



การพัฒนาการสืบค้นข้อมูลทางการแพทย์โดยใช้กฎหมายสัมพันธ์

โดย

นายสุรชัย เทียนสง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

ภาควิชาคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2552

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

การพัฒนาการสืบค้นข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์

โดย

นายสุรชัย เทียนสง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

ภาควิชาคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2552

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

DEVELOPMENT OF MEDICAL SEARCH USING ASSOCIATION RULES

By

Surachai Teansong

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

Department of Computing

Graduate School

SILPAKORN UNIVERSITY

2009

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร อนุมัติให้วิทยานิพนธ์เรื่อง “ การพัฒนาการสืบค้น
ข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์ ” เสนอโดย นายสุรชัย เทียนสง เป็นส่วนหนึ่งของการ
ศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

.....

(รองศาสตราจารย์ ดร.ศิริชัย ชินะตั้งกูร)

คณบดีบัณฑิตวิทยาลัย

วันที่.....เดือน..... พ.ศ.....

อาจารย์ที่ปรึกษาวิทยานิพนธ์

1. อาจารย์ ดร.สุนีย์ พงษ์พินิจภิญโญ

2. เกสัชกร รองศาสตราจารย์ ดร.วีรยุทธ เลิศนที

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.จันทนา ผ่องเพ็ญศรี)

...../...../.....

..... กรรมการ

(รองศาสตราจารย์ ดร.ชวลีรัตน์ จรัสกุลชัย)

...../...../.....

..... กรรมการ

(อาจารย์ ดร.สุนีย์ พงษ์พินิจภิญโญ)

...../...../.....

..... กรรมการ

(เกสัชกร รองศาสตราจารย์ ดร.วีรยุทธ เลิศนที)

...../...../.....

48309329 : สาขาวิชาเทคโนโลยีสารสนเทศ

คำสำคัญ : การสืบค้น/การแพทย์/กฎความสัมพันธ์

สุรัชย์ เทียนสง : การพัฒนาการสืบค้นข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์.

อาจารย์ที่ปรึกษาวิทยานิพนธ์ : อ.ดร.สุณีย์ พงษ์พิณิจภิญโญ และ ภก.รศ.ดร.วีรยุทธ เลิศนที. 94 หน้า.

เนื่องด้วยในปัจจุบัน การสืบค้นข้อมูลทางการแพทย์บนอินเทอร์เน็ตด้วยระบบสืบค้นหรือที่เรียกว่าเสิร์จเอนจินนั้นถือเป็นสิ่งสำคัญ เพราะเป็นเครื่องมือที่ทำให้เราเข้าถึงข้อมูลที่ต้องการได้สะดวกรวดเร็ว แต่ด้วยข้อมูลที่มีปริมาณมากและการระบุคีย์เวิร์ดของผู้ใช้งานทั่วไปที่ขาดความรู้ในด้านการแพทย์และทักษะในการใช้งานเสิร์จเอนจิน ทำให้ไม่สามารถระบุคีย์เวิร์ดที่ชัดเจนให้กับเสิร์จเอนจินทำให้ได้ผลลัพธ์เป็นจำนวนมากและผลลัพธ์ที่ได้อาจไม่เกี่ยวข้องตรงกับความต้องการงานวิจัยนี้จึงได้เสนอแนวทางเพื่อพัฒนารูปแบบการสืบค้นข้อมูลทางการแพทย์ ด้วยการแปลงคีย์เวิร์ดที่ผู้ใช้ในการค้นหาให้เป็นคีย์ที่ชัดเจนเพื่อให้เข้าถึงข้อมูลทางการแพทย์ได้เอกสารที่เกี่ยวข้องตรงกับความต้องการของผู้ใช้โดยใช้ฐานข้อมูลค่าทางความสัมพันธ์ของการเกิดร่วมกันเชิงความถี่ของคำในเอกสารที่เกี่ยวข้องทางการแพทย์ โดยแยกเป็นกลุ่ม โรคที่ชัดเจน

งานวิจัยนี้ได้พัฒนาวิธีการสืบค้นข้อมูลทางการแพทย์ด้วยคำสำคัญที่เกิดขึ้นจากการใช้กฎความสัมพันธ์ ผลการทดลองพบว่าจำนวนเอกสารที่ตรงกับความต้องการของผู้ใช้งานมีจำนวนเพิ่มขึ้น ตัวบ่งชี้ที่ใช้ในการวัดประสิทธิภาพของวิธีการที่นำเสนอ ได้แก่ ค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) โดยค่าเฉลี่ยของค่าความแม่นยำและค่าความครบถ้วนของวิธีการที่นำเสนอจะสูงกว่าวิธีการใช้คำสืบค้นทั่วไปร้อยละ 20.00 และ 39.90 ตามลำดับ

ภาควิชาคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2552

ลายมือชื่อนักศึกษา.....

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ 1..... 2.....

48309329 : MAJOR : INFORMATION TECHNOLOGY

KEY WORDS : MEDICAL SEARCH / ASSOCIATION TECHNIQUE

SURACHAI TEANSONG : DEVELOPMENT OF MEDICAL SEARCH USING ASSOCIATION RULES. THESIS ADVISORS : : SUNEE PONGPINIGPINYO,Ph.D., AND ASSOC.PROF.VERAYUTH LERTNATTEE,Ph.D.. 94 pp.

In relation to the medical profession, at present the use of search engines to retrieve medical information is a desirable and highly important tool, since it provides a rapid access to medical databases and knowledge. However, such information contains a large number of documents and non medical experts may find difficulties choosing the proper keyword for the search engine. As a consequence, without determining a correct or certain keyword, the user may not transform the search into a conceptual search. This study presents a model to develop a medical information retrieval system, by transforming the keyword into certain queries, to enhance performance for the medical information retrieval tasks and increase a relevant and conceptual answer for the user. To achieve this objective, the database was separated into groups of diseases and was structured from the relationship of a word and the frequencies of that word, presented in a medical document.

In this research, a method for searching medical information by suggestion of keywords using association rule, was developed. The result showed that the number of relevant documents, retrieved from a set of related keywords, was increased. The precision and recall were used as performance indicators. The average precision and recall outperformed searching with the simple keywords with gaps of 20.00% and 39.90%, respectively.

Department of Computing Graduate School, Silpakorn University Academic Year 2009
Student's signature

Thesis Advisors' signature 1. 2.

กิตติกรรมประกาศ

ในการวิจัยครั้งนี้สำเร็จลุล่วงไปได้ด้วยดีนั้น ผู้วิจัยต้องขอกราบขอบพระคุณ อาจารย์ที่ปรึกษา อาจารย์ ดร.สุนีย์ พงษ์พินิจภิญโญ อาจารย์ที่ปรึกษาร่วม เกสัชกร รองศาสตราจารย์ ดร.วีรยุทธ์ เลิศนที ประธานกรรมการ รองศาสตราจารย์ ดร.จันทนา ผ่องเพ็ญศรี และกรรมการผู้ทรงคุณวุฒิ รองศาสตราจารย์ ดร.ชวลีรัตน์ จรัสกุลชัย ซึ่งได้ให้คำปรึกษาแนะนำและประสาทวิชา รวมถึงคณาจารย์ภาควิชาคอมพิวเตอร์ และเจ้าหน้าที่ทุกท่าน

สุดท้ายนี้ต้องขอขอบคุณบิดา มารดา รวมถึงภรรยาและบุตร ที่คอยดูแลให้กำลังใจและเป็นแรงผลักดันให้ผู้วิจัยได้ศึกษาต่อจนสำเร็จการศึกษา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง	ฉ
สารบัญภาพ	ญ
บทที่	
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์การวิจัย	2
ประโยชน์ที่คาดว่าจะได้รับ.....	2
ขอบเขตการวิจัย.....	2
เครื่องมือที่ใช้ในการวิจัย.....	2
นิยามศัพท์เฉพาะ	3
2 ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง.....	4
การค้นหากฎความสัมพันธ์	4
การหารากศัพท์.....	6
การกำจัดคำหยุด.....	7
การขยายผลการสอบถาม	8
งานวิจัยที่เกี่ยวข้อง	9
3 วิธีดำเนินการวิจัย.....	12
การเตรียมข้อมูล.....	12
การศึกษาเพื่อหาความสัมพันธ์ของคำ.....	16
การวิเคราะห์และออกแบบระบบ.....	17
การทดสอบระบบ	18
การวิเคราะห์และประเมินผลการทดลอง	18
สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์.....	19
4 ผลการดำเนินงานวิจัย.....	21
การเตรียมข้อมูล.....	21

บทที่	หน้า
การเลือกชุดคำที่มีน้ำหนักคำเชิงความถี่สูง	24
การพิจารณาคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์	26
การสร้างกฎความสัมพันธ์ของคำ.....	28
การออกแบบระบบ	34
การทดสอบระบบ	36
การวิเคราะห์และประเมินผลการทดสอบ	40
สรุปผลการทดสอบ.....	45
5 สรุปผลการดำเนินการวิจัย.....	46
สรุปผลการดำเนินการวิจัย	46
ปัญหาและแนวทางการวิจัยในอนาคต	48
บรรณานุกรม	50
ภาคผนวก	53
ภาคผนวก ก รายการคำหยุดที่ใช้ในงานวิจัย	54
ภาคผนวก ข รายชื่อแหล่งข้อมูลเว็บไซต์ที่ใช้ในงานวิจัย.....	70
ภาคผนวก ค รายการคำที่ผ่านการคัดเลือกโดยผู้เชี่ยวชาญทางการแพทย์.....	75
ภาคผนวก ง คู่มือการใช้งานโปรแกรมช่วยการสืบค้นข้อมูลทางการแพทย์	82
ภาคผนวก จ ตัวอย่างกฎความสัมพันธ์.....	89
ประวัติผู้วิจัย	94

สารบัญตาราง

ตารางที่		หน้า
1	แสดงขั้นตอนและระยะเวลาการดำเนินการวิจัย	12
2	แสดงโรคในกลุ่มโรคระบบหายใจ จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ	13
3	แสดงโรคในกลุ่มโรคระบบไหลเวียนเลือด จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ	14
4	แสดงโรคในกลุ่มเนื้องอก (รวมมะเร็ง) จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ	14
5	แสดงรายละเอียดกลุ่มโรคและจำนวนหน้าเอกสารในแต่ละภาษา	21
6	แสดงจำนวนคำที่ได้จากการคัดเลือกคำเชิงความถี่แยกตามกลุ่มโรคและภาษา...	26
7	แสดงตัวอย่างคำจากการคัดเลือกคำเชิงความถี่กลุ่มโรคระบบหายใจ	26
8	แสดงจำนวนคำจากการคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์	27
9	แสดงจำนวนกฎที่ได้จากการสร้างกฎความสัมพันธ์	33
10	แสดงโครงสร้างฐานข้อมูลกลุ่มโรค	33
11	แสดงโครงสร้างฐานข้อมูลกฎความสัมพันธ์	34
12	แสดงโครงสร้างฐานข้อมูลผู้ใช้	34
13	แสดงคำค้นที่ผู้ทดสอบสุ่มเลือกเพื่อใช้ในการทดสอบ	36
14	แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มโรคระบบหายใจ	37
15	แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มโรคระบบไหลเวียนเลือด	37
16	แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มเนื้องอก (รวมมะเร็ง)	37
17	แสดงกฎที่ผู้ทดสอบเลือกใช้จากฐานข้อมูลคำทางความสัมพันธ์	38
18	แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎทางความสัมพันธ์ กลุ่มโรคระบบหายใจ	39
19	แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎทางความสัมพันธ์ กลุ่มโรคระบบไหลเวียนเลือด	39
20	แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎทางความสัมพันธ์ กลุ่มเนื้องอก (รวมมะเร็ง)	40
21	แสดงผลการวิเคราะห์ผลลัพธ์เปรียบเทียบวิธีธรรมดากับการใช้ฐานข้อมูล กฎความสัมพันธ์	42

สารบัญญภาพ

ภาพที่		หน้า
1	แสดงขั้นตอนการดำเนินการวิจัย.....	20
2	แสดงภาพเอกสารที่ผ่านการกรองข้อมูลและแปลงข้อมูล	22
3	แสดงภาพขั้นตอนการดำเนินการกำจัดคำหยุด การตัดคำ และการหารากศัพท์...	23
4	แสดงเอกสารที่ได้ผ่านกระบวนการตัดคำ.....	23
5	แสดงภาพการจัดกลุ่มเอกสาร	24
6	แสดงการคัดเลือกชุดคำค่าเฉลี่ยโดยใช้พีวเตอร์.....	25
7	แสดงเปรียบเทียบจำนวนคำก่อนและหลังจากผู้เชี่ยวชาญเลือกคำกลุ่มข้อมูล ภาษาไทย.....	27
8	แสดงเปรียบเทียบจำนวนคำก่อนและหลังจากผู้เชี่ยวชาญเลือกคำกลุ่มข้อมูล ภาษาอังกฤษ.....	27
9	แสดงความถี่ของการเกิดคำในกลุ่มข้อมูลภาษาไทย.....	28
10	แสดงความหนาแน่นของคำในกลุ่มข้อมูลภาษาอังกฤษ.....	29
11	แสดงการสร้างกฎความสัมพันธ์ของคำที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาไทย.....	29
12	แสดงการสร้างกฎความสัมพันธ์ของคำที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาไทย.....	30
13	แสดงการสร้างกฎความสัมพันธ์ของคำที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 90 กลุ่มข้อมูลภาษาไทย.....	31
14	แสดงการสร้างกฎความสัมพันธ์ของคำที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาไทย.....	32
15	แสดงการสร้างกฎความสัมพันธ์ของคำที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาไทย.....	32
16	แสดงรูปแบบการทดสอบโดยการนำคำจากฐานข้อมูลคำที่มีความสัมพันธ์มาใช้	35
17	แสดงค่าความแม่นยำและค่าความครบถ้วนที่เพิ่มขึ้นของการสืบค้นการนำ ฐานข้อมูลกฎความสัมพันธ์มาใช้กลุ่มโรคระบบหายใจ.....	43
18	แสดงค่าความแม่นยำและค่าความครบถ้วนที่เพิ่มขึ้นของการสืบค้นการนำ ฐานข้อมูลกฎความสัมพันธ์มาใช้กลุ่มโรคระบบไหลเวียนเลือด.....	44

ภาพที่		หน้า
19	แสดงค่าความแม่นยำและค่าความครบถ้วนที่เพิ่มขึ้นของการสืบค้นการนำ ฐานข้อมูลγκความสัมพันธ์มาใช้กลุ่มเนื้องอก (รวมมะเร็ง).....	44
20	แสดงหน้าจอหลักของระบบการสืบค้นข้อมูลทางการแพทย์.....	83
21	แสดงการระบุคำค้น เต้านม ในกลุ่มเนื้องอก (รวมมะเร็ง).....	84
22	แสดงการเลือกγκที่มีคำว่า เต้านม และ รักแร้.....	84
23	แสดงผลลัษการสืบค้นคำว่า เต้านม และ รักแร้.....	85
24	แสดงหน้าจอเข้าระบบสำหรับผู้เชี่ยวชาญ.....	85
25	แสดงหน้าจอสำหรับเพิ่มγκโดยผู้เชี่ยวชาญ.....	86
26	แสดงหน้าจอสำหรับผู้เชี่ยวชาญเพิ่มγκโดยเลือกกลุ่มโรคที่ต้องการ.....	86
27	แสดงหน้าจอการเพิ่มγκสำหรับผู้เชี่ยวชาญ.....	86
28	แสดงหน้าจอเสร็จสิ้นการเพิ่มγκความสัมพันธ์ของคำโดยผู้เชี่ยวชาญ.....	87
29	แสดงγκที่เพิ่มใหม่โดยมีคะแนนไม่ถึง 3 คะแนน.....	87
30	แสดงγκความสัมพันธ์ของคำที่ผู้เชี่ยวชาญเพิ่มเติม.....	88

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันอินเทอร์เน็ตนับเป็นคลังข้อมูลขนาดใหญ่ที่มีข้อมูลมหาศาลในศาสตร์ทุกแขนง อาทิ ด้านวิชาการ สุขภาพ บันเทิง เทคโนโลยี รวมถึงข่าวสารต่าง ๆ เป็นต้น เนื่องจากการที่มีข้อมูลจำนวนมากและหลากหลายอีกทั้งการจัดเก็บข้อมูลต่าง ๆ ไว้หลากหลายสถานที่ ทำให้การได้มาซึ่งข้อมูลที่เราต้องการและสนใจเป็นไปได้ด้วยความยาก เพราะการเข้าถึงข้อมูลต่างๆ ได้นั้นสิ่งสำคัญคือเราต้องทราบชื่อเรียกที่อยู่ของข้อมูลนั้น ๆ หรือ URL (Uniform Resource Locator) ซึ่งเป็นไปไม่ได้ที่เราจะทราบและจดจำครอบคลุมได้ทั้งหมด ดังนั้นจึงจำเป็นต้องมีเครื่องมือเพื่อช่วยในการเข้าถึงข้อมูลต่างๆ เหล่านั้นได้สะดวกและรวดเร็ว ซึ่งเครื่องมือนั้นก็คือ เสิร์จเอนจิน (Search Engine) ซึ่งเป็นเครื่องมือสำคัญที่ช่วยในการสืบค้นข้อมูลเพื่อให้เราเข้าถึงข้อมูลที่ต้องการได้สะดวกรวดเร็ว และมีความจำเป็นอย่างยิ่งที่จะต้องพัฒนาประสิทธิภาพการทำงานของเสิร์จเอนจินให้สามารถสืบค้นเอกสารที่ผู้ใช้ต้องการและสนใจในรูปแบบของเว็บเพจได้อย่างครบถ้วนและถูกต้องตามความต้องการมากที่สุด

ปัญหาหลักที่เสิร์จเอนจินพบคือ เรื่องของคิวรีที่ระบบค้นหาได้รับมาจากผู้ใช้งาน คิวรีที่มีลักษณะทั่วไปหรือการระบุคีย์เวิร์ดไม่ชัดเจน อาจทำให้ได้รับเอกสารจากการค้นหามากมายและเอกสารส่วนใหญ่ไม่เกี่ยวข้องกับคิวรีนั้นหรือหากมีความเกี่ยวข้องอาจอยู่ในลำดับท้าย ๆ ทำให้เสียเวลาในการตรวจสอบ และอาจมีเอกสารหลาย ๆ ฉบับที่เกี่ยวข้องไม่ถูกค้นพบกรณีที่คิวรีนั้นระบุเฉพาะเจาะจงเกินไปอาจทำให้ไม่ได้รับเอกสารใดๆจากการค้นหาเลย เนื่องมาจากคิวรีนั้นไม่มีคีย์เวิร์ดที่บ่งชี้ไปถึงเอกสารที่ต้องการนั้น

ปัญหาดังที่กล่าวมานั้น อาจกล่าวได้ว่าเป็นปัญหาที่ผู้ใช้งานซึ่งไม่มีความรู้ความเข้าใจในการใช้งานเสิร์จเอนจินพอ อีกทั้งมักพบปัญหาอีกประการหนึ่งที่สำคัญคือผู้ใช้งานนั้นทราบถึงความต้องการของตนแต่ไม่สามารถที่จะสื่อสารออกมาเป็นตัวหนังสือเพื่อระบุเป็นคีย์เวิร์ดในการสืบค้นได้ หลายครั้งเราอาจพบกับปัญหาในการกำหนดคีย์เวิร์ด เพื่อต้องการค้นหาเอกสารนั้น ๆ แต่ไม่รู้จะระบุเช่นไร โดยเฉพาะข้อมูลทางการแพทย์ซึ่งเป็นข้อมูลเฉพาะด้านและมีมิติของข้อมูลเกี่ยวข้องกับหลายโดเมนข้อมูล ผู้ใช้งานจะทราบเฉพาะความต้องการและความรู้สึกที่ตนเองมีและ

ต้องการซึ่งก็จะเป็นคำที่มีลักษณะทั่วไป การใช้คีย์เวิร์ดที่ไม่ชัดเจนเพียงหนึ่งคีย์เวิร์ดจึงเป็นไปได้ยากที่จะได้ข้อมูลจากการสืบค้นตรงความต้องการ

จากเหตุผลดังกล่าว ผู้ทำการวิจัยจึงมีแนวคิดที่จะนำฐานข้อมูลคำที่มีความสัมพันธ์กันที่ได้จากการนำเอกสารหรือเว็บเพจทางการแพทย์ มาหาความสัมพันธ์เพื่อให้เกิดชุดคำที่มีความสัมพันธ์กันในแต่ละคำมาช่วยเพิ่มให้กับคีย์เวิร์ดเดิม เพื่อใช้ในการสืบค้น เพื่อให้ได้เอกสารที่เกี่ยวข้องกับความต้องการของผู้ใช้งานมากยิ่งขึ้น

วัตถุประสงค์การวิจัย

พัฒนาเทคนิคในการสืบค้นข้อมูลทางการแพทย์ด้วยแนวคิดการจัดรูปแบบคิรีโดยใช้คำมาเพิ่มให้กับคีย์เวิร์ดเดิมซึ่งนำมาจากฐานข้อมูลคำที่มีความสัมพันธ์ของการเกิดร่วม ที่ได้จากการเรียนรู้ความสัมพันธ์ของคำที่มีอยู่ในเว็บเพจทางการแพทย์ทั้งภาษาไทยและภาษาอังกฤษ

ประโยชน์ที่คาดว่าจะได้รับ

1. ผู้ใช้งานทั่วไปที่ขาดทักษะการใช้งานเสิร์จเอ็นจินและไม่มีความรู้ทางการแพทย์สามารถใช้งานได้อย่างสะดวก
2. ผู้ใช้ทั่วไปสามารถได้รับเอกสารที่เกี่ยวข้องจากการสืบค้นตรงตามความต้องการมากขึ้น

ขอบเขตการวิจัย

พัฒนาเทคนิคในการสืบค้นข้อมูลทางการแพทย์ด้วยแนวคิดการจัดรูปแบบคิรีโดยใช้คำมาเพิ่มให้กับคีย์เวิร์ดเดิม โดยนำมาจากฐานข้อมูลคำทางความสัมพันธ์ของการเกิดร่วมที่ได้จากการเรียนรู้ความสัมพันธ์ของคำที่มีอยู่ในเว็บเพจทางการแพทย์ทั้งภาษาไทยและภาษาอังกฤษ โดยได้จากการรวบรวมข้อมูลจากเว็บเพจในกลุ่มตัวอย่างข้อมูลทางการแพทย์ 3 กลุ่มโรคที่พบมากที่สุดซึ่งนำมาจากสถิติกลุ่มโรคภาพรวมกรมการแพทย์ ประจำปีงบประมาณ 2550 ซึ่งประกอบด้วย 3 กลุ่มโรคได้แก่ กลุ่มโรกระบบหายใจ กลุ่มโรกระบบไหลเวียนเลือด และกลุ่มเนื้องอก (รวมมะเร็ง) จำนวนรวม 786 เว็บเพจ เพื่อจัดทำฐานข้อมูลคำที่มีความสัมพันธ์ของคำและจัดทำเว็บเสิร์จอินเทอร์เฟซ เพื่อใช้ในการทดสอบเทคนิคดังกล่าว

เครื่องมือที่ใช้ในการวิจัย

1. ฮาร์ดแวร์

1.1 Intel Pentium 4 Core 2 Q 2.63 GHz

1.2 RAM 8 GB

1.3 Hard disk 250 GB

2. ซอฟต์แวร์

2.1 ระบบปฏิบัติการ : Ubuntu 9.04 Server 64 bit

2.2 เครื่องมือในการพัฒนา : Weka , JAVA ,Gnome Datamine, HtmlAsText

โปรแกรมตัดคำ Swath, Portor Stemmer

2.3 ฐานข้อมูล : MySQL

นิยามศัพท์เฉพาะ

1. เสิร์จเอนจิน หมายถึง เครื่องมือสืบค้นที่มีระบบการทำงานโดยใช้โปรแกรมที่เรียกว่า Spider หรือ Web Crawler ซึ่งเป็น Robot ชนิดหนึ่ง ท่องไปในเว็บเพจต่าง ๆ โดยวิ่งไปตามลิงก์ของแต่ละเว็บไซต์ เพื่อไปยังอีกเว็บไซต์หนึ่ง เพื่ออ่านข้อมูลและจัดเก็บเว็บเพจที่พบเข้าสู่ฐานข้อมูล ทำให้ฐานข้อมูลมีขนาดใหญ่ และแสดงผลลัพธ์จากการสืบค้น

2. คีย์เวิร์ด หมายถึง คำหรือข้อความที่ผู้ใช้อินเทอร์เน็ต นิยมพิมพ์ลงไปเพื่อใช้ในการค้นหาเว็บไซต์ และข้อมูลต่าง ๆ ตามความต้องการของผู้ที่ใช้อินเทอร์เน็ต เพื่อความสะดวก และความรวดเร็วในการสืบค้นเว็บเพจที่อยู่ในฐานข้อมูลของเสิร์จเอนจิน

3. การสืบค้น หมายถึง กระบวนการหรือการกระทำใดๆ ในการค้นหาสารสนเทศที่กระทำโดยเสิร์จเอนจินทำให้ได้สารสนเทศที่ต้องการอย่างรวดเร็ว ถูกต้องแม่นยำ ผลจากการสืบค้นเป็นข้อมูลหรือรายการเอกสาร

4. ความสัมพันธ์ หมายถึง ความสัมพันธ์กันของข้อมูลที่ได้จากการค้นหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ โดยใช้เทคนิคการค้นหาความสัมพันธ์ของข้อมูล (Association rule Discovery) ซึ่งเป็นเทคนิคหนึ่งของ Data Mining ที่สำคัญ

5. ค่าสนับสนุน หมายถึง ร้อยละของการดำเนินการที่กฎสามารถนำไปใช้ หรือ เป็นร้อยละของการดำเนินการที่กฎที่ใช้มีความถูกต้อง

6. ค่าความเชื่อมั่น หมายถึง จำนวนของกรณีที่กฎถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ได้

บทที่ 2

ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ ผู้วิจัยได้แบ่งออกเป็นหัวข้อต่าง ๆ ดังนี้

1. การค้นหากฎความสัมพันธ์ (Association Rule Discovery)

การค้นหากฎความสัมพันธ์ คือ การค้นหากฎความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ ที่มีอยู่เพื่อช่วยในการวิเคราะห์และตัดสินใจในธุรกิจตัวอย่างของการค้นหากฎความสัมพันธ์ เช่น การวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า “ Market Basket Analysis ” Jiawei Han and Micheline Kamber (2001 : 225) เป้าหมายของการค้นหากฎความสัมพันธ์ คือ จะแยกและดึงสิ่งที่ซ่อนไว้ในฐานข้อมูลได้อย่างไร และจะหารายการ (Item) ใน Transaction เดียวกันได้อย่างไร ซึ่งจะสามารถบอกได้ว่ารายการใดที่มีแนวโน้มที่จะสามารถพบด้วยกันใน Transaction เดียวกันการใช้เทคนิคการค้นหากฎความสัมพันธ์นอกจากจะนำไปประยุกต์ใช้กับการวิเคราะห์การซื้อสินค้าของลูกค้าแล้วยังสามารถมาประยุกต์ใช้กับด้านต่าง ๆ เช่น

ด้านการศึกษา กฤษณะ ไวยมัย, ชิตชนก ส่งศิริ และธนาวินท์ รักธรรมานนท์ (2001 : 139-141) ได้นำมาประยุกต์ใช้ในการช่วยทำนายเกรดรายวิชาต่างๆในภาคเรียนต่อไปของนิสิตคณะวิศวกรรมศาสตร์ โดยหาความสัมพันธ์ของผลการเรียนในแต่ละวิชาที่ส่งผลต่อกัน ซึ่งทำให้ได้ว่าวิชาใดบ้างที่มีผลต่อวิชาที่ต้องการจะทำนายเกรดล่วงหน้า โมเดลจะทำนายโดยอ้างอิงจากข้อมูลเดิมของนิสิตที่เคยเรียนมาและได้ผลการเรียนเช่นเดียวกับนิสิตคนนั้น

ด้านกฎหมาย กฤษณะ ไวยมัย และธีระวัฒน์ พงษ์ศิริปรีดา (2001 : 143-152) ได้ใช้เทคนิคการค้นหากฎความสัมพันธ์และเทคนิค Data Classification มาประยุกต์ใช้เพื่อการจัดสรรกฎหมายที่เหมาะสมกับการพิจารณาคดีความ โดยนำเทคนิค Data Classification มาสร้างตัวจำแนกข้อมูลจากกฎเกณฑ์ที่ได้จากเทคนิค การค้นหากฎความสัมพันธ์ ตัวจำแนกข้อมูลสามารถนำไปใช้ทำนายคดีความแต่ละคดีว่าควรใช้กฎหมายฉบับใดในการพิจารณา โดยใช้ข้อมูลคดีความของศาลฎีกาซึ่งประกอบด้วยคดีอาญา และคดีแพ่ง และใช้การตัดคำด้วยพจนานุกรมภาษาไทยเพื่อแบ่งคดีความให้เป็นวลีสั้น ๆ โดยใช้เทคนิคการค้นหากฎจากไฟล์ดัชนีและการหากฎเกณฑ์และชุดทดสอบแบ่งเป็น 2 ชุด คือ ระดับกฎหมาย และระดับมาตรา ผลการทดลองได้ผลลัพธ์ในการทำนายที่ดีกว่าการใช้ Data Classification แบบทั่วไปซึ่งอยู่ในรูปของต้นไม้ช่วยตัดสินใจ (Decision Tree)

รูปแบบของการค้นหากฎความสัมพันธ์

รูปแบบทั่วไปของการค้นหากฎความสัมพันธ์ คือ $A \rightarrow B$

โดยที่ A : เป็นเงื่อนไข และ

B : เป็นผลลัพธ์ที่เกิดขึ้นหรืออยู่ในรูปของ “ถ้า.....แล้ว”

