander. (2008). Translating queries into snippets for improved query expansion. Paper presented at the COLING' 08 Proceedings of the 22nd International Conference on Computational Linguistics.
- Ritchie, A., Teufel, Simone and Robertson, Stephen. (2006). Creating a Test Collection for Citation-based IR Experiments. Paper presented at the HLT-NAACL'06 Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics.
- Rivas, A. R., Borrajo, L., Iglesias, E.L., and Romero, R. (2012). Applying Lemur Query Expansion Techniques in Biomedical Information Retrieval. *Advances in Intelligent and Soft Computing Volume 151 : Distributed Computing and Artificial Intelligence*.
- Rose, S., Engel, Dave, Cramer, Nick and Cowley, Wendy. (2010). Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory*.
- Rousseeuw, P. J. (1987). Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1987).
- Rungsawang, A., Tangpong, A., Laohawee, P., and Khampachua, T. (1993). Novel Query Expansion Technique using Apriori Algorithm. Paper presented at the Proceedings of The Eighth Text Retrieval Conference (TREC 8).

- Ruthven, I. (2003). Re-examining the potential effectiveness of interactive query expansion. Paper presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval.
- Ruthven, I. a. L., Mounia. (2003, June). A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(2).
- Sahib, N. G., Tombros, Anastasios and Ruthven, Ian. (2010). Enabling Interactive Query Expansion through Eliciting the Potential Effect of Expansion Terms. *Lecture Notes in Computer Science*, 5993/2010.
- Sajja, K. (2011). Performance Study of Lucene in Parallel and Distributed Environments. (Master of Science's Thesis (Computer Science)), Boise State University.
- Sakai, T., and Mitamura, Teruko. (2010). Boiling Down Information Retrieval Test Collection. *RIAO'10 Adaptivity, Personalization and Fusion of Heterogeneous Information*.
- Sanjuan, E., Ibekwe-SanJuan, Fidelia, Torres-Moreno, Juan-Manuel and Velazquez-Morales, Patricia. (2007). Combining Vector Space Model and Multi Word Term Extraction for Semantic Query Expansion. *Natural Language Processing and Information Systems*, 4592/2007.
- Schenker, A., Last, Mark, Bunke, Horst and Kandel, Abraham. (2004). Comparison of Algorithms for Web Document Clustering Using Graph Representations of Data. *Lecture Notes in Computer Science*, 3138/2004.
- Sesli, M. a. Y., E.D. (2010). Compare various combinations of similarity coefficients and clustering methods for *Olea europaea sativa*. *Scientific Research and Essays*, 5(16).
- Seung, Y. Y. a. H., Achim. (2006). Clustering-based relevance feedback for web pages. Paper presented at the *PRICAI'06 Proceedings of the 9th Pacific Rim international conference on Artificial intelligence*.

- Shapira, B., Taieb-Maimon, Meir, and Nemeth, Yael. (2005). Subjective and objective evaluation of interactive and automatic query expansion. *Online Information Review*, 2, 374-390.
- Sharma, D. K. a. S., A.K. (2010). A Novel Ranking Algorithm of Query Words Stored in QIIIEP Server. *International Journal of Engineering Science an Technology*, 2(11).
- Sharma, N., Bajpai, Aman, and Litoriya, Ratnesh. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emering Technology and Advanced Engineering*, 2(5).
- Shekarpour, S., Hoffner, Konrad, Lehmannm Jens, and Auer, Soren. (2013, 2013, September 16-18). Keyword Query Expansion on Linked Data Using Linguistic and Semantic Features. Paper presented at the 7th IEEE International Conference on Semantic Computing.
- Shiri, A., and Revie, Crawford. (2006). Query Expansion Behavior Within a Thesaurus-Enhanced Search Enviroment : A User-Centered Evaluation. *Journal of the American Society for Information Science and Technology*, 57(4).
- Shrovastava, R., Upadhyay, Khushbu, Bhati, Raman, and Mishra, Durgesh Kumar. (2010, 2010, 28-30 September). Comparison between K-Mean and C-Mean Clustering for CBIR. Paper presented at the Second International Conference on Computational Intelligence, Modelling and Simlation.
- Slonim, N. a. T., Naftali. (2000). Document clustering using word clusters via the inoformation bottleneck method. Paper presented at the Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval.
- Song, M., Song, Il-Yeol, Hu, Xiaohua, and Allen, Robert. (2005). Semantic Query Expansion Combining Association Rules with Ontologies and Information Retrieval Techniques. *Data Warehouse and Knowledge Discovery*, 2589.

- Song, M., Song, Il-Yeol, Hu, Xiaohua, and Allen, Robert B. (2007). Integration of association rules and ontologies for semantic query expansion. *Data & Knowledge Engineering* 63.
- Steinbach, M., Karypis, and Kumar, Vipin. (2000). A Comparison of Document Clustering Techniques. *KDD Workshop on Text Mining*.
- Strohman, T., Metzler, Donald, Turtle, Howard, and Croft, W. Bruce. (2004). Indri : A language-model based search for complex queries. Paper presented at the Proceedings of the International Conference on Intelligence Analysis.
- Takagi, T., and Tajima, Masanori. (2001). Query Expansion Using Conceptual Fuzzy Sets For Search Engine. Paper presented at the IEEE International Fuzzy Systems Conference.
- Tantisripreecha, T. a. S., Nuanwan. (2010). Query expansion algorithm for supreme court sentences retrieval using ontology. Paper presented at the Proceedings of 48th Kasetsart University Annual Conference: Science.
- Tao, Q., Gao, Ling, and Zhang, Zhe. . . . (2011, 2011, 27-29 May). The study of semantic query expansion based on improved vector space model. Paper presented at the Communication Software and Networks (ICCSN).
- Thaiprayoon, S., Kongthon, Alisa, Palingoon, Pornpimon, and Haruechaiyasak, Choochart. . . (2012, 2012, 16-18 May). Search Result Clustering for Thai Twitter Based on Suffix Tree Clustering. Paper presented at the 2012 9th International Conference on Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology (ECTI-COM).
- Theeramunkong, T., Somlertlamvanich, Virach, Tanhermhong, Thanasan, and Chinnan, Wirat. (2000). Character Cluster Based Thai Information Retrieval. Paper presented at the IRAL'00 Proceedings of the fifth international workshop on Information retrieval with Asian Languages.

- Theeramunkong, T., Chinnan, Wirat, Tanhermhong, Thanasan, and Sornlertlamvanich, Virach. (2000). Full-text search for Thai information retrieval systems. Paper presented at the The fourth Symposium on Natural Language Processing 2000.
- Theeramunkong, T. (2000). Passage-based Web text mining. Paper presented at the IRAL'00 Proceedings of the fifth international workshop on.
- Theva-askorn, R. (2003). Applying a Corpus-based Dictionary For Query Expansion with Lexical Semantic Relation. (Master's thesis of Engineering (Computer Engineering)), King Mongkut's University of Technology Thonburi, Bangkok.
- Tianyuan, L., Meina, Song, and Xiaoqi, Zhang. (2010). Research of Massive Heterogeneous Data Integration Based on Lucene and XQuery. 2010 IEEE 2nd Symposium on Web Society (SWS).
- Tibshirani, R., Walther, Guenther, and Hastie, Trevor. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63(2).
- Tuffery, S. (2011). *Data Mining and Statistics for Decision Making*. United Kingdom: John Wiley & Sons, Ltd.
- Turney, P. D., and Pantel, Patrick. (2010). From Frequency to Meaning : Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37.
- Turtle, H., Hegde, Yatish, and Rowe, Steven A. (2012). Yet another comparison of Lucene and Indri performance. Paper presented at the Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval.
- Vega-Pons, S. a. R.-S., Jose. (2011). *Internation Journal of Pattern Recognition and Artificial Intelligence*. 25(3).
- Vilares, J., Vilares, Manuel, and Alonso, Miguel. (2001). Towards the Development of Heuristics for Automatic Query Expansion. *Database and Expert Systems Applications*, 2113.

- Wang, S.-c., and Tanaka, Yuzuru. (2006). Topic-Oriented Query Expansion for Web Search. Paper presented at the WWW'06 Proceedings of the 15th international conference on World Wide Web.
- Wang, Z., and Niu, Qiang. (2012). A Hybrid Query Expansion Algorithm Combined Global Analysis and Ontology Technology. Paper presented at the Advanced Science and Technology Letters (ASTL) : Information Science and Technology (IST), Shanghai, China.
- Weerkamp, W., Balog, Krisztian and Rijke, Maarten de. (2009, 2009, August 2-7). A Generative Blog Post Retrieval Model that Uses Query Expansion based on External Collections. Paper presented at the Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP.
- Weiss, S. M., Indurkha, Nitin, Zhang, Tong, and Damerau, Fred J. (2005). Text Mining. United States of America: Springer Science+Business Media.
- Willson, R. a. G., Lisa M. (2008, 2008, June 5-7). The Effect of Misspellings on Information Retrieval in Online Public Access Catalogues. Paper presented at the Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British, Vancouver, Columbia.
- Willson, R. a. G., Lisa M. (2010). The Effect of Spelling and Retrieval System Familiarity on Search Behavior in Online Public Access Catalogs: A Mixed Methods Study. Journal of the American Society for Information Science and Technology, 61(12).
- Witten, I., Frank, Eibe, Trigg, Len, Hall, Markm Holmes, Geoffrey, and Cunningham, Sally Jo. (1999). Weka : Practical Machine Learning Tools and Techniques with Java Implementations. Paper presented at the Proceedings of the ICONIP/ANZIIS/ANNE'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information System.
- Wu, S. a. M., Sally. (2006). Information Retrieval Evaluation with PartialRelevance Judgment. Paper presented at the BNCOD'06 Proceedings of the 23rd British

National Conference on Databases, conference on Flexible and Efficient Information Handling.

- Wu, X. B., Alvaro. (2008). Keyword Extraction for Contextual Advertisement. Paper presented at the WWW '08 Proceedings of the 17th international conference on World Wide Web.
- Xu, J. a. C., W.Bruce. (1996). Query Expansion Using Local and Global Document Analysis. Paper presented at the SIGIR '96 Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval.
- Xu, X., Zhu, Weizhong, Zhang, Xiaodan, Hu, Xiaohua, and Song, Il-Yeol. (2006, October 8-11,2006). A Comparison of Local Analysis, Global Analysis and Ontology-based Query Expansion Strategies for Bio-medical Literature Search. Paper presented at the SMC'06 IEEE International Conference on System, .
- Xu, X., Xiaodan, Zhu, Hu, Xiaohua, and Song, Il-Yeol. (2006, 2006, October 8-11). A Comparison of Local Analysis, Global Analysis and Ontology-based Query Expansion Strategies for Bio-medical Literature Search. Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics.
- Xu, X. a. H., Xiaohua. (2010). Cluster-based Query Expansion Using Language Modeling in the Biomedical Domain. Paper presented at the 2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops.
- Xu , Y. C. a. C., Zhiwei. (2005, 2005, 03-06 January). User-Oriented Relevance Judgment: A Conceptual Model. Paper presented at the Proceedings of the 38th Hawaii International Conference on System Sciences.
- Xu , Y. C. a. C., Zhiwei. (2006). Relevance Judgment : What Do Information Users Consider Beyond Topicality? Journal of the American Society for Information Science and Technology, 57(7).

- Yan, M. (2005). Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion. (Master's Thesis of Doctor of Philosophy), Virginia Polytechnic Institute and State University.
- Yan, R., and Haupmann, Alexander. (2007). Query Expansion using probabilistic local feedback with application to multimedia retrieval. Paper presented at the CIKM'07 Proceedings of the sixteenth ACM conference on Conference on information and knowledge management.
- Yang, H. a. Z., Minjie. (2002, 18-20 July 2002). Query expansion with naive bayes for searching distributed collections. Paper presented at the Proceedings of the 11th Conference on Intelligent Systems: Emerging Technologies.
- Ye, Z., Huang, Xiangji, and Lin, Hongfei. (2011). A Bayesian network approach to context sensitive query expansion. Paper presented at the SAC'11 Proceedings of the 2011 ACM Symposium on Applied Computing.
- Yilmaz, K. E., Arslan, Ahmet, and Yilmazel, Ozgur. (2009). Turkish Text Retrieval Experiments Using Lemur Toolkit. Paper presented at the IADIS International Conference Applied Computing 2009.
- Yin, P.-Y., Bhanu, Bir, Chang, Kuang-Cheng and Dong, Anlei. (2005). Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning. IEEE Transactions on pattern analysis and machine intelligence, 27(10), 1536-1551.
- Yin, Y., Kaku, Ikou, Tang, Jiafu and Zhu, JianMing. (2011). Data Mining : Concepts, Methods and Applications in Management and Engineering Design. New York: Springer London Dordrecht Heidelberg.
- Yuan, Y., Zhang, Yong, and Xing, Chunxiao. (2007, 2007, 26-27 July). Specific Academic Area based Automatic Query Expansion. Paper presented at the Pervasive Computing and Application, 2007. ICPCA 2007. 2nd International Conference on.

- Yurekil, B., Capan, Gokhan, Yilmazel, Baris, and Yilmazel, Ozgur. (2009). Guided Navigation Using Query Log Mining through Query Expansion. Paper presented at the NSS'09 Proceedings of the 2009 Thrid International Conference on Netword and System Security.
- Zamir, O., and Etzioni, Oren. (1998). Web document clustering : a feasibility demonstration. Paper presented at the SIGIR'98 Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval.
- Zhang, X. (2002). Collaborative Relevance Judgment: A Group ConsensusMethod for Evaluating User Search Performance. *Journal of the American Society for Information Science and Technology*, 53(3).
- Zhang, Y., Zhang, Tao, and Chen, Shijie. (2005). Research on Lucene-based English-Chinese Cross-Language Information Retrieval. *Journal of Chinese Language and Computing*, 15(1).
- Zhang, Y., and Li, Jian-lin. (2009). Research and Improvement of Search Engine Based on Lucene. Paper presented at the International Conference on Intellogent Human-Machine Systems and Cybernetics.
- Zhang, Y., Xu, Xiaofei, and Yunming, Ye. (2010). NSS-AKmeans : An Agglomerative Fuzzy K-Means clustering method with automatic selection of cluster number. Paper presented at the Second International Conference on Advanced Computer Control (ICACC).
- Zhang, Y., Xu, Xiaofei, Liu, Yingqun, Li, Xutao, and Ye, Yumming. (2011). An Agglomerative Fuzzy K-Means Approach to Building Decision Cluster Classifiers. Paper presented at the Second International Confernece on Innovations in Bio-inspired Computing and Application.
- Zhang, Z., Ji, Rongrong, Yao, Hongxun, Xu, Pengfei, and Wang, Jicheng. (2007). Random Sampling SVM Based Soft Query Expansion for Image Retrieval. Paper presented at the Fourth International Conference on Image and Graphics.

- Zhao, S., Wu, Hong-Xia, Lin, Yong-Min. . . : . (2012). Research on Bayesian Network Retrieval Model Based on Query Expansion. Emerging Research in Artificial Intelligence and Computational Intelligence Communications in Computer and Information Science 2012.
- Zhao, Y., Wang, Xiao-Long and Liu, Bing-Quan. (2004). Query Expansion Using Trigger-based Vector Space Model. Paper presented at the Proceedings of the First International Conference on AIRS, Beijing.
- Zhu, W., Xu, Xuheng, Hu, Xiaohua, Song, Il-Yeol, and Allen, Robert B. (2006, 2006, 10-12 May). Using UMLS-based Re-weighting Terms as Query Expansion Strategy. Paper presented at the 2006 IEEE International Conference on Granular Computing.



ภาคผนวก ก ตัวอย่างเอกสารและรายการเอกสารที่นำมาใช้ในงานวิจัย

งานวิจัยนี้ได้กำหนดใช้เอกสารเป็นรูปแบบบทคัดย่อโครงการงานชั้นปี 4 ที่จัดทำโดยนิสิตคณะพาณิชยศาสตร์และการบัญชี สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย ในช่วงปี 2552-2554 จำนวน 100 ฉบับ โดยมีความยาวบทคัดย่อเป็นหน้ากระดาษขนาด A4 จำนวนตั้งแต่ 1 ถึง 3 หน้าโดยที่รูปที่ ก.1 และ ก.2 แสดงตัวอย่างบทคัดย่อโครงการงานที่จัดทำในช่วงปี 2552 รูปที่ ก.3 และ ก.4 แสดงตัวอย่างบทคัดย่อโครงการงานที่จัดทำในช่วงปี 2553 และรูปที่ ก.5 และ ก.6 แสดงตัวอย่างบทคัดย่อโครงการงานที่จัดทำในช่วงปี 2554

จากตัวอย่างเอกสารทั้งหมดที่นำเสนอ นั้น จะเห็นได้ว่า เอกสารนั้นจะมีขอบเขตข้อมูลดังต่อไปนี้

1. รหัสโครงการ
2. รหัสประจำตัวนิสิต
3. ชื่อนิสิต
4. ชื่ออาจารย์ที่ปรึกษา
5. ชื่อหัวข้อภาษาไทย
6. ชื่อหัวข้อภาษาอังกฤษ
7. คำศัพท์สำคัญ
8. บทคัดย่อ

ตารางที่ ก.1 แสดงรายการเอกสารที่นำมาใช้ในการสร้างชุดทดสอบเพื่อนำมาใช้ในการประเมินประสิทธิภาพการค้นคืน และสร้างกลุ่มคำเพื่อนำมาใช้ในการขยายข้อสอบถาม สำหรับการศึกษาในงานวิจัยนี้

52B03

บทสรุปย่อรายงาน

เลขประจำตัวนิสิต : 494 20021 26

ชื่อนิสิต : นางสาว กมลชนก ศิริโพบูลย์ทรัพย์

อาจารย์ที่ปรึกษา : อ.ดร.วัชรา จันทาทับ

หัวข้อโครงการ(ภาษาไทย) : การศึกษาแนวปฏิบัติที่ดีในการพัฒนาโมบายคอมเมิร์ซและตัวอย่างการนำไปใช้

หัวข้อโครงการ(ภาษาอังกฤษ) : A Study of Good Practices in Developing Mobile Commerce and Its Application

คำศัพท์สำคัญ(ภาษาไทย): โมบายคอมเมิร์ซ, โมบายเว็บแอปพลิเคชัน, การสั่งอาหาร

คำศัพท์สำคัญ(ภาษาอังกฤษ): Mobile Commerce, Mobile Web Application

บทคัดย่อ

ในปัจจุบันแนวโน้มของการใช้โทรศัพท์เคลื่อนที่นั้นเพิ่มสูงขึ้น จากกล่าวได้ว่ามีน้อยคนที่จะไม่มีโทรศัพท์เคลื่อนที่ที่เป็นของตัวเอง เมื่อโทรศัพท์เคลื่อนที่เข้ามาเป็นส่วนหนึ่งของชีวิตประจำวัน และมีความสามารถมากกว่าการสื่อสารระหว่างคนสองคน (โทรออกและรับสาย) เช่น การเล่นเกม การถ่ายภาพ การถ่ายวิดีโอ การเข้าชมเว็บไซต์ หรือแม้แต่กระทั่งการทำงานที่เหมือนเครื่องคอมพิวเตอร์ ซึ่งมีโปรแกรมต่าง ๆ เช่น Microsoft Word, Microsoft Excel, Microsoft Power Point เป็นต้น ดังนั้นจึงมีการพัฒนาระบบต่าง ๆ ให้รองรับกับความสามารถของโทรศัพท์เคลื่อนที่ เพื่อให้ผู้ใช้งานเกิดความสะดวกสบาย การดำเนินงานคล่องตัวมากขึ้น ประกอบกับการที่สื่ออินเทอร์เน็ตได้เข้ามามีบทบาทในการประกอบธุรกิจในยุคปัจจุบัน ไม่ว่าจะเป็นการซื้อขายสินค้าหรือบริการ การติดต่อสื่อสารระหว่างกัน การค้นข้อมูลจากแหล่งข้อมูลต่าง ๆ ทั่วโลก ด้วยเหตุนี้ผู้ประกอบการจึงได้หันมาให้ความสนใจช่องทางใหม่ทางธุรกิจนี้ ส่งผลให้เกิดการทำธุรกรรมผ่านทางโทรศัพท์มือถือ หรือที่เรียกว่า โมบายคอมเมิร์ซ (Mobile Commerce) แต่เนื่องจากการพัฒนาระบบโมบายคอมเมิร์ซบนโทรศัพท์เคลื่อนที่นั้นยังคงมีข้อจำกัดมากมาย ทำให้ผู้ประกอบการธุรกิจหลายรายไม่สามารถใช้ประโยชน์จากช่องทางนี้ได้อย่างเต็มที่เท่าที่ควร ดังนั้นผู้จัดทำ

จึงได้ทำการศึกษา รวบรวม และประมวลแนวปฏิบัติที่ดีในการพัฒนาโมบายคอมเมอร์ซึ้นเพื่อให้สามารถนำไปประยุกต์ใช้ให้เกิดประโยชน์อย่างมีประสิทธิภาพมากที่สุด

โครงการนี้ครอบคลุมการศึกษาและทำความเข้าใจเกี่ยวกับหลักการทํางาน ประโยชน์ และข้อจํากัดของระบบโมบายคอมเมอร์ซึ (Mobile Commerce) และประมวลออกมาเป็นแนวปฏิบัติที่ดีในการพัฒนาโมบายคอมเมอร์ซึ (Mobile Commerce) รวมไปถึงศึกษาการเขียนเว็บแอปพลิเคชันและโมบายเว็บแอปพลิเคชันโดยใช้โปรแกรม Microsoft Visual Studio 2008 ร่วมกับ ASP.NET และภาษา C# เพื่อนำไปประยุกต์ใช้ในการพัฒนาระบบโมบายคอมเมอร์ซึ (Mobile Commerce) ตัวอย่าง นั่นคือ ระบบการสั่งซื้ออาหารที่โรงอาหาร อาคารมหิตลาธิเบศร คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ซึ่งโครงการนี้ได้พัฒนาทั้งในส่วนของเว็บแอปพลิเคชันและโมบายเว็บแอปพลิเคชัน สำหรับในส่วนของเว็บแอปพลิเคชันนั้นมีฟังก์ชันการทํางานแบ่งเป็น 3 ส่วน ได้แก่ ส่วนของร้านค้า สามารถตั้งคํารายการอาหาร รับรายการสั่งซื้ออาหาร สรุปยอดการขาย ส่วนของนิสิต อาจารย์ และบุคลากรทั่วไป สามารถสั่งซื้ออาหาร และยกเลิกการสั่งซื้อได้ และส่วนของผู้จัดการระบบ สามารถตั้งคํ่า และเรียกดูยอดรวมผู้ใช้ระบบได้ สำหรับในส่วนของโมบายเว็บแอปพลิเคชันมีฟังก์ชันกานทํางานได้แก่ การสั่งซื้ออาหาร และการเรียกดูรายการอาหารที่สั่ง นอกจากนี้ หากผู้ใดมีความสนใจในโครงการนี้สามารถนำไปพัฒนาและปรับปรุงแก้ไขเพิ่มเติมเพื่อให้มีประสิทธิภาพมากยิ่งขึ้นได้

วิชา 2603496 โครงการเทคโนโลยีสารสนเทศเพื่อธุรกิจ

บทสรุปย่อรายงาน

ปีการศึกษา 2553

1. รหัสประจำตัว : 5042058826
2. ชื่อ นิสิต : นางสาว รัตนาพร เตชะวิริยะจินดา
3. ชื่ออาจารย์ที่ปรึกษา : อาจารย์ ดร.บุรุษย์ ภัทรโกศล
4. หัวข้อโครงการ (ไทย) การสร้างโปรแกรมประยุกต์ที่ช่วยสร้างแบบสอบถามออนไลน์
5. หัวข้อโครงการ (อังกฤษ) : An Online Questionnaire Building Application
6. คำศัพท์สำคัญ :

1. ลิเคิร์ต สเกล (Likert Scale)

เป็นแบบทดสอบที่ใช้ในการประเมินค่าทัศนคติประเภทหนึ่ง ซึ่งผู้ที่เสนอวิธีวัดดังกล่าวนี้คือ Likert โดยผู้พัฒนาแบบสอบถามต้องเสนอข้อความแสดงทัศนคติต่อสิ่งนั้น ๆ หลาย ๆ ข้อความ แล้วให้ผู้รับการทดสอบประเมินค่าแต่ละข้อความ ถือเป็น 1 มาตร แล้วจึงนำคะแนนจากมาตรต่าง ๆ มารวมเป็นคะแนนรวม และยึดคะแนนนี้เป็นหลักในการตีความ นอกจากนี้เนื่องจากถือว่าข้อความต่าง ๆ วัดจากทัศนคติต่อสิ่งเดียวกัน ดังนั้นจึงควรมีข้อความหลายข้อความเพื่อให้ข้อความมีความเที่ยงและน่าเชื่อถือมากขึ้น

2. มัลติเพิล ออร์เทอร์เนทีฟ (Multiple Alternatives)

เป็นแบบสอบถามลักษณะเดียวกับมัลติเพิล ชอยซ์ กล่าวคือมีการกำหนดทางเลือกให้หลายทางเลือก ภายใต้คำถามหนึ่งคำถาม เพื่อให้ทดสอบหรือวัดผลตามเจตนาของผู้พัฒนา ซึ่งแบบสอบถามประเภทนี้เป็นที่นิยมเนื่องจากเป็นที่คุ้นเคยสำหรับบุคคลทั่วไปและเข้าใจง่าย

3. จาวาสคริปต์ (JavaScript)

คือภาษาสคริปต์ใหม่ที่ถูกพัฒนาขึ้นโดย เน็ตสเคปคอมมิวนิเคชันส์ (Netscape Communications Corporation) สามารถทำให้การสร้างเว็บเพจมีลูกเล่นต่าง ๆ มากมาย และยังสามารถโต้ตอบกับผู้ใช้ได้อย่างทันที เช่น การใช้เมาส์คลิกหรือ การกรอกข้อความในฟอร์ม เป็นต้น

4. ซีเอสเอส (CSS)

คือชุดคำสั่งที่ใช้สำหรับการกำหนดการแสดงผลข้อมูลหน้าเว็บเพจ ซึ่งคำเต็ม ๆ ของ CSS คือ Cascading Style Sheets ใช้ในการตกแต่งเว็บเพจโดยเฉพาะ

5. พีเอชพี (PHP)

พีเอชพี (PHP) คือ ภาษาสคริปต์เว็บเพจในลักษณะเซิร์ฟเวอร์-ไซด์ สคริปต์ โดยเป็นภาษาที่ทำให้โปรแกรมประยุกต์สามารถตอบสนองการใช้งานของผู้ใช้ได้อย่างรวดเร็ว และยังมีความสามารถในการจัดการข้อมูลที่มีความซับซ้อนได้ ทำให้โปรแกรมประยุกต์ลักษณะเป็น ไดนามิก เพจ (Dynamic Page)

7. บทคัดย่อ :

ในปัจจุบันแบบสอบถามออนไลน์ได้เข้ามามีบทบาทแทนที่แบบสอบถามแบบดั้งเดิม และถูกนำมาใช้อย่างแพร่หลายเพื่อเก็บและศึกษาข้อมูลที่น่าสนใจ ดังนั้นเพื่อให้ได้โปรแกรมประยุกต์ที่สามารถช่วยสร้างแบบสอบถามและจัดการแบบสอบถามได้ ผู้พัฒนาจึงเลือกพัฒนาโปรแกรมประยุกต์นี้ขึ้น โดยเลือกพัฒนาการสร้างและจัดการแบบสอบถามเฉพาะแบบสอบถามประเภท Likert Scale และ Multiple Alternatives เนื่องจากแบบสอบถามทั้งสองประเภทมีลักษณะการใช้ที่เข้าใจง่ายและเป็นที่ยอมรับใช้อย่างแพร่หลาย