(If.....Then....) เช่น

$A \rightarrow B$; if A Then B เป็นกฎที่ 1

$B \rightarrow A$; if B Then A เป็นกฎที่ 2

การประเมินค่าของกฎจะใช้ค่าสนับสนุน(Support) และค่าความเชื่อมั่น(Confidence) โดยที่

ค่าสนับสนุน คือ เปอร์เซ็นต์ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้อง ตามกฎต่อจำนวนข้อมูลทั้งหมด สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าสนับสนุน(A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction ทั้งหมด}}$$

โดยที่ A หมายถึง เหตุการณ์ที่ใช้เป็นเงื่อนไขในการหาผลลัพธ์

B หมายถึง เหตุการณ์ที่เป็นผลลัพธ์

Transaction (A,B) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A และ B

ค่าความเชื่อมั่น คือเปอร์เซ็นต์ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมดที่เป็นเงื่อนไข สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าความเชื่อมั่น (A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction (A)}}$$

โดยที่ Transaction (A) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A อย่างเดียว ในการเลือกกฎใดนั้นจะต้องพิจารณาค่าสนับสนุน และค่าความเชื่อมั่นที่มีค่าสูงกว่าค่า Threshold ที่ตั้งไว้ นอกจากนี้จะต้องกำหนดค่าสนับสนุนต่ำสุด (Minimum Support) และค่าความเชื่อมั่นต่ำสุด (Minimum Confidence)

2. การหารากศัพท์

Stemming เป็นการหา Stem อันหมายถึง รากศัพท์ คำหลัก หรือต้นตอของศัพท์ ซึ่งในภาษาไทยนั้นคงไม่ต้องมีการหารากศัพท์แต่ประการใดเพราะไม่ได้มีการแจกวิภัติตามกาลหรือตามโครงสร้างของประโยค (ศุภชัย ตั้งวงศ์สานต์ 2551)

ในข้อความมักจะมีการแจกวิภัติต่าง ๆ เหล่านี้ปะปนอยู่ซึ่งเป็นเรื่องปกติของการเขียนเรียงความ แต่เมื่อนำคำเหล่านี้ไปสร้างเป็นกรณีโดยตรงโดยขาดความเข้าใจของตัวภาษาจะเกิดผลเสียอยู่หลายประการ คือทำให้สิ้นเปลืองพื้นที่ในเรื่องของการทำดัชนี โดยปัญหาอีกประเด็นที่สำคัญคือ การสืบค้นคำ ก็จะได้คำที่ตรงตามความต้องการได้น้อยทั้งที่เป็นคำเดียวกัน ซึ่งการทำ stemming มีอยู่หลายวิธีที่สำคัญได้แก่

- Table Lookup
- Affix Removal
- Successor Variety

Porter's Stemmer Algorithm

Algorithm นี้แบ่งเป็น 5 กฎใน 5 ขั้นตอน คำในข้อความที่ต้องการหาตัวรากศัพท์ให้ไล่จากกฎข้อที่ 1 ไปยังกฎข้อที่ 5 ตามลำดับไม่สามารถเริ่มด้วยการแทรกระหว่างกลางหรือข้ามขั้นตอนได้ ตัวคำสั่งใน Algorithm เขียนเป็น Pseudo Code ที่อ่านเข้าใจได้โดยการใช้สัญลักษณ์ที่กำหนดดังนี้ (ศุภชัย ตั้งวงศ์สานต์ 2551)

- สัญลักษณ์ C แทนตัวแปรพยัญชนะซึ่งหมายถึงอักษรในภาษาอังกฤษทุกตัว ยกเว้น a, e, i, o, u และตัว y หากอักษรตัวหน้ามันเป็นพยัญชนะ

- สัญลักษณ์ V แทนตัวแปรสระ ซึ่งหมายถึงอักษรทุกตัวที่ไม่ใช่พยัญชนะ
- สัญลักษณ์ L แทนอักษรที่ไม่เจาะจงจะเป็นพยัญชนะหรือสระก็ได้
- สัญลักษณ์ ϕ แทน string ที่ว่างเปล่า

- การเขียนสัญลักษณ์ต่าง ๆ ที่มี C, V และ L รวม ๆ กัน ใช้เป็นตัวกำหนด pattern
- สัญลักษณ์ * ใช้แทนการซ้ำ 0 ครั้งหรือหลาย ๆ ครั้งของตัว pattern ที่กำหนด
- สัญลักษณ์ + ใช้แทนการซ้ำ 1 ครั้งหรือหลาย ๆ ครั้งของตัว pattern ที่กำหนด
- บรรทัดที่ขึ้นต้นด้วย % ถือว่าเป็น comment

ดังนั้นตามสัญลักษณ์ที่กำหนดข้างต้น การเขียนนิพจน์ (C)* จะหมายถึงตัวพยัญชนะที่เรียงลำดับมีซ้ำกัน 0 ครั้งหรือหลายครั้ง ในขณะที่ ((V)* (C))* หมายถึงตัวสระที่เรียงลำดับมีซ้ำกัน

0 ครั้งหรือหลายครั้งตามด้วยพยัญชนะที่เรียงลำดับมีซ้ำกัน 0 ครั้งหรือหลายครั้ง ซึ่งทั้งหมดสามารถปรากฏ 0 ครั้งหรือหลาย ๆ ครั้งได้แต่หากเขียนว่า (V* C) จะหมายถึงตัวเริ่มต้นเป็นสระมีซ้ำกัน 0 ครั้ง หรือหลาย ๆ ครั้งและตามด้วยพยัญชนะปิดท้าย

3. การกำจัดคำหยุด

การกำจัดคำหยุด (Stop-Word List Removal) เป็นการนำคำที่ไม่มีนัยสำคัญออก โดยที่ไม่ทำให้ความหมายของเอกสารเปลี่ยนแปลงคำที่ไม่มีนัยสำคัญในที่นี้หมายถึง คำที่ใช้กันโดยทั่วไปไม่มีความหมายสำคัญต่อเอกสาร เมื่อตัดออกจากเอกสารแล้วไม่ทำให้ใจความของเอกสารเปลี่ยนแปลง ประเภทของคำที่จัดว่าเป็นคำหยุดในภาษาไทย (ณิชาพร สุระ 2541 : 44 - 60) มีดังต่อไปนี้

3.1 คำบุพบท เป็นคำที่ใช้เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กัน มักใช้นำหน้าคำนาม คำสรรพนาม คำกริยา หรือคำวิเศษณ์ เพื่อแสดงความสัมพันธ์ของคำหรือกลุ่มคำที่อยู่หลังคำบุพบทว่ามีความสัมพันธ์กับคำหรือกลุ่มคำที่อยู่หน้าคำบุพบทอย่างไร ตัวอย่างของคำบุพบท ได้แก่ ของ ใน แก่ แต่ ต่อ ตั้งแต่ โดย เมื่อ กว่า กับ เป็น คู่ก่อน ซ้ำแต่ ทาง คู่ แก่ ฯลฯ

3.2 คำสันธาน เป็นคำที่ทำหน้าที่เชื่อมคำกับคำ ประโยคกับประโยค ข้อความกับข้อความ เพื่อทำให้ประโยคนั้นกระชับ และสละสลวย โดยมักจะใช้เชื่อมใจความที่คล้ายคลึงกัน ใจความที่ขัดแย้งกัน ใจความที่เป็นเหตุเป็นผลกัน หรือใจความที่ให้เลือกร้อยอย่างใดอย่างหนึ่ง ตัวอย่างของคำสันธาน ได้แก่ เพราะ เพราะว่า และ หรือ จึง ดังนั้น มิฉะนั้น ทั้ง แต่ แต่ว่า ครั้น หรือ ไม่ก็ ฯลฯ

3.3 คำสรรพนาม เป็นคำที่ใช้แทนคำนามที่กล่าวถึงมาแล้วในประโยค เพื่อไม่ต้องกล่าวคำนามนั้นซ้ำอีกตัวอย่างของคำสรรพนาม ได้แก่ ฉัน เรา เขา ดิฉัน กระผม กู คุณ ท่าน เธอ ได้เท่ามัน สิ่ง ใครบ้าง ต่างก็ ตัวนั้น อันโน้น อะไร ไหน ไฉ อย่างไร ฯลฯ

3.4 คำวิเศษณ์ เป็นคำที่ใช้ขยายคำอื่น เช่น คำนาม คำสรรพนาม คำกริยา หรือคำวิเศษณ์ เพื่อให้มีความหมายชัดเจนและมีรายละเอียดมากยิ่งขึ้น ตัวอย่างของคำวิเศษณ์ ได้แก่ มาก น้อย ใหญ่ เล็ก มหึมา มโหฬาร อ้วน โต สูง ไพเราะ เยอะ หลาย สวย หอม นุ่ม เผ็ด เช้า คำ นี้ นั้น เอง ทั้งนี้ ค่ะ ครับ ขอรับ จำ จ๊ะ วะ ไม่ หามิได้ บ่ ทำไม อย่างไร หมด อคิด ปัจจุบัน โบราณ หน้า ฯลฯ

3.5 คำอุทาน เป็นคำที่แสดงอารมณ์ อาการ หรือความรู้สึกของผู้พูด รวมทั้งเป็นคำที่ใช้เสริมคำพูดตัวอย่างของคำอุทาน ได้แก่ เอ๊ะ อ๊ะ อ้อ เฮ้อ ว้าย ไร้ โถ อนิจจา สิ หนอย ชะ นะ นำ แหม ตายละ คุณพระช่วย ชิชะ อุวะ ไม่น่าเลย โอ้โฮ ฯลฯ คำหยุดมักเป็นคำที่ปรากฏขึ้นบ่อยครั้งในเอกสาร และปรากฏในเอกสารเกือบทุกฉบับ จึงถือได้ว่าคำหยุดเป็นคุณลักษณะที่ไม่เกี่ยวข้องหรือไม่มีประโยชน์ในการค้นคืนหรือการจำแนกหมวดหมู่ ดังนั้นการกำจัดคำหยุดจึงเป็นกระบวนการที่

ควรทำก่อนการจัดทำดัชนี เพื่อกำจัดคุณลักษณะที่ไม่เป็นประโยชน์ และลดขนาดของดัชนีลง ซึ่งจะช่วยประหยัดทั้งพื้นที่และเวลาในการประมวลผล

4. การขยายผลการสอบถาม

การขยายผลการสอบถาม (Query Expansion) เป็นขั้นตอนสำหรับเสิร์จเอนจินที่มีการเพิ่มเติมส่วนคำสำหรับค้นหาขึ้นกับผู้ใช้เป็นคนให้คำนำหน้าในการค้นหา จุดมุ่งหมายเพื่อปรับปรุงความแม่นยำ และหรือค่าระลึกลับ คำค้นเพิ่มเติมนี้ได้จากพจนานุกรมที่รวมคำศัพท์ที่มีความหมายใกล้เคียง พจนานุกรมอื่นๆ เป็นต้นและได้จากเอกสารที่ผู้ใช้กำหนดไว้ว่าเกี่ยวข้อง นอกจากนี้คำค้นเพิ่มเติมยังสามารถให้คำนำหน้าเป็นบวกรหรือลบได้ด้วย ส่วนใหญ่คำเพิ่มเติมนี้เพิ่มให้กับควรีที่ปรากฏอยู่ การขยายผลการสอบถามนี้เป็นเทคนิคสำหรับให้นำหนักคำค้นใหม่เพื่อประสิทธิภาพการค้นคืนสารสนเทศ

เทคนิคการขยายผลการสอบถามมี 3 แบบ ดังนี้ (รัชนีกร เทวอักษร 2546 : 36) (ศุภชัย ตั้งวงศ์สานต์ 2551)

4.1 Manual query expansion (MQE) เป็นเทคนิคที่ผู้ใช้ดัดแปลงควรีเองโดยที่ระบบไม่ได้ช่วยผู้ใช้เลย

4.2 Automatic query expansion (AQE) เป็นเทคนิคที่ดัดแปลงควรีโดยที่ปราศจากการควบคุมของผู้ใช้ ตัวอย่างเช่น ระบบจะเพิ่ม thesaurus term ให้กับควรี

ปัญหา 3 ประการของ Automatic query expansion

1. จะเลือกหัวข้อที่เหมาะสมได้อย่างไร
2. จะเลือกคำนำหน้าของหัวข้อที่เลือกได้อย่างไร
3. จำนวนเท่าไรของหัวข้อที่จะถูกเพิ่มเติมเข้ากับควรีเดิม

มีการศึกษาเกี่ยวกับ Automatic query expansion มานานถึง 30 ปี และมีหลายวิธีที่ถูกนำมาศึกษา แบ่งได้เป็น 6 ลักษณะดังนี้

การขยายผลการสอบถามโดยคำนึงถึงโครงสร้างประโยคเกี่ยวข้องกับ linguistic knowledge และสถิติที่เกี่ยวข้อง วิธีนี้ใช้หลักไวยากรณ์และพจนานุกรมในการสกัดรายชื่อของเทอมสำหรับแต่ละควรีจะถูกขยายโดยการเพิ่มเติมเทอมที่มีความใกล้เคียงมากที่สุดวิธีนี้ดีกว่าควรีเดิม เล็กน้อย

การขยายผลการสอบถามด้วยข้อมูลที่เกี่ยวข้องข้อมูลที่เกี่ยวข้องถูกใช้ในการสร้างสารสนเทศโดยรวม เช่น thesaurus หรือ minimum spanning tree ควรีจะถูกขยายโดยใช้โครงสร้างสารสนเทศโดยรวม ปัญหาสำคัญคือข้อมูลที่เกี่ยวข้องไม่ได้มีไว้ให้ใช้อย่างสม่ำเสมอ

การจัดแบ่งประเภทของเทอมอัตโนมัติความคล้ายคลึงระหว่างเทอมนั้นได้ถูกนำมาคำนวณเป็นอันดับแรกบนพื้นฐานของสมมติฐานที่เกี่ยวข้องและจากนั้นจะจัดเทอมโดยใช้ similarity threshold value ซึ่งคิวรีจะถูกขยายโดยการเติมเทอมทั้งหมดของกลุ่มของคำซึ่งวิธีการนี้เป็นวิธีหนึ่งที่มีประโยชน์

การจัดแบ่งประเภทของเอกสาร เอกสารจะถูกจัดกลุ่มก่อนเป็นอันดับแรกและคำที่พบไม่บ่อยในกลุ่มของเอกสารจะถูกแบ่งกลุ่มในเทอมและ thesaurus class เดียวกัน คิวรีจะถูกขยายเพิ่มเติมโดยการแทนที่ด้วยคำโดย thesaurus class หรือ โดยการเพิ่ม thesaurus ในดัชนี อย่างไรก็ตาม สมรรถนะของการค้นคืนขึ้นอยู่กับค่าที่สำคัญบางรายการที่มีความยากที่จะวัดได้วิธีนี้ค่อนข้าง มีราคาสูง

ขยายผลการสอบถามด้วยฐานข้อมูลรูปภาพ บัญชีคำที่คล้ายคลึงกันนั้นเป็น คำ กับคำที่มีเมตริกที่คล้ายคลึงกันซึ่งเป็นคำที่ถูกสร้างมาจากคำที่มีในนั้นถูกทำเป็นดัชนีอย่างไร วิธีนี้ถูกใช้ในการคาดคะเนความน่าจะเป็นของคำที่เหมือนกันของคิวรีใน vector space model วิธีนี้ทำให้สมรรถนะของการค้นคืนดีขึ้นอย่างมีนัยสำคัญ

ขยายผลการสอบถามด้วยคำกับวลีที่เกี่ยวข้องกับบัญชีคำถูกสร้างโดย identifying phrase ในตัวอักษรและแสดง โดยใช้เทอมที่เกี่ยวข้องกับคำเหล่านั้นอย่างมาก วิธีนี้สามารถ simulate ทุกชนิดของ term-based association thesauri เช่น term VS term , term VS noun term , term VS verb term และ term VS adjective, adverb term association thesauri คิวรีถูกขยายโดยการเพิ่มวลีที่มีความสัมพันธ์อย่างมากกับคิวรีจากบัญชีคำ การค้นพบนี้ทำให้ สมรรถนะของการค้นคืนดีขึ้นอย่างมีนัยสำคัญ

4.3 Interactive query expansion (IQE) เป็นวิธีที่ผู้ใช้มีส่วนร่วมบางประการกับระบบในขบวนการขยายผลการสอบถาม วิธีนี้เป็นส่วนหนึ่งในค่าความเกี่ยวข้องป้อนกลับ

งานวิจัยที่เกี่ยวข้อง

รุจิรา อาชวานันทกุล และยีน ภูววรรณ (2546 : 44) ได้พัฒนารูปแบบการสืบค้นใหม่โดยใช้หลักการของความสัมพันธ์ทางความหมายของคำจากฐานความรู้รูปภาพ ที่จัดเก็บในโครงสร้างข้อมูลที่มีประสิทธิภาพในการเรียกค้นสูง นอกจากนี้ยังช่วยประหยัดเวลาในการคัดเลือกเอกสารที่ผู้ใช้ต้องการจากชุดเอกสารผลลัพธ์ เนื่องจากเอกสารผลลัพธ์ที่ถูกต้องจะถูกนำมาแสดงในอันดับแรกๆ ของชุดเอกสารผลลัพธ์และผลลัพธ์ที่ได้จากการสืบค้นจะมีความครบถ้วนและความถูกต้องสูงขึ้นเมื่อเปรียบเทียบกับเทคนิคในการสืบค้นของเสิร์จเอนจินทั่วไป โดยมีความถูกต้องเพิ่มขึ้น โดยเฉลี่ย 15-36 เปอร์เซ็นต์ และมีความครบถ้วนเพิ่มขึ้น โดยเฉลี่ย 40.75-66.13 เปอร์เซ็นต์

รัชนิกร เทวอักษร (2546:36) ได้พัฒนาเทคนิคการขยายคำค้นเพื่อเพิ่มประสิทธิภาพให้กับระบบสืบค้นเอกสาร โดยมีวัตถุประสงค์เพื่อเพิ่มความถูกต้องและความครอบคลุมให้ระบบสืบค้นเอกสารภาษาไทยด้วยคุณสมบัติเด่นของระบบการขยายคำค้นภาษาไทย โดยเน้นการเพิ่มประสิทธิภาพของระบบโดยอิงฐานข้อมูลพจนานุกรม LEXITRON ซึ่งใช้ความสัมพันธ์ 3 รูปแบบคือ คำเหมือน คำเกี่ยวข้อง และคำแปล (สำหรับสืบค้นข้ามภาษา) ระบบการขยายคำค้นนี้ประกอบด้วยส่วนสำคัญ 2 ส่วนคือ ส่วนความสามารถในการสืบค้นและส่วนความสามารถในการเลือก ในส่วนของความสามารถในการสืบค้นนั้นแบ่งออกเป็น 4 ส่วนคือ ส่วนนิพจน์ตรรกะแบบบูลีน ส่วนขยายคำค้นโดยอาศัยฐานความรู้พจนานุกรม ส่วนแปลคำค้นและส่วนการวัดค่าความเหมือน โดยมีวิธีการวัดประสิทธิภาพ 2 ระบบ ได้แก่ การวัดโดยใช้ค่าความถูกต้อง และวิธีการวัดความพึงพอใจโดยการให้คะแนนเปรียบเทียบผลการทดสอบปรากฏว่า ระบบการขยายคำค้นนี้สามารถดึงเอกสารที่ตรงและเกี่ยวข้องกับคำค้นออกมาได้มากและถูกต้อง ซึ่งทั้งนี้ขึ้นอยู่กับค่าที่ตั้งค่า threshold กล่าวคือ ตั้งค่า threshold ถ้าตั้งค่าต่ำอาจทำให้ได้เอกสารที่ไม่ตรงกับความต้องการมาก แต่ในทางตรงกันข้ามถ้าตั้งค่า threshold สูง ระบบขยายคำค้นก็ทำงานไม่ได้เต็มที่ ซึ่งชื่อเฉพาะ คำค้นที่ไม่เกี่ยวข้องกับเอกสาร การขาดความสัมพันธ์ของคำในแบบ generic และ specific และระบบตัดคำ

Fonseca and others (2005 : 696-703) ได้พัฒนาเทคนิคการจัดรูปแบบคิวรีแบบใหม่ที่กำหนดคิวรีที่ชัดเจนขึ้นให้กับเสิร์จเอนจินด้วยวิธีการวิเคราะห์และค้นหาซ้ำ ๆ ในความสัมพันธ์ของคิวรีแบบพิเศษนี้ ซึ่งได้จากการตรวจสอบความสัมพันธ์จากนั้นนำมาสร้างรูปแบบความสัมพันธ์เพื่อระบุเป็นผังความคิดความสัมพันธ์ ซึ่งจะแสดงแนวคิดที่ความสัมพันธ์กับคิวรีปัจจุบันและแสดงให้เห็นและเลือกแนวความคิดที่ต้องการและมีความหมายที่สัมพันธ์กับคิวรีนั้น แนวคิดนี้เป็นวิธีการขยายคิวรีโดยได้คิวรีใหม่แทนที่คิวรีเดิม โดยให้เลือกใช้งาน ซึ่งการทำวิธีการนี้จะสามารถเพิ่มค่าความถูกต้องแม่นยำจากเดิม 32 เปอร์เซ็นต์ เพิ่มขึ้นเป็น 52 เปอร์เซ็นต์

Nguyen and others (2003 : 156-165) ได้เสนอทฤษฎีในการค้นหาความหมายของวลีที่ถูกซ่อนจากเว็บเพจใน meta tags และ titles page ด้วยเฟรมเวิร์คใหม่ซึ่งใช้ตัวอย่างลำดับสูงสุดที่เข้ากันได้กับวลีที่ซ่อนอยู่ด้วยการใช้คำค้นบนเสิร์จเอนจิน จากนั้นนำกลุ่มของวลีที่เกิดร่วมกันที่มีความถี่สูงสุดจากเว็บเพจด้วยวิธีการ clustering และ association rule ร่วมกันค้นหาและเรียนรู้คำจำกัดความของวลีนั้น เริ่มต้นด้วยการใช้ข้อมูลที่มีปริมาณมาก ประมาณ 1000 วลีต่อวลีที่ถูกซ่อน 1 วลี จากนั้นทำการลดปริมาณวลีลงด้วยการใช้งานร่วมด้วยการใช้ clustering ในการแบ่งคำจำกัดความของวลีและหาความสัมพันธ์ของวลีเหล่านั้นด้วย association rule

Navigli and others (2003 : 22-31) ได้ประยุกต์ใช้เทคนิคนี้ในการสร้างกฎเพื่อใช้ในการ

ตัดสินใจสร้างความสัมพันธ์แบบเฉพาะเจาะจง (Specific relation) ระหว่างคำหลัก และคำขยาย จากเอกสารในโดเมนการท่องเที่ยว ได้มีการกำหนดความสัมพันธ์ที่สนใจไว้ 10 ความสัมพันธ์ เช่น Time (Afternoon<-Time<-Tea), Matter (ceramics<-Matter<-tile) หลังจากทีระบบได้ใช้วิธีทางสถิติ คำนวณหาโมโนภาพที่ต้องการแล้วจะนำคำที่ได้มาหาความสัมพันธ์ที่กำหนด โดยในการเรียนรู้จะ เรียนรู้จากตัวอย่างที่แสดงในรูปแบบของ feature vector ของโมโนภาพที่ความสัมพันธ์กันและ โมโนภาพเหล่านี้จะถูกกำกับด้วยความหมายที่เป็นไปได้ทั้งหมดในบริบทนี้และลิสต์ของคำ hypernym ซึ่งได้จากเวิร์ดเน็ตเมื่อนำเข้าไปเรียนรู้ด้วยวิธีของต้นไม้ช่วยตัดสินใจ (Decision tree) จะ ได้ผลลัพธ์ที่เป็นกฎเพื่อใช้ในการพิจารณาความสัมพันธ์ของโมโนภาพ

Maedche and Staab (2000 : 321-325) ได้มีการประยุกต์ใช้เทคนิคการหาความสัมพันธ์สัมพันธ์ซึ่งเป็นวิธีการหนึ่งในการทำดาต้ามายนิ่งมาใช้ในการหาความสัมพันธ์ โดยมองว่า คำ คือ ไอเท็มที่ต้องการนำมาค้นหา และประโยคที่คู่คำเกิดด้วยกันคือทรานเซคชันที่สนใจ และได้มีการนำ ความรู้พื้นฐานจากเวิร์ดเน็ตเพื่อดึงคำที่มีความหมายโดยทั่วไปกว้างกว่าคำที่พิจารณาเพิ่มเข้าไปใน ทรานเซคชันเพื่อให้คำสนับสนุนที่ได้มีค่าสูงขึ้น เป็นการนำคำสถิติการเกิดคู่กันของโมโนภาพมาใช้ ในการหาความสัมพันธ์ ซึ่งไม่ใช่ความสัมพันธ์ในเชิงความหมาย (semantic relation)

Abdelmgeid Amin Aly (2008 : 343 - 348) ใช้เทคนิคการขยายแบบสอบถามด้วยวัตถุ ประสงค์ที่จะหาเทอมที่เหมาะสมเพิ่มเติม งานวิจัยนี้ได้นำเสนอ 2 วิธีของการขยายผลแบบสอบถาม ที่ประยุกต์ใช้ในลำดับขั้น การทดลองกระทำโดย 3 กลุ่มข้อมูลทดสอบ คือ CISI, NPL และ CACM ที่มีขนาดที่แตกต่างกันและชนิดของเอกสารที่แตกต่างกัน การประเมินสมรรถนะของการค้นคืนโดย หาค่าเฉลี่ยของความแม่นยำในการวัด ซึ่งความแม่นยำหมายถึงสัดส่วนของเอกสารที่เกี่ยวข้องการ ค้นคืนต่อจำนวนที่เกี่ยวข้องทั้งหมด ผลการทดลองแสดงให้เห็นว่า วิธีที่พัฒนาขึ้นนี้มีประสิทธิภาพที่ สูงขึ้นของการค้นคืน เมื่อมีการประยุกต์ใช้การขยายผลแบบสอบถามนี้ ทั้งในด้านขนาดของคอเรคชัน และจำนวนของเทอมการสืบค้นเมื่อเปรียบเทียบกับควิรีแบบต้นตำรับ ใน vector space model แบบ ปกติ

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคการหาความสัมพันธ์ของการเกิดร่วมของคำที่มีอยู่ในเว็บเพจทางการแพทย์ โดยนำผลที่ได้มาใช้ในการพัฒนาการสืบค้นข้อมูลทางการแพทย์บนอินเทอร์เน็ตโดยใช้เสิร์จเอนจิน ซึ่งได้กำหนดขั้นตอนและระยะเวลาการดำเนินการวิจัยดังแสดงในตารางที่ 1

ตารางที่ 1 แสดงขั้นตอนและระยะเวลาการดำเนินการวิจัย

ลำดับที่	ขั้นตอนการดำเนินการ	ระยะเวลา (เดือน)												
		1	2	3	4	5	6	7	8	9	10	11	12	
1	เตรียมข้อมูล	←————→												
2	ศึกษาเพื่อหาค่าความสัมพันธ์ของคำ				←————→									
3	วิเคราะห์ผลและออกแบบระบบ						←————→							
4	ทดสอบระบบ								←————→					
5	วิเคราะห์และประเมินผลการทดสอบ									←————→				
6	สรุปผลการวิจัยและจัดทำรายงาน วิทยานิพนธ์												←————→	

จากขั้นตอนและระยะเวลาดำเนินการวิจัยมีรายละเอียดดังนี้

1. การเตรียมข้อมูล

1.1 การคัดเลือกข้อมูล (Data Selection) งานวิจัยนี้ได้ดำเนินการเก็บรวบรวม ข้อมูลในกลุ่มข้อมูลทางการแพทย์ในรูปแบบเอกสารเอชทีเอ็มแอล ซึ่งมีเนื้อหาทั่วไปเกี่ยวกับโรค อาทิ ความหมาย อาการ การรักษา เป็นต้น จำนวน 786 เพจ ซึ่งนำมาจากเว็บไซต์ทางการแพทย์ หน่วยงานทางการแพทย์ และองค์กรที่ไม่หวังผลทางการค้า ทั้งในและต่างประเทศ โดยแยกเป็นภาษาไทยและภาษาอังกฤษ นำมาพิจารณาสร้างฐานข้อมูลคำที่มีความสัมพันธ์ซึ่งเนื้อหาประกอบด้วยกลุ่มโรคที่มีจำนวนผู้ป่วยมากที่สุดตามสถิติข้อมูลด้านการป่วยของผู้ป่วยนอกแยกตามสาเหตุ 21 กลุ่มโรค

(กรมการแพทย์ 2550) จำนวน 3 กลุ่มโรค ได้แก่ กลุ่มโรคระบบหายใจ กลุ่มโรคระบบไหลเวียนเลือด และกลุ่มเนื้องอก (รวมมะเร็ง)

จากสถิติข้อมูลด้านการป่วยดังกล่าว ในแต่ละกลุ่มโรคนั้นประกอบด้วยรายโรคต่าง ๆ ที่มีผู้ป่วยมากที่สุดในแต่ละกลุ่มตามจำนวนผู้ป่วยในตามรายโรค (75 รายโรค/กลุ่มโรค) ในภาพรวม (กรมการแพทย์ 2550) สามารถแยกตามกลุ่มโรคได้ ดังนี้

กลุ่มโรคระบบหายใจ ประกอบด้วย ปอดอักเสบ (Pneumonia) ระบบหายใจส่วนบน ติดเชื้อเฉียบพลัน และโรคอื่นของระบบหายใจส่วนบน และโรคหืด (Asthma) ซึ่งโรคที่อยู่ในกลุ่มของระบบหายใจส่วนบนติดเชื้อเฉียบพลัน และโรคอื่นของระบบหายใจส่วนบนนั้นประกอบด้วย ไข้หวัด (Common Cold) คออักเสบ (Pharyngitis) และ ไซนัส (Sinusitis) ดังนั้นรายโรคที่อยู่ในกลุ่มโรคระบบหายใจที่นำมาดำเนินการวิจัย จึงประกอบด้วย 5 รายโรค ดังแสดงรายละเอียดชื่อโรค จำนวนแหล่งข้อมูล และจำนวนเว็บเพจ ดังแสดงในตารางที่ 2

ตารางที่ 2 แสดงโรคในกลุ่มโรคระบบหายใจ จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ

ชื่อโรค	ภาษาอังกฤษ		ภาษาไทย	
	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ
โรคปอดอักเสบ	7	34	8	8
โรคไข้หวัด	8	34	11	14
โรคคออักเสบ	9	15	7	7
โรคไซนัส	7	21	15	18
โรคหืด	8	48	16	37
รวม	152		84	

กลุ่มโรคระบบไหลเวียนเลือด ประกอบด้วย 3 รายโรค คือ โรคความดันโลหิตสูง (Hypertension) โรคหัวใจขาดเลือด (Heart Attack) และโรคหลอดเลือดในสมอง (Stroke) ดังแสดงในตารางที่ 3

ตารางที่ 3 แสดงโรคในกลุ่มโรกระบบไหลเวียนเลือด จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ

ชื่อโรค	ภาษาอังกฤษ		ภาษาไทย	
	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ
โรคความดันโลหิตสูง	9	28	6	11
โรคหัวใจขาดเลือด	6	27	15	17
โรคหลอดเลือดในสมอง	6	21	13	15
รวม	76		43	

กลุ่มเนื้องอก (รวมมะเร็ง) ประกอบด้วยเนื้องอกร้าย 4 ชนิด คือ เนื้องอกร้ายที่เต้านม (Breast Cancer) เนื้องอกร้ายที่ปากมดลูก (Cervical Cancer) เนื้องอกร้ายที่ปอด (Lung Cancer) และเนื้องอกร้ายที่ตับ (Liver Cancer) ดังแสดงรายละเอียดในตารางที่ 4

ตารางที่ 4 แสดงโรคในกลุ่มเนื้องอก (รวมมะเร็ง) จำนวนแหล่งข้อมูลและจำนวนเว็บเพจ

ชื่อโรค	ภาษาอังกฤษ		ภาษาไทย	
	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ	จำนวนแหล่งข้อมูล	จำนวนเว็บเพจ
เนื้องอกร้ายที่เต้านม	5	123	10	30
เนื้องอกร้ายที่ปากมดลูก	5	83	12	28
เนื้องอกร้ายที่ปอด	5	103	9	17
เนื้องอกร้ายที่ตับ	5	66	9	11
รวม	375		86	

1.2 การกรองข้อมูล (Data Cleaning) เนื่องจากเอกสารเอชทีเอ็มแอลที่ได้ดำเนินการรวบรวมมานั้นยังไม่สามารถนำมาดำเนินการในขั้นตอนต่อไปได้ เนื่องจากเอกสารดังกล่าวนั้นมีเนื้อหาที่ไม่ต้องการเป็นจำนวนมากที่ส่งผลกระทบต่อการใช้งาน อาทิ ชื่อเมนู ข้อความประชาสัมพันธ์ เป็นต้น ข้อความดังกล่าวมีจำนวนมากและทุกเอกสาร ดังนั้นจำเป็นต้องขจัดข้อความดังกล่าวออก

หรือดึงข้อมูลที่ต้องการออกมาเท่านั้น นอกจากนี้ในส่วนของเนื้อหาในเอกสารที่ต้องการยังประกอบด้วยอักขระพิเศษและโครงสร้างภาษาเอสทีเอ็มแอล ดังนั้นการดำเนินการในขั้นนี้จำเป็นต้องการองข้อมูลที่ไม่ต้องการหรือคัดเลือกข้อมูลที่เป็นที่ต้องการออกมามากที่สุด

1.3 การแปลงข้อมูล (Data Transformation) ในงานวิจัยนี้มีการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมเพื่อดำเนินการหลายขั้นตอน อาทิ การแปลงข้อมูลจากเอกสารเอสทีเอ็มแอลให้อยู่ในรูปแบบเอกสารข้อความ และการแปลงข้อมูลเพื่อเข้าสู่กระบวนการหาความสัมพันธ์ของคำ

1.4 การตัดคำในภาษาไทย เป็นพื้นฐานที่สำคัญในการวิเคราะห์และประมวลผลภาษา และการพัฒนาระบบที่เกี่ยวข้อง เนื่องจากภาษาไทยนั้นเป็นภาษาที่มีลักษณะการเขียนที่ไม่มีการใช้ตัวอักษรหรือสัญลักษณ์อื่นใดขึ้นระหว่างคำ โดยลักษณะการเขียนอยู่ในรูปแบบของคำอยู่เรียงติดกันเป็นประโยค งานวิจัยนี้จำเป็นที่จะต้องดำเนินการตัดคำเพื่อแยกคำต่าง ๆ ออกจากกัน เพื่อนำคำที่เกิดขึ้นนั้นเข้าสู่กระบวนการหาความสัมพันธ์

1.5 การกำจัดคำหยุด (Stop-Word List Removal) คือเป็นการนำคำที่ไม่มีนัยสำคัญออก โดยที่ไม่ทำให้ความหมายของเอกสารเปลี่ยนแปลง คำที่ไม่มีนัยสำคัญ ในที่นี้หมายถึง คำที่ใช้กันโดยทั่วไปไม่มีความหมายสำคัญต่อเอกสาร เมื่อตัดออกจากเอกสารแล้วไม่ทำให้ใจความของเอกสารเปลี่ยนแปลง ซึ่งคำหยุดที่ใช้ประกอบด้วยคำหยุดภาษาไทย (ฉิชาพร สุระ 2541 : 44 - 60) จำนวน 1335 คำ และภาษาอังกฤษ (Pubmed Stopword List 2009) จำนวน 364 คำ รวมทั้งสิ้น 1699 คำ ซึ่งคำหยุดนั้นมักเป็นคำที่ปรากฏขึ้นบ่อยครั้งในเอกสารและปรากฏในเอกสารเกือบทุกฉบับ จึงถือได้ว่าคำหยุดเป็นคุณลักษณะที่ไม่เกี่ยวข้องหรือไม่มีประโยชน์ในการค้นคืน

1.6 การหารากศัพท์ (Stemming) เนื่องจากแหล่งข้อมูลที่นำมาใช้ในงานวิจัยที่เป็นภาษาอังกฤษนั้นเป็นภาษาที่มีการเปลี่ยนคำเดิมเป็นคำอื่นๆ ด้วยการเติม ตัด หรือเปลี่ยนแปลงโครงสร้างประโยคหรือตัวอักษรท้ายเพิ่มเติม ซึ่งคำต่างๆ เหล่านี้มีความหมายมาจากรากศัพท์เดียวกัน ทำให้มีผลต่อการวิจัยอย่างยิ่งเพราะเนื่องจากงานวิจัยนี้เกี่ยวข้องกับการให้น้ำหนักของคำที่มีความถี่ของคำที่ปรากฏ ซึ่งหากคำที่มีรากศัพท์เดียวกันแต่ยังมีได้ดำเนินการหารากศัพท์ที่มีปริมาณมากในแต่ละเอกสาร จะทำให้คำที่มีความสำคัญหรือน่าสนใจอื่นๆ จะไม่สามารถปรากฏเพื่อนำมาใช้งานได้ โดยอัลกอริทึมที่นำมาใช้ในงานวิจัยนี้คือการหารากศัพท์ของ Porter (Snowball 2009)

1.7 การคัดเลือกชุดคำ เนื่องจากเนื้อหาที่อยู่ในเว็บเพจนั้นมีจำนวนมาก ถึงแม้เราจะทำการลดด้วยการกำจัดคำหยุดและการหารากศัพท์ ปริมาณคำในแต่ละกลุ่มข้อมูลยังคงมีปริมาณมากซึ่งหากนำคำทั้งหมดที่ได้มาดำเนินการหาความสัมพันธ์ก็จะได้กฎที่เกิดขึ้นมีจำนวนมาก จึงจำเป็นต้องคัดเลือกชุดคำ โดยการให้ค่าน้ำหนักของคำโดยใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร

(Document Frequency) ซึ่งหมายความว่าคำดังกล่าวพบได้ในหลายเอกสาร ซึ่งในงานวิจัยนี้ใช้วิธีการหาค่าเฉลี่ยของคำที่เกิดขึ้นต่อจำนวนเอกสารทั้งหมด โดยหาได้จากสมการดังนี้

$$\text{ค่าเฉลี่ย} = \frac{\text{จำนวนเอกสารที่พบคำนั้น}}{\text{จำนวนเอกสารทั้งหมด}}$$

งานวิจัยนี้ได้จัดกลุ่มข้อมูลออกเป็นกลุ่มตามรายโรค อาทิ กลุ่มเนื้องอก (รวมมะเร็ง) แบ่งออกเป็น 4 กลุ่มรายโรคซึ่งผู้วิจัยได้กำหนดจำนวนการคัดเลือกชุดคำไว้ที่กลุ่มละ 70 คำซึ่งมีความเป็นไปได้ที่จะเกิดขึ้นคือเกิดคำที่ซ้ำกันระหว่างกลุ่มข้อมูล และค่าน้ำหนักของคำที่เท่ากันทั้งในกลุ่มรายโรคเดียวกันและระหว่างกลุ่มรายโรค ซึ่งในกรณีดังกล่าวนี้ในงานวิจัยนี้จะนับคำที่เกิดขึ้นในกรณีดังกล่าวเพียงหนึ่งคำเท่านั้น

1.8 การคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์ เนื่องจากคำที่ได้จากการคัดเลือกชุดคำโดยการให้ค่าน้ำหนักของคำโดยใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร (Document Frequency) คำที่ได้นั้นมีทั้งคำที่สื่อความหมายและคำที่ไม่สื่อความหมายใดหรืออาจไม่มีความสำคัญในทางการแพทย์ซึ่งยากต่อการนำไปใช้งาน ดังนั้นขั้นตอนนี้เป็นกรคัดเลือกคำโดยผู้เชี่ยวชาญซึ่งประกอบด้วยเภสัชกร จำนวน 3 ท่าน โดยการเลือกคำดังกล่าวใช้วิธีการคัดเลือกโดยใช้หลักเกณฑ์ในการเลือกคือ คำที่สามารถนำไปใช้งานได้หรือมีประโยชน์ โดยคำนั้นจะต้องมีผู้เชี่ยวชาญเลือกอย่างน้อย 2 ท่านจาก 3 ท่าน

2. การศึกษาเพื่อหาความสัมพันธ์ของคำ

การนำเอาคำที่ได้ดำเนินการจัดเตรียมสมบูรณ์แล้ว นำมาหาค่าความสัมพันธ์ของคำที่ปรากฏว่าคำใดมีความสัมพันธ์กับคำใด หรือกล่าวอีกนัยคือ หากพบคำใดนี้แล้วจะพบคำอีกคำ หรืออีกหลายคำปรากฏอยู่ด้วย ซึ่งในงานวิจัยนี้ เลือกใช้เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule Discovery) ซึ่งเป็นวิธีที่เหมาะสมกับข้อมูลที่น่ามาใช้ เพราะต้องการทราบความสัมพันธ์ของคำต่าง ๆ นั้นสัมพันธ์กันอย่างไร เพื่อนำมาใช้เป็นคำช่วยสืบค้นเพิ่มให้คำค้นเดิมของผู้ใช้ โดยกำหนดค่าความเชื่อมั่นต่ำสุดและค่าสนับสนุนต่ำสุดในช่วงที่เหมาะสมกับข้อมูล ซึ่งทฤษฎีการค้นหากฎความสัมพันธ์นั้นมี ดังนี้

ศึกษาการค้นหากฎความสัมพันธ์ โดยมีรูปแบบของการค้นหากฎความสัมพันธ์สามารถเขียนได้ดังนี้

$$A \rightarrow B$$

โดยที่ A เป็นเงื่อนไข และ B เป็นผลลัพธ์ที่เกิดขึ้น การค้นหากฎความสัมพันธ์ทั้งหมดจะต้องมีค่าสนับสนุนมากกว่าค่าสนับสนุนต่ำสุดที่กำหนดไว้ และมีค่าความเชื่อมั่นมากกว่าค่าความเชื่อมั่นต่ำสุดที่กำหนดไว้ ในงานวิจัยนี้ผู้จัดทำวิจัยต้องการค้นหากฎความสัมพันธ์ของคำที่เกิดขึ้น ซึ่งรูปแบบเงื่อนไขและผลลัพธ์ที่เกิดขึ้นนั้น สามารถเขียนกฎความสัมพันธ์ได้ดังนี้

item1 item2 \rightarrow item3	Support% Confidence%
Item (คำ)	คือคำที่เป็นเงื่อนไขและผลลัพธ์
Support (ค่าสนับสนุน)	คือเปอร์เซ็นต์ของจำนวนข้อมูลที่มีสมาชิกสอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด
Confidence (ค่าความเชื่อมั่น)	คือเปอร์เซ็นต์ของจำนวนข้อมูลที่สอดคล้องตามกฎต่อจำนวนข้อมูลที่เป็นเงื่อนไข

ในงานวิจัยนี้เป็นการสร้างความสัมพันธ์ของคำซึ่งค่าสนับสนุนต่ำสุดและค่าความเชื่อมั่นต่ำสุดที่กำหนดนั้นมีผลต่อจำนวนคำที่นำมาสร้างกฎและจำนวนกฎที่เกิดขึ้น ดังนั้นการกำหนดค่าดังกล่าวต้องเหมาะสมกับข้อมูลแต่ละข้อมูลซึ่งอาจแตกต่างกัน

3. การวิเคราะห์และออกแบบระบบ

เมื่อได้ผลจากการหาค่าความสัมพันธ์ของคำมาแล้ว ผลที่ได้มานั้นจะอยู่ในรูปแบบของความสัมพันธ์ในรูปแบบ $A \rightarrow B$ เช่น จากการทดลองในกลุ่มโรคระบบหายใจ อาจพบว่า ไอ อักเสบ \Rightarrow เสมหะ กล่าวได้ว่าหากพบคำว่า ไอ อักเสบ แล้วนั้นเรามักจะพบคำว่า เสมหะ โดยงานวิจัยนี้สนใจที่จะนำคำที่เกิดขึ้นในกฎมาใช้ทั้งกฎในรูปแบบคำรวมที่เกิดขึ้นด้วยกันเท่านั้นโดยไม่พิจารณาคำที่เกิดขึ้นก่อนที่ส่งผลทำให้เกิดอีกคำหนึ่ง จากตัวอย่างคำว่า ไอ อักเสบ เสมหะ หากผู้ใช้งานต้องการเลือกคำใดคำหนึ่งในกฎก็จะพบคำที่อยู่ในกฎเดียวกันด้วย จากนั้นนำคำหรือชุดคำ

ที่ได้จากการหาความสัมพันธ์ สร้างเป็นฐานข้อมูลคำที่มีความสัมพันธ์และจัดทำเว็บไซต์อินเทอร์เน็ตขึ้นมาเพื่อใช้ในการทดสอบ

4. การทดสอบระบบ

การทดสอบระบบคือ ส่วนของการนำคำหรือชุดคำที่ได้จากการหาค่าความสัมพันธ์ของคำมาใช้ในการสืบค้น โดยผู้ทดสอบจำนวน 12 คน ทดลองใช้คำคีย์เวิร์ดคนละ 1 คำค้นหาการสืบค้นด้วยวิธีธรรมดา และเทคนิคใหม่ที่นำฐานข้อมูลคำที่มีความสัมพันธ์มาใช้โดยทดสอบกฎที่มีจำนวนคำในกฎ 2 และ 3 คำ ผู้ทดสอบจะเป็นผู้ใช้ชุดพินิจตัดสินใจว่าผลลัพธ์ที่ได้จากการสืบค้นใน 50 เอกสารแรก และระบุว่าเอกสารผลลัพธ์นั้น ถูกต้อง, ไม่ถูกต้อง, ไม่แน่ใจ อย่างใดอย่างหนึ่ง

5. การวิเคราะห์และประเมินผลการทดสอบ

การวัดความแม่นยำ(precision)และความครบถ้วน(recall) (รุจิรา อาชวานันทกุล และ ยืน ภู่วรรณ 2546 : 44) ใช้วิธีเปรียบเทียบชุดเอกสารจากการสืบค้น โดยนำฐานข้อมูลคำทางความสัมพันธ์มาใช้ กับชุดเอกสารจากสืบค้นด้วยวิธีธรรมดา โดยกำหนดค่าตัวแปรและนิยามค่าความถูกต้อง (precision) และค่าความครบถ้วน (recall) ดังนี้

ns คือ ชุดเอกสารจากสืบค้นด้วยวิธีธรรมดา

nc คือ ชุดเอกสารที่ถูกต้องจากสืบค้นด้วยวิธีธรรมดา

as คือ ชุดเอกสารจากสืบค้น โดยนำฐานข้อมูลคำทางความสัมพันธ์มาใช้

ac คือ ชุดเอกสารที่ถูกต้องจากสืบค้นฐานข้อมูลคำทางความสัมพันธ์มาใช้

|x| คือ จำนวนเอกสารที่อยู่ในชุดเอกสาร x

ค่าความแม่นยำของเอกสาร (precision) จากการสืบค้นโดยนำฐานข้อมูลคำทางความสัมพันธ์มาใช้ คำนวณจากจำนวนเอกสารที่ถูกต้องหารด้วยจำนวนเอกสารทั้งหมดที่ได้รับจากการสืบค้น

$$\text{ค่าความแม่นยำ(precision)} = |ac| / |as| \quad (1)$$

ค่าความแม่นยำที่เพิ่มขึ้นของเอกสาร คำนวณจากผลต่างของค่าความแม่นยำจากการสืบค้นด้วยวิธีธรรมดา และการสืบค้นฐานข้อมูลคำทางความสัมพันธ์มาใช้

$$\text{ค่าความแม่นยำที่เพิ่มขึ้น} = (|ac| / |as|) - (|nc| / |ns|) \times 100 \quad (2)$$

ค่าความครบถ้วน (recall) คำนวณจากจำนวนเอกสารที่ได้รับหารด้วยจำนวนเอกสารทั้งหมดที่มีอยู่ในระบบ ในงานวิจัยนี้สนใจค่าความครบถ้วนที่เพิ่มขึ้น โดยพิจารณาเปรียบเทียบระหว่างการนำฐานข้อมูลค่าทางความสัมพันธ์มาใช้กับวิธีธรรมดา

$$\text{ค่าความครบถ้วน (recall) ที่เพิ่มขึ้น} = \frac{(|ac - ac \cap nc| - |nc - ac \cap nc|)}{|ac \cup nc|} \times 100 \quad (3)$$

$|ac - ac \cap nc|$ คือ จำนวนเอกสารที่ถูกต้องที่ได้รับเพิ่มขึ้นเมื่อสืบค้นโดยนำฐานข้อมูลค่าทางความสัมพันธ์มาใช้

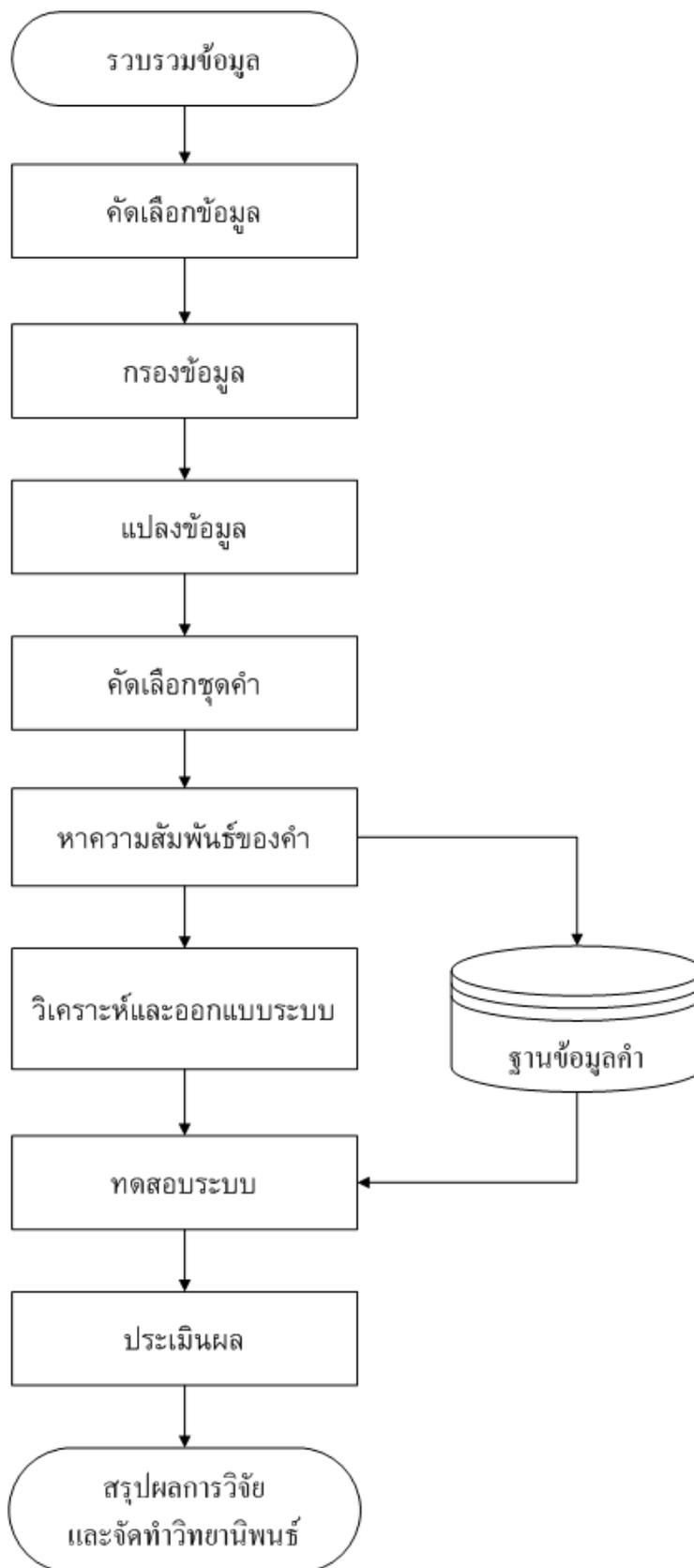
$|nc - ac \cap nc|$ คือ จำนวนเอกสารที่ขาดหายไป เมื่อสืบค้นโดยนำฐานข้อมูลค่าทางความสัมพันธ์มาใช้

$|ac \cup nc|$ คือ จำนวนเอกสารที่ถูกต้องทั้งหมด ที่ได้รับจากการสืบค้นแบบธรรมดาและการสืบค้นโดยนำฐานข้อมูลค่าทางความสัมพันธ์มาใช้

6. สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์

ขั้นตอนสุดท้ายในการทำวิจัยคือ การสรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์สำหรับเผยแพร่ผลงานวิจัย

จากขั้นตอนวิธีการดำเนินงานวิจัยสามารถเขียนเป็นแผนผังแสดงขั้นตอนการดำเนินการวิจัย ดังแสดงในภาพที่ 1



ภาพที่ 1 แสดงขั้นตอนการดำเนินการวิจัย

บทที่ 4

ผลการดำเนินงานวิจัย

จากการที่ศึกษาวิธีการต่างๆ เพื่อนำมาประยุกต์ใช้ในการวิจัย โดยผู้วิจัยได้ดำเนินงานวิจัยจนสำเร็จตามวัตถุประสงค์ ซึ่งสามารถอธิบายผลการดำเนินงานวิจัยออกเป็น 7 ส่วน ดังนี้

1. การเตรียมข้อมูล
2. การเลือกชุดคำที่มีน้ำหนักคำเชิงความถี่
3. การพิจารณาเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์
4. การสร้างกฎความสัมพันธ์ของคำ
5. การออกแบบระบบและทดสอบระบบ
6. การวิเคราะห์และประเมินผลการทดสอบ
7. สรุปผลการทดสอบ

1. การเตรียมข้อมูล

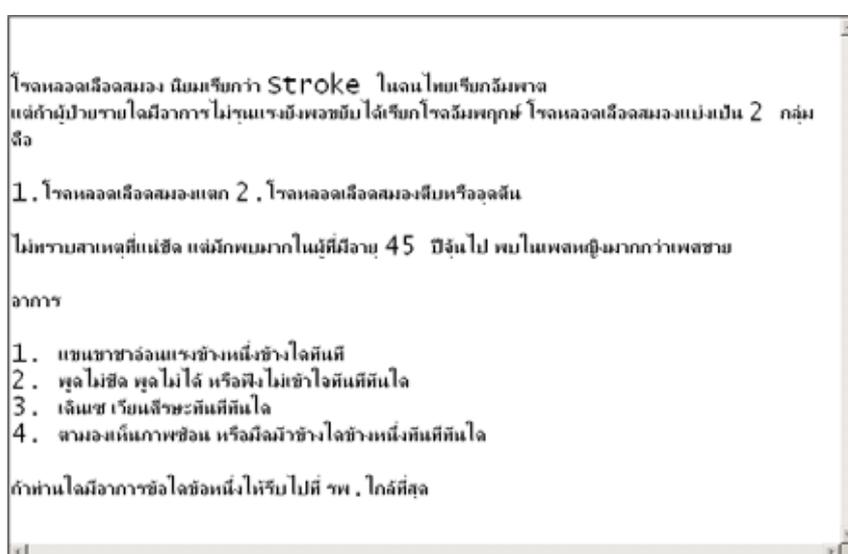
งานวิจัยนี้ได้ดำเนินการเก็บรวบรวมข้อมูลในกลุ่มข้อมูลทางการแพทย์ทั้งภาษาไทยและภาษาอังกฤษในรูปแบบเอกสารเอชทีเอ็มแอล ซึ่งประกอบด้วย 3 กลุ่มโรคแสดงดัง ตารางที่ 5

ตารางที่ 5 แสดงรายละเอียดกลุ่มโรคและจำนวนเอกสารในแต่ละภาษา

กลุ่มโรค	จำนวนเอกสาร	
	ภาษาอังกฤษ	ภาษาไทย
กลุ่มโรคระบบหัวใจ	152	84
กลุ่มโรคระบบไหลเวียนเลือด	76	43
กลุ่มเนื้องอก (รวมมะเร็ง)	375	86
รวม	603	213

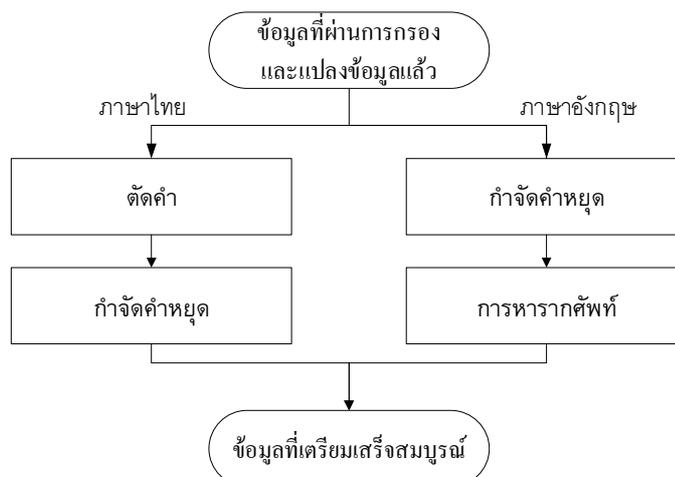
เมื่อได้ดำเนินการรวบรวมเว็บเพจที่ต้องการมาแล้วสิ่งที่จะต้องดำเนินการต่อไปเพื่อให้ได้ข้อมูลที่เราสนใจและต้องการเพื่อนำมาดำเนินการวิจัยในขั้นตอนต่อไป ดังนี้

1.1 กรองข้อมูลและแปลงข้อมูล เนื่องจากเอกสารเว็บเพจนั้นนอกจากจะมีส่วนของเนื้อหาที่น่าสนใจแล้วยังประกอบไปด้วยข้อความที่เป็นองค์ประกอบของเว็บเพจ อาทิ เมนู ข่าวประชาสัมพันธ์ ซึ่งเป็นเนื้อหาซ้ำ ๆ โดยในแต่ละเว็บเพจจะมีปริมาณข้อมูลที่ไม่น่าสนใจที่มาก บางเว็บเพจมีปริมาณมากกว่าเนื้อหาที่ต้องการ ซึ่งหากไม่ดำเนินการดึงข้อมูลที่ต้องการออกมาหรือตัดส่วนที่ไม่น่าสนใจออกไปจะทำให้ผลที่ได้คลาดเคลื่อนและไม่สามารถนำมาใช้งานต่อไปได้ จากนั้นดำเนินการแปลงเว็บเพจดังกล่าวซึ่งอยู่ในรูปแบบแฟ้มเอกสารเอชทีเอ็มแอลให้อยู่ในรูปแบบแฟ้มเอกสารข้อความ จากนั้นจะได้แฟ้มเอกสารที่มีเนื้อหาที่ต้องการ ดังภาพที่ 2



ภาพที่ 2 แสดงภาพเอกสารที่ผ่านการกรองข้อมูลและแปลงข้อมูล

1.2 การกำจัดคำหยุด การตัดคำสำหรับภาษาไทย (Swath 2009) และการหารากศัพท์สำหรับภาษาอังกฤษ (Snowball 2009) ขั้นตอนนี้แนวทางการดำเนินการแตกต่างกันระหว่างภาษาไทยและภาษาอังกฤษเนื่องจากโครงสร้างทางภาษาที่แตกต่างกัน โดยภาษาไทยนั้นจำเป็นต้องทำการตัดคำเพื่อแยกคำหยุดออกจากเนื้อหาที่ต้องการเพื่อกำจัดคำหยุด ในขณะที่ข้อมูลภาษาอังกฤษเป็นข้อมูลที่จะต้องดำเนินการกำจัดคำหยุดก่อนการหารากศัพท์เพราะหากหารากศัพท์ก่อน คำหยุดก็จะถูกเปลี่ยนแปลงรูปไป ดังนั้นผู้วิจัยมีแนวทางการดำเนินการ ดังภาพที่ 3