ผู้ใช้สามารถใช้โปรแกรมประยุกต์นี้เพื่อสร้างแบบสอบถามทั้งสองประเภท โดยในแบบสอบถามประเภท Likert Scale เมื่อผู้ใช้กรอกรายละเอียดแบบสอบถาม ระบบจะสร้างแบบสอบถามที่มีคำถาม พร้อมทั้งแสดงระดับคำตอบ 5 ระดับคือ ตั้งแต่เห็นด้วยน้อยที่สุดไปจนถึงเห็นด้วยมากที่สุด ส่วนในแบบสอบถามประเภท Multiple Alternatives ระบบจะแสดงคำถามพร้อมตัวเลือกต่างๆ นอกจากนี้ผู้ใช้ยังสามารถเพื่อคุณค่าให้แบบสอบถามทั้งสองแบบ กล่าวคือสามารถเลือกรูปแบบสีที่จะใช้ในระดับคำตอบแต่ละระดับของแบบสอบถามประเภท Likert Scale ได้ และสามารถเลือกที่จะใส่รูปภาพประกอบตัวเลือกในแบบสอบถามประเภท Multiple Alternatives ได้ ซึ่งเมื่อสร้างแบบสอบถามเสร็จแล้วผู้ใช้สามารถจัดการกับแบบสอบถามได้ โดยสามารถค้นหา แก้ไข ลบ และนำออกข้อมูลผลสำรวจแบบสอบถามในรูปแบบไฟล์ .csv (Comma Separated Value File) โดยการพัฒนาหน้าที่การทำงานดังกล่าวได้อาศัยภาษาพีเอชพี(PHP)และจาวาสคริปต์(Javascript) เป็นหลัก

ระบบจะแสดงผลผลสำรวจแบบสอบถามด้วยวิธีทางสถิติอย่างง่าย คือแสดงยอดรวมของคำตอบแต่ละข้อ และแสดงสัดส่วนเปรียบเทียบระหว่างคำตอบในรูปแบบเปอร์เซ็นต์ นอกจากนี้ยังสามารถนำออกข้อมูลดังกล่าว ซึ่งผู้ใช้สามารถนำข้อมูลผลสำรวจที่ได้ไปวิเคราะห์ทางสถิติขั้นสูงต่อไปเพื่อใช้ในการศึกษาและวิจัย



รูปที่ ก.5 ตัวอย่างเอกสารบทคัดย่อโครงการที่จัดทำในปี 2554 (หน้าที่หนึ่ง)

บทคัดย่อ

ความสำคัญของการแสดงความคิดเห็น คือ การเปิดโอกาสให้บุคคลอื่นได้แสดงความคิดเห็น ข้อเสนอแนะกับเรื่องต่างๆ เช่น บริษัทแห่งหนึ่งจะพัฒนาระบบรับฟังความคิดเห็นจากพนักงาน ในการปรับปรุงการทำงานในรูปแบบต่างๆ ซึ่งจะเห็นว่าเป็นแนวคิดในการพัฒนาองค์กรที่ดี นอกจากนี้บางสถานการณ์ผู้บริหารอาจจะมีโจทย์ปัญหาบางประการที่หาทางออกได้ลำบาก ผู้บริหารก็สามารถใช้กระบวนการรับฟังความคิดเห็นดังกล่าวเปิดโอกาสให้พนักงานได้แสดงความคิดเห็นหรือหาทางแก้ปัญหาได้ ซึ่งเป็นการเปิดรับทางเลือกที่หลากหลายและอาจจะทางเลือกเป็นประโยชน์ได้นั่นเอง

การแสดงความคิดเห็นบนอินเทอร์เน็ตก็มีความหมายเช่นเดียวกับการแสดงความคิดเห็นทั่วไป แต่แตกต่างกันที่ช่องทางในการแสดงความคิดเห็นนั้นอยู่บนอินเทอร์เน็ตนั่นเอง โครงการการพัฒนาโปรแกรมกล่องคำเพื่อการแสดงความคิดเห็นบนอินเทอร์เน็ตจึงจัดทำขึ้นเพื่อพัฒนาโปรแกรมที่สามารถเพิ่มช่องทางการรับการแสดงความคิดเห็นที่สะดวก รวดเร็ว และสามารถเข้าถึงผู้แสดงความคิดเห็นได้มากขึ้นเพราะสามารถเข้าถึงกล่องรับความคิดเห็นได้ทุกที่ ทุกเวลา ที่สามารถใช้อินเทอร์เน็ตได้ อีกทั้งยังมีโปรแกรมยังมีการจัดการข้อมูลที่ดีขึ้นกว่ารูปแบบเดิม เพราะสามารถเก็บข้อมูลในรูปแบบอิเล็กทรอนิกส์ ที่สามารถนำไปใช้ประโยชน์ได้ทันที และเป็นการประหยัดค่าใช้จ่ายมากกว่าการใช้กล่องที่ใช้การรับกระดาษความคิดเห็นนั่นเอง

โดยจุดเด่นของโปรแกรมกล่องคำเพื่อการแสดงความคิดเห็นบนอินเทอร์เน็ต คือ ผู้ร่วมแสดงความคิดเห็นสามารถเลือกที่จะแสดงข้อมูลของตนเองต่อเจ้าของกล่องหรือไม่แสดงตัวตนของตนเองต่อเจ้าของกล่องรับความคิดเห็นได้ ทำให้ผู้ร่วมแสดงความคิดเห็นสามารถแสดงความคิดเห็นของตนเองเข้าสู่กล่องรับความคิดเห็นได้อย่างเต็มที่ ซึ่งถือเป็นการเปิดรับความคิดเห็นที่กว้างขึ้น มีความปลอดภัยและมีความน่าเชื่อถือได้ ว่ามีเพียงเจ้าของรับความคิดเห็นเท่านั้นที่รับรู้ข้อมูลการแสดงความคิดเห็นในกล่องรับความคิดเห็นนั้นๆ

โดยผู้พัฒนาได้พัฒนาโปรแกรมกล่องคำเพื่อการแสดงความคิดเห็นบนอินเทอร์เน็ตในรูปแบบเว็บแอปพลิเคชันด้วยภาษา PHP, HTML, CSS, JavaScript และมีการใช้เครื่องมือในการพัฒนาโปรแกรม คือ Dreamweaver CS3, phpMyAdmin และ winSCP ซึ่งผู้พัฒนาได้รับความรู้จากการใช้เครื่องมือและภาษาคอมพิวเตอร์ตามวัตถุประสงค์ อีกทั้งยังได้โปรแกรมที่เป็นประโยชน์ต่อผู้ใช้ซึ่งสามารถใช้งานได้จริงบนอินเทอร์เน็ต อันเป็นประโยชน์ต่อโลกธุรกิจและสังคมในปัจจุบัน

ตารางที่ ก.1 แสดงรายการเอกสารที่นำมาใช้ในงานวิจัย

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
52B01	การพัฒนาเว็บไซต์เพื่อส่งเสริมธุรกิจการผลิต เครื่องใช้ไฟฟ้าอิเล็กทรอนิกส์	The Development of a Website to Promote Electronic Products Manufacturing Business
52B02	การศึกษาและประยุกต์ใช้ไมโครซอฟต์เอสคิว แอลเซิร์ฟเวอร์ 2008 อินทิเกรชันเซอร์วิสเซส	A Study and Applications of Microsoft SQL Server 2008 Integration Services
52B03	การศึกษาแนวปฏิบัติที่ดีในการพัฒนาโมบาย คอมเมิร์ซและตัวอย่างการนำไปใช้	A Study of Good Practices in Developing Mobile Commerce and Its Application
52B04	ระบบการสั่งอาหารผ่านโทรศัพท์มือถือใน ร้านอาหาร (กรณีศึกษา:ร้านอาหารสีฟ้า)	Ordering System by Cell Phone in the Restaurant (Case Study: ร้านอาหารสีฟ้า)
52B05	การพัฒนาระบบงานแบบเว็บโดยใช้ เอเอส พีดอทเน็ต 3.5 และเอแจ็ก	การพัฒนาระบบงานแบบเว็บ, ASP.NET 3.5, AJAX ,เอเอสพีดอทเน็ต
52B06	การพัฒนาระบบติดตามยานพาหนะ	The Development of Vehicle Tracking System
52B07	การพัฒนาเว็บไซต์สังคมด้วยซอฟต์แวร์ต้น รหัสเปิดเอลก์	Developing Social Website by Using Elgg Opensource Software
52B08	การมีส่วนร่วมของผู้ซื้อในการปรับแต่งสินค้า กับการประยุกต์ใช้ในเว็บแอปพลิเคชัน	Applying Product Customization in Web Application
52B09	การศึกษาโปรแกรม Microsoft Photosynth และการประยุกต์ใช้กับพาณิชย์- อิเล็กทรอนิกส์	The Study of Microsoft Photosynth and Its applications in e-Commerce
52B10	การพัฒนาเว็บแอปพลิเคชันที่ปลอดภัย โดย ใช้เครื่องมือ OWASP	The Development of Secure Web Application by Using OWASP Tools
52B11	การพัฒนาระบบสารสนเทศโดยใช้ ไมโครซอฟท์ลิงค์	Developing an Information System by Using Microsoft LINQ
52B12	ระบบสารสนเทศเพื่อการบริหารจัดการการ จัดสรรทุนอุดหนุนการศึกษานิสิต	Information System for Student Funding Management

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
52B13	การศึกษาและประยุกต์ใช้ Social Network Application	The Study of Social Network Application and Its Application
52B14	การพัฒนาโปรแกรมเพื่อแสดงเส้นทางที่ดีที่สุดในการเดินทาง	The Development of Best Traveling Path Visualization Program
52B15	การศึกษาการทำงานของ Spring Framework ในการทำงานร่วมกับ JDBC	The study of Spring Framework in collaboration of JDBC
52B16	การศึกษาและประยุกต์ใช้โอเพนเอ็มพี	The Study of OpenMP and Its Application
52B17	การศึกษาและเปรียบเทียบระบบจัดการฐานข้อมูลเอ็กซ์เอ็มแอล	A Study and Comparison of XML Database Management System
52B18	การพัฒนาโปรแกรมแสดงความสัมพันธ์ระหว่างปริมาณการจราจรกับการแข่งขันกีฬา	The Development of Relationship Visualization Software: Traffic and Sport Event
52B20	การประยุกต์ใช้การนำเที่ยวเสมือนจริงสื่อผสมในเว็บแอปพลิเคชันกรณีศึกษา จตุรัสจามจุรี	Applying Multimedia Virtual Tour in Web Application: A Case Study of the Chamchuri Square
52B21	ผลของตัวแปรด้านปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ต่อการตอบกลับแบบสอบถามออนไลน์	Impacts of human-computer-interaction variables on online questionnaires' response rates
52B22	การศึกษาและประยุกต์ใช้ไมโครซอฟท์เอเอสพีดีเอ็นเน็ตเอ็มวีซี	The Study of Microsoft ASP.NET MVC and Its Application
52B23	การพัฒนาโปรแกรมเพื่อแสดงความสัมพันธ์ระหว่างปริมาณการจราจรและมลพิษทางอากาศ	The development of relationship visualization software: traffic and air pollution
52B25	ระบบสนับสนุนเพื่อเรียกค่าสินไหมทดแทนรถยนต์ ณ จุดเกิดเหตุด้วยเทคโนโลยี PDA และ GPS	Car Claiming Support System by Using PDA And GPS Technologies
52B26	การศึกษาและประยุกต์ใช้ไมโครซอฟท์	The Study of Microsoft Windows

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
	วินโดวส์เพาเวอร์เชลล์	PowerShell and Its Application
52B27	การศึกษาโปรแกรม Google SketchUp และการประยุกต์ใช้เพื่อสร้างโมเดล 3 มิติ สำหรับการท่องเที่ยวเสมือนจริง	The Study of Google SketchUp and its Applications in Producing 3D Models for Virtual Tour
52B28	การศึกษามาตรฐานคอมมอนแวร์เฮาส์เมตะโมเดลและการนำไปใช้	A Study of Common Warehouse Metamodel Specification and Its Uses
52B30	การนำเสนอสินค้าแบบเสมือนจริงและการประยุกต์ใช้ในพาณิชย์อิเล็กทรอนิกส์	Product Visualization and Applications in e- Commerce
52B31	การพัฒนาระบบบริหารจัดการรถโดยสารภายในจุฬาลงกรณ์มหาวิทยาลัย	Developing Chulalongkorn University Campus Bus Management System
52B32	การศึกษาและประยุกต์ใช้เค้กพีเอชพีเฟรมเวิร์ก	The Study of CakePHP Framework and Its Application
52B33	เว็บแอปพลิเคชันสำหรับการสร้างประวัติส่วนตัวและแฟ้มผลงานแบบออนไลน์	web application for port-folio and resume online
52B34	การพัฒนาโปรแกรมแสดงความสัมพันธ์ระหว่างการจราจรและการจัดงานแสดงต่างๆ	The development of relationship visualization program: traffic and exhibition events
52B35	ผลของการใช้เสียงในบริบทพาณิชย์อิเล็กทรอนิกส์	Impacts of voice in electronic commerce contexts
53B01	ระบบติดตามการบำรุงรักษาระบบสารสนเทศพัฒนาโดยผู้ให้บริการภายนอก	Information System For Outsourced System Maintenance
53B02	ระบบสารสนเทศสำหรับธุรกิจข้าวกล่อง	Information System for Made-to-order Lunch Box Business
53B03	ระบบสารสนเทศเพื่อการจัดการสำหรับธุรกิจแพปลา	Information system for wholesaler business in fish market
53B04	การศึกษาและประยุกต์ใช้ไมโครซอฟท์แชร์พอยท์ 2010	A Study and Applications of Microsoft SharePoint 2010
53B06	การศึกษาและพัฒนาพจนานุกรมไทยที่ใช้	The Study and Development of

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
	จินตทัศน์	Visualized Thai Dictionary
53B07	การศึกษาและประยุกต์ภาษา AIML	The study of AIML and its applications
53B08	แอปพลิเคชันแสดงจุดน่าสนใจเพื่อการท่องเที่ยวกรณีศึกษาที่เยาวราช	A point of interest visualization application for tourism: A case study at China town
53B09	การศึกษาและการนำไปประยุกต์ใช้งานของระบบตรวจจับการเคลื่อนไหว :โมชันอนาลิซิส	A study and Application of Motion Capture System: MotionAnalysis
53B10	การศึกษาและประยุกต์ใช้นิติคอมพิวเตอร์	The Study and Application of Computer Forensics
53B11	การศึกษาและประยุกต์ใช้เครื่องมือการสร้างโมเดลและวิเคราะห์กระบวนการธุรกิจ: เว็บสเฟียร์ บีซีเนส โมเดลเลอร์ แอดvanced เวอร์ชัน	A study and Application of Business Process Modeling and Analysis Tool: WebSphere Business Modeler Advanced V7
53B12	การศึกษาและประยุกต์เครื่องมือพัฒนาโปรแกรมแบบจินตทัศน์	The Study and Application of Visual Programming Language Tools
53B13	การศึกษาการสร้างหนังสืออิเล็กทรอนิกส์	The Study of E-book Creation
53B14	การศึกษาและประยุกต์ใช้ระบบจัดการเนื้อหาสำหรับจัดการองค์ความรู้	A Study and Applications of Drupal Content Management System for Knowledge Management
53B15	การศึกษาและการประยุกต์ใช้ซอฟต์แวร์สำหรับการตกแต่งบ้าน	The study of Home Interior Design Software and Its Applications
53B16	การศึกษาและเปรียบเทียบซอฟต์แวร์การบริหารลูกค้าสัมพันธ์	A Study and Comparison of Customer Relationship Management Software
53B17	ระบบการจัดการขนส่งอ้อยอัจฉริยะ	Smart Sugarcane Distribution System
53B18	การสร้างโปรแกรมประยุกต์ที่ช่วยสร้างแบบสอบถามออนไลน์	An Online Questionnaire Building Application
53B19	การทำแผนผังด้วย AR	Floor plan construction with AR

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ) (Augmented Reality)
53B20	แอปพลิเคชันที่ช่วยในการเปรียบเทียบ ความสามารถทางการเงินของบริษัท และอุตสาหกรรม	An industry and organization financial performance comparison application
53B21	Mash up application เพื่อสนับสนุนการ ค้นหารายละเอียดของคำบนหน้าเว็บเพจ	A mash up application for supporting information searching of “words” on the web page
53B22	ระบบสนับสนุนการตัดสินใจคัดเลือกผู้จัดหา เม็ดมะขามสำหรับธุรกิจการผลิตและส่งออก แป้งจากเม็ดมะขาม	Decision Support System for Selecting Tamarind Seed Suppliers for Production and Export Business of Tamarind Kernel Powder
53B23	ระบบการเลือกตั้งอิเล็กทรอนิกส์สำหรับการ เลือกคณะกรรมการองค์การบริหารสโมสร นิสิต จุฬาลงกรณ์มหาวิทยาลัย	Electronic Election System for Electing Committee of SGCU
53B24	การศึกษาและประยุกต์กระดานอัจฉริยะโดย ใช้วีรีโมท	The study and application of interactive board using Wii Remote
53B25	ระบบควบคุมการเล่นเพลงโดยใช้เสียง	System Control Song Playing by Voice
53B26	แอปพลิเคชันแสดงภาพประวัติศาสตร์ของจุด น่าสนใจ	Historical pictorial application for point of interest
53B27	การพัฒนาเกมที่ใช้การตรวจจับการ เคลื่อนไหวด้วยเว็บแคม	Game Development using Webcam Motion Detection
53B28	การศึกษาและพัฒนาระบบธุรกิจอัจฉริยะ โดยใช้ไมโครซอฟท์เพอร์ฟอร์แมนซ์พอยท์ เซอร์วิส 2010	A Study and Development of Business Intelligence System by Using Microsoft PerformancePoint Services 2010
53B29	ระบบควบคุมสินค้าคงคลังของธุรกิจขาย เครื่องประดับ	Inventory control system for selling jewelry business
53B30	การศึกษาและพัฒนาระบบธุรกิจอัจฉริยะ โดยใช้ค็อกนอสเอ็กเพรส	A Study and Development of Business Intelligence System by Using Cognos

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
		Express
53B31	การศึกษาและประยุกต์ใช้โปรแกรม ไมโครซอฟท์เอ็กเพรสชันสตูดิโอ 4	A Study and Application of Microsoft Expression Studio 4
53B37	การศึกษาและประยุกต์ใช้เครื่องมือวิศวกรรม ซอฟต์แวร์ใช้คอมพิวเตอร์ช่วย:พาวเวอร์ ดีไซน์เนอร์ 15	A Study and Applications of a Computer Aided Software Engineering Tool: PowerDesigner 15
54B01	การพัฒนาโปรแกรมกล่องดำเพื่อการแสดง ความคิดเห็นบนอินเทอร์เน็ต	The Development of Opinion Black Box on the Internet
54B02	ซอฟต์แวร์ช่วยสนับสนุนการแต่งทำนองเพลง จากเนื้อร้อง	Software for Composing Song's Melody from Lyrics
54B03	การศึกษาและการพัฒนาซอฟต์แวร์โดยใช้ วิธีการแบบอจาายล์	A Study and Development of Software using Agile Methodology
54B04	การใช้ HTML5 สำหรับการพัฒนาเว็บ แอปพลิเคชัน	Using HTML5 for Web Application Development
54B05	เว็บแอปพลิเคชันค้นหาเส้นทางที่เดินทาง ด้วย MRT และ BTS	Web Application For Searching Travelling Route By MRT And BTS
54B06	โปรแกรมอ่าน RSS Feed บนแอนดรอยด์ แพลตฟอร์ม	RSS Feed Reader on Android Platform
54B07	การประยุกต์ใช้สโนอร์ท	The Application of Snort
54B08	การพัฒนาเกมส์แอปพลิเคชันสำหรับไอโฟน	Game Application Development for iPhone
54B09	ซอฟต์แวร์ช่วยออกแบบการจัดพานพุ่ม ดอกไม้	Software for Designing Flower Bouquet (Phaan-Poom)
54B10	การพัฒนาเว็บสำหรับธุรกิจการขายคอนแทค เลนส์	Development of Web Application for contact lens
54B11	เว็บแอปพลิเคชันขายของเล่นเด็กโดย พิจารณาทักษะที่เหมาะสมสำหรับเด็ก	Web Application for Selling toys suitable for Children Development
54B12	การพัฒนาซอฟต์แวร์บนไมโครซอฟท์	The Software Development on Microsoft

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
	วินโดวส์โฟนเซเวน	Windows Phone 7
54B13	กูเกิลแอฟสคริปต์	Google Apps Script
54B14	การพัฒนาไอโฟนแอฟพลิเคชันสำหรับการแสดงส่วนผสมอาหารในซูเปอร์มาร์เก็ต	iPhone Application Development for Displaying Food Ingredient in Supermarket
54B15	การพัฒนาซอฟต์แวร์ที่เป็นเครื่องมือสำหรับช่วยในงานตรวจสอบเทคโนโลยีสารสนเทศ	A Development of a Software Tool Aided in IT Audit
54B16	การประยุกต์ใช้ไมโครซอฟท์ซอลเวอร์ฟาวนด์ชัน	The Application of Microsoft Solver Foundation
54B17	การสร้างวารสารดิจิทัลสำหรับอุปกรณ์แท็บเล็ตโดยใช้อะโดบีอินดีไซน์	Digital Magazine Creation for Tablet Devices Using Adobe InDesign
54B18	ระบบสนับสนุนการตรวจสอบการเข้าและออกอาคารจอดรถโดยการประยุกต์ใช้เทคนิคการรู้จำป้ายทะเบียนแบบอัตโนมัติ	System for Supporting Car Park Control Using Automatic Licence Plate Recognition Technique
54B19	การศึกษาเกี่ยวกับตำแหน่งงานเทคโนโลยีสารสนเทศในองค์กรของประเทศไทย	A Study on Information Technology Jobs in Thailand's Organizations
54B20	ซอฟต์แวร์ช่วยออกแบบสติ๊กเกอร์บนบรรจุภัณฑ์จากภาพจริง	Packaging Sticker Design Support Software Based on Real Image
54B21	การประยุกต์ใช้ไมโครซอฟท์เอนเทอร์ไพรส์ไลบรารี	The Application of Microsoft Enterprise Library
54B22	ตัวแปรที่มีผลต่อการอ่านกราฟเส้นบนจอภาพ	Variables Affecting On-screen Line Graph Visualization
54B23	โปรแกรม Android ให้ข้อมูลสภาพการจราจรในกรุงเทพมหานคร	A Bangkok Traffic Report Application based on Android Technology
54B24	ศูนย์ให้บริการเลือกรถพยาบาลออนไลน์ : ใช้ Traffy ในการจัดเส้นทาง	Ambulance Selecting Service Center Online : Using Traffy for Routing
54B25	เว็บแอฟพลิเคชันสำหรับการจัดส่งการ์ดอวยพรอิเล็กทรอนิกส์	Web Application Development for e-cards

รหัสโครงการ	ชื่อโครงการ (ภาษาไทย)	ชื่อโครงการ (ภาษาอังกฤษ)
54B27	โปรแกรมการสื่อสารหน้าจอบนแอนดรอยด์แพลตฟอร์ม	The screen communication application on Android platform
54B28	การเปรียบเทียบเครื่องมือสำหรับการพัฒนาโปรแกรมประยุกต์บนระบบปฏิบัติการแอนดรอยด์	A Comparison of Development Tools for Android Application
54B29	การเปรียบเทียบคุณลักษณะด้านการตรวจสอบของระบบจัดการฐานข้อมูล	A Comparison of Audit Features in Database Management Systems
54B30	การพัฒนากระบวนการแบบกระแสดำเนินงานโดยใช้เจอบอสเจบีพีเอ็ม 5.1	A Development of Workflow Application Using JBoss jBPM 5.1
54B31	การพัฒนาแอปพลิเคชันโดยใช้เครื่องมือบนโมบายล์ซอฟต์แวร์แพลตฟอร์ม	Application Development Using Mobile Software Platform Tools
54B32	การศึกษาเกี่ยวกับประกาศนียบัตรวิชาชีพด้านเทคโนโลยีสารสนเทศในองค์กรของประเทศไทย	A Study on IT Certifications in Thailand's Organizations
54B33	การศึกษาเกี่ยวกับผู้บริหารเทคโนโลยีสารสนเทศระดับสูงในองค์กรของประเทศไทย	A Study on Chief Information Officers in Thailand's Organizations
54B34	เว็บแอปพลิเคชันสำหรับร้านขายดอกไม้	Web Application for a Flower Shop
54B35	การพัฒนาซอฟต์แวร์ที่เป็นเครื่องมือสำหรับการทำการประเมินซีเอ็มเอ็มไอ	A Development of Software Tool for CMMI Appraisal
54B36	ตัวแปรที่มีผลต่อการอ่านตารางบนจอภาพ	Variables affecting on-screen table visualization
54B37	การพัฒนาแอปพลิเคชันบนไอโฟนสำหรับการทำการตลาดเฉพาะบุคคลของธุรกิจค้าปลีก	iPhone Application Development for Personalized Marketing of Retail Business
54B38	การศึกษาความรู้และการปฏิบัติของกลุ่มผู้ใช้ที่ไม่ใช่พนักงานเทคโนโลยีสารสนเทศเกี่ยวกับความปลอดภัยของเทคโนโลยีสารสนเทศ	A Study of Knowledge and Practices of Non-IT People Towards Information Technology Security

ภาคผนวก ข รายการคำหยุดภาษาไทย (Thai Stop-words List)

ผู้วิจัยได้นำรายการคำหยุดภาษาไทย จากวรวิทย์ เกษร (2553) มาใช้ในการวิเคราะห์คำในรายการเอกสารทั้งหมด ซึ่งมีดังต่อไปนี้

ก็	ก็คือ	ก็แค่	ก็จะ
ก็ดี	ก็ได้	ก็ต่อเมื่อ	ก็ตาม
ก็ตามแต่	ก็ตามที	ก็แล้วแต่	กระทั่ง
กระทำ	กระนั้น	กระผม	กลับ
กล่าว	กล่าวคือ	กลุ่ม	กลุ่มก้อน
กลุ่มๆ	กว่า	กว้าง	กว้างขวาง
กว้างๆ	ก่อน	ก่อนหน้า	ก่อนหน้านี้
ก่อนๆ	กัน	กันดีกว่า	กันดีไหม
กันเถอะ	กันนะ	กันและกัน	กันไหม
กันเอง	กับ	การ	กำลัง
กำลังจะ	กำหนด	กฎ	เก็บ
เกิด	เกิน	เกินๆ	เกี่ยวกัน
เกี่ยวกับ	เกี่ยวข้อ	เกี่ยวเนื่อง	เกี่ยวๆ
เกือบ	เกือบจะ	เกือบๆ	แก
แก่	แก่ใจ	ใกล้	ใกล้ๆ
ไกล	ไกลๆ	ขณะ	ขณะเดียวกัน
ขณะใด	ขณะใดๆ	ขณะที่	ขณะนั้น
ขณะนี้	ขณะนี้	ขวาง	ขวางๆ