ภาพที่ 3 แสดงภาพขั้นตอนการดำเนินการกำจัดคำหยุด การตัดคำและการหารากศัพท์

ข้อมูลกลุ่มโรคภาษาไทยเมื่อผ่านกระบวนการตัดคำแล้ว (Swath 2009) แสดงดังภาพที่ 4

มะเร็งเต้านม การเกิดโรค มะเร็งเต้านม เป็นมะเร็งที่พบบ่อยเป็นอันดับ 1 ของประเทศไทย มักจะเกิดในหญิงอายุ 40 ปีขึ้นไป และพบมากในหญิงที่ไม่ได้มีบุตรหรือมีบุตรน้อย และในผู้ที่มีการสูบบุหรี่หรือดื่มแอลกอฮอล์เป็นประจำ สาเหตุของการเกิดโรคยังไม่ทราบแน่ชัด แต่คาดว่าน่าจะเกิดจากฮอร์โมนเพศหญิงที่มากเกินไป หรือการเปลี่ยนแปลงของเซลล์ในเต้านม

อาการของมะเร็งเต้านมจะเริ่มต้นด้วยก้อนเนื้อที่เต้านม ซึ่งอาจมีลักษณะเป็นก้อนแข็งหรือเป็นก้อนนุ่มก็ได้ ก้อนเนื้อนี้อาจมีขนาดตั้งแต่ไม่กี่มิลลิเมตรไปจนถึงหลายเซนติเมตร ก้อนเนื้อที่เต้านมอาจทำให้เกิดอาการอื่น ๆ ได้ เช่น อาการเจ็บปวดที่เต้านม อาการบวมของผิวหนังที่เต้านม หรืออาการที่ผิวหนังรอบ ๆ เต้านมเปลี่ยนสีหรือเป็นรอยแตก

การวินิจฉัยโรค มะเร็งเต้านมทำได้โดยการตรวจร่างกาย การตรวจเลือด การตรวจเต้านมด้วยมือ การตรวจเต้านมด้วยเครื่องอัลตราซาวด์ การตรวจเต้านมด้วยเครื่องแมมโมแกรม การตรวจชิ้นเนื้อเต้านม และการตรวจหาฮอร์โมนเต้านม

การรักษาโรค มะเร็งเต้านมทำได้โดยการผ่าตัด การฉายรังสี และการใช้ยาต้านมะเร็ง การผ่าตัดเต้านมสามารถทำได้ทั้งการผ่าตัดเต้านมทั้งหมดและการผ่าตัดเต้านมบางส่วน การฉายรังสีเต้านมเป็นการใช้รังสีเอกซ์เพื่อทำลายเซลล์มะเร็งในเต้านม การใช้ยาต้านมะเร็งเป็นการใช้ยาเพื่อทำลายเซลล์มะเร็งในเต้านม

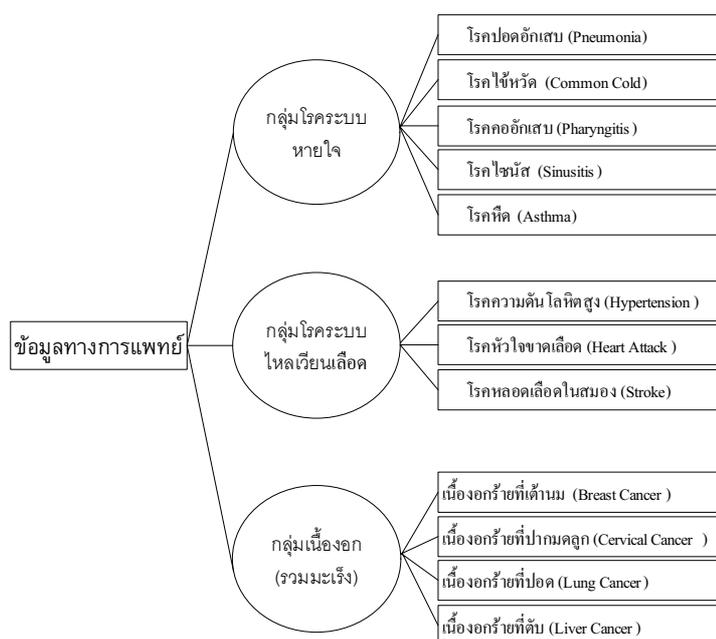
การพยากรณ์โรค มะเร็งเต้านมขึ้นอยู่กับหลายปัจจัย เช่น ขนาดของก้อนเนื้อที่เต้านม ตำแหน่งของก้อนเนื้อที่เต้านม และระยะของโรค มะเร็งเต้านมในระยะเริ่มต้นมีอัตราการรอดชีวิตสูง ในขณะที่อัตราการรอดชีวิตจะลดลงตามระยะของโรค

การป้องกันโรค มะเร็งเต้านมทำได้โดยการตรวจเต้านมด้วยมือเป็นประจำ การตรวจเต้านมด้วยเครื่องอัลตราซาวด์เป็นประจำ และการตรวจหาฮอร์โมนเต้านมเป็นประจำ

MAMMOGRAPHY

ภาพที่ 4 แสดงเอกสารที่ได้ผ่านกระบวนการตัดคำ

1.3 การจัดเตรียมชุดเอกสาร ขั้นตอนนี้เป็นการจัดเตรียมเอกสารเป็นกลุ่มเอกสารเพื่อสำหรับการนำเข้าสู่การคัดเลือกชุดคำ (weka 2009) โดยแบ่งออกเป็นสองภาษาและแต่ละภาษาประกอบด้วยข้อมูลตามกลุ่มโรคซึ่งแต่ละกลุ่มโรคจะประกอบด้วยเอกสารที่เกี่ยวข้องกับโรคในกลุ่มนั้นๆ ดังแสดงดังภาพที่ 5 ซึ่งเมื่อรวมแล้วจะได้ชุดเอกสารเพื่อดำเนินการทดสอบทั้งหมด 6 ชุด แต่ละชุดประกอบด้วยโรคตามกลุ่มที่กำหนด



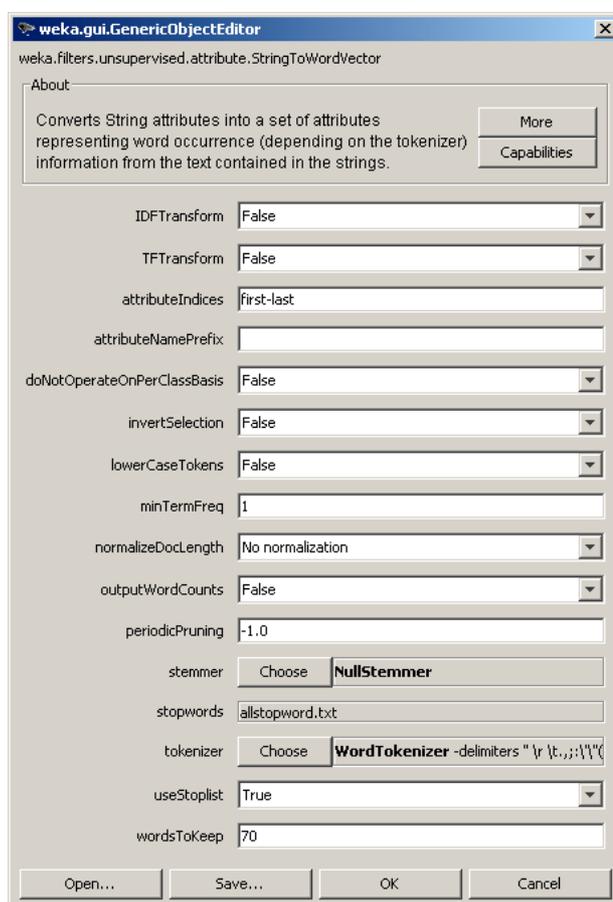
ภาพที่ 5 แสดงภาพการจัดกลุ่มเอกสาร

2. การเลือกชุดคำที่มีน้ำหนักค่าเชิงความถี่สูง

เนื่องจากเนื้อหาที่อยู่ในเว็บเพจนั้นมีจำนวนมาก ถึงแม้จะทำการลดด้วยการกำจัดคำหยุดและการหารากศัพท์ ปริมาณคำในแต่ละกลุ่มข้อมูลยังคงมีปริมาณมากซึ่งหากนำคำทั้งหมดที่ได้มาดำเนินการหาความสัมพันธ์ก็จะได้กฎที่เกิดขึ้นมีจำนวนมาก จึงจำเป็นต้องคัดเลือกชุดคำด้วย (weka 2009) โดยการให้ค่าน้ำหนักของคำโดยใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร (Document Frequency) ซึ่งหมายความว่าคำดังกล่าวพบได้ในหลายเอกสาร ในงานวิจัยนี้ใช้วิธีการหาค่าเฉลี่ยของคำที่เกิดขึ้นต่อจำนวนเอกสารทั้งหมดโดยหาได้จากสมการดังนี้

$$\text{ค่าเฉลี่ย} = \frac{\text{จำนวนเอกสารที่พบคำนั้น}}{\text{จำนวนเอกสารทั้งหมด}}$$

งานวิจัยนี้ได้จัดกลุ่มข้อมูลออกเป็นกลุ่มตามรายโรค อาทิ กลุ่มเนื้องอก (รวมมะเร็ง) แบ่งออกเป็น 4 กลุ่มรายโรค ซึ่งได้กำหนดจำนวนคำต่อกลุ่มโรคไว้ไม่เกิน 200 คำ จากการทดสอบการคัดเลือกชุดคำค่าเฉลี่ยโดยใช้ฟิวเตอร์ StringToWordVector ใน (weka 2009) โดยกำหนดค่า wordsTokeep ในค่าที่ 50, 60 และ 70 พบว่าค่าเฉลี่ยที่เหมาะสมอยู่ที่ 70 ซึ่งมีความเป็นไปได้ที่จะเกิดขึ้นคือเกิดคำที่ซ้ำกันระหว่างกลุ่มข้อมูลและค่านำหนักของคำที่เท่ากันทั้งในกลุ่มรายโรคเดียวกัน และระหว่างกลุ่มรายโรค ซึ่งในกรณีดังกล่าวนี้ในงานวิจัยนี้จะนับคำที่เกิดขึ้นในกรณีดังกล่าวเพียงหนึ่งค่าเฉลี่ยเท่านั้น ซึ่งค่าเฉลี่ยของคำที่เกิดขึ้นอยู่ระหว่าง 0.016 ถึง 1.0 การใช้ฟิวเตอร์ดังกล่าวแสดงดังภาพที่ 6



ภาพที่ 6 การคัดเลือกชุดคำค่าเฉลี่ยโดยใช้ฟิวเตอร์ StringToWordVector ใน (weka 2009)

ซึ่งเมื่อดำเนินการเลือกชุดคำในแต่ละกลุ่มเอกสารได้จำนวนคำแสดงดังตารางที่ 6 และตัวอย่างคำที่ได้จากการคัดเลือกชุดคำเชิงความถี่ แสดงดังตารางที่ 7

ตารางที่ 6 แสดงจำนวนคำที่ได้จากการคัดเลือกคำเชิงความถี่แยกตามกลุ่มโรคและภาษา

ชุดข้อมูล	จำนวนคำ	
	ภาษาไทย	ภาษาอังกฤษ
กลุ่มโรกระบบหายใจ	179	197
กลุ่มโรกระบบไหลเวียนเลือด	125	129
กลุ่มเนื้องอก (รวมมะเร็ง)	155	136

ตารางที่ 7 แสดงตัวอย่างคำจากการคัดเลือกคำเชิงความถี่กลุ่มโรกระบบหายใจ

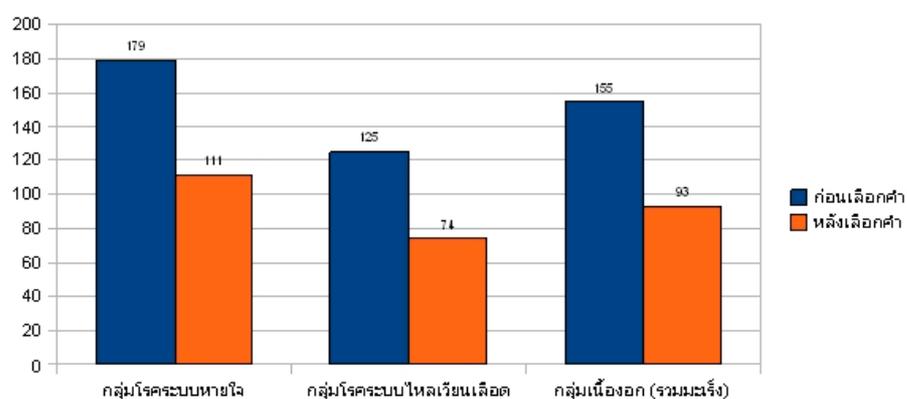
สาร	อักเสบ	แพทย์	เป็นไข้
สาเหตุ	อาการ	แพ้	เย็น
สูด	อากาศ	โรค	เสมหะ
หลอดลม	อายุ	โรคมุมิแพ้	แบคทีเรีย
หลีกเลี่ยง	อาหาร	ไอ	โรคหัด
หอบ	เด็ก	คอ	ไข้
หอบหืด	เรื้อรัง	คัดจมูก	ไข้หวัด
หายใจ	เวลา	งาน	ไวรัส
ออกกำลังกาย	เสียง	จมูก	กรณี
ออกฤทธิ์	เสียชีวิต	จาม	กลิ่น

3. การพิจารณาคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์

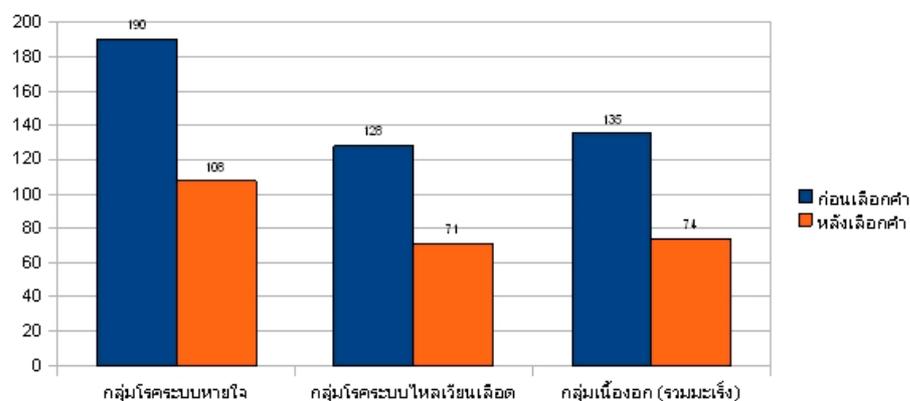
เนื่องจากคำที่ได้จากการคัดเลือกชุดคำโดยการให้ค่าน้ำหนักของคำโดยใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร (Document Frequency) คำที่ได้นั้นมีทั้งคำที่สื่อความหมายและคำที่ไม่สื่อความหมายใดซึ่งยากต่อการนำไปใช้งาน โดยการคัดเลือกคำในงานวิจัยนี้ใช้ผู้เชี่ยวชาญทางการแพทย์ตำแหน่งเภสัชกรจำนวน 3 ท่านทำการเลือกคำที่ผ่านการคัดเลือกชุดคำเชิงความถี่มาแล้ว ซึ่งการคัดเลือกชุดคำนี้ใช้เกณฑ์ในการคัดเลือกคือคำที่ถูกคัดเลือกให้สามารถนำไปใช้งานได้จะต้องมีผู้เชี่ยวชาญทางการแพทย์เลือก 2 ท่านจาก 3 ท่าน ซึ่งผลการคัดเลือกคำโดยผู้เชี่ยวชาญ แสดงดังตารางที่ 8 โดยรายการคำที่ผ่านการคัดเลือกแสดงในภาคผนวก ก

ตารางที่ 8 แสดงจำนวนคำจากการคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์

ชุดข้อมูล	จำนวนคำ	
	ภาษาไทย	ภาษาอังกฤษ
กลุ่มโรคระบบหายใจ	111	108
กลุ่มโรคระบบไหลเวียนเลือด	74	71
กลุ่มเนื้องอก (รวมมะเร็ง)	93	74
รวม	278	253



ภาพที่ 7 แสดงเปรียบเทียบจำนวนคำก่อนและหลังจากผู้เชี่ยวชาญเลือกคำกลุ่มข้อมูลภาษาไทย



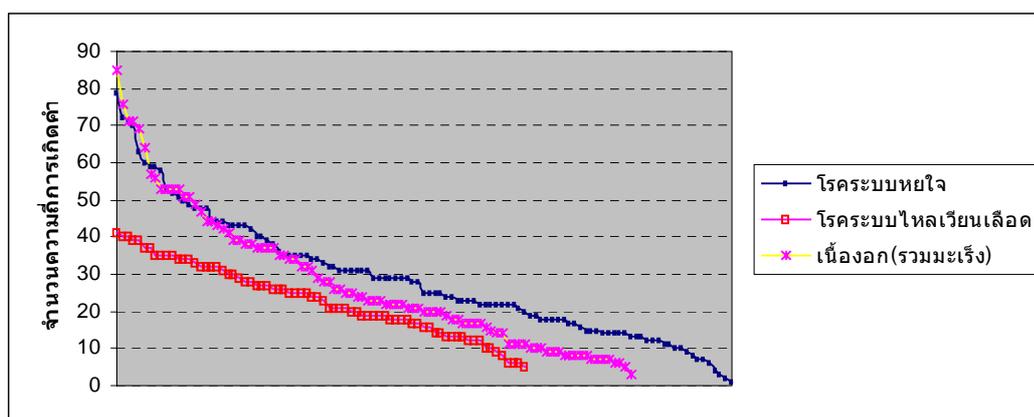
ภาพที่ 8 แสดงเปรียบเทียบจำนวนคำก่อนและหลังจากผู้เชี่ยวชาญเลือกคำกลุ่มข้อมูลภาษาอังกฤษ

4. การสร้างกฎความสัมพันธ์ของคำ

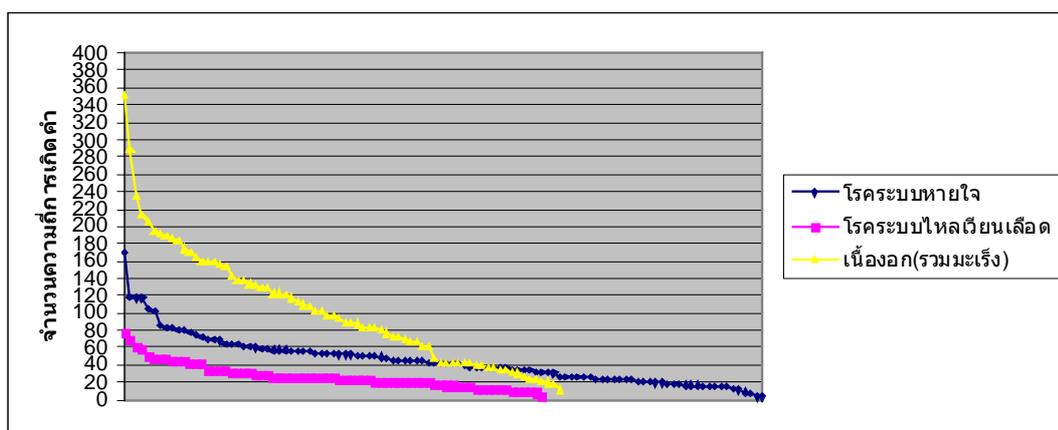
ขั้นตอนนี้เป็นกรนำเอาข้อมูลที่ได้ดำเนินการคัดเลือกคำเชิงความถี่และผ่านการคัดเลือกคำจากผู้เชี่ยวชาญทางการแพทย์นำมาดำเนินการหาความสัมพันธ์ของคำในแต่ละกลุ่มข้อมูลคือ กลุ่มโรคระบบหายใจ กลุ่มโรคระบบไหลเวียนเลือด และกลุ่มเนื้องอก (รวมมะเร็ง) ทั้งภาษาไทยและภาษาอังกฤษ รวม 6 ชุดข้อมูล โดยในขั้นตอนนี้ผู้วิจัยได้ใช้ (Togaware : 2009) เป็นเครื่องมือในการทดลอง

งานวิจัยนี้เป็นการนำเอาการหาความสัมพันธ์มาประยุกต์ใช้เพื่อหาคำที่เกิดร่วมกันโดยสนใจที่จะนำคำที่เกิดขึ้นในกฎมาใช้ทั้งกฎในรูปแบบคำร่วมที่เกิดขึ้นด้วยกันเท่านั้นโดยไม่พิจารณาคำที่เกิดขึ้นก่อนที่ส่งผลทำให้เกิดอีกคำหนึ่ง ดังนั้นจำนวนคำที่เกิดขึ้นที่จะนำมาใช้งานจะต้องมีปริมาณความถี่ของการเกิดในเอกสารอยู่ในระดับที่เหมาะสม และคำที่เกิดร่วมกันนั้นจะต้องมีค่าความเชื่อมั่นที่สูงเนื่องจากต้องการจำนวนคำที่พบร่วมกันมีจำนวนไม่ต่างกัน จากการทดลองพบว่าจำนวนคำในเอกสารที่นำมาทดลองนั้นมีปริมาณความถี่ของคำเกิดขึ้นในแต่ละเอกสารที่แตกต่างกันซึ่งหากนำมาสร้างกฎความสัมพันธ์กันโดยใช้ค่าสนับสนุนที่ต่ำ คำที่มีจำนวนคำที่น้อยกลับมีค่าความเชื่อมั่นที่สูงทั้งที่พบคำดังกล่าวในเอกสารน้อยฉบับ

จากการตรวจสอบข้อมูลพบว่าจำนวนคำที่เกิดขึ้นในเอกสารของข้อมูลภาษาไทยและภาษาอังกฤษ อยู่ในช่วง 10 เอกสารขึ้นไป ดังแสดงในภาพที่ 9 และ 10



ภาพที่ 9 แสดงความถี่ของการเกิดคำในกลุ่มข้อมูลภาษาไทย



ภาพที่ 10 แสดงความหนาแน่นของค่าในกลุ่มข้อมูลภาษาอังกฤษ

ดังนั้นงานวิจัยนี้จึงกำหนดค่าสนับสนุนต่ำสุดที่ 10 ซึ่งจะทำให้ค่าที่ผู้เชี่ยวชาญเลือกมานั้นไม่สามารถนำมาใช้งานได้อันเนื่องมาจากมีค่าสนับสนุนไม่ถึง 10 ซึ่งจากการตรวจสอบข้อมูลทั้งหมดนำมาทำการทดลองพบว่าข้อมูลมีจำนวนน้อยและเป็นข้อมูลที่ผู้เชี่ยวชาญทางการแพทย์เป็นผู้เขียนไว้ซึ่งอาจเป็นคำพูดเฉพาะบุคคลหรือเป็นอาการของโรคในลักษณะเดียวกัน อาทิ คำว่า กาฬโรค บางท่านเรียกว่ากาฬโรคปอด ซึ่งมีลักษณะอาการปอดอักเสบ และเมื่อพูดถึงปอดอักเสบ อาจพูดเชื่อมโยงอาการเดียวกันทำให้มีจำนวนคำว่ากาฬโรคน้อยกว่าคำอื่นๆ

จากการทำการทดสอบด้วยค่าความเชื่อมั่นต่ำสุดที่แตกต่าง โดยเริ่มจากค่าความเชื่อมั่นที่ 80, 85 และ 90 ตามลำดับและกำหนดจำนวนคำต่อกฎไว้ไม่เกิน 3 คำ ทั้งกลุ่มข้อมูลภาษาไทยและกลุ่มข้อมูลภาษาอังกฤษโดยการสร้างกฎความสัมพันธ์ แสดงดังภาพที่ 11 และ 12

```

surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgell/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 th-respi1080.tab th-respi1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-respi1080.tab ... [111 item(s), 84 transaction(s)] done [0.00s].
sorting and recoding items ... [103 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-respi1080.rul ... [4871 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgell/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 th-cardio1080.tab th-cardio1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-cardio1080.tab ... [75 item(s), 43 transaction(s)] done [0.00s].
sorting and recoding items ... [75 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cardio1080.rul ... [8033 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgell/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 th-cancer1080.tab th-cancer1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-cancer1080.tab ... [93 item(s), 85 transaction(s)] done [0.00s].
sorting and recoding items ... [80 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cancer1080.rul ... [5624 rule(s)] done [0.00s].

```

ภาพที่ 11 แสดงการสร้างกฎความสัมพันธ์ของค่าที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาไทย

จากภาพที่ 11 จะเห็นว่าการสร้างกฎที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาไทยนั้นจะได้จำนวนกฎที่เกิดขึ้นมาก โดยกลุ่มโรคระบบหายใจได้จำนวนกฎ 4,871 กฎและจำนวนค่าในการสร้างกฎ 103 ค่า ,กลุ่มโรคระบบไหลเวียนเลือดได้จำนวนกฎ 8,033 กฎและจำนวนค่าในการสร้างกฎ 75 ค่า และกลุ่มเนื้องอก (รวมมะเร็ง) ได้จำนวนกฎ 5,624 กฎและจำนวนค่าในการสร้างกฎ 80 ค่า

จากนั้นทำการสร้างกฎความสัมพันธ์ของค่าโดยตั้งปรับค่าความเชื่อมั่นต่ำสุดที่ 85 แต่ยังคงค่าสนับสนุนต่ำสุดไว้ที่ 10 ผลการสร้างกฎความสัมพันธ์แสดงดังภาพที่ 12

```

surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c85 -m2 -n3 th-respiflg.tab th-respi1085.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02)      (c) 1996-2003 Christian Borgelt
reading th-respiflg.tab ... [111 item(s), 84 transaction(s)] done [0.00s].
sorting and recoding items ... [103 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-respi1085.rul ... [1675 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c85 -m2 -n3 th-cardioflg.tab th-cardio1085.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02)      (c) 1996-2003 Christian Borgelt
reading th-cardioflg.tab ... [75 item(s), 43 transaction(s)] done [0.00s].
sorting and recoding items ... [75 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cardio1085.rul ... [1703 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c85 -m2 -n3 th-cancerflg.tab th-cancer1085.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02)      (c) 1996-2003 Christian Borgelt
reading th-cancerflg.tab ... [93 item(s), 85 transaction(s)] done [0.00s].
sorting and recoding items ... [80 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cancer1085.rul ... [503 rule(s)] done [0.00s].

```

ภาพที่ 12 แสดงการสร้างกฎความสัมพันธ์ของค่าที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาไทย

จากภาพที่ 12 จะเห็นว่าการสร้างกฎที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาไทยนั้นจะได้จำนวนกฎที่เกิดขึ้นลดลงอย่างมากและยังคงรักษาจำนวนค่าในการสร้างกฎไว้เท่ากับการตั้งค่าความเชื่อมั่นต่ำสุดที่ 80 โดยกลุ่มโรคระบบหายใจได้จำนวนกฎ 1675 กฎและจำนวนค่าในการสร้างกฎ 103 ค่า ,กลุ่มโรคระบบไหลเวียนเลือดได้จำนวนกฎ 1703 กฎและจำนวนค่าในการสร้างกฎ 75 ค่า และกลุ่มเนื้องอก (รวมมะเร็ง) ได้จำนวนกฎ 503 กฎและจำนวนค่าในการสร้างกฎ 80 ค่า

จากนั้นทำการสร้างกฎความสัมพันธ์ของค่าโดยตั้งปรับค่าความเชื่อมั่นต่ำสุดที่ 90 แต่ยังคงค่าสนับสนุนต่ำสุดไว้ที่ 10 ผลการสร้างกฎความสัมพันธ์แสดงดังภาพที่ 13

```

surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c90 -m2 -n3 th-respfig.tab th-respi1090.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-respfig.tab ... [111 item(s), 84 transaction(s)] done [0.00s].
sorting and recoding items ... [95 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-respi1090.rul ... [691 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c90 -m2 -n3 th-cardiofig.tab th-cardio1090.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-cardiofig.tab ... [75 item(s), 43 transaction(s)] done [0.00s].
sorting and recoding items ... [52 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cardio1090.rul ... [1457 rule(s)] done [0.00s].
surachai@surachai-desktop:~/gnome-datamine-tools-1.0.1/borgel/apriori/src$ ./apriori -o -th -a -s10 -c90 -m2 -n3 th-cancerfig.tab th-cancer1090.rul
./apriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading th-cancerfig.tab ... [93 item(s), 85 transaction(s)] done [0.00s].
sorting and recoding items ... [78 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing th-cancer1090.rul ... [287 rule(s)] done [0.00s].

```

ภาพที่ 13 แสดงการสร้างกฎความสัมพันธ์ของค่าที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 90 กลุ่มข้อมูลภาษาไทย

จากภาพที่ 13 จะเห็นว่าการสร้างกฎที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 90 กลุ่มข้อมูลภาษาไทยนั้นจะได้จำนวนกฎที่เกิดขึ้นลดลงอย่างมากแต่จำนวนค่าในการสร้างกฎก็ลดลงตามไปด้วยโดยกลุ่มโรกระบบหายใจได้จำนวนกฎ 691 กฎและจำนวนค่าในการสร้างกฎ 95 ค่า ,กลุ่มโรกระบบไหลเวียนเลือดได้จำนวนกฎ 1,457 กฎและจำนวนค่าในการสร้างกฎ 52 ค่า และกลุ่มเนื้องอก (รวมมะเร็ง) ได้จำนวนกฎ 287 กฎและจำนวนค่าในการสร้างกฎ 78 ค่า

ในส่วนของกลุ่มข้อมูลภาษาอังกฤษผู้วิจัยได้ดำเนินการทดลองในรูปแบบเดียวกัน กำหนดค่าสนับสนุนต่ำสุดที่ 10 และทำการทดสอบด้วยค่าความเชื่อมั่นต่ำสุดที่แตกต่าง โดยเริ่มจากค่าความเชื่อมั่นที่ 80, 85 และ 90 ตามลำดับและกำหนดจำนวนค่าต่อกฎไว้ไม่เกิน 3 ค่า แสดงดังภาพที่ 14

```

surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 en-respfig.tab en-resp1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-respfig.tab ... [108 item(s), 152 transaction(s)] done [0.00s].
sorting and recoding items ... [103 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-resp1080.rul ... [895 rule(s)] done [0.00s].
surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 en-cardiofig.tab en-cardio1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-cardiofig.tab ... [71 item(s), 76 transaction(s)] done [0.00s].
sorting and recoding items ... [70 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-cardio1080.rul ... [1463 rule(s)] done [0.00s].
surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c80 -m2 -n3 en-cancerfig.tab en-cancer1080.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-cancerfig.tab ... [74 item(s), 375 transaction(s)] done [0.00s].
sorting and recoding items ... [64 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-cancer1080.rul ... [218 rule(s)] done [0.00s].