ขอ	ของ	ขึ้น	เข้า
ข้าง	ข้างเคียง	ข้างต้น	ข้างบน
ข้างล่าง	ข้างๆ	ขาด	เข้าพเจ้า
เข้าๆ	ขึ้น	เขา	เข้า
เข้าใจ	เขียน	คง	คงจะ
คงอยู่	ครบ	ครบครัน	ครบถ้วน
ครั้ง	ครั้งกระนั้น	ครั้งก่อน	ครั้งครา
ครั้งคราว	ครั้งใด	ครั้งที่	ครั้งนั้น
ครั้งนี้	ครั้งละ	ครั้งหนึ่ง	ครั้งหลัง
ครั้งคราว	ครั้งใด	ครั้งที่	ครั้งนั้น
ครั้งนี้	ครั้งละ	ครั้งหนึ่ง	ครั้งหลัง
ครั้งหลังสุด	ครั้งไหน	ครั้งๆ	ครั้น
ครึบ	ครา	คราใด	คราที่
ครานั้น	ครานี้	คราหนึ่ง	คราไหน
คราว	คราวก่อน	คราวใด	คราวที่
คราวนั้น	คราวนี้	คราวโน้น	คราวละ
คราวหน้า	คราวหนึ่ง	คราวหลัง	คราวไหน
คราวๆ	คล้าย	คล้ายกัน	คล้ายกันกับ
คล้ายกับ	คล้ายกับว่า	คล้ายว่า	ควร
ความ	ค่อนข้าง	ค่อนข้าง	ค่อนข้างจะ
ค่อยไปทาง	ค่อนข้างมาทาง	ค่อย	ค่อยๆ
คะ	คะ	คำ	คิด
คิดว่า	คือ	คุณ	คุณๆ

เคย	เคยๆ	แค่	แค่จะ
แค่นั้น	แค่นี้	แค่เพียง	แค่ว่า
แค่นั้น	ใคร	ใคร	ใครจะ
ง่าย	ง่ายๆ	ไง	จง
จด	จนกระทั่ง	จนกว่า	จนขณะนี้
จนตลอด	จนถึง	จนทั่ว	จนบัดนี้
จนเมื่อ	จนเมื่อ	จนเมื่อ	จรด
จรดกับ	จริง	จริงจัง	จริงๆ
จริงๆ จังๆ	เช่นไร	จน	จนจะ
จนเจียน	จน	จนกับ	จนจน
จะ	จะ	จะ	จะได้
จัง	จังๆ	จัด	จัดการ
จัดงาน	จัดแจง	จัดตั้ง	จัดทำ
จัดหา	จัดให้	จับ	จ้ำ
จ้ำ	จาก	จากนั้น	จากนี้
จากนี้ไป	จำเป็น	จำเป็น	จำเป็น
จึง	จึงจะ	จึงเป็น	จึงๆ
ฉะนั้น	ฉะนั้น	ฉัน	เฉกเช่น
เฉพาะ	เฉย	เฉยๆ	ไหน
ช่วง	ช่วงก่อน	ช่วงต่อไป	ช่วงถัดไป
ช่วงท้าย	ช่วงที่	ช่วงนั้น	ช่วงนี้
ช่วงระหว่าง	ช่วงแรก	ช่วงหน้า	ช่วงหลัง
ช่วงๆ	ช่วย	ซ้ำ	ซ้ำนาน

ชาว	ข้าๆ	เช่น	เช่นก่อน
เช่นกัน	เช่นเคย	เช่นดัง	เช่นดังก่อน
เช่นดังเก่า	เช่นดังที่	เช่นดังว่า	เช่นเดียวกัน
เช่นเดียวกับ	เช่นใด	เช่นที่	เช่นที่เคย
เช่นที่ว่า	เช่นนั้น	เช่นนั้นเอง	เช่นนี้
เช่นเมื่อ	เมื่อ	เมื่อถือ	เมื่อมัน
เชื่อว่า	ใช่	ใช่ใหม่	ใช่
ซะ	ซะก่อน	ซะจน	ซะจนกระทั่ง
ซะจนถึง	ซึ่ง	ซึ่งก็	ซึ่งก็คือ
ซึ่งกัน	ซึ่งกันและกัน	ซึ่งได้แก่	ซึ่งๆ
ณ	ด้วย	ด้วยกัน	ด้วยเช่นกัน
ด้วยที่	ด้วยประการฉะนี้	ด้วยเพราะ	ด้วยว่า
ด้วยเหตุที่	ด้วยเหตุนั้น	ด้วยเหตุนี้	ด้วยเหตุเพราะ
ด้วยเหตุว่า	ด้วยเหมือนกัน	ดัง	ดัง
ดังกล่าว	ดังกับ	ดังกับ	ดังกับว่า
ดังกับว่า	ดังเก่า	ดังเก่า	ดังเคย
ดังเคย	ดังจะ	ดังจะ	ดังเช่น
ดังเช่น	ดังเช่นที่	ดังเช่นที่	ดังเดิม
ดังเดิม	ดังต่อไปนี้	ดังแต่ก่อน	ดังแต่ก่อน
ดังที่	ดังที่	ดังที่กล่าว	ดังที่เคย
ดังที่จะเป็น	ดังนั้น	ดังนี้	ดังนี้เช่น
ดังนี้เพราะ	ดังแม้	ดังแม้	ดังว่า
ดังว่า	ดังเหมือน	ดังเหมือน	ด้าน

ด้านๆ	ดำเนิน	ดำเนินการ	ดำเนินงาน
ดำเนินไป	ดิฉัน	ดี	ดีๆ
ดู	ดูจะ	ดูแล	ดูแล้ว
ดูว่า	ดูเหมือน	ดูเหมือนว่า	ดูๆ
เดิม	เดิมที	เดิมๆ	เดี๋ยว
เดี๋ยวกัน	เดี๋ยวกับ	เดี๋ยว	เดี๋ยวก่อน
เดี๋ยวนั้น	เดี๋ยวนี	แต่	โดย
โดยง่าย	โดยเฉพาะ	โดยเฉพาะอย่าง	โดยดี
โดยดุษฎี	โดยตลอด	โดยทั่ว	โดยทั่วกัน
โดยทั่วถึง	โดยทั่วไป	โดยที่	โดยแท้
โดยแท้จริง	โดยนัย	โดยปกติ	โดยมัก
โดยมักจะ	โดยมาก	โดยเมื่อ	โดยรวม
โดยรวมๆ	โดยเร็ว	โดยละม่อม	โดยลำดับ
โดยส่วนมาก	โดยส่วนรวม	โดยส่วนใหญ่	ใด
ใดๆ	ได้	ได้แก่	ได้แต่
ได้ที่	ได้มา	ได้รับ	ตน
ตนเอง	ตนๆ	ตรง	ตรงๆ
ตลอด	ตลอดกาล	ตลอดกาลนาน	ตลอดจน
ตลอดถึง	ตลอดทั้ง	ตลอดทั่ว	ตลอดทั่วถึง
ตลอดทั่วทั้ง	ตลอดปี	ตลอดไป	ตลอดมา
ตลอดระยะเวลา	ตลอดวัน	ตลอดเวลา	ตลอดศก
ต่อ	ต่อกัน	ต่อกับ	ต่อจาก
ต่อไป	ต่อไปนี้	ต่อมา	ต่อเมื่อ

ต่อว่า	ต่อให้	ต่อๆ	ต้อง
ต้องการ	ตอน	ตอนก่อน	ตอนใด
ตอนต่อ	ตอนต่อไป	ตอนต่อมา	ตอนถัดไป
ตอนถัดมา	ตอนที่	ตอนที่แล้ว	ตอนนั้น
ตอนนี้	ตอนแรก	ตอนสุดท้าย	ตอนหน้า
ตอนหลัง	ตอนไหน	ตอนๆ	ตะหาก
ตั้ง	ตั้งต้น	ตั้งแต่	ตั้งแต่นั้น
ตั้งแต่นี้	ตั้งแต่แรก	ตั้งที่	ตั้งอยู่
ตัว	ตัวใด	ตัวที่	ตัวนั้น
ตัวนี้	ตัวไหน	ตัวละ	ตัวไหน
ตัวอย่างเช่น	ตัวเอง	ตัวๆ	ต่าง
ต่างก็	ต่างหาก	ต่างๆ	ตาม
ตามด้วย	ตามแต่	ตามที่	ตามๆ
เต็มไปด้วย	เต็มไปหมด	เต็มๆ	แต่
แต่ก็	แต่ก่อน	แต่จะ	แต่เดิม
แต่ต้อง	แต่ถ้า	แต่ทว่า	แต่ที่
แต่นั้น	แต่เพียง	แต่เมื่อ	แต่ไร
แต่ละ	แต่ว่า	แต่ไหน	แต่อย่างใด
โต	โตๆ	โต	ถ้า
ถ้าจะ	ถ้าหาก	ถึง	ถึงแก่
ถึงจะ	ถึงบัดนั้น	ถึงบัดนี้	ถึงเมื่อ
ถึงเมื่อใด	ถึงเมื่อไร	ถึงแม้	ถึงแม้จะ
ถึงแม้ว่า	ถึงอย่างไร	ถือ	ถือว่า

ถูก	ถูกต้อง	ถูกๆ	เถอะ
เถิด	ทรง	ทว่า	ทั้ง
ทั้งคน	ทั้งตัว	ทั้งที่	ทั้งนั้น
ทั้งนั้นด้วย	ทั้งนั้นเพราะ	ทั้งนี้	ทั้งปวง
ทั้งเป็น	ทั้งมวล	ทั้งสิ้น	ทั้งหมด
ทั้งหลาย	ทั้งๆ	ทั้งๆ ที่	ทัน
ทันใดนั้น	ทันที่	ทันที่ทันใด	ทั่ว
ทำไม	ทำไม	ทำให้	ทำๆ
ที่	ที่	ที่จริง	ที่ซึ่ง
ทีเดียว	ที่ใด	ที่ใด	ที่ได้
ที่เถอะ	ที่แท้	ที่แท้จริง	ที่นั่น
ที่นี่	ที่ไร	ที่ละ	ที่ละ
ที่แล้ว	ที่ว่า	ที่สุด	ที่แห่งนั้น
ที่ไหน	ที่ๆ	ที่ๆ	ทุก
ทุกคน	ทุกครั้ง	ทุกครา	ทุกคราว
ทุกชิ้น	ทุกตัว	ทุกทาง	ทุกที่
ทุกที่	ทุกเมื่อ	ทุกวัน	ทุกวันนี้
ทุกสิ่ง	ทุกหน	ทุกแห่ง	ทุกอย่าง
ทุกอัน	ทุกๆ	เท่า	เท่ากัน
เท่ากับ	เท่าใด	เท่าที่	เท่า นั้น
เท่านี้	เท่าไร	เท่าไหร่	แท้
แท้จริง	เธอ	นอก	นอกจาก
นอกจากที่	นอกจากนั้น	นอกจากนี้	นอกจากว่า

นอกนั้น	นอกเหนือ	น้อย	น้อยกว่า
น้อยๆ	นะ	นะ	นัก
นั้น	นั้นไฉ	นับจากนั้น	นับจากนี้
นับแต่นี้	น่า	นาง	นางสาว
น่าจะ	นาน	นานๆ	นาย
นำ	นำพา	นำมา	นิด
นิดหน่อย	นิดๆ	นี่	นี้
นี่ไง	นี่นา	นี่แน่ะ	นี่แหละ
นี่แหละ	นี่เอง	นี่เอง	นู่น
ผู้	เน้น	เน้น	เน้นเอง
ใน	ในช่วง	ในที่	ในเมื่อ
ในระหว่าง	บน	บอก	บอกแล้ว
บอกว่า	บ่อย	บ่อยกว่า	บ่อยครั้ง
บ่อยๆ	บัดดล	บัดเดี๋ยวนี้	บัดนั้น
บัดนี้	บาง	บ้าง	บางกว่า
บางขณะ	บางครั้ง	บางครั้ง	บางครั้ง
บางที่	บางที่	บางแห่ง	บางๆ
แบบ	ปฏิบัติ	ประกอบ	ประการ
ประการฉะนี้	ประการใด	ประการหนึ่ง	ประมาณ
ประสพ	ปรับ	ปรากฏ	ปรากฏว่า
ปัจจุบัน	ปิด	เป็น	เป็นด้วย
เป็นดัง	เป็นต้น	เป็นแต่	เป็นเพื่อ
เป็นอัน	เป็นอันมาก	เป็นอาทิ	เปลี่ยน

เปลี่ยนแปลง	ไป	ผ่าน	ผ่านๆ
ผิด	ผิดๆ	ผู้	ผู้ใด
เผื่อ	เผื่อจะ	เผื่อที่	เผื่อว่า
ฝ่าย	ฝ่ายใด	ฯพณฯ	พบ
พบว่า	พยายาม	พร้อม	พร้อมกัน
พร้อมกัน	พร้อมด้วย	พร้อมทั้ง	พร้อมที่
พร้อมเพียง	พวก	พวกกัน	พวกกู
พวกแก	พวกเขา	พวกคุณ	พวกฉัน
พวกท่าน	พวกที่	พวกเธอ	พวกนั้น
พวกนี้	พวกนั้น	พวกโน้น	พวกมัน
พวกมึง	พอ	พอกัน	พอควร
พอจะ	พอดี	พอดี	พอที่
พอที่	พอเพียง	พอแล้ว	พอสม
พอสมควร	พอเหมาะ	พอๆ	พา
ฟัง	ฟัง	ฟังๆ	พูด
เพราะ	เพราะฉะนั้น	เพราะว่า	เพิ่ง
เพิ่งจะ	เพิ่ม	เพิ่มเติม	เพียง
เพียงแค่ว่า	เพียงใด	เพียงแต่	เพียงพอ
เพียงเพราะ	เพียงเพื่อ	เพียงไหน	เพื่อ
เพื่อที่	เพื่อว่า	เพื่อให้	ภาค
ภาย	ภายใต้	ภายนอก	ภายหลัง
มอง	มองว่า	มัก	มักจะ
มัน	มัย	มา	มาก

มากกว่า	มากมาย	มี	มีฉะนั้น
มิใช่	มิได้	มี	มีแต่
มิ่ง	มุง	มุงเน้น	มุงหมาย
เมื่อ	เมื่อก่อน	เมื่อครั้ง	เมื่อครั้งก่อน
เมื่อคราว	เมื่อคราวก่อน	เมื่อคราวที่	เมื่อคืน
เมื่อเช้า	เมื่อใด	เมื่อนั้น	เมื่อนี้
เมื่อเย็น	เมื่อไร	เมื่อวันวาน	เมื่อวาน
เมื่อไหร่	แม้	แม้กระทั่ง	แม้แต่
แม้ว่า	แม้ว่า	ไม่	ไม่ค่อย
ไม่ค่อยจะ	ไม่ค่อยเป็น	ไม่ใช่	ไม่เป็นไร
ไม่ว่า	ยก	ยกให้	ยอม
ยอมรับ	ยอม	ยอม	ยัง
ยังคง	ยังนั่น	ยังนี่	ยังเงิ่น
ยังไฉ	ยังจะ	ยังแต่	ยาก
ยาว	ยาวนาน	ยิ่ง	ยิ่งกว่า
ยิ่งขึ้น	ยิ่งขึ้นไป	ยิ่งจน	ยิ่งจะ
ยิ่งนัก	ยิ่งเมื่อ	ยิ่งแล้ว	ยิ่งใหญ่
เยอะ	เยอะแยะ	แยะ	รวด
รวดเร็ว	รวม	ร่วม	รวมกัน
ร่วมกัน	รวมด้วย	ร่วมด้วย	รวมถึง
รวมทั้ง	ระยะ	ระหว่าง	รับ
รู้	รู้	รู้ว่า	เร็ว
เร็ว ๆ	เรา	เรา ๆ	เริ่ม

เรียก	เรียก	เรื่อย	เรื่อย ๆ
ไร	ล้วน	ล้วนจน	ล้วนแต่
ละ	ล่าสุด	เล็ก	เล็กน้อย
เล็กๆ	เลย	แล้วว่า	แล้ว
แล้วกัน	แล้วแต่	แล้วเสร็จ	วันใด
วันนั้น	วันนี้	วันไหน	สบาย
สมัย	สมัยก่อน	สมัยนั้น	สมัยนี้
สมัยโน้น	ส่วน	ส่วนเกิน	ส่วนน้อย
ส่วนดี	ส่วนใด	ส่วนที่	ส่วนน้อย
ส่วนนั้น	ส่วนมาก	ส่วนใหญ่	สั้น
สั้น ๆ	สามารถ	สำคัญ	สิ่ง
สิ่งใด	สิ่งนั้น	สิ่งนี้	สิ่งไหน
สั้น	สุด	เสร็จ	เสร็จแล้ว
เสียจน	เสียด้วย	เสียนี้	เสียแล้ว
แสดง	แสดงว่า	หน	หนอ
หนอย	หนอย	หมด	หมดกัน
หมดสิ้น	หรือไง	หรือเปล่า	หรือไม่
หรือยัง	หรือไร	หลังจาก	หาก
หากแม้	หากแม้	หากแม้ว่า	หากว่า
หาความ	หาใช่	หรือ	เหตุ
เหตุหมด	เหตุนั้น	เหตุนี้	เหตุไร
เห็นแก่	เห็นควร	เห็นจะ	เห็นว่า
เหลือ	เหลือเกิน	เหล่า	เหล่านั้น

เหล่านี้	แห่ง	แห่งใด	แห่งนั้น
แห่งนี้	แห่งนี้	แห่งนี้	แหละ
ให้	ให้แก่	ใหญ่	ใหญ่โต
ไหน	ไหนๆ	อดีต	อนึ่ง
อยาก	อย่าง	อย่างเช่น	อย่างดี
อย่างเดียว	อย่างใด	อย่างที่	อย่างน้อย
อย่างนั้น	อย่างนี้	อย่างนั้น	อย่างมาก
อย่างยิ่ง	อย่างไร	อย่างไร้	อย่างไรก็ได้
อย่างไรเสีย	อย่างละ	อย่างหนึ่ง	อย่างไร
ต่างๆ	อัน	อันจะ	อันใด
อันได้แก่	อันที่	อันที่จริง	อันที่จะ
อันเนื่องมาจาก	อันละ	อันไหน	อันๆ
อาจ	อาจจะ	อาจเป็น	อาจเป็นด้วย
อีก	อื่น	อื่นๆ	เอง
เอ็ง	เอา	ฯลฯ	



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ค รายการคำที่ได้จากการวิเคราะห์เอกสาร

ผู้วิจัยได้วิเคราะห์คำที่ปรากฏอยู่ในเอกสาร ซึ่งเป็นบทคัดย่อโครงการงานนิสิตชั้นปีที่ 4 ที่จัดทำโดยนิสิตคณะพาณิชยศาสตร์และการบัญชี ภาควิชาสถิติ สาขาเทคโนโลยีสารสนเทศเพื่อธุรกิจ จุฬาลงกรณ์มหาวิทยาลัยในช่วงปีการศึกษา 2552-2554 ทั้งนี้ผู้วิจัยได้นำรายการคำหยุดของวรวีร์ เกษร (2550) และพจนานุกรมอิเล็กทรอนิกส์ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติมาที่ปรับปรุงล่าสุดปี 2552 มาประกอบการวิเคราะห์ รายการคำที่ได้มาจะนำมาใช้ในการวิเคราะห์กลุ่มคำและกำหนดดัชนีในชุดทดสอบ

about	access	action	activity
address	air	analysis	animation
apple	application	artificial	association
attitude	audit	average	basic
best	bit	black	blend
block	board	book	builder
building	bus	business	c
capture	cash	cell	center
chat	chief	china	class
classic	collaboration	comma	command
commerce	commission	common	communication
community	compare	computer	construction
contact	container	control	copy
core	creation	creator	crime
cross	current	cutting	dashboard
debt	decision	degree	dependency
design	designer	designing	dimension
e	earth	eclipse	edit

end	enterprise	error	event
evidence	exception	exhibition	exist
export	express	expression	extract
feed	find	fish	flash
flex	floor	flower	folio
format	forum	foundation	free
fuel	function	game	good
grade	guidance	hardware	historical
home	human	image	industry
infrared	injection	intelligence	interest
interior	international	inventory	java
kernel	key	language	learning
lens	let	library	licence
license	limited	lion	load
lunch	made	magazine	management
manager	manual	map	margin
marker	market	mash	mass
maya	melody	memory	mobile
model	modeling	motion	mouse
multimedia	multiple	native	network
noble	object	objective	office
officer	online	open	opinion
oracle	order	organization	original
package	parallel	park	pattern
people	performance	pictorial	plan
plate	platform	play	plot
pocket	point	pollution	powder
power	process	processing	product
production	profit	program	programming

project	promote	prompt	public
publication	query	r	reader
rebate	relationship	remote	report
reporter	response	resume	retail
return	ribbon	route	sale
screen	script	search	secure
security	seed	senior	server
service	share	shell	shop
site	skeleton	smart	snort
social	source	speech	sport
spring	square	standard	stop
stream	structure	studio	study
suit	supply	support	surface
sweet	table	tablet	tamarind
task	technique	technology	ten
text	thai	thread	time
tool	total	touch	tour
tourism	town	traffic	transform
turnover	tutorial	up	us
user	v	value	variable
vehicle	version	video	view
visual	visualize	voice	warehouse
web	wide	window	wireless
without	word	words	world
x	กฎ	กรณี	กรรมการ
กรอก	กระจาย	กระดาน	กระทบ
กระบวน	กระแสน	กล	กล่อก
กล่อก	กลาง	กล่อก	กลาย
ก่อก	กะ	ก่อก	กาน

กาย	การก	การนำ	กาล
กึ่ง	กึ่ง	ขนส่ง	ขอบเขต
ข้าง	ข้าง	ขาย	ขาย
ข้าว	ข้าว	ข้าวสาร	ชี้
ขีด	ขโมย	คณะ	คดี
ครอบ	คล้อง	คล้องตัว	คลัง
คลุม	ควบ	ควบคุม	ความถี่
ความเร็ว	ค็อก	คอม	คัด
คัน	ค้ำ	ค่า	คา
คำขอ	คิว	คืน	คุย
คู	คู่	คู่มือ	งาน
งาม	จน	จรรยา	จอ
จังหวะ	จันทร์	จ้าง	จาน
จาม	จ่าย	จำหน่าย	จี
จุด	ฉบับ	ฉาย	ชม
ช่อง	ชัด	ชัดเจน	ชั้น
ชั้น	ชั่วโมง	ชำระ	ชิน
ชั้น	ชี้	ชี้	ชื้อ
ชุด	ชุบ	ชู	ชอย
ชัก	ชี้	ชื้อ	ด้วน
ด้วน	ด้อย	ดิบ	ตั้ง
ตั้งดู	ตรงกัน	ตรวจ	ตอบ
ต่อ	ตัวกลาง	ตัวผู้	ตัวอย่าง
ตัวเงิน	ตัวเลข	ตัวแปร	ตาราง
ต่ำ	ตำแหน่ง	ติดต่อ	ติดตาม
ถนัด	ถาน	ถึงกัน	ทดสอบ
ทน	ทบ	ทอ	ห้อง
ทอง	ห้องถิ่น	ทันสมัย	ทั่วไป
ท่า	ท่า	ท่าทาง	ท่า

ทึ่ม	ทูน	ธรรมดา	นว
น้อง	นั้ง	นับ	นัย
นาค	นาวิก	น้ำตาล	นิยม
นุ่มนวล	บกพร่อง	บ่ง	บท
บรรจุ	บัญญัติ	บาน	บ้าน
บาย	บำรุง	บี	บุ
ปฏิทิน	ป็น	ประกาศ	ประจำ
ประชาสัมพันธ์	ประดิษฐ์	ประตุ	ประยุกต์
ประโยชน์	ปลัก	ปลีก	ปลุก
ปัจจัย	ป้าย	ปี	ผล
ผลิต	ผลงาน	ผ้า	ผู้ใหญ่
ฝึกสอน	พัก	พัน	พยาบาล
พลาด	พาดินชัย	พาด	พิมพ์
พิเศษ	พี	พื้นฐาน	พื้นที่
ฟรี	ฟอร์ม	ฟัง	ฟ้า
ภาคภูมิ	ภายใน	มด	มวด
มอบ	มัด	มั่นคง	มันใจ
มาร	มีเสียง	มือ	มือ
มุ่ม	มูล	ย่อ	ยอด
ยื้อน	ยา	ย่าน	ยาน
ย้าย	यी	ยึดหยุ่น	ยื่น
ยุ่งยาก	รม	รส	ร้อง
รอง	รอด	ร้อน	รอบ
รักษา	รับประกัน	รับสั่ง	รา
รากฐาน	ร้าง	ร้าง	ร้าน
ราย	รายงาน	รู้	รู้
รุ่น	รุนแรง	รู้	รู้
รูป	ลง	ลงรอยกัน	ลด
ลบ	ล้ม	ลอก	ล้ม

ลักษณะ	ลั่น	ลับ	ลาย
ลำดับ	ลิงค์	ลี	ลีก
วงจร	วัด	วาง	ว่าง
วิจัย	วิชาการ	วิถี	วิทยาศาสตร์
วิสัย	วุฒิ	ศร	ศัภษา
ศาล	ศาสตร์	ศาสนา	ศึกษา
ศูนย์	ศูนย์การค้า	สงสัย	ส่งเสริม
สถาน	สถานี	สถาบัน	สถาปัตยกรรม
สถิติ	สน	สนทนา	สนุกสนาน
สภาวะ	สมควร	สมบูรณ์	สมัคร
สมาคม	สมัครเสมอ	สมุด	สมเหตุสมผล
สร้าง	สร้างสรรค์	สรุป	ส่วนผสม
สวย	สอง	สอดคล้อง	สอน
สอบ	สอบสวน	สะดวก	สังกัด
สังคม	สังเคราะห์	สัญลักษณ์	สัดส่วน
สปดาห์	สัมผัส	สัมพันธ์	สาขา
สาธารณสุข	สาธารณะ	สาธิต	สารสนเทศ
สำนัก	สำรวจ	สำหรับ	สำเร็จ
สำเร็จรูป	สิ่งแวดล้อม	สิทธิ	สิทธิ์
สินค้า	สิ้นสุด	สืบ	สื่อ
สื่อสาร	สื่อสารมวลชน	สุขภาพ	สุดท้าย
สู่	สูง	สูญ	สโมสร
หน่วย	หนัก	หนังสือ	หน้า
หนาแน่น	หนี	หมวด	หมาย
หมู่	หยาบ	หรือ	หลัก
หลัง	ห้อง	ห้องสมุด	หัน
หา	ห่าง	ห่าง	หาย
อก	องค์	องศา	อด
อบรม	อวย	ออ	ออกกำลัง

ออกเสียง	อ้อย	อะไร	อักษร
อัตรา	อัตราในมิติ	อันดับ	อา
อากาศ	ข้างอิง	อาจารย์	อาทิตย์
อาศัย	อำ	อิสระ	อี
อู่	ฮา	เก	เกม
เก่า	เกี่ยว	เข้	เข้าถึง
เข้ารหัส	เขี้ยว	เครือ	เครื่อง
เครื่องจักร	เครื่องประดับ	เงิน	เจ
เจริญ	เจ้าหน้าที่	เฉลี่ย	เจ็ด
เช่า	เด่น	เดิน	เดินทาง
เดือน	เต็ม	เต็มที่	เติบโต
เท	เทียบ	เบิก	เปรียบเทียบ
เปอร์เซ็นต์	เป้าหมาย	เปิด	เม็ด
เย็น	เยี่ยม	เร่งรีบ	เรียน
เรียบร้อย	เรียบเรียง	เรือ	เรือ
เลข	เล็ง	เล่า	เลี้ยง
เลียน	เลื่อน	เลื่อน	เว้น
เวน	เวลา	เสถียร	เส้น
เสมอ	เสริม	เสียง	เสียง
เชื้อ	เหนือ	เหมาะสม	เหลว
เชื้อ	แข่ง	แข่งขัน	แขนง
แง	แจ้ง	แดง	แต่ง
แถบ	แถว	แท่ง	แนว
แบ	แบ่ง	แปรรูป	แปล
แปลก	แผนก	แผนการ	แพ
แพง	แพร	แพร์	แพร์หลาย
แพ้ม	แมน	แมน	มัน
แม่นยำ	แรก	แลกเปลี่ยน	แวดล้อม
แหวง	แหล่ง	แอ	โครง

โควตา	โต้ตอบ	โทษ	ไป
โปรแกรม	โม	โยง	โรงงาน
โรงแรม	โลก	ไหว้	โอ
โอกาส	โอน	ไฮ	ใจ
ใจแตก	ไบ	ไส้	ใหม่
ไซ	ไถ	ไทย	ไว้
ไวยากรณ์	ไอ		