```

ภาพที่ 14 แสดงการสร้างกฎความสัมพันธ์ของค่าที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาอังกฤษ

จากภาพที่ 14 จะเห็นว่าการสร้างกฎที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 80 กลุ่มข้อมูลภาษาอังกฤษนั้นจะได้จำนวนกฎที่แตกต่างจากกลุ่มข้อมูลภาษาไทยเนื่องมาจากการกระจายตัวของคำ และรูปแบบทางโครงสร้างภาษาที่แตกต่าง โดยกลุ่มโรกระบบหายใจได้จำนวนกฎ 895 กฎและจำนวนคำในการสร้างกฎ 103 คำ ,กลุ่มโรกระบบไหลเวียนเลือดได้จำนวนกฎ 1,463 กฎและจำนวนคำในการสร้างกฎ 70 คำ และกลุ่มเนื้องอก (รวมมะเร็ง) ได้จำนวนกฎ 218 กฎและจำนวนคำในการสร้างกฎ 64 คำ

จากนั้นทำการสร้างกฎความสัมพันธ์ของค่าโดยตั้งปรับค่าความเชื่อมั่นต่ำสุดที่ 85 แต่ยังคงค่าสนับสนุนต่ำสุดไว้ที่ 10 ผลการสร้างกฎความสัมพันธ์แสดงดังภาพที่ 15

```

surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c85 -m2 -n3 en-respfig.tab en-resp1085.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-respfig.tab ... [108 item(s), 152 transaction(s)] done [0.00s].
sorting and recoding items ... [68 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-resp1085.rul ... [406 rule(s)] done [0.00s].
surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c85 -m2 -n3 en-cardiofig.tab en-cardio1085.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-cardiofig.tab ... [71 item(s), 76 transaction(s)] done [0.00s].
sorting and recoding items ... [61 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-cardio1085.rul ... [522 rule(s)] done [0.00s].
surachai@surachai-desktop:~/.gnome-datamine-tools-1.0.1/borgelt/apriori/src$ Japriori -o -th -a -s10 -c85 -m2 -n3 en-cancerfig.tab en-cancer1085.rul
Japriori - find association rules with the apriori algorithm
version 4.18 (2003.12.02) (c) 1996-2003 Christian Borgelt
reading en-cancerfig.tab ... [74 item(s), 375 transaction(s)] done [0.00s].
sorting and recoding items ... [58 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing en-cancer1085.rul ... [148 rule(s)] done [0.00s].

```

ภาพที่ 15 แสดงการสร้างกฎความสัมพันธ์ของค่าที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาอังกฤษ

จากภาพที่ 15 จะเห็นว่าการสร้างกฎที่ค่าสนับสนุน 10 และค่าความเชื่อมั่น 85 กลุ่มข้อมูลภาษาอังกฤษนั้นจะได้จำนวนกฎที่เกิดขึ้นลดลงอย่างมากแต่จำนวนค่าในการสร้างกฎก็ลดลงตามไปด้วยโดยกลุ่มโรคระบบหายใจได้จำนวนกฎ 406 กฎและจำนวนค่าในการสร้างกฎ 68 ค่า , กลุ่มโรคระบบไหลเวียนเลือดได้จำนวนกฎ 522 กฎและจำนวนค่าในการสร้างกฎ 61 ค่า และกลุ่มเนื้องอก (รวมมะเร็ง) ได้จำนวนกฎ 148 กฎและจำนวนค่าในการสร้างกฎ 58 ค่า

ดังนั้นงานวิจัยนี้จึงเลือกใช้ค่าความเชื่อมั่นต่ำสุดที่ 85 เนื่องจากจำนวนค่าที่นำมาสร้างกฎในแต่ละกลุ่มโรครยังอยู่ครบถ้วนและจำนวนกฎไม่มากเกินไปสำหรับข้อมูลภาษาไทยและค่าความเชื่อมั่นต่ำสุดที่ 80 สำหรับข้อมูลภาษาอังกฤษจำนวนกฎที่ได้จากการทดสอบแสดงดังตารางที่ 9

ตารางที่ 9 แสดงจำนวนกฎที่ได้จากการสร้างกฎความสัมพันธ์

ชุดข้อมูล	จำนวนค่า	
	ภาษาไทย	ภาษาอังกฤษ
กลุ่มโรคระบบหายใจ	1,675	895
กลุ่มโรคระบบไหลเวียนเลือด	1,703	1,463
กลุ่มเนื้องอก (รวมมะเร็ง)	503	218
รวม	3,881	2,576

พิจารณาผลเพื่อจัดทำฐานข้อมูลค่าทางความสัมพันธ์การจัดทำฐานข้อมูลประกอบด้วย ตารางข้อมูล 2 ตาราง คือ 1) ตารางกลุ่มโรค ซึ่งประกอบด้วย 3 กลุ่มโรคได้แก่ โรคระบบหายใจ โรคระบบไหลเวียนเลือด และเนื้องอก (รวมมะเร็ง) 2) ตารางกฎความสัมพันธ์ โดยการออกแบบโครงสร้างแสดงดังตารางที่ 10 และ 11

ตารางที่ 10 แสดงโครงสร้างฐานข้อมูลกลุ่มโรค (Disease)

ชื่อฟิลด์	ประเภทข้อมูล	ความกว้าง	คำอธิบาย	ประเภทคีย์
D_ID	Int	11	รหัสกลุ่มโรค	PK
H_NAME_T	varchar	50	ชื่อกลุ่มโรคภาษาไทย	
H_NAME_E	varchar	50	ชื่อกลุ่มโรคภาษาอังกฤษ	

ตารางที่ 11 แสดงโครงสร้างฐานข้อมูลกฎความสัมพันธ์ (Rules)

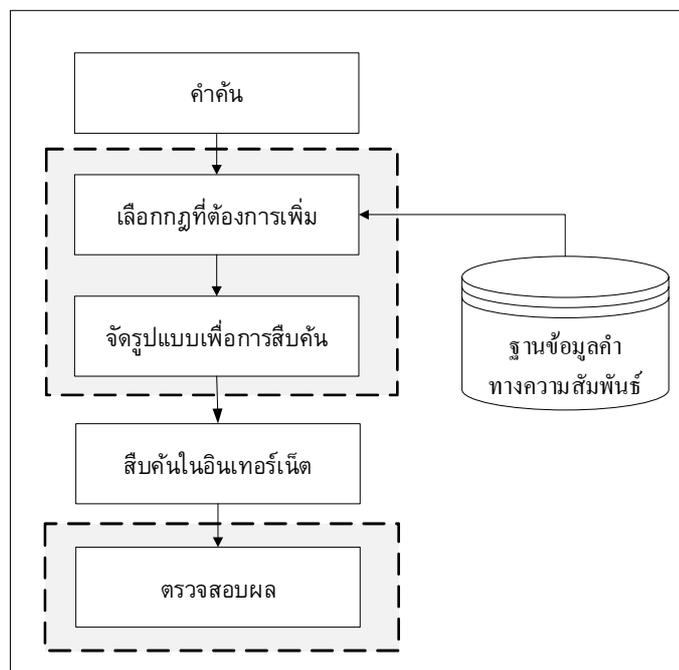
ชื่อฟิลด์	ประเภทข้อมูล	ความกว้าง	คำอธิบาย	ประเภทคีย์
ID	Int	11	รหัสกฎ	PK
RULES	varchar	200	กฎความสัมพันธ์	
D_ID	Int	11	รหัสกลุ่มโรค	FK

ตารางที่ 12 แสดงโครงสร้างฐานข้อมูลผู้ใช้ (User)

ชื่อฟิลด์	ประเภทข้อมูล	ความกว้าง	คำอธิบาย	ประเภทคีย์
ID	Int	11	รหัส	PK
USERNAME	varchar	50	ชื่อบัญชีผู้ใช้	
PWD	varchar	50	รหัสบัญชีผู้ใช้	

5. การออกแบบระบบ

เมื่อได้ผลของกฎที่มีความสัมพันธ์ของค่าแล้วนั้น นำมาพิจารณาสร้างระบบเพื่อใช้ในการทดสอบการใช้งานในการสืบค้น โดยรูปแบบแสดงดังภาพที่ 10



ภาพที่ 16 แสดงรูปแบบการทดสอบโดยการนำคำจากฐานข้อมูลคำที่มีความสัมพันธ์มาใช้

5.1 คำค้น หรือ คีย์เวิร์ดที่ผู้ทดสอบสุ่มเลือกกฎที่ต้องการตามกลุ่มโรค โดยระบุคำที่เกี่ยวข้องตรงกับคำในกฎของระบบสืบค้นของกลุ่มโรคนั้น ๆ ซึ่งประกอบด้วย โรคระบบทางเดินหายใจ โรคระบบไหลเวียนเลือด และเนื้องอก (รวมมะเร็ง)

5.2 เลือกกลุ่มคำที่มีความสัมพันธ์ที่ต้องการ ซึ่งเมื่อผู้ใช้งานเลือกคำดังกล่าวแล้วเลือกค้นหา ระบบจะดำเนินการเลือกคำที่มีความสัมพันธ์นั้นจากฐานข้อมูลคำทางความสัมพันธ์มาใช้

5.3 จัดรูปแบบเพื่อการสืบค้น คือการจัดรูปแบบคิวรี ประกอบด้วย คำค้นหรือคีย์เวิร์ด โอเปอเรเตอร์ คำที่มีความสัมพันธ์กับคำค้น ตัวอย่างเช่น ในระบบสืบค้นเกี่ยวกับกลุ่มโรคทางเดินหายใจ หากผู้ทดสอบระบุคำที่ต้องการคือ เสมหะ ระบบจะทำงานค้นหากฎที่มีคำที่มีความสัมพันธ์กับเสมหะจากฐานข้อมูลคำทางความสัมพันธ์มาใช้ ซึ่งการจัดรูปแบบเพื่อการสืบค้นจะประกอบด้วย คำว่า เสมหะ และ ไอ และ อักเสบ โดยการเรียงลำดับของคำในกฎนั้นอาจมีการสลับตำแหน่ง ไอ และ อักเสบ และ เสมหะ เป็นต้น

5.4 การตรวจสอบผล ผู้ทำการทดสอบใช้ดุลพินิจตัดสินใจว่าผลลัพธ์ที่ได้จากการสืบค้นใน 50 เอกสารแรก และระบุว่าเอกสารผลลัพธ์นั้น ถูกต้องกับโรค, ไม่ถูกต้อง, ไม่แน่ใจ อย่างใดอย่างหนึ่ง โดยทดสอบในกลุ่มโรคทั้งสองภาษาคือภาษาไทยและภาษาอังกฤษ โดยกำหนดความหมายดังนี้

ถูกต้อง หมายถึง ต้องมีเนื้อหาเกี่ยวกับโรคในกลุ่มโรคที่ทดสอบนั้น โดยระบุข้อมูล
ทั่วไป สาเหตุ อาการ การรักษา ทั้งหมดหรืออย่างใดอย่างหนึ่ง
ไม่ถูกต้อง หมายถึง เนื้อหาที่ไม่มีเนื้อหาเกี่ยวกับ สาเหตุ อาการ การรักษา
ไม่แน่ใจ หมายถึง ไม่สามารถตัดสินใจได้อย่างแน่ชัด

5.5 การประมวลผลที่ได้จากการทดสอบ โดยพิจารณาจากความถูกต้องของเนื้อหาของเอกสารที่ได้รับโดยเปรียบเทียบค่าความต่างระหว่างวิธีการสืบค้นด้วยวิธีธรรมดากับการสืบค้นด้วยวิธีการนำคำจากฐานข้อมูลคำทางความสัมพันธ์มาช่วยในการสืบค้น

6. การทดสอบระบบ

ทดสอบโดยผู้ทดสอบ 12 คน แบ่งเป็น 3 กลุ่มที่เท่ากัน โดยแบ่งทดสอบกลุ่มละภาษาคือ ภาษาไทยและภาษาอังกฤษ โดยทดสอบคนละ 1 คำค้น ที่ไม่ซ้ำกัน จากนั้นนำคำค้นนั้นไปสืบค้นทางอินเทอร์เน็ตด้วยวิธีธรรมดา ซึ่งในที่นี้ใช้ระบบเสิร์จเอนจินของ google.co.th ซึ่งเป็นที่นิยมและรู้จักอย่างแพร่หลาย ผู้ตรวจสอบจะเป็นผู้ใช้ดุลพินิจตัดสินใจว่าผลลัพธ์ที่ได้จากการสืบค้นใน 50 เอกสารแรก และระบุว่าเอกสารผลลัพธ์นั้น ถูกต้องกับโรค,ไม่ถูกต้อง,ไม่แน่ใจ อย่างใดอย่างหนึ่ง เปรียบเทียบกับการสืบค้นด้วยเทคนิคใหม่โดยใช้ฐานข้อมูลคำทางความสัมพันธ์ ด้วยวิธีการเดียวกัน คำค้นที่ผู้ทดสอบสุ่มเลือกแสดงดังตารางที่ 13

ตารางที่ 13 แสดงคำค้นที่ผู้ทดสอบสุ่มเลือกเพื่อใช้ในการทดสอบ

คิวรี	กลุ่มโรกระบบ หายใจ	คิวรี	กลุ่มโรกระบบไหล เวียนเลือด	คิวรี	กลุ่มเนื้องอก (รวมมะเร็ง)
Res1	ไอ	Car1	ความดัน	Can1	เต้านม
Res2	น้ำมูก	Car2	เครียด	Can2	สูบบุหรี่
Res3	mucus	Car3	Cholesterol	Can3	tumor
Res4	cough	Car4	diabetes	Can4	cervical

จากนั้นนำคำค้นดังกล่าวมาทดสอบด้วยวิธีธรรมดา โดยนำคำค้นนั้นไปทำการสืบค้นในเสิร์จเอนจิน ของ google.co.th โดยผลการสืบค้นทั้ง 3 กลุ่มโรคแสดงดังตารางที่ 14 ,15 และ 16

ตารางที่ 14 แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มโรคระบบหายใจ

คิวรี	สืบค้นด้วยวิธีธรรมดา		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Res1	3	47	0
Res2	7	42	1
Res3	1	47	2
Res4	3	45	2

ตารางที่ 15 แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มโรคระบบไหลเวียนเลือด

คิวรี	สืบค้นด้วยวิธีธรรมดา		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Car1	22	18	10
Car2	0	50	0
Car3	0	46	4
Car4	0	50	0

ตารางที่ 16 แสดงผลการสืบค้นด้วยวิธีธรรมดาในกลุ่มเนื้องอก (รวมมะเร็ง)

คิวรี	สืบค้นด้วยวิธีธรรมดา		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Can1	34	11	5
Can2	1	49	0
Can3	5	31	4
Can4	19	15	6

จากนั้นนำคำค้นดังกล่าวมาทดสอบด้วยเทคนิคใหม่โดยนำคำจากฐานข้อมูลกฎทางความสัมพันธ์มาช่วยขยายคำค้นโดยการจัดรูปแบบคิวรี ประกอบด้วย คำค้นหรือคีย์เวิร์ด โอเปอเรเตอร์ “และ” คำที่มีความสัมพันธ์กับคำค้น โดยการทดสอบนั้นจะทดสอบโดยใช้กฎที่มีจำนวนคำ 2 และ 3 คำ อย่างละ 1 กฎ โดยกฎที่ผู้ทดสอบเลือกแสดงดังตารางที่ 17

ตารางที่ 17 แสดงกฎที่ผู้ทดสอบเลือกใช้จากฐานข้อมูลคำทางความสัมพันธ์

คิวรี	กลุ่มโรคระบบหายใจ	คิวรี	กลุ่มโรคระบบไหลเวียนเลือด	คิวรี	กลุ่มเนื้องอก (รวมมะเร็ง)
Res1 2	ไอ ไข้	Car 12	ความดัน อ่อนแรง	Can 12	เต้านม รักแร้
Res1 3	ไข้ ไอ อักเสบ	Car 13	อ่อนแรง สมอง ความดัน	Can 13	รักแร้ เต้านม ขนาด
Res2 2	น้ำมูก เหลือง	Car 22	เครียด แน่นหน้าอก	Can 22	เสี่ยง สูบหรี่
Res2 3	เหลือง น้ำมูก ติดเชื้อ	Car 23	เจ็บ หน้าอก เครียด	Can 23	สูบบุหรี่ อายุ เสี่ยง
Res3 2	Mucus healthcare	Car 32	Cholesterol Smoke	Can 32	Tumor lymph
Res3 3	Warm fluid mucus	Car 33	Cholesterol Smoke pressure	Can 33	Node lymph tumor
Res4 2	Cough wheeze	Car 42	Cholesterol Diabetes	Can 42	Cervical pap
Res4 3	Wheeze cough chest	Car 43	Cholesterol Diabetes heart	Can 43	Vaccine hpv cervical

โดยผลการทดสอบการสืบค้นทั้ง 3 กลุ่มโรค แสดงดังตารางที่ 18, 19 และ 20

ตารางที่ 18 แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎหมายทางความสัมพันธ์กลุ่มโรคระบบหายใจ

คิวรี	สืบค้นด้วยฐานข้อมูลกฎหมายทางความสัมพันธ์		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Res12	12	38	0
Res13	32	18	0
Res22	35	13	2
Res23	40	8	2
Res32	9	36	5
Res33	2	44	4
Res42	9	40	1
Res43	6	44	0

ตารางที่ 19 แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎหมายทางความสัมพันธ์กลุ่มโรคระบบไหลเวียนเลือด

คิวรี	สืบค้นด้วยฐานข้อมูลกฎหมายทางความสัมพันธ์		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Car12	32	15	3
Car13	40	8	2
Car22	6	30	4
Car23	9	21	10
Car32	5	45	0
Car33	16	30	4
Car42	5	45	0
Car43	10	37	3

ตารางที่ 20 แสดงผลการสืบค้นโดยใช้ฐานข้อมูลกฎหมายความสัมพันธ์กลุ่มเนื้องอก (รวมมะเร็ง)

คิวรี	สืบค้นฐานข้อมูลกฎหมายความสัมพันธ์		
	ถูกต้อง	ไม่ถูกต้อง	ไม่แน่ใจ
Can12	37	11	2
Can13	41	5	4
Can22	6	44	0
Can23	8	42	0
Can32	6	44	0
Can33	10	40	0
Can42	21	21	8
Can43	25	20	5

6. การวิเคราะห์และประเมินผลการทดสอบ

วิเคราะห์ผลลัพธ์ที่ได้จากการสืบค้น ผู้ตรวจสอบจำนวน 12 คน จะทดลองใช้คำศัพท์วิวัฒนาการ 1 คำแตกต่างกันสืบค้นด้วยวิธีธรรมดา และเทคนิคใหม่ที่น่าฐานข้อมูลทางความสัมพันธ์มาใช้ โดยแสดงข้อมูลในตารางที่ 21 โดย

T คือ จำนวนเอกสารที่ผู้ใช้พิจารณาแล้วเห็นว่าถูกต้อง

F คือ จำนวนเอกสารที่ผู้ใช้พิจารณาแล้วเห็นว่าผิด

P คือ เปอร์เซ็นต์ความแม่นยำ (precision)

R คือ เปอร์เซ็นต์ความครบถ้วน (recall)

C คือ จำนวนเอกสารที่ถูกต้องและตรงกันของการสืบค้นด้วยการนำฐานข้อมูลกฎหมายสัมพันธ์มาใช้ เปรียบเทียบกับวิธีธรรมดา

L คือ จำนวนเอกสารที่ถูกต้องและขาดหายไปเมื่อนำฐานข้อมูลกฎหมายสัมพันธ์มาใช้ เปรียบเทียบกับการสืบค้นด้วยวิธีธรรมดา

I คือ จำนวนเอกสารที่ถูกต้องและเพิ่มขึ้น เมื่อนำฐานข้อมูลกฎหมายสัมพันธ์มาใช้ เปรียบเทียบกับการสืบค้นด้วยวิธีธรรมดา

%R คือ เปอร์เซนต์ความครบถ้วน (recall) ที่เพิ่มขึ้นเมื่อสืบค้นด้วยการนำฐานข้อมูลจากความสัมพันธ์มาใช้

%P คือ เปอร์เซนต์ความแม่นยำ (precision) ที่เพิ่มขึ้นเมื่อสืบค้นด้วยการนำฐานข้อมูลจากความสัมพันธ์มาใช้

การวิเคราะห์ผลลัพธ์ที่ได้จากการสืบค้นตามตารางที่ 21 นั้นคือ ความแม่นยำ (precision) ของผลลัพธ์และความครบถ้วนของผลลัพธ์ (recall) โดย

6.1 ความความแม่นยำของผลลัพธ์ (precision)

เมื่อนำจำนวนเอกสารที่ถูกต้อง (T1, T2, T3) ในตารางที่ 21 มาคำนวณค่าความแม่นยำ และค่าความแม่นยำที่เพิ่มขึ้นตามสมการ 1 และสมการ 2 ตามลำดับแล้ว จะได้ค่าเปอร์เซนต์ความแม่นยำ (P1, P2, P3) และค่าเปอร์เซนต์ความแม่นยำที่เพิ่มขึ้น (%P)

6.2 ความครบถ้วนของผลลัพธ์ (recall)

ค่าความครบถ้วน (recall) ของการสืบค้น (R) หาได้จากสมการ ดังนี้

ค่าความครบถ้วน (recall) การสืบค้นด้วยวิธีธรรมดา (R1) = $T1 / I+L$

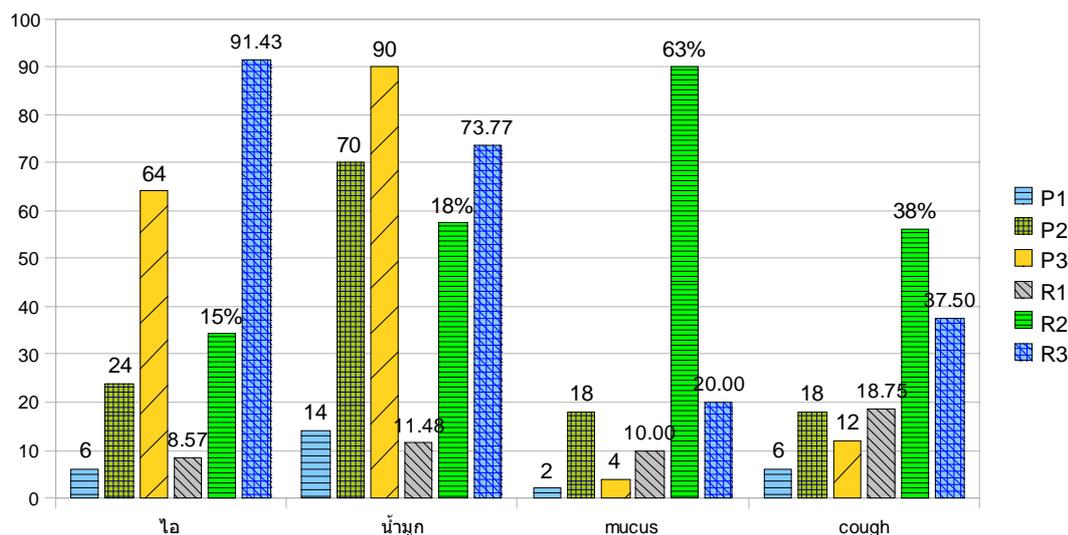
การสืบค้นโดยใช้ฐานข้อมูลจากความสัมพันธ์ (R2) = $T2 / I+L$, (R3) = $T3 / I+L$

เมื่อนำเอกสารผลลัพธ์ที่ได้จากการสืบค้นด้วยวิธีธรรมดา และการนำเทคนิคใหม่ที่ใช้ฐานข้อมูลจากความสัมพันธ์มาใช้เปรียบเทียบกัน และคำนวณค่าความครบถ้วน (recall) ที่เพิ่มขึ้นตามสมการ 3 จะได้ข้อมูลผลลัพธ์ดังแสดงในตาราง 21 (%R)

ตารางที่ 21 แสดงผลการวิเคราะห์ผลลัพธ์เปรียบเทียบวิธีธรรมดากับการใช้อำนาจข้อมูลจากความสัมพันธ์

คิวรี	วิธีธรรมดา				ใช้กฎความสัมพันธ์								เปรียบเทียบ				
	T1	F1	P1	R1	T2	F2	P2	R2	T3	F3	P3	R3	C	L	I	%R	%P
Res1	3	47	6	8.57	12	38	24	34.29	32	18	64	91.43	6	-3	38	54.29	38
Res2	7	42	14	11.48	35	13	70	57.38	40	8	80	73.77	13	-6	67	54.10	66
Res3	1	47	2	10.00	9	36	18	90.00	2	44	4	20.00	1	0	10	45.00	9
Res4	3	45	6	18.75	9	40	18	56.25	6	44	12	37.50	1	2	14	28.13	9
Car1	22	18	44	35.48	32	15	64	51.61	40	8	80	64.52	16	6	56	22.58	28
Car2	0	50	0	0.00	6	30	12	54.55	9	21	18	81.82	2	-2	13	68.18	15
Car3	0	46	0	0.00	5	45	10	29.41	16	30	32	94.12	2	-2	19	61.76	21
Car4	0	50	0	0.00	5	45	10	38.46	10	37	20	76.92	1	-1	14	57.69	15
Can1	34	11	68	37.78	37	11	74	41.11	41	5	82	45.56	11	23	67	5.56	10
Can2	1	49	2	9.09	6	44	12	54.55	8	42	16	72.73	2	-1	12	54.55	12
Can3	5	31	10	29.41	6	44	12	35.29	10	40	20	58.82	2	3	14	17.65	6
Can4	19	15	38	44.19	21	21	42	48.84	25	20	50	58.14	11	8	35	9.30	8

จากข้อมูลตารางที่ 21 ข้อมูลเปอร์เซ็นต์ความแม่นยำ (precision) และความครบถ้วน (recall) ของการสืบค้นด้วยวิธีธรรมชาติเปรียบเทียบกับ การสืบค้นโดยการนำฐานข้อมูลความสัมพัทธ์มาใช้ โดยแสดงในรูปแบบกราฟแยกตามกลุ่มโรคดังภาพที่ 17, 18 และ 19



P1 คือ ค่าความแม่นยำ (precision) ของการสืบค้นด้วยวิธีธรรมชาติ

P2 คือ ค่าความแม่นยำ (precision) ของการใช้อัลกอริทึมความสัมพันธ์ที่มีจำนวนคำ 2 คำต่อกลุ่ม

P3 คือ ค่าความแม่นยำ (precision) ของการใช้อัลกอริทึมความสัมพันธ์ที่มีจำนวนคำ 3 คำต่อกลุ่ม

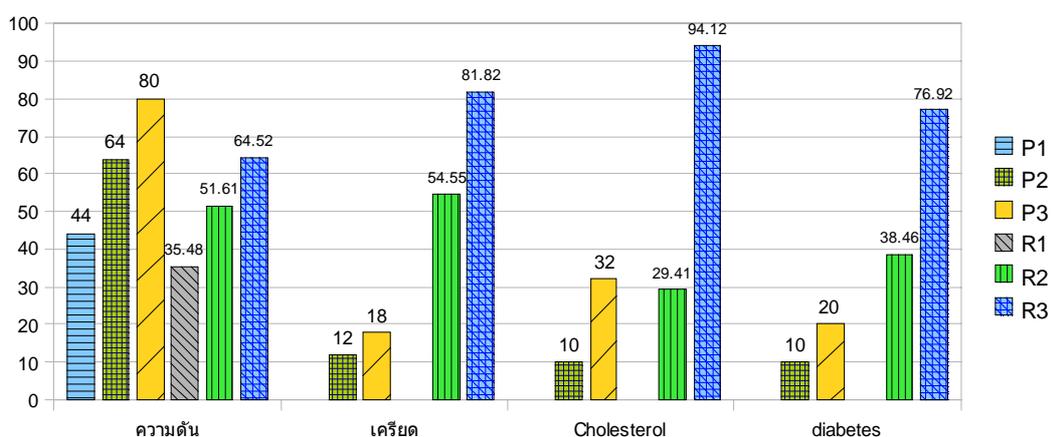
R1 คือ ค่าความครบถ้วน (recall) ของการสืบค้นด้วยวิธีธรรมชาติ

R2 คือ ค่าความครบถ้วน (recall) ของการใช้อัลกอริทึมความสัมพันธ์ที่มีจำนวนคำ 2 คำต่อกลุ่ม

R3 คือ ค่าความครบถ้วน (recall) ของการใช้อัลกอริทึมความสัมพันธ์ที่มีจำนวนคำ 3 คำต่อกลุ่ม

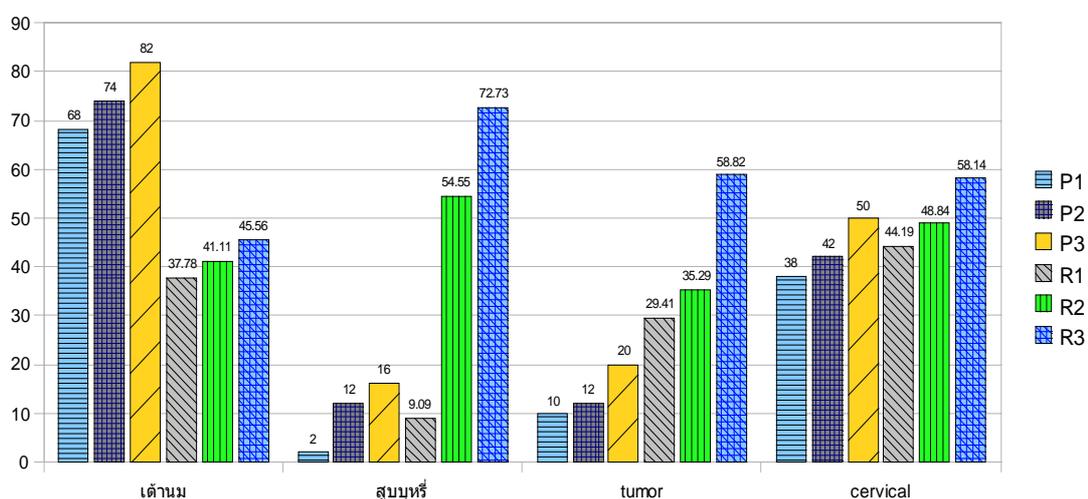
ภาพที่ 17 แสดงค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) ที่เพิ่มขึ้นของการสืบค้น การนำฐานข้อมูลความสัมพัทธ์มาใช้กลุ่มโรคระบบหายใจ

จากภาพที่ 17 แสดงเปรียบเทียบค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) ของการสืบค้นด้วยวิธีธรรมชาติเปรียบเทียบกับ การนำคำจากฐานข้อมูลคำทางความสัมพันธ์ มาใช้กลุ่มโรคระบบหายใจระหว่างจำนวนคำในกฎ 2 คำ และ 3 คำ ซึ่งมีค่าความแม่นยำ (precision) เพิ่มขึ้นเฉลี่ย 31% และค่าความครบถ้วน (recall) เพิ่มขึ้นเฉลี่ย 45.38%



ภาพที่ 18 แสดงค่าความแม่นยำ (precision) และความครบถ้วน (recall) ที่เพิ่มขึ้นของการสืบค้น การนำฐานข้อมูลกฎความสัมพันธ์มาใช้กลุ่มโรกระบบไหลเวียนเลือด

จากภาพที่ 18 แสดงเปรียบเทียบค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) ของการสืบค้นด้วยวิธีธรรมดาเปรียบเทียบกับการนำค่าจากฐานข้อมูลค่าทางความสัมพันธ์ มาใช้กลุ่มโรกระบบไหลเวียนเลือดระหว่างจำนวนค่าในกฎ 2 ค่า และ 3 ค่า ซึ่งมีค่าความแม่นยำ (precision) เพิ่มขึ้นเฉลี่ย 20% และค่าความครบถ้วน (recall) เพิ่มขึ้นเฉลี่ย 61.43%



ภาพที่ 19 แสดงค่าความแม่นยำ (precision) และความครบถ้วน (recall) ที่เพิ่มขึ้นของการสืบค้น การนำฐานข้อมูลกฎความสัมพันธ์มาใช้กลุ่มเนื้องอก (รวมมะเร็ง)

จากภาพที่ 13 แสดงเปรียบเทียบค่าความแม่นยำ (precision) และค่าความครบถ้วน (recall) ของการสืบค้นด้วยวิธีธรรมดาเปรียบเทียบกับการนำคำจากฐานข้อมูลคำทางความสัมพันธ์มาใช้กลุ่มเนื้องอก (รวมมะเร็ง) ระหว่างจำนวนคำในกฎ 2 คำ และ 3 คำ ซึ่งมีค่าความแม่นยำ (precision) เพิ่มขึ้นเฉลี่ย 22% และค่าความครบถ้วนเพิ่มขึ้น (recall) เฉลี่ย 21.76%

8. สรุปผลการทดสอบ

การนำฐานข้อมูลคำทางความสัมพันธ์มาช่วยเพิ่มให้กับคำค้นเดิมของผู้ใช้งานเพื่อการสืบค้นข้อมูลทางการแพทย์นั้น สามารถลดจำนวนเอกสารที่ไม่ถูกต้องลงได้มาก ทำให้ผลลัพธ์ที่ได้จากการสืบค้นมีจำนวนเอกสารที่ถูกต้องกับความต้องการของผู้ใช้มากขึ้น จากภาพที่ 17, 18 และ 19 แสดงให้เห็นว่าค่าความแม่นยำ (precision) ของเอกสารที่ได้รับจากการสืบค้นที่เพิ่มขึ้นจากการใช้คำจากฐานข้อมูลคำทางความสัมพันธ์มาใช้ระหว่างกฎที่มีจำนวนคำ 3 คำนั้นสามารถได้รับเอกสารจากการสืบค้นตรงความต้องการผู้ใช้โดยมีค่าความแม่นยำ (precision) เพิ่มขึ้นเฉลี่ย 10% และค่าความครบถ้วนที่เพิ่มขึ้น (recall) เฉลี่ยเท่ากับ 15.30% และเมื่อเปรียบเทียบการสืบค้นด้วยวิธีธรรมดากับการนำฐานข้อมูลคำทางความสัมพันธ์มาใช้พบที่พบว่าผู้ใช้งานสามารถได้รับเอกสารเกี่ยวข้องตรงความต้องการมากขึ้นโดยมีค่าความแม่นยำ (precision) เพิ่มขึ้นโดยเฉลี่ยเท่ากับ 20% และค่าความครบถ้วนที่เพิ่มขึ้น (recall) โดยเฉลี่ยเท่ากับ 39.90% เทคนิคใหม่นี้นอกจากช่วยให้ผลลัพธ์ที่ได้มีความถูกต้องกับความต้องการของผู้ใช้มากขึ้นแล้วยังช่วยประหยัดเวลาในการคัดเลือกเอกสารที่ผู้ใช้ต้องการจากชุดเอกสารผลลัพธ์ เนื่องจากเอกสารผลลัพธ์ที่ถูกต้องจะถูกนำมาแสดงในอันดับแรก ๆ ของชุดเอกสารผลลัพธ์

บทที่ 5

สรุปผลการดำเนินการวิจัย

สรุปผลการดำเนินการวิจัย

ผู้วิจัยได้เสนอขั้นตอนวิธีการพัฒนาการสืบค้นข้อมูลทางการแพทย์ ด้วยตระหนักว่าการเข้าถึงข้อมูลทางการแพทย์นั้นมีส่วนสำคัญในการดำเนินชีวิตในปัจจุบัน การทราบถึงข้อมูลพื้นฐานทางด้านดังกล่าวอย่างสะดวกและรวดเร็วจะเป็นประโยชน์ในการดูแลรักษาสุขภาพ

แต่ด้วยข้อมูลในอินเทอร์เน็ตในปัจจุบันมีจำนวนมาก อีกทั้งผู้ใช้งานทั่วไปขาดความรู้ความเข้าใจด้านการแพทย์ รวมถึงมีทักษะในการใช้งานเสิร์จเอนจินที่น้อยทำให้การระบุคีย์เวิร์ดเพื่อการสืบค้นเพื่อให้ได้เอกสารที่เกี่ยวข้องกับความต้องการของตนเป็นไปได้ยาก ผู้วิจัยจึงได้เสนองานวิจัยชิ้นนี้เพื่อช่วยให้ได้รับเอกสารที่มีความเกี่ยวข้องตรงตามความต้องการผู้ใช้งานมากที่สุด ซึ่งขั้นตอนในการดำเนินการสรุปได้ดังนี้

ขั้นตอนที่ 1 เป็นขั้นตอนในการจัดเตรียมข้อมูลและปรับโครงสร้างข้อมูล การรวบรวมเว็บเพจทางการแพทย์เพื่อการจัดทำฐานข้อมูลคำที่สัมพันธ์กัน, การกรองหรือคัดเลือกข้อมูลเพื่อให้ได้ข้อมูลที่เราต้องการ, การแปลงข้อมูลเอกสารเอสทีเอ็มแอลให้อยู่ในรูปแบบเอกสารข้อความเนื่องจากเนื้อหาที่เรานำมาดำเนินการมีด้วยกันสองภาษาซึ่งในแต่ละภาษานั้นจะมีลักษณะเฉพาะของภาษาแตกต่างกัน ซึ่งได้แก่ ภาษาไทยนั้นจำเป็นต้องดำเนินการตัดคำออกจากกันเพื่อใช้ในการสร้างความสัมพันธ์ของคำต่าง ๆ งานวิจัยนี้ใช้ (SWATH 2009) เป็นเครื่องมือในการดำเนินการ จากนั้นนำเข้าสู่กระบวนการกำจัดคำหยุด (ณิชาพร สุระ 2541 : 44 - 60) ในส่วนของภาษาอังกฤษนั้นจำเป็นต้องดำเนินการกำจัดคำหยุด (Pubmed Stopword List 2009) ก่อน จากนั้นจึงเข้าสู่กระบวนการหารากศัพท์โดยการหารากศัพท์ของ Porter (Snowball 2009) เพื่อลดทอนคำที่เกิดขึ้นที่มีรากศัพท์เดียวกันเพื่อให้ผลในขั้นตอนการคัดเลือกคำที่มีความถี่ได้คำที่หลากหลายขึ้น

ขั้นตอนที่ 2 การคัดเลือกคำที่มีความถี่ เนื่องจากเอกสารที่นำมาดำเนินการวิจัยประกอบด้วยคำจำนวนมากถึงแม้เราจะดำเนินการกำจัดคำหยุดออกไปบ้างแล้วแต่ปริมาณคำที่พบก็ยังคงมีปริมาณที่มาก หากนำมาสร้างกฎความสัมพันธ์จะได้กฎที่มีจำนวนมาก จึงจำเป็นต้องคัดเลือกคำจำนวนหนึ่งในแต่ละกลุ่มข้อมูลโดยการให้ค่าน้ำหนักของคำซึ่งใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร (Document Frequency) ซึ่งหมายความว่าคำดังกล่าวพบได้ในหลายเอกสาร

ขั้นตอนที่ 3 การคัดเลือกคำโดยผู้เชี่ยวชาญทางการแพทย์ เนื่องจากคำที่ได้จากการคัดเลือกคำโดยการให้ค่าน้ำหนักของคำโดยใช้ค่าเฉลี่ยเชิงความถี่ของคำนั้นต่อทุกเอกสาร (Document Frequency) คำที่ได้นั้นมีทั้งคำที่สื่อความหมายและคำที่ไม่สื่อความหมายใดหรืออาจไม่มีความสำคัญในทางการแพทย์ซึ่งยากต่อการนำไปใช้งาน ดังนั้นขั้นตอนนี้เป็นการคัดเลือกคำโดยผู้เชี่ยวชาญ ซึ่งประกอบด้วยเภสัชกร จำนวน 3 ท่าน โดยการเลือกคำดังกล่าวใช้วิธีการคัดเลือกโดยใช้หลักเกณฑ์ในการเลือกคือ คำที่สามารถนำไปใช้งานได้หรือมีประโยชน์ โดยคำนั้นจะต้องมีผู้เชี่ยวชาญเลือกอย่างน้อย 2 ท่านจาก 3 ท่าน ซึ่งขั้นตอนนี้ถือว่ามีความเหมาะสมเพราะนอกจากจะทำให้เกิดความสัมพันธ์ที่ถูกต้องขึ้นมามีคำที่ดีแล้ว ยังช่วยให้เกิดความสัมพันธ์ที่เกิดขึ้นมีปริมาณลดลงได้มาก

ขั้นตอนที่ 4 นำข้อมูลเอกสารเข้าสู่กระบวนการการสร้างความสัมพันธ์กระบวนการนี้จะสร้างกฎความสัมพันธ์ของคำที่ผ่านการคัดเลือกโดยผู้เชี่ยวชาญทางการแพทย์ โดยใช้ค่าสนับสนุนที่ 10 และค่าความเชื่อมั่นที่ 85 สำหรับข้อมูลภาษาไทย และค่าความเชื่อมั่นที่ 80 สำหรับกลุ่มข้อมูลภาษาอังกฤษ และกำหนดจำนวนคำหรือไอเท็มที่เกิดขึ้นในกฎแต่ละกฎไม่เกิน 3 คำหรือไอเท็ม ดังนั้นกฎที่เกิดขึ้นก็จะประกอบด้วยกฎที่มีจำนวนคำหรือไอเท็ม 2 คำและ 3 คำต่อกฎเท่านั้น ซึ่งการใช้ค่าสนับสนุนดังกล่าวส่งผลให้คำบางคำไม่สามารถนำมาใช้งานได้เนื่องจากมีจำนวนน้อยซึ่งอาจมีผลมาจากการรวบรวมข้อมูลที่ทำมาทำการทดลองมีปริมาณน้อย อีกทั้งข้อมูลอาจกล่าวถึงปัจจัยบางประการที่เฉพาะตามสถานการณ์ทำให้มีจำนวนที่น้อยกว่าข้อมูลทั่วไป

ขั้นตอนที่ 5 การวิเคราะห์ ออกแบบและทดสอบ งานวิจัยนี้เป็นการนำเอาการหาความสัมพันธ์มาประยุกต์ใช้เพื่อหาคำที่เกิดร่วมกัน โดยสนใจที่จะนำคำที่เกิดขึ้นในกฎมาใช้ทั้งกฎในรูปแบบคำรวมที่เกิดขึ้นด้วยกันเท่านั้น โดยไม่พิจารณาคำที่เกิดขึ้นก่อนที่ส่งผลทำให้เกิดอีกคำหนึ่ง การนำผลที่ได้จากการหาความสัมพันธ์ของคำมาวิเคราะห์พบว่ากฎที่เกิดขึ้นมีจำนวนมากและเมื่อพิจารณาพบว่ามีคำในกฎเกิดการซ้ำกันโดยค่าสนับสนุนและค่าความเชื่อมั่นเท่ากัน และค่าสนับสนุนกับค่าความเชื่อมั่นไม่เท่ากันจึงจำเป็นลดจำนวนกฎที่ซ้ำกันให้เหลือเพียงกฎเดียวทำให้จำนวนกฎมีปริมาณลดลงอย่างมาก ซึ่งค่าที่เท่ากันสามารถเลือกคำใดคำหนึ่งนำมาใช้งานได้เนื่องจากในการใช้งานจริง ผู้ใช้งานสามารถระบุคำที่อยู่ในกฎเพียงคำเดียวก็สามารถระบุคำที่มีความสัมพันธ์กันได้ จากนั้นนำรายการกฎความสัมพันธ์ของคำมาจัดทำฐานข้อมูลคำที่มีความสัมพันธ์กัน และจัดทำเว็บไซต์อินเทอร์เน็ตเพื่อทดสอบ โดยเว็บไซต์อินเทอร์เน็ตยังออกแบบให้ผู้เชี่ยวชาญทางการแพทย์สามารถเพิ่มคำเพื่อสร้างกฎความสัมพันธ์ได้อีกทางหนึ่งด้วย

ขั้นตอนที่ 6 การวิเคราะห์และประเมินผลการทดสอบ โดยผู้ทดสอบ 12 คน แบ่งเป็น 3 กลุ่มที่เท่ากัน โดยแบ่งทดสอบสองภาษาคือภาษาไทยและภาษาอังกฤษ ผู้ทดสอบสุ่มเลือกคำจาก

ฐานข้อมูลคนละ 1 คำค้นที่ไม่ซ้ำกัน จากนั้นนำคำค้นนั้นไปสืบค้นทางอินเทอร์เน็ตด้วยวิธีธรรมดา ซึ่งในที่นี้ใช้ระบบเสิร์จเอนจินของ google.co.th ซึ่งเป็นที่นิยมและรู้จักอย่างแพร่หลาย ผู้ตรวจสอบจะเป็นผู้ใช้ดุลพินิจตัดสินใจว่าผลลัพธ์ที่ได้จากการสืบค้นใน 50 เอกสารแรกและระบุว่าเอกสารผลลัพธ์นั้น ถูกต้องกับโรค, ไม่ถูกต้อง, ไม่แน่ใจ ใดๆอย่างหนึ่ง เปรียบเทียบกับการสืบค้นด้วยเทคนิคใหม่โดยใช้ฐานข้อมูลคำทางความสัมพันธ์ด้วยวิธีการเดียวกัน โดยการทดสอบจะเปรียบเทียบระหว่างการสืบค้นด้วยวิธีธรรมดากับการใช้กฎความสัมพันธ์มาช่วยในการขยายแบบสอบถาม ด้วยกฎความสัมพันธ์ที่มีจำนวนคำ 2 คำหรือ ไอเท็มและ 3 คำหรือ ไอเท็ม โดยการวัดค่าความแม่นยำ (precision) และความครบถ้วน (recall)

ขั้นตอนที่ 7 สรุปผล ผู้วิจัยได้ดำเนินการพัฒนาการสืบค้นข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์ ผลการทดสอบพบว่าเมื่อเอกสารที่เกี่ยวข้องตรงความต้องการผู้ใช้งานมากขึ้นเมื่อเทียบกับการสืบค้นด้วยวิธีธรรมดา จากการทดสอบแสดงให้เห็นว่าค่าความถูกต้องของเอกสารที่ได้รับจากการสืบค้นที่เพิ่มขึ้นจากการใช้คำจากฐานข้อมูลทางความสัมพันธ์มาใช้ระหว่างกฎที่มีจำนวนคำ 3 คำหรือ ไอเท็มนั้นสามารถได้รับเอกสารจากการสืบค้นตรงความต้องการผู้ใช้งานมากกว่ากฎที่มีจำนวนคำ 2 คำหรือ ไอเท็มโดยมีค่าความแม่นยำเพิ่มขึ้นเฉลี่ย 10% และค่าความครบถ้วนที่เพิ่มขึ้น (recall) เฉลี่ยเท่ากับ 15.30% และเมื่อเปรียบเทียบการสืบค้นด้วยวิธีธรรมดากับการนำฐานข้อมูลคำทางความสัมพันธ์มาใช้ นั้น พบว่าผู้ใช้งานสามารถได้รับเอกสารที่เกี่ยวข้องตรงความต้องการมากขึ้น โดยมีค่าความแม่นยำเพิ่มขึ้นโดยเฉลี่ยเท่ากับ 20% และค่าความครบถ้วนที่เพิ่มขึ้น (recall) โดยเฉลี่ยเท่ากับ 39.90% เทคนิคใหม่นี้นอกจากช่วยให้ผลลัพธ์ที่ได้มีความถูกต้องกับความต้องการของผู้ใช้มากขึ้นแล้วยังช่วยประหยัดเวลาในการคัดเลือกเอกสารที่ผู้ใช้ต้องการจากชุดเอกสารผลลัพธ์ เนื่องจากเอกสารผลลัพธ์ที่ถูกต้องจะถูกนำมาแสดงในอันดับแรกๆ ของชุดเอกสารผลลัพธ์

ปัญหาและแนวทางการวิจัยในอนาคต

การทดลองในกลุ่มข้อมูลภาษาไทยนั้นการตัดคำถือเป็นปัจจัยหนึ่งที่สำคัญ เนื่องจากประโยคบางประโยค เป็นลักษณะของคำประสม เมื่อผ่านกระบวนการตัดคำอาจได้คำที่เกิดขึ้นจำนวนมากแต่สื่อความหมายไม่ได้ และคำบางคำเป็นการเขียนทับศัพท์ด้วยภาษาไทยทำให้เสมือนเป็นคำประสม อาทิ แร่แอสเบสตอส ซึ่งเป็นแร่ที่นำไปสู่อាកอร์ในกลุ่มโรคปอด เมื่อผ่านการตัดคำจะได้ผลเป็น 5 คำ ได้แก่ แร่, แอ, สเบส, ตอ, ส ซึ่งจะส่งผลให้จำนวนคำที่เกิดขึ้นและส่งผลในการสร้างความสัมพันธ์กันเองในคำเดียวกัน และสื่อความหมายไม่ได้เช่นกัน จากปัญหาดังกล่าวหากมีการพัฒนาระบบการตัดคำเฉพาะด้านการแพทย์ที่มีประสิทธิภาพจะสามารถนำมาพัฒนาเพื่อให้ได้คำที่ดีและ

สามารถพัฒนาการสืบค้นจนครอบคลุมข้อมูลทางการแพทย์ทุกแขนงในอนาคต

อีกประเด็นที่น่าสนใจ เนื่องจากการใช้ฐานข้อมูลคำที่มีความสัมพันธ์ในงานวิจัยนี้อาจมีข้อดีโดยตรงที่การสื่อความหมายของคำซึ่งไม่ใช่ความสัมพันธ์กันเชิงความหมาย (semantic relation) ซึ่งหากนำไปพัฒนาหาความสัมพันธ์ร่วมกับฐานข้อมูลทางการแพทย์เพื่อสร้างเป็นฐานความรู้เชิงมโนภาพก็จะสามารถเพิ่มประสิทธิภาพการสืบค้นข้อมูลทางการแพทย์ได้ดียิ่งขึ้น

บรรณานุกรม

ภาษาไทย

กระทรวงสาธารณสุข. กรมการแพทย์. สถิติบริการข้อมูลด้านการป่วยของผู้ป่วยนอกและใน
ปีงบประมาณ 2550 [ออนไลน์]. เข้าถึงเมื่อ 15 มกราคม 2552. เข้าถึงได้จาก
<http://www.dms.moph.go.th/dmsstat/diseas50.htm>

กฤษณะ ไวยมัย, จิตชนก ส่งศิริ และธนาวิทย์ รักธรรมานนท์. “การใช้เทคนิคดาต้าไมน์นิ่ง
 เพื่อพัฒนาคุณภาพการศึกษานิสิตคณะวิศวกรรมศาสตร์.” The Nectec Technical
 Journal 3, 11 (2001) : 134-142.

กฤษณะ ไวยมัย และธีระวัฒน์ พงษ์ศิริปริดา. “การใช้เทคนิค Association Rule Discovery
 เพื่อการจัดสรรกฎหมายในการพิจารณาคดีความ.” The Nectec Technical Journal 3, 11
 (2001) : 143-152.

ณิชพร สุระ. “การจำแนกหมวดหมู่เอกสารภาษาไทยอัตโนมัติโดยใช้อัลกอริทึม FPTC.”
 วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย
 สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2541.

รุจิรา อาชวานันทกุล และ ยืน ภู่วรรณ. “การเพิ่มประสิทธิภาพของระบบสืบค้นโดยใช้ฐานความรู้
 มโนภาพ.” ใน Conference on Internet Technology 2003 & Workshop on UniNet
 Network and Computer Applications. n.p. , 2003.

รัชนิกร เทวอักษร. “การประยุกต์ใช้พจนานุกรมอิงฐานข้อมูลคลังประโยคสำหรับระบบการขยายคำ
 ค้นด้วยความสัมพันธ์ทางอรรถศาสตร์ของคำ”วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขา
 วิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี, 2546.

ศุภชัย ตั้งวงศ์สานต์. ระบบการจัดเก็บและการสืบค้นสารสนเทศด้วยคอมพิวเตอร์. กรุงเทพฯ: โรง
 พิมพ์พิทักษ์การพิมพ์, 2551

ภาษาต่างประเทศ

Aly,A.A.Using aQuery Expansion Technique to Improve Document Retrieval.Information
 Technologies and Knowledge.2008,343-348

- Fonseca, Bruno M et al. "ConceptBased Interactive Query Expansion." In Proceedings of the 14th ACM international conference on Information and knowledge management. n.p. 2005.
- Han Jiawei, and Micheline Kamber. Data Mining Concepts and Techniques. USA : Morgan Kaufman, 2001.
- K.Shin, S.-Y.Han, A.Gelbukh,J.Park. Advanced Relevance Feedback Query Expansion Strategy for Information Retrieval in MEDLINE. Progress in Pattern Recognition, Image Analysis and Applications, 9th Iberoamerican Congress on Pattern Recognition, CIARP 2004, Puebla, Mexico, October 26-29, 2004, Proceedings 2004
- Maedche, Alexander, and Staab Steffen. "Discovering Conceptual Relations from Text : W.Horn (ed.): ECAI 2000." In Proceedings of the 14th European Conference on Artificial Intelligence. Berlin : IOS Press, Amsterdam, 2000.
- Navigli R Velardi P, and Gangemi A. "Ontology Learning and Its Application to Automated Terminology Translation." In IEEE Intelligent Systems & Their Applications. n.p. : 22-31, 2003.
- Nguyen, Hung V et al. "Mining "Hidden Phrase" Definitions from the Web." In APWeb 2003: LNCS, Springer Verlag. n.p. :156-165, 2003.
- NirSoft [pseud]. HTMLAsText v1.11 - HTML to text freeware converter [Online]. Accessed 19 August 2009. Available from <http://www.nirsoft.net/utils/htmlastext.html/>
- Porter M.F. An algorithm for suffix stripping ,1980 [Online]. Accessed 19 January 2009. Available from <http://tartarus.org/~martin/PorterStemmer/def.txt>
- Snowball [pseud]. Software: Several tarballs of the snowball sources are available [Online]. Accessed 19 October 2009. Available from <http://snowball.tartarus.org/>
- SWATH [pseud]. Software: SWATH - Thai Word Segmentation [Online]. Accessed 20 May 2009. Available from <http://www.cs.cmu.edu/~paisarn/software.html>
- Togaware [pseud]. Gnome Data Mine Tools [Online]. Accessed 19 October 2009. Available from <http://www.togaware.com/datamining/gdatamine/>
- U.S. National Library of Medicine. Pubmed Stopword List [Online]. Accessed 15 January 2009. Available from <http://www.medparse.com/umlsstop.htm>

Weka 3 [pseud]. Data Mining Software in Java [Online]. Accessed 19 January 2009. Available from <http://www.cs.waikato.ac.nz/~ml/weka/index.html/>

ภาคผนวก

ภาคผนวก ก
รายการคำหยุดที่ใช้ในงานวิจัย

รายการคำหยุดที่ใช้ในงานวิจัย

a	anything	came	enough
about	anywhere	can	especially
above	applicable	cannot	et
abs	apply	cc	etc
accordingly	are	cm	ever
across	arise	come	every
after	around	compare	everyone
afterwards	as	could	everything
again	assume	de	everywhere
against	at	dealing	except
all	be	department	find
almost	became	depend	for
alone	because	did	found
along	become	discover	from
already	becomes	dl	further
also	becoming	do	gave
although	been	does	get
always	before	done	give
am	beforehand	due	go
among	being	during	gone
amongst	below	each	got
an	beside	ec	gov
analyze	besides	ed	had
and	between	effected	has
another	beyond	eg	have
any	both	either	having
anyhow	but	else	he
anyone	by	elsewhere	hence

her	it	moreover	often
here	its	most	on
hereafter	itself	mostly	only
hereby	just	mr	onto
herein	keep	much	other
hereupon	kept	mug	others
hers	kg	must	otherwise
herself	km	my	ought
him	last	myself	our
himself	latter	namely	ours
his	latterly	nearly	ourselves
how	lb	necessarily	out
however	ld	neither	over
hr	letter	never	overall
i	like	nevertheless	owing
ie	ltd	next	own
if	made	no	oz
ii	mainly	nobody	particularly
iii	make	noone	per
immediately	many	nor	perhaps
importance	may	normally	pm
important	me	nos	precede
in	meanwhile	noted	predominantly
inc	mg	nothing	present
incl	might	now	presently
indeed	ml	nowhere	previously
into	mm	obtained	primarily
investigate	mo	of	promptly
is	more	off	pt

quickly	so	thence	upon
quite	some	there	us
rather	somehow	thereafter	use
readily	someone	thereby	used
really	something	therefore	usefully
recently	sometime	therein	usefulness
refs	sometimes	thereupon	using
regarding	somewhat	these	usually
relate	somewhere	they	various
said	soon	this	very
same	specifically	thorough	via
seem	still	those	was
seemed	strongly	though	we
seeming	studied	through	were
seems	sub	throughout	what
seen	substantially	thru	whatever
seriously	such	thus	when
several	sufficiently	to	whence
shall	take	together	whenever
she	tell	too	where
should	th	toward	whereafter
show	than	towards	whereas
showed	that	try	whereby
shown	the	type	wherein
shown	their	ug	whereupon
shows	theirs	under	wherever
significantly	them	unless	whether
since	themselves	until	which
slightly	then	up	while

whither	ก็ได้อ	กันและกัน	ขณะเดียวกัน
who	ก็ต่อเมื่อ	กันใหม่	ขณะใด
whoever	ก็ตาม	กันเอง	ขณะใดๆ
whom	ก็ตามแต่	กับ	ขณะที่
whose	ก็ตามที	การ	ขณะนั้น
why	กระทั่ง	กำลัง	ขณะนี้
will	กระทำ	กำลังจะ	ขณะหนึ่ง
with	กระนั้น	กำหนด	ขวาง
within	ระดม	ดู	ขวางๆ
without	กลับ	เก็บ	ขอ
wk	กล่าว	เกิด	ของ
would	กล่าวคือ	เกิน	ขึ้น
wt	กลุ่ม	เกินๆ	ซ้ำ
yet	กลุ่มก้อน	เกี่ยวกัน	ข้าง
you	กลุ่มๆ	เกี่ยวกับ	ข้างเคียง
your	ก็แล้วแต่	เกี่ยวข้อง	ข้างต้น
yours	กว่า	เกี่ยวเนื่อง	ข้างบน
yourself	กว้าง	เกี่ยวๆ	ข้างล่าง
yourselves	กว้างขวาง	เกือบ	ข้างๆ
yr	กว้างๆ	เกือบจะ	ขาด
เ	ก่อน	เกือบๆ	ซ้ำพเจ้า
แ	ก่อนหน้า	แก่	ซ้ำๆ
ใ	ก่อนหน้านี้	แก่	ขึ้น
ไ	ก่อนๆ	แก้ไข	เขา
ก็	กัน	ใกล้	เข้า
ก็คือ	กันดีกว่า	ใกล้ๆ	เข้าใจ
ก็แค่	กันดีไหม	ไกล	เขียน
ก็จะ	กันเถอะ	ไกลๆ	คง
ก็ดี	กันนะ	ขณะ	คงจะ