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ง รายการกลุ่มคำที่ได้จากการวิเคราะห์กลุ่ม

ตารางที่ ง.1 แสดงผลการจัดกลุ่มที่ได้จากการวิเคราะห์กลุ่มคำด้วยจำนวนกลุ่ม 30 กลุ่ม
สมการค่าความเหมือนโคไซน์ และสมการเกณฑ์การเชื่อมโยงเชิงสมบูรณ์

กลุ่มที่ 1

ไซ	องค์	สถาบัน	พาล
----	------	--------	-----

ฉิ

กลุ่มที่ 2

โป	original	ส่งเสริม	เครื่อง
----	----------	----------	---------

ประชาสัมพันธ

กลุ่มที่ 3

ด้อย	ไฮ	กระดาน	อวย
อี	คำขอ	service	พยาบาล
center	ขอบเขต	ถึงกัน	วงจร
เครื่องจักร	คู	project	เจ
error	รักษา	access	license
แดง	บำรุง	security	block
public	พิเศษ	สมเหตุสมผล	source
community	เชี่ยวชาญ	exception	อักษร
ลับ	เข้ารหัส	cutting	สัปดาห์
report	ย่อ	copy	china
แ่ง	โรงแรม	stream	tourism
หยาบ	ศาล	bit	interest
format	ชอย	crime	market
บ่ง	ย่าน	skeleton	fish
สื่อสารมวลชน	town	ข้าว	บท
screen	เว้น	หาย	lunch
cross	รวม	สูญ	order

ไถ	edit	modeling	made
evidence	ฝึกสอน	sale	กล
animation	ทำทาง	แพร	คุย
capture	นุ่มนวล	ribbon	สนทนา
motion	บ้าน	site	artificial
สภาวะ	สงสัย	นั่ง	ประดิษฐ์
ศักยภาพ	เรียบง่าย	อก	อบรม
lens	เลข	วิถี	contact
แฉดวง	ชี	flower	ทอ
office	เวน	designing	reader
ตัวกลาง	สาธารณสุข	หนัก	feed
เทียบ	สังกัด	snort	ปลั๊ก
เติบโต	มัด	สำนัก	ซื้อ
เจริญ	กล่อม	ข่าว	นับ
ตัวผู้	ขึ้น	pattern	suit
เสียง	พัน	technique	ขนาด
เลื่อน	ชิน	plate	ฮา
objective	ทดสอบ	licence	tablet
เด่น	เสริม	park	ยืดหยุ่น
game	ถาน	ป้าย	วิสัย
คัด	ท้องถิ่น	ขีด	โทษ
enterprise	image	audit	ด่วน
ตัวเงิน	บรรจุ	tool	จาน
พิมพ์	เต็ม	version	standard
จ้าง	creation	manager	guidance
ทันสมัย	magazine	พก	มั่นคง
ปี	ร่าง	ล้ม	สมควร
prompt	technology	กระแส	บุ
พลาด	association	สามิต	classic

foundation	มือ	คู่มือ	เหลว
สอดคล้อง	ซั๊ก	กฎ	end
ฟอร์ม	ผ้า	สม่ำเสมอ	task
เสื่อ	chief	find	voice
shop	tutorial	ประโยค	mouse
core	hardware	ว่าง	stop
วุฒิ	มาร	เลียน	ร้อง
officer	communication	ลอก	historical
ห้าง	play	text	สน
ลิงค์	กาล	common	เหนือ
นาวิก	เลื่อน	เจ้าหน้าที่	processing
สมุด	pictorial	blend	designer
ทอง	infrared	expression	bus
ฐ	รับสั่ง	reporter	folio
สังเคราะห์	interior	publication	margin
average	us	manual	floor
ศูนย์การค้า	about	creator	เปอร์เซ็นต์
exhibition	grade	container	สัดส่วน
plot	senior	structure	value
มอบ	ศาสนา	apple	comma
activity	video	noble	ลบ
resume	learning	forum	ใส่
ค็อก	chat	international	building
เบิก	เล่า	แฟ้ม	น้ำตาล
lion	แขนง	สอบถาม	plan
x	sweet	board	construction
เสถียร	ปลูก	seed	ผสวน
เข้	โรงงาน	tamarind	profit
retail	smart	decision	debt

ปลีก	อ้อย	แป้ง	inventory
สัญลักษณ์	บกพร่อง	address	current
วิชาการ	ลี	แถบ	r
table	รอด	window	यी
ความถี่	function	search	แปลก
people	compare	คลัง	คลัง
เรียงเรียง	ถนัด	production	ฉาย
ซูป	free	export	ดิบ
กึ่ง	without	kernel	ฉบับ
เครื่องประดับ	limited	ชี้	network
เสมอ	ร้าง	สิทธิ์	social
dashboard	กระบวนการ	touch	ตาราง
กาย	analysis	คา	รู้
ออกกำลัง	น้อง	โคเวตา	อา
สนุกสนาน	รุ่น	ยา	เวลา
โอน	สอน	แมน	เปิด
รับประกัน	คดี	มั่นใจ	ลง
remote	ด่วน	powder	อะไร
แปรรูป	ฟรี	return	เครีอ
ซ้าย	ลงรอยกัน	ชั่วโมง	แบ
แพร่หลาย	ป็น	พื้นที่	จำหน่าย
รา	ก่อ	ต่ำ	ภายใน
สำเร็จรูป	โหว	นิยม	ข้าม
สัมพันธ์	secure	degree	ท่า
กระจาย	share	พาณิชย์	งาน
แลกเปลี่ยน	รอบ	โอกาส	โครง
โต้ตอบ	ลักษณะ	ปัจจัย	design
สร้างสรรค์	มุ่ม	เป้าหมาย	home
สังคม	สัมพันธ์	wireless	สร้าง

เช็ค	ชี้	มือ	flash
เข้าถึง	ชัดเจน	ตรงกัน	ภาคภูมิ
ค่า	ลด	สำหรับ	รู้
ค้า	สืบ	load	ไว้
บัญชี	query	transform	ทำ
ศร	language	extract	ควบคุม
c	เยี่ยม	server	script
studio	ปี	study	action
visual	แรก	คิว	user
คลุม	อันดับ	ประยุกต์	ติดต่อ
ครอบ	ten	ศึกษา	promote
สรุป	ข่าวสาร	เปรียบเทียบ	แม่นยำ
ยอด	ประกาศ	หรือ	ใหม่
จ่าย	เก่า	เรื่อง	สื่อสาร
สมบูรณ์	product	แผนก	ประจำ
web	แต่ง	หน่วย	สูง
รากฐาน	ปฏิทิน	สถิติ	application
cell	กลาง	รายงาน	commerce
ฟ้า	ศูนย์	ชุด	mobile
งาน	commission	รูป	good
ทั่วไป	rebate	มูล	การนำ
อาจารย์	supply	ตั้ง	แนว
เต็มที	มวล	แมน	สอง
หมาย	mass	ใจ	เกม
ช่อง	นัย	ชั้น	หนาแน่น
หัน	เย็น	มีเสียง	รุนแรง
โลก	ร้อน	เย็น	เสียง
คล้องตัว	วิจัย	บาน	ล้อม
สะดวก	attitude	ประตุ	pollution

point	response	ซึบ	ข้างอิง
power	online	เดิน	ห่าง
word	support	กระทบ	shell
โอ	จี	marker	ซัด
สถาน	แจ้ง	สำเร็จ	maya
ค้น	model	อาศัย	dimension
ขโมย	สถาปัตยกรรม	จังหวะ	class
ผล	เก	melody	let
เช่า	นว	สู่	command
vehicle	อ้า	black	หมวด
ยาน	basic	opinion	ไวยากรณ์
อด	pocket	กล่อง	air
ต่อ	จอ	แพร์	relationship
เรียน	computer	route	หลัง
เกี่ยว	human	เร่งรีบ	เท
เล็ง	คู่	สถานี	memory
เหมาะสม	เฉลี่ย	ไอ	parallel
หนี้	สมาคม	สาธารณะ	สาขา
คืน	แข่ง	ขนส่ง	ศาสตร์
อัตโนมัติ	ผู้ใหญ่	แพง	surface
หลัก	map	กะ	หมู่
สิทธิ	event	มด	eclipse
รอง	sport	ท้อง	วิทยาศาสตร์
การก	traffic	แวดล้อม	ความเร็ว
กรณี	multimedia	earth	builder
อู่	ดิ่งดูด	warehouse	ตัวแปร
ตัวอย่าง	จาม	อาทิตย์	variable
view	ควบ	จันทร์	programming
อิสระ	สำรวจ	สิ่งแวดล้อม	ลำดับ

ตัวเลข	ทุน	world	best
thread	wide	flex	spring
object	oracle	กรรอก	collaboration
เดือน	exist	สมัคร	control
ทบ	V	ใบ	dependency
ทน	ลึก	ยุ่งยาก	injection
สุขภาพ	library	จน	java
time	platform	พื้นฐาน	แถว
ย้อน	native	เอื้อ	สิ้นสุด
แผนการ	ทีม	ลาย	package
program			
กลุ่มที่ 4			
business	วัด	process	แม่นยำ
performance	ชี้	แข่งขัน	key
กลุ่มที่ 4			
ก้าน	กึ่ง	ไทย	โยง
กลาย	แหล่ง	หา	โปรแกรม
งาม	สวย	thai	ห้อง
visualize	แปล	ห้องสมุด	
กลุ่มที่ 6			
ชั้น			
กลุ่มที่ 7			
สอบ	ตรวจ		
กลุ่มที่ 8			
เดินทาง	ติดตาม		
กลุ่มที่ 9			
เรือ	management	cash	ชำระ
fuel	แพ		
กลุ่มที่ 10			

express

intelligence

วาง

รส

กลุ่มที่ 24

จุด

กลุ่มที่ 25

อากาศ

กลุ่มที่ 26

ออกเสียง

เสียง

ธรรมดา

speech

ฟัง

กลุ่มที่ 27

ส่วนผสม

กลุ่มที่ 28

รู้

กลุ่มที่ 29

เดิน

กลุ่มที่ 30

เม็ด

ทา



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก จ ผลการวิเคราะห์เมตริกซ์แสดงคำที่ปรากฏในเอกสาร

เนื่องจากในงานวิจัยนี้ ผู้วิจัยสามารถวิเคราะห์เอกสารจำนวน 100 รายการ ได้คำศัพท์ทั้งสิ้น 794 คำ ดังนั้นเมตริกซ์แสดงคำที่ปรากฏอยู่ในเอกสารมีขนาดถึง 794x100 จึงไม่สามารถนำเสนอได้ในรูปของเมตริกซ์ ผู้วิจัยจึงได้นำเสนอผลการวิเคราะห์เมตริกซ์แสดงค่าน้ำหนักที่ปรากฏในเอกสาร ดังตารางที่ จ.1 โดยแต่ละแถวจะแสดงถึงค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร ทั้งนี้ ผู้วิจัยได้นำเสนอเฉพาะผลการวิเคราะห์เมตริกซ์แสดงคำที่ปรากฏในเอกสาร ที่มีค่าน้ำหนักมากกว่าศูนย์เท่านั้น

ตารางที่ จ.1 ตารางแสดงค่าน้ำหนักของคำที่ปรากฏอยู่ในเอกสาร

รหัสเอกสาร	คำ	ค่าน้ำหนัก	รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B01	ไซ	0.4183	52B01	ท่า	0.4000
52B01	ส่งเสริม	0.4887	52B01	ข้าม	0.3046
52B01	ผลิต	1.0458	52B01	ภายใน	0.1315
52B01	เครื่อง	0.9586	52B01	จำหน่าย	0.3046
52B01	promote	0.4000	52B01	แบ	0.3398
52B01	business	0.1842	52B01	สร้าง	0.1103
52B01	home	0.3398	52B01	โปรแกรม	0.1170
52B01	original	0.8000	52B01	flash	0.2796
52B01	design	0.2444	52B01	ติดต่อ	0.2340
52B01	โครง	0.1005	52B01	user	0.2602
52B01	งาน	0.0178	52B01	action	0.4000
52B01	ประชาสัมพันธ์	0.6796	52B01	script	0.2444
52B01	ไป	0.4887	52B01	ควบคุม	0.1708
52B01	ไทย	0.2000	52B01	ทำ	0.0282
52B01	วัด	0.2602	52B01	ไว้	0.1648

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B01	รู้	0.0774
52B01	ภาคภูมิ	0.4000
52B01	ใจ	0.2796
52B01	แมน	0.4000
52B02	ศึกษา	0.0850
52B02	ประยุกต์	0.0371
52B02	คิว	0.1859
52B02	ชั้น	0.2176
52B02	study	0.0634
52B02	server	0.5262
52B02	extract	0.5714
52B02	transform	0.5714
52B02	load	0.4854
52B02	สำหรับ	0.0430
52B02	สร้าง	0.0394
52B02	เปรียบเทียบ	0.1567
52B02	แข่งขัน	0.1177
52B02	หรือ	0.2441
52B02	ตั้ง	0.6527
52B02	มูล	0.1567
52B02	แหล่ง	0.1266
52B02	รูป	0.0617
52B02	ทำ	0.0605
52B02	ชุด	0.1859
52B02	รายงาน	0.1369
52B02	สถิติ	0.1315

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B02	หน่วย	0.1315
52B02	งาน	0.0191
52B02	แผนก	0.2427
52B02	เรื่อง	0.0939
52B02	ตรงกัน	0.2176
52B02	ตรวจ	0.1137
52B02	สอบ	0.1177
52B02	แม่นยำ	0.2176
52B02	โครง	0.0479
52B02	ตัวอย่าง	0.1672
52B02	เกี่ยว	0.0600
52B02	เรียน	0.1220
52B02	รู้	0.0553
52B02	ต่อ	0.2427
52B02	อด	0.2176
52B06	ติดตาม	0.5229
52B06	ยาน	0.3807
52B06	vehicle	0.1667
52B06	เช่า	1.5229
52B06	โครง	0.0140
52B06	งาน	0.0037
52B06	สำหรับ	0.0251
52B06	ผล	0.0562
52B06	ขโมย	0.1416
52B06	ประยุกต์	0.0216
52B06	ตั้ง	0.0871

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B06	เหมาะสม	0.0738
52B06	คู่	0.1416
52B06	กรณี	0.0833
52B06	ศึกษา	0.0124
52B06	การก	0.1084
52B06	รอง	0.0799
52B06	สิทธิ	0.1018
52B06	หลัก	0.0871
52B06	อัตโนมัติ	0.1165
52B06	ตั้ง	0.1269
52B06	คืน	0.1416
52B06	ชำระ	0.1165
52B06	หนี้	0.1667
52B06	ตำแหน่ง	0.1084
52B06	ค้น	0.1416
52B06	สถาน	0.0738
52B03	ศึกษา	0.0496
52B03	แนว	0.1463
52B03	โม	0.9586
52B03	บาย	1.5229
52B03	คอม	1.0153
52B03	ตัวอย่าง	0.0975
52B03	การนำ	0.0332
52B03	study	0.0370
52B03	good	0.1667
52B03	mobile	0.5825

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B03	commerce	0.5091
52B03	application	0.0281
52B03	แอ	0.1859
52B03	สูง	0.0532
52B03	ประจำ	0.0712
52B03	สื่อสาร	0.1423
52B03	สอง	0.0738
52B03	เกม	0.1084
52B03	โอ	0.1018
52B03	ชม	0.0799
52B03	หรือ	0.0610
52B03	ทำ	0.0412
52B03	งาน	0.0130
52B03	เครื่อง	0.0799
52B03	โปรแกรม	0.0487
52B03	word	0.1667
52B03	power	0.1269
52B03	point	0.1269
52B03	รอง	0.0799
52B03	สะดวก	0.0332
52B03	คล้องตัว	0.1269
52B03	สื่อ	0.0738
52B03	สินค้า	0.0582
52B03	ติดต่อ	0.0488
52B03	แหล่ง	0.0738
52B03	โลก	0.0641

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B03	หัน	0.1084
52B03	ช่อง	0.1743
52B03	ใหม่	0.0712
52B03	ผล	0.0281
52B03	มือ	0.0871
52B03	ราย	0.0565
52B03	เต็มที่	0.1416
52B03	ประยุกต์	0.0433
52B03	ครอบ	0.1018
52B03	คลุม	0.1018
52B03	เกี่ยว	0.0350
52B03	visual	0.0601
52B03	studio	0.0641
52B03	c	0.0871
52B03	ชื่อ	0.3206
52B03	ศร	0.1667
52B03	คณะ	0.1018
52B03	บัญชี	0.0914
52B03	สำหรับ	0.0502
52B03	ร้าน	0.0767
52B03	ค้า	0.1084
52B03	ค่า	0.1033
52B03	สรุป	0.0799
52B03	ยอด	0.2538
52B03	ขาย	0.0687
52B03	อาจารย์	0.1165

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B03	ทั่วไป	0.0565
52B03	งาน	0.1667
52B03	โครง	0.0140
52B04	มือ	1.0458
52B04	ร้าน	0.7367
52B04	กรณี	0.1000
52B04	ศึกษา	0.0149
52B04	ฟ้า	1.0000
52B04	cell	0.2000
52B04	study	0.0444
52B04	mobile	0.1398
52B04	application	0.1686
52B04	สารสนเทศ	0.1075
52B04	สำหรับ	0.0301
52B04	โครง	0.0167
52B04	งาน	0.0111
52B04	สะดวก	0.0796
52B04	รากฐาน	0.1301
52B04	ประยุกต์	0.0260
52B04	web	0.1973
52B04	หลัก	0.0523
52B04	สมบูรณ์	0.1398
52B04	สอง	0.0886
52B04	รอง	0.1917
52B04	ทำ	0.0212
52B04	จ่าย	0.1708

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B04	เงิน	0.2602
52B04	เรื่อง	0.0658
52B04	รูป	0.0432
52B04	เข้าถึง	0.0796
52B04	เช็ค	0.2000
52B04	อะไร	0.1699
52B04	ค่า	0.0620
52B05	งาน	0.0223
52B05	พี	0.3333
52B05	ลง	0.1769
52B05	แอ	0.2833
52B05	เปรียบ	0.1219
52B05	ไซ	0.0581
52B05	หา	0.1553
52B05	หรือ	0.0543
52B05	เปิด	0.1831
52B05	เวลา	0.1100
52B05	รู้	0.0430
52B05	สารสนเทศ	0.0597
52B05	สร้าง	0.0306
52B05	สำหรับ	0.0669
52B05	โครง	0.0186
52B05	ครอบ	0.1358
52B05	คลุม	0.1358
52B05	ศึกษา	0.0331
52B05	แนว	0.0650

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B05	สะดวก	0.0884
52B05	ประจำ	0.0949
52B05	ทั่วไป	0.0753
52B05	ทำ	0.0157
52B05	เกี่ยว	0.0467
52B05	อา	0.1888
52B05	รู้	0.1692
52B05	ตาราง	0.1283
52B07	ไซ	0.5229
52B07	เปิด	0.0458
52B07	social	0.1888
52B07	network	0.0944
52B07	เครือข่าย	0.2168
52B07	ข่าย	0.2168
52B07	สังคม	0.2905
52B07	ขอ	0.1241
52B07	ประจำ	0.0474
52B07	สร้างสรรค์	0.0944
52B07	ลง	0.0442
52B07	โต้ตอบ	0.0944
52B07	สร้าง	0.0460
52B07	แลกเปลี่ยน	0.2330
52B07	กระจาย	0.1111
52B07	หรือ	0.0271
52B07	ผล	0.0375
52B07	สำหรับ	0.1171

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B07	ภายใน	0.1827
52B07	ติดต่อ	0.0975
52B07	สัมพันธ์	0.0492
52B07	รู้	0.0430
52B07	ทำ	0.0274
52B07	สำเร็จรูป	0.0777
52B07	รูป	0.0240
52B07	ทั่วไป	0.0377
52B07	ประยุกต์	0.0433
52B07	รา	0.1692
52B07	แพร่หลาย	0.0581
52B07	เวลา	0.0825
52B07	ค่า	0.0344
52B07	จ่าย	0.0474
52B07	โครง	0.0093
52B07	งาน	0.0099
52B07	ฉบับ	0.0679
52B07	ศึกษา	0.0248
52B07	หมาย	0.0949
52B07	เกี่ยว	0.0233
52B07	wireless	0.1111
52B07	supply	0.1111
52B07	แผนก	0.1888
52B07	rebate	0.1111
52B07	commission	0.1111
52B07	ศูนย์	0.0679

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B07	กลาง	0.0777
52B07	หน้า	0.1065
52B07	ไป	0.0679
52B07	ปฏิทิน	0.2222
52B07	สื่อสาร	0.0474
52B07	กรณี	0.0556
52B07	สถาน	0.0492
52B08	ชื่อ	0.1776
52B08	แต่ง	1.0455
52B08	สินค้า	0.6990
52B08	ประยุกต์	0.0399
52B08	product	0.4301
52B08	web	0.1012
52B08	application	0.0519
52B08	e	0.0888
52B08	commerce	0.0940
52B08	แอ	0.0245
52B08	การนำ	0.0918
52B08	เก่า	0.1171
52B08	มวล	0.1307
52B08	mass	0.1538
52B08	เป้าหมาย	0.4699
52B08	ราย	0.0521
52B08	ทำ	0.0109
52B08	ปัจจัย	0.0769
52B08	ตอบ	0.1146

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B08	เปิด	0.1268
52B08	โอกาส	0.1417
52B08	สร้าง	0.0212
52B08	ตัวอย่าง	0.0450
52B08	ไซ	0.0402
52B08	โครง	0.0129
52B08	งาน	0.0017
52B08	ศึกษา	0.0114
52B08	ลง	0.1224
52B09	ศึกษา	0.0297
52B09	โปรแกรม	0.2047
52B09	ประยุกต์	0.0519
52B09	พาณิชย์	0.3903
52B09	study	0.0444
52B09	e	0.3465
52B09	commerce	0.3666
52B09	degree	0.2000
52B09	product	0.1398
52B09	หรือ	0.1953
52B09	นิยม	0.0585
52B09	ทำ	0.0212
52B09	ต่ำ	0.2000
52B09	สะดวก	0.0796
52B09	ทั่วไป	0.1356
52B09	เรื่อง	0.0658
52B09	ชำระ	0.1398

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B09	คำ	0.0620
52B09	สินค้า	0.6990
52B09	รอง	0.0959
52B09	เช่า	0.1523
52B09	พื้นที่	0.1699
52B09	ร้าน	0.0921
52B09	ขาย	0.0824
52B09	เข้าถึง	0.1592
52B09	เวลา	0.0495
52B09	สถาน	0.0886
52B09	เปิด	0.0824
52B09	ชั่วโมง	0.1155
52B09	ติดต่อ	0.0585
52B09	ประชาสัมพันธ์	0.1699
52B09	ประกาศ	0.1699
52B09	ข่าวสาร	0.1301
52B09	รูป	0.0864
52B09	การนำ	0.1194
52B09	สัมผัส	0.1398
52B09	มุม	0.3666
52B09	สร้าง	0.0827
52B09	ลักษณะ	0.0770
52B09	สำหรับ	0.0301
52B09	รอบ	0.1301
52B09	ไซ	0.1569
52B09	share	0.2000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B09	ไว้	0.0824
52B09	ชม	0.0959
52B09	ผล	0.0337
52B09	งาน	0.0022
52B10	แอ	0.3188
52B10	secure	0.2857
52B10	web	0.1409
52B10	application	0.0723
52B10	ช่อง	0.5229
52B10	โหว	0.7143
52B10	สังคม	0.0747
52B10	ติดต่อ	0.0418
52B10	สื่อสาร	0.0610
52B10	ก่อ	0.1567
52B10	ป็น	0.1429
52B10	ทำ	0.0252
52B10	เล็ง	0.0747
52B10	คำ	0.2788
52B10	เป้าหมาย	0.0873
52B10	สำหรับ	0.0215
52B10	ซี	0.0873
52B10	โอ	0.0873
52B10	สินค้า	0.0499
52B10	กรณี	0.0714
52B10	ศึกษา	0.0106
52B10	หรือ	0.0174

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B10	ชัดเจน	0.0610
52B10	ten	0.2857
52B10	อันดับ	0.0929
52B10	แรก	0.0929
52B10	รายงาน	0.0685
52B10	ไว้	0.0589
52B10	ปี	0.0714
52B10	เครื่อง	0.0685
52B10	server	0.0658
52B10	หลัก	0.0373
52B10	งาน	0.0016
52B10	เยี่ยม	0.1429
52B10	ชม	0.0685
52B10	ทั่วไป	0.0484
52B10	ขาย	0.0589
52B11	สารสนเทศ	0.2150
52B11	language	0.1301
52B11	query	0.2000
52B11	visual	0.0721
52B11	studio	0.0770
52B11	สืบ	0.4569
52B11	โปรแกรม	0.2924
52B11	ลง	0.1592
52B11	แหล่ง	0.3544
52B11	รูป	0.0864
52B11	ลด	0.1973

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B11	ลงรอยกัน	0.4000
52B11	โลก	0.3078
52B11	การนำ	0.0398
52B11	ยุ่งยาก	0.1301
52B11	จน	0.0796
52B11	ทำ	0.0141
52B11	งาน	0.0089
52B11	โครง	0.0335
52B11	ศึกษา	0.0595
52B11	เกี่ยว	0.0420
52B11	แนว	0.0585
52B11	พื้นฐาน	0.1539
52B11	หลัก	0.1046
52B11	หา	0.0699
52B11	รู้	0.0387
52B11	ตัวอย่าง	0.0585
52B11	แอ	0.0319
52B11	ภายใน	0.0658
52B11	ติดต่อ	0.0585
52B12	สารสนเทศ	0.1434
52B12	ทุน	1.4214
52B12	ศึกษา	0.1487
52B12	management	0.0667
52B12	ทำ	0.0094
52B12	คณะ	0.0815
52B12	บัญชี	0.0731

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B12	สะดวก	0.0265
52B12	จน	0.1061
52B12	visual	0.0481
52B12	หลัก	0.0349
52B12	web	0.0877
52B12	application	0.0899
52B12	รูป	0.0288
52B12	ใบ	0.1133
52B12	สมัคร	0.5592
52B12	กรอก	0.1015
52B12	พื้นฐาน	0.0513
52B12	รายงาน	0.0639
52B13	ศึกษา	0.0149
52B13	ประยุกต์	0.0260
52B13	social	1.0194
52B13	network	1.0194
52B13	application	0.3372
52B13	study	0.0444
52B13	flex	0.4000
52B13	อันดับ	0.1301
52B13	เปิด	0.0824
52B13	ผล	0.0337
52B13	สร้าง	0.0827
52B13	platform	0.5592
52B13	script	0.2444
52B13	flash	0.2796

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B13	library	0.3046
52B13	สำหรับ	0.0301
52B13	package	0.2000
52B13	ติดต่อ	0.0585
52B13	ทำ	0.0141
52B13	builder	0.1523
52B13	web	0.0658
52B13	พื้นฐาน	0.0770
52B13	eclipse	0.1222
52B13	design	0.1222
52B13	surface	0.2000
52B14	โปรแกรม	0.0487
52B14	เดินทาง	0.2397
52B14	best	0.3333
52B14	program	0.2832
52B14	จรรยา	1.2218
52B14	ผล	0.0281
52B14	สุขภาพ	0.1165
52B14	ทน	0.1667
52B14	จน	0.0663
52B14	กลายเป็น	0.1165
52B14	ทบ	0.1018
52B14	สูง	0.0532
52B14	สื่อ	0.1477
52B14	รายงาน	0.0799
52B14	การนำ	0.1658