คงอยู่	คราวที่	คำ	จนถึง
ครบ	คราวนั้น	คิด	จนทั่ว
ครบครัน	คราวนี้	คิดว่า	จนบัดนี้
ครบถ้วน	คราวโน้น	คือ	จนเมื่อ
ครั้ง	คราวละ	คุณ	จนแม่
ครั้งกระนั้น	คราวหน้า	คุณๆ	จนแม่น
ครั้งก่อน	คราวหนึ่ง	เคย	จรด
ครั้งครา	คราวหลัง	เคยๆ	จรดกับ
ครั้งคราว	คราวไหน	แค่	จริง
ครั้งใด	คราวๆ	แค่จะ	จริงจัง
ครั้งที่	คราหนึ่ง	แค่นั้น	จริงๆ
ครั้งนั้น	คราไหน	แค่นี้	จริงๆจังๆ
ครั้งนี้	คล้าย	แค่เพียง	จน
ครั้งละ	คล้ายกัน	แค่ว่า	จนจะ
ครั้งหนึ่ง	คล้ายกันกับ	แค่ไหน	จนเจียน
ครั้งหลัง	คล้ายกับ	ใคร	จบ
ครั้งหลังสุด	คล้ายกับว่า	ใคร่	จบกับ
ครั้งไหน	คล้ายว่า	ใคร่จะ	จบจน
ครั้งๆ	ควร	ใครๆ	จะ
ครัน	ความ	ง่าย	จ๊ะ
ครับ	ค่อนข้าง	ง่ายๆ	จ๊ะ
ครา	ค่อนข้าง	ไง	จะได้
คราใด	ค่อนข้างจะ	จง	จัง
คราที่	ค่อนข้าง	จด	จังๆ
ครานั้น	ค่อนข้าง	จน	จัด
คราวนี้	ค่อย	จนกระทั่ง	จัดการ
คราว	ค่อยๆ	จนกว่า	จัดงาน
คราวก่อน	คะ	จนขณะนี้	จัดแจง
คราวใด	คะ	จนตลอด	จัดตั้ง

จัดทำ	ช่วงท้าย	เช่นที่ว่า	ด้วยที่
จัดหา	ช่วงที่	เช่นนั้น	ด้วยประการฉะนี้
จัดให้	ช่วงนั้น	เช่นนั้นเอง	ด้วยเพราะ
จับ	ช่วงนี้	เช่นนี้	ด้วยว่า
จับ	ช่วงระหว่าง	เช่นเมื่อ	ด้วยเหตุที่
จับ	ช่วงแรก	เช่นไร	ด้วยเหตุนี้
จาก	ช่วงหน้า	เชื่	ด้วยเหตุนี้
จากนั้น	ช่วงหลัง	เชื่อถือ	ด้วยเหตุเพราะ
จากนี้	ช่วงๆ	เชื่อมั่น	ด้วยเหตุว่า
จากนี้ไป	ช่วย	เชื่อว่า	ด้วยเหมือนกัน
จำ	ช่วยๆ	ใช่	ดั่ง
จำเป็น	ซ้ำ	ใช่	ดั่ง
จำพวก	ซ้ำนาน	ใช่ใหม่	ดั่งกล่าว
จึง	ชาว	ชะ	ดั่งกับ
จึงจะ	ซ้ำๆ	ชะก่อน	ดั่งกับ
จึงเป็น	เช่น	ชะจน	ดั่งกับว่า
จู่ๆ	เช่นก่อน	ชะจนกระทั่ง	ดั่งกับว่า
ฉะนั้น	เช่นกัน	ชะจนถึง	ดั่งเก่า
ฉะนั้น	เช่นเคย	ซึ่ง	ดั่งเก่า
ฉนั้น	เช่นดั่ง	ซึ่งก็	ดั่งเคย
เฉกเช่น	เช่นดั่งก่อน	ซึ่งก็คือ	ดั่งเคย
เฉพาะ	เช่นดั่งเก่า	ซึ่งกัน	ดั่งจะ
เฉย	เช่นดั่งที่	ซึ่งกันและกัน	ดั่งจะ
เฉยๆ	เช่นดั่งว่า	ซึ่งได้แก่	ดั่งเช่น
ไฉน	เช่นเดียวกัน	ซึ่งๆ	ดั่งเช่น
ช่วง	เช่นเดียวกับ	ณ	ดั่งเช่น
ช่วงก่อน	เช่นใด	ด้วย	ดั่งเช่น
ช่วงต่อไป	เช่นที่	ด้วยกัน	ดั่งเช่นที่
ช่วงถัดไป	เช่นที่เคย	ด้วยเช่นกัน	ดั่งเช่นที่

ดั้งเดิม	ดิฉัน	โดยทั่ว	ได้รับ
ดั้งเดิม	ดี	โดยทั่วกัน	ตน
ดังต่อไปนี้	ดี ๆ	โดยทั่วถึง	ตนเอง
ตั้งแต่ก่อน	ดู	โดยทั่วไป	ตน ๆ
ตั้งแต่ก่อน	ดูจะ	โดยทั่ว ๆ ไป	ตรง
ดังที่	ดูแล	โดยที่	ตรง ๆ
ดังที่	ดูแลแล้ว	โดยแท้	ตลอด
ดังที่กล่าว	ดูว่า	โดยแท้จริง	ตลอดกาล
ดังที่เคย	ดูเหมือน	โดยนัย	ตลอดกาลนาน
ดังที่จะ	ดูเหมือนว่า	โดยปกติ	ตลอดจน
ดังที่เป็น	ดู ๆ	โดยมัก	ตลอดถึง
ดั่งนั้น	เดิม	โดยมักจะ	ตลอดทั้ง
ดั่งนี้	เดิมที	โดยมาก	ตลอดทั่ว
ดั่งนี้เช่น	เดิม ๆ	โดยเมื่อ	ตลอดทั่วถึง
ดั่งนี้เพราะ	เดียว	โดยรวม	ตลอดทั่วทั้ง
ดั่งแม่	เดียว	โดยรวม ๆ	ตลอดปี
ดั่งแม่	เดียวก่อน	โดยเร็ว	ตลอดไป
ดั่งแม่น	เดียวกัน	โดยละม่อม	ตลอดมา
ดั่งแม่น	เดียวกับ	โดยลำดับ	ตลอดระยะเวลา
ดั่งว่า	เดียนั้น	โดยส่วนมาก	ตลอดวัน
ดั่งว่า	เดียนี้	โดยส่วนรวม	ตลอดเวลา
ดั่งเหมือน	แต่	โดยส่วนใหญ่	ตลอดศก
ดั่งเหมือน	โดย	ใด	ต่อ
ด้าน	โดยง่าย	ใดๆ	ต่อกัน
ด้าน ๆ	โดยเฉพาะ	ได้	ต่อกับ
ดำเนิน	โดยเฉพาะอย่างยิ่ง	ได้แก่	ต้อง
ดำเนินการ	โดยดี	ได้แต่	ต้องการ
ดำเนินงาน	โดยคุณ	ได้ที่	ต่อจากนั้น
ดำเนินไป	โดยตลอด	ได้มา	ต่อจากนี้

ต่อด้วย	ตั้ง	ตามๆ	ถึงจะ
ต่อแต่นี้	ตั้งต้น	เต็มไปด้วย	ถึงบัดนั้น
ตอน	ตั้งแต่	เต็มไปหมด	ถึงบัดนี้
ตอนก่อน	ตั้งแต่นั้น	เต็มๆ	ถึงเมื่อ
ตอนใด	ตั้งแต่นี้	แต่	ถึงเมื่อใด
ตอนต่อ	ตั้งแต่แรก	แต่ก็	ถึงเมื่อไร
ตอนต่อไป	ตั้งที่	แต่ก่อน	ถึงแม้
ตอนต่อมา	ตั้งอยู่	แต่จะ	ถึงแม้จะ
ตอนถัดไป	ตั้งๆ	แต่เดิม	ถึงแม้ว่า
ตอนถัดมา	ตัว	แต่ต้อง	ถึงอย่างไร
ตอนที่	ตัวใด	แต่ถ้า	ถือ
ตอนที่แล้ว	ตัวที่	แต่ทว่า	ถือว่า
ตอนนั้น	ตัวนั้น	แต่ที่	ถูก
ตอนนี้	ตัวนี้	แต่นั้น	ถูกต้อง
ตอนแรก	ตัวโน้น	แต่เพียง	ถูกๆ
ตอนสุดท้าย	ตัวละ	แต่เมื่อ	เถอะ
ตอนหน้า	ตัวไหน	แต่ไร	เถิด
ตอนหลัง	ตัวอย่างเช่น	แต่ละ	ทรง
ตอนไหน	ตัวเอง	แต่ว่า	ทว่า
ตอนๆ	ตัวๆ	แต่ไหน	ทั้ง
ต่อเนื่อง	ต่าง	แต่อย่างใด	ทั้งคน
ต่อไป	ต่างก็	โต	ทั้งตัว
ต่อไปนี้	ต่างหาก	โตๆ	ทั้งที่
ต่อมา	ต่างๆ	ใต้	ทั้งที่
ต่อเมื่อ	ตาม	ถ้า	ทั้งนั้น
ต่อว่า	ตามด้วย	ถ้าจะ	ทั้งนั้นด้วย
ต่อให้	ตามแต่	ถ้าหาก	ทั้งนั้นเพราะ
ต่อๆ	ตามที่	ถึง	ทั้งนี้
ตะหาก	ตามที่	ถึงแก่	ทั้งนี้เช่น

ทั้งนี้ด้วย	ทำไม	ทุก	แท้จริง
ทั้งนี้เพราะ	ทำไม	ทุกคน	เธอ
ทั้งปวง	ทำให้	ทุกครั้ง	นอก
ทั้งเป็น	ทำๆ	ทุกครา	นอกจาก
ทั้งมวล	ที่	ทุกคราว	นอกจากที่
ทั้งสิ้น	ที่	ทุกชั้น	นอกจากนั้น
ทั้งหมด	ที่จริง	ทุกตัว	นอกจากนี้
ทั้งหลาย	ที่ซึ่ง	ทุกทาง	นอกจากว่า
ทั้งๆ	ที่เดียว	ทุกที่	นอกนั้น
ทั้งๆที่	ที่ใด	ทุกที่	นอกเหนือ
ทัน	ที่ใด	ทุกเมื่อ	นอกเหนือจาก
ทันใด	ที่ได้	ทุกวัน	น้อย
ทันใดนั้น	ที่เถอะ	ทุกวันนี้	น้อยกว่า
ทันที	ที่แท้	ทุกสิ่ง	น้อยๆ
ทันทีทันใด	ที่แท้จริง	ทุกหน	นะ
ทั่ว	ที่นั่น	ทุกแห่ง	นะ
ทั่วกัน	ที่นี่	ทุกอย่าง	นัก
ทั่วถึง	ที่นี้	ทุกอัน	นักๆ
ทั่วถึงกัน	ที่นี้	ทุกๆ	นั้น
ทั่วทั้ง	ที่ไร	เท่า	นั้น
ทั่วไป	ที่ละ	เท่ากัน	นั้น ใจ
ทั่วๆ	ที่ละ	เท่ากับ	นั้นเป็น
ทั่วๆไป	ที่แล้ว	เท่าใด	นั้นแหละ
ทาง	ที่ว่า	เท่าที่	นั้นเอง
ทางๆ	ที่สุด	เท่านั้น	นั้นๆ
ท่าน	ที่แห่งนั้น	เท่านี้	นับ
ท่ามกลาง	ที่ไหน	เท่าไร	นับจากนั้น
ทำ	ที่ๆ	เท่าไร	นับจากนี้
ทำงาน	ที่ๆ	เท่า	นับตั้งแต่

นับแต่	นับ	บ่อยๆ	ปรากฏว่า
นับแต่ที่	เน้น	บัดคล	ปัจจุบัน
นับแต่นั้น	เน้น	บัดเดียว	ปิด
นับแต่นี้	เน้นเอง	บัดเดี๋ยวนี้	เป็น
นำ	เนื่อง	บัดนั้น	เป็นด้วย
นาง	เนื่องจาก	บัดนี้	เป็นดัง
นางสาว	เนื่องด้วย	บาง	เป็นต้น
น่าจะ	เนื่องถึง	บ้าง	เป็นต้นไป
นาน	เนื่องมาจาก	บางกว่า	เป็นต้นมา
นานๆ	แน่	บางขณะ	เป็นแต่
นาย	แน่ะ	บางครั้ง	เป็นแต่เพียง
นำ	โน่น	บางครา	เป็นที่
นำพา	โน้น	บางคราว	เป็นที่
นำมา	โน่นไง	บางที่	เป็นที่สุด
นำมาซึ่ง	โน่นแน่ะ	บางที่	เป็นเพราะ
นิ	ใน	บางหน	เป็นเพราะว่า
นิค	ในช่วง	บางแห่ง	เป็นเพียง
นิคหน่อย	ในที่	บางๆ	เป็นเพียงว่า
นิตๆ	ในเมื่อ	แบบ	เป็นเพื่อ
นี้	ในระหว่าง	ปฏิบัติ	เป็นอัน
นี้	ในอันที่	ประกอบ	เป็นอันมาก
นี้ไง	ในอันที่จะ	ประการ	เป็นอันว่า
นี้นา	บน	ประการละนี้	เป็นอันๆ
นี้แน่ะ	บอก	ประการใด	เป็นอาทิ
นี้แหละ	บอกแล้ว	ประการหนึ่ง	เป็นๆ
นี้แหละ	บอกว่า	ประมาณ	เปลี่ยน
นี้เอง	บ่อย	ประสพ	เปลี่ยนแปลง
นี้เอง	บ่อยกว่า	ปรับ	เปิด
นุ่น	บ่อยครั้ง	ปรากฏ	เปิดเผย

ไป	พร้อมกับ	พวกเขา	เพิ่ง
ไป	พร้อมด้วย	พอ	เพิ่งจะ
ผ่าน	พร้อมทั้ง	พอกัน	เพิ่ม
ผ่านๆ	พร้อมที่	พอควร	เพิ่มเติม
ผิด	พร้อมที่จะ	พอจะ	เพียง
ผิดๆ	พร้อมเพียง	พอดี	เพียงแต่
ผู้	พร้อมๆ	พอดี	เพียงใด
ผู้ใด	พร้อมๆกัน	พอทำเนา	เพียงแต่
เดิน	พร้อมๆกับ	พอที่	เพียงพอ
เดินๆ	พร้อมๆด้วย	พอที่	เพียงเพราะ
เพื่อ	พร้อมๆทั้ง	พอที่จะ	เพียงเพื่อ
เพื่อจะ	พวก	พอเพียง	เพียงไร
เพื่อที่	พวกกัน	พอแล้ว	เพียงไหน
เพื่อว่า	พวกกู	พอสมควร	เพื่อ
เพื่อๆ	พวกแก	พอสมควร	เพื่อที่
ฝึกฝาย	พวกเขา	พอเหมาะ	เพื่อที่จะ
ฝึกฝั	พวกคุณ	พอๆ	เพื่อว่า
ฝาย	พวกฉัน	พอๆกัน	เพื่อให้
ฝายใด	พวกท่าน	พัฒนา	ภาค
ฝายๆ	พวกที่	พา	ภาคๆ
พณฯ	พวกเธอ	พิจารณา	กาย
พบ	พวกนั้น	พึง	กายได้
พบว่า	พวกนี้	พึง	ภายนอก
พยายาม	พวกนั้น	พื่นๆ	ภายใน
พร้อม	พวกโน้น	พูด	กายภาค
พร้อมกัน	พวกมัน	เพราะ	กายภาคหน้า
พร้อมกันกับ	พวกมึง	เพราะฉะนั้น	กายหน้า
พร้อมกันนั้น	พวกเรา	เพราะว่า	กายหลัง
พร้อมกันนี้	พวกไหน	เพราะๆ	มอง

มองว่า	เมื่อคราว	ไม่เป็นไร	ยื่นเตือน
มัก	เมื่อคราวก่อน	ไม่ว่า	ยื่น
มักจะ	เมื่อคราวที่	ไม่เสียที	ยื่นกว่า
มัน	เมื่อคืน	ไม่ไหว	ยื่นขึ้น
มันๆ	เมื่อเช้า	ยก	ยื่นขึ้นไป
มีย	เมื่อใด	ยกให้	ยื่นจน
มียนะ	เมื่อนั้น	ยอม	ยื่นจะ
มียนั่น	เมื่อนี้	ยอม	ยื่นัก
มียเนี่ย	เมื่อเฮ็น	ยอมรับ	ยื่นไปกว่า
มียละ	เมื่อไร	ยอมรับว่า	ยื่นไปกว่านั้น
มา	เมื่อวันวาน	ยอมๆ	ยื่นเมื่อ
มาก	เมื่อวาน	ยอม	ยื่นแล้ว
มากกว่า	เมื่อไหร่	ยอมๆ	ยื่นใหญ่
มากมาย	แม่	ยัง	ยื่นๆ
มากๆ	แม่กระทั่ง	ยังคง	ยื่นนาน
มิ	แม่แต่	ยังงั้น	ยื่นยง
มิฉะนั้น	แม่น	ยังงี้	ยื่นยัน
มิใช่	แม่นกระทั่ง	ยังงั้น	ยื่นยาว
มิได้	แม่นว่า	ยังง	เยอะ
มิ	แม่นหาก	ยังงั้นกัน	เยอะแยะ
มีแต่	แม่นเหมือน	ยังงั้นชะ	เยอะๆ
มีง	แม้ว่า	ยังงั้นเสีย	เยะ
มุ่ง	แม่หาก	ยื่นจะ	เยะๆ
มุ่งเน้น	ไม่	ยังแต่	รวด
มุ่งหมาย	ไม่ค่อย	ยาก	รวดเร็ว
เมื่อ	ไม่ค่อยจะ	ยาว	รวม
เมื่อก่อน	ไม่ค่อยเป็น	ยาวนาน	ร่วม
เมื่อครั้ง	ไม่ใช่	ยาวๆ	รวมกัน
เมื่อครั้งก่อน	ไม่แต่	ยื่น	รวมกัน

รวมด้วย	แรก	วันไหน	ส่วนมาก
ร่วมด้วย	แรกๆ	วันๆ	ส่วนหนึ่ง
รวมถึง	ไร	ว่า	ส่วนใหญ่
รวมทั้ง	ลง	วาง	ส่วนไหน
ร่วมมือ	ล้วน	วางไว้	สอง
รวมๆ	ล้วนจน	ว่าด้วย	สะดวก
ระยะ	ล้วนแต่	ไว้	สั่ง
ระยะๆ	ล้วนแล้ว	ส่ง	สั่ง
ระหว่าง	ล้วนแล้วแต่	ส่งๆ	สั่งๆ
รับ	ล้วนๆ	สนใจ	สามารถ
รับรอง	ละ	สบาย	สำคัญ
รู้	ละ	สบายๆ	สำหรับ
รู้ว่า	ล่าง	สมัย	สิ่ง
รู้	ล่างสุด	สมัยก่อน	สิ่งใด
รู้ว่า	เล็ก	สมัยที่	สิ่งนั้น
เร็ว	เล็กน้อย	สมัยนั้น	สิ่งนี้
เร็วๆ	เล็กๆ	สมัยนี้	สิ่งไหน
เรา	เลย	สมัยโน้น	สิ้น
เราๆ	เล่า	สมัยเมื่อ	สิ้นกาลนาน
เริ่ม	เล่าว่า	สร้าง	สืบเนื่อง
เรียก	เลือก	ส่วน	สุด
เรียกเรื่อง	แล้ว	ส่วนเกิน	สุดๆ
เรียบ	แล้วกัน	ส่วนค้อย	สู่
เรียบร้อย	แล้วแต่	ส่วนดี	สูง
เรียบๆ	แล้วเสร็จ	ส่วนใด	สูงกว่า
เรื่อง	และ	ส่วนที่	สูงส่ง
เรื่องๆ	วันใด	ส่วนน้อย	สูงสุด
เรื่อย	วันนั้น	ส่วนนั้น	สูงๆ
เรื่อยๆ	วันนี้	ส่วนนี้	เสมอกับ

เสมือนว่า	หนึ่ง	หากว่า	เหมือนกัน
เสร็จ	หมด	หากแม่	เหมือนกันกับ
เสร็จกัน	หมดกัน	หาก मैं	เหมือนกันว่า
เสร็จแล้ว	หมดสิ้น	หาก मैंว่า	เหมือนกับ
เสร็จสมบูรณ์	หมายความว่า	หาก मैंว่า	เหมือน मैं
เสร็จสิ้น	หมายความว่าถึง	หากว่า	เหมือนว่า
เสีย	หมายความว่า	หากว่า	เหล่า
เสียก่อน	หมายใจ	หาใช่	เหล่านั้น
เสียจน	หมายถึง	หาใช่ไม่	เหล่านี้
เสียจนกระทั่ง	หรือ	หาไม่	เหลือ
เสียจนถึง	หรือไง	หาหรือ	เหลือเกิน
เสียด้วย	หรือเปล่า	เหตุใด	แห่ง
เสียนั้น	หรือไม่	เหตุ นั้น	แห่งใด
เสีย นั่นเอง	หรือยัง	เหตุ นี้	แห่ง นั้น
เสีย นี้	หรือไร	เหตุ ไร	แห่ง นี้
เสีย นี้ ไร	หรือว่า	เห็น	แห่ง ไนั้น
เสีย ยิ่ง	หรืออย่างไร	เห็น แก่	แห่ง ไไหน
เสีย ยิ่ง นึก	หลัก	เห็น ควร	แหละ
เสีย แล้ว	หลักๆ	เห็น จะ	ให้
เสีย สิ้น	หลังจาก	เห็น ดี	ให้ แก่
แสดง	หลังๆ	เห็น ว่า	ใหญ่
แสดง ว่า	หลาก	เห็น สมควร	ใหญ่ โต
หน	หลากหลาย	เห็น เหมาะ	ใหญ่ ๆ
หนอ	หลาย	เห็น ๆ	ให้ ดี
หนอย	หลายๆ	เหมาะ	ให้ แค่
หน้อย	หา	เหมาะ ควร	ให้ ไป
หนอย แน่	หาก	เหมาะ สม	ใหม่
หนอย แน่ะ	หาก กระนั้น	เหมาะ ๆ	ให้ มา
หน้อย ๆ	หาก แต่	เหมือน	ใหม่ ๆ

ไหน	อย่างไรก็	อะไร	อาจเป็นเพราะ
ไหนๆ	อย่างไรก็ดี	อ่ะๆ	อาจเพราะ
อดีต	อย่างไรก็ตาม	อัน	อาทิ
อนึ่ง	อย่างไรกัน	อันจะ	อาทิเช่น
อยาก	อย่างไรชะ	อันใด	อ่าน
อย่าง	อย่างไรเล่า	อันได้แก่	อีก
อย่างเช่น	อย่างไรเสีย	อันที่	อื่น
อย่างดี	อย่างละ	อันที่จริง	อื่นๆ
อย่างเดียว	อย่างหนึ่ง	อันที่จะ	เอง
อย่างไร	อย่างไร	อันเนื่องจาก	เอ็ง
อย่างที่	ต่างๆ	อันเนื่องมาจาก	เอา
อย่างน้อย	อยู่	อันละ	เอาแต่
อย่างนั้น	อยู่ๆ	อันไหน	ฯ
อย่างนี้	ออก	อันๆ	ฯล
อย่างโน้น	อะ	อาจ	ฯลฯ
อย่างมาก	อ่ะ	อาจจะ	ะ
อย่างยิ่ง	อ๊ะ	อาจเป็น	า
อย่างไร	อ๊ะ	อาจเป็นด้วย	ๆ

ภาคผนวก ข

รายชื่อแหล่งข้อมูลเว็บไซต์ที่ใช้ในงานวิจัย

รายชื่อแหล่งข้อมูลเว็บไซต์ภาษาไทย

ชื่อหน่วยงาน	ชื่อเรียกเว็บไซต์
คณะแพทยศาสตร์ศิริราชพยาบาล	http://www.si.mahidol.ac.th/project/geriatrics/index.htm
ทะเบียนมะเร็ง โรงพยาบาลสงขลานครินทร์	http://medinfo2.psu.ac.th/cancer/index.php
มหาวิทยาลัยสุโขทัยธรรมมาธิราช	http://www.stou.ac.th/study/sumrit/index.html
ศูนย์มะเร็งลำปาง	http://www.lampangcancer.com/lpccWebnew/index.php
ศูนย์โรคตับและปลูกถ่ายตับ โรงพยาบาลจุฬาลงกรณ์	http://www.chulalivercenter.com
สถานวิทยามะเร็งศิริราช	http://www.si.mahidol.ac.th/th/department/cancer/
สถาบันมะเร็งแห่งชาติ	http://www.nci.go.th
สมาคมความดันโลหิตสูงแห่งประเทศไทย	http://www.thaihypertension.org
สมาคมรังสีรักษาและมะเร็งวิทยาแห่งประเทศไทย	http://thastro.org
สาขารังสีรักษาและมะเร็งวิทยา ฝ่ายรังสีวิทยา โรงพยาบาลจุฬาลงกรณ์	http://www.chulacancer.net
สถาบันประสาทวิทยา	http://pni.go.th/pnigoth/?p=848
เครือข่ายการจัดการความรู้เพื่อสนับสนุน การดำเนินงานป้องกันควบคุมโรคไม่ติดต่อ โรงพยาบาลจุฬาลงกรณ์	http://kmncd.org/
โรงพยาบาลไทยนครินทร์	http://www.thainakarin.co.th
โรงพยาบาลปิยะเวท	http://www.piyavate.com
โรงพยาบาลมหาชัย	http://www.mahachaihospital.com
โรงพยาบาลรามคำแหง	http://www.ram-hosp.co.th
โรงพยาบาลลาดพร้าว	http://www.ladpraohospital.com
โรงพยาบาลวิภาวดี	http://www.vibhavadi.com

โรงพยาบาลสงขลา	http://www.sk-hospital.com
โรงพยาบาลสบ้าย้อย	http://www.sabayoihospital.go.th
โรงพยาบาลสูงเม่น	http://www.smhos.com
โรงพยาบาลหนองแสง	http://www.nongsaenghospital.go.th
โรงพยาบาลเวชธานี	http://www.vejthani.com

รายชื่อแหล่งข้อมูลเว็บไซต์ภาษาต่างประเทศ

ชื่อหน่วยงาน	ชื่อเรียกเว็บไซต์
AARP	http://www.aarp.org/health/Health_Tools/
Agency for Healthcare Research and Quality	http://www.ahrq.gov/
American Academy of Allergy Asthma & Immunology	http://www.aaaai.org/
American Academy of Family Physicians	http://familydoctor.org
American Cancer Society	http://www.cancer.org
American Heart Association	http://www.americanheart.org
American Liver Foundation	http://www.liverfoundation.org/
American Lung Association	http://www.lungusa.org
American Stroke Association	http://www.strokeassociation.org
Asthma UK	http://www.asthma.org.uk/
Breastcancer	http://www.breastcancer.org/
Centers for Disease Control and Prevention	http://www.cdc.gov/
Common Cold	http://www.commoncold.org/
EPA	http://www.epa.gov/
Food and Drug Administration	http://www.fda.gov
Heart Disease	http://www.heartd.org/
KidsHealth Comes From Nemours	http://kidshealth.org/
Liver cancer treatment	http://www.livercancertreatment.org/
Lungcancer.org	http://www.lungcancer.org
MedHelp	http://www.medhelp.org/
Memorial Sloan-Kettering Cancer Center	http://www.mskcc.org
National Cancer Institute	http://www.cancer.gov
National Heart, Lung, and Blood Institute	http://www.nhlbi.nih.gov
National Institute of Neurological Disorders and Stroke	http://www.ninds.nih.gov

National Jewish Health	http://www.nationaljewish.org/
National Stroke Association	http://www.stroke.org
NHS Choices	http://www.nhs.uk
Stroke Alliance For Europe	http://www.safestroke.org/
The Abramson Cancer Center of the University of Pennsylvania	http://www.oncolink.org
The Canadian Stroke Network's	http://www.canadianstrokenetwork.ca
The National Cervical Cancer Public Education Campaign	http://www.cervicalcancercampaign.org/
The National Institute of Allergy and Infectious Diseases (NIAID)	http://www3.niaid.nih.gov
The Ohio State University Medical Center	http://www.medicalcenter.osu.edu
The Stroke Association	http://www.stroke.org.uk
The University of Texas Medical Branch	http://www.utmb.edu
University of Chicago	http://www.uchicago.edu
University of Maryland Medical System	http://www.umm.edu
University of Michigan Health System	http://www.med.umich.edu/
University of Virginia Health System	http://www.healthsystem.virginia.edu

ภาคผนวก ค

รายการคำที่ผ่านการคัดเลือกโดยผู้เชี่ยวชาญทางการแพทย์

รายการคำกลุ่มโรคระบบหายใจ

กาฬโรค	อูดตัน	หนอง	เชื้อ
เบตาสเตรปโต	ก่อ	กิน	หลักเลี้ยง
ค็อกคัส	โพรง	หอบ	ไหล
ช่องคอ	กระดูก	ขนาด	จุมก
รูมาติก	เป็นไข	ทางเดินหายใจ	ไข
วิตามิน	โพรงอากาศ	แบคทีเรีย	น้ำ
Pneumonia	ละออง	ปวดศีรษะ	น้ำจุมก
ทอนซิล	สมอง	เยื่อ	บวม
ออกซิเจน	กล้ามเนื้อ	ไวรัส	ปอด
ต่อมน้ำเหลือง	ฝุ่น	สี	หอบหืด
ต่อมทอนซิล	โรคแทรกซ้อน	ควบคุม	คอ
หัวใจ	วิธีการ	เชื้อไวรัส	เด็ก
ช่องจุมก	กลืน	ติดต่อ	ร่างกาย
เนื้องอก	กำเริบ	ภูมิแพ้	หวัด
คอหอย	ไซนัส	วินิจฉัย	เรื้อรัง
นับปล้น	คิม	สัมผัส	อากาศ
ริดสีดวง	ประสิทธิภาพ	พัน	ป้องกัน
ตาย	ป่วย	โรคภูมิแพ้	หลอดลม
ลำคอ	โพรงจุมก	สูด	เสมหะ
ออกฤทธิ์	แดง	คัดจุมก	หายใจ
ปอดบวม	แทรกซ้อน	แพ้	ติดเชื้อ
ภูมิคุ้มกัน	ปวด	เขียว	ไอ
โรคปอด	เหลือง	เจ็บ	สาเหตุ
ลงคอ	พักผ่อน	เจ็บปล้น	รุนแรง
ส่อง	อ่อนเพลีย	ภาวะ	ยา
โรคหวัด	ไขหวัด	ยาปฏิชีวนะ	รักษา
โรคหัวใจ	จาม	ออกกำลังกาย	อักเสบ
หู	เชื้อโรค	รับประทาน	อาการ

acute	cold	hospital	rhinovirus
adult	common	hsv	risk
affect	condition	ill	runny
age	control	immune	sinus
agonist	corticosteroid	increase	smoke
air	cough	infect	sneeze
airway	culture	inflammation	sore
allergen	diagnose	influenza	spray
allergist	diphtheria	inhale	spread
allergy	disease	lung	strep
antibiotic	drug	lymph	streptococcus
aspirin	exercise	medical	surgery
asthma	exposure	medicine	swallow
attack	eye	mucus	swollen
avoid	face	nasal	symptom
bacteria	facial	neck	system
block	feel	nose	throat
breath	fever	oxygen	tonsil
care	first	penicillin	touch
catch	flow	pharynx	treatment
cause	flu	pneumonia	vaccine
cell	fluid	prevent	viral
chest	fungi	recommend	virus
child	hand	reduce	vitamin
children	headach	relieve	vomit
chronic	health	respiratory	warm
clinic	healthcare	result	wheeze