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B14	ไซ	0.0436
52B14	ไทย	0.0833
52B14	รูป	0.0360
52B14	รู้	0.1613
52B14	เรียน	0.2846
52B14	ลักษณะ	0.0641
52B14	ชัดเจน	0.1423
52B14	ทำ	0.0059
52B14	โครง	0.0140
52B14	งาน	0.0019
52B14	หรือ	0.0203
52B14	แอ	0.0797
52B14	แนว	0.0488
52B14	ออ	0.0621
52B14	เปิด	0.0687
52B14	ค่า	0.0516
52B14	จ่าย	0.0712
52B14	ย้อน	0.1269
52B14	หลัง	0.0914
52B14	เดือน	0.3333
52B14	ปี	0.0833
52B14	มูล	0.0914
52B14	time	0.1416
52B14	สัมพันธ์	0.0738
52B14	เวลา	0.0412
52B14	วาง	0.0914

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B14	แผนการ	0.1165
52B15	ศึกษา	0.0425
52B15	ทำ	0.0605
52B15	งาน	0.0223
52B15	spring	1.7143
52B15	study	0.0634
52B15	collaboration	0.2857
52B15	control	0.1859
52B15	dependency	0.2857
52B15	injection	0.2857
52B15	โปรแกรม	0.0418
52B15	นิยม	0.1672
52B15	แอส	0.1821
52B15	สำหรับ	0.0860
52B15	ติดต่อ	0.3343
52B15	หรือ	0.0698
52B15	java	0.1650
52B15	แนว	0.2507
52B15	แถว	0.7281
52B15	ตาราง	0.4950
52B15	ลด	0.0939
52B15	object	0.2857
52B15	รูป	0.1234
52B15	สร้าง	0.0394
52B15	ตัวอย่าง	0.0836
52B15	สรุป	0.1369

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B16	ศึกษา	0.0135
52B16	ประยุกต์	0.0590
52B16	โอ	0.4998
52B16	study	0.0202
52B16	application	0.0153
52B16	โปรแกรม	0.2658
52B16	thread	0.0909
52B16	เท	0.0591
52B16	memory	0.0909
52B16	หน่วย	0.1256
52B16	parallel	0.0909
52B16	ผล	0.2453
52B16	อัตรา	0.0591
52B16	สูง	0.0870
52B16	สาขา	0.0772
52B16	ทำ	0.0706
52B16	งาน	0.0192
52B16	ศาสตร์	0.0772
52B16	นิยม	0.0266
52B16	หมู่	0.0772
52B16	วิทยาศาสตร์	0.0772
52B16	ความเร็ว	0.1818
52B16	เครื่อง	0.1307
52B16	รอง	0.0436
52B16	ลด	0.0299
52B16	เวลา	0.0450

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B16	หรือ	0.0333
52B16	หมาย	0.0388
52B16	ลง	0.0362
52B16	ตอบ	0.0339
52B16	เรื่อง	0.0299
52B16	ตัวแปร	0.0591
52B16	variable	0.0909
52B16	programming	0.0692
52B16	c	0.0475
52B16	visual	0.0328
52B16	studio	0.0350
52B16	ลำดับ	0.3332
52B16	ตัวเลข	0.1384
52B16	หา	0.0953
52B16	วัด	0.0591
52B16	จน	0.0362
52B16	สิ้นสุด	0.0909
52B16	การนำ	0.0181
52B16	หลัง	0.0499
52B16	ทีม	0.1818
52B16	สมบูรณ์	0.1271
52B16	แนว	0.0266
52B17	ศึกษา	0.1487
52B17	study	0.1109
52B17	management	0.5000
52B17	language	0.6505

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B17	native	1.0000
52B17	ติดต่อ	0.1463
52B17	สื่อสาร	0.2135
52B17	แลกเปลี่ยน	0.6990
52B17	หรือ	0.1221
52B17	ลาย	1.0000
52B17	งาน	0.0167
52B17	ทั่วไป	0.1694
52B17	world	0.5000
52B17	wide	0.5000
52B17	web	0.1644
52B17	รูป	0.1079
52B17	เชื้อ	0.3807
52B17	ทำ	0.0176
52B17	สัมพันธ์	0.2215
52B17	server	0.4604
52B17	oracle	0.3495
52B17	exist	0.5000
52B17	ลึก	0.5000
52B17	สร้าง	0.2068
52B17	v	0.5000
52B17	สารสนเทศ	0.4032
52B17	ตัวอย่าง	0.4388
52B17	แอ	0.0797
52B18	โปรแกรม	0.0798
52B18	สัมพันธ์	0.2417

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B18	จรรยา	1.2218
52B18	แข่งขัน	0.7490
52B18	relationship	0.2542
52B18	traffic	0.3813
52B18	sport	0.5455
52B18	event	0.4634
52B18	map	0.1271
52B18	ผู้ใหญ่	0.1818
52B18	นิยม	0.0532
52B18	สถาน	0.2417
52B18	แข่ง	0.3089
52B18	เดินทาง	0.2614
52B18	ทำ	0.0192
52B18	สมาคม	0.1818
52B18	ไทย	0.1818
52B18	สะดวก	0.0362
52B18	สื่อ	0.1611
52B18	รายงาน	0.0871
52B18	การนำ	0.1809
52B18	ไซ	0.0475
52B18	รูป	0.0393
52B18	รู้	0.1408
52B18	เรียน	0.2329
52B18	ลักษณะ	0.0700
52B18	ชัดเจน	0.0776
52B18	โครง	0.0152

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B18	งาน	0.0020
52B18	หรือ	0.0444
52B18	แอ	0.0869
52B18	แนว	0.0532
52B18	ออ	0.0677
52B18	เปิด	0.0749
52B18	ค่า	0.0563
52B18	จ่าย	0.0776
52B18	เฉลี่ย	0.2542
52B18	ผล	0.0307
52B18	วาง	0.0997
52B18	แผนการ	0.1271
52B20	ประยุกต์	0.0944
52B20	การนำ	0.1447
52B20	สื่อ	0.2417
52B20	แอ	0.3188
52B20	กรณี	0.0909
52B20	ศึกษา	0.0406
52B20	multimedia	0.1818
52B20	tour	0.4634
52B20	web	0.1196
52B20	application	0.0613
52B20	study	0.0403
52B20	square	0.3636
52B20	ดึงดูด	0.1183
52B20	ใจ	0.1271

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B20	รูป	0.0785
52B20	นิยม	0.0532
52B20	ราย	0.3081
52B20	สถาน	0.1611
52B20	จาม	0.1818
52B20	สร้าง	0.0501
52B20	ทำ	0.0385
52B20	ทั่วไป	0.0616
52B20	สอบ	0.0749
52B20	ชม	0.2614
52B20	เดิน	0.1384
52B20	ประตู	0.1545
52B20	บาน	0.1818
52B20	ลักษณะ	0.0700
52B20	ยื่น	0.1818
52B20	ภายใน	0.1196
52B20	องศา	0.3636
52B20	รอบ	0.1183
52B20	ไว้	0.0749
52B20	โครง	0.0152
52B20	งาน	0.0041
52B20	โปรแกรม	0.0532
52B20	สำเร็จรูป	0.1271
52B20	flash	0.1271
52B20	มีเสียง	0.1818
52B20	ร้าน	0.0837

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B20	ค้า	0.1183
52B20	ชั้น	0.1818
52B20	ผล	0.0613
52B20	ขอ	0.0677
52B20	สำรวจ	0.1111
52B20	ควบ	0.1818
52B20	คู่	0.1818
52B20	รู้	0.1056
52B20	หรือ	0.0444
52B21	ผล	0.1226
52B21	ตัวแปร	0.1183
52B21	ตอบ	0.7447
52B21	ขอ	0.4739
52B21	human	0.1545
52B21	computer	0.1384
52B21	online	0.2222
52B21	response	0.3636
52B21	ลักษณะ	0.3498
52B21	หน้า	0.4357
52B21	web	0.0598
52B21	สมบุรณ์	0.3813
52B21	attitude	0.1818
52B21	เวลา	0.1799
52B21	time	0.1545
52B21	โครง	0.0152
52B21	งาน	0.0020

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B21	วิจัย	0.1183
52B21	ร้อน	0.3636
52B21	เย็น	0.3636
52B21	สำหรับ	0.0274
52B21	หลัง	0.2992
52B21	หน่วย	0.0837
52B21	คณะ	0.1111
52B21	บัญชี	0.0997
52B21	เฉลี่ย	0.3813
52B21	นัย	0.4634
52B21	สถิติ	0.0837
52B21	อิสระ	0.1818
52B25	จุด	1.0458
52B25	support	0.2176
52B25	จุด	0.2857
52B25	พี	0.2857
52B25	โครง	0.0718
52B25	งาน	0.0223
52B25	ประยุกต์	0.0742
52B25	ภายใน	0.0939
52B25	แรก	0.1859
52B25	ทำ	0.0605
52B25	ศึกษา	0.0425
52B25	เกี่ยว	0.0600
52B25	แจ้ง	0.6527
52B25	ศูนย์	0.1745

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B25	เรื่อง	0.0939
52B25	หา	0.3994
52B25	ตำแหน่ง	1.1152
52B25	โลก	0.1099
52B25	แพร่หลาย	0.1494
52B25	แม่นยำ	0.2176
52B25	แนว	0.0836
52B25	สอง	0.2532
52B25	หลัก	0.1494
52B25	จน	0.1137
52B25	ผล	0.1445
52B25	โปรแกรม	0.0836
52B25	หน้า	0.1369
52B25	จอ	0.1429
52B25	แฉ	0.0911
52B25	เครื่อง	0.2739
52B25	pocket	0.5714
52B25	visual	0.1030
52B25	basic	0.1997
52B25	ตรวจ	0.1137
52B25	สอบ	0.1177
52B25	คู่	0.2427
52B25	ค่า	0.0885
52B25	อ่ำ	0.2857
52B25	นว	0.2857
52B25	สะดวก	0.0568

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B22	ศึกษา	0.0892
52B22	ประยุกต์	0.1039
52B22	พี	1.0000
52B22	study	0.0444
52B22	application	0.0337
52B22	โครง	0.0335
52B22	งาน	0.0200
52B22	สารสนเทศ	0.1075
52B22	เรื่อง	0.0658
52B22	แพ	0.1523
52B22	เก	0.1699
52B22	รูป	0.0432
52B22	สถาปัตยกรรม	0.2796
52B22	แอ	0.2869
52B22	model	0.8388
52B22	view	0.8000
52B22	ผล	0.2023
52B22	หลัก	0.1046
52B22	ทำ	0.0423
52B22	ติดต่อ	0.1170
52B22	หรือ	0.0488
52B22	ชัดเจน	0.2562
52B22	กระทบ	0.1699
52B22	สัมพันธ์	0.0886
52B22	ออ	0.0745
52B22	ลด	0.0658

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B22	ลง	0.0796
52B22	เท	0.1301
52B22	สำหรับ	0.0301
52B22	โปรแกรม	0.0292
52B22	ตัวอย่าง	0.0585
52B22	ศูนย์	0.1222
52B22	คณะ	0.1222
52B22	บัญชี	0.1097
52B22	visual	0.1442
52B22	studio	0.1539
52B22	server	0.1842
52B22	พื้นฐาน	0.0770
52B23	สัมพันธ์	0.3938
52B23	จราจร	0.9503
52B23	อากาศ	1.6990
52B23	relationship	0.3107
52B23	traffic	0.3107
52B23	air	0.4444
52B23	pollution	0.4444
52B23	map	0.3107
52B23	เรื่อง	0.0731
52B23	ล้อม	0.2222
52B23	เสียง	0.1888
52B23	ทบ	0.1358
52B23	สุขภาพ	0.1553
52B23	รุนแรง	0.2222

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B23	หนาแน่น	0.1692
52B23	การนำ	0.2211
52B23	ไซ	0.1162
52B23	ไทย	0.1111
52B23	แหล่ง	0.0985
52B23	กลาง	0.1553
52B23	ควบคุม	0.1897
52B23	สิ่งแวดล้อม	0.1692
52B23	รูป	0.1919
52B23	ตาราง	0.1283
52B23	ค่า	0.0689
52B23	สถิติ	0.1023
52B23	ปัจจัย	0.1111
52B23	หรือ	0.0814
52B23	ทำ	0.0157
52B23	โครง	0.0186
52B23	แอ	0.0708
52B23	แนว	0.0650
52B23	มูล	0.1219
52B23	ผล	0.0375
52B23	ลง	0.0884
52B23	เฉลี่ย	0.1553
52B23	จันทร์	0.2222
52B23	อาทิตย์	0.2222
52B23	เปิด	0.0915
52B23	โอกาส	0.1023

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B23	อ้างอิง	0.1283
52B23	ห่าง	0.1888
52B23	วาง	0.1219
52B26	ศึกษา	0.0676
52B26	ประยุกต์	0.1180
52B26	study	0.0403
52B26	application	0.0307
52B26	shell	0.1818
52B26	script	0.1111
52B26	โครง	0.0305
52B26	งาน	0.0223
52B26	แนว	0.1064
52B26	ทำ	0.0257
52B26	โปรแกรม	0.0532
52B26	พื้นฐาน	0.0700
52B26	เครื่อง	0.0871
52B26	เครือข่าย	0.2366
52B26	ข่าย	0.2366
52B26	หนังสือ	0.1050
52B26	หรือ	0.0222
52B26	เรียน	0.0776
52B26	รู้	0.0704
52B26	ไวยากรณ์	0.1818
52B26	หมวด	0.1818
52B26	หมู่	0.1545
52B26	จุด	0.1901

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B26	สะดวก	0.0362
52B26	ลักษณะ	0.1399
52B26	command	0.1545
52B26	let	0.1818
52B26	class	0.1818
52B27	ศึกษา	0.0425
52B27	โปรแกรม	0.2924
52B27	ประยุกต์	0.0556
52B27	สร้าง	0.1773
52B27	สำหรับ	0.1075
52B27	study	0.0317
52B27	tour	0.2427
52B27	dimension	0.1429
52B27	model	0.1997
52B27	สื่อ	0.1266
52B27	งาน	0.0095
52B27	สถาปัตยกรรม	0.0999
52B27	เกม	0.0929
52B27	หน้า	0.0685
52B27	สวย	0.0929
52B27	งาม	0.0929
52B27	เครื่อง	0.1369
52B27	สูง	0.0456
52B27	ลง	0.0568
52B27	สะดวก	0.0284
52B27	เชื้อ	0.1088

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B27	นิยม	0.0418
52B27	maya	0.1214
52B27	เหมาะสม	0.0633
52B27	ผล	0.0241
52B27	โครง	0.0120
52B27	มือ	0.0747
52B27	สมัคร	0.0999
52B27	ชัด	0.1429
52B27	หา	0.0499
52B27	warehouse	0.2176
52B27	หรือ	0.0349
52B27	ไว้	0.0589
52B27	โม	0.2054
52B27	วาง	0.0784
52B27	earth	0.2857
52B27	แวดล้อม	0.1214
52B27	รอบ	0.0929
52B27	ชัดเจน	0.0610
52B27	ทำ	0.0050
52B27	เลี้ยง	0.0747
52B27	การนำ	0.0568
52B27	ไซ	0.0373
52B27	เรื่อง	0.0470
54B01	โปรแกรม	0.2339
54B01	กล่อง	1.6990
54B01	opinion	0.2000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B01	black	0.2000
54B01	เปิด	0.3296
54B01	โอกาส	0.1842
54B01	เรื่อง	0.0658
54B01	ฟัง	0.2796
54B01	ทำ	0.0141
54B01	งาน	0.0045
54B01	รูป	0.1727
54B01	แนว	0.0585
54B01	หา	0.1398
54B01	หรือ	0.0488
54B01	หมาย	0.0854
54B01	ช่อง	0.2092
54B01	โครง	0.0167
54B01	สะดวก	0.0398
54B01	เข้าถึง	0.1592
54B01	เวลา	0.0495
54B01	สู่	0.0959
54B01	เต็มที่	0.1699
54B01	รู้	0.0774
54B01	แอ	0.0319
54B01	โลก	0.0770
54B01	สังคม	0.1046
54B02	แต่ง	1.6990
54B02	melody	0.6667
54B02	สู่	0.3195

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B02	จังหวะ	0.6667
54B02	หรือ	0.0814
54B02	รูป	0.1439
54B02	รู้	0.1291
54B02	อาศัย	0.3069
54B02	เรื่อง	0.2192
54B02	คำ	0.2066
54B02	เสียง	0.4337
54B02	ทำ	0.0471
54B02	โครง	0.0558
54B02	งาน	0.0074
54B03	ศึกษา	0.1487
54B03	อา	1.6990
54B03	study	0.4437
54B03	สารสนเทศ	0.5376
54B03	สำเร็จ	1.3979
54B03	ปัจจัย	1.0000
54B03	งาน	0.0223
54B03	เวลา	0.4949
54B03	คำ	0.6198
54B03	จ่าย	0.8539
54B04	สำหรับ	0.0376
54B04	แอ	0.0398
54B04	web	0.0822
54B04	application	0.0422
54B04	นิยม	0.0731

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B04	แพร่	0.2500
54B04	แพร่หลาย	0.2614
54B04	ช่อง	0.1307
54B04	ทำ	0.0088
54B04	เข้าถึง	0.0995
54B04	ไซ	0.3268
54B04	การนำ	0.0497
54B04	สินค้า	0.6990
54B04	ซื้อ	0.0962
54B04	ราย	0.0847
54B04	หรือ	0.0610
54B04	ชม	0.1198
54B04	สร้าง	0.0689
54B04	โครง	0.0209
54B04	งาน	0.0028
54B04	ฉบับ	0.1527
54B04	ศึกษา	0.0372
54B04	ประยุกต์	0.0325
54B05	แอ	0.1417
54B05	หา	0.3107
54B05	เดินทาง	0.9586
54B05	web	0.0365
54B05	application	0.0187
54B05	route	0.1111
54B05	โครง	0.0093
54B05	งาน	0.0037

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B05	สะดวก	0.0884
54B05	เร่งรีบ	0.1111
54B05	ห้อง	0.1111
54B05	หรือ	0.0271
54B05	จราจร	0.0679
54B05	มล	0.1111
54B05	อากาศ	0.0944
54B05	เวลา	0.1650
54B05	ชั่วโมง	0.0642
54B05	ตาราง	0.0642
54B05	กะ	0.1111
54B05	หมาย	0.0474
54B05	ค่า	0.0689
54B05	จ่าย	0.0949
54B05	แพง	0.0944
54B05	จน	0.0442
54B05	ขนส่ง	0.0846
54B05	สาธารณะ	0.0944
54B05	นิยม	0.0325
54B05	พี	0.1667
54B05	ไอ	0.0556
54B05	สถานี	0.4719
54B05	สถาน	0.1969
54B05	ชัดเจน	0.0474
54B05	กรอก	0.1692
54B05	marker	0.1358

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B05	ห่าง	0.0944
54B05	ราย	0.0377
54B05	เดิน	0.0846
54B05	ขับ	0.0846
54B05	ถาน	0.0944
54B05	ทำ	0.0039
54B05	ซื้อ	0.0609
54B05	ติดตาม	0.0581
54B05	ข่าวสาร	0.0723
54B05	ปลั๊ก	0.1111
54B05	วาง	0.0609
54B06	โปรแกรม	0.2924
54B06	feed	1.4545
54B06	แอด	0.1739
54B06	reader	0.1545
54B06	platform	0.1271
54B06	นิยม	0.1064
54B06	สูง	0.0580
54B06	ตอบ	0.1354
54B06	งาน	0.0101
54B06	สะดวก	0.0362
54B06	ติดตาม	0.1901
54B06	ข่าวสาร	0.2366
54B06	ทอ	0.1818
54B06	เวลา	0.1799
54B06	ข่าว	1.4545

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B06	อะไร	0.1545
54B06	ใหม่	0.0776
54B06	หรือ	0.0222
54B06	ตรวจ	0.0724
54B06	สอบ	0.0749
54B06	ไซ	0.0475
54B06	สำนัก	0.3089
54B06	ประจำ	0.0776
54B06	ตั้ง	0.1384
54B06	สำหรับ	0.0274
54B06	eclipse	0.1111
54B06	java	0.1050
54B06	ควบคุม	0.0776
54B06	ทำ	0.0128
54B06	ผล	0.0307
54B06	โครง	0.0152
54B06	ไว้	0.0749
54B07	ประยุกต์	0.0779
54B07	application	0.0337
54B07	snort	0.6000
54B07	โลก	0.1539
54B07	เครือข่าย	0.1301
54B07	ข่าย	0.1301
54B07	เครื่อง	0.1917
54B07	หรือ	0.1953
54B07	หนัก	0.2000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B07	ขโมย	0.1699
54B07	จน	0.2388
54B07	มูล	0.1097
54B07	ทบ	0.2444
54B07	กระทบ	0.3398
54B07	นับ	0.1097
54B07	ไต่	0.2000
54B07	องค์	0.1398
54B07	หา	0.1398
54B07	ตัวผู้	0.2000
54B07	สืบ	0.3046
54B07	ลด	0.0658
54B07	เสี่ยง	0.1301
54B07	โปรแกรม	0.2047
54B07	โอ	0.1222
54B07	นิยม	0.0585
54B07	งาน	0.0223
54B07	มุม	0.1222
54B07	โครง	0.0335
54B07	ศึกษา	0.0297
54B07	รูป	0.0864
54B07	ทำ	0.0353
54B07	เสริม	0.3398
54B07	ทดสอบ	0.1398
54B07	ขึ้น	0.4000
54B07	รู้	0.0387

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B07	พัน	0.2000
54B07	ขึ้น	0.1699
54B08	เกม	1.3010
54B08	สำหรับ	0.1806
54B08	game	0.1699
54B08	application	0.0337
54B08	มือ	0.2092
54B08	นิยม	0.2340
54B08	ยอด	0.1523
54B08	ขาย	0.0824
54B08	สูง	0.0638
54B08	ลักษณะ	0.0770
54B08	งาน	0.0111
54B08	เด่น	0.1222
54B08	โอ	0.3000
54B08	สัมผัส	0.5592
54B08	หน้า	0.3834
54B08	จอ	0.4000
54B08	จุด	0.1046
54B08	แฉ	0.0956
54B08	ตอบ	0.0745
54B08	ศึกษา	0.0446
54B08	เข้าถึง	0.0796
54B08	หลัก	0.0523
54B08	ทำ	0.0282
54B08	ประยุกต์	0.0260

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B08	objective	0.1699
54B08	c	0.1046
54B08	เลื่อน	0.1699
54B09	designing	0.4000
54B09	flower	0.3398
54B09	โครง	0.0335
54B09	งาน	0.0223
54B09	ทำ	0.0565
54B09	แฉดววง	0.4000
54B09	แข่งขัน	0.3296
54B09	สูง	0.1277
54B09	รากฐาน	0.2602
54B09	ประยุกต์	0.0519
54B09	แฉ	0.0638
54B09	หลัก	0.1046
54B09	โปรแกรม	0.0585
54B09	สำเร็จรูป	0.2796
54B09	flash	0.2796
54B09	ตรวจ	0.3184
54B09	สอบ	0.3296
54B09	สิทธิ	0.4887
54B09	ร้าน	0.5525
54B09	หรือ	0.0488
54B09	ขาย	0.3296
54B09	อัตโนมัติ	0.2796
54B09	สินค้า	0.1398

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B09	โลก	0.1539
54B10	สำหรับ	0.1003
54B10	ขาย	0.2746
54B10	web	0.3288
54B10	application	0.1686
54B10	contact	0.3333
54B10	lens	0.3333
54B10	เข้าถึง	0.1326
54B10	ศัภยภพ	0.2832
54B10	ชม	0.3195
54B10	ชั่วโมง	0.1925
54B10	สะดวก	0.0663
54B10	โครง	0.0279
54B10	งาน	0.0186
54B10	แฉ	0.2125
54B10	โลก	0.2565
54B10	สภวะ	0.2832
54B10	แข่งขัน	0.2746
54B10	สูง	0.1064
54B10	รากฐาน	0.2168
54B10	การนำ	0.0663
54B10	หลัก	0.2614
54B10	โปรแกรม	0.0975
54B10	ไซ	0.0871
54B10	ทำ	0.0471
54B10	ซื้อ	0.7696

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B10	ติดต่อ	0.2925
54B10	สินค้า	0.4660
54B10	เลข	0.2832
54B10	แจ้ง	0.5076
54B10	ชำระ	0.4660
54B10	เงิน	0.4337
54B10	เรียบร้อย	0.2538
54B10	ร้าน	0.1535
54B10	สอบ	0.2746
54B10	ราย	0.1130
54B10	หรือ	0.0814
54B10	สงสัย	0.3333
54B10	user	0.2168
54B10	หลัง	0.1828
54B10	บ้าน	0.2832
54B10	office	0.3333
54B10	ชัดเจน	0.1423
54B10	รอง	0.1598
54B11	แฉ	0.3188
54B11	ขาย	0.3090
54B11	เหมาะสม	0.5538
54B11	สำหรับ	0.2258
54B11	web	0.0822
54B11	application	0.0422
54B11	โครง	0.0209
54B11	งาน	0.0167

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B11	สร้าง	0.1034
54B11	หรือ	0.0610
54B11	ซื้อ	0.6734
54B11	แรก	0.4879
54B11	ปี	0.3750
54B11	โปรแกรม	0.1097
54B11	ไซ	0.0654
54B11	นิยม	0.0731
54B11	สะดวก	0.0497
54B11	หลัก	0.0654
54B11	ตัวกลาง	0.1904
54B11	ติดต่อ	0.0731
54B11	เรียน	0.1067
54B11	รู้	0.0484
54B11	เวลา	0.0619
54B11	เทียบ	0.2124
54B11	ทำ	0.0088
54B11	ขอ	0.2793
54B11	กล้ำม	0.5000
54B11	มัด	0.5000
54B11	สังคม	0.1307
54B11	สถาบัน	0.1904
54B11	หน่วย	0.1151
54B11	สังกัด	0.1904
54B11	สุขภาพ	0.1747
54B11	สาธารณสุข	0.2500

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B12	เวน	2.0000
54B12	โครง	0.0670
54B12	ศึกษา	0.0892
54B12	รู้	0.1549
54B12	พื้นฐาน	0.1539
54B12	ทั่วไป	0.1356
54B12	เกี่ยว	0.2521
54B12	ตัวอย่าง	0.3510
54B12	โปรแกรม	0.0585
54B12	ประยุกต์	0.0519
54B12	สร้าง	0.1103
54B12	งาน	0.0089
54B12	ชี	0.2444
54B12	visual	0.1442
54B12	studio	0.1539
54B12	express	0.3046
54B12	ถาน	0.3398
54B12	สถานี	0.3398
54B12	สถาน	0.1772
54B12	ประตู	0.3398
54B13	script	0.1527
54B13	รูป	0.1079
54B13	ทำ	0.0529
54B13	งาน	0.0195
54B13	เครือข่าย	0.1626
54B13	ข่าย	0.1626

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B13	นิยม	0.1463
54B13	สะดวก	0.0497
54B13	โปรแกรม	0.0731
54B13	ขอ	0.1862
54B13	เข้าถึง	0.0995
54B13	เวลา	0.0619
54B13	ศึกษา	0.0930
54B13	เกี่ยว	0.1051
54B13	ชี	0.2500
54B13	การนำ	0.0497
54B13	โยง	0.1626
54B13	โครง	0.0419
54B13	ประยุกต์	0.2596
54B13	เด่น	0.1527
54B13	ตัวอย่าง	0.2925
54B13	แรก	0.1626
54B13	สอง	0.1108
54B13	ติดต่อ	0.2194
54B13	สุดท้าย	0.1747
54B14	ไอ	0.5000
54B14	สำหรับ	0.2408
54B14	ส่วนผสม	1.6000
54B14	application	0.0337
54B14	ปัจจัย	0.1000
54B14	เจริญ	0.3046
54B14	เติบโต	0.4000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B14	สมบูรณ์	0.2796
54B14	เหมาะสม	0.0886
54B14	สุขภาพ	0.2796
54B14	วิถี	0.2000
54B14	จน	0.0796
54B14	มือ	0.2000
54B14	ผล	0.0674
54B14	ชื่อ	0.5387
54B14	สะดวก	0.0796
54B14	งาน	0.0223
54B14	ส่งเสริม	0.1222
54B14	หัน	0.1301
54B14	เลี้ยง	0.1699
54B14	งาน	0.1699
54B14	ด่วน	0.2000
54B14	โทษ	0.2000
54B14	แอ	0.1594
54B14	พก	0.2000
54B14	วาง	0.1097
54B14	แผนการ	0.1398
54B14	เวลา	0.0495
54B14	ทำ	0.0282
54B14	หลัก	0.1046
54B14	โปรแกรม	0.0877
54B14	objective	0.1699
54B14	c	0.1046