รายการคำกลุ่มโรคระบบไหลเวียนเลือด

กล้ามเนื้อ	ปกติ	สุขภาพ	cause
กล้ามเนื้อหัวใจ	ป้องกัน	สูบบุหรี่	check
กิน	ปัจจัย	เส้นเลือด	chest
เกลือ	ปัสสาวะ	เสี่ยง	cholesterol
แข็งตัว	ผิดปกติ	เสียชีวิต	clot
แขน	ผู้ป่วย	หน้าอก	control
ไขมัน	พฤติกรรม	หลอดเลือด	coronary
ควบคุม	พัก	หลอดเลือดแดง	damage
ความดัน	ยา	หายใจ	death
เค็ม	ขาดความดัน	อ้วน	diabetes
เครียด	ระดับ	ออกกำลัง	diastole
งด	รักษา	ออกซิเจน	die
เจ็บ	รับประทาน	อ่อนแรง	diet
ชา	ร่างกาย	อัมพฤกษ์	disease
ซีก	รุนแรง	อัมพาต	drug
เดิน	โรค	อาการ	emergency
ตรวจ	โรคความดัน	อายุ	exercise
ต่ำ	โลหิตสูง	อาหาร	factor
ตีบ	โรคหัวใจ	อุดตัน	failure
ตีบตัน	ลด	age	fat
เต็น	ลิ้ม	angioplasty	feel
ไต	เลือด	area	health
น้ำตาล	โลหิต	artery	healthy
น้ำหนัก	วิธี	aspirin	heart
แน่นหน้าอก	เวลา	attack	hemorrhage
บริเวณ	สมอง	bleed	higher
บีบ	สม่ำเสมอ	blood	hypertension
เบาหวาน	สาเหตุ	brain	include

increase	muscle	procedure	symptom
ischemic	oxygen	reduce	systole
kidney	pain	renal	therapy
level	patient	result	treatment
lifestyle	physic	risk	weak
low	plaque	salt	weight
lower	pressure	sign	
measure	prevent	smoke	
mmhg	problem	stroke	
women			

รายการคำกลุ่มเนื้องอก(รวมมะเร็ง)

HPV	น้ำดี	รังสี	หูด
กระจาย	น้ำเหลือง	รับประทาน	อาการ
กระดูก	เนื้องอก	เรื้อรัง	อายุ
ก้อน	บริเวณ	โรค	อาหาร
ขนาด	บำบัด	โรคตับ	เอกซเรย์
คลำ	บุหรี	โรคมะเร็ง	เอสโตรเจน
คว้นบุหรี	ประจำเดือน	ลักษณะ	แอลกอฮอล์
คัดกรอง	ป้องกัน	ลูกกลม	โอกาส
เคมี	ปอด	วัคซีน	ฮอร์โมน
ฉายแสง	ปัจจัย	วิธี	abnormal
ฉีด	ปากมดลูก	วิธีการ	adult
ช่องคลอด	แป๊ปสเมียร์	วินิจฉัย	age
ช่องท้อง	ผ่าตัด	เวลา	artery
ชิ้นเนื้อ	ผิดปกติ	ไวรัส	biopsy
เชื้อรา	ผิวหนัง	ไวรัสตับอักเสบบ	blood
ซ้ำ	พยาธิใบไม้	ไวรัสตับอักเสบบี	body
เซลล์	เพศสัมพันธ์	สตรี	breast
ตรวจ	แพร่กระจาย	ส่อง	cancer
ต่อมน้ำเหลือง	ภาวะ	สาเหตุ	carcinoma
ตัด	ภูมิคุ้มกัน	สุขภาพ	cause
ตับ	มดลูก	สุรา	cell
ตับอ่อน	มะเร็ง	สูบ	cervical
ติดเชื้อ	แมมโมแกรม	สูบบุหรี	cervix
ติดตาม	ยา	เสมหะ	change
เต้านม	ระดับ	เสียง	chemotherapy
ทรวงอก	รักแร้	หญิง	chest
ทางเดิน	รักษา	หน้าอก	cinch
นม	รังไข่	หลอดลม	develop
			diagnose

disease	liver	procedure	therapy
dose	lung	radiate	tissue
drug	lymph	ray	toxic
estrogen	mammogram	risk	transplant
factor	mastectomy	scan	treatment
gene	node	sexual	trial
genetics	nscl	small	tumor
health	organ	smoke	ultrasound
hepatocellular	pain	spread	uterus
hormone	pap	stage	vaccine
hpv	patient	surgery	vagina
hysterectomy	pelvic	survive	woman
infect	prevent	symptom	
inject	primary	technique	

ภาคผนวก ง

คู่มือการใช้งานโปรแกรมช่วยการสืบค้นข้อมูลทางการแพทย์

คู่มือการใช้งาน โปรแกรมช่วยการสืบค้นข้อมูลทางการแพทย์



ภาพที่ 20 แสดงหน้าจอหลักของระบบการสืบค้นข้อมูลทางการแพทย์

โปรแกรมช่วยการสืบค้นข้อมูลทางการแพทย์ แบ่งเป็น 3 ส่วนคือ

1. ส่วนของเมนูกลุ่มโรค
2. ส่วนของการระบุคำค้น
3. ส่วนสำหรับผู้เชี่ยวชาญ

1. ส่วนของเมนูกลุ่มโรค แบ่งออกเป็น 3 กลุ่มโรค คือ กลุ่มโรคระบบหายใจ กลุ่มโรคระบบไหลเวียนเลือด กลุ่มเนื้องอก ผู้ใช้ต้องเลือกกลุ่มโรคให้ตรงกับความต้องการ

2. ส่วนของการระบุคำค้น คือส่วนที่ผู้ใช้งานต้องระบุคำค้นที่ต้องการ ระบบจะนำคำค้นที่ผู้ใช้งานต้องการไปค้นหาคำที่ผู้ใส่ระบุ ซึ่งหากพบกฎที่มีคำค้นของผู้ใช้งานระบบจะแสดงกฎเพื่อให้ผู้ใช้งานเลือก จากนั้นทำการคลิกปุ่มค้นหา ระบบจะทำการส่งคำค้นไปยัง google และแสดงผลการสืบค้น

ตัวอย่างเช่น ผู้ใช้งานเลือกกลุ่มเนื้องอก(รวมมะเร็ง) จากนั้นระบุคำค้นว่า เต้านม แสดงดังภาพที่ 21

Web Images Videos Maps News Shopping Gmail more ▼

Google เต้านม รักษารึ Search

About 168,000 results (0.26 seconds) Advanced search

Everything More Show search tools

เต้านมเกิน หรือ นม 4 เต้า - Siriraj E-Public Library ☆ - [Translate this page]
7 ก.ย. 2009 ... หลายท่านอาจเคยได้ยินเรื่อง นม 4 เต้ามาก่อน คือสภาพหลังรับทานอาหารนมเต้านมอีก 2 เต้าเกิดขึ้นที่รักแร้ ในความเป็นจริงนั้น อาจเป็นมากกว่า 4 ...
www.si.mahidol.ac.th/sidoctor/e-pl/article/detail.asp?id=309 - Cached - Similar

เมื่อฉันมีนม 4 เต้า - | Ya & You ยักษ์กับคุณ | สืบค้นไว้ อันใจเมื่อใช้ขุ ☆ - [Translate this page]
เต้านมรักแร้? (axillary breast). ลักษณะและอาการของเต้านมเกิน ... การพบก้อนที่รักแร้ จำเป็นต้องแยกให้ออกว่าเป็นเนื้อเต้านมเกินหรือเป็นภาวะที่ต่อมน้ำเหลืองที่ ...
www.yaandyou.net/druglist2.php?...เมื่อฉันมีนม%204%20เต้า... - Cached

มะเร็งเต้านม - วิกิพีเดีย ☆ - [Translate this page]
มะเร็งเต้านม (Breast cancer) เป็นมะเร็งที่พบบ่อยในผู้หญิงไทย ... นิวซี นิวกลาง และนิวเวง) คำศัพท์เต้านมและรักแร้ ที่สำคัญอีกห้าคำกับเต้านม ...
th.wikipedia.org/wiki/มะเร็งเต้านม - Cached - Similar

การผ่าตัดมะเร็งเต้านมแบบ Sentinel node โดยไม่ต้องเลาะต่อมน้ำเหลือง ... ☆ - [Translate this page]
การรักษา มะเร็งเต้านมที่ดีที่สุด คือ การผ่าตัดเต้านมพร้อมเลาะต่อมน้ำเหลืองที่รักแร้ออก เพราะ มะเร็งเต้านมจะมีการแพร่กระจายไปยังต่อมน้ำเหลืองที่รักแร้ ...
guru.sanook.com > ... > พบพจนานุกรม - Cached - Similar

โรคมะเร็ง | มะเร็งเต้านม ☆ - 2 visits - 11/9/09 - [Translate this page]
คล้ายกับเต้านมหรือที่รักแร้, มีการเปลี่ยนแปลงของขนาดเต้านม; มีไข้ไหลออกจากหัวนม หรือเจ็บหัวนมถูกสิ่งรบกวนในเต้านม; ผิวที่เต้านมจะมีลักษณะเหมือนเปลือก ...
www.siamhealth.net/Disease/cancer/breastcancer.htm - Cached - Similar

มีก้อนใต้รักแร้ - คนยอมรับมาตีปาดับลู ☆ - [Translate this page]
16 พ.ย. 2006 ... ก้อนใต้รักแร้อาจเป็น 1.ต่อมน้ำเหลืองในระยะกำลังอักเสบ 2.นมเต้าที่ตาม 3.ต่อมน้ำเหลืองจากมะเร็ง ชนิดแรก ที่สำคัญเป็นก้อนอยู่ข้างใต้รักแร้หนึ่ง ...
www.balaw.com/webboard/qa/view.asp?id=6962 - Cached - Similar

ภาพที่ 23 แสดงผลลัพธ์การสืบค้นคำว่า เต้านม และ รักแร้

3. ส่วนสำหรับผู้เชี่ยวชาญ

ส่วนนี้เป็นส่วนสำหรับผู้เชี่ยวชาญในการเพิ่มกฎความสัมพันธ์ให้กับระบบ เมื่อผู้เชี่ยวชาญเลือกเมนูนี้ จะขึ้นหน้าลือคอินมาให้ ดังภาพที่ 24

การพัฒนาการสืบค้นข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์
Development of Medical Search Using Association Rules

หน้าหลัก
กลุ่มโรคระบบหัวใจ
กลุ่มโรคระบบไหลเวียนเลือด
กลุ่มเนื้อเยื่อ (รวมมะเร็ง)
สำหรับผู้เชี่ยวชาญ

กรุณา Login ก่อนเข้าระบบทุกครั้ง

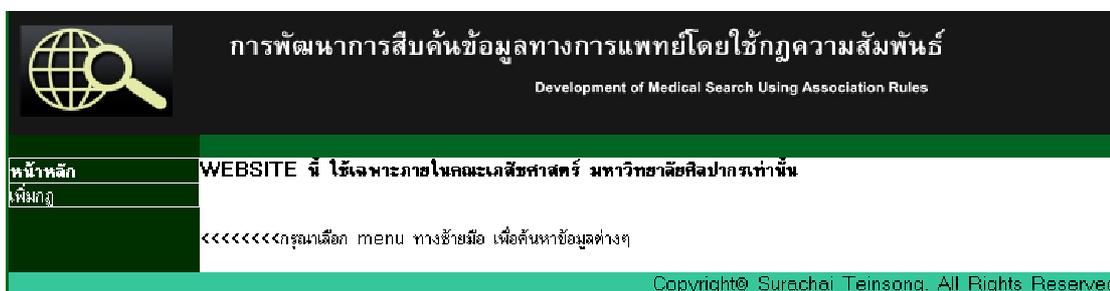
username :

password :

Copyright© Surachai Teinsong. All Rights Reserved

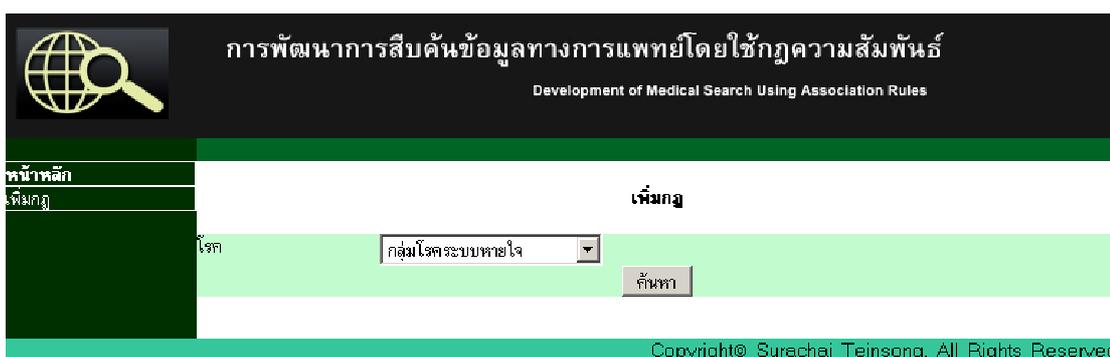
ภาพที่ 24 แสดงหน้าจอเข้าระบบสำหรับผู้เชี่ยวชาญ

เมื่อผู้เชี่ยวชาญเข้าสู่ระบบแล้ว จะปรากฏหน้าจอดังภาพที่ 25



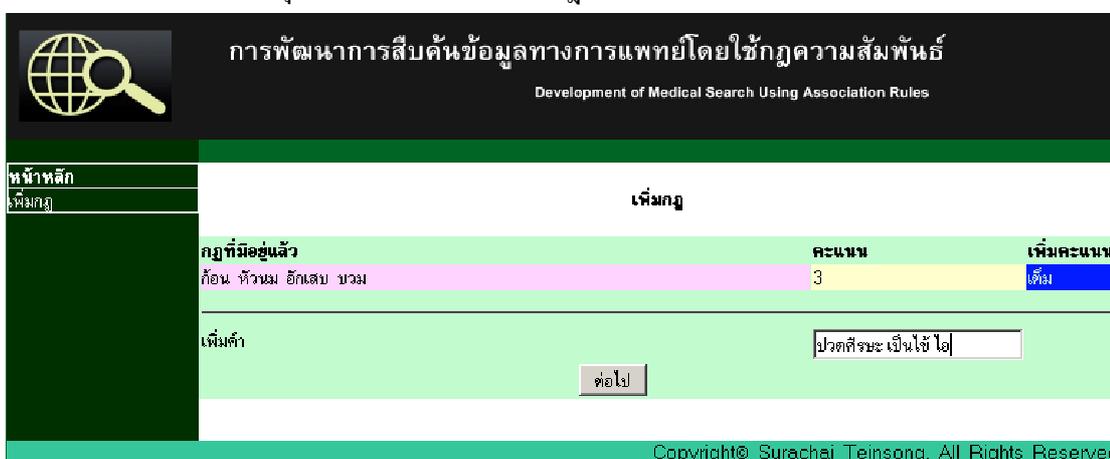
ภาพที่ 25 แสดงหน้าจอสำหรับเพิ่มกฎโดยผู้เชี่ยวชาญ

จากภาพที่ 25 ผู้เชี่ยวชาญคลิกเมนู เพิ่มกฎ จะปรากฏหน้าจอสำหรับเลือกกลุ่มโรคที่ต้องการเพิ่มกฎดังภาพที่ 26



ภาพที่ 26 แสดงหน้าจอสำหรับผู้เชี่ยวชาญเพิ่มกฎ โดยเลือกกลุ่มโรคที่ต้องการ

จากนั้นคลิกปุ่มค้นหา จากนั้นจะปรากฏหน้าจอดังภาพที่ 27



ภาพที่ 27 แสดงหน้าจอการเพิ่มกฎสำหรับผู้เชี่ยวชาญ

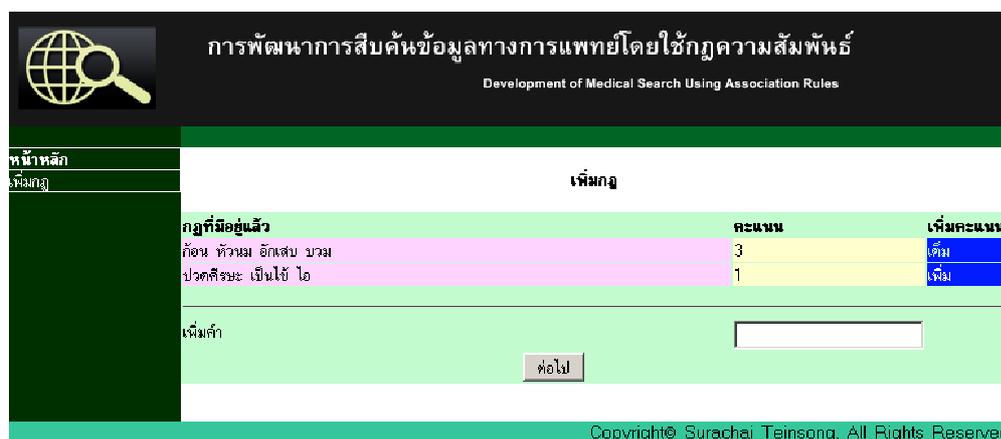
จากภาพที่ 27 แสดงหน้าจอการเพิ่มกฎสำหรับผู้เชี่ยวชาญ ผู้เชี่ยวชาญสามารถเพิ่มคะแนนให้กับกฎที่มีอยู่แล้ว โดยกฎดังกล่าวจะสามารถให้ผู้ใช้งานได้จะต้องมีคะแนนเท่ากับ 3 คะแนน โดยสามารถคลิกปุ่ม เพิ่มคะแนน (สีน้ำเงิน) แต่หากคะแนนเท่ากับ 3 คะแนนแล้ว ปุ่มดังกล่าวจะเปลี่ยนเป็นคำว่า เต็ม และไม่สามารถคลิกได้

ผู้เชี่ยวชาญสามารถสร้างกฎใหม่ขึ้นมาโดยการกรอกค่าลงไปในช่วง เพิ่มค่า จากตัวอย่างทดสอบโดยการเพิ่มคำว่า ปวดศีรษะ เป็นไข้ ไอ จากนั้นคลิกปุ่ม ต่อไป จะปรากฏหน้าจอดังภาพที่ 28



ภาพที่ 28 แสดงหน้าจอเสร็จสิ้นการเพิ่มกฎความสัมพันธ์ของค่าโดยผู้เชี่ยวชาญ

จากนั้นกฎดังกล่าวจะยังไม่สามารถให้ผู้ใช้งานนำไปใช้งานได้ เพราะมีคะแนนไม่ถึง 3 คะแนน โดยสามารถเข้าไปตรวจสอบได้โดยการคลิกที่เมนู เพิ่มกฎ จากนั้นเลือกกลุ่มโรคที่ต้องการจะปรากฏกฎที่ยังมีคะแนนไม่ถึง 3 คะแนน ดังภาพที่ 29



ภาพที่ 29 แสดงกฎที่เพิ่มใหม่โดยมีคะแนนไม่ถึง 3 คะแนน

จากภาพที่ 29 เมื่อทำการเพิ่มคะแนนให้กฎ ปวดศีรษะ เป็นไข้ ไอ ดังกล่าวครบ 3 คะแนนแล้ว กฎดังกล่าวผู้ใช้งานจะสามารถนำไปใช้งานได้ ดังแสดงดังภาพที่ 30

การพัฒนาระบบสืบค้นข้อมูลทางการแพทย์โดยใช้กฎความสัมพันธ์
Development of Medical Search Using Association Rules

หน้าหลัก | | ค้นหา

คลังโรคระบบหัวใจ
 คลังโรคระบบไหลเวียนเลือด
 คลังเชิงลึก(รวมหนังสือ)
 ดัชนีผู้เชี่ยวชาญ

Copyright© Surachai Teinsong. All Rights Reserved

ภาพที่ 30 แสดงกฎความสัมพันธ์ของคำที่ผู้เชี่ยวชาญเพิ่มเติม

ภาคผนวก จ
ตัวอย่างกฎความสัมพันธ์

ตัวอย่างกฎความสัมพันธ์

ไข่ <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 คอ <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 ยา <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 อักเสบ <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 รักษา <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 อาการ <- ต่อม้ำเหลือง (10.7%/9, 100.0%)
 ควบคุม <- ออกฤทธิ์ (14.3%/12, 92.3%)
 หอบหืด <- ออกฤทธิ์ (15.5%/13, 100.0%)
 หลอดลม <- ออกฤทธิ์ (15.5%/13, 100.0%)
 หายใจ <- ออกฤทธิ์ (13.1%/11, 84.6%)
 รุนแรง <- ออกฤทธิ์ (14.3%/12, 92.3%)
 ยา <- ออกฤทธิ์ (15.5%/13, 100.0%)
 แดง <- ต่อมทอนซิล (10.7%/9, 90.0%)
 หนอง <- ต่อมทอนซิล (10.7%/9, 90.0%)
 แบคทีเรีย <- ต่อมทอนซิล (11.9%/10, 100.0%)
 เจ็บ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 ยาปฏิชีวนะ <- ต่อมทอนซิล (10.7%/9, 90.0%)
 เชื้อ <- ต่อมทอนซิล (10.7%/9, 90.0%)
 ไข่ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 คอ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 ป้อนกัน <- ต่อมทอนซิล (10.7%/9, 90.0%)
 สาเหตุ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 ยา <- ต่อมทอนซิล (10.7%/9, 90.0%)
 อักเสบ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 รักษา <- ต่อมทอนซิล (11.9%/10, 100.0%)
 อาการ <- ต่อมทอนซิล (11.9%/10, 100.0%)
 ริดสีดวง <- เนื้องอก (11.9%/10, 90.9%)
 เนื้องอก <- ริดสีดวง (11.9%/10, 83.3%)

ต่อมทอนซิล <- แดง หนอง (10.7%/9, 85.4%)
 ต่อมทอนซิล <- แดง แบททีเรีย (10.7%/9, 82.1%)
 ต่อมทอนซิล <- หนอง แบททีเรีย (10.7%/9, 81.0%)
 ต่อมทอนซิล <- แบททีเรีย คอ (11.9%/10, 81.2%)
 เสมหะ <- ไอ อักเสบ (48.8%/41, 85.7%)
 ไอ <- ติดเชื้อ ยา (50.0%/42, 84.9%)
 ไอ <- ติดเชื้อ อักเสบ (52.4%/44, 86.8%)
 ไข้ <- เสมหะ ติดเชื้อ (39.3%/33, 84.8%)
 ไข้ <- เสมหะ อักเสบ (39.3%/33, 81.7%)
 ไข้ <- เสมหะ รักษา (39.3%/33, 81.2%)
 ไข้ <- เสมหะ อากา (41.7%/35, 83.3%)
 ไข้ <- สาเหตุ ไอ (36.9%/31, 80.5%)
 ไข้ <- สาเหตุ ยา (39.3%/33, 80.2%)
 ไข้ <- สาเหตุ อักเสบ (41.7%/35, 80.8%)
 ไข้ <- สาเหตุ รักษา (42.9%/36, 84.3%)
 ไข้ <- ไอ ติดเชื้อ (40.5%/34, 85.1%)
 ไข้ <- ไอ อักเสบ (41.7%/35, 82.4%)
 ไข้ <- ไอ รักษา (41.7%/35, 80.8%)
 บวม <- ไข้ เสมหะ (33.3%/28, 81.1%)
 บวม <- ไข้ ไอ (33.3%/28, 80.9%)
 บวม <- ไข้ ติดเชื้อ (33.3%/28, 81.6%)
 บวม <- ไข้ อักเสบ (34.5%/29, 81.2%)
 บวม <- ไข้ รักษา (35.7%/30, 82.1%)
 บวม <- ไข้ อากา (35.7%/30, 80.4%)
 บวม <- คอ อากา (31.0%/26, 81.6%)
 บวม <- ป้องกัน เสมหะ (32.1%/27, 80.0%)
 เรื้อรัง <- ไอ ติดเชื้อ (39.3%/33, 82.0%)
 เรื้อรัง <- ไอ ยา (45.2%/38, 83.3%)
 เรื้อรัง <- ไอ อักเสบ (44.0%/37, 82.0%)
 เรื้อรัง <- ยา อักเสบ (50.0%/42, 85.2%)

symptom <- attack (22.4%/34, 94.4%)
cause <- attack (21.1%/32, 88.9%)
infect <- flu (20.4%/31, 93.9%)
nose <- eye (21.1%/32, 88.9%)
cold <- eye (20.4%/31, 86.1%)
common <- eye (20.4%/31, 86.1%)
symptom <- eye (20.4%/31, 86.1%)
infect <- eye (23.0%/35, 97.2%)
nose <- runny (21.7%/33, 97.1%)
cold <- runny (21.1%/32, 94.1%)
symptom <- runny (21.1%/32, 94.1%)
infect <- runny (21.7%/33, 97.1%)
treatment <- hospital (20.4%/31, 93.9%)
cause <- hospital (20.4%/31, 93.9%)
infect <- spread (22.4%/34, 91.9%)
cause <- spread (21.1%/32, 86.5%)
symptom <- relieve (21.1%/32, 100.0%)
treatment <- diagnose (20.4%/31, 86.1%)
symptom <- diagnose (23.7%/36, 100.0%)
cause <- diagnose (23.7%/36, 100.0%)
nose <- sneeze (23.7%/36, 94.7%)
cold <- sneeze (22.4%/34, 89.5%)
cough <- sneeze (23.0%/35, 92.1%)
infect <- sneeze (23.7%/36, 94.7%)
cause <- sneeze (20.4%/31, 81.6%)
infect <- hand (20.4%/31, 88.6%)
cause <- hand (20.4%/31, 88.6%)
air <- airway (22.4%/34, 81.0%)
breath <- airway (24.3%/37, 88.1%)
treatment <- airway (23.0%/35, 83.3%)

allergist <- allergen allergy (10.5%/16, 84.5%)
allergist <- allergen medical (11.2%/17, 84.4%)
allergist <- allergen treatment (11.2%/17, 81.0%)
allergist <- allergen symptom (11.2%/17, 81.0%)
allergist <- allergy asthma (10.5%/16, 81.9%)
allergist <- allergy medical (10.5%/16, 80.7%)
rhinovirus <- virus cold (11.8%/18, 80.3%)
neck <- fever throat (11.8%/18, 81.3%)
inflammatory <- drug common (11.8%/18, 84.1%)
inflammatory <- drug treatment (11.8%/18, 80.1%)
inflammatory <- drug symptom (11.8%/18, 82.5%)
inflammatory <- drug cause (11.8%/18, 82.0%)
oxygen <- pneumonia chest (11.2%/17, 80.8%)
oxygen <- chest bacteria (10.5%/16, 81.3%)
decongest <- recommend nasal (11.2%/17, 81.6%)
decongest <- nasal treatment (13.2%/20, 80.2%)
culture <- sore antibiotic (11.8%/18, 80.8%)
exercise <- avoid smoke (13.2%/20, 80.3%)
vaccine <- hand prevent (11.8%/18, 80.1%)
vaccine <- pneumonia recommend (10.5%/16, 80.4%)
acute <- sinus chronic (12.5%/19, 80.7%)
attack <- airway wheeze (12.5%/19, 81.6%)
attack <- wheeze chest (13.2%/20, 80.6%)
warm <- hand sore (10.5%/16, 84.7%)
warm <- headach nose (10.5%/16, 80.0%)
warm <- headach cold (11.2%/17, 80.1%)
warm <- sore ill (11.8%/18, 86.2%)
warm <- sore avoid (11.2%/17, 81.1%)
warm <- sore risk (13.2%/20, 86.4%)
warm <- sore fever (13.8%/21, 81.3%)

ประวัติผู้วิจัย

ชื่อ – สกุล	นายสุรชัย เทียนสง
ที่อยู่	25/7 หมู่ 4 ต.บ้านใหม่ อ.บางใหญ่ จ.นนทบุรี 11140
ประวัติการศึกษา	
พ.ศ. 2541	สำเร็จการศึกษาปริญญาครุศาสตรบัณฑิต สาขาวิชาคอมพิวเตอร์ศึกษา จากคณะครุศาสตร์ มหาวิทยาลัยราชภัฏนครปฐม
พ.ศ. 2548	ศึกษาต่อระดับปริญญาโท สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร
ประวัติการทำงาน	
พ.ศ. 2527	บริษัท รอยัลอินครัสตรีส์ (ประเทศไทย) จำกัด
พ.ศ. 2540 – ปัจจุบัน	คณะเกษตรศาสตร์ มหาวิทยาลัยศิลปากร