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B14	ควบคุม	0.0854
54B14	ทดสอบ	0.1398
54B14	ลง	0.0796
54B14	manager	0.2000
54B14	version	0.1699
54B14	สร้าง	0.0276
54B14	โครง	0.0167
54B14	เรียบร้อย	0.1523
54B14	ไว้	0.0824
54B15	สำหรับ	0.0753
54B15	งาน	0.0149
54B15	ตรวจ	0.5306
54B15	สอบ	0.5493
54B15	สารสนเทศ	0.5376
54B15	tool	0.1165
54B15	audit	0.2832
54B15	โลก	0.0641
54B15	นับ	0.0914
54B15	ลด	0.0548
54B15	ขีด	0.1269
54B15	แข่งขัน	0.0687
54B15	การนำ	0.0332
54B15	เสียง	0.2168
54B15	ควบคุม	0.1423
54B15	ผล	0.0562
54B15	ติดตาม	0.0871

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B15	เรียน	0.0712
54B15	รู้	0.0323
54B15	เกี่ยว	0.0350
54B15	จน	0.1326
54B15	หัน	0.1084
54B15	เรื่อง	0.0548
54B15	standard	0.1667
54B15	guidance	0.1667
54B15	มั่นคง	0.1269
54B15	โครง	0.0419
54B15	หรือ	0.0610
54B15	ข้างเคียง	0.0962
54B15	control	0.1084
54B15	association	0.1667
54B15	สถาบัน	0.1269
54B15	technology	0.1084
54B15	สู่	0.0799
54B15	ค่า	0.0516
54B15	ร่าง	0.1667
54B15	ฟอร์ม	0.1667
54B15	การก	0.1084
54B15	ทำ	0.0059
54B15	ครอบ	0.1018
54B15	คลุม	0.1018
54B15	สอดคล้อง	0.1084
54B15	รูป	0.0360

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B15	รายงาน	0.0799
54B16	ประยุกต์	0.0556
54B16	ชั้น	1.4141
54B16	application	0.0241
54B16	foundation	0.2857
54B16	สารสนเทศ	0.0384
54B16	โครง	0.0359
54B16	งาน	0.0223
54B16	ศึกษา	0.0425
54B16	เกี่ยว	0.0900
54B16	มือ	0.1494
54B16	สำหรับ	0.0215
54B16	เสี่ยง	0.0929
54B16	พลาด	0.1745
54B16	เรื่อง	0.0470
54B16	ลด	0.0470
54B16	ผลิต	0.0747
54B16	ราย	0.0484
54B16	โลก	0.0550
54B16	ทำ	0.0454
54B16	หา	0.0499
54B16	แนว	0.0836
54B16	visual	0.0515
54B16	studio	0.0550
54B16	อาศัย	0.0658
54B16	ตัวอย่าง	0.0418

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B16	ตรวจ	0.0568
54B16	สอบ	0.0589
54B16	ตรงกัน	0.1088
54B16	พื้นฐาน	0.0550
54B16	สู่	0.0685
54B16	รูป	0.0308
54B16	ผล	0.0241
54B16	คอม	0.1088
54B16	command	0.1214
54B16	prompt	0.1429
54B17	สร้าง	0.2757
54B17	ปี	0.1553
54B17	magazine	0.4444
54B17	creation	0.1888
54B17	tablet	0.2222
54B17	โครง	0.0186
54B17	งาน	0.0050
54B17	ศึกษา	0.1157
54B17	เกี่ยว	0.0467
54B17	รูป	0.0960
54B17	ประยุกต์	0.0288
54B17	โยง	0.1446
54B17	เสียง	0.1446
54B17	ทำ	0.0235
54B17	หา	0.0777
54B17	สรุป	0.2130

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B17	หมาย	0.0949
54B17	เท	0.1446
54B17	สำหรับ	0.0669
54B17	ฮา	0.1692
54B17	นาค	0.2222
54B17	อาศัย	0.1023
54B17	แหล่ง	0.1969
54B17	หลัก	0.0581
54B17	ไซ	0.0581
54B17	ผลิต	0.1162
54B17	ชื่อ	0.1219
54B17	suit	0.2222
54B17	สุดท้าย	0.1553
54B17	หรือ	0.0271
54B18	ตรวจ	0.2653
54B18	สอบ	0.2746
54B18	ประยุกต์	0.1298
54B18	รู้	0.0645
54B18	ป้าย	0.6667
54B18	อัตโนมัติ	0.2330
54B18	park	0.3333
54B18	control	0.2168
54B18	licence	0.3333
54B18	plate	0.3333
54B18	technique	0.3333
54B18	ยาน	1.0153

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B18	ปัจจัย	0.1667
54B18	สะดวก	0.0663
54B18	สำหรับ	0.1003
54B18	เดินทาง	0.1598
54B18	ประจำ	0.1423
54B18	ผล	0.2248
54B18	อัตรา	0.2168
54B18	ควบคุม	0.1423
54B18	ศึกษา	0.0248
54B18	ลด	0.1096
54B18	พลาด	0.2036
54B18	ทำ	0.0235
54B18	งาน	0.0223
54B18	คืน	0.5663
54B18	โครง	0.0558
54B18	visual	0.1202
54B18	studio	0.1283
54B18	โปรแกรม	0.0487
54B18	ติดต่อ	0.0975
54B18	c	0.1743
54B18	สู่	0.1598
54B18	สรุป	0.1598
54B18	ยืดหยุ่น	0.2832
54B18	การนำ	0.0663
54B18	จอ	0.1667
54B19	ศึกษา	0.0453

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B19	เกี่ยว	0.0548
54B19	ตำแหน่ง	0.8485
54B19	งาน	0.0223
54B19	สารสนเทศ	0.2805
54B19	ไทย	0.1304
54B19	study	0.0193
54B19	technology	0.0566
54B19	อัตรา	0.0566
54B19	โลก	0.1004
54B19	เรียน	0.0371
54B19	รู้	0.0168
54B19	ใหม่	0.0371
54B19	สร้าง	0.0120
54B19	เปรียบเทียบ	0.0477
54B19	แข่งขัน	0.0358
54B19	เหมาะสม	0.0770
54B19	วิสัย	0.0870
54B19	ขีด	0.1324
54B19	คัด	0.0870
54B19	ทำ	0.0061
54B19	เต็ม	0.0739
54B19	โครง	0.0364
54B19	หลัก	0.0227
54B19	สถิติ	0.0400
54B19	ปี	0.0435
54B19	แหล่ง	0.0770

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B19	อ้างอิง	0.1004
54B19	ลักษณะ	0.0335
54B19	ไซ	0.0227
54B19	สรุป	0.0834
54B19	ผล	0.0293
54B19	สำรวจ	0.1062
54B19	รูป	0.0375
54B19	แอ	0.0554
54B19	ทั่วไป	0.0884
54B19	หน้า	0.0417
54B19	ตัวอย่าง	0.0254
54B20	เก	1.6990
54B20	บรรจุ	1.2000
54B20	design	0.0815
54B20	support	0.1015
54B20	image	0.1333
54B20	ประจำ	0.0569
54B20	ท้องถิ่น	0.1333
54B20	ผลิต	0.2092
54B20	สินค้า	0.0466
54B20	ทันสมัย	0.1133
54B20	เปรียบเทียบ	0.0731
54B20	แข่งขัน	0.0549
54B20	ค่า	0.1240
54B20	จ้าง	0.3398
54B20	จ่าย	0.0569

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B20	สูง	0.0426
54B20	ทำ	0.0141
54B20	ตั้ง	0.0697
54B20	โครง	0.0112
54B20	งาน	0.0015
54B20	แนว	0.0390
54B20	สร้าง	0.0368
54B20	ดึงดูด	0.0867
54B20	ใจ	0.0932
54B20	พิมพ์	0.1133
54B20	ลด	0.0877
54B20	ร้าน	0.0614
54B20	โปรแกรม	0.0585
54B20	eclipse	0.0815
54B20	หลัก	0.0349
54B20	ปี	0.0932
54B20	หน้า	0.0639
54B20	แอ	0.0213
54B20	ตัวเงิน	0.1333
54B20	เวลา	0.0330
54B20	ขาย	0.0549
54B21	ประยุกต์	0.0708
54B21	รี	1.2218
54B21	application	0.0613
54B21	enterprise	0.5455
54B21	library	0.2769

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B21	ทำ	0.0321
54B21	รู้	0.1056
54B21	แอ	0.2608
54B21	ศึกษา	0.0811
54B21	ชื่อ	0.0997
54B21	เกี่ยว	0.0764
54B21	พื้นฐาน	0.1399
54B21	แนว	0.0532
54B21	เรื่อง	0.0598
54B21	แพ	0.4153
54B21	design	0.1111
54B21	pattern	0.1818
54B21	อก	0.5455
54B21	cross	0.1818
54B21	cutting	0.1818
54B21	เข้าถึง	0.0724
54B21	access	0.1818
54B21	ลง	0.0724
54B21	exception	0.1818
54B21	ตรวจ	0.0724
54B21	สอบ	0.0749
54B21	สมเหตุสมผล	0.1818
54B21	มั่นคง	0.1384
54B21	security	0.1384
54B21	เข้ารหัส	0.1818
54B21	ลับ	0.1545

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B21	รูป	0.0393
54B21	community	0.1818
54B21	public	0.1818
54B21	license	0.1818
54B21	งาน	0.0061
54B21	ค่า	0.0563
54B21	จ่าย	0.0776
54B21	ก่อ	0.0997
54B21	ทั่วไป	0.0616
54B21	block	0.1818
54B21	ไซ	0.0475
54B21	สื่อ	0.1611
54B21	ตัวอย่าง	0.0532
54B21	แจ้ง	0.0951
54B21	สะดวก	0.0362
54B21	open	0.1271
54B21	source	0.1271
54B21	ด้อย	0.1384
54B22	ตัวแปร	0.4337
54B22	ผล	0.0749
54B22	เส้น	1.7778
54B22	จอ	0.2222
54B22	screen	0.1553
54B22	สื่อ	0.1969
54B22	การนำ	0.2211
54B22	สถิติ	0.1023

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B22	แพร่หลาย	0.1162
54B22	งาน	0.0050
54B22	วิทยาศาสตร์	0.1888
54B22	ศึกษา	0.0331
54B22	หรือ	0.1628
54B22	สื่อสารมวลชน	0.2222
54B22	รูป	0.4318
54B22	แนว	0.0650
54B22	สร้างสรรค์	0.1888
54B22	มุม	0.1358
54B22	user	0.1446
54B22	ตัวกลาง	0.1692
54B22	สร้าง	0.1225
54B22	โปรแกรม	0.0325
54B22	รา	0.1692
54B22	เหมาะสม	0.0985
54B22	ชัดเจน	0.0949
54B22	แปล	0.1692
54B22	หมาย	0.0949
54B22	วิจัย	0.4337
54B22	ห้อง	0.1888
54B22	ทดสอบ	0.3107
54B22	หน่วย	0.2046
54B22	ตอบ	0.0827
54B22	บ่ง	0.2222
54B22	ลักษณะ	0.0855

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B22	จุด	0.2324
54B22	marker	0.1358
54B22	format	0.1888
54B22	ก่อ	0.1219
54B22	สูง	0.0709
54B22	หายาบ	0.2222
54B22	เอื้อ	0.1692
54B22	ประยุกต์	0.0288
54B22	ประจำ	0.0949
54B22	แง่	0.2222
54B23	โปรแกรม	0.0688
54B23	จราจร	1.2218
54B23	traffic	0.0822
54B23	report	0.1176
54B23	application	0.0198
54B23	technology	0.0765
54B23	หนาแน่น	0.0896
54B23	แนว	0.0344
54B23	รายงาน	0.0564
54B23	หรือ	0.0144
54B23	ยาน	0.1792
54B23	เดินทาง	0.1692
54B23	เวลา	0.1164
54B23	ผล	0.1190
54B23	ย้อน	0.3583
54B23	หลัง	0.2581

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B23	ชั่วโมง	0.0679
54B23	สัปดาห์	0.1176
54B23	รูป	0.0254
54B23	อักษร	0.3998
54B23	เข้าถึง	0.0936
54B23	สะดวก	0.0234
54B23	สำหรับ	0.0354
54B23	งาน	0.0013
54B23	แดง	0.1176
54B23	คล้องตัว	0.0896
54B23	เขียว	0.1176
54B23	java	0.0679
54B24	ศูนย์	0.4752
54B24	พยาบาล	2.0000
54B24	ขอ	0.1241
54B24	service	0.2222
54B24	center	0.1111
54B24	online	0.0679
54B24	โครง	0.0093
54B24	งาน	0.0050
54B24	แอส	0.0708
54B24	สำหรับ	0.1505
54B24	ทำ	0.0118
54B24	เดินทาง	0.1598
54B24	visual	0.0401
54B24	studio	0.0428

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B24	โปรแกรม	0.0162
54B24	server	0.0512
54B24	สร้าง	0.0153
54B24	web	0.0365
54B24	แจ้ง	0.2538
54B24	หรือ	0.0407
54B24	หา	0.0388
54B24	คำขอ	0.0944
54B24	ตอบ	0.0827
54B24	ค้น	0.0944
54B24	ลด	0.0731
54B24	เวลา	0.0550
54B25	แอส	0.2318
54B25	สำหรับ	0.1095
54B25	web	0.0598
54B25	application	0.0307
54B25	e	0.2100
54B25	โครง	0.0152
54B25	งาน	0.0101
54B25	สะดวก	0.0362
54B25	นิยม	0.0532
54B25	ไซ	0.1426
54B25	หรือ	0.1110
54B25	ทำ	0.0192
54B25	อี	1.6990
54B25	ทั่วไป	0.0616

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B25	เครือข่าย	0.1183
54B25	ข่าย	0.1183
54B25	สังคม	0.0951
54B25	ขอ	0.0677
54B25	อวย	0.3636
54B25	กระดาน	0.1384
54B25	เช่า	0.1384
54B25	ไฮ	0.1818
54B25	ชื่อ	0.0997
54B25	ไว้	0.1498
54B25	ขอบเขต	0.1545
54B25	ศึกษา	0.0135
54B25	พิมพ์	0.1545
54B25	ลง	0.0724
54B25	ช่อง	0.0951
54B25	ติดต่อ	0.0532
54B25	สื่อสาร	0.1552
54B25	ถึงกัน	0.3636
54B25	ประจำ	0.0776
54B25	โอกาส	0.0837
54B25	พิเศษ	0.1384
53B01	ติดตาม	0.6100
53B01	บำรุง	0.3333
53B01	รักษา	0.2036
53B01	สารสนเทศ	0.2240
53B01	พลาด	0.1018

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B01	error	0.1416
53B01	ไป	0.1018
53B01	เจ	0.1018
53B01	project	0.1165
53B01	แถม	0.0531
53B01	web	0.0548
53B01	application	0.0281
53B01	คู	0.1667
53B01	เครื่องจักร	0.1667
53B01	กล	0.1667
53B01	วงจร	0.1416
53B01	สินค้า	0.0582
53B01	ครบ	0.1018
53B01	คลุม	0.1018
53B01	งาน	0.0223
53B01	ไถ	0.3333
53B01	เดิน	0.1269
53B01	นั่ง	0.1667
53B01	ขับ	0.1269
53B01	เกี่ยว	0.0350
53B01	จ้าง	0.2832
53B01	หลัก	0.0436
53B01	หน่วย	0.0767
53B01	คำขอ	0.1416
53B01	ทำ	0.0235
53B01	มือ	0.0871

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B01	สูญ	0.1165
53B01	หาย	0.1667
53B01	เวลา	0.0412
53B01	ช่อง	0.0871
53B01	การนำ	0.0332
53B01	ประยุกต์	0.0216
53B01	โปรแกรม	0.0487
53B01	visual	0.0601
53B01	studio	0.0641
53B01	คิว	0.1084
53B01	server	0.0767
53B01	ขาย	0.0687
53B01	บัญชี	0.0914
53B01	ควบคุม	0.0712
53B01	เข้าถึง	0.1326
53B01	สิทธิ	0.1018
53B02	สารสนเทศ	0.1955
53B02	สำหรับ	0.0821
53B02	ข้าว	0.7273
53B02	กล่อง	0.6178
53B02	made	0.1818
53B02	order	0.1818
53B02	lunch	0.1818
53B02	business	0.0837
53B02	ลด	0.1793
53B02	ทำ	0.0385

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B02	งาน	0.0223
53B02	ลง	0.2171
53B02	โครง	0.0457
53B02	visual	0.0656
53B02	basic	0.1271
53B02	ภายใน	0.1196
53B02	บท	0.1545
53B02	ขาย	0.2247
53B02	ติดต่อ	0.0532
53B02	ซื้อ	0.2798
53B02	ตรวจ	0.0724
53B02	สอบ	0.0749
53B02	ผลิต	0.2852
53B02	ผล	0.0920
53B02	สรุป	0.1743
53B02	ชัดเจน	0.0776
53B02	จอ	0.0909
53B02	เก่า	0.1384
53B02	ช่อง	0.0951
53B03	สารสนเทศ	0.2150
53B03	สำหรับ	0.0301
53B03	แพ	0.6092
53B03	business	0.0921
53B03	fish	0.2000
53B03	market	0.2000
53B03	ตลาด	0.2444

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B03	error	0.1699
53B03	ไป	0.1222
53B03	เจ	0.1222
53B03	project	0.1398
53B03	เท	0.1301
53B03	กลาย	0.1398
53B03	งาน	0.0223
53B03	ลด	0.0658
53B03	โครง	0.0167
53B03	หา	0.0699
53B03	ทำ	0.0071
53B03	แนว	0.0585
53B03	modeling	0.1523
53B03	language	0.1301
53B03	visual	0.0721
53B03	studio	0.0770
53B03	server	0.0921
53B03	รอง	0.0959
53B03	ครอบ	0.1222
53B03	คลุม	0.1222
53B03	security	0.1523
53B03	สู่	0.0959
53B03	sale	0.2000
53B03	management	0.7000
53B03	ขาย	0.2472
53B03	สินค้า	0.0699

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B03	ชำระ	0.5592
53B03	เงิน	0.5204
53B03	เรือ	0.6000
53B03	cash	0.4000
53B03	fuel	0.4000
53B03	หลัก	0.0523
53B03	เข้าถึง	0.0796
53B04	ศึกษา	0.0458
53B04	ประยุกต์	0.0599
53B04	study	0.0341
53B04	site	0.1538
53B04	รี	0.0940
53B04	library	0.1171
53B04	ribbon	0.1538
53B04	web	0.0506
53B04	ทำ	0.0597
53B04	งาน	0.0223
53B04	นับ	0.1688
53B04	ปัจจัย	0.0769
53B04	สำเร็จ	0.1075
53B04	เป้าหมาย	0.0940
53B04	ไว้	0.0634
53B04	ติดต่อ	0.0900
53B04	สื่อสาร	0.1314
53B04	แลกเปลี่ยน	0.1075
53B04	สารสนเทศ	0.0414

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B04	สูง	0.0491
53B04	แข่งขัน	0.1268
53B04	ศักยภาพ	0.1307
53B04	การนำ	0.0612
53B04	โปรแกรม	0.0675
53B04	ก่อ	0.0844
53B04	คล้องตัว	0.1171
53B04	ส่งเสริม	0.0940
53B04	เปรียบเทียบ	0.0844
53B04	สะดวก	0.0306
53B04	เข้าถึง	0.0612
53B04	ขนส่ง	0.1171
53B04	มวล	0.1307
53B04	แพร่	0.1538
53B04	เหมาะสม	0.0682
53B04	รูป	0.0332
53B04	สถาน	0.0682
53B04	เวลา	0.0761
53B04	โครง	0.0258
53B04	เกี่ยว	0.0323
53B04	หมาย	0.0657
53B04	สร้าง	0.0212
53B04	ไซ	0.0402
53B04	ศูนย์	0.0940
53B04	กลาง	0.1075
53B04	ขีด	0.1171

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B04	ตอบ	0.0573
53B04	ลด	0.1012
53B04	ค่า	0.0477
53B04	จ่าย	0.0657
53B04	อบรม	0.1538
53B04	รักษา	0.0940
53B06	ศึกษา	0.1116
53B06	ไทย	0.5000
53B06	study	0.1109
53B06	thai	0.5000
53B06	visualize	0.5000
53B06	รู้	0.1936
53B06	หา	0.5242
53B06	แหล่ง	0.4430
53B06	ห้องสมุด	0.5000
53B06	ห้อง	0.4247
53B06	เรียน	0.2135
53B06	หรือ	0.0610
53B06	กลายเป็น	0.3495
53B06	ช่อง	0.2614
53B06	ติดต่อ	0.1463
53B06	สื่อสาร	0.2135
53B06	โอกาส	0.2302
53B06	โปรแกรม	0.2924
53B06	กึ่ง	0.5000
53B06	ก้าน	0.5000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B06	งาน	0.0167
53B06	แปล	0.3807
53B06	หมาย	0.2135
53B06	สัมพันธ์	0.2215
53B06	โยง	0.3253
53B06	สวย	0.3253
53B06	งาม	0.3253
53B06	ทำ	0.0176
53B07	ศึกษา	0.0496
53B07	ประยุกต์	0.1298
53B07	study	0.0739
53B07	ทบ	0.8146
53B07	สำหรับ	0.1505
53B07	ตอบ	0.2482
53B07	หรือ	0.0814
53B07	ชั่วโมง	0.1925
53B07	กรณี	0.1667
53B07	ประดิษฐ์	1.0000
53B07	artificial	0.6667
53B07	intelligence	0.5076
53B07	สร้าง	0.1379
53B07	งาน	0.0223
53B07	การนำ	0.0663
53B07	สนทนา	1.1650
53B07	คุย	0.5663
53B07	โปรแกรม	0.1462

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B07	อาศัย	0.1535
53B07	ทำ	0.0353
53B07	language	0.2168
53B07	เวลา	0.1650
53B07	เรื่อง	0.1096
53B07	รูป	0.0720
53B07	พื้นฐาน	0.1283
53B07	เกี่ยว	0.0700
53B08	แอ	0.1700
53B08	จุด	1.0458
53B08	กรณี	0.2667
53B08	ศึกษา	0.0397
53B08	point	0.1015
53B08	interest	0.1133
53B08	application	0.0225
53B08	tourism	0.1333
53B08	study	0.0296
53B08	china	0.1333
53B08	town	0.1333
53B08	พี	0.1333
53B08	ไอ	0.1333
53B08	map	0.2796
53B08	java	0.0770
53B08	script	0.0815
53B08	เจ	0.0815
53B08	คิว	0.0867

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B08	รู้	0.1015
53B08	โครง	0.0112
53B08	งาน	0.0015
53B08	ส่งเสริม	0.0815
53B08	ย่าน	0.5663
53B08	ร้าน	0.1842
53B08	ค้า	0.0867
53B08	สถาน	0.0591
53B08	นิยม	0.0390
53B08	หรือ	0.0326
53B08	ชอย	0.1333
53B08	เดินทาง	0.5113
53B08	ราย	0.0452
53B08	เวลา	0.0660
53B08	หลัก	0.0349
53B08	ศาล	0.1333
53B08	วัด	0.0867
53B08	โรงแรม	0.1333
53B08	marker	0.1629
53B08	ขอ	0.0496
53B08	ฉบับ	0.1629
53B08	ย่อ	0.1133
53B08	เต็ม	0.1133
53B08	เลื่อน	0.1133
53B08	สะดวก	0.0265
53B09	ศึกษา	0.1487

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B09	การนำ	0.0497
53B09	ประยุกต์	0.0649
53B09	งาน	0.0139
53B09	ตรวจ	0.0995
53B09	โม	0.2397
53B09	ชั้น	1.1422
53B09	study	0.0555
53B09	application	0.0422
53B09	motion	0.6371
53B09	capture	0.7500
53B09	animation	0.2500
53B09	แอ	0.1992
53B09	เว้น	0.2500
53B09	อาศัย	0.2302
53B09	ทำ	0.0441
53B09	แนว	0.0731
53B09	ใหม่	0.1067
53B09	สร้าง	0.0345
53B09	หรือ	0.0305
53B09	รม	0.5000
53B09	พิเศษ	0.1904
53B09	ฮา	0.1904
53B09	ผล	0.0422
53B09	นุ่มนวล	0.2500
53B09	อ้างอิง	0.1444
53B09	ลักษณะ	0.0962

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B09	ท่าทาง	0.2500
53B09	ฝึกสอน	0.2500
53B09	ไอ	0.1250
53B09	ค่า	0.1549
53B09	project	0.1747
53B09	ตำแหน่ง	0.1626
53B09	จุด	0.1307
53B09	marker	0.1527
53B09	process	0.1747
53B09	model	0.1747
53B09	edit	0.2500
53B09	visual	0.0902
53B09	basic	0.1747
53B09	user	0.1626
53B09	โครง	0.0209
53B09	skeleton	0.2500
53B09	โปรแกรม	0.0366
53B09	builder	0.1904
53B10	ศึกษา	0.1259
53B10	ประยุกต์	0.0399
53B10	study	0.0341
53B10	application	0.0259
53B10	computer	0.3514
53B10	crime	0.1538
53B10	bit	0.1538
53B10	stream	0.1538

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B10	copy	0.1538
53B10	evidence	0.1538
53B10	ตรวจ	0.3673
53B10	สอบ	0.3803
53B10	งาน	0.0206
53B10	โครง	0.0515
53B10	ทำ	0.0706
53B10	หา	0.1075
53B10	ไซ	0.0804
53B10	หนังสือ	0.0888
53B10	จน	0.1224
53B10	คุย	0.2614
53B10	ซัก	0.1538
53B10	เกี่ยว	0.1293
53B10	การก	0.1001
53B10	องค์	0.1075
53B10	รู้	0.0894
53B10	ทั่วไป	0.0521
53B10	หมาย	0.0657
53B10	แนว	0.1350
53B10	บท	0.1307
53B10	ตัวอย่าง	0.0450
53B10	ฮา	0.1171
53B10	ฉบับ	0.1880
53B10	โอกาส	0.0708
53B10	สอง	0.0682

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B10	สังกัด	0.1171
53B10	สอบสวน	0.1538
53B10	คดี	0.1538
53B10	พิเศษ	0.1171
53B10	ปี	0.0769
53B10	อาจารย์	0.1075
53B10	สอน	0.1307
53B10	คณะ	0.1880
53B10	มุม	0.0940
53B10	เรื่อง	0.1012
53B10	สาขา	0.1307
53B10	สารสนเทศ	0.0414
53B10	สถิติ	0.0708
53B10	บัญชี	0.0844
53B10	พื้นฐาน	0.0592
53B10	รุ่น	0.1307
53B10	น้อง	0.1538
53B10	หน้า	0.0737
53B11	ศึกษา	0.0541
53B11	ประยุกต์	0.0472
53B11	สร้าง	0.0251
53B11	โม	0.6100
53B11	ปี	0.3813
53B11	study	0.0403
53B11	application	0.0307
53B11	business	0.5023

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B11	process	0.6354
53B11	modeling	0.2769
53B11	analysis	0.1818
53B11	tool	0.1271
53B11	management	0.0909
53B11	แข่งขัน	0.1498
53B11	สูง	0.0580
53B11	ผล	0.1226
53B11	การนำ	0.0724
53B11	สารสนเทศ	0.0489
53B11	จน	0.0724
53B11	ควบคุม	0.0776
53B11	งาน	0.0223
53B11	ภายใน	0.1196
53B11	เปรียบเทียบ	0.0997
53B11	ยุ่งยาก	0.1183
53B11	กระบวนการ	0.1818
53B11	สัมพันธ์	0.0806
53B11	โครง	0.0152
53B11	เกี่ยว	0.0382
53B11	แนว	0.0532
53B11	พื้นฐาน	0.0700
53B11	อันดับ	0.1183
53B11	รู้	0.0352
53B11	ตัวอย่าง	0.1064
53B11	เดินทาง	0.1743

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B11	ติดต่อ	0.0532
53B11	สำนัก	0.1545
53B11	สู่	0.0871
53B11	เหมาะสม	0.0806
53B11	โปรแกรม	0.0266
53B11	ร่าง	0.1818
53B11	แผนการ	0.1271
53B11	limited	0.1818
53B11	ทำ	0.0128
53B12	ศึกษา	0.1157
53B12	ประยุกต์	0.0288
53B12	โปรแกรม	0.1625
53B12	study	0.0493
53B12	application	0.0375
53B12	visual	0.6411
53B12	programming	1.3537
53B12	language	0.7228
53B12	without	0.4444
53B12	technology	0.1446
53B12	ทำ	0.0392
53B12	รู้	0.0860
53B12	สร้าง	0.2757
53B12	แอ	0.3188
53B12	งาน	0.0124
53B12	สะดวก	0.0884
53B12	ด่วน	0.2222

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B12	open	0.1553
53B12	source	0.1553
53B12	ฟรี	0.2222
53B12	ค่า	0.0689
53B12	จ่าย	0.0949
53B12	สอง	0.2954
53B12	studio	0.1710
53B12	เกี่ยว	0.0467
53B12	ไซ	0.0581
53B12	ตัวอย่าง	0.0650
53B12	ด้อย	0.1692
53B13	ศึกษา	0.0811
53B13	สร้าง	0.2507
53B13	หนังสือ	1.1549
53B13	study	0.0403
53B13	e	0.7349
53B13	book	1.0909
53B13	creation	0.1545
53B13	โครง	0.0152
53B13	งาน	0.0081
53B13	เกี่ยว	0.0382
53B13	publication	0.3636
53B13	รูป	0.1178
53B13	แฟ้ม	0.3089
53B13	international	0.1818
53B13	forum	0.1545

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B13	ทำ	0.0128
53B13	สื่อ	0.0806
53B13	หน่วย	0.0837
53B13	หา	0.1271
53B13	สรุป	0.1743
53B13	หมาย	0.0776
53B13	ด้วย	0.1384
53B13	ทั่วไป	0.0616
53B13	นิยม	0.1064
53B13	ตัวอย่าง	0.0532
53B13	reader	0.3089
53B13	noble	0.1818
53B13	apple	0.1818
53B13	open	0.6354
53B13	structure	0.1818
53B13	format	0.3089
53B13	container	0.1818
53B13	อาศัย	0.0837
53B13	แหล่ง	0.0806
53B13	หลัก	0.0475
53B13	ไซ	0.0475
53B13	creator	0.1818
53B13	manual	0.1818
53B13	ชื่อ	0.1994
53B13	free	0.3636
53B13	source	0.2542

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B13	ชัดเจน	0.0776
53B13	หรือ	0.0444
53B13	มุม	0.1111
53B13	ถนัด	0.1818
53B13	ผล	0.0307
53B14	ศึกษา	0.0620
53B14	ประยุกต์	0.1082
53B14	ฐ	0.6667
53B14	พาล	0.6667
53B14	สำหรับ	0.0753
53B14	องค์	0.5825
53B14	ฐ	0.1936
53B14	study	0.0370
53B14	management	0.5000
53B14	แปล	0.1269
53B14	สูง	0.0532
53B14	ไซ	0.5229
53B14	สนทนา	0.1165
53B14	chat	0.1667
53B14	หรือ	0.1017
53B14	ติดตาม	0.0871
53B14	ข่าวสาร	0.2168
53B14	เรียน	0.0712
53B14	อ	0.0621
53B14	e	0.1925
53B14	learning	0.1667

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B14	video	0.1667
53B14	ติดต่อ	0.0488
53B14	สื่อสาร	0.0712
53B14	สะดวก	0.0332
53B14	งาน	0.0130
53B14	วิจัย	0.1084
53B14	แหล่ง	0.0738
53B14	โลก	0.1283
53B14	สื่อ	0.0738
53B14	สินค้า	0.0582
53B14	commerce	0.1018
53B14	หน่วย	0.0767
53B14	สถาบัน	0.6345
53B14	สังคม	0.0871
53B14	ศาสนา	0.1667
53B14	เลี้ยง	0.1743
53B14	สร้าง	0.0230
53B14	ทำ	0.0176
53B14	แนว	0.0975
53B14	โครง	0.0419
53B14	เกี่ยว	0.0350
53B14	ลักษณะ	0.0641
53B14	สถิติ	0.1535
53B14	senior	0.1667
53B14	project	0.1165
53B14	load	0.1416

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B14	forum	0.1416
53B14	grade	0.1667
53B14	about	0.1667
53B14	us	0.1667
53B14	อาจารย์	0.1165
53B15	ศึกษา	0.0270
53B15	ประยุกต์	0.1416
53B15	สำหรับ	0.2463
53B15	บ้าน	1.6990
53B15	study	0.0403
53B15	home	1.2356
53B15	interior	1.0909
53B15	design	0.5554
53B15	เจริญ	0.1384
53B15	ก่อ	0.0997
53B15	เล่า	0.1545
53B15	เรื่อง	0.0598
53B15	เรียน	0.0776
53B15	รู้	0.0352
53B15	ศาสตร์	0.1545
53B15	แขนง	0.1818
53B15	ทำ	0.0257
53B15	งาน	0.0061
53B15	สถาปัตยกรรม	0.2542
53B15	ภายใน	0.1196
53B15	ค่า	0.0563

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B15	จ่าย	0.0776
53B15	ซื้อ	0.3498
53B15	เหมาะสม	0.1611
53B15	หรือ	0.0222
53B15	sweet	0.5455
53B15	จำหน่าย	0.2769
53B15	โปรแกรม	0.0532
53B15	สร้าง	0.1003
53B15	เวลา	0.0450
53B15	ตอบ	0.0677
53B15	ปัจจัย	0.1818
53B15	นับ	0.0997
53B15	ผล	0.0307
53B16	ศึกษา	0.0947
53B16	สัมพันธ์	0.8861
53B16	study	0.0403
53B16	relationship	0.1271
53B16	management	0.0909
53B16	compare	0.1818
53B16	function	0.1818
53B16	oracle	0.1271
53B16	แนว	0.1064
53B16	สร้าง	0.0501
53B16	ใหม่	0.0776
53B16	ภายใน	0.0598
53B16	สอดคล้อง	0.1183

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B16	ตอบ	0.0677
53B16	สูง	0.0580
53B16	รอด	0.1384
53B16	ประยุกต์	0.0472
53B16	อาศัย	0.0837
53B16	งาน	0.0182
53B16	ลด	0.0598
53B16	อัตรา	0.1183
53B16	สูญ	0.1271
53B16	ทำ	0.0321
53B16	ขอ	0.2031
53B16	รี	0.1111
53B16	ลี	0.1545
53B16	แอ	0.0290
53B16	สอง	0.1611
53B16	ตัวอย่าง	0.0532
53B16	อ้างอิง	0.1050
53B16	ลักษณะ	0.0700
53B16	เด่น	0.1111
53B16	บกพร่อง	0.1818
53B16	โครง	0.0152
53B16	หรือ	0.0222
53B17	ขนส่ง	0.6092
53B17	อ้อย	2.0000
53B17	smart	0.2000
53B17	ผล	0.2023

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B17	ผลิต	0.8366
53B17	product	0.1398
53B17	แอ	0.0319
53B17	web	0.1315
53B17	application	0.0674
53B17	โครง	0.0167
53B17	งาน	0.0045
53B17	สารสนเทศ	0.1613
53B17	ไว้	0.0824
53B17	รากฐาน	0.1301
53B17	สำหรับ	0.0602
53B17	ประยุกต์	0.0260
53B17	หลัก	0.0523
53B17	โรงงาน	0.6796
53B17	สู่	0.1917
53B17	ทำ	0.0141
53B17	พื้นที่	0.1699
53B17	ปลูก	0.2000
53B17	เวลา	0.0495
53B17	เปิด	0.0824
53B17	ลำดับ	0.1222
53B17	สอดคล้อง	0.1301
53B17	ชื่อ	0.1539
53B17	แปรรูป	0.2000
53B17	ดิบ	0.4000
53B17	น้ำตาล	0.4000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B17	รายงาน	0.0959
53B17	ตรวจ	0.0796
53B17	สอบ	0.0824
53B17	แข่งขัน	0.0824
53B18	สร้าง	0.2757
53B18	โปรแกรม	0.1671
53B18	ประยุกต์	0.1484
53B18	ขอ	0.2128
53B18	online	0.1745
53B18	building	0.2857
53B18	application	0.0482
53B18	multiple	1.1429
53B18	ชื่อ	0.1745
53B18	พี	0.5714
53B18	แพร่หลาย	0.2988
53B18	ศึกษา	0.0425
53B18	สอง	0.3797
53B18	ลักษณะ	0.1099
53B18	นิยม	0.0836
53B18	กรอก	0.2176
53B18	ราย	0.0968
53B18	ตอบ	0.4256
53B18	ค่า	0.0885
53B18	รูป	0.1851
53B18	ใส่	0.2427
53B18	หา	0.0999

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B18	ลบ	0.2857
53B18	ผล	0.1445
53B18	สำรวจ	0.5236
53B18	comma	0.2857
53B18	value	0.2857
53B18	ทำ	0.0101
53B18	งาน	0.0032
53B18	อาศัย	0.1315
53B18	หลัก	0.0747
53B18	สถิติ	0.2631
53B18	ยอด	0.2176
53B18	สัดส่วน	0.2857
53B18	เปอร์เซ็นต์	0.2857
53B18	สูง	0.0912
53B18	วิจัย	0.1859
53B19	ทำ	0.0039
53B19	floor	0.1111
53B19	plan	0.1111
53B19	construction	0.1111
53B19	โลก	0.7696
53B19	ผสวน	1.0000
53B19	เสริม	0.0944
53B19	สื่อ	0.0985
53B19	ศึกษา	0.0331
53B19	สถาปัตยกรรม	0.0777
53B19	เกม	0.0723

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B19	รูป	0.1679
53B19	โม	0.2130
53B19	โปรแกรม	0.0487
53B19	สร้าง	0.0766
53B19	จอ	0.0556
53B19	หรือ	0.0407
53B19	ฉาย	0.1111
53B19	เครื่อง	0.0533
53B19	ไป	0.0679
53B19	เจ	0.0679
53B19	การนำ	0.1326
53B19	สัมผัส	0.0777
53B19	โต้ตอบ	0.0944
53B19	ราย	0.0377
53B19	กล้อง	0.1111
53B19	ลง	0.0442
53B19	สำหรับ	0.0334
53B19	งาน	0.0087
53B19	แพง	0.0944
53B19	สะดวก	0.0442
53B19	ประยุกต์	0.0288
53B19	ทั่วไป	0.0377
53B19	สินค้า	0.0388
53B19	แปลก	0.1111
53B19	ใหม่	0.0474
53B19	ดึงดูด	0.0723

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B19	ไทย	0.0556
53B19	จีน	0.1888
53B19	แพร่หลาย	0.0581
53B19	เทียบ	0.0944
53B19	รู้	0.0215
53B19	maya	0.0944
53B19	ผล	0.0187
53B19	เท	0.0723
53B19	ล	0.0944
53B19	ย	0.1111
53B19	เล็ง	0.0581
53B19	ภายใน	0.0731
53B19	สอง	0.0492
53B20	แอ	0.1594
53B20	industry	0.5000
53B20	organization	0.5000
53B20	performance	0.2124
53B20	application	0.0422
53B20	r	0.2500
53B20	การนำ	0.1990
53B20	เกี่ยว	0.1051
53B20	เงิน	1.3010
53B20	ตาราง	0.1444
53B20	สร้าง	0.2413
53B20	ตัวเลข	0.1904
53B20	ชุด	0.1626

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B20	ชัดเจน	0.1067
53B20	แห่ง	0.7500
53B20	current	0.2500
53B20	อัตรา	0.9758
53B20	สินค้า	0.0874
53B20	inventory	0.2124
53B20	turnover	0.5000
53B20	total	1.0000
53B20	debt	0.2500
53B20	อด	0.1904
53B20	ขาย	0.1030
53B20	profit	0.2500
53B20	margin	0.2500
53B20	ผล	0.1265
53B20	งาน	0.0084
53B20	basic	0.1747
53B20	power	0.1904
53B20	return	0.2500
53B20	คล่อง	0.2500
53B20	หา	0.0874
53B20	ทุน	0.1904
53B20	หรือ	0.0305
53B20	สะดวก	0.0497
53B20	รู้	0.1527
53B20	ภายใน	0.0822
53B20	เสี่ยง	0.1626

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B21	mash	1.0000
53B21	up	1.0000
53B21	application	0.0843
53B21	หา	0.6990
53B21	ราย	0.1694
53B21	หน้า	0.7190
53B21	words	0.5000
53B21	web	0.0822
53B21	search	0.2500
53B21	สะดวก	0.0995
53B21	สร้าง	0.0345
53B21	งาน	0.0223
53B21	ศึกษา	0.0372
53B21	แหล่ง	0.1108
53B21	พาณิชย์	0.1626
53B21	รู้	0.0968
53B21	หนังสือ	0.1444
53B21	เปิด	0.2060
53B21	ฟัง	0.1747
53B21	หรือ	0.1831
53B21	ชม	0.1198
53B21	สื่อ	0.1108
53B21	จอ	0.1250
53B21	หมาย	0.1067
53B21	โอ	0.1527
53B21	แอ	0.1195

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B21	ภายใน	0.0822
53B21	ใหม่	0.1067
53B21	window	0.1904
53B21	โครง	0.0209
53B21	ทำ	0.0265
53B21	แถบ	0.2500
53B21	ไซ	0.0654
53B21	นิยม	0.0731
53B21	ประจำ	0.1067
53B21	เครื่อง	0.1198
53B21	address	0.2500
53B21	เวลา	0.0619
53B21	โอกาส	0.1151
53B22	เม็ด	2.0000
53B22	ผลิต	0.2461
53B22	แป็ง	0.3529
53B22	decision	0.1176
53B22	support	0.0896
53B22	tamarind	0.2353
53B22	seed	0.1176
53B22	production	0.1176
53B22	export	0.1176
53B22	business	0.0542
53B22	kernel	0.1176
53B22	powder	0.1176
53B22	จน	0.0936

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B22	ทา	0.9994
53B22	जान	0.0999
53B22	งาน	0.0170
53B22	หรือ	0.0287
53B22	มั่นคง	0.0896
53B22	สร้าง	0.0324
53B22	มั่นใจ	0.0999
53B22	แม่	0.1176
53B22	ยา	0.1176
53B22	ลด	0.0387
53B22	เวลา	0.0582
53B22	ลง	0.0468
53B22	ผล	0.0397
53B22	โครง	0.0099
53B22	ศึกษา	0.0087
53B22	ปี	0.0588
53B22	บัญชี	0.0645
53B22	โควตา	0.4706
53B22	ราย	0.0797
53B22	เหมาะสม	0.0521
53B22	สารสนเทศ	0.0316
53B22	visual	0.0424
53B22	studio	0.0453
53B22	แอด	0.0563
53B22	โปรแกรม	0.0172
53B22	คิว	0.0765

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B22	server	0.0542
53B22	ที่	0.0719
53B22	c	0.0615
53B22	นิยม	0.0344
53B22	แพร่หลาย	0.0615
53B22	รอง	0.0564
53B22	ครอบ	0.0719
53B22	คลุม	0.0719
53B22	เปิด	0.0485
53B22	คา	0.1176
53B22	ชื่อ	0.0453
53B22	ใหม่	0.0502
53B22	จอ	0.0588
53B23	สำหรับ	0.0502
53B23	คณะ	1.0182
53B23	กรรมการ	2.0000
53B23	สโมสร	2.0000
53B23	online	0.2036
53B23	touch	0.3333
53B23	screen	0.2330
53B23	window	0.2538
53B23	application	0.0562
53B23	บัญชี	0.3656
53B23	ปี	0.3333
53B23	เวลา	0.1650
53B23	นับ	0.5485

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B23	ข้าวโมง	0.1925
53B23	ผล	0.1124
53B23	ประกาศ	0.2832
53B23	ลง	0.2653
53B23	แม่นยำ	0.2538
53B23	ตรวจ	0.1326
53B23	สร้าง	0.0460
53B23	สะดวก	0.0663
53B23	สิทธิ์	0.2832
53B23	ราย	0.1130
53B23	สมัคร	0.4660
53B23	ลด	0.1096
53B23	สิทธิ	0.2036
53B23	สรุป	0.1598
53B23	ชี้	0.3333
53B23	ต่อ	0.2832
53B23	อด	0.2538
53B23	องค์	0.2330
53B23	รู้	0.0645
53B23	หน่วย	0.1535
53B24	ศึกษา	0.0541
53B24	ประยุกต์	0.0472
53B24	กระดาน	1.5229
53B24	รี	0.5554
53B24	โม	0.4357
53B24	study	0.0403

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B24	application	0.0307
53B24	board	0.5455
53B24	remote	0.5455
53B24	infrared	0.1818
53B24	ประจำ	0.0776
53B24	ทำ	0.0577
53B24	งาน	0.0182
53B24	สะดวก	0.0724
53B24	แจ้ง	0.0951
53B24	ใหม่	0.0776
53B24	สื่อ	0.0806
53B24	สอน	0.1545
53B24	มือ	0.0951
53B24	หรือ	0.0222
53B24	สัมผัส	0.1271
53B24	ควบคุม	0.1552
53B24	mouse	0.1545
53B24	สูง	0.0580
53B24	โครง	0.0152
53B24	เกี่ยว	0.0382
53B24	โปรแกรม	0.0798
53B24	สร้าง	0.0501
53B24	สุดท้าย	0.1271
53B24	สถาน	0.0806
53B25	ควบคุม	0.3202
53B25	เสียง	1.1384

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B25	control	0.1626
53B25	voice	0.2124
53B25	speech	1.2500
53B25	application	0.0422
53B25	programming	0.1904
53B25	text	0.5000
53B25	สื่อสาร	0.3202
53B25	ทำ	0.0353
53B25	งาน	0.0223
53B25	ผล	0.0422
53B25	สำเร็จ	0.1747
53B25	อาศัย	0.1151
53B25	mouse	0.2124
53B25	หรือ	0.0305
53B25	หลัก	0.0654
53B25	ลอก	0.2500
53B25	เลียน	0.2500
53B25	สำหรับ	0.1129
53B25	โรงงาน	0.2124
53B25	ฟัง	0.6990
53B25	มือ	0.2614
53B25	ว่าง	0.2500
53B25	สะดวก	0.0497
53B25	สอง	0.2215
53B25	ลักษณะ	0.0962
53B25	ออกเสียง	0.8495

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B25	ธรรมดา	0.7500
53B25	ใส่	0.4247
53B25	โปรแกรม	0.1097
53B25	ประโยค	0.5000
53B25	สารสนเทศ	0.0672
53B25	รายงาน	0.1198
53B25	visual	0.0902
53B25	studio	0.0962
53B25	c	0.1307
53B25	server	0.1151
53B25	สร้าง	0.0345
53B25	ภายใน	0.0822
53B25	สิทธิ์	0.4247
53B25	ลง	0.0995
53B25	ลำดับ	0.1527
53B25	find	0.2500
53B25	play	0.2500
53B25	stop	0.2500
53B25	ชื่อ	0.2742
53B25	ร้อง	0.2500
53B26	แอด	0.1913
53B26	จุด	1.0458
53B26	historical	0.2667
53B26	pictorial	0.1333
53B26	application	0.0225
53B26	point	0.2031

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B26	interest	0.2265
53B26	พี	0.3333
53B26	ไอ	0.2000
53B26	map	0.0932
53B26	สถาน	0.2363
53B26	รุ่น	0.2265
53B26	หลัง	0.0731
53B26	หรือ	0.0651
53B26	ทั่วไป	0.0452
53B26	เลื่อน	0.5333
53B26	กาล	0.1333
53B26	เวลา	0.0990
53B26	ขอ	0.1489
53B26	อ้างอิง	0.0770
53B26	หนังสือ	0.0770
53B26	ชู	0.1333
53B26	ทอง	0.1133
53B26	เล่า	0.1133
53B26	เรื่อง	0.0438
53B26	สมุด	0.1333
53B26	นาวิก	0.1333
53B26	มูล	0.1463
53B26	หลัก	0.0349
53B26	ปี	0.1333
53B26	สู่	0.0639
53B26	ลิงค์	0.1133

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B26	ศึกษา	0.0198
53B26	อาศัย	0.1228
53B26	version	0.1133
53B26	marker	0.0815
53B26	window	0.1015
53B26	การนำ	0.0531
53B26	ชัดเจน	0.0569
53B26	รู้	0.0258
53B26	ประยุกต์	0.0173
53B26	java	0.0770
53B26	script	0.1629
53B26	เจ	0.0815
53B26	คิว	0.0867
53B26	รู้	0.1015
53B26	เก่า	0.1015
53B26	หา	0.0466
53B26	สน	0.1333
53B26	สังคม	0.0697
53B26	สืบ	0.1015
53B26	ภายใน	0.0438
54B27	โปรแกรม	0.1462
54B27	สื่อสาร	0.4981
54B27	หน้า	0.7190
54B27	แอ	0.3188
54B27	screen	0.1165
54B27	communicatio	0.1667

รหัสเอกสาร	คำ	ค่าน้ำหนัก
	n	
54B27	application	0.0281
54B27	platform	0.1165
54B27	ศึกษา	0.0620
54B27	รู้	0.0645
54B27	พื้นฐาน	0.0641
54B27	ทั่วไป	0.1130
54B27	เกี่ยว	0.0350
54B27	หนังสือ	0.0962
54B27	ทำ	0.0412
54B27	ติดต่อ	0.0488
54B27	สร้าง	0.0460
54B27	การนำ	0.0332
54B27	ไซ	0.0436
54B27	ตัวอย่าง	0.0975
54B27	สำหรับ	0.0753
54B27	จอ	0.4167
54B27	ชม	0.1598
54B27	หรือ	0.0407
54B27	เปิด	0.2060
54B27	งาน	0.0167
54B27	แนว	0.0488
54B27	open	0.1165
54B27	source	0.1165
54B27	สะดวก	0.0663
54B27	มาร	0.1667

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B27	สนทนา	0.1165
54B27	ภายใน	0.0548
54B27	java	0.0962
54B27	eclipse	0.1018
54B27	classic	0.1667
54B27	บุ	0.1667
54B27	user	0.1084
54B27	ประยุกต์	0.0216
54B27	สมควร	0.1416
54B28	สำหรับ	0.0463
54B28	โปรแกรม	0.2924
54B28	ประยุกต์	0.1598
54B28	แ	0.0981
54B28	application	0.0259
54B28	ใหม่	0.0657
54B28	ตอบ	0.0573
54B28	สร้าง	0.0212
54B28	สะดวก	0.0306
54B28	ทำ	0.0109
54B28	ศึกษา	0.0572
54B28	เปิด	0.0634
54B28	สังคม	0.0804
54B28	โอ	0.1880
54B28	ค่า	0.1430
54B28	จ่าย	0.1970
54B28	เหมาะสม	0.0682

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B28	eclipse	0.1880
54B28	หนังสือ	0.0888
54B28	สื่อ	0.0682
54B28	มูล	0.0844
54B28	เรียน	0.0657
54B28	รู้	0.0298
54B28	ติดต่อ	0.0450
54B28	งาน	0.0051
54B28	สอดคล้อง	0.1001
54B28	ลักษณะ	0.1184
54B28	ชัดเจน	0.0657
54B29	ตรวจ	0.7959
54B29	สอบ	0.8239
54B29	audit	0.1545
54B29	management	0.0909
54B29	ภายใน	0.0598
54B29	การนำ	0.0724
54B29	เกี่ยว	0.0382
54B29	หรือ	0.0444
54B29	สะดวก	0.0362
54B29	รักษา	0.1111
54B29	ลับ	0.1545
54B29	เข้าถึง	0.0724
54B29	ลำดับ	0.1111
54B29	ผล	0.0307
54B29	งาน	0.0142

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B29	สูญ	0.1271
54B29	สำหรับ	0.0274
54B29	ติดตาม	0.1901
54B29	ควบคุม	0.1552
54B29	สม่ำเสมอ	0.1818
54B29	มั่นใจ	0.1545
54B29	กฎ	0.1818
54B29	ไว้	0.0749
54B29	ทำ	0.0257
54B29	คู่มือ	0.3636
54B29	เหมาะสม	0.0806
54B29	ทบท	0.1111
54B29	ศึกษา	0.0541
54B29	oracle	0.2542
54B29	server	0.0837
54B29	สอง	0.0806
54B29	นิยม	0.0532
54B29	ประยุกต์	0.0236
54B29	กรณี	0.0909
54B29	สถิติ	0.1818
54B30	งาน	0.0223
54B30	กระแส	0.8125
54B30	เจ	0.3818
54B30	ปี	0.2184
54B30	application	0.0105
54B30	ภายใน	0.0616

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B30	เวลา	0.0155
54B30	ปี	0.0312
54B30	ล้ม	0.0625
54B30	เหลว	0.0625
54B30	จน	0.0249
54B30	รอด	0.0476
54B30	แวดล้อม	0.0531
54B30	business	0.0576
54B30	process	0.1311
54B30	สารสนเทศ	0.0336
54B30	หรือ	0.0153
54B30	management	0.0312
54B30	end	0.1250
54B30	ใหม่	0.0267
54B30	พื้นฐาน	0.0240
54B30	ทำ	0.0176
54B30	ลำดับ	0.0382
54B30	สำหรับ	0.0094
54B30	ควบคุม	0.0267
54B30	อัตโนมัติ	0.1311
54B30	โครง	0.0105
54B30	ศึกษา	0.0139
54B30	เกี่ยว	0.0131
54B30	หมาย	0.0267
54B30	แนว	0.0366
54B30	รู้	0.0121

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B30	ประยุกต์	0.0081
54B30	ตัวอย่าง	0.0183
54B30	ร้าน	0.1151
54B30	เสียง	0.0407
54B30	ตัวกลาง	0.0476
54B30	ชื่อ	0.0481
54B30	หลัก	0.0163
54B30	human	0.0531
54B30	task	0.0625
54B30	ก่อ	0.0343
54B30	ลด	0.0205
54B30	ติดตาม	0.0327
54B31	แอ	0.2833
54B31	โม	0.9586
54B31	บาย	1.5229
54B31	application	0.0375
54B31	mobile	0.1553
54B31	platform	0.1553
54B31	มือ	0.3486
54B31	ประจำ	0.0949
54B31	ติดต่อ	0.0650
54B31	สื่อสาร	0.0949
54B31	พื้นฐาน	0.1710
54B31	สร้าง	0.0306
54B31	งาน	0.0050
54B31	ฟัง	0.1553

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B31	เด่น	0.1358
54B31	ทำ	0.0157
54B31	โครง	0.0186
54B31	ศึกษา	0.0992
54B31	รู้	0.0860
54B31	ทั่วไป	0.0753
54B31	เกี่ยว	0.0467
54B31	hardware	0.2222
54B31	tutorial	0.2222
54B31	ตัวอย่าง	0.1300
54B31	สำหรับ	0.0334
54B31	eclipse	0.1358
54B31	java	0.1283
54B31	c	0.2324
54B31	visual	0.0801
54B31	studio	0.0855
54B31	express	0.1692
54B31	ปัจจัย	0.1111
54B31	หรือ	0.0271
54B31	โอกาส	0.2046
54B31	นิยม	0.0650
54B31	สำเร็จ	0.1553
54B32	ศึกษา	0.0248
54B32	เกี่ยว	0.1401
54B32	สารสนเทศ	0.5376
54B32	ไทย	0.2500

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B32	study	0.0370
54B32	งาน	0.0204
54B32	โลก	0.1924
54B32	ทันสมัย	0.1416
54B32	เวลา	0.0412
54B32	ทบท	0.1018
54B32	ทำ	0.0235
54B32	รู้	0.1291
54B32	ใหม่	0.0712
54B32	สูง	0.0532
54B32	โอกาส	0.0767
54B32	โครง	0.0140
54B32	ขอบเขต	0.1416
54B32	สัมพันธ์	0.0738
54B32	ไป	0.1416
54B32	รี	0.2036
54B32	สอบ	0.2746
54B32	สำรวจ	0.1018
54B32	เรื่อง	0.0548
54B32	ไซ	0.0871
54B32	แหล่ง	0.0738
54B32	ราย	0.0565
54B32	จน	0.0663
54B32	สำหรับ	0.0251
54B32	ทั่วไป	0.0565
54B32	สร้าง	0.0230

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B33	ศึกษา	0.1082
54B33	เกี่ยว	0.1528
54B33	สารสนเทศ	0.5376
54B33	สูง	0.5222
54B33	ไทย	0.1818
54B33	study	0.0403
54B33	chief	0.3636
54B33	สังคม	0.0951
54B33	รอบ	0.1183
54B33	ที่	0.1545
54B33	สู่	0.0871
54B33	ห้าง	0.1818
54B33	ร้าน	0.0837
54B33	กลาง	0.1271
54B33	จน	0.0724
54B33	หน่วย	0.0837
54B33	งาน	0.0020
54B33	ผล	0.0307
54B33	วาง	0.0997
54B33	รู้	0.0352
54B33	หรือ	0.1997
54B33	ตำแหน่ง	0.7097
54B33	officer	0.1818
54B33	วุฒิ	0.1818
54B33	แหล่ง	0.1611
54B33	หลัก	0.0951

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B33	core	0.1818
54B33	สรุป	0.0871
54B33	ทั่วไป	0.0616
54B33	ทดสอบ	0.1271
54B33	โปรแกรม	0.0266
54B34	แอด	0.0319
54B34	สำหรับ	0.0903
54B34	ร้าน	0.9208
54B34	ขาย	0.6591
54B34	web	0.0658
54B34	application	0.0337
54B34	flower	0.1699
54B34	shop	0.2000
54B34	เจริญ	0.3046
54B34	รอด	0.1523
54B34	เปิด	0.0824
54B34	หรือ	0.0977
54B34	ทำ	0.0494
54B34	ค่า	0.0620
54B34	เหมาะสม	0.0886
54B34	ย่าน	0.3398
54B34	โลก	0.0770
54B34	สะดวก	0.1592
54B34	ชื่อ	0.4617
54B34	เงิน	0.1301
54B34	การนำ	0.0398

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B34	ไซ	0.4706
54B34	หน้า	0.2876
54B34	สวย	0.1301
54B34	งาม	0.1301
54B34	งาน	0.0067
54B34	ดึงดูด	0.1301
54B34	เดินทาง	0.1917
54B34	ชม	0.1917
54B34	สินค้า	0.1398
54B34	ก่อ	0.1097
54B34	สร้าง	0.0827
54B34	เสื้อ	0.2000
54B34	ผ้า	0.2000
54B34	ตอบ	0.0745
54B34	ปัจจัย	0.2000
54B34	นับ	0.1097
54B34	ผล	0.0337
54B35	สำหรับ	0.2007
54B35	ทำ	0.0471
54B35	tool	0.2330
54B35	สารสนเทศ	0.1792
54B35	ซี	1.2218
54B35	โลก	0.1283
54B35	ค่า	0.1033
54B35	จ่าย	0.1423
54B35	รับประกัน	0.3333

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B35	ศึกษา	0.0992
54B35	สถิติ	0.1535
54B35	พี	0.3333
54B35	กรณี	0.1667
54B35	งาน	0.0037
54B35	ผล	0.0562
54B35	เกี่ยว	0.0700
54B35	ตอบ	0.1241
54B35	ไว้	0.1373
54B36	ตัวแปร	0.2710
54B36	ผล	0.0984
54B36	ตาราง	1.1549
54B36	จอ	0.0417
54B36	screen	0.0582
54B36	table	0.0833
54B36	แข่ง	0.0708
54B36	เวลา	0.0206
54B36	ข่าวสาร	0.0542
54B36	รูป	0.0540
54B36	สะดวก	0.0166
54B36	การนำ	0.0663
54B36	หลัก	0.0218
54B36	แถว	0.0708
54B36	จุด	0.0871
54B36	งาน	0.0056
54B36	เรียน	0.0356

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B36	รู้	0.0484
54B36	พื้นฐาน	0.0321
54B36	เรื่อง	0.0274
54B36	สูง	0.0266
54B36	เป้าหมาย	0.0509
54B36	สร้าง	0.0230
54B36	สื่อสาร	0.0356
54B36	ปัจจัย	0.0417
54B36	ลักษณะ	0.1603
54B36	ตัวเลข	0.2538
54B36	หลัง	0.0914
54B36	ชุด	0.1084
54B36	อักษร	0.1416
54B36	ลำดับ	0.1018
54B36	นิยม	0.0488
54B36	เลข	0.4955
54B36	วิชาการ	0.0833
54B36	ไทย	0.2083
54B36	สัญลักษณ์	0.0833
54B36	ประจำ	0.0356
54B36	ค่า	0.0258
54B36	ไว้	0.0343
54B36	วิจัย	0.2168
54B36	ข้าม	0.0635
54B36	ศึกษา	0.0124
54B36	วัด	0.1084

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B36	โครง	0.0070
54B36	หน่วย	0.0767
54B36	ตอบ	0.0310
54B36	หรือ	0.0102
54B36	สอง	0.0369
54B37	แฉ	0.2869
54B37	สำหรับ	0.1204
54B37	ทำ	0.0706
54B37	คำ	0.6505
54B37	ปลึก	1.0000
54B37	application	0.0337
54B37	retail	0.2000
54B37	business	0.0921
54B37	โครง	0.0670
54B37	งาน	0.0134
54B37	ศึกษา	0.0297
54B37	ตัวอย่าง	0.0585
54B37	ไอ	0.4000
54B37	สภาวะ	0.1699
54B37	แข่งขัน	0.0824
54B37	ราย	0.0678
54B37	หัน	0.1301
54B37	สารสนเทศ	0.0538
54B37	ลด	0.1315
54B37	สร้าง	0.1103
54B37	แบ	0.1699

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B37	สินค้า	0.1398
54B37	หรือ	0.0977
54B37	เข้า	0.2000
54B37	มูล	0.1097
54B37	เวลา	0.0495
54B37	เสถียร	0.2000
54B37	รอง	0.0959
54B37	ติดต่อ	0.1170
54B37	สวย	0.1301
54B37	งาม	0.1301
54B37	ภายใน	0.0658
54B37	x	0.2000
54B37	lion	0.2000
54B37	โปรแกรม	0.0292
54B37	ประยุกต์	0.0519
54B37	การนำ	0.0398
54B37	ไป	0.3666
54B37	โม	0.2876
54B37	ชื่อ	0.2309
54B37	ความถี่	0.2000
54B37	สื่อสาร	0.0854
54B38	ศึกษา	0.0149
54B38	รู้	0.1162
54B38	สารสนเทศ	0.5376
54B38	เกี่ยว	0.1261
54B38	study	0.0222

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B38	people	0.2000
54B38	technology	0.0651
54B38	security	0.0761
54B38	สะดวก	0.0398
54B38	ตอบ	0.0372
54B38	ติดต่อ	0.0293
54B38	สื่อสาร	0.0427
54B38	โยง	0.0651
54B38	ประจำ	0.0427
54B38	ผล	0.0337
54B38	งาน	0.0134
54B38	ข้าม	0.0761
54B38	ตั้ง	0.0523
54B38	รักษา	0.1833
54B38	โอกาส	0.0460
54B38	เป้าหมาย	0.0611
54B38	โครง	0.0084
54B38	ทำ	0.0106
54B38	เรียบเรียง	0.1000
54B38	สำรวจ	0.1833
54B38	การก	0.0651
54B38	เสียง	0.0651
54B38	ไซ	0.0523
54B38	โปรแกรม	0.0292
54B38	รส	0.0761
54B38	โอน	0.1000

รหัสเอกสาร	คำ	ค่าน้ำหนัก
54B38	การนำ	0.0199
53B27	เกม	1.3010
53B27	ตรวจ	0.5969
53B27	game	0.2832
53B27	motion	0.2832
53B27	นิยม	0.0488
53B27	สร้าง	0.0230
53B27	สนุกสนาน	0.1667
53B27	รู้	0.0323
53B27	ออกกำลัง	0.1667
53B27	กาย	0.1667
53B27	แพร่หลาย	0.0871
53B27	เกี่ยว	0.0350
53B27	ผลิต	0.0871
53B27	ส่งเสริม	0.1018
53B27	จำหน่าย	0.1269
53B27	สูง	0.1064
53B27	โครง	0.0140
53B27	งาน	0.0037
53B27	ฉบับ	0.1018
53B27	ทำ	0.0118
53B27	ศึกษา	0.0248
53B27	หา	0.0582
53B27	ชื่อ	0.0641
53B27	สะดวก	0.0332
53B27	จุด	0.0871

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B27	ประยุกต์	0.0216
53B27	โปรแกรม	0.0244
53B27	รา	0.1269
53B28	ศึกษา	0.1487
53B28	แม่นยำ	0.6667
53B28	study	0.0739
53B28	business	0.3069
53B28	intelligence	0.5076
53B28	คลัง	0.2832
53B28	warehouse	0.2538
53B28	ที่	0.8495
53B28	วัด	0.6505
53B28	หลัก	0.1743
53B28	key	0.6667
53B28	performance	0.5663
53B28	dashboard	0.3333
53B28	แข่งขัน	0.4120
53B28	เสมอ	0.3333
53B28	สูง	0.1064
53B28	สมควร	0.2832
53B28	เรื่อง	0.1096
53B28	กลายเป็น	0.4660
53B28	สำหรับ	0.1003
53B28	โอกาส	0.1535
53B28	ใหม่	0.1423
53B28	เปรียบเทียบ	0.1828

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B28	โครง	0.0558
53B28	งาน	0.0074
53B28	เกี่ยว	0.0700
53B28	หมาย	0.1423
53B28	แนว	0.0975
53B28	รู้	0.0645
53B28	แอ	0.0531
53B28	ประยุกต์	0.0433
53B28	การนำ	0.0663
53B28	ตัวอย่าง	0.0975
53B28	กรณี	0.1667
53B28	สร้าง	0.1379
53B28	แหล่ง	0.1477
53B28	รายงาน	0.1598
53B28	รูป	0.0720
53B28	สรุป	0.1598
53B28	ลด	0.1096
53B28	ราย	0.1130
53B28	ทำ	0.0118
53B29	ควบคุม	0.4269
53B29	สินค้า	0.6990
53B29	ขาย	0.2746
53B29	เครื่องประดับ	0.6667
53B29	inventory	0.2832
53B29	control	0.2168
53B29	business	0.0767

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B29	web	0.1096
53B29	application	0.0562
53B29	โครง	0.0140
53B29	งาน	0.0186
53B29	ทำ	0.0059
53B29	แข่งขัน	0.1373
53B29	โลก	0.1283
53B29	เกี่ยว	0.0350
53B29	เรียบร้อย	0.1269
53B29	รากฐาน	0.1084
53B29	สำหรับ	0.0251
53B29	ประยุกต์	0.0216
53B29	รอง	0.2397
53B29	ชื่อ	0.1283
53B29	ผลิต	0.5229
53B29	กึ่ง	0.5000
53B29	สำเร็จรูป	0.6990
53B29	ซูป	0.3333
53B29	เบิก	0.5000
53B29	ติดตาม	0.1743
53B29	ตรวจ	0.3316
53B29	สู่	0.1598
53B29	เข้าถึง	0.1990
53B29	สิทธิ	0.1018
53B29	สังกัด	0.1269
53B29	ร้าน	0.0767

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B30	ศึกษา	0.1487
53B30	คือ	0.5000
53B30	รส	1.1422
53B30	study	0.1109
53B30	business	0.4604
53B30	intelligence	0.7614
53B30	express	1.1422
53B30	ผล	0.2529
53B30	ขอ	0.3724
53B30	online	0.3055
53B30	processing	0.5000
53B30	นับ	0.2742
53B30	วาง	0.8227
53B30	สิ่งแวดล้อม	0.3807
53B30	โอกาส	0.4604
53B30	เหนือ	0.5000
53B30	ยุ่งยาก	0.3253
53B30	เข้าถึง	0.1990
53B30	ค่า	0.4648
53B30	จ่าย	0.2135
53B30	รักษา	0.3055
53B30	ไว้	0.2060
53B30	การนำ	0.0995
53B30	พลาด	0.3055
53B30	ก่อ	0.2742
53B30	สูญ	0.3495

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B30	สารสนเทศ	0.1344
53B30	การก	0.3253
53B30	ล้น	0.5000
53B30	เกี่ยว	0.1051
53B30	หมาย	0.4269
53B30	อันดับ	0.3253
53B30	ทำ	0.0353
53B30	งาน	0.0111
53B30	ไต่	0.2500
53B30	หลัก	0.1307
53B30	โปรแกรม	0.0731
53B30	แรก	0.3253
53B30	สัมพันธ์	0.2215
53B30	มุม	0.3055
53B30	reporter	0.5000
53B30	รายงาน	0.2397
53B30	รูป	0.1079
53B30	สุดท้าย	0.3495
53B30	สรุป	0.2397
53B30	ลด	0.1644
53B30	ราย	0.1694
53B30	สร้าง	0.0689
53B30	โครง	0.0419
53B31	ศึกษา	0.0297
53B31	ประยุกต์	0.0692
53B31	โปรแกรม	0.2924

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B31	รส	0.1015
53B31	study	0.0592
53B31	application	0.0450
53B31	expression	1.2000
53B31	studio	0.2052
53B31	เป้าหมาย	0.0815
53B31	ตอบ	0.0496
53B31	สะดวก	0.0531
53B31	ชุด	0.2602
53B31	ทำ	0.0235
53B31	งาน	0.0163
53B31	รูป	0.0576
53B31	การนำ	0.0265
53B31	มูล	0.0731
53B31	web	0.0438
53B31	blend	0.1333
53B31	design	0.0815
53B31	แอ	0.0850
53B31	วงจร	0.1133
53B31	สวย	0.0867
53B31	งาม	0.0867
53B31	ดีงดูด	0.0867
53B31	โครง	0.0223
53B31	เกี่ยว	0.0280
53B31	สร้าง	0.0184
53B31	ตัวอย่าง	0.0390

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B31	สารสนเทศ	0.0358
53B31	สำหรับ	0.0201
53B31	สถิติ	0.1228
53B31	อาจารย์	0.0932
53B31	เจ้าหน้าที่	0.1133
53B31	โยง	0.0867
52B28	ศึกษา	0.0372
52B28	คอม	1.5229
52B28	โม	0.7190
52B28	การนำ	0.0995
52B28	study	0.0370
52B28	common	0.1667
52B28	warehouse	0.3807
52B28	คลัง	1.2742
52B28	แลกเปลี่ยน	0.2330
52B28	แข่งขัน	0.0687
52B28	อาศัย	0.0767
52B28	ทำ	0.0294
52B28	สารสนเทศ	0.0448
52B28	งาน	0.0056
52B28	ประจำ	0.0712
52B28	สูง	0.0532
52B28	ชุด	0.1084
52B28	ยืดหยุ่น	0.1416
52B28	สร้าง	0.0919
52B28	ย้าย	0.3333

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B28	ขาย	0.0687
52B28	เรื่อง	0.0548
52B28	ยุ่งยาก	0.1084
52B28	ใหม่	0.0712
52B28	รูป	0.0360
52B28	รอง	0.2397
52B28	พลาด	0.1018
52B28	โครง	0.0279
52B28	oracle	0.1165
52B28	builder	0.1269
52B28	power	0.1269
52B28	designer	0.1667
52B28	studio	0.0641
52B28	server	0.0767
52B28	เรียน	0.0712
52B28	รู้	0.0323
52B28	ประยุกต์	0.0216
52B28	ผล	0.0281
52B28	ทั่วไป	0.0565
52B30	การนำ	0.1326
52B30	สินค้า	0.6990
52B30	ประยุกต์	0.0577
52B30	พาณิชย์	0.8674
52B30	product	0.3107
52B30	e	0.3850
52B30	commerce	0.2715

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B30	business	0.1023
52B30	นิยม	0.0650
52B30	แพร่หลาย	0.2324
52B30	ช่อง	0.1162
52B30	ทำ	0.0078
52B30	เข้าถึง	0.0884
52B30	ซื้อ	0.0855
52B30	ราย	0.0753
52B30	หรือ	0.0543
52B30	ชม	0.1065
52B30	ไซ	0.1743
52B30	สร้าง	0.0306
52B30	โครง	0.0186
52B30	งาน	0.0025
52B30	สถิติ	0.1023
52B30	ฉบับ	0.1358
52B30	ศึกษา	0.0331
52B30	โลก	0.0855
52B31	ภายใน	0.2192
52B31	bus	0.3333
52B31	management	0.1667
52B31	แอด	0.0531
52B31	ซี	0.1018
52B31	c	0.1743
52B31	application	0.1124
52B31	server	0.2302

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B31	โม	0.0799
52B31	บาย	0.1269
52B31	mobile	0.2330
52B31	รักษา	0.1018
52B31	สิ่งแวดล้อม	0.2538
52B31	สร้าง	0.0230
52B31	ลด	0.0548
52B31	หัน	0.1084
52B31	สาธารณะ	0.1416
52B31	สะดวก	0.0332
52B31	ติดต่อ	0.0975
52B31	งาน	0.0223
52B31	เจ้าหน้าที่	0.2832
52B31	ปี	0.0833
52B31	การนำ	0.0332
52B31	สารสนเทศ	0.0448
52B31	ทำ	0.0176
52B31	ส่งเสริม	0.1018
52B31	หลัก	0.0436
52B31	เดินทาง	0.0799
52B31	ตาราง	0.1925
52B31	ขับ	0.2538
52B31	สำรวจ	0.1018
52B31	สำหรับ	0.0251
52B31	ควบคุม	0.0712
52B31	visual	0.0601

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B31	studio	0.0641
52B32	ศึกษา	0.0372
52B32	ประยุกต์	0.0162
52B32	พี	1.0000
52B32	study	0.0277
52B32	application	0.0211
52B32	แอด	0.1195
52B32	นิยม	0.0731
52B32	เครื่อง	0.0599
52B32	งาน	0.0153
52B32	สะดวก	0.0497
52B32	ตัวอย่าง	0.0731
52B32	พาณิชย์	0.0813
52B32	ไซ	0.1961
52B32	สินค้า	0.0437
52B32	สนทนา	0.2621
52B32	สอง	0.0554
52B32	ทำ	0.0265
52B32	หรือ	0.0153
52B32	โปรแกรม	0.0183
52B32	ตรงกัน	0.1904
52B32	เรื่อง	0.0411
52B32	การนำ	0.0497
52B32	รูป	0.0270
52B32	ลด	0.0411
52B32	เวลา	0.0309

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B32	เรียน	0.0534
52B32	รู้	0.0484
52B32	โครง	0.0314
52B32	สำหรับ	0.0188
52B32	เด่น	0.0764
52B32	ราย	0.0424
52B32	หลัก	0.0327
52B32	พื้นฐาน	0.0481
52B32	สิทธิ	0.0764
52B32	กระดาน	0.0952
52B32	อี	0.1062
52B33	แฉ	0.0869
52B33	สำหรับ	0.1095
52B33	สร้าง	0.0251
52B33	แฟ้ม	1.0812
52B33	ผล	0.2759
52B33	งาน	0.0223
52B33	ขอ	0.1354
52B33	web	0.1196
52B33	application	0.0613
52B33	folio	0.1818
52B33	resume	0.9091
52B33	online	0.1111
52B33	e	0.2100
52B33	activity	0.1818
52B33	ย่อ	0.7723

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B33	สมัคร	0.2542
52B33	หรือ	0.0888
52B33	ศึกษา	0.0947
52B33	ภายใน	0.2391
52B33	ข้างเคียง	0.2100
52B33	เด่น	0.1111
52B33	มอบ	0.1818
52B33	ทุน	0.1384
52B33	ทำ	0.0321
52B33	ประยุกต์	0.0236
52B33	ราย	0.1232
52B33	ชื่อ	0.0997
52B33	สถาน	0.0806
52B33	รูป	0.0393
52B33	plot	0.1818
52B33	การนำ	0.0362
52B34	โปรแกรม	0.0162
52B34	สัมพันธ์	0.3446
52B34	จรรยา	0.8146
52B34	งาน	0.0223
52B34	relationship	0.0777
52B34	program	0.0944
52B34	traffic	0.1553
52B34	exhibition	0.1111
52B34	event	0.0944
52B34	พี	0.2222

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B34	ไอ	0.1111
52B34	หนาแน่น	0.0846
52B34	แอ	0.0354
52B34	ออ	0.0414
52B34	สถาน	0.1477
52B34	นิยม	0.0325
52B34	ศูนย์	0.2036
52B34	สินค้า	0.0777
52B34	ทอง	0.0944
52B34	ศูนย์การค้า	0.1111
52B34	การนำ	0.0221
52B34	ย้อน	0.0846
52B34	หลัง	0.0609
52B34	รอบ	0.0723
52B34	ไว้	0.0915
52B34	อาศัย	0.1023
52B34	marker	0.0679
52B34	รูป	0.0240
52B34	ประยุกต์	0.0144
52B34	ชัดเจน	0.0949
52B34	ลักษณะ	0.0428
52B34	ราย	0.1506
52B34	เฉลี่ย	0.1553
52B34	average	0.2222
52B34	ผล	0.0562
52B34	เวลา	0.1100

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B34	หัวโม่ง	0.0642
52B34	เกี่ยว	0.0233
52B34	ชื่อ	0.1219
52B34	ลิงค์	0.1888
52B34	เดินทาง	0.0533
52B34	ภายใน	0.0365
52B35	ผล	0.1475
52B35	เสียง	1.3010
52B35	พาณิชย์	0.0813
52B35	voice	0.1062
52B35	commerce	0.0764
52B35	ใจ	0.1747
52B35	ชม	0.2996
52B35	ออกเสียง	0.5309
52B35	สอดคล้อง	0.5692
52B35	โครง	0.0314
52B35	งาน	0.0042
52B35	ไซ	0.2614
52B35	ตัวแปร	0.0813
52B35	สังเคราะห์	0.1250
52B35	สร้าง	0.0172
52B35	หน่วย	0.0576
52B35	ตัวอย่าง	0.0366
52B35	โปรแกรม	0.0183
52B35	ทา	0.6371
52B35	ใจแตก	0.3750

รหัสเอกสาร	คำ	ค่าน้ำหนัก
52B35	น้อย	0.5309
52B35	สถิติ	0.1727
53B37	ศึกษา	0.1062
53B37	ประยุกต์	0.0742
53B37	study	0.0634
53B37	computer	0.2176
53B37	tool	0.1997
53B37	โม	0.4108
53B37	business	0.5262
53B37	process	0.7988
53B37	management	0.1429
53B37	Modeling	0.4351
53B37	แข่งขัน	0.2354
53B37	สูง	0.0912
53B37	จน	0.1137
53B37	งาน	0.0223
53B37	ภายใน	0.0939
53B37	เปรียบเทียบ	0.1567
53B37	การนำ	0.0568
53B37	นับ	0.1567
53B37	ยุ่งยาก	0.1859
53B37	สัมพันธ์	0.1266
53B37	ผล	0.0482
53B37	โครง	0.0239
53B37	เกี่ยว	0.0600
53B37	แนว	0.0836

รหัสเอกสาร	คำ	ค่าน้ำหนัก
53B37	พื้นฐาน	0.1099
53B37	อันดับ	0.1859
53B37	ทั่วไป	0.0968
53B37	model	0.1997
53B37	รู้	0.0553
53B37	ตัวอย่าง	0.0836
53B37	รับส่ง	0.2857
53B37	ชื่อ	0.1099
53B37	สินค้า	0.0999
53B37	ทำ	0.0202



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก จ รายการข้อสอบถามที่ได้จากการกำหนดจากผู้ประเมิน ความเกี่ยวข้อง

รายการข้อสอบถามที่ผู้ประเมินความเกี่ยวข้องได้กำหนดไว้ เพื่อนำมาใช้ในงานวิจัยนี้ ซึ่งนำมาใช้ในการขยายข้อสอบถาม การค้นคืนสารสนเทศ และการประเมินประสิทธิภาพการค้นคืนสารสนเทศ

ตารางที่ จ.1 แสดงรายการข้อสอบถามที่ได้จากการกำหนดจากผู้ประเมินความเกี่ยวข้อง

ลำดับ	ข้อสอบถาม			
1	TWCAQ	2010	HTML5	คนพิการ
	CCS3	เว็บไซต์		
2	Application	แอนดรอยด์	แนะนำ	การแต่งกาย
	SQLite	Eclipse	Java	
3	อาหาร	แอนดรอยด์	Application	แนะนำ
	Eclipse	Database	SQLite	
4	search	google	term	query
	session			
5	Window 8	Application	travel	bangkok
	bing	map		
6	ระดับชั้น	คณิตศาสตร์	ศูนย์กวดวิชา	ผู้เรียน
	ความสามารถ	เว็บ	ระบบสารสนเทศ	
7	voice	command	through	smart
	phone			
8	Cake	Decorating	Web	
9	Mobile	App	Android	teach
	folding	paper		
10	GO	google	app	engine

	cloud	ร้านทอง	web	
11	c#	sql	บรรพชา	เว็บ
	app	asp.net	จุฬาลงกรณ์	
12	ระบบโรงงาน	asp.net	sql server	
13	block	editor	android	design
	component			
14	BI	opensource	DB	
15	IBM	ALM	rational	developer
16	ศึกษา	ไมโครซอฟต์	ไพรวท	คลาวด์
17	Phone	app	html5	java
18	opencv	image	retrieval	template
	matching	c		
19	การเปรียบเทียบ	คุณภาพ	เว็บไซต์	องค์การ
	ท่องเที่ยว	AEC		
20	แอปพลิเคชัน	แอนดรอยด์	เลียนแบบ	เสียงมนุษย์
	eclipse	5 วินาที	ความถี่เสียง	java
21	GPS	สยาม	ร้านค้า	android
22	เข้าถึง	ส่วนบุคคล	คุณภาพ	ระเบียบ
	กฎหมาย			
23	CRM	mobile	ธุรกิจ	
24	BI	คลิกวิว	Qlikview	พัฒนาระบบ

	ธุรกิจอัจฉริยะ			
25	Cognos	mobile	ibm	



ประวัติผู้เขียนวิทยานิพนธ์

นายพีรภัทร กุลประดิษฐ์ เกิดเมื่อวันที่ 21 กรกฎาคม พ.ศ. 2530 ที่จังหวัดนนทบุรี สำเร็จการศึกษาปริญญาตรีวิทยาศาสตร์บัณฑิต จากโครงการขยายโอกาสทางการศึกษา สาขาวิชาการพัฒนาซอฟต์แวร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552 และเข้าศึกษาต่อในหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาการพัฒนาระบบคอมพิวเตอร์ด้านธุรกิจ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี ในปีการศึกษา 2553



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY